

FILOGENOMSKA ANALIZA RNA VIRUSOV PRI NEVRETENČARJIH

Sabina Ott Rutar

Doktorska disertacija
Mednarodna podiplomska šola Jožefa Stefana
Ljubljana, Slovenija

Mentor: prof. dr. Dušan Kordiš, Institut »Jožef Stefan«, Ljubljana, Slovenija

Komisija za oceno doktorske disertacije:

prof. dr. Igor Križaj, predsednik, Institut »Jožef Stefan«, Ljubljana, Slovenija

prof. dr. Boris Rogelj, član, UL, Fakulteta za kemijo in kemijsko tehnologijo, Ljubljana, Slovenija

akad. prof. dr. Tatjana Avšič-Županc, član, UL, Medicinska fakulteta, Ljubljana, Slovenija

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL



Sabina Ott Rutar

FILOGENOMSKA ANALIZA RNA VIRUSOV PRI
NEVREtenčARJIH

Doktorska disertacija

PHYLOGENOMIC ANALYSIS OF RNA VIRUSES IN
INVERTEBRATES

Doctoral Dissertation

Mentor: prof. dr. Dušan Kordiš

Ljubljana, Slovenija, oktober 2018

Zahvala

Doktorat je pot, ki se ne more prehoditi brez sodelovanja, vztrajnosti, potrpežljivosti in podpore. Doktorsko delo je ekipno delo, s katerim zamisli dobijo realno obliko. Zato gre moja prva zahvala mentorju prof. dr. Dušanu Kordišu za vso pomoč, potrpežljivost in izjemno vodenje pri nastajanju dela.

Iskrena zahvala gre tudi prof. dr. Igorju Križaju (predsedniku komisije), prof. dr. Borisu Roglju in akad. prof. dr. Tatjani Avšič-Županc za strokovno obravnavo in kritično oceno mojega dela.

Seveda uspeh ne bi bil možen brez podpore domačih in najbližjih: mame Lilijane, očeta Erika, brata Erika in moža Grega. Zato bi se jim na tem mestu iz srca zahvalila za vso motivacijo, razumevanje in vzpodbudne besede.

Nenazadnje gre zahvala tudi vsem sodelavcem z Odseka za molekularne in biomedicinske znanosti (B2) za vso tehnično pomoč in sodelovanje tekom mojega doktorata.

Povzetek

RNA virusi so največja skupina virusov. Med njimi so številni povzročitelji bolezni pri ljudeh, živalih in rastlinah. Nekatere skupine žuželk in gliste so znani vektorji rastlinskih in živalskih RNA virusov. **Najnovejše analize raznolikosti RNA virusov pri nevretenčarjih so dokazale, da skupina Protostomia predstavlja center raznolikosti RNA virusov.** Številne skupine nevretenčarjev so bile doslej slabo analizirane z vidika raznolikosti in razširjenosti RNA virusov, in sicer: ploščati črvi, kožekrilci, metulji, postrance, ceponožci, najstarejše skupine žuželk in najstarejše skupine živali. Namen mojega doktorskega dela je bil analizirati RNA virome zgoraj omenjenih skupin nevretenčarjev; dokazati, da je raznolikost RNA virusov prisotna že pri najstarejših skupinah živali; dokazati, da je raznolikost RNA virusov ohranjena čez celotne nevretenčarje; ter pojasniti evolucijsko in genomsko dinamiko RNA virusnih skupin. Da bi pojasnila razširjenost, evolucijo, genomsko organizacijo in endogenizacijo RNA virusov, sem naredila obsežno filogenomsko analizo RNA virusov v številnih transkriptomih in genomih zgoraj omenjenih skupin nevretenčarjev. V bogatih transkriptomskih podatkovnih bazah sem pridobila celotne genome različnih predstavnikov ssRNA(-), ssRNA(+) in dsRNA virusov. V genomih zgoraj omenjenih skupin nevretenčarjev sem analizirala prisotnost celotnih oz. okvarjenih RNA virusov, kar je pokazalo na pretekle infekcije teh organizmov z različnimi RNA virusi. Z obsežno filogenetsko analizo sem klasificirala nove viruse in analizirala evolucijo RNA virusov pri zgoraj omenjenih skupinah nevretenčarjev.

Našla sem 815 novih RNA virusov in endogeniziranih RNA virusov, ki so spadali v 22 od 25 definiranih klad. RNA virusne klade, ki so prisotne pri vseh skupinah nevretenčarjev, so: Hepe-Virga, Narna-Levi, Mono-Chu, Orthomyxo, Partiti-Picobirna, Picorna-Calici, Tombus-Noda and Toti-Chryso. **Najbolj zastopane virusne klade pri analiziranih nevretenčarjih so: Mono-Chu, Partiti-Picobirna in Picorna-Calici.** Znotraj klade Picorna-Calici sem ugotovila, da so najbolj zastopane skupine pikornavirusov dicistrovirusi, vodni pikornavirusi in iflavirusi; da so dicistrovirusi in iflavirusi prisotni pri kopenskih in vodnih nevretenčarjih; vodni pikornavirusi pa so omejeni na vodne nevretenčarje. S pomočjo **filogenetske analize sem dokazala, da so številni novi RNA virusi znotraj posamezne klade oblikovali nove skupine.** Našla sem številne raznolike strukture RNA virusnih genomov, ki so se pojavile že pri predniku živali in so se ohranile pri vseh nevretenčarjih. Najboljši primer za to predstavljajo pikornavirusi. Ugotovila sem, da je večina genomskih struktur novih RNA virusov podobna genomom znanih virusov, ki spadajo v isto skupino. Našla sem tudi najstarejše predstavnike številnih RNA virusnih klad v transkriptomih najstarejših skupin živali z drugačnimi strukturami virusnih genomov. V analiziranih genomih sem našla endogenizirane predstavnike 12 RNA virusnih klad. Ugotovila sem, da je samo klada Mono-Chu endogenizirana pri vseh analiziranih skupinah nevretenčarjev. V doktorskem delu sem dokazala, da so se prvi predstavniki RNA virusnih klad pojavili že pri najstarejših skupinah živali; da se je že pri predniku živali pojavila »eksplozija« **raznolikih genomskih struktur RNA virusov; da je center raznolikosti RNA virusov že pri najstarejših skupinah živali; da je raznolikost RNA virusov ohranjena čez celotno skupino nevretenčarjev.**

Abstract

RNA viruses are the largest group of viruses. Among them are some of the most notorious human, animal and plant pathogens. Many insects and nematodes are vectors of plant and animal RNA viruses. **The latest studies of RNA viral diversity in invertebrates have demonstrated that the Protostomia lineage represents the center of RNA viral diversity.** RNA viruses in numerous invertebrate groups have not yet been thoroughly analysed, such as in platyhelminths, hymenopterans, lepidopterans, amphipods, copepods, the oldest hexapod lineages and the oldest animal lineages. The aim of my thesis was to analyse the RNA viromes of the above-mentioned invertebrate lineages; to demonstrate that the »explosion« of the RNA viral diversity was present in the oldest metazoan lineages; to demonstrate that the diversity of the RNA viromes is conserved in all invertebrate lineages; and to infer the genome and evolutionary dynamics of RNA viral clades. An extensive phylogenomic analysis of RNA viruses in numerous transcriptomic and genomic libraries of the above-mentioned invertebrate groups has been used to explore the distribution, evolution, genomic organization and endogenization of RNA viruses. The complete genomes of diverse representatives of the ssRNA (-), ssRNA (+), and dsRNA viruses have been found in rich transcriptome databases. The analysis of endogenized RNA viruses in the genomes of the above-mentioned invertebrate groups has shown the past infections of these organisms with various RNA viruses. By extensive phylogenetic analysis, I classified new RNA viruses and analyzed the evolution of RNA viruses in the above-mentioned invertebrate groups.

815 novel RNA viruses have been found and belong to the 22 out of 25 RNA viral clades. The following RNA viral clades were found in all analysed invertebrate groups: Hepe-Virga, Narna-Levi, Mono-Chu, Orthomyxo, Partiti-Picobirna, Picorna-Calici, Tombus-Noda and Toti-Chryso. **The most abundant groups among ss(-)RNA, dsRNA, and ss(+)RNA viruses are: Mono-Chu, Partiti-Picobirna and Picorna-Calici.** In the Picorna-Calici clade dicistroviruses, iflaviruses and aquatic picornaviruses are the most abundant. Dicistroviruses and iflaviruses were found in terrestrial and aquatic invertebrates, while aquatic picornaviruses are restricted to aquatic invertebrates. **Phylogenetic analysis has shown that a number of novel RNA viruses formed new groups inside the defined viral clades. A rich repertoire of diverse RNA viral genomes has been found, many of them originated in the oldest metazoan lineages and remained conserved in all invertebrate groups. This can be most clearly seen in picornaviruses. The majority of the viral genome structures are similar to the known viruses of the same group. Complete genomes of the numerous oldest representatives of diverse RNA viral clades have been found in basal metazoans, they show also diverse characteristics of their genome structure.** Representatives of the 12 RNA viral clades have been found endogenized in invertebrate genomes. Mono-Chu is the only clade that is endogenized in all analysed invertebrate phyla. **In my thesis, I demonstrated that the first representatives of RNA viral clades originated in metazoan ancestors; the explosion of diverse RNA viral genomes occurred in the metazoan ancestor; the center of RNA viral diversity has been shifted from Protostomia to the oldest metazoan lineages; and RNA viral diversity is conserved in all invertebrate lineages.**

Kazalo

Slike	xvii
Tabele	xxviii
Seznam dodatkov in slik v okviru rezultatov	xxix
Kratice	xxx
Simboli	xxxiii
Slovar	xxxv
1 Uvod	1
1.1 Virusi	1
1.1.1 Definicija virusov	1
1.1.2 Genetske in evlucijske prednosti virusov	2
1.1.2.1 Številčnost populacije	2
1.1.2.2 Hitrost mutacij	2
1.1.2.3 Način razmnoževanja	3
1.1.2.4 Raznolikost virosfere	3
1.1.2.5 Prilagodljivost (angl. » <i>adaptability</i> «) in sposobnost razvijanja (angl. » <i>evolvability</i> «)	5
1.1.3 Klasifikacija virusov	5
1.1.3.1 Sistemi klasifikacije	5
1.1.3.2 Lastnosti virusnih skupin	8
1.1.3.2.1 dsDNA	8
1.1.3.2.2 ssDNA	9
1.1.3.2.3 ss(+)RNA	9
1.1.3.2.4 ss(-)RNA	9
1.1.3.2.5 dsRNA	9
1.1.3.2.6 ssRNA – RT	9
1.1.3.2.7 dsDNA – RT	10
1.1.3.3 Izzivi pri klasifikaciji in karakterizaciji virusov	10
1.1.4 Izvor virusov	11
1.1.5 Razširjenost virusov pri bakterijah, arhejah in evkariontih	13
1.1.6 Vloga virusov	15
1.1.6.1 Okoljski vidik	15
1.1.6.2 Medicinski vidik	16
1.1.6.3 Agrikulturni vidik	17
1.1.6.4 Vektorji rastlinskih in živalskih patogenov	18
1.2 RNA virusi	20
1.2.1 Strukturna raznolikost RNA virusov	20
1.2.1.1 Klada Astro	20
1.2.1.2 Klada Birna	21
1.2.1.3 Klada Bunya-Arena	21
1.2.1.4 Klada Hepe-Virga	22
1.2.1.5 Klada Hypo	23
1.2.1.6 Klada Luteo-Sobemo	24
1.2.1.7 Klada Narna-Levi	25
1.2.1.8 Klada Mono-Chu	27

1.2.1.9	Klada Orthomyxo	28
1.2.1.10	Klada Nido	28
1.2.1.11	Klada Permutotetra	30
1.2.1.12	Klada Partiti-Picobirna	30
1.2.1.13	Klada Picorna-Calici	31
1.2.1.14	Klada Reo	35
1.2.1.15	Klada Tombus-Noda	35
1.2.1.16	Klada Toti-Chryso	37
1.2.1.17	Klada Ophio	38
1.2.1.18	Klada Flavi	38
1.2.1.19	Klada Cysto	40
1.2.1.20	Klada Poty	40
1.2.1.21	Weivirus	41
1.2.1.22	Yuevirus41	
1.2.1.23	Yanvirus42	
1.2.1.24	Qinvirus 42	
1.2.1.25	Zhaovirus	42
1.2.2	Od RNA odvisna RNA polimeraza (RdRp)	42
1.2.3	Razmnoževanje RNA virusov	45
1.2.3.1	Prepoznavanje gostiteljske celice in pritrditev virusa	45
1.2.3.2	Vstop v gostiteljsko celico	46
1.2.3.3	Odstranjevanje virusne ovojnice in/ali kapside	47
1.2.3.4	Podvojevanje in prepisovanje	49
1.2.3.5	Sestavljanje in zorenje virusa	53
1.2.3.6	Izstop virusa iz celice	53
1.2.4	Endogenizacija RNA virusov in primerjava z retroelementi	54
1.3	Nevretenčarji	57
1.3.1	Skupine nevretenčarjev	57
1.4	Pregled raznolikosti in razširjenosti RNA virusov pri nevretenčarjih	60
1.4.1	Pregled	60
1.4.2	Preostala odprta vprašanja	64
2	Namen dela	67
2.1	Namen	67
2.1.1	Cilji in utemeljitev znanstvene relevantnosti	67
2.1.2	Hipoteze	68
2.1.3	Metodologija	68
2.2	Izzivi	69
2.2.1	Izzivi, povezani s podatkovnimi bazami	69
2.2.2	Izzivi, povezani z analizo strukture genoma	70
2.2.3	Izzivi pri filogenetski analizi	70
3	Metode dela	71
3.1	Pridobivanje podatkov	71
3.2	Iterativno iskanje homolognih zaporedij (RNA virusov) v transkriptomskih bazah podatkov nevretenčarjev in endogeniziranih RNA virusov v genomskih bazah podatkov nevretenčarjev	72
3.2.1	Iterativno iskanje homolognih zaporedij v transkriptomskih podatkovnih bazah nevretenčarjev	72
3.2.2	Iterativno iskanje homolognih zaporedij v genomskih podatkovnih bazah nevretenčarjev	74
3.3	Filogenetska analiza RNA virusov in endogeniziranih RNA virusov	77
3.3.1	Pridobivanje podatka o virusni kladi	77
3.3.2	Filogenetska analiza	77
3.3.2.1	Izbor zaporedij za filogenetsko analizo	77
3.3.2.2	Poravnava zaporedij	78
3.3.2.3	Izbor optimalnega modela	78
3.3.2.4	Izris filogenetskega drevesa	79
3.3.2.5	Ocena filogenetskega drevesa	79
3.4	Primerjalna analiza genomske organizacije RNA virusov pri raziskanih skupinah nevretenčarjev	79

4 Rezultati	81
4.1 Analiza zastopanosti virusnih klad.....	81
4.1.1 Analiza zastopanosti posameznih virusnih klad, kot jih je definirala skupina Shi in sod. (2016a).....	81
4.1.1.1 Rezultati Shi in sod. (2016a).....	82
4.1.1.2 Rezultati, nastali v okviru doktorata.....	84
4.1.2 Analiza zastopanosti virusnih skupin znotraj virusne klade Picorna-Calici.....	89
4.1.2.1 Rezultati Shi in sod. (2016a).....	89
4.1.2.2 Rezultati, nastali v okviru doktorata.....	90
4.2 Endogenizirani RNA virusi v genomih nevretenčarjev.....	92
4.3 Filogenetska analiza novih RNA virusov in endogeniziranih RNA virusov pri raziskanih skupinah nevretenčarjev.....	95
4.3.1 Kopenski nevretenčarji.....	95
4.3.1.1 Kožekrilci (Hymenoptera).....	95
4.3.1.1.1 Klada Bunya-Arena.....	95
4.3.1.1.2 Klada Flavi.....	96
4.3.1.1.3 Klada Hepe-Virga.....	96
4.3.1.1.4 Klada Luteo-Sobemo.....	99
4.3.1.1.5 Klada Mono-Chu.....	99
4.3.1.1.6 Klada Narna-Levi.....	100
4.3.1.1.7 Klada Tombus-Noda.....	101
4.3.1.1.8 Klada Ophio.....	101
4.3.1.1.9 Klada Orthomyxo.....	101
4.3.1.1.10 Klada Partiti-Picobirna.....	101
4.3.1.1.11 Klada Permutotetra.....	105
4.3.1.1.12 Klada Picorna-Calici.....	105
4.3.1.1.13 Klada Qinvirus.....	106
4.3.1.1.14 Klada Reoviridae.....	106
4.3.1.1.15 Klada Toti-Chryso.....	106
4.3.1.2 Metulji (Lepidoptera).....	107
4.3.1.2.1 Klada Bunya-Arena.....	107
4.3.1.2.2 Klada Astro.....	107
4.3.1.2.3 Klada Flavi.....	107
4.3.1.2.4 Klada Hepe-Virga.....	108
4.3.1.2.5 Klada Luteo-Sobemo.....	108
4.3.1.2.6 Klada Mono-Chu.....	108
4.3.1.2.7 Klada Narna-Levi.....	109
4.3.1.2.8 Klada Tombus-Noda.....	109
4.3.1.2.9 Klada Orthomyxo.....	110
4.3.1.2.10 Klada Partiti-Picobirna.....	110
4.3.1.2.11 Klada Permutotetra.....	112
4.3.1.2.12 Klada Picorna-Calici.....	112
4.3.1.2.13 Klada Reo.....	114
4.3.1.2.14 Klada Toti-Chryso.....	114
4.3.1.3 Najstarejše skupine žuželk: skakači (Collembola), dvorepke (Diplura), ščetinorepke (Zygentoma), Monocondylia.....	114
4.3.1.3.1 Klada Mono-Chu.....	114
4.3.1.3.2 Klada Narna-Levi.....	114
4.3.1.3.3 Klada Hepe-Virga.....	114
4.3.1.3.4 Klada Orthomyxo.....	115
4.3.1.3.5 Klada Picorna-Calici.....	115
4.3.1.3.6 Klada Qinvirus.....	117
4.3.1.3.7 Klada Tombus-Noda.....	117
4.3.1.3.8 Klada Toti-Chryso.....	118
4.3.1.3.9 Klada Reo.....	118
4.3.2 Vodni nevretenčarji.....	118
4.3.2.1 Raki (Crustacea).....	118
4.3.2.1.1 Klada Bunya-Arena.....	119
4.3.2.1.2 Klada Birna.....	119
4.3.2.1.3 Klada Flavi.....	119
4.3.2.1.4 Klada Hepe-Virga.....	120
4.3.2.1.5 Klada Luteo-Sobemo.....	121
4.3.2.1.6 Klada Mono-Chu.....	121

4.3.2.1.7	Klada Narna-Levi.....	123
4.3.2.1.8	Klada Tombus-Noda.....	124
4.3.2.1.9	Kladi Ophio in Qin.....	124
4.3.2.1.10	Klada Orthomyxo.....	124
4.3.2.1.11	Klada Partiti-Picobirna.....	125
4.3.2.1.12	Klada Permutotetra.....	125
4.3.2.1.13	Klada Picorna-Calici.....	125
4.3.2.1.14	Klada Reo.....	126
4.3.2.1.15	Klada Toti-Chryso.....	126
4.3.2.1.16	Klada Zhaovirus.....	127
4.3.2.2	Skupina Spiralia: mehkužci (Mollusca), ploščati črvi (Platyhelminthes), kolobarniki (Annelida) in kotačniki (Rotifera).....	127
4.3.2.2.1	Klada Astro.....	127
4.3.2.2.2	Klada Birna.....	128
4.3.2.2.3	Klada Bunya-Arena.....	128
4.3.2.2.4	Klada Flavi.....	128
4.3.2.2.5	Klada Hepe-Virga.....	128
4.3.2.2.6	Klada Luteo-Sobemo.....	129
4.3.2.2.7	Klada Narna-Levi.....	129
4.3.2.2.8	Klada Tombus-Noda.....	129
4.3.2.2.9	Klada Mono-Chu.....	130
4.3.2.2.10	Klada Orthomyxo.....	131
4.3.2.2.11	Klada Partiti-Picobirna.....	131
4.3.2.2.12	Klada Picorna-Calici.....	133
4.3.2.2.13	Klada Qinvirus.....	134
4.3.2.2.14	Klada Toti-Chryso.....	135
4.3.2.3	Najstarejše skupine živali: ožigalkarji (Cnidaria), spužve (Porifera) in rebrače (Ctenophora).....	135
4.3.2.3.1	Klada Bunya-Arena.....	135
4.3.2.3.2	Klada Birna.....	135
4.3.2.3.3	Klada Hepe-Virga.....	136
4.3.2.3.4	Klada Luteo-Sobemo.....	136
4.3.2.3.5	Klada Narna-Levi.....	136
4.3.2.3.6	Klada Tombus-Noda.....	137
4.3.2.3.7	Klada Ophio.....	137
4.3.2.3.8	Klada Mono-Chu.....	137
4.3.2.3.9	Klada Orthomyxo.....	138
4.3.2.3.10	Klada Partiti-Picobirna.....	138
4.3.2.3.11	Klada Picorna-Calici.....	138
4.3.2.3.12	Klada Reo.....	139
4.3.2.3.13	Klada Toti-Chryso.....	139
4.3.2.3.14	Klada Poty.....	139
4.3.2.3.15	Klada Qinvirus.....	139
4.3.2.3.16	Klada Weivirus.....	139
4.3.2.3.17	Klada Yuevirus.....	140
4.4	Strukturna raznolikost genomov RNA virusov pri raziskanih skupinah nevretenčarjev.....	140
4.4.1	Kopenski nevretenčarji.....	141
4.4.1.1	Kožekrilci (Hymenoptera).....	141
4.4.1.1.1	Klada Bunya-Arena.....	141
4.4.1.1.2	Klada Flavi.....	141
4.4.1.1.3	Klada Hepe-Virga.....	141
4.4.1.1.4	Klada Luteo-Sobemo.....	142
4.4.1.1.5	Klada Mono-Chu.....	142
4.4.1.1.6	Klada Narna-Levi.....	142
4.4.1.1.7	Klada Tombus-Noda.....	142
4.4.1.1.8	Klada Orthomyxo.....	142
4.4.1.1.9	Klada Partiti-Picobirna.....	142
4.4.1.1.10	Klada Permutotetra.....	143
4.4.1.1.11	Klada Picorna-Calici.....	143
4.4.1.1.12	Klada Qinvirus.....	143
4.4.1.1.13	Klada Toti-Chryso.....	143
4.4.1.2	Metulji (Lepidoptera).....	143
4.4.1.2.1	Klada Bunya-Arena.....	144

4.4.1.2.2	Klada Astro.....	144
4.4.1.2.3	Klada Hepe-Virga.....	144
4.4.1.2.4	Klada Luteo-Sobemo.....	145
4.4.1.2.5	Klada Mono-Chu.....	145
4.4.1.2.6	Klada Narna-Levi.....	146
4.4.1.2.7	Klada Tombus-Noda.....	146
4.4.1.2.8	Klada Orthomyxo.....	147
4.4.1.2.9	Klada Partiti-Picobirna.....	147
4.4.1.2.10	Klada Permutotetra.....	147
4.4.1.2.11	Klada Picorna-Calici.....	148
4.4.1.2.12	Klada Toti-Chryso.....	148
4.4.1.3	Najstarejše skupine žuželk: skakači (Collembola), dvorepke (Diplura), ščetinorepke (Zygentoma), Monocondylia.....	148
4.4.1.3.1	Klada Narna-Levi.....	148
4.4.1.3.2	Klada Hepe-Virga.....	148
4.4.1.3.3	Klada Orthomyxo.....	148
4.4.1.3.4	Klada Picorna-Calici.....	149
4.4.1.3.5	Klada Qinivirus.....	149
4.4.1.3.6	Klada Tombus-Noda.....	149
4.4.2	Vodni nevretenčarji.....	149
4.4.2.1	Raki (Crustacea).....	149
4.4.2.1.1	Klada Bunya-Arena.....	149
4.4.2.1.2	Klade Birna, Qinivirus in Ophio.....	149
4.4.2.1.3	Klada Flavi.....	150
4.4.2.1.4	Klada Hepe-Virga.....	150
4.4.2.1.5	Klada Luteo-Sobemo.....	151
4.4.2.1.6	Klada Mono-Chu.....	151
4.4.2.1.7	Klada Narna-Levi.....	152
4.4.2.1.8	Klada Orthomyxo.....	152
4.4.2.1.9	Klada Partiti-Picobirna.....	152
4.4.2.1.10	Klada Permutotetra.....	152
4.4.2.1.11	Klada Picorna-Calici.....	153
4.4.2.1.12	Klada Toti-Chryso.....	153
4.4.2.2	Skupina Spiralia: mehkužci (Mollusca), ploščati črvi (Platyhelminthes), kolobarniki (Annelida) in kotačniki (Rotifera).....	153
4.4.2.2.1	Klada Astro.....	153
4.4.2.2.2	Klada Birna.....	154
4.4.2.2.3	Klada Bunya-Arena.....	154
4.4.2.2.4	Klada Flavi.....	154
4.4.2.2.5	Klada Hepe-Virga.....	154
4.4.2.2.6	Klada Luteo-Sobemo.....	155
4.4.2.2.7	Klada Narna-Levi.....	155
4.4.2.2.8	Klada Tombus-Noda.....	155
4.4.2.2.9	Klada Mono-Chu.....	155
4.4.2.2.10	Klada Partiti-Picobirna.....	156
4.4.2.2.11	Klada Picorna-Calici.....	156
4.4.2.2.12	Klada Toti-Chryso.....	156
4.4.2.3	Najstarejše skupine živali: ožigalkarji (Cnidaria), spužve (Porifera) in rebrače (Ctenophora).....	157
4.4.2.3.1	Klada Bunya-Arena.....	157
4.4.2.3.2	Klade Birna, Qinivirus, Weivirus in Yuevirus.....	157
4.4.2.3.3	Klada Hepe-Virga.....	158
4.4.2.3.4	Klada Luteo-Sobemo.....	158
4.4.2.3.5	Klada Narna-Levi.....	158
4.4.2.3.6	Klada Tombus-Noda.....	158
4.4.2.3.7	Klada Mono-Chu.....	159
4.4.2.3.8	Klada Partiti-Picobirna.....	159
4.4.2.3.9	Klada Picorna-Calici.....	159
4.4.2.3.10	Klada Reo.....	162
4.4.2.3.11	Klada Toti-Chryso.....	162
5	Razprava	163
5.1	Kožekrilci (Hymenoptera).....	164
5.2	Metulji (Lepidoptera).....	166

5.3	Najstarejše skupine žuželk: skakači (Collembola), dvorepke (Diplura), ščetinorepke (Zygentoma), Monocondylia.....	169
5.4	Raki (Crustacea)	170
5.5	Skupina Spiralia: mehkužci (Mollusca), ploščati črvi (Platyhelminthes), kolobarniki (Annelida) in kotačniki (Rotifera).....	174
5.6	Najstarejše skupine živali: ožigalkarji (Cnidaria), spužve (Porifera) in rebrače (Ctenophora).....	177
5.7	Nevretenčarji.....	180
5.8	Endogenizirani RNA virusi v genomih nevretenčarjev.....	183
6	Zaključki	187
Dodatek A	Analiza zastopanosti virusnih klad	191
A.1	Analiza zastopanosti posameznih virusnih klad na podlagi rezultatov Shi in sod. (2016a)	191
A.2	Analiza zastopanosti virusnih skupin znotraj klade Picorna-Calici na podlagi rezultatov Shi in sod. (2016a)	193
Dodatek B	Filogenetska analiza	195
B.1	Kožekrilci (Hymenoptera).....	195
B.2	Metulji (Lepidoptera)	204
B.3	Najstarejše skupine žuželk: skakači (Collembola), dvorepke (Diplura), ščetinorepke (Zygentoma), Monocondylia.....	212
B.4	Raki (Crustacea)	217
B.5	Skupina Spiralia: mehkužci (Mollusca), ploščati črvi (Platyhelminthes), kolobarniki (Annelida) in kotačniki (Rotifera).....	227
B.6	Najstarejše skupine živali: ožigalkarji (Cnidaria), spužve (Porifera) in rebrače (Ctenophora).....	237
Dodatek C	Analiza organizacije novih virusnih genomov	249
C.1	Kožekrilci (Hymenoptera).....	249
C.2	Metulji (Lepidoptera)	258
C.3	Najstarejše skupine žuželk: skakači (Collembola), dvorepke (Diplura), ščetinorepke (Zygentoma), Monocondylia.....	261
C.4	Raki (Crustacea)	264
C.5	Skupina Spiralia: mehkužci (Mollusca), ploščati črvi (Platyhelminthes), kolobarniki (Annelida) in kotačniki (Rotifera).....	270
C.6	Najstarejše skupine živali: ožigalkarji (Cnidaria), spužve (Porifera) in rebrače (Ctenophora).....	275
	Literatura	281
	Bibliografija	293
	Biografija	295

Slike

Slika 1.1: Morfološka raznolikost virusov (Cann, 2012, povzeto po F.A. Murphy, School of veterinary medicine, University of California, Davis: http://www.vetmed.ucdavis.edu/viruses/VirusDiagram.html). Angleški izrazi: DNA viruses – DNA virusi; RNA viruses – RNA virusi; Reverse-transcribing viruses – retrovirusi.	4
Slika 1.2: Razširjenost virusov po Baltimoru med prokarionti (bakterije, arheje) in evkarionti. a) Koonin in sod., 2015, b) Nasir in Caetano-Anollés, 2015. Angleški izrazi: Prokaryotes – prokarionti; Eukaryotes – evkarionti; Retro-transcribing – retrovirusi; Plants – rastline; Fungi – glive. Oznaka IP – Invertebrates and plants.	14
Slika 1.3: Genomska arhitektura nestrukturnega dela astrovirusov. Trp_SpC – tripsinu-podobna serin proteaza; DUF – domena z neznano funkcijo (angl. »domain of unknown function«); RdRp_1 – od RNA odvisna RNA polimeraza družine 1, »hsdR superfamily« – naddružina proteinov hsdR (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD).	20
Slika 1.4: Segment A skupine Birnaviridae (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD). Angleški izrazi; Superfamily – naddružina.	21
Slika 1.5: Primer strukture virusov reda Bunyavirales – Feraviridae (Hulo in sod., 2011 – ViralZone). Angleški izrazi: Cleavage by host protease – cepitev z gostiteljsko proteazo; alternative initiation – alternativna iniciacija.	21
Slika 1.6: Primer strukture virusov reda Arenaviridae (Hulo in sod., 2011 – ViralZone).	22
Slika 1.7: Struktura poliproteina, tipičnega za Hepe-Virga klado (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD). Okrajšave: Vmethyltransf – virusna metiltransferaza, Viral_helicase1 superfam – virusna helikaza naddružine 1, RdRp_2 – od RNA odvisna RNA polimeraza družine 2.	22
Slika 1.8: ORF B skupine Hypoviridae: N-terminalni del proti C-terminalnemu delu: Peptidaza C8-DUF- RdRp- Helikaza DEAD/DEAH (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD).	24
Slika 1.9: Struktura Luteoviridae. Luteo P0 – luteovirusni P0 protein, Pept_S39 – peptidaza S39, RdRp (od RNA odvisna RNA polimeraza), Luteo_Ct-»luteovirus coat protein« (Shi in sod., 2016a).	24
Slika 1.10: Struktura Sobemovirusov. Pept_S39- peptidaza S39, RdRp (od RNA odvisna RNA polimeraza), S domain – »viral coat protein (S domain)«, Tryp- tripsinu podobna peptidaza (angl. »trypsin-like peptidase«) (Shi in sod., 2016a).	24
Slika 1.11: Struktura Barnaviridae. RdRp (od RNA odvisna RNA polimeraza), Tryp – tripsinu podobna peptidaza (angl. »trypsin-like peptidase«) (Shi in sod., 2016a).	25
Slika 1.12: Segmentiran virus – Luteo-Sobemo klada. RdRp (od RNA odvisna RNA polimeraza), Tryp – tripsinu podobna peptidaza (angl. »trypsin-like peptidase«), S domain – »viral coat protein (S domain)« (Shi in sod., 2016a).	25
Slika 1.13 Struktura narnavirusov in mitovirusov (Koonin in Dolja, 2014). Okrajšave: CP – protein plašča, MP – protein MP (angl. »movement protein«). Angleški izraz: Floating genus – »plavajoči rod« oz. neklasificirana virusna skupina.	26
Slika 1.14: Struktura Leviviridae - primer faga Qβ enterobakterij (angl. »enterobacteria phage Qβ«) (Hulo in sod., 2011 – ViralZone). Angleški izrazi: Maturation/lysis A2 – maturacijskega proteina mat A2, Minor CP A1 – daljša verzija proteina plašča CP, ki se prevede s pojavom zaviranja terminacije, Ribosomal read-through – zaviranje terminacije.	27
Slika 1.15: Struktura virusov iz reda Mononegavirales (Shi in sod. 2016a). Oznake: N – nukleoprotein; P – fosfoprotein; M – matriks protein. Angleški izrazi: Fusion glycoprotein – fuzijski glikoprotein; RNA capping domain – domena, odgovorna za dodajanje kape.	27
Slika 1.16 Struktura Chuvirusa (Shi in sod., 2016a). Oznake: N – nukleoprotein; G – glikoprotein. Angleški izraz »putative« – domnevni.	28
Slika 1.17: Primer strukture skupine Orthomyxoviridae (Shi in sod., 2016a). Oznake: HA – hemagglutinin, NA – nevraminidaza, M – matriks protein in NS – nestrukturni protein. Angleški izraz: Putative glycoprotein – domnevni glikoprotein.	28
Slika 1.18: Struktura jedrnega dela genoma humanega SARS koronavirusa Coronaviridae, Nidovirales) (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD).	29
Slika 1.19: Genomske strukture tipičnih predstavnikov družine Permutotetraviridae (Shi in sod., 2016a). Oznake: C_VNN – »nodavirus capsid protein«; S_domain – »Viral coat protein (S domain)«, Peptidase_A21 – »Peptidase A21 family coat protein« (Shi in sod., 2016a).	30

Slika 1.20: Primeri genomske strukture družine Partitiviridae (Shi in sod., 2016a). Angleški izraz: Capsid – kapsida.....	31
Slika 1.21: Genomska struktura »Beihai barnacle virus 14« (Shi in sod., 2016a).....	31
Slika 1.22: Primeri genomskih struktur virusne skupine Picobirnaviridae (Shi in sod., 2016a).....	31
Slika 1.23: Struktura Dicistroviridae (Shi in sod., 2016a). Strukturni poliprotein Rhv-Dicistro VP4-Rhv-CRPV je označen kot Picorna_Ct (angl. »picorna-like coat domains«), Hel – helikaza naddružine 3, Pept – peptidaza C3.....	32
Slika 1.24: Primer virusa z evolucijsko afiniteto do Dicistroviridae in genomom, kodiranim z enim ORF (Shi in sod., 2016a). Strukturni poliprotein Rhv-Dicistro VP4-Rhv-CRPV je označen kot Picorna_Ct (angl. »picorna-like coat domains«), Hel – helikaza naddružine 3, Pept – peptidaza C3.....	32
Slika 1.25: Struktura Iflavoridae. Strukturni poliprotein Rhv-Rhv-CRPV je označen kot Picorna_Ct (angl. »picorna-like coat domains«), Hel – helikaza naddružine 3, Pept – peptidaza C3 (Shi in sod., 2016a).....	32
Slika 1.26: Struktura genomov vodnih pikornavirusov (Shi in sod., 2016a). Niz Rhv-Rhv-CRPV je označen kot Picorna_Ct (angl. »picorna-like coat domains«), Hel – helikaza naddružine 3, Pept – peptidaza C3 (Shi in sod., 2016a).....	33
Slika 1.27: Struktura Nora virusa. Hel – helikaza naddružine 3, Pept – peptidaza C3 (Shi in sod., 2016a).....	33
Slika 1.28: Strukture kelp fly virusu sorodnih pikornavirusov. KfV_Ct – »Kelp fly virus coat proteins«, Hel – helikaza naddružine 3, Pept – peptidaza C3 (Shi in sod., 2016a).....	33
Slika 1.29: Struktura skupine Secoviridae. Picorna_Ct (angl. »picorna-like coat domains«) – modra, Hel – helikaza naddružine 3, Pept – peptidaza C3 (Shi in sod., 2016a).....	34
Slika 1.30: Struktura Arivirusa 1 in Arivirusa 2. Strukturni poliprotein Helikaza naddružine 3-Peptidaza C3-RdRp. Nestrukturni poliprotein vključuje dve Rhv domeni (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD).....	34
Slika 1.31: Struktura virusne skupine Picornaviridae. Picorna_Ct (angl. »picorna-like coat domains«), Hel – helikaza naddružine 3, Pept – peptidaza C3 (Shi in sod., 2016a).....	34
Slika 1.32: Struktura skupine Secoviridae. Picorna_Ct (angl. »picorna-like coat domains«) – modra, Hel – helikaza naddružine 3, Pept – peptidaza C3 (Shi in sod., 2016a).....	34
Slika 1.33: Struktura skupine Caliciviridae. Picorna_Ct (angl. »picorna-like coat domains«), Hel – helikaza naddružine 3, Pept – peptidaza C3 (Shi in sod., 2016a).....	34
Slika 1.34: Primer genomske strukture skupine Reoviridae (Hulo in sod., 2011 – ViralZone, Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD). Angleški izraz: genome – genom.....	35
Slika 1.35: Osrednji del genoma skupine Tombusviridae (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD). Oznaka: Tombus_P33 – tombusvirusni protein P33; RT_like in RdRp_3 – naddružini RdRp. Angleški izraz: Superfamily – naddružina.....	35
Slika 1.36: Struktura virusne skupine Carmotetraviridae (Hulo in sod., 2011 – ViralZone). Oznake: P130 in P40 – proteina P130 in P40; CP – protein plašča. Angleški izraz: replicase poliprotein – replikazni poliprotein.....	37
Slika 1.37: Primeri struktur virusne skupine Totiviridae (Shi in sod., 2016a). Kapsidna proteina (domeni): Totivirus_Coat, LA-virus_Coat. RdRp – od RNA odvisna RNA polimeraza.....	37
Slika 1.38: Primer strukture Quadriviridae (Shi in sod., 2016a). Angleški izraz: Structural protein – strukturni protein.....	38
Slika 1.39: Primer genomske strukture virusne skupine Chrysovirus (Shi in sod., 2016a). Angleški izraz: Major Capsid Protein – glavni protein kapside.....	38
Slika 1.40: Struktura genoma iz družine Ophioviridae (Hulo in sod., 2011 – ViralZone; Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD; Shi in sod., 2016a). Oznaka CP – protein kapside.....	38
Slika 1.41: a) Primer genomske strukture Yellow fever virus (Hulo in sod., 2011 – ViralZone; Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD; Shi in sod., 2016a). b) Primer genomske strukture virusa denga vročice (angl. »Dengue virus«) (Gebhard in sod., 2011). Angleški izraz: cap – kapa.....	39
Slika 1.42: Genomska struktura Cystoviridae (Hulo in sod., 2011 – ViralZone; Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD; Shi in sod., 2016a). Angleški izraz: prokaryotic operon: prokariontski operon.....	40
Slika 1.43: Struktura Potyviridae (Marchler – Bauer in sod., 2017 – CDD). Oznake: Peptidase_C3 – peptidaza C3; Peptidase_C6 – peptidaza C6; Potyviriid_P3 – protein P3, specifičen za Potyviridae; DEXDc in P-loop – helikaza; Poty_PP – poliprotein skupine potyviridae (angl. »Potyviridae poliprotein«); Peptidase_C4 – peptidaza C4; Poty_coat – protein plašča, specifičen za Potyviridae. Angleški izraz: superfamily – naddružina.....	40
Slika 1.44 : a) Zaporedje krajšega poliproteina P3N-PIPO. Rdeče in modro označena dela se prepisujeta v različnih bralnih okvirjih, b) struktura krajšega poliproteina P3N-PIPO; oznake: Peptidase_S30 – peptidaza S30; Peptidase_C6 – peptidaza C6; Potyviriid_P3 – protein P3, specifičen za Potyviridae (Marchler – Bauer in sod., 2017 – CDD, NCBI – https://www.ncbi.nlm.nih.gov/).....	41
Slika 1.45: Genomska struktura Weivirusov (Shi in sod., 2016a). Oznaka: Alverna_Coat – protein plašča homologen proteinu plašča skupine Alvernaviridae. Angleški izraz new – nov.....	41
Slika 1.46: Genomska struktura Yuevirusov (Shi in sod., 2016a). Angleški izraz new – nov.....	42

Slika 1.47: Genomska struktura Yanvirusov (Shi in sod., 2016a). Angleški izraz new – nov.....	42
Slika 1.48: Genomska struktura Qinvirusov (Shi in sod., 2016a). Angleški izraz new – nov.....	42
Slika 1.49: Genomska struktura Zhaovirusov (Shi in sod., 2016a). Oznaka RT – angl. »stop codon read-through«.	42
Slika 1.50: Povzetek mehanizmov vstopa virusov v gostiteljsko celico. Angleški izrazi: Receptor-mediated endocytosis – receptorsko-posredovana endocitoza, enveloped viruses – virusi z virusno ovojnico, pili – pili, peptidoglycan – peptidoglikan, clathrin or caveolin-mediated endocytosis – s klatrinom oz. kaveolinom posredovana endocitoza, most RNA viruses – večina RNA virusov, G-bacterial cell wall – celična stena gram negativnih bakterij, direct injection of genomic RNA – direktno injiciranje genomske RNA, direct translocation of virus particle into the cytoplasm – direktna translokacija virusnega delca v citoplazmo, fusion with cell membrane – fuzija s celično membrano.....	49
Slika 1.51: Podvojevanje, prepisovanje in prevajanje ss(+)RNA virusov (Ahlquist, 2006). Angleški izrazi: Translation – prevajanje, Template recruitment – rekrutiranje matrice, replication proteins – replikacijski protein, synthesis – sinteza, virion proteins – protein viriona.....	50
Slika 1.52: Podvojevanje, prepisovanje in prevajanje ss(-)RNA virusov (Cann, 2012). Angleški izrazi: Attachment and entry – vezava in vstop, uncoating – odstranjevanje virusnega plašča in/ali kapside, expression – izražanje, replicative intermediate – podvojitveni intermediat, replication – podvojevanje, proteins – protein, assembly – sestavljanje, release – sproščanje virusnega delca.....	51
Slika 1.53: Podvojevanje, prepisovanje in prevajanje dsRNA virusov (Ahlquist, 2006). Angleški izrazi: synthesis – sinteza, translation – prevajanje, packaging – pakiranje.....	52
Slika 1.54: Povzetek mehanizmov izstopa virusov iz celic. Angleški izrazi: budding – brstenje, ESCRT-dependent budding – od ESCRT odvisno brstenje, ESCRT-independent budding – od ESCRT neodvisno brstenje, complex – kompleks, viroporins – viroporini, cytoplasm – citoplazma, Endoplasmic reticulum – endoplazemski retikel, Golgi apparatus – Golgijev aparat, exocytosis – eksocitoza, nucleus – jedro, cell lysis – liza celic, budding – brstenje, mitochondrion – mitohondrij, hypha – hifa, direct cell to cell transfer – direkten prehod med celicami, hyphal anastomosis – hifalna anastomoza, fungal viruses – glivni virusi, plant viruses – rastlinski virusi, cytoplasm – citoplazma.....	54
Slika 1.55: Genom lentivirusa HIV-1.....	55
Slika 1.56: Povzetek metod, uporabljenih pri identifikaciji, karakterizaciji novih RNA virusov pri raziskovanju RNA virusne diverzitete (Halstead, 1966; Achong, 1972; Bauer in Murphy, 1975; Angly in sod., 2006; Culley in sod., 2006; Desnues in sod., 2008; Hogenhout in sod., 2008; Cantalupo in sod., 2011; Junglen in Drosten, 2013; Mizuno in sod., 2013; Virgin, 2014; Carapeta in sod., 2015; Paez-Espino in sod., 2016; Nouri in sod., 2015; Kawakami in sod., 2016; Ladner in sod., 2016; Rosani in Gerdol, 2017; Galinier in sod., 2017; Philosof in sod., 2017).....	62
Slika 1.57: Število metatranskriptomskih knjižnic glede na taksonomske skupine v referenci Shi in sod. (2016a).....	64
Slika 3.1: Število knjižnic v bazi podatkov TSA (angl. »Transcriptome Shotgun Assembly«) za analizirane skupine nevretenčarjev (8. 1. 2018).....	73
Slika 3.2: Število transkriptomskih (TSA, angl. »Transcriptome Shotgun Assembly«) knjižnic pri posamezni skupini živali (Metazoa). Analiza je bila narejena 7. 2. 2018.....	73
Slika 3.3: Število knjižnic v bazi podatkov WGS (angl. »Whole Genome Shotgun«) za analizirane skupine nevretenčarjev (8. 1. 2018).....	75
Slika 3.4: Število knjižnic v podatkovni bazi WGS (angl. »Whole Genome Shotgun«) za različne skupine živali (Metazoa). Analiza je bila narejena 7. 2. 2018 (do 17h).....	76
Slika 4.1: Zastopanost posameznih virusnih klad pri najstarejših skupinah živali (Metazoa).....	85
Slika 4.2: Zastopanost posameznih virusnih klad pri raki (Crustacea).....	86
Slika 4.3: Zastopanost posameznih virusnih klad pri skupini Lophotrochozoa (Spiralia).....	87
Slika 4.4: Zastopanost virusnih klad pri kožekrilcih (Hymenoptera).....	88
Slika 4.5: Zastopanost posameznih virusnih klad pri metuljih (Lepidoptera).....	89
Slika 4.6: Zastopanost posameznih virusnih skupin klade Picorna-Calici pri nevretenčarjih. Angleški izraz: related – podoben, npr. posavirus-related; posavirusu podobni virusi.....	92
Slika 4.7: Filogenetsko drevo virusov klade Hepe-Virga , ki smo jih našli v transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«) v programu IQTREE (Nguyen in sod., 2015). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro.....	98
Slika 4.8: Filogenetsko drevo virusov klade Mono-Chu , ki smo jih našli v genomih in transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«) v programu IQTREE (Nguyen in sod., 2015). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro. Endogenizirani virusi so označeni z zvezdico.....	100
Slika 4.9: Filogenetsko drevo virusov klade Partiti-Picobirna , ki smo jih našli v genomih in transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«) v programu IQTREE (Nguyen in sod., 2015). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali	

organizmi Vespoidea (ki niso Formicidae) s svetlo modro. Endogenizirani virusi so označeni z zvezdico.....	104
Slika 4.10: Filogenetsko drevo virusov klade Mono-Chu , ki smo jih našli v genomih in transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo. Endogenizirani virusi so označeni z zvezdico.....	109
Slika 4.11: Filogenetsko drevo virusov klade Partiti-Picobirna , ki smo jih našli v genomih in transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo. Endogenizirani virusi so označeni z zvezdico.....	111
Slika 4.12: Filogenetsko drevo virusov klade Picorna-Calici , ki smo jih našli v transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo.	113
Slika 4.13: Filogenetsko drevo virusov klade Orthomyxo , ki smo jih našli v genomih in transkriptomih najstarejših skupin žuželk. Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Skakači (Collembola) so označeni z rdečo, dvorepke (Diplura) z vijolično, ščetinorepke (Zygentoma) z modro in skupina Monocondylia z zeleno barvo. Endogenizirani RNA virusi so označeni z zvezdico.	115
Slika 4.14: Filogenetsko drevo virusov klade Picorna-Calici , ki smo jih našli v transkriptomih najstarejših skupin žuželk. Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Skakači (Collembola) so označeni z rdečo, dvorepke (Diplura) z vijolično, ščetinorepke (Zygentoma) z modro in skupina Monocondylia z zeleno barvo.	117
Slika 4.15: Filogenetsko drevo virusa klade Toti-Chryso , ki smo jih našli v transkriptomih najstarejših skupin žuželk. Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Skakači (Collembola) so označeni z rdečo, dvorepke (Diplura) z vijolično, ščetinorepke (Zygentoma) z modro in skupina Monocondylia z zeleno barvo.	118
Slika 4.16: Filogenetsko drevo virusov klade Flavi , odkritih v genomih in transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Skupina rakov enakonožcev (Isopoda) je označena z rdečo, skupina ceponožcev (Copepoda) je označena z zeleno, skupina rakov postranic (Amphipoda) je označena z modro, skupina deseteronožcev (Decapoda) je označena z vijolično, skupina Branchiopoda (Daphnia) je označena z rjavim odtenkom. Endogenizirani virusi so označeni z zvezdico.	120
Slika 4.17: Filogenetsko drevo virusov klade Mono-Chu , ki smo jih našli v genomih in transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) je z zeleno, postranice (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično. Endogenizirani virusi so označeni z zvezdico.	123
Slika 4.18: Filogenetsko drevo virusov klade Toti-Chryso , ki smo jih našli v genomih in transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postranice (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično. Endogenizirani virusi so označeni z zvezdico.	127
Slika 4.19: Filogenetsko drevo virusov klade Mono-Chu , ki smo jih odkrili v genomih in transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo. Endogenizirani virusi so označeni z zvezdico. Virusi, odkriti v referenci Waldron in sod. (2017), so označeni s krogcem.	131
Slika 4.20: Filogenetsko drevo virusov klade Partiti-Picobirna , ki smo jih odkrili v genomih in transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo. Endogenizirani virusi so označeni z zvezdico.	132
Slika 4.21: Filogenetsko drevo virusov klade Picorna-Calici , ki smo jih našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«) v programu IQTREE (Nguyen in sod., 2015). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo.	134
Slika 4.22: Filogenetsko drevo klade Birna . Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) z modro barvo.	136
Slika 4.23: Filogenetsko drevo virusov klade Mono-Chu , ki smo jih našli v genomih in transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo. Endogenizirani virusi so označeni z zvezdico.	138
Slika 4.24: Filogenetsko drevo virusov klade Weivirus , ki smo jih našli v genomih in transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno,	

rebrače (Ctenophora) pa z modro barvo. Endogenizirani RNA virusi so označeni z zvezdico. Virusi iz Waldron in sod. (2017) so označeni s krogci.....	140
Slika 4.25: Genomska organizacija virusov klade Flavi , ki smo jih našli pri kožekrilcih (Hymenoptera). Endogenizirana RNA virusa sta označena z zvezdico. Novi RNA virusi so označeni z rdečo pisavo.	141
Slika 4.26: Genomska organizacija virusov klade Partiti-Picobirna , ki smo jih našli pri kožekrilcih (Hymenoptera).	143
Slika 4.27: Genomska organizacija virusov klade Hepe-Virga , ki smo jih našli pri metuljih in veščah (Lepidoptera). Novi RNA virusi so označeni z rdečo pisavo.	144
Slika 4.28: Genomska organizacija virusov klade Luteo-Sobemo , ki smo jih našli pri metuljih in veščah (Lepidoptera).	145
Slika 4.29: Genomska organizacija virusov klade Mono-Chu , ki smo jih našli pri metuljih in veščah (Lepidoptera).	146
Slika 4.30: Genomska organizacija virusov klade Tombus-Noda , ki smo jih našli pri metuljih in veščah (Lepidoptera).	147
Slika 4.31: Genomska organizacija virusov klade Orthomyxo , ki smo jih našli pri najstarejših skupinah žuželk (Collembola, Diplura, Zygentoma, Monocondylia). Novi RNA virusi so označeni z rdečo pisavo.	149
Slika 4.32: Genomska organizacija virusov klad Birna , Qinvirus in Ophioviridae , ki smo jih našli pri raki (Crustacea).	150
Slika 4.33: Genomska organizacija virusov klade Luteo-Sobemo , ki smo jih našli pri raki (Crustacea).	151
Slika 4.34: Genomska organizacija virusov klade Narna-Levi , ki smo jih našli pri raki (Crustacea).	152
Slika 4.35: Genomska organizacija virusov klade Astro , ki smo jih našli pri skupini Spiralia. Novi RNA virusi so označeni z rdečo pisavo.	154
Slika 4.36: Genomska organizacija virusov klade Tombus-Noda , ki smo jih našli pri skupini Spiralia.	155
Slika 4.37: Genomska organizacija virusov klade Toti-Chryso , ki smo jih našli pri skupini Spiralia.	156
Slika 4.38: Genomska organizacija virusov klad Birna , Qinvirus , Weivirus in Yuevirus , ki smo jih našli pri najstarejših skupinah živali (Porifera, Cnidaria, Ctenophora). Novi RNA virusi so označeni z rdečo pisavo.	157
Slika 4.39: Genomska organizacija virusov klade Mono-Chu , ki smo jih našli pri najstarejših skupinah živali (Porifera, Cnidaria, Ctenophora).	159
Slika 4.40: Primerjava Tripsinu podobnih serinskih proteaz (angl. »Trypsin-like serine protease«) med <i>Nematostella vectensis</i> (HADN01010937) in »Hubei virga-like virusom 23« (Katoh in Standley, 2013 – MAFFT).	160
Slika 4.41: Genomska organizacija virusov klade Picornal-Calici , ki smo jih našli pri najstarejših skupinah živali (Porifera, Cnidaria, Ctenophora). Virusi, odkriti v referenci Waldron in sod. (2017), so označeni s krogcem.	161
Slika 5.1: Jedrni lokalizacijski signali pri virusni kladi Mono-Chu.	184
Slika 5.2: Jedrni lokalizacijski signali pri skupini Orthomyxoviridae.	185
Slika 5.3: Jedrni lokalizacijski signali pri virusni kladi Weivirus.	185
Slika A.1: Zastopanost virusnih klad pri posameznih redovih žuželk (Insecta) (Shi in sod., 2016a).	191
Slika A.2: Zastopanost virusnih klad pri posamezni skupini rakov (Crustacea) (Shi in sod., 2016a).	192
Slika A.3: Zastopanost virusnih klad pri posamezni skupini Spiralia (Lophotrochozoa) (Shi in sod., 2016a).	192
Slika A.4: Zastopanost posameznih virusnih skupin klade Picornal-Calici pri žuželkah (Insecta), raki (Crustacea) in mehkužci (Mollusca) (Shi in sod., 2016a). Angleški izraz: related – podoben, npr. posavirus-related; posavirusu podobni virusi.	193
Slika B.1: Filogenetsko drevo virusov klade Bunya-Arena , ki smo jih našli v transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro.	195
Slika B.2: Filogenetsko drevo virusov klade Flavi , ki smo jih našli v genomih in transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro. Endogenizirani virusi so označeni z zvezdico.	196
Slika B.3: Filogenetsko drevo virusov klade Luteo-Sobemo , ki smo jih našli v transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro.	197
Slika B.4: Filogenetsko drevo virusov klade Narna-Levi , ki smo jih našli v genomih in transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, ...	

Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro. Endogenizirani virusi so označeni z zvezdico.	198
Slika B.5: Filogenetsko drevo virusov klade Tombus-Noda , ki smo jih našli v transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro.	199
Slika B.6: Filogenetsko drevo virusa klade Ophio , ki smo ga našli v transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro.	200
Slika B.7: Filogenetsko drevo virusov klade Orthomyxo , ki smo jih našli v genomih in transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro. Endogenizirani virusi so označeni z zvezdico.	200
Slika B.8: Filogenetsko drevo virusov klade Permutotetra , ki smo jih našli v transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro.	201
Slika B.9: Filogenetsko drevo virusov klade Picornia-Calici , ki smo jih našli v transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro.	202
Slika B.10: Filogenetsko drevo virusov klade Qinvirus , ki smo jih našli v genomih in transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro. Endogenizirani virusi so označeni z zvezdico.	203
Slika B.11: Filogenetsko drevo virusov klade Reo , ki smo jih našli v genomih in transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro. Endogenizirani virusi so označeni z zvezdico.	203
Slika B.12: Filogenetsko drevo virusov klade Toti-Chryso , ki smo jih našli v genomih in transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro. Endogenizirani virusi so označeni z zvezdico.	204
Slika B.13: Filogenetsko drevo virusov klade Astro , ki smo jih našli v transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo.	204
Slika B.14: Filogenetsko drevo virusov klade Bunya-Arena , ki smo jih našli v transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skupino več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo.	205
Slika B.15: Filogenetsko drevo virusa klade Flavi , ki smo ga odkrili v transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo.	206
Slika B.16: Filogenetsko drevo virusov klade Hepe-Virga , ki smo jih našli v genomih in transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skupino več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo. Endogenizirani virusi so označeni z zvezdico.	207
Slika B.17: Filogenetsko drevo virusov klade Luteo-Sobemo , ki smo jih našli v transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skupino več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo.	207
Slika B.18: Filogenetsko drevo virusa družine Narnaviridae (klada Narna-Levi), ki smo ga odkrili v transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skupino več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo.	208
Slika B.19: Filogenetsko drevo virusov klade Tombus-Noda , ki smo jih našli v genomih in transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje	

verjetnosti (angl. » <i>maximum likelihood</i> «) v programu IQTREE (Nguyen in sod., 2015). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo. Endogenizirani virusi so označeni z zvezdico.	209
Slika B.20: Filogenetsko drevo virusov klade Orthomyxo , ki smo jih našli v genomih in transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo. Endogenizirani virusi so označeni z zvezdico.	210
Slika B.21: Filogenetsko drevo virusov klade Permutotetra , ki smo jih našli v transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo.	210
Slika B.22: Filogenetsko drevo virusov klade Toti-Chryso , ki smo jih našli v genomih in transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo. Endogenizirani virusi so označeni z zvezdico.	211
Slika B.23: Filogenetsko drevo virusov klade Reo , ki smo jih našli v transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo.	211
Slika B.24: Filogenetsko drevo virusov redu Tymovirales (klada Hepe-Virga), ki smo jih odkrili v transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo.	212
Slika B.25: Filogenetsko drevo virusa klade Mono-Chu , ki smo ga našli v genomu najstarejših skupin žuželk. Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skakači (Collembola) so označeni z rdečo, dvorepke (Diplura) z vijolično, ščetinorepke (Zygentoma) z modro in skupina Monocondylia z zeleno barvo. Endogeniziran RNA virus je označen z zvezdico.	212
Slika B.26: Filogenetsko drevo virusa družine Narnaviridae (klade Narna-Levi), ki smo ga odkrili v transkriptomih najstarejših skupin žuželk. Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skakači (Collembola) so označeni z rdečo, dvorepke (Diplura) z vijolično, ščetinorepke (Zygentoma) z modro in skupina Monocondylia z zeleno barvo.	213
Slika B.27: Filogenetsko drevo virusa klade Qinvirus , ki smo ga našli v transkriptomih najstarejših skupin žuželk. Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skakači (Collembola) so označeni z rdečo, dvorepke (Diplura) z vijolično, ščetinorepke (Zygentoma) z modro in skupina Monocondylia z zeleno barvo.	214
Slika B.28: Filogenetsko drevo virusov skupine Negevirus (klada Hepe-Virga), ki smo jih našli v genomih in transkriptomih najstarejših skupin žuželk. Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skakači (Collembola) so označeni z rdečo, dvorepke (Diplura) z vijolično, ščetinorepke (Zygentoma) z modro in skupina Monocondylia z zeleno barvo. Endogenizirani RNA virusi so označeni z zvezdico. Angleški izraz: group – skupina.	214
Slika B.29: Filogenetsko drevo virusov družine Tombusviridae (klada Tombus-Noda), ki smo jih našli v transkriptomih najstarejših skupin žuželk. Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skakači (Collembola) so označeni z rdečo, dvorepke (Diplura) z vijolično, ščetinorepke (Zygentoma) z modro in skupina Monocondylia z zeleno barvo.	215
Slika B.30: Filogenetsko drevo virusov klade Reo , ki smo jih našli v transkriptomih najstarejših skupin žuželk. Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Skakači (Collembola) so označeni z rdečo, dvorepke (Diplura) z vijolično, ščetinorepke (Zygentoma) z modro in skupina Monocondylia z zeleno barvo.	216
Slika B.31: Filogenetsko drevo virusov klade Bunya-Arena , ki smo jih odkrili v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postranice (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično.	217
Slika B.32: Filogenetsko drevo virusa klade Birna , ki smo ga našli v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postranice (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično.	217
Slika B.33: Filogenetsko drevo virusov klade Hepe-Virga , ki smo jih našli v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postranice (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično.	218
Slika B.34: Filogenetsko drevo virusov klade Luteo-Sobemo , ki smo jih našli v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postranice (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično.	219
Slika B.35: Filogenetsko drevo virusov klade Narna-Levi , ki smo jih našli v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum</i>	

- likelihood*). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postrance (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično. 220
- Slika B.36: Filogenetsko drevo virusov klade **Tombus-Noda**, ki smo jih odkrili v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postrance (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično. 221
- Slika B.37: Filogenetsko drevo virusov klade **Ophio in Qin** (skupine **Qinvirus**, **Yuevirus**, **Ophioviridae**, **Mononegavirales**, **Orthomyxoviridae** in **Bunyavirales**), ki smo jih našli v genomih in transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postrance (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično. Endogenizirani virusi so označeni z zvezdico. 222
- Slika B.38: Filogenetsko drevo virusov klade **Orthomyxo**, ki smo jih odkrili v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postrance (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično. 222
- Slika B.39: Filogenetsko drevo virusov klade **Partiti-Picobirna**, ki smo jih našli v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postrance (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično. 223
- Slika B.40: Filogenetsko drevo virusa klade **Permutotetra**, ki smo ga odkrili v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postrance (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično. 224
- Slika B.41: Filogenetsko drevo virusov klade **Picornia-Calici**, ki smo jih našli v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postrance (Amphipoda) z modro, deseteronožci (Decapoda) z vijolično in skupina Branchiura (*Argulus*) z magenta barvnim odtenkom. 225
- Slika B.42: Filogenetsko drevo virusov klade **Reo**, ki smo jih našli v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«) v programu IQTREE (Nguyen in sod., 2015). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postrance (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično. 226
- Slika B.43: Filogenetsko drevo virusa klade **Zhaovirus**, ki smo ga odkrili v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postrance (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično. 226
- Slika B.44: Filogenetsko drevo virusov klade **Astro**, ki smo jih odkrili v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo. 227
- Slika B.45: Filogenetsko drevo virusa klade **Birna**, ki smo jih našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo. 227
- Slika B.46: Filogenetsko drevo virusov klade **Bunya-Arena**, ki smo jih našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo. 228
- Slika B.47: Filogenetsko drevo virusov klade **Flavi**, ki smo jih našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo. 229
- Slika B.48: Filogenetsko drevo virusa klade **Luteo-Sobemo**, ki smo ga našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«) v programu IQTREE (Nguyen in sod., 2015). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo. 230
- Slika B.49: Filogenetsko drevo virusa družine **Hepeviridae** (klada **Hepe-Virga**), ki smo ga našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«) v programu IQTREE (Nguyen in sod., 2015). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, skupina kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo. 230
- Slika B.50: Filogenetsko drevo virusov družine **Narnaviridae** (klada **Narna-levi**), ki smo jih našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo. 231
- Slika B.51: Filogenetsko drevo virusov skupine **Negevirus** (klada **Hepe-Virga**), ki smo jih našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi

- (Platyhelminthes) z vijolično, skupina kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo. Angleška izraza: new – nova; group: skupina..... 232
- Slika B.52: Filogenetsko drevo virusov družine **Nodaviridae** (klada **Tombus-Noda**), ki smo jih našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, skupina kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo..... 233
- Slika B.53: Filogenetsko drevo virusa klade **Orthomyxo**, ki smo jih našli v genomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo. Endogenizirani virusi so označeni z zvezdico..... 234
- Slika B.54: Filogenetsko drevo virusov klade **Qinvirus**, ki smo ga našli v genomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo. Endogeniziran virus je označen z zvezdico..... 234
- Slika B.55: Filogenetsko drevo virusov družine **Tombusviridae** (klada **Tombus-Noda**), ki smo jih našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«) v programu IQTREE (Nguyen in sod., 2015). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo..... 235
- Slika B.56: Filogenetsko drevo virusov klade **Toti-Chryso**, ki smo jih našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo. 236
- Slika B.57: Filogenetsko drevo virusa družine **Togaviridae** (klada **Hepe-Virga**), ki smo ga našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo..... 236
- Slika B.58: Filogenetsko drevo virusov klade **Bunya-Arena**, ki smo jih našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo. Virus, ki smo ga našli v transkriptomu *Mnemiopsis leidyi* (GFAT01115010 + GFAT01138919 + GFAT01007784 + GFAT01017202 + GFAT01031184 + GFAT01090215 + GFAT01127526), smo označili s pravokotnikom..... 237
- Slika B.59: Filogenetsko drevo virusov družine **Hepeviridae** (klada **Hepe-Virga**), ki smo jih odkrili v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo..... 237
- Slika B.60: Filogenetsko drevo virusov klade **Luteo-Sobemo**, ki smo jih našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo. Virusi iz Waldron in sod. (2017) so označeni s krogci. 238
- Slika B.61: Filogenetsko drevo virusov družine **Narnaviridae** (klada **Narna-Levi**), ki smo jih našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo. Virusi iz Waldron in sod. (2017) so označeni s krogci..... 239
- Slika B.62: Filogenetsko drevo virusa družine **Nodaviridae** (klada **Tombus-Noda**), ki smo ga našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo..... 240
- Slika B.63: Filogenetsko drevo virusa klade **Ophioviridae**, ki smo ga našli v genomu najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo. Endogeniziran virus je označen z zvezdico. 240
- Slika B.64: Filogenetsko drevo virusov klade **Orthomyxo**, ki smo jih našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo. 241
- Slika B.65: Filogenetsko drevo virusov klade **Partiti-Picobirna**, ki smo jih našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo. Virusi iz Waldron in sod. (2017) so označeni s krogci. 242
- Slika B.66: Filogenetsko drevo virusov klade **Picornal-Calici**, ki smo jih našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo. Virusi iz Waldron in sod. (2017) so označeni s krogci. 243

Slika B.67: Filogenetsko drevo virusa klade Qinvirus , ki smo ga našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo.	244
Slika B.68: Filogenetsko drevo virusa klade Potyviriidae , ki smo ga našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo.	244
Slika B.69: Filogenetsko drevo virusa klade Reo , ki smo ga našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo.	245
Slika B.70: Filogenetsko drevo virusa redu Tymovirales (klada Hepe-Virga), ki smo ga našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo.	245
Slika B.71: Filogenetsko drevo virusov klade Yuevirus , ki smo jih našli v transkriptomih najstarejših skupin živali (Porifera, Cnidaria, Ctenophora). Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo.	246
Slika B.72: Filogenetsko drevo virusov družine Tombusviridae (klada Tombus-Noda), ki smo jih našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo. Virusi iz Waldron in sod. (2017) so označeni s krogi.	247
Slika B.73: Filogenetsko drevo virusov klade Toti-Chryso , ki smo jih našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo.	248
Slika B.74: Filogenetsko drevo virusov družine Leviviridae (klada Narna-Levi), ki smo jih našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. » <i>maximum likelihood</i> «). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo.	248
Slika C.1: Genomska organizacija virusov klade Bunya-Arena , ki smo jih našli pri kožekrilcih (Hymenoptera). Novi RNA virusi so označeni z rdečo pisavo.	249
Slika C.2: Genomska organizacija virusov klade Hepe-Virga , ki smo jih našli pri kožekrilcih (Hymenoptera).	250
Slika C.3: Genomska organizacija virusov klade Luteo-Sobemo , ki smo jih našli pri kožekrilcih (Hymenoptera).	251
Slika C.4: Genomska organizacija virusov klade Mono-Chu , ki smo jih našli pri kožekrilcih (Hymenoptera).	252
Slika C.5: Genomska organizacija virusov klade Narna-Levi , ki smo jih našli pri kožekrilcih (Hymenoptera).	253
Slika C.6: Genomska organizacija virusov klade Tombus-Noda , ki smo jih našli pri kožekrilcih (Hymenoptera).	254
Slika C.7: Genomska organizacija virusov klade Orthomyxo , ki smo jih našli pri kožekrilcih (Hymenoptera).	255
Slika C.8: Genomska organizacija virusov klade Permutotetra , ki smo jih našli pri kožekrilcih (Hymenoptera).	255
Slika C.9: Genomska organizacija virusov klade Picornal-Calici , ki smo jih našli pri kožekrilcih (Hymenoptera).	256
Slika C.10: Genomska organizacija virusov klade Qinvirus , ki smo jih našli pri kožekrilcih (Hymenoptera).	257
Slika C.11: Genomska organizacija virusov klade Toti-Chryso , ki smo jih našli pri kožekrilcih (Hymenoptera).	257
Slika C.12: Genomska organizacija virusov klade Astro , ki smo jih našli pri metuljih in veščah (Lepidoptera). Novi RNA virusi so označeni z rdečo pisavo.	258
Slika C.13: Genomska organizacija virusov klade Bunya-Arena , ki smo jih našli pri metuljih in veščah (Lepidoptera).	258
Slika C.14: Genomska organizacija virusov klade Orthomyxo , ki smo jih našli pri metuljih in veščah (Lepidoptera).	258
Slika C.15: Genomska organizacija virusov klade Narna-Levi , ki smo jih našli pri metuljih in veščah (Lepidoptera).	259
Slika C.16: Genomska organizacija virusov klade Permutotetra , ki smo jih našli pri metuljih in veščah (Lepidoptera).	259
Slika C.17: Genomska organizacija virusov klade Partiti-Picobirna , ki smo jih našli pri metuljih in veščah (Lepidoptera).	260
Slika C.18: Genomska organizacija virusov klade Picornal-Calici , ki smo jih našli pri metuljih in veščah (Lepidoptera).	260
Slika C.19: Genomska organizacija virusov klade Toti-Chryso , ki smo jih našli pri metuljih in veščah (Lepidoptera).	261

Slika C.20: Genomska organizacija virusov klade Narna-Levi , ki smo jih našli pri najstarejših skupinah žuželk (Collembola, Diplura, Zygentoma, Monocondylia). Novi RNA virusi so označeni z rdečo pisavo.	261
Slika C.21: Genomska organizacija virusov klade Hepe-Virga , ki smo jih našli pri najstarejših skupinah žuželk (Collembola, Diplura, Zygentoma in Monocondylia).	262
Slika C.22: Genomska organizacija virusov klade Tombus-Noda , ki smo jih našli pri najstarejših skupinah žuželk (Collembola, Diplura, Zygentoma, Monocondylia).	263
Slika C.23: Genomska organizacija virusov klade Qinvirus , ki smo jih našli pri najstarejših skupinah žuželk (Collembola, Diplura, Zygentoma, Monocondylia).	263
Slika C.24: Genomska organizacija virusov klade Picornal-Calici , ki smo jih našli pri najstarejših skupinah žuželk (Collembola, Diplura, Zygentoma, Monocondylia).	264
Slika C.25: Genomska organizacija virusov klade Bunya-Arena , ki smo jih našli pri rakih (Crustacea). Novi RNA virusi so označeni z rdečo pisavo.	264
Slika C.26: Genomska organizacija virusov klade Flavi , ki smo jih našli pri rakih (Crustacea). Endogeniziran virus je označen z zvezdico.	265
Slika C.27: Genomska organizacija virusov klade Hepe-Virga , ki smo jih našli pri rakih (Crustacea).	266
Slika C.28: Genomska organizacija virusov klade Orthomyxo , ki smo jih našli pri rakih (Crustacea).	266
Slika C.29: Genomska organizacija virusov klade Mono-Chu , ki smo jih našli pri rakih (Crustacea).	267
Slika C.30: Genomska organizacija virusov klade Partiti-Picobirna , ki smo jih našli pri rakih (Crustacea).	267
Slika C.31: Genomska organizacija virusov klade Permutotetra , ki smo jih našli pri rakih (Crustacea).	268
Slika C.32: Genomska organizacija virusov klade Picornal-Calici , ki smo jih našli pri rakih (Crustacea).	269
Slika C.33: Genomska organizacija virusov klade Toti-Chryso , ki smo jih našli pri rakih (Crustacea).	270
Slika C.34: Genomska organizacija virusov klade Birna , ki smo jih našli pri skupini Spiralia. Novi RNA virusi so označeni z rdečo pisavo.	270
Slika C.35: Genomska organizacija virusov klade Bunya-Arena , ki smo jih našli pri skupini Spiralia.	271
Slika C.36: Genomska organizacija virusov klade Flavi , ki smo jih našli pri skupini Spiralia.	271
Slika C.37: Genomska organizacija virusov klade Hepe-Virga , ki smo jih našli pri skupini Spiralia.	272
Slika C.38: Genomska organizacija virusov klade Luteo-Sobemo , ki smo jih našli pri skupini Spiralia.	273
Slika C.39: Genomska organizacija virusov klade Mono-Chu , ki smo jih našli pri skupini Spiralia.	273
Slika C.40: Genomska organizacija virusov klade Narna-Levi , ki smo jih našli pri skupini Spiralia.	274
Slika C.41: Genomska organizacija virusov klade Picornal-Calici , ki smo jih našli pri skupini Spiralia.	274
Slika C.42: Genomska organizacija virusov klade Partiti-Picobirna , ki smo jih našli pri skupini Spiralia.	275
Slika C.43: Genomska organizacija virusov klade Bunya-Arena , ki smo jih našli pri najstarejših skupinah živali (Porifera, Cnidaria, Ctenophora). Novi RNA virusi so označeni z rdečo pisavo.	275
Slika C.44: Genomska organizacija virusov klade Hepe-Virga , ki smo jih našli pri najstarejših skupinah živali (Porifera, Cnidaria, Ctenophora). Virusi iz Waldron in sod. (2017) so označeni s krogci.	276
Slika C.45: Genomska organizacija virusov klade Luteo-Sobemo , ki smo jih našli pri najstarejših skupinah živali (Porifera, Cnidaria, Ctenophora). Virusi iz Waldron in sod. (2017) so označeni s krogci.	277
Slika C.46: Genomska organizacija virusov klade Narna-Levi , ki smo jih našli pri najstarejših skupinah živali (Porifera, Cnidaria, Ctenophora). Virusi iz Waldron in sod. (2017) so označeni s krogcem.	277
Slika C.47: Genomska organizacija virusov klade Tombus-Noda , ki smo jih našli pri najstarejših skupinah živali (Porifera, Cnidaria, Ctenophora). Virusi iz Waldron in sod. (2017) so označeni s krogcem.	278
Slika C.48: Genomska organizacija virusov klade Partiti-Picobirna , ki smo jih našli pri najstarejših skupinah živali (Porifera, Cnidaria, Ctenophora). Virusi iz Waldron in sod. (2017) so označeni s krogcem.	278
Slika C.49: Genomska organizacija virusov klade Toti-Chryso , ki smo jih našli pri najstarejših skupinah živali (Porifera, Cnidaria, Ctenophora).	279
Slika C.50: Genomska organizacija virusov klade Reo , ki smo jih našli pri najstarejših skupinah živali (Porifera, Cnidaria, Ctenophora).	279

Tabele

Tabela 1.1: Endogenizirani RNA virusi, ki jih opisuje literatura. 1 – (Shi in sod., 2016a), 2 – (Li in sod., 2015), 3 – (Cotton in sod., 2016), 4 – (Katzourakis in Gifford, 2010), 5 – (Theze in sod., 2014), 6 – (Suzuki in sod., 2017), 7 – (Crochu in sod., 2004), 8 – (Aiewsakun in Katzourakis, 2015), 9 – (Cui in Holmes, 2011), 10 – (Feschotte in Gilbert, 2012), 11 – (Fort in sod., 2012), 12 – (Holmes, 2011), črna pika – prisotnost, bela pika – odsotnost.	56
Tabela 4.1: Na novo odkriti endogenizirani RNA virusi v genomih nevretenčarjev. Črna pika pomeni prisotnost, bela pika pomeni odsotnost posameznih virusnih klad.	93
Tabela 4.2: Odkrite endogenizirane RNA virusne klade pri členonožcih in ploščatih črvih v okviru analize Shi in sod. (2016a). Črna pika pomeni prisotnost, bela pika pomeni odsotnost posameznih virusnih klad.	94
Tabela 5.1: Prisotnost in odsotnost virusnih klad pri žuželkah, rakah, skupini Spiralia in najstarejših skupinah živali. Združili smo podatke iz literature z rezultati naših analiz. Črna pika – prisotnost znana že iz literature, kljukica – prisotnost, ki je bila na novo ugotovljena v okviru naših raziskav, bela pika – odsotnost virusne klade.	182

Seznam dodatkov in slik v okviru rezultatov

POGLAVJE	PODPOGLAVJE	PODPOGLAVJE	ZAPOREDNA SLIKA/DODATEK	STRAN
4.1 Analiza zastopanosti virusnih klad	4.1.1 Analiza zastopanosti posameznih virusnih klad	4.1.1.1 Rezultati Shi in sod. (2016a)	Dodatek A; A.1, Slika A.1 - Slika A.3	191-192
		4.1.1.2 Rezultati nastali v okviru doktorata	Slika 4.1 - Slika 4.5	85-89
	4.1.2 Analiza zastopanosti skupin znotraj klade Picorna-Calici	4.1.2.1 Rezultati Shi in sod. (2016a)	Dodatek A; A.2, Slika A.4	193
		4.1.1.2 Rezultati nastali v okviru doktorata	Slika 4.6	92
4.3 Filogenetska analiza RNA virusnih klad	4.3.1 Kopenski nevretenčarji	4.3.1.1 Kožekrilci	Slika 4.7 - Slika 4.9	98-104
			Dodatek B; B.1, Slika B.1 - Slika B.12	195-204
		4.3.1.2 Metulji	Slika 4.10 - Slika 4.12	109-113
			Dodatek B; B.2 Slika B.13 - Slika B.24	204-212
		4.3.1.3 Najstarejše skupine žuželk	Slika 4.13 - Slika 4.15	115-118
			Dodatek B; B.3 Slika B.25 - Slika B.30	212-216
	4.3.2 Vodni nevretenčarji	4.3.2.1 Raki	Slika 4.16 - Slika 4.18	120-127
			Dodatek B; B.4, Slika B.31 - Slika B.43	217-226
		4.3.2.2 Spiralia	Slika 4.19- Slika 4.21	131-134

			Dodatek B; B.5 Slika B.44 – Slika B.57	227-236
		4.3.2.3 Najstarejše skupine živali	Slika 4.22 - Slika 4.24	136-140
			Dodatek B; B.6, Slika B.58 – Slika B.74	237-248
4.4 Strukturna raznolikost genomov RNA virusov	4.4.1 Kopenski nevretenčarji	4.4.1.1 Kožekrilci	Slika 4.25, Slika 4.26	141-143
			Dodatek C; C.1, Slika C.1 – Slika C.11	249-257
		4.4.1.2 Metulji	Slika 4.27 - Slika 4.30	144-147
			Dodatek C; C.2 Slika C.12 – Slika C.19	258-261
		4.4.1.3 Najstarejše skupine žuželk	Slika 4.31	149
			Dodatek C; C.3 Slika C.20 – Slika C.24	261-264
	4.4.2 Vodni nevretenčarji	4.4.2.1 Raki	Slika 4.32 - Slika 4.34	150-152
			Dodatek C; C.4, Slika C.25 – Slika C.33	264-270
		4.4.2.2 Spiralia	Slika 4.35 - Slika 4.37	154-156
			Dodatek C; C.5 Slika C.34 – Slika C.42	270-275
		4.4.2.3 Najstarejše skupine živali	Slika 4.38, Slika 4.39, Slika 4.41	157-161
			Dodatek C; C.6, Slika C.43 – Slika C.50	275-279

Kratice

AIC	... angl. » <i>Aikaike Information Criterion</i> «
ATP	... Adenozin trifosfat, angl. » <i>adenosine triphosphate</i> «
BIC	... angl. » <i>Bayesian Information Criterion</i> «
BTV	... Virus bolezni modrikastega jezika, angl. » <i>Blue tongue virus</i> «
CaMV	... Virus cvetačnega mozaika, angl. » <i>Cauliflower mosaic virus</i> «
CDD	... angl. » <i>Conserved Domain Database</i> «
COG	... angl. » <i>Clusters of Orthologous Groups of proteins database</i> «
DDBI	... angl. » <i>DNA Database of Japan</i> «
DNA	... Deoksiribonukleinska kislina, angl. » <i>deoxyribonucleic acid</i> «
dsDNA	... Dvovijačna DNA, angl. » <i>Double-stranded DNA</i> «
dsDNA- RT	... angl. » <i>Double-stranded DNA retro-transcribing</i> «
dsRNA	... Dvovijačna RNA, angl. » <i>Double-stranded RNA</i> «
EBI	... Evropski inštitut za bioinformatiko, angl. » <i>European Bioinformatics Institute</i> «
EMBL	... Evropski laboratorij za molekularno biologijo, angl. » <i>European Molecular Biology Laboratory</i> «
ESCRT	... angl. » <i>Endosomal Sorting Complexes Required for Transport</i> «
EST	... angl. » <i>Expressed Sequence Tags</i> «
FF	... angl. » <i>fold families</i> «
FSF	... angl. » <i>fold superfamilies</i> «
GSS	... angl. » <i>Genomic Survey Sequences</i> «
GTR	... angl. » <i>General Time Reversible</i> «
HFMD	... Bolezen rok, nog in ust, angl. » <i>hand, foot and mouth disease</i> «
HIV	... Virus humane imunske pomanjkljivosti, angl. » <i>Human Immunodeficiency Virus</i> «
hli	... angl. » <i>high light inducible</i> «
HK97	... angl. » <i>bacteriophage Hong Kong 97</i> «
HTC	... angl. » <i>High-Throughput cDNA</i> «
HTG	... angl. » <i>High-Throughput Genomic Sequences</i> «
ICTV	... Mednarodni komite za taksonomijo virusov, angl. » <i>International Committee on Taxonomy of Viruses</i> «
ICVCN	... angl. » <i>International Code of Virus Classification and Nomenclature</i> «
INSDC	... angl. » <i>International Nucleotide Sequence Database</i> «
IRES	... angl. » <i>Internal Ribosomal Entry Sites</i> «
JTT	... angl. » <i>Jones-Taylor-Thornton model</i> «
LUCA	... angl. » <i>Last Universal Common Ancestor</i> «
LECA	... angl. » <i>Last Eukaryotic Common Ancestor</i> «
LRT, aLRT	... angl. » <i>Likelihood Ratio Test</i> «, » <i>approximate Likelihood Ratio Test</i> «
MVB	... Multivezikularna telesa, angl. » <i>Multivesicular bodies</i> «

NADH	... Nikotinamid adenin dinukleotid, angl. » <i>nicotinamide adenine dinucleotide</i> «
NCBI	... Nacionalni center za biotehnološko informacijo, angl. » <i>National Center for Biotechnology Information</i> «
NES	... angl. » <i>Nuclear Export Signal</i> «
NIG	... angl. » <i>National Institute of Genetics</i> «
NIH	... angl. » <i>National Institute of Health</i> «
NLS	... Jedrni lokalizacijski signal, angl. » <i>Nuclear Localization Signal</i> «
NNI	... angl. » <i>Nearest Neighbour Interchange</i> «
nr	... angl. » <i>non-redundant</i> «
NSP	... Nestrukturni protein, angl. » <i>non-structural protein</i> «
NTP	... Nukleozid trifosfat, angl. » <i>nucleotide triphosphate</i> «
OMIM/OMIA	... angl. » <i>Online Mendelian Inheritance in Man</i> «, » <i>Online Mendelian Inheritance in Animals</i> «
ORF	... Odprt bralni okvir, angl. » <i>Open Reading Frame</i> «
PAT	... angl. » <i>Patented sequences</i> «
PCR	... Verižna reakcija s polimerazo (angl. » <i>Polymerase Chain Reaction</i> «)
PDB	... angl. » <i>Protein Data Bank</i> «
PFAM	... angl. » <i>Protein Family Database</i> «
PIR	... angl. » <i>Protein Information Resource</i> «
PRD1/Adeno	... angl. » <i>Prototype member of the tectiviridae bacteriophages</i> «
RNA	... Ribonukleinska kislina, angl. » <i>ribonucleic acid</i> «
SIB	... Švicarski inštitut za bioinformatiko, angl. » <i>Swiss Institute of Bioinformatics</i> «
SMART	... angl. » <i>Simple Modular Architecture Research Tool</i> «
SPR	... angl. » <i>Subtree Pruning and Regrafting</i> «
SRA	... angl. » <i>Sequence Read Archive</i> «
ssDNA	... Enovijačna DNA, angl. » <i>Single-stranded DNA</i> «
ss(+)RNA	... Enovijačna, smiselna RNA, angl. » <i>Single stranded sense/positive strand RNA</i> «
ss(-)RNA	... Enovijačna, protismiselna RNA, angl. » <i>Single stranded antisense/negative strand RNA</i> «
ssRNA-RT	... angl. » <i>Single-stranded RNA retro-transcribing</i> «
STS	... angl. » <i>Sequence Tagged Sites</i> «
TBR	... angl. » <i>Tree Bisection and Reconnection</i> «
TNFR	... Receptor za dejavnik tumorske nekroze, angl. » <i>tumor necrosis factor receptor</i> «
TPA	... angl. » <i>Third Party Annotation</i> «
TrEMBL	... angl. » <i>Translated Nucleotide Sequence Database</i> «
TSA	... angl. » <i>Transcriptome Shotgun Database</i> «
UNA	... angl. » <i>Unannotated Sequences</i> «
UPGMA	... angl. » <i>Unweighted Pair Method with Arithmetic mean</i> «
WAG	... angl. » <i>Whelan-Goldman model</i> «
WGS	... angl. » <i>Whole Genome Shotgun Database</i> «

Simboli

L	... maksimalna verjetnost za model (AIC, BIC)
N	... število neodvisnih parametrov
n_1	... število podatkov za omenjen model

Slovar

ANGLEŠKO IME	SLOVENSKO IME	PROTEINSKE DRUŽINE V PODATKOVNI BAZI PFAM
Capsid protein	Protein kapside	pfam11729, pfam01829, pfam00980, pfam08762, pfam11492
Coat protein	Protein plašča	pfam00286, pfam02495, pfam05892, pfam00983, pfam00721, pfam01785, pfam00729, pfam00894, pfam01819, pfam08462, pfam05518
Glycoprotein	Glikoprotein	pfam05161, pfam07245, pfam03563, pfam01589, pfam00943, pfam01563, pfam16506, cl05624, cl03280, pfam00523, pfam00974, pfam01004, pfam00869, pfam01539, pfam03273
Helicase	Helikaza	pfam01443, pfam00910, pfam07652
Methyltransferase	Metiltransferaza	pfam01660, pfam04521
Movement protein	MP protein	pfam00803, pfam01307, pfam01107, pfam95318, pfam04817

Non-structural protein	Nestrukturni proteini	pfam11073, TIGR04210, pfam03231, pfam00948, pfam01560, pfam01005, pfam01002, pfam01350, pfam01006, pfam01538, pfam01506, cl07056, pfam08301, cl15181
Nucleocapsid protein	Nukleokapsidni protein	pfam00846, pfam02477, pfam05733, pfam00952, pfam01533, cl05749, pfam05505, cl20066, pfam00945
Peptidase	Peptidaza	pfam05381, pfam05417, pfam05415, pfam05533, pfam01707, pfam00944, pfam13365, pfam05416, pfam00863, cl21584, pfam05578, pfam12387, pfam05550
RNA-dependent Polymerase (RdRp)	RNA Od RNA odvisna RNA polimeraza	pfam04196, pfam00978, pfam02123, pfam00998, pfam00680, pfam00946, pfam00972, pfam00602
Structural protein	Strukturni proteini	pfam03014, pfam16504

Poglavje 1

Uvod

1.1 Virusi

1.1.1 Definicija virusov

Človeško poznavanje obolenj, povezanih z virusi ima precej daljšo zgodovino kot poznavanje virusov samih. Prva definicija virusov je bila precej enostavna. S širjenjem repertoarja raziskanih virusov in hkrati znanja o njihovi morfologiji, mehanizmi podvojevanja in strukturi genoma je definicija virusov postajala vedno bolj kompleksna. Za pionirja znanstvenega dela z virusnimi obolenji velja Louis Pasteur, ki je med drugimi projekti preučeval tudi steklino. Kot »snov ali agens«, ki povzroča steklino, je imenoval »virus« (iz latinskega izraza za strup, sluz, infektivna, škodljiva snov) (Cann, 2012). Vendar ni razlikoval bakterij od ostalih povzročiteljev obolenj. Mejniki v definiciji je postavil Dimitri Ivanovski, ki je opazil, da se simptomi prenašajo med rastlinami tobaka tudi ob inokulaciji z ekstraktom, ki je prešel skozi keramični bakteriološki filter. Opazil je, da bolezen rastlin tobaka povzroča »snov«, ki je manjša od bakterij. Friderick Loeffler in Paul Frosch (1898) sta ugotovila, da podobne »snovi« povzročajo obolenje goveda z boleznijo slinavke in parkljevke oz. boleznijo rok, nog in ust (HFMD, angl. »hand, foot and mouth disease«). Leta 1909 sta Karl Landsteiner in Erwin Popper odkrila, da tudi pri človeku poliomielitis povzroča podobna snov. Postavila se je prva definicija virusa kot »snovi, ki povzroča bolezen in je manjša od bakterije« (angl. »filterable disease agent«). Oblikovala se je nova perspektiva, ki je odprla nadaljnja vprašanja o morfologiji, strukturi in načinu razmnoževanja virusov. Frederick Twort je leta 1915 odkril bakteriofage, ki so postali glavni modelni sistem za proučevanje strukture in mehanizmov razmnoževanja virusov (Cann, 2012). Z odkritjem metode sekvenciranja po Sangerju, PCR (angl. »Polymerase Chain Reaction«) in kasneje tretje oz. naslednje generacije sekvenciranja (angl. »next generation sequencing«) se je število novo odkritih virusov precej povečalo, s tem pa se je razširila tudi kompleksnost njihove definicije in virusi so postali ločena domena življenja. Z odkritjem novih DNA virusov, med njimi leta 2003 t. i. »orjaških virusov« (angl. »giant viruses«), kot so Mimivirusi z 1,2 Mbp dolгим genomom, se je razširilo znanje o njihovi raznolikosti in definicija je postala izziv za virologe – kaj loči virus od celičnih organizmov (Raoult in sod., 2004; Koonin in sod., 2015; Cann, 2012). Po odkritju »orjaških virusov« sta Raoult in Fortere predlagala delitev organizmov na organizme, ki kodirajo ribosome (celični organizmi – bakterije, arheje in evkarionti), in organizme, ki kodirajo kapside (virusi). S tem pa v zgodbo niso bili vključeni virusom podobni elementi brez kapside – viroidi in virusoidi (Koonin in sod., 2015; Raoult in Fortere, 2008). Viroidi so namreč dolgi od 200 do 400 bp, virusoidi pa okrog 1000 bp dolge satelitne RNA (Cann, 2012). Primerna definicija, ki bi ločila viruse od celičnih organizmov se tako razvija še danes.

Virusi niso samo obligatni znotrajcelični paraziti (Balique in sod., 2015; Cann, 2012). Nekateri prokarionti, kot so bakterije Rickettsiae in Chlamydiae, so se tudi specializirali za znotrajcelično parazitsko življenje. Ravno tako niso vsi virusi manjši od bakterij. Najdaljši virusi namreč lahko dosežejo dolžino tudi do 400 nm (Mimivirusi). Najmanjše bakterije so dolge do 300 nm. Ravno tako vsi virusi nimajo enostavnejšega genoma od bakterij. Genom najkompleksnejših virusov obsega 2,5 Mbp (Pandoravirus), medtem ko genom najmanjših prokariontov (paraziti) obsega 0,58 Mbp (Mycoplasma genitalium) in evkariontov 2,3 Mbp (Encephalitozoon) (Cann, 2012). Za razliko od celičnega kraljestva življenja se virusi

»sestavljajo« in ne rastejo kot integrirana vsota lastnih komponent. Hkrati nimajo orodij za sintezo ribosomov in proizvodnjo metabolne energije, zato so popolnoma odvisni od gostiteljskih celic. Številni virusi podobni elementi, kot so viroidi in virusoidi, nimajo virusnega ovoja niti kapside. Vsekakor so virusi najštevilčnejše, povsod prisotne biološke tvorbe na Zemlji (Koonin in sod., 2006; Paez-Espino in sod., 2016). Ne obstaja celica brez lastnega repertoarja virusov. Kot kraljestvo, ločeno od celičnega kraljestva življenja (arheje, bakterije in evkarionti), se pomikajo med biomi. S tem kritično vplivajo na okolje, medicino, kmetijstvo in posledično na globalno gospodarstvo (Koonin in sod., 2006; Nasir in Caetano-Anollés, 2015; Balique in sod., 2015). Tretja generacija sekvenciranja (angl. »next generation sequencing«) je omogočila, da so se raziskave pomaknile od genetike, morfologije in strukture posameznega virusa na strukturo, genomiko in delovanje viroma in celotne virosfere. Virosfera je nabor vseh virusov in virusom podobnih delcev v vseh ekosistemih in v vseh gostiteljih (biosfera). Virom je del mikrobioma, sestavljenega iz vseh organizmov, ki naseljujejo biosfero, vključno z bakterijami in arhejami. Zajema nabor vseh virusov in virusom podobnih elementov, ki naseljujejo celične gostitelje in biosfero (je del virosfere, ki naseljuje biosfero) (Virgin, 2014). Z elementi biosfere (nabor vseh celičnih gostiteljev) so elementi viroma prepleteni v mrežo interakcij, ki je glavni vir vzporedne evolucije gostitelja in virusa. Z vpletenostjo v mrežo interakcij z vsemi biomi vplivajo na vse ravni biološke organizacije, od sestave genoma gostiteljev do delovanja ekosistema, zato se upravičeno pojavljajo argumenti o viromu kot najuspešnejši domeni življenja na Zemlji (Wasik in Turner, 2013).

Po sedanjí oceni naj bi bil z vidika analize zaporedij raziskan le 1 % virosfere. Biološka analiza virosfere seveda krepko zaostaja (Mokili in sod., 2012; Virgin, 2014). Raznolikost virusov smo torej šele začeli spoznavati kot »temno snov virusnega vesolja« in bogat vir prihodnjih odkritij (Mokili in sod., 2012).

1.1.2 Genetske in evolucijske prednosti virusov

Vprašanje biološke uspešnosti virusov je večplastno in ni rešljivo samo z uporabo enega merila. Gre za mrežo vzajemno odvisnih dejavnikov, ki privedejo do njihove končne prilagodljivosti v primeru nenadnih biotskih in abiotskih sprememb tako v ekosistemih kot v gostiteljih.

1.1.2.1 Številčnost populacije

Kot posledica kratkih generacijskih časov virusov je ena izmed njihovih ključnih adutov, ki vplivajo na njihovo biološko uspešnost, sama **številčnost virosfere**. Zaradi mehanizma podvojevanja, ki bazira na sestavljanju in ne na rasti, so sposobni ustvariti številno potomstvo v enem samem generacijskem ciklu. Pri tem virosfera izkorišča celični aparat za proizvodnjo metabolne energije in njihovih sestavnih delov (Wasik in Turner, 2013).

Virusi so najštevilčnejše tvorbe v oceanih. V vodnem okolju najdemo 10 milijonov virusnih delcev na mililiter oz. 10^{10} delcev na liter morske vode (Breitbart in sod., 2007; Mokili in sod., 2012; Bergh in sod., 1989). Virusi po številčnosti dominirajo tudi na kopenskih okoljih. Na gram vzorca prsti najdemo 10^9 virusov in virusom podobnih delcev (Mokili in sod., 2012). Če pogledamo človeka kot na element biosfere, ki je sestavljen iz 10^{13} celic, na njem hkrati najdemo 100-krat več virusov in 10-krat več bakterij (Mokili in sod., 2012). Če štejemo samo fage, se v katerem koli času na Zemlji nahaja približno 10^{31} – 10^{32} fagnih delcev (Wasik in Turner, 2013), kar je 10–100-krat več od števila celic v zemeljskih ekosistemih (Simmonds in sod., 2017). Biomasa virusov je trenutno ocenjena na 200 milijard ton, kar je masa 75 milijonov sinjih kitov (angl. »*blue whale*«). Če bi jih nanizali enega zraven drugega, bi merili v dolžino, ki je primerljiva z dolžino 65 galaksij (Simmonds in sod., 2017).

1.1.2.2 Hitrost mutacij

Poleg sposobnosti produkcije številčnega potomstva v kratkem časovnem obdobju je eden izmed adutov virusov, ki narekuje njihovo raznolikost in prilagodljivost, njihova **hitrost mutacije**. Mutacije so vir genetske raznolikosti in podlaga za genetski potencial, ki vodi v prilagodljivost posamezne vrste (Duffy in sod., 2008). Hitrost mutacij predstavlja število točkovnih mutacij, insercij in delecij, ki se kopičijo na časovno enoto ali na generacijski čas. Hitrost mutacij lahko poteka linearno, kar pomeni, da vse mutacije izvirajo iz istega

starševskega organizma, ali geometrično, kar pomeni, da je vsak potomec lahko nov vir mutacij. Število nakopičenih mutacij je odvisno od natančnosti polimeraze (Duffy in sod., 2008; Sanjuán in Domingo-Calap, 2016). Od RNA odvisna RNA polimeraza je manj natančna kot od RNA odvisna DNA polimeraza. Od RNA odvisna DNA polimeraza je manj natančna kot od DNA odvisna DNA polimeraza. Drugi dejavnik, ki vpliva na število mutacij, so kemijski mutageni in ostali celični encimi, ki so vzrok za kemijske modifikacije nukleotidnih baz (metilacija, oksidacija, deaminacija). Tretji dejavnik, ki vpliva na hitrost mutacij, je biokemijski aparat, ki izvaja popravilne mehanizme na genomu (Duffy in sod., 2008; Sanjuán in Domingo-Calap, 2016). Pomemben dejavnik, ki je vzrok delecij in insercij v virusnih genomih, pa je sam mehanizem podvojevanja in rekombinacije, ki ga izkorišča posamezna skupina virusov. Vsi omenjeni dejavniki vplivajo na dejstvo, da imajo virusi v povprečju višje hitrosti mutacij kot celični organizmi, hkrati pa imajo zaradi raznolikosti v mehanizmih podvojevanja tudi najširši razpon hitrosti mutacij, ki je med $1,5 \times 10^{-3}$ in $1,8 \times 10^{-8}$ mutacij na nukleotid na cikel podvojevanja (Duffy in sod., 2008). Pri genomu, dolgem 1000 bp, to pomeni 2 mutaciji na cikel podvojevanja, pri povprečno dolgem genomu virusov, 10000bp, to pomeni 15–20 mutacij na cikel podvojevanja, če vzamemo najvišjo hitrost podvojevanja. Hitrost, s katero se mutacije fiksirajo oz. stabilizirajo v populaciji, imenujemo hitrost nukleotidnih zamenjav (angl. »*substitution rate*«). Meri se kot število fiksiranih sprememb na določeni poziciji znotraj genoma v enoti časa. Odvisna je od genetskega drifta, generacijskega časa, naravne selekcije oz. selekcijskega pritiska in same hitrosti mutacij v populaciji. V vseh sistemih selekcijski pritisk optimizira hitrost mutacij oz. določa spodnjo mejo hitrosti mutacije. Spontane mutacije povečajo verjetnost, da se najde optimalna rešitev za minimiziranje omenjenih negativnih učinkov selekcijskega pritiska. Po drugi strani je večina spontanih mutacij škodljivih in organizmi ne preživijo. Statično okolje z nizkim selekcijskim pritiskom minimizira obremenitev škodljivih mutacij. Zato je hitrost mutacij hkrati omejena s stopnjo škodljivih mutacij in s tem izgubo viabilnosti. Zato mora vsak organizem z danimi mehanizmi najti ravnotežje med dvema nasprotno delujočima silama – genetskim driftom oz. selekcijskim pritiskom in izgubo viabilnosti zaradi škodljivih mutacij. Virusi so obligatni paraziti celic in nikoli niso izpostavljeni statičnemu okolju. Meje hitrosti mutacij so zato odvisne od naravne selekcije in tendence virusov, da se izognejo izgubi viabilnosti (Duffy in sod., 2008). Hkrati obstaja velikost genoma kot omejitveni faktor. Če bi bil pri določeni hitrosti mutacij virusov njihov genom večji, bi pogostejše prihajalo do usodnih mutacij. Virusi so s selekcijo prisiljeni k zmanjšanju neesencialnih genomskega elementov. Maksimum njihove hitrosti mutacij je določen z recipročno vrednostjo velikosti njihovega genoma (Duffy in sod., 2008; Wasik in Turner, 2013).

1.1.2.3 Način razmnoževanja

Eden izmed adutov, ki virusom omogočajo uspešno preživetje, je širok spekter mehanizmov razmnoževanja, ki imajo dve ključni lastnosti. Vsem mehanizmom je skupna njihova odvisnost od celic in integriranje virusnega biokemijskega aparata s celičnim. Hkrati sami mehanizmi razmnoževanja omogočajo mozaičnost genomske arhitekture.

Parazitizem je na nek način dominantni slog življenja na planetu. Omogoča namreč elegantno in ekonomično rešitev skladiščenja metabolne energije, ki jo proizvajajo gostitelji, in izkoriščanje njihovega biokemijskega aparata. Virusi pri svojem načinu razmnoževanja preusmerjajo pretok metabolne energije k nastanku sestavnih delov, sestavljanju in končnemu zorenju. Za ohranjanje homeostaze virusov je hkrati potrebno bistveno manj energije kot za celično homeostazo – podvojevanje DNA, sintezo proteinov, sintezo lipidov in sintezo ATP.

Poleg ekonomičnosti načina razmnoževanja virusni genetski sistemi interagirajo z genetskimi sistemi gostitelja in drugih virusov, ki so v istem časovnem intervalu v isti ekološki niši. Z gostiteljem in med virusi lahko pride do izmenjave posameznih genov ali pa več genov. Tako lahko določeni geni, ki sestavljajo virus, prihajajo iz oddaljeno sorodnih ali nesorodnih virusov (Wasik in Turner, 2013). Ob pridobitvi novih genov pride do premika, do spremembe hitrosti mutacij zaradi spremembe v genetskem driftu ob prehodu gena na nov virusni sistem.

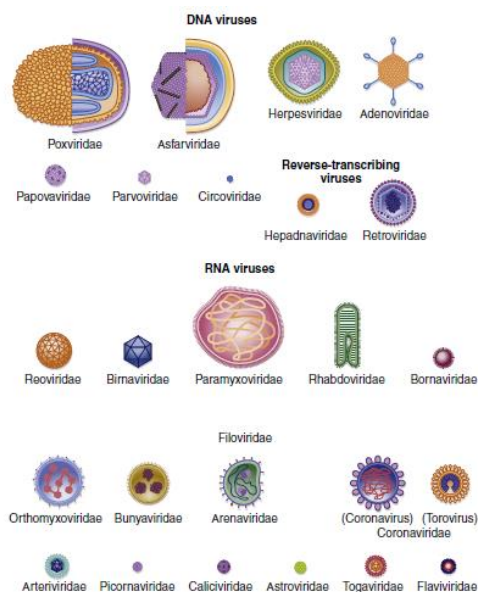
1.1.2.4 Raznolikost virosfere

Virosfera je del biosfere, ki ima eno najvišjih hitrosti mutacij, s tem pa največjo hitrost nukleotidnih zamenjav v biosferi. Hkrati je odvisna od biokemijskih aparatov celičnih

domen življenja in je s tem vpeta v mrežo interakcij na nivoju genoma in proteoma z vsakim njenim elementom (in z naborom ostalih virusov, ki v istem trenutku okužujejo gostitelja). Oba dejavnika narekujeta vzorec evolucije, katere posledica je **raznolikost virosfere**, ki za več razredov presega raznolikost celične domene življenja. Raznolikost virusov se kaže tako na nivoju genoma kot na nivoju morfologije (Dolja in Koonin, 2011; Wasik in Turner, 2013).

V nasprotju z univerzalnostjo celičnega genoma, ki temelji na DNA, virusni genom lahko hkrati predstavlja bodisi DNA bodisi RNA. Oba tipa nukleinskih kislin sta lahko v dvoverižni (angl. »*double-stranded*«) ali enoverižni (angl. »*single-stranded*«) obliki (Dolja in Koonin, 2011; Hulo in sod., 2011 – ViralZone). Ne glede na tip in obliko nukleinskih kislin je lahko genom virusov linearen ali krožen. Genom virusov je lahko sestavljen samo iz enega (nesegmentiran) ali iz več segmentov (segmentiran). Velikost virusnih genomov obsega tri velikostne razrede od najmanjših cirkovirusov (*Circoviridae*), katerih genom je velik 1,8 kb, do največjih odkritih pandoravirusov, katerih genom meri okoli 2,5 Mbp. Najenostavnejši genomi virusov kodirajo samo en protein, medtem ko najkompleksnejši genom kodira tudi do 2600 proteinov (pandoravirusi). V svojem ciklu razmnoževanja lahko virusi izkoriščajo vse možne DNA-RNA pretvorbe (Dolja in Koonin, 2011; Hulo in sod., 2011 – ViralZone).

Za viruse je značilna velika strukturna raznolikost. Najenostavnejše virusom podobne delce, kot so viroidi in virusoidi, sestavlja samo satelitna RNA molekula. Genom virusa je lahko obdan s kapsido, ki je lahko enojna ali dvojna. Kapsida je struktura, sestavljena iz ene ali več proteinskih podenot, ki se okrog virusnega genoma virusa ovijejo in uredijo v strukturo raznolikih simetrij na podlagi hidrofobnih in elektrostatskih interakcij. Genom s kapsido imenujemo ribonukleokapsida (angl. »*ribonucleocapsid*«). Pri segmentiranih virusih lahko kapsida obdaja samo en segment genoma ali celoten virusni genom. Kapsida je lahko obdana z virusno ovojnico (angl. »*viral envelope*«). Virusna ovojnica je lipidna struktura, ki izvira iz celične membrane gostitelja in ima torej podobno sestavo lipidnega dvosloja. Za prepoznavo virusnega delca s celičnimi receptorji je virusna ovojnica delno spremenjena z matriks proteini (angl. »*matrix proteins*«), ki virusno ovojnico povezujejo s kapsido in glikoproteini, katerih vloga je interakcija z receptorji na gostiteljski celici (Cann, 2012). Strukture nekaterih virusov so kompleksne in sestavljene iz več kot 100 različnih proteinov (*Poxviridae*). Poxvirusi so eni največjih virusov in lahko v dolžino merijo do 450 nm, v širino pa do 260 nm (Cann, 2012; Hulo in sod., 2011 – ViralZone). Eni najmanjših virusov so cirkovirusi, ki imajo premer samo 20 nm. Slika 1.1 predstavlja morfološko raznolikost virusov (Slika 1.1).



Slika 1.1: Morfološka raznolikost virusov (Cann, 2012, povzeto po F.A. Murphy, School of veterinary medicine, University of California, Davis: <http://www.vetmed.ucdavis.edu/viruses/VirusDiagram.html>). Angleški izrazi: DNA viruses – DNA virusi; RNA viruses – RNA virusi; Reverse-transcribing viruses – retrovirusi.

Omenjena raznolikost virusov odpira pogled na celotno virosfero z dveh perspektiv. Prvo perspektivo predstavljajo številčnost, hitrost mutacij in način razmnoževanja, ki narekujejo njen genetski potencial (Breitbart in sod., 2007; Wasik in Turner, 2013). Drugo perspektivo predstavlja visoko dinamično celično okolje. Genetski potencial virosfere je vzročno-posledično povezan z okoljem kot faktorjem virusne raznolikosti. Genetski drift, selekcijski pritisk celične ekološke niše in visoka dinamika celičnega okolja vodi v visoke hitrosti mutacij. Posledica je nastanek širokega spektra variant virusov, dominira tista varianta, ki se najbolj odziva na dano nišo. Biosfera ima na razpolago širok spekter habitatov oz. ekoloških niš. Določena vrsta organizmov je lahko fenotipsko plastična, kar pomeni, da lahko naseli več raznolikih niš. Lahko je tudi visoko specifična in povezana s točno določeno ekološko nišo. Virosfera ima številne viruse, ki so specializirani za določen ekosistem oz. ozek nabor gostiteljev. Ker so odvisni od celic, bodo naseljevali ekološke niše, ki podpirajo rast celične domene življenja. Genetska in fenotipska raznolikost virosfere omogoča naseljevanje številnih ekosistemov in gostiteljev, ki so na razpolago, torej dominirajo v biosferi (Wasik in Turner, 2013).

1.1.2.5 Prilagodljivost (angl. »*adaptability*«) in sposobnost razvijanja (angl. »*evolvability*«)

Dejavniki, od katerih je odvisna raznolikost virosfere (način razmnoževanja, številčnost populacije, hitrost mutacije), so hkrati dejavniki, ki omogočajo maksimiziranje učinkovitosti pri izkoriščanju njihove genetske kapacitete v dinamičnem ekološkem okolju. Virusi se med svojim življenjskim ciklom stalno premikajo med gostitelji in ekosistemi (voda, zrak, zemlja), ki so bolj ali manj primerni za njihovo razširjenost. Vendar virusi kot biološke tvorbe oz. delci med njimi ne ločijo. Zato lahko pasivno preidejo v ekološko past (okolje, ki slabo podpira ali ne podpira njihove rasti in razmnoževanja). Tudi v ustreznih ekoloških nišah pride do nihanj ekoloških dejavnikov (imunski sistem gostitelja). Skupek mehanizmov, s katerimi so virusi sposobni preiti čez ekološke pasti in čez dinamično okolje gostitelja, definira prilagodljivost organizma (angl. »*adaptability*«). Dejavniki, ki vodijo v raznolikost virosfere, virusom hkrati omogočijo večjo robustnost in prilagodljivost ob pogosti menjavi med gostitelji in različnimi ekološkimi okolji. Tudi bakterije namreč pasivno prehajajo v ekološke pasti in ne ločijo med primernimi in neprimernimi ekološkimi nišami, vendar zaradi slabše razvitih »obrambnih mehanizmov« pogosteje pride do »lokalnega izumrtja« v ekoloških pasteh (Wasik in Turner, 2013).

Prilagodljivost organizmov kot lastnost organizmov nam torej pove, kako hitro je določen organizem sposoben razviti mehanizme znotraj lastnih genetskih kapacitet, ki minimizirajo negativne učinke ob spremembi ekološkega dejavnika. Dinamično okolje vedno rezultira v nastanku variacij organizma znotraj populacije. Kako hitro nastane subpopulacija organizma z optimalno rešitvijo na stresorje dinamičnega okolja tekom naravne selekcije, meri sposobnost razvijanja organizma (angl. »*evolvability*«) (Wasik in Turner, 2013). Številčnost populacije in hitrost mutacij omogočita virusom hitro prilagoditev na stresorje specifične ekološke niše. Zaradi stalne menjave niš in stalnih sprememb znotraj ustrezne niše (ustreznega gostitelja) je tendenca virusov k stalnemu prilagajanju in razvijanju izmed vseh sistemov znotraj biosfere največja.

Vsi omenjeni dejavniki – številčnost, hitrost mutacij, način razmnoževanja, izpostavljenost dinamičnemu okolju, raznolikost, prilagodljivost in evolucija – so glavne lastnosti, ki virosferi omogočajo genetsko in evolucijsko prednost pred celično domeno življenja. Raznolikost virusov in evolucijske spremembe, ki so jo omogočile, so še vedno slabo raziskani. Številni znanstveniki se ukvarjajo s ključnimi vprašanji 1) o izvoru virusne domene življenja, 2) kako je profil faktorjev, ki opisujejo specifično dinamično okolje, vplival na potek njihove evolucije in 3) kakšna je raznolikost virusov v specifični dinamični ekološki niši v primerjavi z ostalimi nišami. Klasifikacija in karakterizacija virusov je ključnega pomena za odgovor na omenjena vprašanja, hkrati pa izziv zaradi izjemne virusne raznolikosti.

1.1.3 Klasifikacija virusov

1.1.3.1 Sistemi klasifikacije

Od odkritja »snovi, ki so manjše od bakterij«, do sodobnih raziskav se je spreminjal pogled na raznolikost virusov. Začetne metode raziskav so omogočale odkritje in karakterizacijo

omejenega števila virusov. Spremenljivke, ki so opisovale značilnosti posamezne skupine virusov (dejavniki klasifikacije), so bile diskretnega značaja. Skupine virusov so se jasno ločile med seboj. Začetek tretje generacije sekvenciranja, virusne metagenomike in metatranskriptomike so omogočili karakterizacijo in kvantifikacijo večjega števila virusov. Spremenljivke, ki so opisovale posamezno skupino virusov, so postajale vedno bolj kontinuirnega značaja. Pojavilo se je vedno več novih skupin virusov, ki so imele značilnosti med dvema prvotno opisanima virusnima skupinama (Shi in sod., 2016a). Ločnice med virusnimi skupinami so se začele brisati in prvotni sistemi klasifikacije so počasi začeli izgubljati na veljavi. Zaradi širokega razpona raznolikosti je bilo vse težje najti in določiti (nove) ključne faktorje, na podlagi katerih bi se definirane skupine virusov ločile med seboj. Dejavniki klasifikacije morajo namreč biti tiste virusne značilnosti, ki so dovolj ohranjene znotraj skupine, vendar dovolj divergentne, da ločijo različne klade ali različne skupine virusov med seboj.

Prvi učinkovit sistem klasifikacije virusov je vzpostavil David Baltimore leta 1971. Za ta sistem klasifikacije je leta 1975 dobil Nobelovo nagrado za fiziologijo in medicino. Prvi je skupne vzorce obnašanja virusov razdelil v skupine na podlagi skupnih značilnosti (Baltimore, 1971). Vsi virusni sistemi opravljajo dve ključni funkciji, podvojevanje genomskega materiala in sintezo mRNA. Opazil je, da sinteza mRNA pri virusih poteka po sedmih možnih poteh. Razred I predstavljajo virusi z dvoverižnim DNA genomom (dsDNA, angl. »double-stranded DNA«), pri katerih, tako kot pri celičnih organizmih, dsDNA služi kot matrika za sintezo mRNA. Razred II predstavljajo virusi z enoverižnim DNA genomom (ssDNA, angl. »single-stranded DNA«), ki je lahko v »smiselni« oz. »sense« ali »nasprotnosmiselni« oz. »antisense« obliki. Preko dsDNA intermediata nastane mRNA, ki je enake polarnosti kot začetna ssDNA (sinteza mRNA mora vedno potekati na novonastali verigi dvoverižne DNA molekule. Razred III predstavljajo virusi z dvoverižnim RNA genomom (dsRNA, angl. »double-stranded RNA«). Večina omenjenih virusov je segmentiranih in sinteza mRNA poteka na »antisense« RNA verigi. Razred IV predstavljajo virusi z enoverižno RNA molekulo, katerih mRNA je enake polarnosti kot genomska RNA virusa (ss(+)RNA, angl. »single-stranded sense/positive strand RNA«). mRNA v tem primeru nastane preko dsRNA intermediata. Razred V predstavljajo RNA virusi z enoverižno RNA komplementarno mRNA (ss(-)RNA, angl. »single-stranded antisense/negative strand RNA«). V tem primeru že genomska RNA lahko služi kot matrika za sintezo mRNA. Razred VI predstavljajo virusi z enoverižnim RNA genomom, ki se razmnožujejo preko dvoverižnega DNA intermediata (ssRNA-RT, angl. »single-stranded RNA retrotranscribing«). Razred VII predstavljajo virusi z dvoverižnim DNA genomom, ki se razmnožujejo preko RNA intermediata (dsDNA-RT, angl. »double-stranded DNA retrotranscribing«). Baltimore je definiral samo šest izmed sedmih naštetih skupin v delu iz leta 1971; »Expression of animal virus genomes«. Sedma skupina dsDNA-RT virusov se je priključila kasneje. Virus cvetačnega mozaika – CaMV (angl. »Cauliflower Mosaic Virus«) je bil prvi rastlinski virus, za katerega je bilo opaženo, da vsebuje DNA (Hull in Shepherd, 1976). Harrison in sod. (1971) so dokazali, da spada med caulimoviruse (Caulimoviridae).

Baltimorjev sistem klasifikacije je prevzela ICTV (angl. »International Committee on Taxonomy of Viruses«), ki je leta 1971 izdala svojo prvo poročilo o karakteriziranih in klasificiranih virusnih skupinah (Lefkowitz in sod., 2018). ICTV je organizacija, ki je bila ustanovljena z namenom taksonomske klasifikacije virusov in z uporabo najsodobnejšega dostopnega sistema za njihovo klasifikacijo (Lefkowitz in sod., 2018; Simmonds in sod., 2017). Od leta 1971 periodično objavlja poročila o trenutno znanih redovih, družinah, rodovih in vrstah virusov na podlagi mednarodnih pravil o klasifikaciji virusov ICVCN (angl. »International Code of Virus Classification and Nomenclature«). Prvotni sistem klasifikacije virusov je zajemal več kriterijev (angl. »multiple criteria system«); mehanizem podvojevanja, morfologijo virusov, serologijo, zaporedje, rang gostiteljev, patogenost, epidemiologijo, epizootiologijo (Simmonds in sod., 2017). Uvrstitev virusa v višje taksonomske skupine je zahtevala potrditev prisotnosti ustreznih bioloških karakteristik, značilnih za te skupine.

Tretja generacija sekvenciranja in doba virusne metagenomike je korenito spremenila virologijo. Z virusno metagenomiko je bilo več virusov klasificiranih samo z genomskim zaporedjem kot pa z eksperimentalno karakterizacijo in klasifikacijo. Z metagenomskimi podatki je bilo sestavljenih veliko število popolnih ali skoraj popolnih virusnih genomov. Med njimi so bile samo na podlagi zaporedij identificirane številne nove skupine, ki lahko igrajo ključno vlogo v biosferi. Uravnavajo lahko ekosistem, ki ga zasedajo. Vendar pa

manjkajo podatki o gostiteljih in ostali biološki podatki virusov, da bi ugotovili dejansko vlogo virusa v ekosistemu. Hkrati je problem velik prepad med trenutno sliko skupin, v katere se potencialno razvrstijo novi virusi, in številom skupin, ki so prepoznane s strani ICTV, saj potrditev virusne skupine v ICTV sistem poteka s časovnim zamikom od takrat, ko je bila nova virusna skupina identificirana s strani specifičnega raziskovalca. Torej uradna slika ICTV nazaduje glede na trenutno sliko hierarhije virosfere. Leta 2016 je zato imel izvršni odbor (angl. »executive committee«) ICTV zasedanje na temo novih sistemov za klasifikacijo in karakterizacijo virusov, identificiranih na podlagi virusne metagenomike in metatranskriptomike. S pričetkom tretje generacije sekvenciranja je klasifikacija virusov postala orientirana predvsem na zaporedje virusnega genoma (Lefkowitz in sod., 2018). Vendar se je z naraščanjem števila znanih virusnih genomov pojavila ključna problematika prekrivanja intervala podobnosti zaporedij med dvema sorodnima virusnima skupinama z intervalom podobnosti zaporedij virusov, ki sodijo v isto virusno. Zaporedja virusov, ki so bila uvrščena v isto skupino so bila med seboj samo 20% podobna, tako kot zaporedja dveh virusov, ki sta bila uvrščena v različni propadajoči skupini (Cann, 2012). Zato je z naraščanjem števila raziskav na tematiko virusne raznolikosti naraščala potreba po vključevanju več plasti informacij v klasifikacijo novih odkritih virusov, anotacija z dodatnimi biološkimi informacijami. Raziskovalci so upoštevali tudi podobne organizacije strukturnih in nestrukturnih genov znotraj genomov virusov, ki pripadajo isti skupini (Koonin in sod., 2006; Dolja in Koonin, 2011; Koonin in Dolja, 2013; Shi in sod., 2016a). Vsaka skupina virusov pa je privzela tudi dodatne gene, predvsem gene vključene v obrambo pred imunskim sistemom gostitelja in premikanje po sistemu gostitelja. Med družinami običajno ni bilo vidnih podobnosti v organizaciji genoma, vseeno pa so prisotni homologni geni, kot so od RNA odvisne RNA polimeraze, helikaze, proteaze in Rep kodirajoči proteini. Samo na podlagi zaporedij je bilo z začetkom virusne metagenomike in metatranskriptomike vse težje ugotoviti biološke korelacije med virusnimi skupinami in med virusi. Zato taksonomska klasifikacija ni trivialna in v literaturi so se pojavili konflikti glede primernosti uporabe specifičnih proteinskih zaporedij (npr. RdRp) kot filogenetskih markerjev (Zanotto in sod., 1996). Vseeno pa lahko iz nukleotidnega zaporedja in njegovih prevodov v vseh šestih bralnih okvirjih posredno pridobimo podatke o 1) evolucijski oddaljenosti med raziskanimi skupinami, 2) organizaciji genoma, 3) ravni izraženosti 4) ponavljajočih motivih, 5) prisotnosti in odsotnosti določenih motivov (IRES (angl. »*internal ribosomal entry sites*«), terminalne sekvence, cepitvena mesta proteinov), 6) vsebnost GC, 7) pogostnost dinukleotidov, 8) frekvenčna razporeditev rabe kodonov (Bahir in sod., 2009).

Biološki podatki torej postanejo sekundarna plast informacij, lahko pa jih predvidimo posredno iz podatkov, pridobljenih na podlagi sekvenc in njihovih prevodov. Skladno s premikom iz usmerjenosti na biološke karakteristike na zaporedje virusnega genoma se je prilagodila tudi ICVCN definicija višjih taksonomskih enot. Po ICVCN je rod definiran kot skupina vrst s skupnimi karakteristikami, družina kot skupina rodov s skupnimi karakteristikami in red kot skupina družin s skupnimi karakteristikami (Simmonds in sod., 2017). ICTV je poskusila mapirati Baltimorjevo klasifikacijsko shemo na shemo nabora gostiteljev.

Iz prve plasti informacij – sekvence genoma lahko torej posredno pridobimo nadaljnje plasti informacij, kot je npr. sekvenca proteoma iz prevodov sekvence genoma, predvidena sekundarna struktura virusnega proteoma in strukturna organizacija genoma. Razvitih je bilo več sistemov klasifikacije virusov in vsak sistem klasifikacije je upošteval svojo plast ali kombinacijo plasti informacij. Prvi sistem klasifikacije je na podlagi sekvence proteinov, ki so 1) dovolj raznoliki, da omogočajo ločevanje med virusnimi skupinami, 2) dovolj ohranjeni znotraj iste virusne skupine in 3) hkrati razširjeni čez vse virusne skupine, ki jih želimo ločiti med seboj. Ti proteini so ponavadi ključni za učinkovite življenjske cikle seta virusnih skupin in predstavljajo jedro seta virusnih skupin. Primer takšnega proteina je od RNA odvisna RNA polimeraza, ki je hkrati ključni gen RNA virusov. Skupina Shi in sod. (2016a) je opravila primer omenjene klasifikacije na podlagi proteina RdRp.

Druga plast informacij je lahko sekundarna struktura virusnega proteoma. S hkratnim razvojem metod sekvenciranja in analize sekvenc so se razvile tudi tehnike povezave sekvenc s sekundarno strukturo proteinov in njihovimi biofizikalnimi lastnostmi (Abrescia in sod., 2012; Abrescia in sod., 2010; Tokuriki in sod., 2009). Na podlagi sistematičnih strukturnih primerjav virusnih ortolognih proteinov, npr. proteinov kapsid, se je razvil sistem klasifikacije virusnih skupin, definiranih po ICTV v štiri skupine: 1) skupine, podobne

BTV (angl. »BTV-like, bluetongue virus-like«), 2) skupine, podobne adenovirusom (PRD1/Adeno (angl. »prototype member of the Tectiviridae DNA bacteriophages«)), 3) skupine, podobne HK97 (bakteriofag Hong Kong 97), in 4) skupine, podobne pikornavirusom (angl. »picorna-like«) (Abrescia in sod., 2012; Abrescia in sod., 2010).

Tretjo plast informacij predstavljajo genomska organizacija virusnih skupin in njena odstopanja. Omenjena informacija nastopa kot dodatna informacija k informaciji o zaporedju genoma za klasifikacijo virusov. Skupina Koonin in sod. (2014, 2015) je na podlagi zaporedja RdRp v kombinaciji z analizo strukturnih karakteristik genoma virusne skupine, klasificirane po ICTV, razdelila na štiri naddružine: 1) Pikornavirusom podobni virusi: Picornavirales, Luteoviridae, Partitiviridae, Caliciviridae, Hypoviridae, Potyviridae in Astroviridae idr., 2) Alphavirusom podobni virusi: Virgaviridae, Closteroviridae, Virgaviridae, Bromoviridae, Togaviridae, Hepeviridae, Endornaviridae in Tymovirales idr. 3), Flavivirusom podobni virusi: Tombusviridae, Umbravirus in Flaviviridae in 4) RNA bakteriofagom podobni virusi: Leviviridae, Narnaviridae, Ourmiavirus, 5) ss (-) RNA virusi, ki naj bi izvirali iz Flavivirusom podobne naddružine in 6) Cystovirusom podobna skupina (dsRNA bakteriofagi), ki so evolucijsko sorodni Reovirusom. Karakteristična za alfavirusom - podobne viruse je helikaza naddružine I (angl. »superfamily I helicase, alphavirus-like superfamily helicase«). Vseeno imajo nekateri Endornavirusi prisotno helikazo naddružine II, tipično za Flaviviridae, Potyviridae in Hypoviridae (Koonin in Dolja, 2014, Koonin in sod., 2015).

Iz narejenih raziskav je razvidno, da samo ena plast informacij, pridobljenih iz sekvenc, ni dovolj za natančno klasifikacijo, ampak je potrebno pri klasifikaciji uporabiti kombinacijo več informacij.

1.1.3.2 Lastnosti virusnih skupin

Z obdobjem tretje generacije sekvenciranja se je: 1) nabor pridobljenih podatkov bistveno povečal, vendar so bili podatki samo enega tipa – sekvence, 2) čas pridobivanja podatkov zmanjšal, 3) biološka anotacija se je lahko predvidela posredno iz samih sekvenc in 4) globlja plast informacij, ki je povezana z zaporedji in se je pridobila z ustreznim razvojem tehnik, je bila informacija o strukturi proteinov (Bahir in sod., 2009; Simmonds in sod., 2017; Lefkowitz in sod., 2018). Vseeno pa so se ohranili zametki Baltimorjeve klasifikacije. Viruse po Baltimorjevem sistemu klasifikacije delimo na dsDNA ssDNA, ss(+)RNA, ss(-)RNA, dsRNA, ssRNA-RT in dsDNA-RT. Potrebno je poudariti, da obstaja širok spekter morfoloških oblik virusov, ki izkoriščajo variabilne mehanizme podvojevanja (Hulo in sod., 2011 – ViralZone; Cann, 2012; Sobhy, 2016). Ne obstaja »povprečen ali konsenzni virus«, ki bi reprezentativno povzel biološke mehanizme podvojevanja, morfologijo ali strukturo razmnoževanja katere koli izmed omenjenih skupin virusov, saj ima vsak virus set svojih unikatnih lastnosti. Vseeno lahko povzamemo glavne značilnosti oz. posebnosti vsake od omenjenih skupin.

1.1.3.2.1 dsDNA

dsDNA virusi so virusi z dvovertičnim DNA genomom. Trenutno so definirani trije redovi: Herpesvirales, Caudovirales in Ligamenvirales. 28 skupin ni klasificiranih v redove: Adenoviridae, Polyomaviridae, Papillomaviridae, Baculovirusom podobni virusi (angl. »Baculo-like viruses«) – Baculoviridae in Nudoviridae, Hytrosaviridae, Lavidaviridae, Nimaviridae, Polydnoviridae, Ampullaviridae, Bicaudaviridae, Clavaviridae, Corticoviridae, Fuselloviridae, Globuloviridae, Guttaviridae, Plasmaviridae, Sphaerolipoviridae, Tectiviridae, Tristromaviridae, Turriviridae in »jedrno-citoplazemske večje DNA« (angl. »nucleocytoplasmic large DNA«) – Ascoviridae, Asfarviridae, Iridoviridae, Marseilleviridae, Mimiviridae, Phycodnaviridae in Poxviridae. Dinodnavirus, Pandoravirus in Pithovirus so zaekrat še nevrščeni v družine oz. neklasificirani (Hulo in sod., 2011 – ViralZone). Njihovi genomi se gibljejo od 4,5 kb s 5–9 ORF-ji do 2,3 Mbp s približno 2600 ORF-ji (Hulo in sod., 2011 – ViralZone). Večinoma so sferične, lahko so tudi opečnate, paličaste, jajčaste ali fibrilarne oblike, ali pa imajo »tripartitno« (oz. »trodelno«) sestavo iz glave, vratu in nog (bakteriofagi). Kapsida je večinoma obdana z virusno ovojnico, v primeru reda Caudovirales virus sestavlja samo gola kapsida. Njihov genom je nesegmentiran, razen v primeru Polidnaviridae. Razmnoževanje se prične s prepoznavo sialne kisline ali heparan sulfata. V celico preidejo s kalveolinom posredovano endocitozo ali s fuzijo s plazmalemo gostitelja. V zgodnjih fazah gostiteljska polimeraza prepíše gene, vključene v podvojevanje in obrambo

pred gostiteljem. Virusna RNA polimeraza nato prepíše še pozne gene. Proces podvojevanja poteka v »jedrnih tovarnah« (angl. »nuclear factories«). Povprečne hitrosti mutacij DNA virusov so bistveno nižje od RNA virusov. Gostiteljske in virusne DNA polimeraze so namreč zaradi popravljalnih (angl. »*proofreading*«) aktivnosti bistveno natančnejše (Duffy in sod., 2008).

1.1.3.2.2 ssDNA

ssDNA virusi imajo enoverižno DNA molekulo. Trenutno ni virusnih skupin, ki bi bile uvrščene v skupni red. Po ICTV klasifikaciji so ssDNA virusi uvrščeni v naslednje skupine (družine): Anelloviridae, Bidnaviridae, Circoviridae, Geminiviridae, Inoviridae, Nanoviridae, Microviridae, Genomoviridae, Parvoviridae, Pleolipoviridae, Spiraviridae, Tolecusatellitidae in skupina, ki še ni klasificirana v družino Bacilladnavirus (Hulo in sod., 2011 – ViralZone). Njihov genom je večinoma krožen, razen pri skupinah Parvoviridae, Microviridae in Bidnaviridae je linearen (Hulo in sod., 2011 – ViralZone). Kapsida večinoma ni ovita z virusno ovojnico, razen v primeru arhejskih virusov Pleolipoviridae. Velikost genoma je med 1,8 kb (Circoviridae, najmanjši znani virusi) in 24,9 kb (Hulo in sod., 2011 – ViralZone). Velikost genoma za razliko od dsDNA virusov ne korelira s številom kodirajočih ORF-jev. Po vstopu virusa v gostiteljsko celico gostiteljska DNA polimeraza sintetizira komplementarno verigo DNA, tako da nastane dsDNA intermediat. Omenjena veriga hkrati služi kot matrica za podvojevanje DNA in za prepisovanje v mRNA, prevajanje in končno sestavljanje delcev. Hitrost mutacij je višja od dsDNA virusov in nižja od RNA virusov (Duffy in sod., 2008; Sanjuán in Domingo-Calap, 2016), čeravno so znani primeri DNA virusov, ki hitreje mutirajo.

1.1.3.2.3 ss(+)RNA

Genom ss(+)RNA virusov predstavlja enoverižna *sense* veriga RNA. Definirajo jih trije redovi: Nidovirales, Tymovirales in Picornavirales in 22 skupin, ki niso klasificirane v specifičen red: Astroviridae, Alphatetraviridae, Alvernnaviridae, Barnaviridae, Benyviridae, Bromoviridae, Caliciviridae, Carmotetraviridae, Closteroviridae, Flaviviridae, Hepeviridae, Leviviridae, Luteoviridae, Narnaviridae, Nodaviridae, Permutotetraviridae, Potyvoviridae, Sarthroviridae, Solinviviridae, Togaviridae, Tombusviridae in Virgaviridae. Nekaj ss(+) RNA virusov ni uvrščenih v virusno družino (Hulo in sod., 2011 – ViralZone). Njihov genom je velik od 2,3 kb do 31 kb. Najkrajši genom ima Ourmiavirus, najdaljši pa Coronaviridae. Genom je izključno linearen. Lahko ga obdaja samo kapsida ali pa tudi virusna ovojnica. Hitrosti nukleotidnih zamenjav RNA virusov se gibljejo med 10^{-2} in 10^{-5} zamenjav/mesto/leto (Duffy in sod., 2008).

1.1.3.2.4 ss(-)RNA

Genom ss(-)RNA virusov predstavlja enoverižna *antisense* veriga RNA. Poleg redov Mononegavirales in Bunyavirales tri skupine niso uvrščene v redove: Ophioviridae, Arenaviridae in Orthomyxoviridae (Hulo in sod., 2011 – ViralZone). Velikost njihovega genoma je od 11 kb do 17 kb, kar je nekoliko ožji rang kot pri ss(+)RNA virusih (Hulo in sod., 2011 – ViralZone). Genom je izključno linearen, lahko je segmentiran ali nesegmentiran. Sestavlja ga od 2 do 8 segmentov. ss(-)RNA in ss(+) RNA virusi imajo največje hitrosti mutacij (Duffy in sod., 2008; Sanjuán in Domingo-Calap, 2016).

1.1.3.2.5 dsRNA

dsRNA virusi so virusi z dvoverižnim RNA genomom, veliki so od 3,7 kb do 30 kb. Genom je lahko segmentiran (2–12 segmentov) ali nesegmentiran. Hitrost mutacij virusov z dvoverižnim RNA genomom je nekoliko manjša od hitrosti mutacij virusov z enoverižnim RNA genomom (Sanjuán in Domingo-Calap, 2016).

1.1.3.2.6 ssRNA – RT

ssRNA-RT viruse sestavlja enoverižni RNA genom, ki se tekom cikla podvojevanja prepíše v DNA s pomočjo reverzne transkriptaze (od RNA odvisne DNA polimeraze). Poleg edine znane virusne skupine Retroviridae, ki jo sestavljata podskupini Orthoretrovirinae in Spumaretrovirinae, ssRNA-RT sestavljajo retrotranspozoni Pseudoviridae in Metaviridae. Genom Retroviridae je dolg med 7 in 11 kb in je obdan s kapsido in virusno ovojnico. Genom retrotranspozonov je dolg od 4,9 do 7,5 kb (Hulo in sod., 2011 – ViralZone). Njihov genom

predstavljajo trije deli: 1) gag: set genov (matriks protein, nukleokapsidni protein v primeru virusa HIV), ki interagirajo z virusno RNA in so odgovorni za njen prehod skozi jedrne pore, 2) pol: predstavlja jedro virusa in predstavlja set nestrukturiranih proteinov, odgovornih za replikacijo virusa – reverzna transkriptaza (prepiše RNA v DNA v dvostopenjskem procesu), RNazaH (razgradi RNA pri RNA/DNA hibridu) in integraza (integrira nastalo DNA v genom gostitelja) in 3) env: set strukturalnih proteinov, ki ovijejo ribonukleokapsido in omogočijo, da virus interagira z virusno ovojnico (gp120, »retroviral envelope protein« v primeru virusa HIV) (Hulo in sod., 2011 – ViralZone). Virusna ovojnica vsebuje površinske glikoproteine, ki interagirajo s celičnimi receptorji (transportni proteini (FeTHT1, XPR, FLVCR1 idr.), lipoproteinski receptorji (Tva), TNFR podobni receptorji (Tvb)), s pomočjo katerih sprožijo od kaveolina (angl. »caveolin«) odvisno endocitozo na glikolipidnih raftih (Hulo in sod., 2011 – ViralZone; Nabi in Le, 2003).

1.1.3.2.7 dsDNA – RT

dsDNA-RT viruse sestavlja izključno krožna dsDNA, dolga od 3 do 8,3 kb. V skupino dsDNA-RT sta klasificirani dve skupini; Caulimoviridae in Hepadnaviridae (Hulo in sod., 2011 – ViralZone). Skupina Caulimoviridae ima izključno rastlinske gostitelje, skupina Hepadnaviridae pa vretenčarske gostitelje. Caulimoviridae imajo samo kapsido, Hepadnaviridae pa kot vse živalske viruse sestavlja tudi virusna ovojnica. Po interakciji s celičnimi receptorji in prehodu skozi plazmalemo virusna DNA preko mikrotubulov preide skozi jedro, kjer poteka transkripcija s pomočjo celičnega transkripcijskega aparata. RNA nato preide skozi jedro, kjer poteka translacija novih virusnih proteinov. Hkrati RNA deluje kot matrika za reverzno transkripcijo v dsDNA v citoplazmi (Hulo in sod., 2011 – ViralZone).

1.1.3.3 Izzivi pri klasifikaciji in karakterizaciji virusov

S pričetkom dobe virusne metagenomike se je izjemno skrajšal čas, v katerem pridobimo dovolj informacij za klasifikacijo in karakterizacijo virusa. Iz dolgotrajnih seroloških, mikroskopskih metod, ki so temeljile na izolaciji in gojenju enega samega virusa, smo se premaknili na hitre metode tretje generacije sekvenciranja, ki zajamejo celoten virom določenega ekosistema in pridobijo ogromne količine podatkov. Pri tem smo se pomaknili od biološko usmerjene klasifikacije na sekvenčno usmerjeno klasifikacijo virusov. Podatkovne baze za skladiščenje genomskih, transkriptomskih, proteomskih in ostalih podatkov so rastle eksponentno. Vendar prednosti pridobitve izjemne količine podatkov prinašajo izzive v vsaki fazi identifikacije novih virusov, ki jih moramo upoštevati pri sami klasifikaciji virusov.

Prva skupina izzivov temelji na **sistemu zajemanja podatkov**. V metagenomiki in metatranskriptomiki gre namreč za sistem vzorčenja, pri katerem vzorčimo posamezen organizem, organizme, ki so sorodni med seboj, ali organizme, ki so v danem trenutku v isti ekološki niši. Iz teh organizmov po specifičnih protokolih nato izoliramo RNA ali DNA (ki lahko pripada DNA virusu, RNA virusu ali celičnemu organizmu, iz katerega smo nukleinsko kislino izolirali). Postopek zajema spiranje, homogenizacijo, izolacijo in čiščenje nukleinske kisline. Po izolaciji RNA lahko posamične vzorce (iz sorodnih organizmov ali organizmov iz iste ekološke niše) tudi združimo. Izolirani DNA ali RNA nato fragmentiramo in pripravimo genomske ali transkriptomske knjižnice (Shi in sod., 2016a). Dolžina fragmentov je odvisna od uporabljene metode sekvenciranja. S tem v končno knjižnico zajamemo skupino organizmov znotraj ekološke niše ali skupino organizmov, ki so sorodni med seboj. Poleg tega je bil virus lahko v trenutku vzorčenja prisoten v celici organizma, medcelični določenega tkiva ali samo na organizmu. Seveda hkrati z zajemanjem več organizmov v genomsko ali transkriptomsko knjižnico lahko zajamemo tudi mešano populacijo virusov, ki so lahko hkrati prisotni na enem ali več vzorčenih organizmih (Simmonds in sod., 2017). Zato **direktna korelacija z gostiteljem praktično ni mogoča**. Dodaten problem predstavljajo segmentirani virusi. Znotraj enega delca so lahko oviti vsi segmenti njihovega genoma (celoten genom) ali samo en segment. **Segmenti se lahko nahajajo na različnih gostiteljih** v določenem trenutku, zato je pridobivanje njihovega celotnega genoma dodatno oteženo. Sekvenčni podatki predstavljajo kratka zaporedja DNA ali RNA, katerih dolžina je odvisna od uporabljene metode.

Druga skupina izzivov izvira iz nadaljnjega **združevanja DNA ali RNA zaporedij**. Oligonukleotidi (angl. »reads«) so kratke sekvence izbrane dolžine (dolžine k) nukleotidov. Do 3 % nukleotidnih baz znotraj sekvence je nejasnih, dvoumnih, hkrati se soočamo še s

kontaminacijo s sekvenco vektorja, v katerem se nahaja tarčna sekvenca. Hkrati repetitivna zaporedja otežijo sestavljanje (Grabherr in sod., 2011; Havlak in sod., 2004). Za samo sestavljanje zaporedij obstajata dva osnovna pristopa. Prvi predstavlja direktno poravnavo zaporedij z referenčnim genomom (angl. »*top-down*«). Drugi zajema prvotno sestavljanje in nato mapiranje z referenčnim genomom (angl. »*bottom-up*«). Sistemi »*bottom-up*« zduževanja temeljijo predvsem na iskanju prekrivanja med kratkimi oligonukleotidi (Havlak in sod., 2004; Compeau in sod., 2011; Grabherr in sod., 2011; Schulz in sod., 2012). Zaradi stopnje negotovosti zaporedij, konfliktnih oligonukleotidov in ponavljajočih zaporedij lahko dobimo **hibridne genome**, ki ne pripadajo samo enemu virusu, temveč več virusom hkrati. Repetitivna zaporedja lahko privedejo tudi do **napačno sestavljenega virusnega genoma**, ki se kaže kot fragmentirani, okvarjeni proteini ob prevodu.

Tretja skupina izzivov izhaja iz same **analize zaporedij**. Transkriptomске in genomske baze podatkov namreč vsebujejo številne knjižnice iz različnih bakterijskih, rastlinskih in živalskih vrst. Število knjižnic eksponentno raste in knjižnice hkrati predstavljajo črno skrinjo potencialnih novih virusov. Po pridobitvi zaporedij in sestavljanju genoma novega virusa se sestavljene viruse primerja s proteinsko podatkovno bazo in s CDD (angl. »*Conserved Domain Database*«). S tem se poiščejo lažno-pozitivni rezultati (Shi in sod., 2016a). Po končni združitvi se analizira genomska struktura virusa – baza CDD. Baza CDD prepozna ohranjene proteinske domene na podlagi ohranjenih motivov, zabeleženih v podatkovnih bazah po principu HMM (angl. »*Hidden Markov Model*«). Divergentne proteinske domene brez karakterističnih motivov se ne bodo prepoznale. Problemi nastopijo, ker ravno zaradi tega številni **prepoznani virusi nimajo proteinskih domen** in jih vidimo v strukturi genoma kot neznane ORF-je. Zato je klasifikacija novih virusov otežena. Drug nivo izzivov predstavlja dejstvo, da so si virusi tekom evolucije s horizontalnim prenosom izmenjevali genetske informacije. Posamezni virusi so lahko **privzeli specifičen gen drugih virusov** ob okužbi istega gostitelja, kar ovira klasifikacijo virusa v višje taksonomske skupine.

Vsi izzivi klasifikacije težijo k temu, da raziskovalci kombinirajo več informacij in poskusijo klasificirati z več perspektiv.

1.1.4 Izvor virusov

Z razvojem tretje generacije sekvenciranja in napredkom virusne metagenomike in metatranskriptomike sta postali možni identifikacija in karakterizacija številnih novih virusov iz enormne količine sekvenčnih podatkov. Princip vzorčenja, ki poleg organizmov vzorči tudi viruse, njihove simbioante in parazite, je spremenil naše razumevanje virusne raznolikosti (Dolja in Koonin, 2018). Večina genetske raznolikosti v biosferi je namreč skoncentrirane v virusih kot posledica njihove 1) sposobnosti, da v kratkem časovnem intervalu nastane številno potomstvo, 2) visoke hitrosti mutacij in 3) njihovega načina razmnoževanja, ki vključuje horizontalni prenos genov iz celičnega kraljestva in omogoča modularnost njihove genomske arhitekture. Horizontalni prenos genov iz celičnega kraljestva na virosfero (iz relativno statičnega na dinamično okolje z visokim genetskim driftom) je hkrati povzročil premik v hitrosti njihovih mutacij. Zato je oteženo določanje izvora vsakega od virusnih genov in s tem določanje izvora virusov. Pomembno vprašanje v filogenomiki virusov je, ali so virusi monofiletskega ali polifiletskega izvora. Odgovore na problematiko so raziskovalci iskali z različnima pristopoma.

Prvi pristop predstavlja komparativna genomika z iskanjem homolognih ohranjenih genov znotraj vseh virusnih skupin in homolognih genov vsake od virusnih skupin s celično domeno življenja. Na podlagi omenjenega pristopa so se pojavljali različni modeli o izvoru virusov. Vseeno so predhodne raziskave prišle do skupnih zaključkov: 1) univerzalno ohranjen gen znotraj vseh virusov, viroidov in virusoidov ne obstaja in 2) virusi so najverjetneje polifiletskega izvora (Koonin in sod., 2006; Dolja in Koonin, 2011; Koonin in Dolja, 2013; Koonin in Dolja, 2014; Koonin in sod., 2015) in 3) klasificiranje in neklasificiranje virusne skupine so bile glede na strukturo njihovih genomov oz. razširjenost družin njihovih nestrukturiranih proteinov (helikaze in peptidaze) razdeljene na alfavirusom, pikornavirusom in flavivirusom podobne naddružine. Sodobna razlaga izvora virusov na podlagi komparativne genomike vključuje dva ključna evlucijska pojava. Prvi pojav predstavlja evlucijski nastanek evkariontov. Najverjetnejša razlaga zajema sobivanje arhejskega gostitelja in bakterijskega endosimbionta. Kasnejši razvoj jedra in endomembranskega sistema je bil podlaga za premik v sestavi evkariontskega viroma od

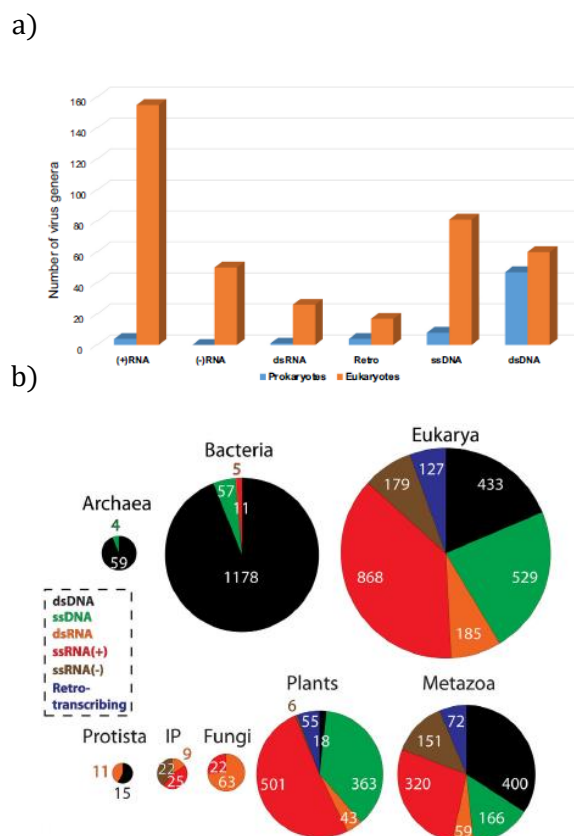
DNA viroma k RNA viromu. Jedrna bariera je namreč predstavljala izziv za podvojevanje DNA virusov. Hkrati je pojav endomembranskega sistema predstavljal ugodno okolje za podvojevanje RNA virusov (Dolja in Koonin, 2018). Drugi ključni pojav v evoluciji virusov, predvsem RNA virusov je horizontalni prenos virusov med gostitelji (angl. »*horizontal virus transfer*«) bodisi od rastlin na vektorje rastlinskih virusov, iz organizmov na njihove simbiote ali parazite (iz rastlinskih patogenih gliv na rastline, iz parazitoidnih os na žuželke, iz koral na njihove simbiote itd.). Nevretenčarji (organizmi kraljestva živali (Metazoa), ki ne sodijo v skupino vretenčarjev (Chordata)) so pri prenosu virusov med gostitelji igrali ključno vlogo. Vsekakor so virusi s svojimi gostitelji povezani v ko-evolucijsko mrežo. Glede na večjo evolucijsko starost bakterij bi lahko sklepali, da večina evkariontskih RNA virusov izvira iz ene izmed dveh znanih skupin RNA bakteriofagov, skupine Leviviridae (poleg Leviviridae so znani RNA bakteriofagi tudi Cystoviridae). Poleg skupine Leviviridae so Cystoviridae znani dsRNA bakteriofagi. Nasprotno pa zaporedje RdRp kaže, da so evkariontske RdRp bolj sorodne reverznim transkriptazam prokariontskih mobilnih elementov kot levivirusnim RdRp. Razširjenost bakterijskih mobilnih elementov glede na zelo omejeno razširjenost levivirusov podpira omenjeno domnevo, da so bakterijske reverzne transkriptaze evolucijski prednik RdRp evkariontskih RNA virusov. Poseben primer virusne skupine so Mitovirusi, ki imajo mehanizem podvojevanja povezan z mitohondriji. Glede na endosimbiontsko teorijo nastanka mitohondrijev iz α -proteobakterijskega endosimbionta in sorodstvene povezave proteinskih zaporedij RdRp mitovirusov s skupino Leviviridae lahko sklepamo na levivirusom podobnega prednika mitovirusov. Narnavirusi bi lahko izvirali iz mitovirusom podobnega prednika, ki je prešel iz podvojevanja v mitohondriju na podvojevanje v citosolu. Sorodstveni odnos ourmiavirusom podobnih virusov z glivnimi narnavirusi in mitovirusi in prevzem genov za proteine kapside iz rastlinskih tombusvirusov nakazuje na horizontalni prenos virusov iz rastlinskih patogenih gliv na rastlinske gostitelje najverjetneje preko nevretenčarskih vektorjev. Številni narnavirusom, mitovirusom in ourmiavirusom podobni virusi so bili najdeni pri nevretenčarjih in so povezani z glivnimi virusi iz glivnih patogenov, ki so nevretenčarski paraziti (Dolja in Koonin, 2018). Levivirusom podobni virusi so torej najverjetneje predniki ločene skupine narnavirusom, mitovirusom in ourmiavirusom podobnih virusov. Podoben naj bi bil tudi scenarij nastanka ostalih virusnih naddružin. Pikornavirusom podobna naddružina, ki je evolucijsko najstarejša virusna naddružina, naj bi nastala z izmenjavo genetskega materiala virusov, ki so izvirali iz bakterijskega endosimbionta in evkariontskega genoma (Dolja in Koonin, 2018). Na evolucijsko starost pikornavirusne naddružine nakazuje sam razpon gostiteljev (enocelični in večcelični evkarionti) virusov, ki sodijo v omenjeno naddružino. Flavivirusom podobna naddružina in alfavirusom podobna naddružina sta verjetno evolucijsko mlajši naddružini. Predstavniki obeh družin namreč za razliko od pikornavirusom podobne naddružine še niso bili najdeni pri enoceličnih evkariontih. Verjetno pa je njihov izvor povezan z nevretenčarji (Dolja in Koonin, 2018). Skupina dsRNA virusov je najverjetneje polifiletskega izvora. Predniki dsRNA virusov so najverjetneje ss(+)RNA virusi in skupine dsRNA virusov so sorodstveno povezane z ločenimi skupinami ss(+)RNA virusov. ss(-)RNA virusi so verjetno evolucijsko mlajša skupina RNA virusov, katerih izvor še vedno ni popolnoma pojasnjen. Sklepamo lahko, da je tudi pri omenjeni skupini RNA virusov horizontalni prenos med gostitelji igral ključno vlogo, najverjetneje v obliki prenosa vretenčarskega ss(-)RNA viroma na rastlinskega preko nevretenčarskih vektorjev (Dolja in Koonin, 2018).

Drugi pristop k raziskavi izvora virusov predstavlja strukturna biologija z iskanjem ohranjenih zvitij proteinov. Filogenetska analiza izvora RNA virusov se je namreč v raziskavah osredotočila samo na en gen ali na manjšo skupino genov in ne na celoten virusni proteom. Hkrati visoke hitrosti mutacij virusov in nizki odstotki podobnosti med zaporedji virusnih proteinov otežujejo filogenomske analize virusnih skupin na podlagi zaporedij (Nasir in Caetano-Anollés, 2015). Izkaže se, da v nasprotju z zaporedjem sekundarna struktura in zvitje proteinov ostaneta enaka. Zvitja proteinov (angl. »*folds*«) se združijo oz. klasificirajo v družine zvitij (angl. »*fold families* – FF«) in naddružine zvitij (angl. »*fold superfamilies* – FSF«). Družine zvitij zajemajo sekundarne strukture, ki so na nivoju sekvenc tipično nad 30 % identične. Naddružine zvitij (FSF) združujejo družine sekundarnih struktur s podobnimi strukturnimi jedri in biokemijskimi karakteristikami, na nivoju sekvenc pa so običajno manj kot 15 % identični (Nasir in Caetano-Anollés, 2015). Nasir in Caetano-Anollés (2015) sta analizirala 5080 proteomov arhej, bakterij, evkariontov in virusov, jih uvrstila v naddružine (FSF) in analizirala skupna zvitja, družine (FF) in naddružine zvitij (FSF) in zvitja, ki so specifična za določeno kraljestvo (bakterije, arheje,

evkarionti) in virosfero. Znotraj 5080 proteomov (cca. milijon proteinov) so identificirali 1995 naddružin zvitij (FSF). Izrisali so Venn-ov diagram s številom naddružin zvitij (FSF), ki so skupne vsem raziskanim domenam življenja (bakterijam, arhejam, evkariontom in virusom), dvema ali trem izmed njih ali pa so specifični samo za eno izmed domen. Manj kot desetina naddružin zvitij (66) je bilo specifičnih za viruse (Nasir in Caetano-Anollés, 2015). Slednje so predstavljale predvsem naddružine zvitij (FSF) proteinov, vključenih v patogenost oz. virulenco virusov in obrambo pred imunskim sistemom gostitelja. Visok delež naddružin zvitij (FSF), ki so skupne vsem trem celičnim kraljestvom, nakazuje na »soobstoj« prvotnih virusov (angl. »*ancient viruses*«) in prednika celične domene življenja (Nasir in Caetano-Anollés, 2015). Nasir in Caetano-Anollés (2015) sta izmed naddružin zvitij (FSF), ki si jih deli vsako izmed celičnih kraljestev z virusi, izbrala tiste, ki se pojavijo pri manj kot 2 % organizmov vsakega od celičnih kraljestev. Ti naj bi praviloma izvirali iz virusov in naj bi bili kritični za viruse kot gradbena enota kapside, virusne membrane ali so vključeni v pritrditev in vstop virusa v celico in nakazujejo na lastnosti prvotnega virusa. Dve izmed naddružin zvitij, ki sta bili najverjetneje prisotni pri prvotnem virusu, sta »double jelly-roll« zvitje in »VP5 kapsidni protein«. Nasir in Caetano-Anollés sta prišla do podobnih zaključkov kot Koonin in sod. (2014, 2015). Najstarejši virusi in virusom podobne entitete so imeli RNA genom.

1.1.5 Razširjenost virusov pri bakterijah, arhejah in evkariontih

Če pogledamo virome bakterij, arhej in evkariontov, opazimo, da noben virus ni prisoten hkrati pri evkariontih in katerem koli od prokariotov (bakterij ali arhej) (Koonin in sod., 2015). V arhejah so razširjeni izključno DNA virusi, večinoma dsDNA in v malem deležu ssDNA. V bakterijah so poleg dsDNA v največjem deležu in ssDNA v manjšem deležu prisotni tudi redki ss(+)RNA virusi (Leviviridae) (Nasir in Caetano-Anollés, 2015). Kooninova ekipa je k sliki o razširjenosti virusov med prokarioti dodala tudi ss(-)RNA, dsRNA viruse in retroelemente, ki so izjemno redki (Koonin in sod., 2015). Pri evkariontih je slika bistveno bolj raznolika. Približno polovico celotnega viroma predstavljajo RNA virusi, četrtno viroma ssDNA virusi, preostalo četrtno pa zajemajo dsDNA virusi in v majhnem deležu retrovirusi in retroelementi (Nasir in Caetano-Anollés, 2015). Podobno sliko je prikazala tudi Kooninova ekipa (Koonin in sod., 2015). Od RNA virusov pri evkariontih večinski delež poberejo ss(+)RNA virusi (Nasir in Caetano-Anollés, 2015). Če razdelimo evkarionte na živali (Metazoa), rastline, glive, protiste, vidimo bistveno večjo raznolikost v razširjenosti virusov v rastlinah in živalih kot pri glivah in protistih. Pri rastlinah in živalih so prisotni ss(+)RNA, in dsRNA, dsDNA, ssDNA in retrovirusi. Vendar ss(-)RNA viruse najdemo praktično samo pri živalih, dsDNA virusi pa so bistveno bolj razširjeni pri živalih kot pri rastlinah. Retrovirusi so v živalih in rastlinah razširjeni v približno enakem deležu. Glive in protiste v večinskem deležu okužujejo dsRNA virusi. Pri glivah so razširjeni tudi ss(+)RNA virusi, pri protistih pa dsDNA virusi. Če ločeno pogledamo viruse s hkrati rastlinskimi in živalskimi (nevretenčarskimi) gostitelji, opazimo, da nevretenčarje in rastline hkrati okužujejo RNA virusi (Slika 1.2 – oznaka IP, angl. »*invertebrates and plants*«) (Nasir in Caetano-Anollés, 2015). Nevretenčarji so namreč znani vektorji rastlinskih virusov in s tem ekološki partnerji rastlin. Slika 1.2 prikazuje razširjenost virusnih skupin po Baltimoru v arhejah, bakterijah in evkariontih glede na Kooninovo skupino in skupino Nasir in sod. (2014) (Slika 1.2). Števila znotraj grafov pri Sliki 2b prikazujejo število virusov s sekvenciranimi celotnimi genomi v bazi NCBI Viral Genomes, ki okužujejo specifičnega gostitelja (Nasir in Caetano-Anollés, 2015; Bao in sod., 2004 - NCBI; Hulo in sod., 2011 - ViralZone).



Slika 1.2: Razširjenost virusov po Baltimoru med prokarionti (bakterije, arheje) in evkarionti. a) Koonin in sod., 2015, b) Nasir in Caetano-Anollés, 2015. Angleški izrazi: Prokaryotes – prokarionti; Eukaryotes – evkarionti; Retro-transcribing – retrovirusi; Plants – rastline; Fungi - glive. Oznaka IP – Invertebrates and plants.

Podatki nakazujejo, da je nabor gostiteljev povezan z mehanizmom razmnoževanja (Nasir in Caetano-Anollés, 2015). Zakaj obstaja takšna razlika v razširjenosti virusov v evkariontih in prokariontih, ni znano, predpostavlja pa se, da je jedrna bariera pri evkariontih zožila podvojevalno (replikacijsko) nišo dsDNA virusov in otežila njihovo zorenje. Podvojevanje večine evkariontskih dsDNA virusov namreč poteka v citoplazmi evkariontov in tiste skupine dsDNA virusov, katerih mehanizem podvojevanja je vezan na jedro, naj bi se razvile kot mehanizem prilagoditve na jedrno bariero (Koonin in sod., 2015). V nasprotju z dsDNA virusi so retroelementi proliferajoče entitete, katerih mehanizem podvojevanja je vezan na integriranje v genom gostitelja. Evkariontski genomi so v primerjavi s prokariontskimi bistveno bolj tolerantni na povečanje velikosti genoma in vključevanje mobilnih elementov, zato so primernejša niša za retroviruse in retroelemente (Koonin in sod., 2015). Hkrati je znotrajcelični membranski sistem evkariontov ugodno dinamično okolje za podvojevanje RNA virusov (Koonin in sod., 2015). Za moderne arheje je značilno uspevanje v ekstremnih temperaturah. Predvideva se, da naj bi izguba RNA virusnih linij povzročila pomik prednikov arhej v ekstremne ekološke niše in njihov razvoj v evolucijsko sodobne oblike arhej (Nasir in Caetano-Anollés, 2015). RNA je namreč bistveno bolj nestabilna pri visokih temperaturah kot DNA (Nasir in Caetano-Anollés, 2015).

Tukaj moramo hkrati upoštevati dejstvo, da raznolikost genomov bakteriofagov (ssDNA virusov) še ni dosegla stopnje nasičenja. Trenutni podatki glede na število raznolikih morfotipov virusov namreč predpostavljajo, da je število raznolikih virusnih skupin (angl. »viral lineages«), ki okužujejo arheje, večje kot pri bakterijah. Hkrati raznolikost retroelementov ni v celoti zajeta z obstoječim sistemom klasifikacije, tako da je dejanska slika o raznolikosti in razširjenosti lahko precej drugačna (Nasir in Caetano-Anollés, 2015; Koonin in sod., 2015).

1.1.6 Vloga virusov

Okoljska genomika virusov, ki temelji na njihovi številčnosti v dinamičnih ekoloških nišah je prišla do ključnega odkritja, da je virosfera eden izmed glavnih geoloških dejavnikov, ki vplivajo na zemeljsko energetske konverzijo in energetske pretoke (Koonin in sod., 2015). Z vstopom v gostitelje pretvorijo gostiteljsko celico v virocelico – sistem, v katerega sta integrirana virusni in gostiteljski (celični) genom (Nasir in Caetano-Anollés, 2015). Genoma znotraj sistema sta integrirana v mrežo interakcij, ki povečujejo mozaičnost genomske arhitekture virusa in vplivajo tako na genotipske kot na fenotipske lastnosti virocelice. Virusi ustvarjajo evolucijsko in energetske sotočje z gostitelji, ki vpliva na globalno ekonomijo, medicino, agrikulturo in ekologijo (Nasir in Caetano-Anollés, 2015). Izguba katere koli od virusnih skupin bi pomenila vidno spremembo v energetskem pretoku ekosistema in evolucijskem krogotoku, kar bi povzročilo znatno spremembo v gospodarstvu, medicini in agrikulturi.

1.1.6.1 Okoljski vidik

Eden najbolj dovzetnih sistemov za vpliv viroma so morski ekosistemi. Oceani namreč pokrivajo 70 % površine Zemlje (Suttle, 2007). Številne raziskave so se ukvarjale s strukturo virusne populacije v morskih ekosistemih (Brum in sod., 2015). V 100 l morske vode lahko najdemo 5000 različnih virusnih genotipov, v 1 kg morskega sedimenta pa do milijon različnih virusnih genotipov (Rohwer in Turber, 2009). Ključni mehanizem, preko katerega virusi vplivajo na morski ekosistem, je preko njihovega načina razmnoževanja. Njihov način razmnoževanja ima dve bistveni lastnosti: 1) Virusni genomi delujejo kot prenosljivi vektorji, ki se pomikajo med spektri različnih gostiteljev in izkoriščajo gostiteljski biokemijski aparat za izmenjavo genetskega materiala z gostiteljem. S tem povečujejo modularnost lastnega genoma in hkrati modificirajo genom gostitelja. 2) Virusi med razmnoževanjem niso sposobni generirati lastne metabolne energije, zato izkoriščajo metabolno energijo in hkrati biokemijski aparat gostitelja za sestavljanje lastnih strukturnih enot in kasnejše zorenje. S tem metabolni in energetske pretok gostitelja popolnoma preusmerijo v virusni življenjski cikel. Z modifikacijo genoma gostitelja vplivajo na nabor sestavnih delov gostiteljskega biokemijskega aparata, s tem pa spreminjajo osnovno zasnovo gostiteljskega fenotipa. S preusmerjanjem metabolnega in energetskega pretoka gostitelja na lastno izgradnjo vplivajo na celično homeostazo. V obeh primerih posredno ali neposredno spreminjajo fiziologijo gostitelja in vplivajo na njegovo viabilnost. Ker so gostitelji med seboj povezani v interakcijsko mrežo, ki skupaj z okoljem tvori ekosistem, sprememba fiziologije katerega koli gostitelja znotraj ekosistema pomeni spremembo delovanja ekosistema.

Navedemo lahko več primerov izmenjave genetskega materiala med virusi in gostitelji, ki imajo končen vpliv na delovanje ekosistema. Fagi cianobakterij (cianofagi) (angl. »*cyanophages*«) v svojem genomu kodirajo gene za fotosintetske proteine: hli (angl. »*high-light inducible*«), *psbA*, *psbD* – glavna proteina reakcijskega centra fotosistema II (protein D1 in D2). Med litično fazo okužbe cianofag »ustavi« gostiteljev transkripcijski in translacijski sistem. Transkripcija fagnega proteina *psbA* (protein D1) med litičnim ciklom omogoči ohranjanje protonskega gradienta za produkcijo NADH in generiranje energije, potrebne za lizo bakterijskih celic in izgradnjo sestavnih delov faga. Hkrati so cianofagi centralni vektorji, ki sodelujejo pri horizontalnem prenosu gena *psbA* med cianobakterijami *Synechococcus* in *Prochlorococcus*. Cianobakterije kot eden najpomembnejših geoloških dejavnikov prispevajo četrtino globalnega fotosintetskega sistema. Drugi primer izmenjave genetskega materiala je med bakterijo *Vibrio cholerae* (njen habitat so morske obale) in fagom, ki v genomu vsebuje gen za toksin kolere (angl. »*cholera toxin*«) – gen *ctx*. Ob prevzemu gena se bakterija *Vibrio cholerae* iz simbionta pretvori v patogena. Genomi številnih fagov vsebujejo virulenčne gene, ki kodirajo zapis za 1) rezistenco na antibiotike, 2) adhezijske proteine, 3) proteine, vključene v sintezo kofaktorjev in vitaminov, 4) proteine, vključene v odziv na stres, 5) proteine, vključene v bakterijsko kemotakso. V oceanih se na leto med bakterijami in fagi izmenja 10^{24} genov. Vendar se ocenjuje, da je dejanski pretok genov med bakterijami v morskih ekosistemih bistveno večji zaradi virusom podobnih delcev (angl. »*generalized transduction agents*«), ki so po strukturi podobni fagom, vendar so manjši in vsebujejo izključno gostiteljske gene. *Elysia chlorotica* (angl. »solar-powered sea slug«), je mehkužec, ki v epitelnih celicah kopiči kloroplaste alg, s katerimi se prehranjuje. Del energije, potrebne za njegovo homeostazo zato pridobi

direktno s fotosintezo. Genom kloroplastov predstavlja zalogo samo frakcije genov, ki so potrebni za celoten biokemijski proces fotosinteze. Del genov je torej kodiranih v samem genomu *Elysia chlorotica*. Organizem jih je pridobil z virusno posredovanim horizontalnim prenosom genov, ki izvirajo iz alg, s katerimi se mehkužec prehranjuje. Virus, ki okužuje mehkužca, je hkrati odgovoren za omenjen horizontalni prenos. Sinhrona smrt mehkužca po odlaganju jajčec sovpada s pojavom virusnih delcev. Na osnovi virusom podobnih delcev poteka izmenjava genetskega materiala ne samo znotraj kraljestev bakterij, arhej in evkariontov, temveč tudi med kraljestvi. S stalnim pretokom genskega sklada (angl. »gene pool«) virusi preoblikujejo fenotipsko zasnovo celotne ekološke niše, posredno pa tudi spremenijo »fiziologijo« celotne ekološke niše (Rohwer in Thurber, 2009).

Navedemo lahko tudi več konkretnih primerov, v katerih okužba gostitelja z virusom in posledično preusmerjanje energetskega in metabolnega pretoka novonastale virocelice vpliva na delovanje ekosistema, v katerem se gostitelj (in tudi virus) nahaja. Skupina evkariontskih enoceličnih alg Coccolithales (evkariontski fitoplankton) vpliva na gradient globalne temperature. Njihovo »cvetenje« in »ponikanje« namreč deluje kot biološka črpalka, ki vpliva na kroženje in debelino plasti CO₂ na Zemlji in hkrati na odboj sončnih žarkov z Zemlje (Zemeljski »albedo«), kar uravnava zemeljsko temperaturo. Virusni s preusmerjanjem njihovega energetskega in metabolnega pretoka vplivajo na njihovo homeostazo, fiziologijo in viabilnost. S tem uravnavajo cikel njihovega cvetenja oz. ponikanja in posledično vplivajo na uravnavanje zemeljske temperature (Rohwer in Thurber, 2009). Z vplivom na metabolni in energetski pretok gostitelja virusi spreminjajo njegovo fiziologijo, njegov življenjski cikel in številčnost populacije. Naraščanje populacije gostitelja povzroči naraščanje gostiteljsko-specifičnih virusov, kar je z negativno povratno zanko povezano z upadom populacije gostiteljev (Suttle, 2007). Ker je gostitelj povezan z ostalimi organizmi v ekosistemu, virusi z uravnavanjem številčnosti populacije gostitelja hkrati uravnajo strukturo ekosistema. Kot primer lahko navedemo viruse, ki okužujejo rake deseteronožce (Decapoda, angl. »penaeid shrimp«) in ribe, ki so hkrati hrana drugim morskim organizmom (Suttle, 2007). Znanе so bile masovne okužbe delfinov in tulnjev z virusi skupine Morbillivirus (»Phocine distemper virus«) (Suttle, 2007). Nenazadnje so številni virusi humani patogeni, ki ključno vplivajo na strukturo človeške populacije. Virusni gradbene enote organizmov znotraj ekosistema pretvarjajo v razgrajen organski material in ogljikov dioksid, ki se nato inkorporira v nove živeče organizme. S tem vplivajo na kroženje ogljika znotraj vsakega od ekosistemov (Suttle, 2007).

Gostitelji virusov so z okoljem torej povezani v interakcijsko mrežo ekosistema. Vsak ekosistem kot vozeli v interakcijski mreži tudi sodeluje z drugimi ekosistemi. Z vplivom na katerega koli gostitelja virusi zato vplivajo na energetski in metabolni pretok (pretok organskih molekul, pretok ogljika) celotne mreže ekosistemov na Zemlji. Vplivajo na zemeljske biogeokemijske procese in strukturo zemeljske populacije (Suttle, 2007). Naše gospodarstvo, agrikultura, ekonomija in medicina so tesno povezani s celotno mrežo ekosistemov. Zato so virusi »glavni urejevalci« globalnega gospodarstva, agrikulture, medicine in ekonomije. Vse bolj očitno postaja, da je v znanje biogeokemije in biologije oceanov potrebno vgraditi tudi znanje o virusno posredovanih procesih (Suttle, 2007).

1.1.6.2 Medicinski vidik

Virusi preko mehanizma podvojevanja, ki je povezan z jedrom, endomembranskim sistemom ali citoplazmo, vplivajo na gostitelja na dva načina: 1) tako genom gostitelja kot genom virusa sta vključena v izmenjavo genetskega materiala, kar vpliva na himerno strukturo genoma virusa in hkrati modificira genom gostitelja; 2) virusi niso sposobni producirati lastne metabolne energije. S tem izkoriščajo metabolno energijo gostiteljskega biokemijskega aparata za izgradnjo lastnih komponent. Virus in gostiteljska celica sta spojena v »virocelico« s spremenjenim metabolnim pretokom in fiziologijo, ki deluje kot tovarna virusov. Fenotipsko se ob aktivni okužbi izrazijo bolezenski simptomi. Virofera se stalno premika med komponentami biosfere, kar hkrati vzpostavlja stalen genetski drift na virosfero in spreminja energetski pretok celotne biosfere. Virusne zoonoze (kot del virosfere) se premikajo med živalskimi gostitelji in človekom in so ključni povzročitelji človeških pandemij (Olival in sod., 2017). Okrog milijardo človeških okužb na leto (Han in sod., 2015) povzročajo zoonoze. Primati (Primates), netopirji (Chiroptera), sodoprsti kopitarji (Cetartiodactyla), lihoprsti kopitarji (Perissodactyla), ptice (Aves) in glodalci (Rodentia) so najzloglasnejši primeri prenašalcev človeških patogenov in številne raziskave so se posvečale faktorjem, ki omogočajo ali pospešujejo prenose zoonoz na človeka, in

raziskavi virusne raznolikosti v omenjenih skupinah živali, ki bi privedle do odkritja novih potencialnih zoonoz (Olival in sod., 2017; Mandl in sod., 2015). Netopirji (Chiroptera) so znani prenašalci koronavirusov, kot sta SARS koronavirus (virus sindroma akutne respiratorne stiske, angl. »*Severe Acute Respiratory Syndrome*«), MERS (angl. »*Middle East Respiratory Syndrome*«) in virusov skupine Mononegavirales, kot so virus Ebole, Hendra virus in Nipah virus. Koronavirus SARS povzroča bolezen sindrom akutne respiratorne stiske. Gre za obliko atipične pljučnice, ki se kapljično prenaša. Dihalno stisko pri boleznih spremlja visoka vročina, kašelj, hripavost, vneto žrelo, glavobol in visoka vročina. Prva pandemija boleznih se je pojavila med marcem in julijem leta 2003 in je bila najbolj problematična zaradi hitrosti razvoja simptoma akutne respiratorne stiske, ki je povzročila 50 % smrtnost pri potrjenih primerih. Zdravljenje je vključevalo interferon alfa (IFN α), ribavirin in proteazne inhibitorje (Stockman in sod., 2006). Septembra leta 2012 je Svetovna Zdravstvena Organizacija (angl. »*World Health Organization*«, WHO) poročala o prvih primerih atipične pljučnice, ki jo povzroča nov koronavirus MERS (Assiri in sod., 2013). Simptomi so podobni kot pri boleznih, ki jo povzroča virus SARS. Bolezen ebola povzroča virus ebola, ki sodi v družino Filoviridae (Mononegavirales). Odkrit je bil že leta 1976, vendar so ga šele kasneje povezali z netopirji (Mandl in sod., 2015). Trenutno je znanih pet različnih tipov virusa ebola: 1) Bundibugyo, 2) Zaire, 3) Reston, 4) Sudan in 5) »Tai forest«. Virusi se prenašajo bodisi preko direktnega kontakta bodisi preko telesnih tekočin. Simptomi vključujejo visoko vročino, izpuščaje, boleče grlo, glavobol, bolečine v mišicah, bruhanje in drisko. Simptomi se razvijejo v virusno hemoragično mrzlico, ki lahko poškoduje jetra in ledvice (Jadav in sod., 2015). Hendra in Nipah virus sta virusa skupine Mononegavirales (Paramyxovirus), ki sta se prvič pojavila v Maleziji leta 1998. Vzrok za smrtnost boleznih, ki ju povzročata, je najpogosteje encefalitis (vnetje možganskega tkiva) (Ang in sod., 2018). Virus influence tipa A je primer zoonoze, ki je široko razširjen tako med pticami kot prašiči (sodoprsti kopitarji). Gre za skupino virusov iz družine Orthomyxoviridae z visokimi hitrostmi mutacij. Gre za družino s segmentiranim ss(-)RNA genomom. Dva seva, ki okužita istega gostitelja, lahko reorganizirata genome, pri čemer lahko nastane nov sev s himerno strukturo genoma (Lauterbach in sod., 2018). Virusno ovojnico med drugim sestavljata proteina hemaglutinin (H) in nevraminidaza (N), ki obstajata v več variantah. Glede na kombinacijo variant hemaglutinina in nevraminidaze (HN) obstaja več virusnih serotipov. Ptice so prenašalke več serotipov, med njimi sta najbolj znana seva H5N1, ki se je pojavil leta 1997 v Hong Kongu, in H7N9, ki se je pojavil leta 2013 v Šanghaju (povzročitelj »ptičje gripe«), prašiči pa so znani prenašalci serotipov H1N1 (znan povzročitelj »prašičje gripe«), H3N2 in H1N2, ki so se leta 2011 in 2012 pojavili v ZDA (Bui in sod., 2017). Virus SIV (angl. »*Simian Immunodeficiency Virus*«) (lentivirus) je naslednji primer zloglasne zoonoze, odkrite pri šimpanzih. Virus je prednik človeškega virusa humane imunske pomanjkljivosti, virusa HIV (angl. »*Human Immunodeficiency Virus*«). Poleg omenjenih primerov lahko človeški patogeni na človeka preidejo preko vmesnega gostitelja, kot je prašič pri Nippah virusu, ali preko žuželk, kot je virus vročice denge ali virus zahodnega Nila (angl. »*West Nile Virus*«) (Mandl in sod., 2015). Vsa omenjena obolenja so izziv in aktualna tematika sodobne medicine.

1.1.6.3 Agrikolturni vidik

Sistem razmnoževanja virusov, ki spreminja genome gostiteljev in preusmerja energetske pretok novonastale virocelice v proizvodnjo virusnih strukturnih in nestrukturnih proteinov, ima lahko ključen vpliv tudi na globalno kmetijsko oskrbo. Kot gostitelji virusov lahko namreč nastopajo rastline in živali, ki so ključni za sodobno agronomijo – poljedelstvo, sadjarstvo, vinarstvo, čebelarstvo in živinorejo. Energetsko preusmerjanje celice v virocelico izrazi patogene simptome na živini, perutnini, čebelah, vinski trti, sadju in poljščinah.

Rastline so bile prvi znani gostitelji virusov od odkritja virusa tobačnega mozaika pred 120 leti (Rossnick in sod., 2015). Glede na deveto poročilo ICTV, ki je izšlo leta 2009, naj bi bilo znanih okrog 1000 vrst rastlinskih virusov. Gostitelji večine izmed njih so poljščine in krmne rastline. Torej se je rastlinska virologija fokusirala preferenčno na poljščine. Tipični primeri so virus TEV (angl. »*Tobacco Etch Virus*«) (Potyviridae), virus y krompirja PVY (angl. »*Potato Virus Y*«) (Potyviridae), CMV (angl. »*Cucumber Mosaic Virus*«) (Bromoviridae, Cucumovirus), virus cvetačnega mozaika CaMV (angl. »*Cauliflower Mosaic Virus*«) (Ortervirales, Caulimoviridae), LIYV (angl. »*Lettuce Infectious Yellow Virus*«) (Closteroviridae), BYDV (angl. »*Barley Yellow Dwarf Virus*«) (Luteoviridae), TYLCV (angl.

»Tomato Yellow Leaf Curl Virus«) (ssDNA, Geminiviridae), TSWV (angl. »Tomato Spotted Wilt Virus«) (Bunyaviridae), RDV (angl. »Rice Dwarf Virus«) (Reoviridae) in MMV (angl. »Maize Mosaic Virus«) (Mononegavirales, Rhabdoviridae) (Whitfield in sod., 2015). Številne raziskave so se ukvarjale z diverziteto, mehanizmom okužbe rastlin z rastlinskimi virusi in mehanizmom rastlinske obrambe pred virusnimi patogeni. Nekateri virusi, kot je CMV, so generalisti in okužujejo okrog 1200 rastlinskih vrst. Virus BYDV je omejen na floemsko tkivo rastline. Virus PEMV (angl. »Pea Enation Mosaic Virus«) (Luteoviridae) je segmentiran virus z dvema segmentoma: RNA1 in RNA2. RNA2 predstavlja himerni genom, ki izvira iz virusov skupine Umbravirus. MP protein (angl. »movement protein«) na RNA2 virusu omogoča, da ni omejen samo na floemsko tkivo. Virus tobačnega mozaika TMV se prenaša mehanistično preko poškodb rastlinskega tkiva. Metagenomske raziskave so pokazale tudi bogato raznolikost virusov pri »nekultiviranih« rastlinah, ki pa so običajno asimptomatske (Roossnick in sod., 2015; Whitfield in sod., 2015). Predvsem v okviru kultiviranih rastlin virusi znižajo hitrosti rasti rastlinskih populacij. Kultivirane rastline imajo namreč nižjo stopnjo populacijske variacije oz. predstavljajo bolj homogene populacije v smislu alelnih variant. Nekultivirane rastline so bolj heterogene z vidika alelnih variant, zato je viden manjši vpliv virusov na fiziologijo njihovih populacij. Obratno imajo lahko virusi tudi pozitiven, mutualističen vpliv na rastlinske gostitelje, saj s sprožanjem rastlinskih obrambnih mehanizmov povečajo in pospešijo rastlinske odgovore na biotski in abiotski stres. Toda virusi so z rastlinami povezani v ekološko interakcijsko mrežo, ki je dvosmerna. Tako kot virusi vplivajo na fiziologijo in populacijo rastlinskih gostiteljev, tudi rastlinski gostitelji lahko vplivajo na širjenje virusa. Invazivne vrste trav, ki so bolj robustne, pospešujejo širjenje virusa BYDV na native vrste. Ne glede na mehanizem razmnoževanja in vpliv virusa na rastlinsko ekologijo so rastline kot gostitelji sesilni organizmi, kar pomeni, da vektorji rastlinskih virusov igrajo ključno vlogo pri prenosu rastlinskih virusov med rastlinskimi gostitelji (Roossnick in sod., 2015; Whitfield in sod., 2015).

Poleg rastlin kot gostiteljev virusi igrajo ključno vlogo v fiziologiji in patologiji domačih oz. udomačenih živali. Med živalmi, na katere se močno zanaša sodobna hortikultura, so tudi čebele (*Apis mellifera*). Izvirajo iz vzhodne Azije (*A. cerana*) in so bile kot ključni opraševalci prenesene v Ameriko, Evropo in Afriko (Wilfert in sod., 2016). Svetovna zaloga udomačenih čebel raste bistveno počasneje od zahtev po opraševanju. Zato je ključnega pomena poznavanje njihovega viroma. Znanih je več čebeljih patogenih virusov, kot so virus akutne paralize čebel ABPV (angl. »Acute Bee Paralysis Virus«) (Dicistroviridae), virus kronične paralize čebel CBPV (angl. »Chronic Bee Paralysis Virus«), virus deformiranih kril DWV (angl. »Deformed Wing Virus«) (Iflaviridae), virus črnih matičnikov BQCV (angl. »Black Queen Cell Virus«) (Dicistroviridae), virus mešičkaste zalege SBV (angl. »Sacbrood virus«) (Iflaviridae), čebelji virus X BVX, čebelji virus Y BVY, filamentozni virus FV, virus motnih kril CWV (angl. »Cloudy Wing Virus«). Večinoma sodijo v red Picornavirales (ss(+)RNA virusi) (Hulo in sod., 2011 – ViralZone; NCBI Resource Coordinators, 2016; NCBI Resource Coordinators, 2017; NCBI – <https://www.ncbi.nlm.nih.gov/>). Virus deformiranih kril DWV (angl. »Deformed Wing Virus«) je iflavirus, ki ga na čebele prenaša ektoparazit varoja – zajedalska pršica (*Varroa destructor*). Definiran je bil kot ključni patogen, odgovoren za zimski pogin čebel. Virus se akumulira v črevesju varoje, ki ga injicira direktno v čebeljo hemolimfo. Invazije varoje so zato povezane z naraščanjem prevalence in virulence DWV virusa in hkrati mortalitete čebeljih kolonij (Martin in sod., 2012; Wilfert in sod., 2016). Poleg omenjenih virusov influence tipa A, ki okužujejo prašiče (Bui in sod., 2017), številni flavivirusi okužujejo govedo in prašiče, kot sta pestivirus BVDV (angl. »Bovine Viral Diarrhea Virus«) in CSFV (angl. »Classical Swine Fever Virus« ali »Hog Cholera Virus«) in nidovirus prašičjega reprodukcijskega in respiratornega sindroma PRRSV (angl. »Porcine Reproductive and Respiratory Syndrome Virus«) (Hulo in sod., 2011 – ViralZone; NCBI Resource Coordinators, 2016; NCBI Resource Coordinators, 2017; NCBI – <https://www.ncbi.nlm.nih.gov/>), kar ima znaten vpliv na pridelavo in predelavo prašičjega in govejega mesa. Tako kot pri rastlinskih virusih tudi pri živalskih virusih vektorji živalskih virusov igrajo ključno vlogo pri njihovi evoluciji.

1.1.6.4 Vektorji rastlinskih in živalskih patogenov

Tako rastlinski kot živalski virusi preko načina razmnoževanja in pretvorbe celice v virocelico spremenijo pretok metabolne energije in ga preusmerijo v izgradnjo lastnih komponent. S tem vplivajo na fiziologijo rastlinskih in živalskih gostiteljev.

Rastlinski gostitelji so sesilni organizmi, zato igrajo vektorji rastlinskih virusov pri njihovem prenosu med rastlinsko biosfero kritično vlogo. Večina znanih vektorjev rastlinskih virusov spada v red Hemiptera. Ključno vlogo pri njihovi učinkovitosti kot vektorji igra igli podobno bodalo, ki je sestavljeno iz maksil in mandibul (angl. »*mandibular and maxillar stylets*«). Znotraj skupine Hemiptera so pomembnejše skupine Aphidoidea (listne uši), Aleyrodoidea (moljevke), Psylloidea (bolšice), Cicadoidea (mali škrtatki) in Fulgoroidea. Poleg žuželk skupine Hemiptera so znani vektorji rastlinskih virusov tudi redovi Thysanoptera (resarji), Coleoptera (hrošči), Acari (pršice) in Nematoda (gliste) in glive. Omenjeni organizmi po različnih mehanizmi prenašajo rastlinske viruse. Po prvem, neperzistentnem (angl. »*non-persistent*«) mehanizmu žuželka izgubi virus nekaj minut po pridobitvi, zato lahko le v tem časovnem intervalu okuži rastlino. Pri omenjenem mehanizmu virusi načeloma potrebujejo proteinski aparat, ki jim omogoča, da se pritrdijo na maksile vektorja. Virus CaMV (angl. »*Cauliflower Mosaic Virus*«) (Ortervirales, Caulimoviridae) se prenaša neperzistentno z listnimi ušmi preko rastlinskih transmissijskih telesc (angl. »*transmission bodies*«) – organelov znotraj rastlinskih celic. Ob hranjenju postanejo aktivirana in listna uš privzame virus in prenese na drugo rastlino (Roossinck, 2015). Virus CMV (angl. »*Cucumber Mosaic Virus*«) (Bromoviridae, Cucumovirus) se prenaša preko listnih uši. Listne uši privlačijo okužene rastline, ki po negativni povratni zanki ob začetku hranjenja začnejo proizvajati molekule, ki preprečujejo njihovo hranjenje. Mehanizem omogoča hitrejše širjenje virusa. Drugi mehanizem je perzistentni mehanizem, pri katerem lahko žuželka rastlino inokulira z virusom skozi celotni življenjski cikel, tudi po levitvi. Mehanizem se deli na dva podtipa – pri prvem se virusi ne razmnožujejo znotraj vektorja (angl. »*persistant circulative*«), pri drugem se virusi razmnožujejo znotraj vektorja (angl. »*persistent propagative*«). Virusi, ki se tipično ne razmnožujejo znotraj gostitelja, so virusi skupine Luteoviridae, Nanoviridae in Geminiviridae. Omenjene skupine imajo ikozahedralno kapsido brez virusne ovojnice. Geminiviruse prenašajo moljevke. Virusi, ki se razmnožujejo znotraj vektorja, večinoma spadajo v skupine Bunyavirales, Rhabdoviridae, Reoviridae in Tymovirales (Hogenhout in sod., 2008). Skupina Bunyavirales se prenaša z resarji. Resarji poškodujejo rastlinsko tkivo, kar sproži nastanek molekul, ki preprečujejo njihovo hranjenje. Tretji mehanizem je delno perzistentni, pri katerem žuželka lahko virus prenese v rastlino nekaj ur ali dni po pridobitvi. Virus »se izgubi« ob levitvi (Hogenhout in sod., 2008; Whitfield in sod., 2015; Roossinck, 2015).

Po nekoliko drugačnem principu vektorji živalskih virusov prenašajo živalske viruse. Najbolj znani vektorji živalskih virusov so žuželke iz reda Diptera – komarji (Culicidae), vinske mušice (*Drosophila spp.*) in pajkovci (Chelicerata), kot so klopi (Ixodida) (Li in sod., 2015; Hackenberg in Kotsyfakis, 2018; Shi in sod., 2018b; Xia in sod., 2018). Pri klopah naj bi ključno vlogo pri prenosu virusov igrali eksosomi. Eksosomi so izvenselični vezikli, ki se sprostijo iz celičnih multivezikularnih telesc v klopno slino iz klopne slin v kri končnega živalskega gostitelja (Hackenberg in Kotsyfakis, 2018). Virusne skupine, ki se prenašajo na živalske gostitelje, so Bunyavirales, Rhabdoviridae, Orthomyxoviridae, Flaviviridae (Shi in sod., 2018b). Komarji so znani prenašalci človeških zoonoz, kot so virus vročice denge (angl. »*Dengue virus*«) (Flaviviridae), virus zahodnega Nila (angl. »*West Nile virus*«) (Flaviviridae), virus rumene mrzlice (angl. »*Yellow fever virus*«) (Flaviviridae), Zika virus (Flaviviridae) in Chikungunya virus (Togaviridae). Večina virusov, ki se prenašajo preko komarjev, ravno tako sodi v skupino Bunyavirales, Flaviviridae in Reoviridae, poleg omenjenih pa tudi Togaviridae in Nidovirales – Mesoniviridae (Xia in sod., 2018). Nekateri izmed virusov, ki so specifični za komarje, pa so patogeni za komarje in se ne razmnožujejo v vretenčarjih (Xia in sod., 2018).

V večini dosedanjih raziskav so členonožci (Arthropoda) igrali ključno vlogo kot vektorji rastlinskih in živalskih virusov. Večina raziskav se je nanje usmerila kot na center raznolikosti virusov. RNA virusi izmed vseh skupin, definiranih po Baltimoru, predstavljajo večinski delež rastlinskih in živalskih virusov. Omenjeno prepričanje pa je lahko zmotno, saj številne ostale skupine nevretenčarjev še vedno niso bile raziskane v kontekstu raznolikosti RNA virusov. Preden se usmerimo na samo raznolikost in razširjenost RNA virusov, se moramo usmeriti na lastnosti RNA virusov, ki so ključni faktorji njihove raznolikosti, to sta njihove genomske strukture in njihov mehanizem razmnoževanja.

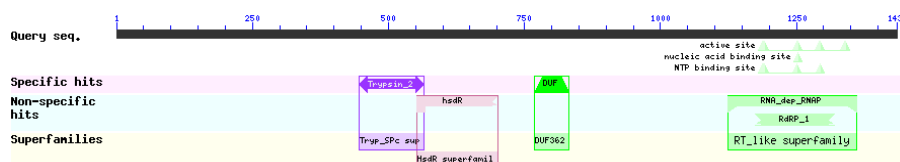
1.2 RNA virusi

1.2.1 Strukturna raznolikost RNA virusov

RNA virusne skupine lahko uvrstimo v 25 filogenetskih skupin oz. klad glede na izvor in skupne karakteristike: 1) ss(+)RNA: Astro, Picorna-Calici, Hepe-Virga, Narna-Levi, Tombus-Noda, Luteo-Sobemo, Nido, Permutotetra, Weivirus, Zhaovirus, Yanvirus, Flavi, in Poty 2) ss(-)RNA: Bunya-Arena, Ophio, Orthomyxo, Mono-Chu, Yuevirus in Qinvirus 3) dsRNA: Birna, Partiti-Picobirna, Toti-Chryso, Reo, Hypo in Cysto (Shi in sod., 2016a). Vsaka klada vsebuje več po ICTV definiranih virusnih skupin poleg t. i. »plavajočih skupin« (angl. »floating genera«), ki niso uvrščene znotraj definiranih virusnih skupin, vendar glede na uporabljene kriterije klasifikacije sodijo v obravnavano klado.

1.2.1.1 Klada Astro

Astroviridae so sferični vretenčarski virusi brez virusne ovojnice z linearnim monopartitnim ss (+) RNA genomom, dolgim od 6,8–7 kb. Enodelni oz. monopartitni (angl. »monopartite«) genom je sestavljen iz treh ORF-jev. Prva dva ORF-ja predstavljata jedrni nestrukturirni del astrovirusov. Kodirata tripsinu podobno peptidazo (angl. »trypsin-like peptidase«) in RdRp. Sta v različnih bralnih okvirjih. Klasični, dobro okarakterizirani astrovirusi vsebujejo med prvima dvema ORF-jema motiv tipične strukture X XXY YYZ (angl. »slippery sequence«), ki ji 5–9 nukleotidov na 3' sledi psevdovozel (angl. »pseudoknot«). Pri astrovirusih je omenjen signal ...C AAA AAA C... . Ko RNA polimeraza prepozna omenjen signal, se bodisi vrne en nukleotid nazaj ali pa ga preskoči. S tem preide v isti bralni okvir, kot je drugi ORF, in nadaljuje prepisovanje. ORF-ja se tako prepiseta kot monocistronska mRNA in en poliprotein. Pojav se angleško imenuje »ribosomal frameshifting«. Alternativne verzije nestrukturnega dela genoma se lahko prepisujejo kot dva ORF-ja v različnih bralnih okvirjih. Nestrukturirni del zajema dve klasični domeni: peptidazna domena (tripsinu podobna peptidazna domena (klan PA) in peptidaza S32, kot je npr. pri Changjiang astro-like virusu) in sledeča RdRp domena (**Slika 1.3**).



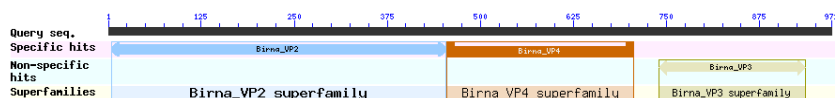
Slika 1.3: Genomska arhitektura nestrukturnega dela astrovirusov. Trp_SPC – tripsinu-podobna serin proteaza; DUF – domena z neznano funkcijo (angl. »domain of unknown function«); RdRp_1 – od RNA odvisna RNA polimeraza družine 1, »hsdR superfamily« – naddružina proteinov hsdR (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD).

Strukturirni del predstavlja tipično en ORF (tretji ORF), ki kodira zapis za kapsidni protein. Prepisuje se kot subgenomska RNA. Tekom evolucije je bil omenjen ORF vpel v mrežo horizontalnih prenosov med različnimi virusi ob prisotnosti na istem gostitelju. Tipični domeni za astrovirusne kapsidne proteine sta: »astrovirus capsid protein precursor« in »turkey astrovirus capsid protein«. Omenjeni proteinski domeni pa sta pri astrovirusih nadomestili domeni: »peptidase family A21 viral coat protein«, tipičen za permutotetraviruse, ali pa »viral coat protein (S domain)«, ki so ju astrovirusi prevzeli od tombusvirusov in sobemovirusov (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD).

Shi in sod. (2016a) so odkrili astroviruse, pri katerih med peptidazno in RdRp proteinsko domeno ni premika bralnega okvirja. V nekaterih primerih je tripsinu podobna peptidaza nadomestila peptidazo S32, tipično za skupino Nidovirales. Proteini plašča, ki so tipični za astroviruse, so nadomestili protein plašča družine peptidaza A21 (angl. »peptidase A21 family coat protein«), razširjen pri Alphatetraviridae, in virusni protein plašča (angl. »viral coat protein; S domain«), razširjen pri rastlinskih virusih.

1.2.1.2 Klada Birna

Birnaviridae so segmentirana virusna skupina. Večinoma so vretenčarski virusi, le Entomobirnavirus je nevretenčarska skupina birnavirusov. So sferične oblike. Genom z dolžino 5,6–6,5 kb je sestavljen iz dveh segmentov, segmenta A in B. Segment A predstavlja strukturni segment z enim ORF, ki kodira en poliprotein, sestavljen iz treh strukturnih delov: VP2, VP3, VP4 (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD) (**Slika 1.4**). VP4 protein je proteaza, ki ob maturaciji poliprotein proteolitsko razcepi, pri čemer nastaneta še VP2 in VP3 gradbeni enoti virusne kapside.

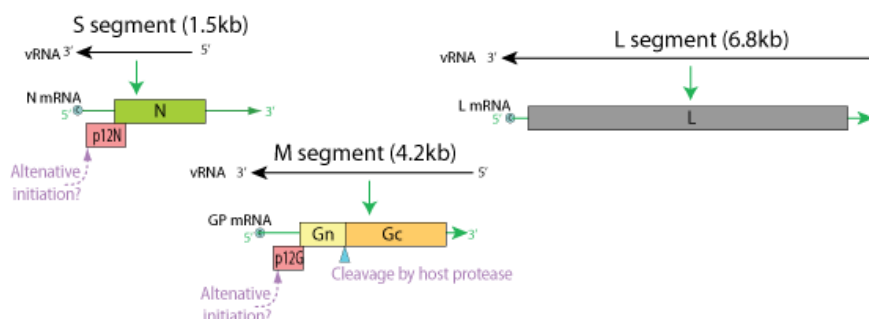


Slika 1.4: Segment A skupine Birnaviridae (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD). Angleški izrazi; Superfamily – naddružina.

Segment B je nestrukturni segment z enim ORF, ki kodira zapis za RdRp (VP1 protein).

1.2.1.3 Klada Bunya-Arena

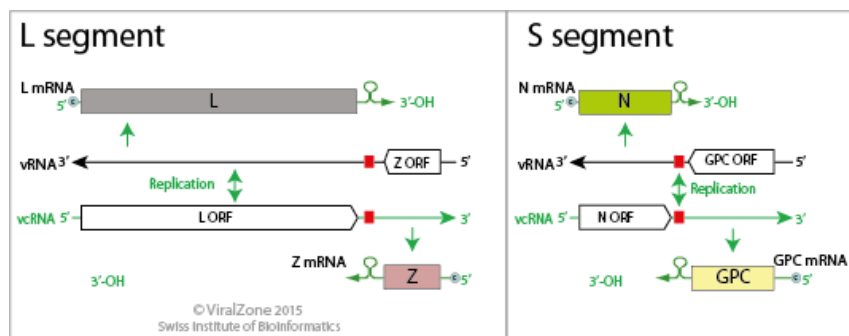
Klada Bunya-Arena je sestavljena iz dveh po ICTV prepoznanih virusnih skupin: reda Bunyvirales in družine Arenaviridae. Red Bunyvirales je sestavljena iz virusnih družin: Feraviridae, Fimoviridae, Hantaviridae, Jonviridae, Nairoviridae, Phasmaviridae, Phenuiviridae, Peribunyaviridae in Tospoviridae (Hulo in sod., 2011 – ViralZone). Genomi klade Bunya-Arena so segmentirani in vsebujejo dva, tri ali več segmentov. Pri virusih, ki so tipično sestavljeni iz treh segmentov, predstavlja segment S majhen segment, dolg okrog 1,5 kb, segment M srednji segment, dolg okrog 4 kb, in segment L največji segment, dolg okrog 7 kb (**Slika 1.5**). Segment L zajema ORF, ki kodira zapis za RNA polimerazo. mRNA segmenta L je lahko bicistronska in kodira zapis za L protein. Gre za endonukleazno domeno na N-terminalnem delu L segmenta. Omenjen segment je soroden L segmentu skupine Mononegavirales, Ophioviridae, Qinvirusu in Yuevirusu (Shi in sod. (2016a)). Sledi segment M, monocistronska RNA, ki kodira bodisi poliprotein z dvema glikoproteinskima domenama bodisi samo eno glikoproteinsko domeno. Segment S je monocistronska RNA, ki kodira zapis za nukleoprotein. Pri nekaterih virusih je nestrukturni protein kodiran z ORF v drugem bralnem okvirju v nasprotni smeri (Rift valley fever virus). Nekateri bunyavirusi so tekom evolucije pridobili dodatne segmente (Shi in sod., 2016a). Pri družini Fimoviridae je dodaten segment neokarakteriziran (Hulo in sod., 2011 – ViralZone).



Slika 1.5: Primer strukture virusov reda Bunyvirales – Feraviridae (Hulo in sod., 2011 – ViralZone). Angleški izrazi: Cleavage by host protease – cepitev z gostiteljsko proteazo; alternative initiation – alternativna iniciacija.

Skupina Arenaviridae ima segmentiran genom sestavljen iz dveh segmentov; segmenta S in segmenta L. Segment S je dolg 3,5 kb in kodira nukleoprotein, segment L je dolg 7,5 kb

in kodira RdRp (**Slika 1.6**). Oba segmenta se lahko prepisujeta v nasprotnih okenskih okvirjih, torej 5'–3' in 3'–5'. Na L segment se v obratni smeri prepisuje »RING finger protein Z«. Na L segment se v obratni smeri prepisuje glikoprotein (GPC).

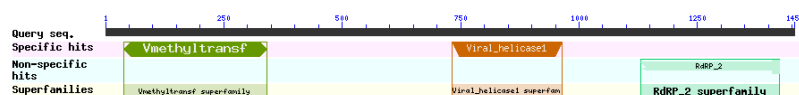


Slika 1.6: Primer strukture virusov reda Arenaviridae (Hulo in sod., 2011 – ViralZone).

1.2.1.4 Klada Hepe-Virga

Hepe-Virga je strukturno raznolika klada, ki zajema virusne družine: Hepeviridae, Virgaviridae, Togaviridae, Bromoviridae, Closteroviridae, Idaeovirus, Negevirus, Endornaviridae, Alphatetraviridae, Benyviridae in red Tymovirales, ki zajema družine: Alphaflexiviridae, Betaflexiviridae, Gammaflexiviridae in Tymoviridae.

Vsem predstavnikom klade Hepe-Virga je skupno prepisovanje nestrukturnega poliproteina, ki predstavlja »jedrni poliprotein klade«. Nestrukturni poliprotein vsebuje tri ključne domene, ki so ohranjene pri vseh predstavnikih klade Hepe-Virga. Na N-terminalnem delu je virusna metiltransferaza alfavirusnega tipa (angl. »*viral methyltransferase*« ali »*alphavirus-like methyltransferase*«) (pfam01660). Protein je vključen v dodajanje kape mRNA, ki je razširjena pri virusih alfavirusnega tipa (Koonin in sod., 2014, Koonin in sod., 2015): Tymovirales, Hordeivirus, Tobravirus, Tobamovirus, Bromovirus, Closterovirus, Virgaviridae, Hepeviridae, Hepevirusom podobni virusi, Togaviridae, Benyviridae, Idaeovirus, Negevirus, Benyviridae in Calicivirus. Drugi protein, skupen vsem virusom Hepe-Virga klade, je virusna helikaza naddružine 1 (angl. »*Superfamily 1 helicase*«, SF1) (pfam01443) (angl. »*viral helicase*«). Tretji protein (na C-terminalnem delu), skupen kladi, je RdRp družine 2 (RdRp_2) (**Slika 1.7**).



Slika 1.7: Struktura poliproteina, tipičnega za Hepe-Virga klado (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD). Okrajšave: Vmethyltransf – virusna metiltransferaza, Viral_helicase1 superfam – virusna helikaza naddružine 1, RdRp_2 – od RNA odvisna RNA polimeraza družine 2.

Poleg omenjenih skupnih proteinskih domen so v genomu vsake od virusnih skupin prisotne tudi specifični virusni kladi - lastne proteinske domene.

Družina Endornaviridae ima najenostavnejšo genomsko strukturo. Ima enodelen (monopartitni) genom, ki kodira poliprotein z dvema proteinskima domenama; virusno helikazo in RdRp_2.

Red Tymovirales je red rastlinskih in glivnih virusov. Ima monopartitni genom z dvema ali več ORF-ji. Na drugem ORF-ju se nahajajo domene, kot so MP protein (angl. »*movement protein*«) ali plaščna proteina »Trichovirus coat protein« oz. »Tymovirus coat protein«. Domena »Trichovirus coat protein« se lahko nahaja tudi na istem ORF-ju kot virusna metiltransferaza, virusna helikaza in RdRp. Eden izmed ORF-jev lahko kodira zapis za strukturni protein SP2.

Družina Hepeviridae ima na enodelnem genomu dva ali več ORF-jev. Na glavnem nestrukturiranem poliproteinu ima poleg skupnih domen kodirano tudi Macro domeno (angl. »*Appr-1 processing enzyme*«), modul, sestavljen iz 180 aminokislin, ki veže ADP ribozo, peptidazo C41 (cistein proteaza) (ki se nahajata na glavnem nestrukturiranem poliproteinu) in SP2 protein (angl. »*structural protein 2*«), ki naj bi imel vlogo pri enkapsidaciji virusov.

Družine Benyviridae, Bromoviridae, Virgaviridae in Closteroviridae so tako kot red Tymovirales skupina rastlinskih virusov, zato so v njihov genom vključeni za rastlinske virusne tipični MP proteini (angl. »*plant movement protein*«). Družini Benyviridae in Bromoviridae imata segmentiran genom, kjer je nestrukturiran poliprotein z metiltransferazo, helikazo in RdRp kodiran na ločenem segmentu. Pri družini Benyviridae se na glavnem nestrukturiranem poliproteinu nahaja tudi peptidaza C36. Dodatni segmenti pa pri obeh družinah kodirajo zapis za različne proteine oz. družine proteinov: protein, odgovoren za premikanje virusov – MP protein naddružine 3A (angl. »*movement protein*«), »Bromovirus coat protein« (Bromoviridae), protein plašča družine »TMV-like« (angl. »*TMV-like coat protein*«), rastlinski MP protein (angl. »*plant viral movement protein*«), 14 kDa velik protein ali protein P25/P26, tipičen za Benyviridae. Družina Virgaviridae ima lahko monopartitni ali segmentiran genom. Dodatne domene tukaj v glavnem predstavljajo MP proteini naddružine MP, naddružine 3A protein plašča družine »TMV-like« (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD). Družina Closteroviridae na monopartitnem genomu kodira zapis za več ORF-jev, na katerih so domene za HSP70, HSP90 protein, »Closterovirus coat protein« in »RNA silencing supressor P21«. Na glavnem nestrukturiranem poliproteinu se N-terminalno od virusne metiltransferaze nahaja peptidaza C42.

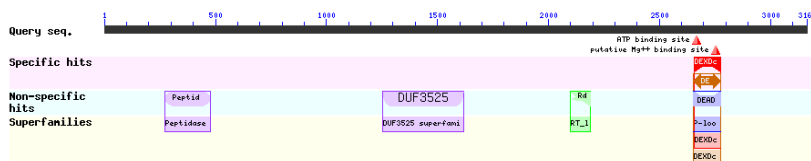
Družina Togaviridae ima poleg Macro domene na glavnem nestrukturiranem poliproteinu (skupaj s helikazo, RdRp in virusno metiltransferazo), ki je razširjena tudi pri Hepeviridae, v genomu kodiran tudi dodaten ORF s proteinskimi domenami, kot so peptidaza S3 in alfavirusni glikoproteini E1, E2 in E3.

Skupina Negevirus se deli na štiri podtipa. Klaster I negevirusov zajema 3 ORF-je. Prvi ORF vključuje poleg glavnih domen (helikaza, virusna metiltransferaza in RdRp) tudi FtsJ metiltransferazo, ki jo lahko srečamo pri skupini Flavivirida in Flavivirusom podobnih virusih. Drugi ORF kodira DiSB – ORF glikoprotein, tipičen za negevirus, in tretji ORF kodira strukturni protein SP24, ki je tipičen za negevirus. Klaster II se od klastera I razlikuje samo po divergentnem ORF II. Najverjetneje gre za DiSB-ORF2 glikoproteinu soroden glikoprotein, vendar CDD ne zazna domene. Klaster III ima štiri ORF-je, klaster IV pa pet ORF-jev.

V drevesu rastlinski Virgaviridae tvorijo svoj klaster, rastlinski Bromo-, Clostero- in Idaeovirusi klastirajo skupaj z vretenčarskimi togavirusi, rastlinski dsRNA virusi skupine Endornaviridae, Tymovirales in Benyviridae pa klastirajo skupaj z vretenčarskimi Hepeviridae.

1.2.1.5 Klada Hypo

Hypoviridae so dsRNA virusi brez kapside z linearnim enodelnim genomom, dolgim 9–13 kb, ki kodira enega ali dva ORF-ja. Geni skupine Hypoviridae so samo nestrukturirani geni. ORF A (če je prisoten) kodira zapis za peptidazo C7, ORF B kodira zapis za nestrukturiran poliprotein, ki na N-terminalnem delu sestavlja peptidazo C8, sledi RdRp in na C-terminalnem delu DEAD/DEAH tip helikaze (naddružina 2), tipičen za skupini Flaviviridae in Potyviridae (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD) (**Slika 1.8**).



Slika 1.8: ORF B skupine Hypoviridae: N-terminalni del proti C-terminalnemu delu: Peptidaza C8- DUF- RdRp- Helikaza DEAD/DEAH (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD).

1.2.1.6 Klada Luteo-Sobemo

Luteo-Sobemo je klada, ki zajema pet virusnih skupin: 1) Luteoviridae – Luteovirus, Enamovirus, Ploverovirus, 2) Sobemovirus in 3) Barnaviridae in 4) Carascovirus. Skupina Barnaviridae je skupina glivnih virusov, medtem ko so ostale skupine rastlinskih virusov.

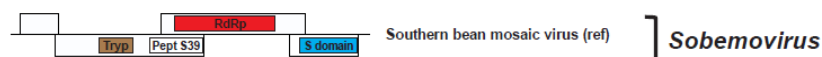
Luteoviridae je družina z linearnim monopartitnim genomom velikosti od 5,3 do 5,7 kb. Jedrni del genoma skupine Luteoviridae predstavljajo štirje zaporedni ORF-ji. Med prvima dvema se nahaja zaporedje tipa X XXY YYZ s sledečim psevdovozlom 5–9 nukleotidov proti 3'. RNA polimeraza se ob zaznavanju sekvenčnega signala vrne en nukleotid nazaj ali preskoči en nukleotid. Pri tem preideta oba ORF na isti bralni okvir in se prepiseta kot ena mRNA in prevedeta kot en poliprotein (angl. »*ribosomal frameshifting*«). Prvi ORF kodira peptidazo, kot je peptidaza S39 ali »Luteovirus RNA polymerase P1-P2/replicase«, drugi ORF pa kodira RdRp. Naslednja dva ORF-ja sta v istem bralnem okvirju in se prepisujeta kot policistronska mRNA. Stop kodon za tretjim ORF delno prepozna tRNA, zato se tretji in četrti ORF prevedeta kot en protein (angl. »*suppression of termination*« ali »*stop codon read-through*«). Tretji ORF kodira zapis za luteovirusni protein plašča (angl. »*luteovirus coat protein*«), četrti pa za »potato leaf roll virus readthrough protein« kot podaljšek luteovirusnih proteinov kapside. Pri Luteovirusu in Ploverovirusu se s tretjim ORF-jem prekriva tudi Vpg gen (angl. »*luteovirus putative Vpg genome linked protein*«). mRNA je pri omenjenih skupinah zato na N-terminalnem delu prekrita z Vpg proteinom kot nadomestkom »cap« proteina. Slika predstavlja reprezentativno strukturo Luteoviridae (Slika 1.9).



Slika 1.9: Struktura Luteoviridae. Luteo P0 – luteovirusni P0 protein, Pept_S39 – peptidaza S39, RdRp (od RNA odvisna RNA polimeraza), Luteo_Ct-»luteovirus coat protein« (Shi in sod., 2016a).

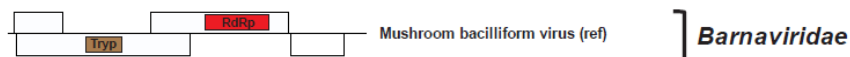
Carascovirusi kodirajo štiri ORF-je. Prvi ORF, ki kodira nestrukturni poliprotein, vsebuje tripsinu podobno peptidazo (angl. »*trypsin-like peptidase*«) in RdRp, zadnji ORF pa kodira »viral coat protein (S domain)«.

Sobemovirusi imajo monopartitno RNA s štirimi ORF. Drugi ORF kodira dve peptidazni domeni – peptidazo S39 in »tripsinu podobno peptidazo« (angl. »*trypsin-like peptidase domain*«), tretji kodira RdRp, zadnji pa »viral coat protein (S domain)« (Slika 1.10).



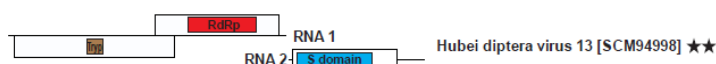
Slika 1.10: Struktura Sobemovirusov. Pept_S39- peptidaza S39, RdRp (od RNA odvisna RNA polimeraza), S domain – »viral coat protein (S domain)«, Tryp- tripsinu podobna peptidaza (angl. »*trypsin-like peptidase*«) (Shi in sod., 2016a).

Barnaviridae imajo monopartitno ss(+) RNA s štirimi ORF, pri čemer je prvi ORF nekarakteriziran in se prekriva z drugim. Drugi ORF kodira zapis za tripsinu podobno peptidazo in tretji za RdRp (**Slika 1.11**).



Slika 1.11: Struktura Barnaviridae. RdRp (od RNA odvisna RNA polimeraza), Tryp – tripsinu podobna peptidaza (angl. »trypsin-like peptidase«) (Shi in sod., 2016a).

Shi in sod. (2016) so odkrili primere segmentiranih virusov, ki sodijo v Luteo-Sobemo klado (**Slika 1.12**).

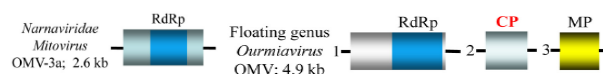


Slika 1.12: Segmentiran virus – Luteo-Sobemo klada. RdRp (od RNA odvisna RNA polimeraza), Tryp – tripsinu podobna peptidaza (angl. »trypsin-like peptidase«), S domain – »viral coat protein (S domain)« (Shi in sod., 2016a).

Skupina Shi in sod. (2016) je odkrila nove luteoviruse z drugačno strukturo genomske organizacije kot pri tipičnih, prvotno znanih luteovirusih. Nekateri so bili segmentirani, kot so »Wuhan house centipede virus 4« (WHYY19909), »Hubei diptera virus 13« (SCM94998), »Xingshan nematode virus 5« (XSNXC27943) in »Wenzhou shrimp virus 9« (shrimp6189). Pri omenjenih primerih virusov prvi segment zajema peptidazno in RdRp domeno na ločenih ORF-jih. Drugi segment zajema protein plašča. Nekateri genomi so monopartitni in zajemajo peptidazno in RdRp domeno na istem ORF-ju in protein plašča na ločenem ORF. Primera sta »Changjiang sobemo-like virus 2« (CJLX29674) in »Beihai sobemo-like virus 8« (BWBFG36635). »Changjiang sobemo-like virus 2« (CJLX29674) na drugem ORF vsebuje virusni protein plašča družine S domene (angl. »viral coat protein; S domain«). »Beihai sobemo-like virus 8« (BWBFG36635) na drugem ORF vsebuje alvernavirusom homologen protein plašča (angl. »alvernaviridae coat-like«). Odkrili so tudi primere virusov, ki so imeli peptidazno domeno in RdRp domeno na ločenih zaporednih ORF-jih, kot je na primer pri »Beihai sobemo-like virusu 1« (BHZY60709). Virus ima na treh zaporednih ORF niz domen: »trypsin-like peptidase« – »RdRp« – »viral coat protein S domain«. Znotraj skupine so zaznali tri različne družine plaščev; luteovirusni protein plašča (angl. »luteovirus coat protein«), virusni protein plašča (domena S) (angl. »viral coat protein S domain«) in protein plašča, tipičen za alvernaviruse (angl. »alvernavirus coat-like protein«). Zaznani sta bili tudi dve različni peptidazni domeni; tripsinu podobna peptidazna domena in peptidaza S39 (Shi in sod., 2016a).

1.2.1.7 Klada Narna-Levi

Klada Narna-Levi vsebuje dve klasificirani virusni družini: Narnaviridae in Leviviridae. **Družina Narnaviridae** se deli na tri funkcionalne podskupine: Narnavirus, Mitovirus in neklasificirano skupino (angl. »floating genus«) Ourmiavirus. Narnavirusi in mitovirusi so elementi brez kapside s ss(+)RNA genomom dolžine med 2,5 in 2,9 kb. Genom kodira en funkcionalen protein, RdRp (od RNA odvisno RNA polimerazo), ki predstavlja jedro vseh RNA virusov (**Slika 1.13**). Skupino Ourmiavirus sestavljajo trije segmenti z enim ORF. Prvi segment predstavlja gen za RdRp, drugi segment gen za protein kapside (angl. »capsid protein«) in tretji segment protein, vključen v gibanje virusa, MP protein (angl. »movement protein«) MP naddružine (Hulo in sod., 2011 – ViralZone; Koonin in Dolja, 2014).



Slika 1.13 Struktura narnavirusov in mitovirusov (Koonin in Dolja, 2014). Okrajšave: CP – protein plašča, MP – protein MP (angl. »movement protein«). Angleški izraz: Floating genus – »plavajoči rod« oz. neklasificirana virusna skupina.

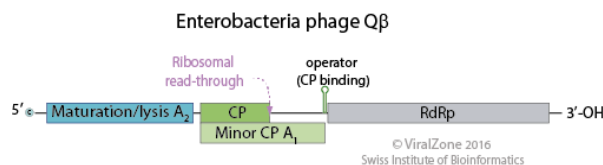
Skupina Shi in sod. (2016a) je odkrila nove narnaviruse s kompleksnejšo genomsko strukturo. Odkrili so primere narnavirusov, ki kodirajo ORF-je, ki se berejo v nasprotnih straneh (5'-3' in od 3'-5'), kot so: »Hubei mosquito virus 3« (3mos6141), »Beihai narna-like virus 24« (BHXG26812) in »Hubei narna-like virus 21« (QCM13322). V filogenetskem drevesu Shi in sod. (2016a) se omenjeni virusi ne grupirajo v isto skupino. Skupina je odkrila tudi primere narnavirusov, ki imajo poleg RdRp domene na ločenem ORF-ju dodatne proteinske domene, kot so: »alvernavirusom homologen protein plašča« (angl. »alvernavirus-like coat protein«) – »Beihai narna-like virus 10« (BHZY60512) in »virusni protein plašča (domena S)« (angl. »viral coat protein; S domain«) – »Changjiang narna-like virus 2« (CJLX29055). Odkrili so tudi primere narnavirusov z dodatno helikazno domeno na nestrukturiranem ORF-ju z RdRp – »Wenzhou shrimp virus 10« (shrimp12509).

Za rod Narnavirus je bila prvotno znana njihova razširjenost izključno pri nižjih evkariontih in glivah. Podvojujejo se v citoplazmi sliv in oomicet. *Phytophthora infestans* je patogen krompirja in narnavirus »Phytophthora infestans RNA virus 4« se uporablja kot agent za njegovo biološko kontrolo. *Leptomonas seymouri* je evkariont, ki sodi v deblo Euglenozoa (bičkarji). Je bližnji sorodnik *Leishmania spp.* in vrteljca (Trypanosoma). »Leptomonas seymouri narna-like virus 1« okužuje glivo *Leptomonas seymouri*. *Fusarium poae* je rastlinska patogena gliva, ki napada žito in je gostitelj »Fusarium poae narnavirusa«. Saccharomyces 20S RNA narnavirus in Saccharomyces 23S RNA narnavirus pa napadata glivo *Saccharomyces cerevisiae*.

Mitovirusi so, podobno kot narnavirusi, razširjeni med glivami, kot so *Macrophomina phaseolina*, *Buergeriula spartinae*, *Cryphonectria parasitica*, *Sclerotinia sclerotium* idr. Gre za primere rastlinskih patogenih gliv. Znano je, da se mitovirusi podvojujejo znotraj mitohondrijev gliv. Skupina Shi in sod. (2016a) je dokazala njihovo bogato razširjenost pri rakah in mehkužcih in v nekoliko manjšem obsegu tudi pri žuželkah.

Za Ourmiaviruse, v nasprotju z Narnavirusi in mitovirusi, je bila znana njihova razširjenost pri rastlinah. K evoluciji mitovirusov naj bi prispevala koevolucija evkariontov iz arhejam podobnega skupnega prednika (endosimbiontska teorija) in mitovirusov iz RNA bakterifagu podobnega skupnega prednika. Prokariontski prednik mitohondrija naj bi bil gostitelj RNA – bakteriofagu podobnega prednika mitovirusom podobnih virusov. Po endosimbiontski teoriji naj bi se prokariontski prednik »vključil« v arhejam podobnega prednika evkariontov. Tekom evolucije prokariontskega endosimbionta kot »gostitelja RNA bakteriofagov« naj bi se odvijala koevolucija »RNA bakteriofaga« v mitovirusom podobne viruse. Narnavirusi so prekopili na podvojevanje v citoplazmi. Večina evkariontov, ki jih okužuje skupina Narnaviridae so rastlinski patogeni, torej povezani z glivo in z rastlino, ki jo gliva napada. Predniki današnjih Ourmiavirusov naj bi s pomočjo horizontalnega prenosa genov za MP protein (angl. »movement protein«) in CP protein (angl. »coat protein«) od Tombusvirusov tekom evolucije prekopili iz glivnih na rastlinske gostitelje (Koonin in sod., 2014).

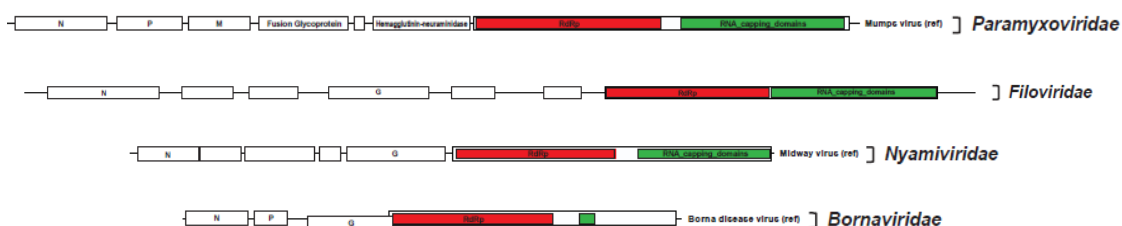
Levivirusi so eni izmed redkih RNA bakteriofagov (Daehnel in sod., 2005). Njihov genom je enodelen, dolg 4 kb in sestavljen iz maturacijskega proteina mat A₂ (angl. »maturation/lysis protein A₂«). Sledi protein kapside, ki se lahko prevaja v dveh alternativnih verzijah z istim začetkom (CP in CP A₁). Drugo verzijo predstavlja protein s podaljšanim C terminalnim delom. Stop kodoni namreč niso prepoznani s sprostitvenimi dejavniki (angl. »release factor«), ampak delno s tRNA, kar nadaljuje prevajanje. Pojav se imenuje »zaviranje terminacije« (angl. »suppression of termination« ali »stop codon read-through« ali »ribosomal read-through«). Na C-terminalnem delu sledi RdRp (Slika 1.14).



Slika 1.14: Struktura Leviviridae - primer faga Q β enterobakterij (angl. »enterobacteria phage Q β «) (Hulo in sod., 2011 – ViralZone). Angleški izrazi: Maturation/lysis A2 – maturacijskega proteina mat A2, Minor CP A1 – daljša verzija proteina plašča CP, ki se prevede s pojavom zaviranja terminacije, Ribosomal read-through – zaviranje terminacije.

1.2.1.8 Klada Mono-Chu

Klada Mono-Chu je skupina ss(-)RNA virusov, ki zajema neklasificirano skupino Chuvirus (Li in sod., 2015) in klasificirane virusne družine reda Mononegavirales: Bornaviridae, Rhabdoviridae, Filoviridae, Mymonaviridae, Nyamiviridae, Paramyxoviridae, Pneumoviridae in Sunviridae. Velikost njihovega linearnega enodelnega genoma se giblje od 10 do 20 kb. Sestavljen je iz niza ORF-jev, ki se prepisuje v treh izmed šestih možnih bralnih okvirjev. Jedrni del genoma zajema niz običajno neprekrivajočih ORF-jev, ki od N-terminalnega proti C-terminalnemu delu kodirajo: 1) Nukleoprotein ali nukleokapsidni protein (N) (angl. »*nucleocapsid protein*«) – z virusno genomsko RNA in virusno kapsido tvori ribonukleokapsido, 2) fosfoprotein (P), 3) matriks protein (angl. »*matrix protein*«) – virulenčni protein, 4) glikoprotein (G) – sodeluje pri odstranjevanju virusne ovojnice in 5) ORF L, ki kodira zapis za »mRNA capping region«, RdRp, in »mRNA capping enzyme – paramyxovirus family«. »mRNA capping enzyme« in »mRNA capping region« sta združena kot domena, odgovorna za dodajanje kape (angl. »*RNA capping domain*«). L domena z RdRp in aparatom, vključenim v generiranje kape, torej predstavlja jedro virusov klade Mono-Chu (Shi in sod., 2016a). ORF z L poliproteinom pri skupini Mono-Chu je soroden segmentu L pri skupini Ophioviridae in Yuevirusu. Nekateri izmed ORF-jev (kot so glikoproteini in matriks proteini, so tekom evolucije lahko izgubili značilne motive, zato lahko niso prepoznani (Slika 15). Poleg omenjenih proteinov lahko virusi klade Mono-Chu kodirajo zapis za fuzijski protein (angl. »*fusion protein*«), SH protein (angl. »*small hydrophobic protein*«), dodatne matriks proteine (angl. »*matrix protein*«), hemaglutinin nevraminidazo (angl. »*hemagglutinin neuraminidase*«), s cisteini bogat protein (angl. »*cysteine-rich protein*«) ali nestrukturni protein NS1 (primer Pneumoviridae), Slika 15 prikazuje nekaj primerov virusov iz reda Mononegavirales (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD) (Slika 1.15).



Slika 1.15: Struktura virusov iz reda Mononegavirales (Shi in sod. 2016a). Oznake: N – nukleoprotein; P – fosfoprotein; M – matriks protein. Angleški izrazi: Fusion glycoprotein – fuzijski glikoprotein; RNA capping domain – domena, odgovorna za dodajanje kape.

Chuvirus je skupina ss(-)RNA virusov, podobnih redu Mononegavirales, vendar iz manjšega števila verižno nanizanih ORF in v obratnem vrstnem redu kot pri skupini Mononegavirales (Li in sod., 2015). N-terminalnemu delu je namreč kodiran L segment, proti C-terminalnemu delu pa sledita glikoprotein in nukleoprotein (Slika 1.16).

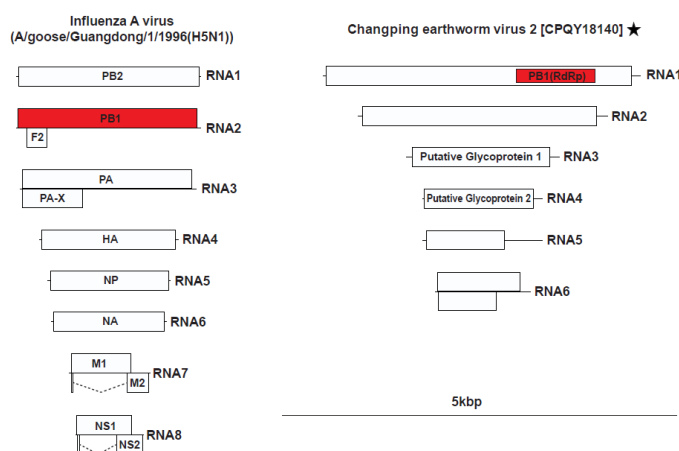


Slika 1.16 Struktura Chuvirusa (Shi in sod., 2016a). Oznake: N – nukleoprotein; G – glikoprotein. Angleški izraz »putative« – domnevni.

Glede na omenjeno strukturo posameznih virusnih skupin znotraj klade Mono-Chu so se verjetno tekom evolucije ostali ORF-ji pri družinah Nyamiviridae in Rhabdoviridae dodajali na N-terminalnem koncu L ORF-ja, pri chuvirusih pa na C-terminalnem koncu.

1.2.1.9 Klada Orthomyxo

Orthomyxoviridae so ss(-)RNA virusna skupina z linearnim segmentiranim genomom skupne dolžine od 10 do 15 kb. Vsak segment predstavlja monocistronsko RNA z enim ORF. Orthomyxoviridae imajo šest ali osem segmentov. Prvi trije segmenti kodirajo tripartitno RdRp, sestavljeno iz PB1, PB2 in PA podenote. Vsaka podenota je kodirana na svojem segmentu z enim ORF. Kot primer virusa z genomom lahko navedemo Quaranjavirus. Poleg prvih treh segmentov, ki kodirajo RdRp, sta četrti in zadnji segment slabo okarakterizirana, peti segment pa kodira glikoprotein. Kot primer virusa z genomom, sestavljenim iz osmih segmentov, lahko navedemo virus gripe (angl. »*Influenza virus*«). Poleg prvih treh segmentov, ki kodirajo tripartitno (tridelno) RdRp, četrti segment kodira hemaglutinin (HA), peti kodira nukleoprotein (NP), šesti kodira encim nevraminidazo (NA), sedmi kodira matriks protein in osmi kodira nestrukturni protein. Hemaglutinin in nevraminidaza sodelujeta pri pritrditvi virusa na celico na osnovi sialne kisline, ki mu sledi s klatrinom posredovan vstop virusa v celico. Nukleoprotein ima podobno vlogo kot pri skupini Mononegavirales, pri čemer z RNA in proteini kapside tvori ribonukleokapsido (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD). Slika 17 prikazuje primer zgradbe skupine Orthomyxoviridae (Slika 1.17).

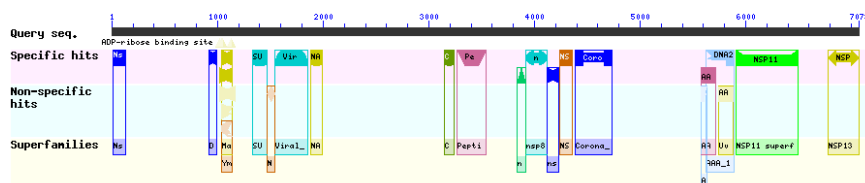


Slika 1.17: Primer strukture skupine Orthomyxoviridae (Shi in sod., 2016a). Oznake: HA – hemagglutinin, NA – nevraminidaza, M – matriks protein in NS – nestrukturni protein. Angleški izraz: Putative glycoprotein – domnevni glikoprotein.

1.2.1.10 Klada Nido

Skupina (red) Nidovirales je skupina vretenčarskih in nevretenčarskih virusov z najdaljšim linearnim monopartitnim ss(+)RNA genomom izmed RNA virusov. Dolžina njihovega genoma je med 19 kb in 30 kb. Zajema virusne skupine Arteriviridae, Coronaviridae, Mesoniviridae in Ronviridae (Hulo in sod., 2011 – ViralZone). Njihov genom je lahko kodiran s 5 do 15 ORF-ji. Jedrni del genoma predstavljata prva dva ORF-ja, ki sta v različnih bralnih okvirjih, se prepisujeta kot ena mRNA in prevajata kot en daljši poliprotein (**pp1ab**). Prepiše se tudi krajša verzija, ki vključuje samo prvi ORF (**pp1a**). Tudi pri redu Nidovirales se namreč med prvima dvema ORF nahaja heptanukleotidno zaporedje X XXY YYZ, ki jim 5–9 nukleotidov proti 3' sledi psevdovozel. Ko RNA polimeraza zazna omenjeno zaporedje, se bodisi vrne en nukleotid nazaj bodisi preskoči en nukleotid. Tako drugi ORF pride v enak

bralni okvir kot prvi in oba ORF se prepiseta kot en poliprotein (angl. »*ribosomal frameshifting*«) (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD). Mesto spremembe bralnega okvirja je tik pred RdRp domeno poliproteina. Strukturi prvih dveh ORF-jev (jedrnega dela genoma) reda Nidovirales se med seboj razlikujeta glede na število nestrukturnih domen, ki jih kodirata. Kot primer lahko vzamemo genomsko strukturo človeškega SARS (angl. »*Severe Acute Respiratory Syndrome*«) koronavirusa (angl. »*human SARS coronavirus*«). Na N-terminalnem delu poliproteina se nahaja Nsp 1 (angl. »*non-structural protein 1*«), ki mu sledi niz domen »Appr1 processing enzyme« – Nsp3 (angl. »*non-structural protein 3*«) – papainu podobna virusna proteaza (angl. »*papain-like viral protease*«) – NAR domena (angl. »*nucleic acid binding domain*«) – Nsp4 (angl. »*non-structural protein 4*«) – peptidaza C30 – Nsp7 (angl. »*non-structural protein 7*«) – Nsp8 (angl. »*non-structural protein 8*«) – Nsp9 (angl. »*non-structural protein 9*«) – Nsp10 (angl. »*non-structural protein 10*«) – RdRp (Nsp12)- S1H (helikaza superdružine 1; angl. »*superfamily 1 DNA and/or RNA helicase*«) – Nsp11 (angl. »*non-structural protein 11*«) – Nsp13 (angl. »*non-structural protein 13*«). Funkcija Nsp1 ni znana. Nsp3 deluje kot *cis* proteaza. Izvaja tri izmed vseh cepitev, potrebnih za nastanek produktov iz poliproteina. Pri samem nastanku ustreznih produktov iz poliproteina sodeluje tudi proteaza C30 (PA klan) (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS). Ostali nestrukturni proteini – Nsp4, Nsp7, Nsp8, Nsp9, Nsp10, Nsp13 poleg RdRp (Nsp12) – sodelujejo pri mehanizmu podvojevanja nidovirusov. Podvojevanje v primeru nidovirusov je izvedeno s pomočjo mreže interakcij virusnih proteinov in dveh polimeraz. Nsp12 je glavna polimeraza nidovirusov s kanoničnimi motivi. Nsp7 in Nsp8 ob interakciji tvorita heksadekamerni cilindrični kompleks in njihova funkcijska porazdelitev elektrostatskega naboja poveča procesivnost RdRp (Nsp12) (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD). Njun kompleks predstavlja drugo polimerazo pri nidovirusih. Ob vstopu ss(+)RNA v citoplazmo gostiteljskih celic lahko ta deluje kot glavna matrica za prevajanje zgodnjih nestrukturnih genov. Med njima sta omenjeni polimerazi. Kompleks Nsp7/Nsp8 ima nižjo afiniteto do ssRNA kot RdRp in naj bi sodeloval pri *de novo* sintezi začetnih oligonukleotidov na ss(+)RNA matrici, pri čemer cilindrična struktura zaobjema nastajajočo dsRNA verigo (te Velthuis in sod., 2012). Zato kompleks Nsp7/8 deluje kot primaza (te Velthuis in sod., 2012). Glavna RdRp (Nsp12) nato nadaljuje sintezo komplementarne antisense verige RNA, pri čemer nastane dsRNA intermediat. Nsp4 je membranski protein, ki pritrdi replikacijsko-transkripcijski kompleks na »prilagojeno« membrano endoplazemskega retikla gostitelja. Na dsRNA intermediatu nato potekata podvojevanje genomske RNA in prepisovanje mRNA poznih genov (sestavnih virusnih delov). Pri prepisovanju mRNA Nsp13 naj bi delovala kot »cap-1« metiltransferaza. Slika 18 prikazuje zaporedje domen na glavnem nestrukturinem poliproteinu nidovirusov (**Slika 1.18**) (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD). Pri drugih skupinah nidovirusov se znotraj pp1ab lahko nahaja več alternativnih peptidaznih domen kot so: peptidaza S32, C107, C62, ki sodijo v klan PA, in peptidaze C31, C32, C33, ki sodijo v CA klan cisteinskih peptidaz (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD).



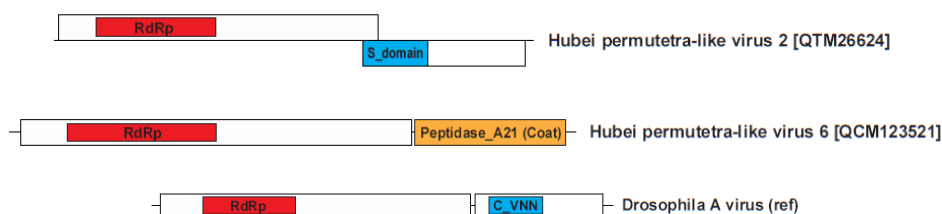
Slika 1.18: Struktura jedrnega dela genoma humanega SARS koronavirusa Coronaviridae, Nidovirales) (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD).

Ostali proteini, ki so v glavnem a) gradbene enote nidovirusov ali b) virulenčni faktorji nidovirusov (sodelujejo pri izogibanju imunskega sistema gostiteljev ali pri propadu gostiteljske celice), se prepisujejo kot subgenomska RNA (monocistronska RNA z enim

ORF). Primeri pri koronavirusih (kot je humani SARS koronavirus) so 1) glikoproteini, kot je S2 glikoprotein koronavirusov (angl. »spike glycoprotein«), ki interagira z receptorji na gostiteljskih celicah, 2) matriks protein, 3) »lipid binding protein«, ki sodeluje pri pritrditvi virusa na plazmalemo gostitelja, 4) proteina 3a (APA3 viroporin) in 3b, ki sodelujeta pri apoptozi gostiteljske celice, 5) protein 7a (SARS_X4), ki sodeluje pri obrambi pred gostiteljevim imunskim sistemom (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD).

1.2.1.11 Klada Permutotetra

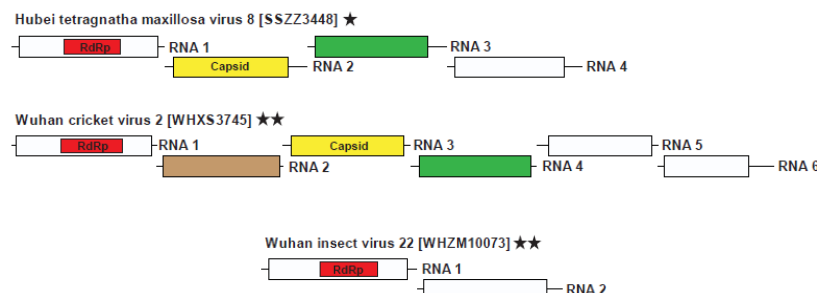
Permutotetraviridae so ss(+)RNA nevretenčarska virusna družina brez virusne ovojnice. Njihov linearen genom je dolg 5,6 kb in je sestavljen iz dveh prekrivajočih ORF-jev. Prvi ORF kodira RdRp, drugi ORF kodira proteine kapside, ki so različnih družin: »Viral coat protein (S domain)«, tipičen za rastlinsko skupino Tombusviridae in Sobemovirus, »Peptidase A21 family coat protein«, tipičen za Alphatetraviridae, Carmotetraviridae in Permutotetraviridae, ali »nodavirus capsid protein (Capsid-VNN, C_VNN) – kapsidni protein, tipičen za nodaviruse. Slika 19 prikazuje primere genomske strukture družine Permutotetraviridae (Hulo in sod., 2011 – ViralZone, Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD) (Slika 1.19).



Slika 1.19: Genomske strukture tipičnih predstavnikov družine Permutotetraviridae (Shi in sod., 2016a). Oznake: C_VNN – »nodavirus capsid protein«; S_domain – »Viral coat protein (S domain)«, Peptidase_A21 – »Peptidase A21 family coat protein« (Shi in sod., 2016a).

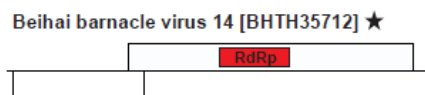
1.2.1.12 Klada Partiti-Picobirna

Partiti-Picobirna virusna klada je skupina vretenčarskih glivnih in rastlinskih virusov, ki je sestavljena iz virusnih družin Partitiviridae in Picobirnaviridae (Shi in sod., 2016a). **Družina Partitiviridae** ima večinoma linearen genom, sestavljen iz dveh segmentov. Genom je dolg od 3,7 do 4,3 kb, vsak segment je predstavljen kot monocistronska RNA. Prvi segment kodira RdRp, drugi segment, ki je bistveno bolj divergenten, kodira gen za proteine kapside. Slednji je pri partitivirih zaradi spremenjenih motivov bistveno slabše okarakteriziran (Hulo in sod., 2011 – ViralZone, Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD). Shi in sod. (2016a) so odkrili nove partitiviruse s štirimi (»Hubei tetragratha maxillosa virus 8« – SSZZ3448) ali šestimi segmenti (»Wuchan cricket virus 2« – WHXS3745). Prvi segment kodira zapis za RdRP. Drugi segment pri »Hubei tetragratha maxillosa virusu 8« in tretji segment pri »Wuchan cricket virusu 2« kodirata zapis za protein kapside. Ostali segmenti so slabše okarakterizirani (Shi in sod., 2016a) (Slika 1.20).



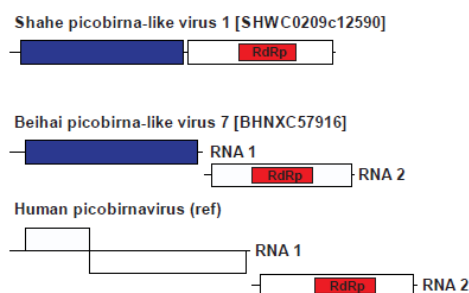
Slika 1.20: Primeri genomske strukture družine Partitiviridae (Shi in sod., 2016a). Angleški izraz: Capsid – kapsida.

Shi in sod. (2016a) so odkrili tudi primere partitivirusov, ki so nesegmentirani in vsebujejo samo monopartitno RNA z dvema ORF-jema – prvim, ki kodira slabše okarakteriziran protein, in drugim, ki kodira RdRp (»Beihai barnacle virus 14«) (Shi in sod., 2016a) (Slika 1.21).



Slika 1.21: Genomska struktura »Beihai barnacle virus 14« (Shi in sod., 2016a).

Družina Picobirnaviridae je virusna družina z linearnim genomom velikosti okrog 4 kb, sestavljenim večinoma iz dveh segmentov. Pri človeškem picobirnavirusu (angl. »*Human picobirnavirus*«) je prvi segment (RNA1) bicistronska RNA, kodirana z dvema ORF-jema. Večji izmed njih kodira kapsidni protein. Drugi segment (RNA2) je monocistronska RNA z enim ORF, ki kodira RdRp. RNA1 je lahko tudi monocistronska, kodirana samo z enim ORF. Shi in sod. (2016a) so odkrili primere nesegmentiranih picobirnavirusov. Prvi ORF njihove monopartitne RNA je slabše okarakteriziran, drugi kodira RdRp (»Shahe picobirna-like virus 1«) (Shi in sod., 2016a) (Slika 1.22).



Slika 1.22: Primeri genomskih struktur virusne skupine Picobirnaviridae (Shi in sod., 2016a).

1.2.1.13 Klada Picorna-Calici

Picorna-Calici je ena od najbolj raznolikih virusnih skupin s hkrati najširšim naborom enoceličnih in mnogoceličnih evkariontskih gostiteljev. Klado sestavlja vsaj 29 različnih rodov (angl. »*genera*«), med katerimi so tudi pomembni humani patogeni (Ullmer in Semler, 2016). So sferične oblike in nimajo virusne ovojnice. Njihov genom ima razpon velikosti nekje od 7 do 11 kb. Je enodelen in ga sestavlja eden ali več ORF. Delimo jih na več podskupin: vodni pikornavirusi (angl. »*aquatic picornavirales*«), Dicistroviridae, Iflaviridae, Posavirusom sorodni pikornavirusi (angl. »*Posavirus-related picornaviruses*«), Arivirusom sorodni pikornavirusi (angl. »*Arivirus-related picornaviruses*«), Kelp fly virusu

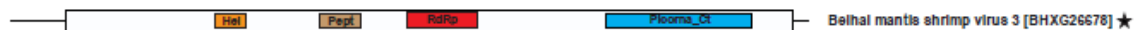
sorodni pikornavirusi (angl. »*Kelp fly virus – related picornaviruses*»), Nora virusu sorodni pikornavirusi (angl. »*Nora virus – related picornaviruses*»), Secoviridae, Picornaviridae, Caliciviridae (Hulo in sod., 2011 – ViralZone). Vsem je skupna RNA helikaza naddružine 3 (angl. »*SF3 helicase; superfamily 3 helicase*»), RdRp družine 1 (RdRp_1) in peptidaza C3 – 3C cistein proteaza (klan PA) in domene VP1-VP4 - proteini plašča, ki so organizirani v psevdo 3 ikozahedralno simetrijo (prikazane kot Rhv, Dicistro-VP4 in CRPV domene pri CDD) (Marchler-Bauer in sod., 2017 – CDD). Razlika je le v številu prisotnih kopij in njihovi organizaciji.

Dicistroviridae je pikornavirusna skupina z dvema ORF-jema. Vsak ORF kodira zapis za dva poliproteina. Prvi kodira poliprotein s proteinskimi domenami v nizu: helikazo naddružine 3, C3 peptidazo (3CPro peptidaza, angl. »poliovirus-type picornain 3C«) in RdRp. C3 peptidaza sodi v PA klan peptidaz in je razdeljena na poddružine C3A, C3B, C3C, C3D, C3E, C3F, C3G in C3H. Drugi ORF kodira strukturni poliprotein Rhv-Dicistro VP4-Rhv-CRPV. Proteinske domene Rhv, CRPV (angl. »cricket paralysis virus capsid protein«) in Dicistro VP4 so osnovne gradbene enote kapside pikornavirusov (**Slika 1.23**).



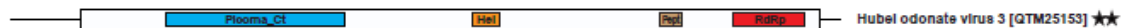
Slika 1.23: Struktura Dicistroviridae (Shi in sod., 2016a). Strukturni poliprotein Rhv-Dicistro VP4-Rhv-CRPV je označen kot Picorna_Ct (angl. »picorna-like coat domains«), Hel – helikaza naddružine 3, Pept – peptidaza C3.

Shi in sod. (2016a) so odkrili nove dicistrovire z genomom, ki kodira en sam poliprotein, in enakim zaporedjem niza domen kot tipični dicistrovirusi: SF3 helikaza-3CPro-RdRp- Rhv-Dicistro VP4-Rhv-CRPV (**Slika 1.24**).



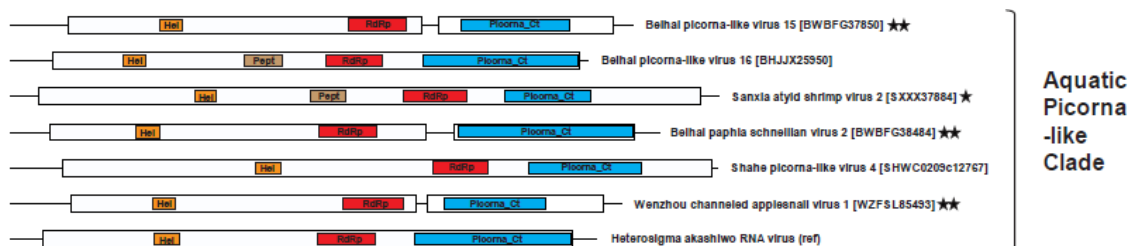
Slika 1.24: Primer virusa z evolucijsko afiniteto do Dicistroviridae in genomom, kodiranim z enim ORF (Shi in sod., 2016a). Strukturni poliprotein Rhv-Dicistro VP4-Rhv-CRPV je označen kot Picorna_Ct (angl. »picorna-like coat domains«), Hel – helikaza naddružine 3, Pept – peptidaza C3.

Iflaviridae je pikornavirusna skupina z monopartitnim genomom, kodiranim z enim ORF. ORF kodira poliprotein, ki je sestavljen iz strukturnega in nestrukturnega dela (**Slika 1.25**). N-terminalni del je strukturni in kodira proteine kapside (Rhv in CRPV). C-terminalni del je nestrukturni in kodira helikazo in RdRp_1.



Slika 1.25: Struktura Iflaviridae. Strukturni poliprotein Rhv-Rhv-CRPV je označen kot Picorna_Ct (angl. »picorna-like coat domains«), Hel – helikaza naddružine 3, Pept – peptidaza C3 (Shi in sod., 2016a).

Genom »vodnih pikornavirusov« (angl. »*aquatic picornavirales*«) je lahko zastopan z enim ali dvema ORF-jema. Peptidazne domene in CRPV domene so lahko odsotne ali pa manjkajo njihovi karakteristični motivi in so zato neprepoznave (**Slika 1.26**). Skupina »vodnih pikornavirusov« ima genomsko strukturo, podobno dicistrovirusom, vendar gre za z vidika zaporedja ločeno in sorodno skupino znotraj skupine Picornavirales. Pri obeh skupinah je torej tekom evolucije prišlo do zlepljanja ORF v en poliprotein.



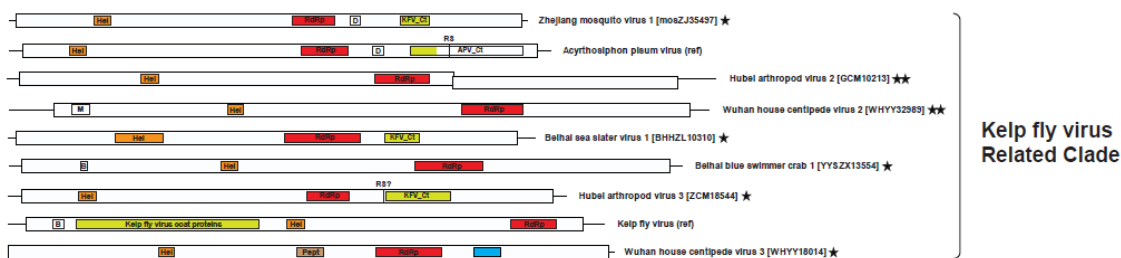
Slika 1.26: Struktura genomov vodnih pikornavirusov (Shi in sod., 2016a). Niz Rhv-Rhv-CRPV je označen kot Picorna_Ct (angl. »picorna-like coat domains«), Hel – helikaza naddružine 3, Pept – peptidaza C3 (Shi in sod., 2016a).

Nora virusom sorodni pikornavirusi imajo genom, sestavljen iz več ORF-jev. Nekateri izmed njih so slabo okarakterizirani. Nestrukturni ORF kodira proteine v nizu helikaza naddružine 3- Peptidaza C4- RdRp. Peptidaza C4 je divergenten del Nora virusu sorodnih virusov. Motivi omenjenega proteina so namreč lahko spremenjeni in domena zato ni prepoznana. Značilen je tudi divergenten ORF, ki kodira »Nora virus coat protein«. Znotraj omenjenega ORF se zazna samo Rhv domena (**Slika 1.27**). Pri Nora virusu podobnih virusih se pojavi tudi HAM1 proteinu podoben protein.



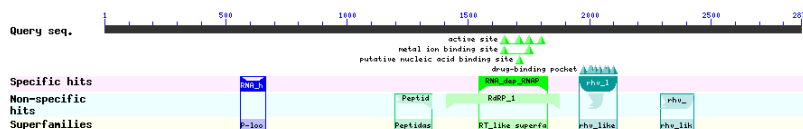
Slika 1.27: Struktura Nora virusa. Hel – helikaza naddružine 3, Pept – peptidaza C3 (Shi in sod., 2016a).

Kelp fly virusu sorodni pikornavirusi so divergentna skupina virusov z monopartitnim genomom, kodiranim z enim ali več ORF. Strukturni del »Kelp fly virus coat proteins (KFV_Ct)« je lahko kodiran na C terminalnem ali N-terminalnem delu genoma in je najbolj divergenten del genoma. Vključuje lahko BIR protein (angl. »*Baculoviral inhibition of apoptosis protein repeat domain protein*«) ali »Calicivirus putative RNA polymerase/capsid protein«. Poleg peptidaze C3 imajo lahko drugo družino peptidaze – peptidazo C37 (**Slika 1.28**).



Slika 1.28: Strukture kelp fly virusu sorodnih pikornavirusov. KFV_Ct – »Kelp fly virus coat proteins«, Hel – helikaza naddružine 3, Pept – peptidaza C3 (Shi in sod., 2016a).

Fisa- ali posavirusu sorodni pikornavirusi imajo strukturo, podobno dicistrovirusom. Razlikujejo se predvsem v sami sekvenci RdRp (**Slika 1.29**).



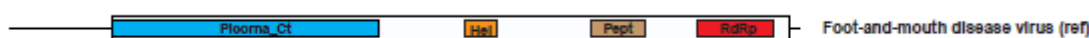
Slika 1.29: Struktura Fisavirusa 1. Helikaza naddružine 3, Peptidaza C3, RdRp in dve Rhv domeni (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD).

Arivirusu sorodni pikornavirusi imajo strukturo, podobno dicistrovirusom. Njihov monopartitni genom sestavljata nestrukturni in strukturni poliprotein. Lahko imajo dodatne slabo okarakterizirane ORF, kot npr. Arivirus 1 (**Slika 1.30**).



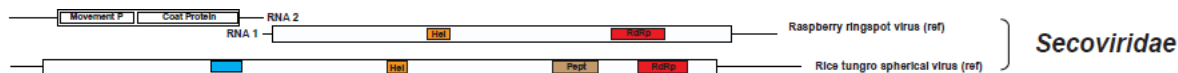
Slika 1.30: Struktura Arirusa 1 in Arirusa 2. Strukturni poliprotein Helikaza naddružine 3-Peptidaza C3-RdRp. Nestrukturni poliprotein vključuje dve Rhv domeni (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD).

Picornaviridae imajo strukturo, podobno skupini Iflaviridae. Monopartitni genom kodira en ORF, ki je sestavljen iz strukturnega (N-terminalni del) in nestrukturnega dela (C-terminalni del). Njihov strukturni del (Picorna_ct) sestavljajo tri Rhv domene (**Slika 1.31**).



Slika 1.31: Struktura virusne skupine Picornaviridae. Picorna_Ct (angl. »picorna-like coat domains«), Hel – helikaza naddružine 3, Pept – peptidaza C3 (Shi in sod., 2016a).

Secoviridae so družina rastlinskih pikornavirusov z večinoma segmentirano, lahko pa tudi nesegmentirano RNA. Segmentirane viruse sestavljata dva segmenta. Segment 1 je nestrukturni segment, ki kodira en ORF oz. en poliprotein Helikaza naddružine 3 – peptidaza C3 – RdRp. Segment 2 je strukturni segment, ki kodira proteine kapside in MP protein. Kot primer lahko navedemo »large coat protein (Como_LCP)« in »small coat protein (Como_SCP)« pri Comovirinae (podskupina Secoviridae). Nesegmentirani virusi skupine Secoviridae, kot je npr. Rice tungro spherical virus, imajo strukturo, podobno skupini Iflaviridae (Rhv-Rhv-helikaza naddružine 3-peptidaza C3G-RdRp) (**Slika 1.32**).



Slika 1.32: Struktura skupine Secoviridae. Picorna_Ct (angl. »picorna-like coat domains«) – modra, Hel – helikaza naddružine 3, Pept – peptidaza C3 (Shi in sod., 2016a).

Caliciviridae so družina vretenčarskih virusov, ki so sorodni Picornavirales. Njihov genom je sestavljen iz dveh ali več ORF-jev. N-terminalni del genoma je strukturni del s helikazo (naddružina 3)-peptidazo C37-RdRp. Strukturni del je lahko kodiran na ločenem ali istem ORF-ju na C-terminalnem delu – »Calicivirus coat protein« (**Slika 1.33**).

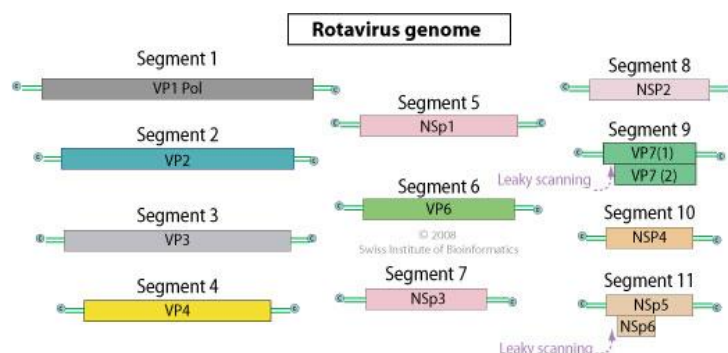


Slika 1.33: Struktura skupine Caliciviridae. Picorna_Ct (angl. »picorna-like coat domains«), Hel – helikaza naddružine 3, Pept – peptidaza C3 (Shi in sod., 2016a).

Pri kladi Picorna-Calici je med evolucijo očitno prišlo do »eksplozije« v raznolikosti genomskih struktur. Razvile so se raznolike genomske organizacije znotraj posameznih podskupin pikornavirusov. Znotraj iste podskupine pikornavirusov (npr. dicistrovirusov) imajo lahko virusi genom iz različnega števila ORF-jev. Pojavile so se proteinske domene, ki so unikatne za posamezno podskupino virusov. Posamezne proteinske domene so zamenjale proteinske domene, ki spadajo v druge družine, kot so različne družine peptidaz – poleg C3 tudi C37, C4. Zato je samo na osnovi genomske strukture težje sklepati, kateri skupini znotraj klade Picorna-Calici pripada analizirani virus. Potrebujemo tudi podatek o podobnosti njegove RdRp z ostalimi že znanimi virusi.

1.2.1.14 Klada Reo

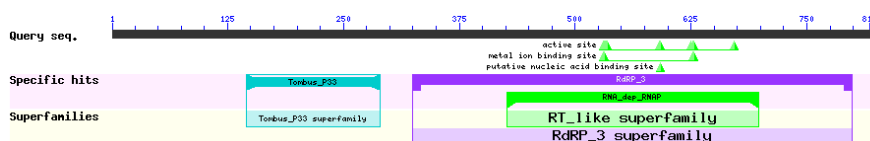
Reoviridae je dsRNA virusna družina z linearnim segmentiranim genomom skupne dolžine od 18 do 31 kb. Vsak segment predstavlja monocistronska RNA, ki vsebuje en ORF. Določeni segmenti imajo lahko dva alternativna start kodona v različnih bralnih okvirjih (angl. »initiation codon«), pri čemer je prvi iniciacijski kodon šibek, npr. ACG ali ATG v kontekstu s šibko Kozakovo sekvenco. Zato 40S podenota ribosomov prvi ORF lahko preskoči in začne prevajati drugega. Pojav se imenuje »leaky scanning« ali »termination-reinitiation«. Skupina Reoviridae se deli na dve poddružini: Sedoreovirinae in Spinareovirinae. Razlikujeta se po organizaciji in družinah strukturnih in nestrukturnih proteinov, ki jih segmenti kodirajo. Raznolikost je posledica raznolikih gostiteljev (ki predstavlja različne dinamične ekološke niše), ki jih reovirusi okužujejo. Gostitelji za reoviruse so človek, drugi vretenčarji, nevretenčarji in evkarionti, kot so glive. Slika 34 prikazuje primer strukture genoma skupine Reoviridae (**Slika 1.34**).



Slika 1.34: Primer genomske strukture skupine Reoviridae (Hulo in sod., 2011 – ViralZone, Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD). Angleški izraz: genome – genom.

1.2.1.15 Klada Tombus-Noda

Klado Tombus-Noda predstavljajo virusi brez virusne ovojnice, premera okrog 30 nm in ikozahedralne simetrije. Sestavljena je iz treh virusnih skupin; Tombusviridae, Nodaviridae in Carmotetraviridae. **Družina Tombusviridae** je divergentna družina rastlinskih virusov. Sestavlja jih enodelni genom s štirimi ali več ORF-ji ali segmentiran genom. Glavni poliprotein, skupen vsem tombusvirusom, predstavljata prva dva ORF, ki kodirata zapis za protein P33 in RdRp (**Slika 1.35**).



Slika 1.35: Osrednji del genoma skupine Tombusviridae (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD). Oznaka: Tombus_P33 – tombusvirusni protein P33; RT_like in RdRp_3 – naddružini RdRp. Angleški izraz: Superfamily – naddružina.

Protein P33 sodeluje pri podvojevanju tombusvirusov v rastlinah. Protein je tipičen za tombusviruse, vseeno pa je verjetno zaradi genetskega drifta (in posledično višje hitrosti mutacij v primerjavi z RdRp) prišlo do izgube karakterističnih motivov, zato domena ni zaznana pri vseh virusih oz. ni ohranjena čez celotno virusno skupino. Posebnost prvih dveh ORF-jev je, da se tipično prepišeta kot ena bicistronska mRNA. Med njima je namreč stop kodon. V fazi prevajanja UAG stop kodona za proteinskim zaporedjem -ESLGVK- na koncu prvega ORF (npr. pri virusu »tomato bushy stunt virus«) (ki kodira zapisa za protein P33)

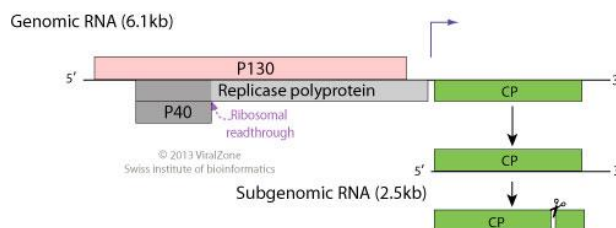
ne prepoznajo sprostitveni dejavniki (eRF), temveč delno tRNA. Gre torej za pojav »suppression of termination« oz. »stop codon read-through«. Ostali ORF-ji se prepisujejo kot subgenomska RNA (angl. »subgenomic RNA«). Drugi glavni ORF kodira proteine plašča, kot sta virusni protein plašča (S domena) (angl. »viral coat protein; S domain«), »alfacarmovirusni protein plašča« (angl. »carmovirus coat protein«). Tretji in četrti glavni ORF-ji kodirajo tombusvirusni P19 protein (angl. »tombusvirus P19 core protein«) in tombusvirusni protein P21/P22 (angl. »TBSV core protein P21/P22«). Nekateri tombusvirusi kot rastlinski virusi kodirajo tudi MP proteine (angl. »movement protein«), kot so: tombusvirusni MP protein (angl. »tombusvirus movement protein«), MP protein naddružine 3A in »Umbravirus long distance movement protein«. Hkrati so nekateri tombusvirusi, kot npr. »Chronic bee paralysis virus«, »Diaphorina citri-associated C virus« in »Red clover necrotic mosaic virus«, segmentirani. Prvi segment pri »Red clover necrotic mosaic virusu« kodira RdRp in virusni protein plašča (angl. »Viral coat protein; S domain«) na dveh ločenih ORF. Drugi segment kodira MP protein. Pri »Chronic bee paralysis virusu« je protein plašča (plaščni protein) kodiran na ločenem segmentu. Pri »Diaphorina citri-associated C virusu« je prvi segment kodiran s funkcionalno okarakteriziranim N-terminalnim ORF in C-terminalnim ORF z RdRp. Drugi segment je kodiran z N-terminalnim slabo okarakteriziranim N-terminalnim ORF, C-terminalni ORF pa kodira DiSB-ORF2 glikoprotein, tipičen za Negevirusse.

Shi in sod. (2016a) so odkrili nove tombusviruse z monopartitno RNA, pri katerih prvi ORF na N-terminalnem koncu genoma ne predstavlja RdRp, temveč drug ORF, ki funkcionalno ni okarakteriziran (»Sanxia tombus like virus 7« (SXSSP2567), »Beihai tombus like virus 12« (BHTSS17936), »Wenzhou tombus-like virus 12« (WZFSL84432) idr.). Omenjeni virusi tudi v njihovem filogenetskem drevesu tvorijo svoj klaster. Nekateri tombusvirusi (Shahe isopoda virus 5) namesto klasičnega in virusnega proteina plašča (angl. »Viral coat protein; S domain«) kodirajo protein plašča naddružine peptidaze A6, ki je tipična za nodaviruse. Skupina tombusvirusov je divergentna skupina, pri kateri je tekom evolucije očitno pomembno vlogo igrala tako segmentacija, pridobivanje novih karakterističnih lastnosti (novih ORF) kot tudi horizontalni prenos genov.

Družina Nodaviridae je družina vretenčarskih virusov. Načeloma so segmentirani. Segmentirani virusi so sestavljeni iz dveh segmentov. Prvi segment je nestrukturni segment in zajema RdRp in vezavni protein B2. Drugi segment je strukturni in kodira zapis za proteine, ki gradijo virusno kapsido in so tipični za nodaviruse: prekursor kapsidnega proteina družine peptidaza A6 (angl. »peptidase A6 family capsid protein precursor«) (tipični samo za nodaviruse; alfanodaviruse), »Nodavirus capsid protein«, Capsid_VNN (tipičen za nodaviruse; betanodaviruse) ali pa so jih nodavirusi prevzeli od drugih RNA virusnih skupin: »VP6 kapsidni protein« (tipičen za rotaviruse; Reoviridae), virusni kapsidni protein (angl. »viral coat protein; S domain«). Nekateri nodavirusi so torej prevzeli gen za kapsidne proteine od virusne klade Reoviridae ali drugih rastlinskih RNA virusov.

Shi in sod. (2016a) so odkrili nove nesegmentirane nodaviruse, pri katerih sta gena za RdRp in protein kapside kodirana na istem segmentu – kot dva ORF na isti monopartitni RNA. Primer je »Beihai noda-like virus 1«, ki poleg gena za RdRp na monopartitni RNA v ločenem ORF-ju vsebuje tudi gen za »Rotavirus major capsid protein VP6«, ki je sicer tipičen za dsRNA virusno skupino Reoviridae (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD). Tekom evolucije je pri skupini Nodaviridae lahko prišlo do segmentacije genoma ali do zlepljanja genoma. Ne moremo namreč podati natančne razlage na podlagi drevesa, koreninjenega na sredini. Težje je namreč določiti najbližjega sorodnika virusne skupine.

Družina Carmotetraviridae je nevretenčarska virusna družina brez virusne ovojnice in z linearnim monopartitnim ss(+)RNA genomom dolžine okrog 6,1 kb. Strukturno se nekoliko razlikuje od Tombusviridae in Nodaviridae. Monopartitna RNA kodira več prekrivajočih ORF. Na N-terminalnem delu je p130 protein. S p130 proteinom se v drugem, vendar med seboj enakem ORF prekrivata 2 ORF, ki se po mehanizmu »suppression of termination« oz. »stop codon read-through« ali »ribosomal read-through« prevajata kot en protein. UAG in UGA kodona sta namreč namesto s sprostitvenimi dejavniki delno prepoznana s tRNA in prevajanje se nadaljuje iz enega na drug ORF. ORF-ja kodirata zapis za Peptidazo (peptidaza C15) in RdRp. Tretji ORF se prepisuje kot subgenomska RNA in kodira strukturni del – proteine plašča (Slika 1.36).

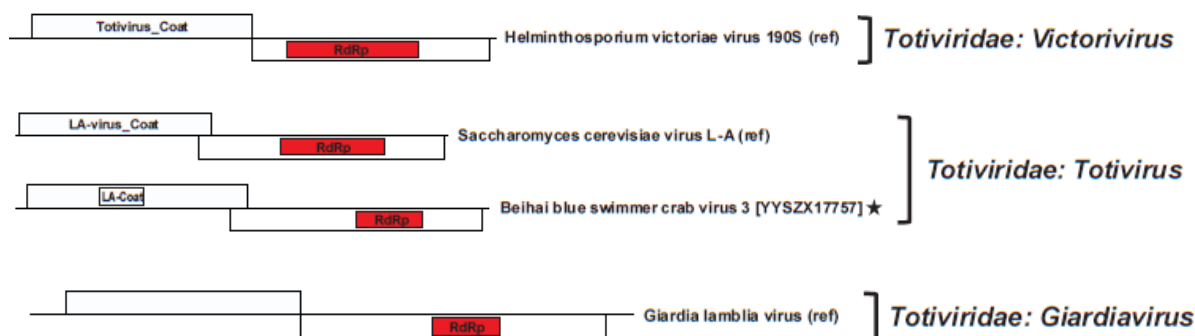


Slika 1.36: Struktura virusne skupine Carmotetraviridae (Hulo in sod., 2011 – ViralZone). Oznake: P130 in P40 – proteina P130 in P40; CP – protein plašča. Angleški izraz: replicase poliprotein – replikazni poliprotein.

1.2.1.16 Klada Toti-Chryso

Toti-Chryso virusna klada je dsRNA klada, sestavljena iz treh virusnih skupin: Totiviridae, Quadriviridae in Chrysovirus.

Družina Totiviridae je sestavljena iz petih virusnih rodov: Giardiavirus, Leishmanivirus, Totivirus, Trichomonasvirus in Victorivirus. Vse imajo linearen nesegmentiran dsRNA genom dolžine med 4 in 7 kb, sestavljen iz dveh ORF. Prvi kodira protein plašča, totivirusni plaščni protein (angl. »totivirus coat protein«) ali L-A virusni protein plašča (angl. »L-A virus major coat protein«). Drugi ORF kodira RdRp. ORF-ja sta v različnih bralnih okvirjih in se med seboj prekrivata. Omenjena ORF (na bicistronski mRNA) se lahko prevajata na dva različna načina. Pri določenih totivirusih se za prvim ORF nahaja zaporedje X XXY YYZ (pri Giardiavirusu je to zaporedje C CCU UUA) in psevdovozelj, kar povzroči, da se ribosom vrne en nukleotid nazaj ali preskoči en nukleotid, pri čemer pride do pojava, pri katerem prideta oba zaporedna ORF pod isti bralni okvir in se lahko prevedeta kot en fuzijski protein (angl. »ribosomal frameshifting«). Pri totivirusih z omenjenim sistemom prevajanja nastane kapsidni protein in fuzijski protein kapsidni protein – RdRp. Lahko pa se na začetku drugega ORF nahaja reiniciacijski motiv (angl. »reinitiation motif«), ki omogoča, da kljub stop kodonu na koncu prvega ORF določen delež 40S podenot ribosomov ostane vezanih in ponovno zažene prevajanje na začetnem kodonu drugega ORF (angl. »RNA termination reinitiation« or »stop-start«). Tako se drugi ORF (RdRp) ne prepiše v okviru fuzijskega proteina, temveč kot samostojni protein (Hulo in sod., 2011 – ViralZone). Slika 37 prikazuje nekaj primerov genomskih struktur skupine Totiviridae (Slika 1.37).



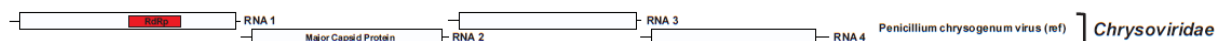
Slika 1.37: Primeri struktur virusne skupine Totiviridae (Shi in sod., 2016a). Kapsidna proteina (domeni): Totivirus_Coat, LA-virus_Coat. RdRp – od RNA odvisna RNA polimeraza.

Quadriviridae so segmentirana virusna družina s tipično štirimi segmenti, od katerih sta dva domnevno strukturna segmenta, eden pa je nestrukturni segment, ki kodira RdRp – Segment 3 (Hulo in sod., 2011 – ViralZone; Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD; Shi in sod., 2016a). Vsak segment je monocistrinska dsRNA (Slika 1.38).



Slika 1.38: Primer strukture Quadriviridae (Shi in sod., 2016a). Angleški izraz: Structural protein – strukturni protein.

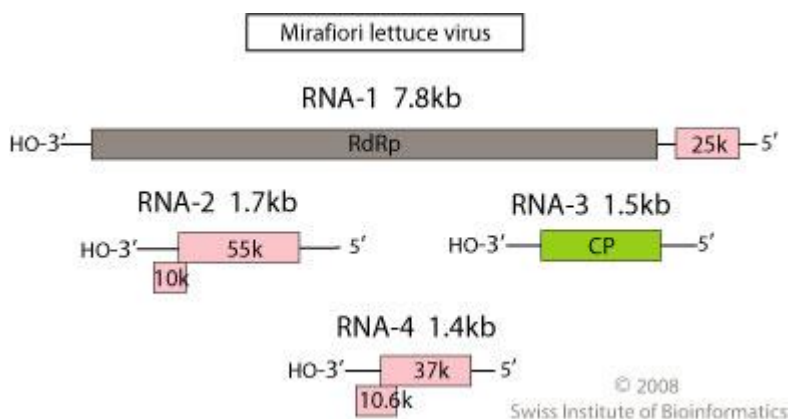
Chrysovirus je rod virusov z linearnim segmentiranim genomom, ki je ravno tako sestavljen iz štirih segmentov skupne dolžine 12,5 kb. Prvi segment je nestrukturen z enim ORF (monocistronska dsRNA), ki kodira RdRp, in drugi segment je strukturni, ki kodira zapis za kapsidni protein (Hulo in sod., 2011 – ViralZone; Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD; Shi in sod., 2016a). Slika 37 prikazuje njihovo strukturo (**Slika 1.39**).



Slika 1.39: Primer genomske strukture virusne skupine Chrysovirus (Shi in sod., 2016a). Angleški izraz: Major Capsid Protein – glavni protein kapside.

1.2.1.17 Klada Ophio

Ophioviridae je družina segmentiranih ss(-)RNA virusov s skupno dolžino genoma od 11,3–12,5 kb. Njihov genom je sestavljen iz treh ali štirih segmentov, pri čemer prvi, najdaljši segment (L segment) kodira zapis za RdRp, drugi segment za kapsidni protein, kot je »Plant viral coat protein nucleocapsid«, ostali segmenti pa so slabše okarakterizirani (Hulo in sod., 2011 – ViralZone; Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD; Shi in sod., 2016a) (**Slika 1.40**).

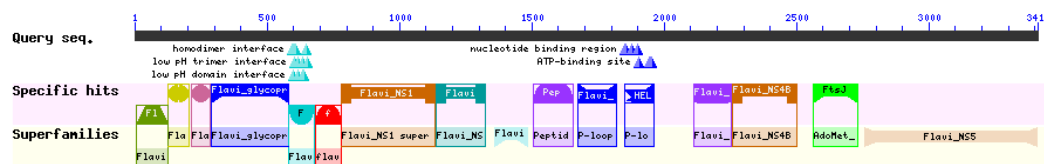


Slika 1.40: Struktura genoma iz družine Ophioviridae (Hulo in sod., 2011 – ViralZone; Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD; Shi in sod., 2016a). Oznaka CP – protein kapside.

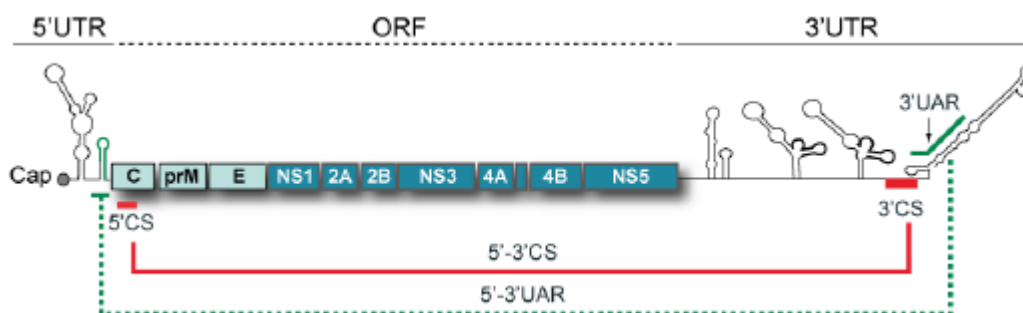
1.2.1.18 Klada Flavi

Flaviviridae so vretenčarska ss(+)RNA virusna družina z linearno monopartitno in monocistronsko RNA. Okarakterizirane so štiri podskupine flavivirusov, ki se glede domen nekoliko razlikujejo: flavivirus, hepacivirus, pegivirus, pestivirus. Nedavno sta bili definirani še dve skupini virusov: flavivirusom podobni virusi in jingmenvirusi (Shi in sod., 2015). Genom flavivirusov je dolg med 9,6 in 12,3 kb in vsebuje samo en ORF, ki kodira zapis za en poliprotein. N terminalni del poliproteina predstavlja strukturni del, C-terminalni pa nestrukturni del (**Slika 1.41**). 5' nekodirajoča regija genoma je dolga do 100 nukleotidov, 3' nekodirajoča regija pa od 400 do 700 nukleotidov (Ng in sod., 2017).

a)



b)



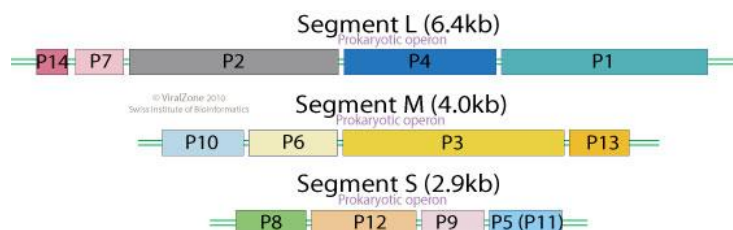
Slika 1.41: a) Primer genomske strukture Yellow fever virus (Hulo in sod., 2011 – ViralZone; Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD; Shi in sod., 2016a). b) Primer genomske strukture virusa denga vročice (angl. »Dengue virus«) (Gebhard in sod., 2011). Angleški izraz: cap – kapa.

Verižni niz domen se začne s kapsidnimi proteini: »Flavivirus capsid protein C«, »Hepatitis C virus capsid protein«, sledijo glikoproteini, ki se nahajajo na virusni ovojnici (»Flavivirus envelope glycoprotein M«, »Flavivirus envelope glycoprotein E«, »Hepatitis C virus envelope glycoprotein E1«). Strukturnemu delu sledi nestrukturni del z nestrukturnimi proteini NS1-5 in vmesnimi peptidaznimi in helikaznimi domenami. Peptidaze delujejo kot cis delujoče regije, ki po prevajanju sodelujejo pri »zorenju« oz. razrezu poliproteina na strukturne in nestrukturne sestavne enote virusa. Pri flavivirusih si na primer sledijo strukturne domene v nizu NS1-NS2A-NS2B-peptidaza S7-helikaza naddružine II-NS4A-NS4B-»FtsJ« podobna metiltransferaza» (angl. »FtsJ-like methyltransferase«) in flavivirusna RdRp, ki predstavlja NS5 domeno pri flavivirusih. Pri hepaticivirusih in pegivirusih je NS5 domena sestavljena iz več podenot, RdRp pa je ločena od NS5 domene. Pestivirusi vsebujejo več peptidaznih domen: peptidaza S32, peptidaza C74, peptidaza C53.

Znotraj klade Flavi je tekom evolucije prednika flavivirusov in jingmenvirusov v jingmenvirusu pri jingmenvirusih prišlo do segmentacije genoma. Helikaza in peptidaza sta na ločenem segmentu, glikoprotein na ločenem bicistronskem segmentu, protein kapside na ločenem bicistronskem segmentu in »FtsJ« podobna metiltransferaza z RdRp na ločenem monocistronskem segmentu. Ostale skupine flavivirusov imajo monopartitni genom, monocistronsko RNA, ki kodira zapis za en poliprotein. Pri vseh virusih je poleg RdRp ohranjena helikaza naddružine 1 (DEAD/DEAH box helikaza). Flavivirusom podobni virusi predstavljajo starejšo obliko skupine Flavivirusov in Jingmenvirusov (Shi in sod., 2016b). Načeloma so pri vseh virusih družine Flaviviridae poleg RdRp ohranjeni tudi »DEAD/DEAH box« helikaza – helikaza naddružine II in »FtsJ« podobna metiltransferaza» (Hulo in sod., 2011 – ViralZone; Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD; Shi in sod., 2016a). Genom flavivirusov in flavivirusom podobnih virusov vsebuje torej »FtsJ« podobno metiltransferazo» oz. »FtsJ metiltransferazo«, ki je sicer tipična za Negeviruse. Vendar se je domena ohranila tudi pri Flavivirusih.

1.2.1.19 Klada Cysto

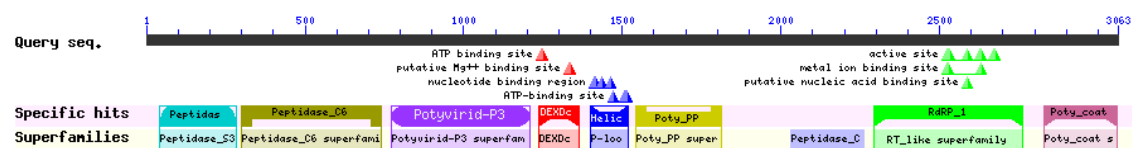
Cystoviridae so bakterijska dsRNA družina z linearnim segmentiranim genomom. Genom je sestavljen iz treh segmentov L (angl. »*large*«), M (angl. »*medium*«) in S (angl. »*small*«) v skupni dolžini 13,3 kb. Segmenti so policistronske mRNA, ki kodirajo 12 proteinov P1-P12 (Slika 1.42). Protein P2 predstavlja polimerazo.



Slika 1.42: Genomska struktura Cystoviridae (Hulo in sod., 2011 – ViralZone; Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD; Shi in sod., 2016a). Angleški izraz: prokaryotic operon: prokariotski operon.

1.2.1.20 Klada Poty

Skupina Potyviridae je skupina filamentoznih rastlinskih virusov brez kapside. Njihov genom je monopartiten in velik od 8,5 do 12 kb. Rodovi, vključeni v družino Potyviridae, so Brambyvirus, Bymovirus, Ipomovirus, Macluravirus, Poacevirus, Potyvirus, Rymovirus in Tritimovirus. Na N-terminalnem delu genoma kodira zapis za dve peptidazi, peptidazo S30 in peptidazo C3. Peptidazi S30 (angl. »*potyvirus P1 peptidase*«) in C3 (angl. »*poliovirus-type picornain 3C*«) sodita v klan PA endopeptidaz mešanega tipa (serinske ali cisteinske peptidaze) s katalitično triado histidin, aspartat/glutamat in cistein/serin. Peptidaza S30 je specifična za Potyviridae. Sledijo protein P3, ki je specifičen za Potyviridae, in »DEAD/DEAH box« helikaza. Helikaze so bile uvrščene v pet naddružin (SF1-SF5, angl. »superfamily helicase 1-5«). Naddružini 1 in 5 sta najbolj razširjeni. »DEAD/DEAH box« helikaza sodi v naddružino 2 helikaz. Sledita domeni Poty_PP (angl. »*Potyviridae polyprotein*«) in peptidaza C4, ki ravno tako sodi v PA klan peptidaz. Na C-terminalnem delu se nahajata od RNA odvisna RNA polimeraza in kapsidni protein, specifičen za Potyviridae (Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD) (Slika 1.43).



Slika 1.43: Struktura Potyviridae (Marchler – Bauer in sod., 2017 – CDD). Oznake: Peptidase_C3 – peptidaza C3; Peptidase_C6 – peptidaza C6; Potyvirus_P3 – protein P3, specifičen za Potyviridae; DEXDc in P-loop – helikaza; Poty_PP – poliprotein skupine potyviridae (angl. »*Potyviridae polyprotein*«); Peptidase_C4 – peptidaza C4; Poty_coat – protein plašča, specifičen za Potyviridae. Angleški izraz: superfamily – naddružina.

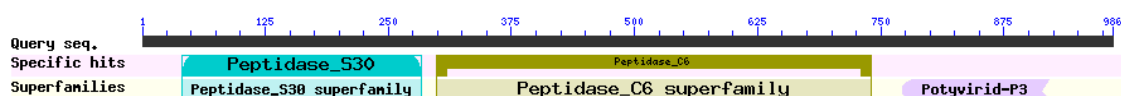
Krajša oblika proteina P3N-PIPO se prepisuje po mehanizmu »polymerase slippage« oz. »RNA editing«. Gre za pojav, pri katerem RNA polimeraza začne prepisovati mRNA v določenem bralnem okvirju od start kodona, dokler ne naleti na specifičen signal »slippery sequence«. Za potyvirus je to tipična sekvenca »...UGGAAAAA...«. Ob prihodu na omenjene signale se RNA polimeraza ustavi, preide za eno mesto nazaj in ponovno ustavi zadnji nukleotid. Nastanejo insercije 1, 2 ali več baz in nadaljevanje prepisovanja v drugem okenskem okvirju do stop kodona. Kot primer lahko navedemo P3N-PIPO protein virusa PVY (angl. »*Potato virus Y*«) (Slika 1.44).

a.)

```
>sp|P0CJ93|MVP_PVYN P3N-PIPO polyprotein OS=Potato virus Y (strain N) PE=3 SV=2
```

```
MATYMTSTICFGSFECKLPYSPASCEHIVKEREVPASVDPFADLETQLSARLLKQKYATVR
VLKNGFTTYRYKTDQIMRIQKKLERKDREYHFQMAAPSIVSKITIAGGDPSPKSEPOA
PRGIHTTPRMKVKTRPIIKLTEGQMNHLIKQIKQIMSEKRGSVHLISKKTTHVQYKKI
LGAYSAAVRTAHMMGLRRRVDFRCDMWTVGLLQRLARTDKWSNQVRTINIRRGDSGVILN
TKSLKGHFGRSSGGLFIVRGSHGKLYDARSRTQSIILNSMIQFSNADNFWKGLDGNWAR
MRYPDHTCVAGLPVEDCGRVAALMAHSILPCYKITCPTCAQQVYASLPVSDLFKLLHKHA
RDGLNRLGADKDRFIHVNKFLIALEHLTEPVDLNLLELFNEIFKSIGEKQQAPFNLVNLN
NFFLKKGKENTAHWEQVAQLSLELAREFQKNRTDNKKGDISFFRNKLSAKANWNLYLSCD
NQLDKKNANFLWGQREYHAKRFFSNFFEEIDPAKGYSAYEIRKHPSGTRKLSIGNLVVPLD
LAEFRQKMGDYRKQPGVSKKCTSSKDGNYVYPCCCTTLDDGSAIESTFYPPTKKHLVIG
NSGDQKFVDLPKGDSEMLYIAKQGYCYINVFLAMLINISEEDAKDFTKKVRDMCVPKLGT
WPTMMDLATCAQMRIFYPDVHDAELPRILVDHDTQTCHVVDSEFGSQTGYHILKASSVS
QLILFANDELES DIKHYRVGGVNPASPELGSTISPFREGGVIMSESAAALKLLKGFIRPK
VMRQLLLDEPYLLILSILSPGILMAMYNNGIFELAVRLWINEKQSIAMIASLLSALALRV
SAAETLVAQR II IDAAATDLLDATCDGFNLHLYPTALMVQLVVKNRNECDDTLFKAGFP
SYNTSVVQIMEKKLSKSLERCLERFNLARKLSATWYSYRAKRSIHSVHKTHRKGRFERVI
QHITTSVLGPKRPGGQRHCLRIERAI
```

b.)

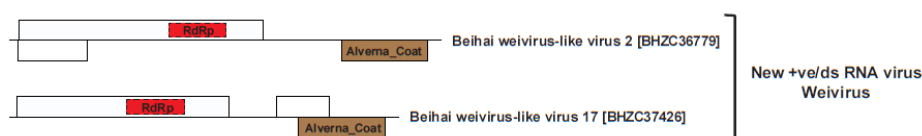


Slika 1.44 : a) Zaporedje krajšega poliproteina P3N-PIPO. Rdeče in modro označena dela se prepisujeta v različnih bralnih okvirjih, b) struktura krajšega poliproteina P3N-PIPO; oznake: Peptidase_S30 – peptidaza S30; Peptidase_C6 – peptidaza C6; Potyvirus_P3 – protein P3, specifičen za Potyviridae (Marchler – Bauer in sod., 2017 – CDD, NCBI – <https://www.ncbi.nlm.nih.gov/>).

Predpostavlja se njihov skupen izvor s Hypovirusi. Najverjetneje je bila prednica skupine Hypoviridae skupina Potyviridae, ki je prestopila gostiteljsko bariero iz rastlin na glive in izgubila kapsido (Koonin in Dolja, 2014). Filogenetska analiza jih uvršča med pikornavirusom podobno naddružino, čeprav helikaza naddružine III nadomešča helikazo naddružine II in klasična peptidaza C3 je nadomeščena s peptidazo C4. Dejstvo nakazuje na izrazito modifikacijo genomske strukture pikornavirusom podobnega prednika (Koonin in sod., 2015).

1.2.1.21 Weivirus

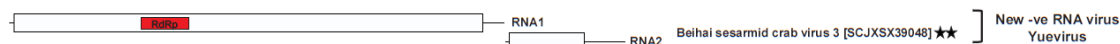
Weivirusi so ss(+)RNA virusi, najdeni pri nevretenčarjih, z linearnim monopartitnim genomom, ki ga sestavlja dva ORF. Prvi predstavlja RdRp, drugi predstavlja protein kapside, ki je homologen kapsidnim proteinom iz alvernavirusne družine (Shi in sod., 2016a) (Slika 1.45).



Slika 1.45: Genomska struktura Weivirusov (Shi in sod., 2016a). Oznaka: Alverna_Coat – protein plašča homologen proteinu plašča skupine Alvernaviridae. Angleški izraz new – nov.

1.2.1.22 Yuevirus

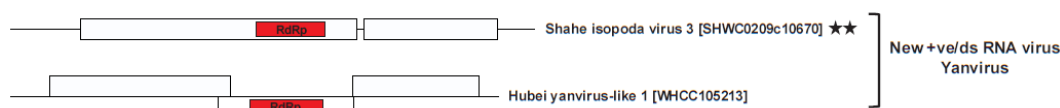
Yuevirusi so ss(-)RNA virusi, najdeni pri nevretenčarjih, z linearnim segmentiranim genomom, ki ga sestavlja dve monocistronski RNA. Eden od segmentov kodira zapis za RdRp (Shi in sod., 2016a) (Slika 1.46).



Slika 1.46: Genomska struktura Yuevirusov (Shi in sod., 2016a). Angleški izraz new – nov.

1.2.1.23 Yanvirus

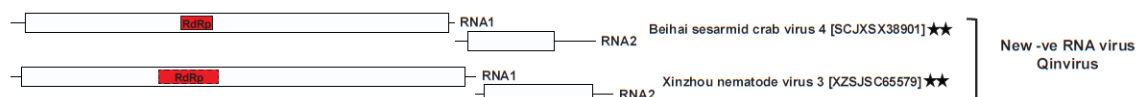
Yanvirusi so ss(+)RNA virusi, najdeni pri nevretenčarjih, z linearnim monopartitnim genomom, ki ga sestavlja dva ali trije ORF. Eden izmed njih predstavlja RdRp (Shi in sod., 2016a) (Slika 1.47).



Slika 1.47: Genomska struktura Yanvirusov (Shi in sod., 2016a). Angleški izraz new – nov.

1.2.1.24 Qinivirus

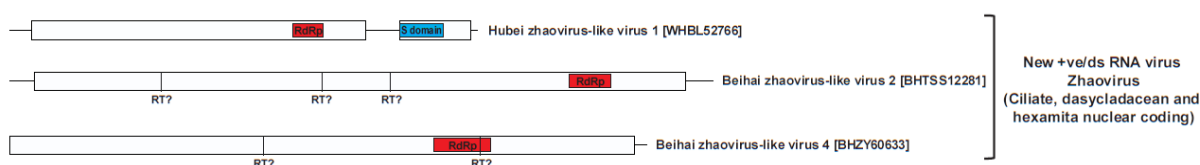
Qinivirusi so ss(-)RNA virusi, najdeni pri nevretenčarjih z genomom, ki ga sestavlja dva segmenta. Prvi segment predstavlja RdRp (Shi in sod., 2016a) (Slika 1.48).



Slika 1.48: Genomska struktura Qinivirusov (Shi in sod., 2016a). Angleški izraz new – nov.

1.2.1.25 Zhaovirus

Zhaovirusi so ss(+) RNA virusi, najdeni pri nevretenčarjih, z linearnim genomom, ki ga načeloma kodira en ORF. ORF je slabo okarakteriziran, vseeno pa je znana domena RdRp. Nekateri zhaovirusi, kot je Hubei zhaovirus-like virus 1, so prevzeli »Viral coat protein (S domain) (S_domain)«, ki je značilen za rastlinske viruse (Shi in sod., 2016a) (Slika 1.49).



Slika 1.49: Genomska struktura Zhaovirusov (Shi in sod., 2016a). Oznaka RT – angl. »stop codon read-through«. Angleška izraza: new – nov; ciliate, dasycladacean and hexamita nuclear coding – ciliatni kod.

1.2.2 Od RNA odvisna RNA polimeraza (RdRp)

Od RNA odvisna RNA polimeraza (RdRp) spada v naddružino od matrice odvisnih polimeraz nukleinskih kislin (angl. »*template-dependent nucleic acid polymerases*«) (te Velthuis, 2014). Odgovorna je za podvojevanje virusov izključno na osnovi RNA (Ferrer-Orta in sod., 2006). Glede na sprejeto hipotezo o izvoru virusov naj bi bil mehanizem podvojevanja izključno na osnovi RNA predhodnik mehanizmu podvojevanja na osnovi obratnega prepisovanja RNA v DNA in mehanizma podvojevanja izključno na osnovi DNA. RNA viroidi, virusoidi in virusi naj bi bili direktni potomci zadnjega skupnega prednika – LUCA (angl. »*Last Universal Common Ancestor*«) (Koonin in sod., 2015). RdRp ss(+)RNA bakteriofagov (Leviviridae) naj bi bili najbližji sorodniki retrovirusnih od RNA odvisnih DNA polimeraz (Černý in sod., 2014). Ustresen začetek sinteze komplementarne verige RNA v procesu podvojevanja je ključnega pomena za ohranjanje integritete virusnega genoma (Ferrer-Orta in sod., 2006). Obstajata dva glavna mehanizma za začetek podvojevanja: 1) od začetnih oligonukleotidov odvisni in 2) od začetnih oligonukleotidov

neodvisni mehanizem podvojevanja. Pri prvem mehanizmu podvojevanja 3' hidroksilno skupino prispeva RNA začetni oligonukleotid ali proteinski oligopeptid, na katerega se nato z nukleofilno adicijo dodajajo nadaljnji nukleotidi glede na matrico. Pri drugem oz. *de novo* mehanizmu prvi nukleotid prispeva 3' hidroksilno skupino za nukleofilno adicijo (Ferrer-Orta in sod., 2006).

RdRp je edini protein, ki je prisoten pri vseh RNA virusih in virusom podobnih elementih, nima pa ustreznega homologa v celičnih organizmih, ki bi imel enako encimsko funkcijo. Celični organizmi sicer kodirajo RdRp družino proteinov, ki je pri njih vključena v pomnoževanje siRNA (angl. »*small interfering RNA*«). RdRp so (oddaljeno) sorodne z od RNA odvisno DNA polimerazo (reverzno transkriptazo) in z od DNA odvisno DNA polimerazo. Hkrati so najstarejša skupina polimeraz s tipično sekundarno strukturo »dlani« (angl. »*palm domain-containing polymerases*«) (te Velthuis, 2014; Koonin in sod., 2015). Velikost RdRp je do 400 aa. So zelo variabilni proteini z manj kot 10 % ohranjenostjo. Hkrati so določene regije znotraj RdRp močno ohranjene. Kooninova ekipa je bila prva ekipa, ki je identificirala ohranjene regije znotraj RdRp. Označili so jih z zaporednimi števili I–VIII (Koonin in sod., 1989; Koonin, 1991, Černý in sod., 2014; Koonin in sod., 2015). Ugotovili so, da je znotraj motivov pri ss(+) in dsRNA virusih ohranjenih samo nekaj aminokislinskih ostankov: Lys3 znotraj motiva I, Asp118 in Asp124 znotraj motiva IV, Ser/Thr190, Gly191, Thr195, Asn/Glu199, Ser/Thr200 znotraj motiva V in Asp268–Asp269 znotraj motiva VI, Phe339, Cys340 znotraj motiva VII in Lys/Arg365 znotraj motiva VIII. Motiv VI predstavlja katalitično domeno od RNA odvisne RNA polimeraze. Pri ss(+)RNA in dsRNA virusih je tipična triada aminokislin GDD, torej Gly267–Asp268–Asp269 (Koonin in sod., 1989; Koonin, 1991; Koonin in sod., 2015). Glicin na mestu 267 lahko zamenja serin, aspartat na mestu 269 (D, Asp) pa lahko zamenja asparagin (N, Asn). Podobni motivi so bili najdeni tudi znotraj dsRNA RdRp in reverznih transkriptaz. Presenetljivo je, da imajo RdRp ss(–)RNA virusov podobne motive, se pa na podlagi zaporedja razlikujejo od RdRp ss(+)RNA, dsRNA virusov in retrovirusov oz. retroelementov (Koonin in sod., 2015). te Velthuis je imenoval sedem ohranjenih regij z značilnimi motivi: Motivi A–F (te Velthuis, 2014), ki si od N proti C terminalnemu delu sledijo v nizu Motiv G, motiv F (1–3), motiv A, motiv B, motiv C, motiv D, motiv E (te Velthuis, 2014). Motivi A, C in D so najpomembnejši motivi na RdRp. Skupaj tvorijo aktivno mesto RdRp. Motiv C predstavlja glavno katalitično jedro od RNA odvisnih RNA polimeraz. Ustreza motivu VI, kot ga je poimenovala Kooninova ekipa, in tipično vsebuje dva zaporedna aspartata za glicinom (motiv GDD). Glicin se je tekom evolucije lahko modificiral v serin (motiv SDD) ali metionin (motiv MDD). Drugi aspartat v motivu je lahko mutiral v asparagin (motiv GDN) (Mononegavirales in Chuvirusi). Motiv A je drugi najpomembnejši motiv na RdRp z ohranjenima dvema aspartatoma v motivu D[x]ND. Aspartata pri obeh motivih vežeta po en magnezijev ion, ki katalizira reakcijo nastanka fosfodiesterne vezi. N-terminalni aspartat v motivu C preko magnezijevega iona namesti α -fosfat na NTP (nukleozid trifosfati) za nukleofilni napad 3'OH skupine na zadnji ribozi novonastajajoče verige. Aminoterminalni aspartat motiva A preko magnezijevega iona namesti preostala dva fosfata (β - in γ -fosfat) na izbranem NTP za nukleofilni napad 3'OH skupine na ribozi na α -fosfat, pri čemer nastane pentakovalentni intermediat. Entropija reakcije nato poteče v smeri nastanka fosfodiesterne vezi (te Velthuis, 2014). Na novo dodan NTP nato prispeva 3'OH skupino riboze za nukleofilni napad na naslednji NTP (glede na matrico). Motiv D z ohranjenim lizinom ali histidinom pomaga aminoterminalnemu aspartatu na motivu A namestiti β -fosfat NTP-ja. Hkrati koordinira izhod pirofosfata (PPi) po reakciji. Pred samo reakcijo mora biti prepoznan pravi NTP glede na matrico RNA. Delno nalogo opravi motiv D, ki z vezavo β -fosfata sproži zapiranje aktivnega mesta. V pravilno konformacijo za vezavo pa lahko preide le, če je NTP glede na matrico pravi. Pred tem motiv B preko ohranjenih aminokislinskih ostankov hkrati loči med NTP in dNTP in veže nukleotidno bazo vstopnega NTP. Pri razlikovanju med NTP and dNTP pomagata ohranjena asparagin in C-terminalni aspartat. Motiv F pri tem pomaga namestiti fosfate na NTP in postavi NTP v pravi položaj za prepoznavo strani motiva B (te Velthuis, 2014).

Vsak od omenjenih sedmih ohranjenih motivov na RdRp zavzame predvideno sekundarno strukturo, ki je bolj ohranjena kot sama sekvenca motiva. Ohranjene sekundarne strukture pripadajočega motiva, ki pripada dvema različnima RdRp na dveh sorodnih virusih, imenujemo homomorfi (angl. »*homomorphs*«). Omenjeni ohranjeni strukturni elementi (homomorfi) zavzamejo 75 % celotne RdRp. Razdeljeni so v tri ločene regije: »prsti«, »dlani« in »palec« z ohranjenim »desno-ročnim zvitjem« (angl. »*right-handed fold*«) in značilno konformacijo zaprte roke (angl. »*closed hand conformation*«)

(Ferrer-Orta in sod., 2006; Černý in sod., 2014; Černý in sod., 2015; Koonin in sod., 2015; te Velthuis, 2014).

»Prsti« zajemajo dva ključna motiva: motiv G in motiv F. Sestavljeni so iz svežnja (angl. »*bundle*«) α -heliksov. Vključijo (zapakirajo) se v veliki žleb matrice RNA. »Dlan« je ključni in najbolj ohranjen del polimeraze, ki zajema motive A, B, C in D. Sestavljena je iz trovijačnega antiparalelnega β -prepognjenega lista (angl. »*three-stranded antiparallel β -sheet*«), ki je prekrit s tremi α -heliksi. Premikanje β -vijčnic je ključnega pomena za celotno reakcijo polimerizacije, saj motiva A in C preideta v pravilno konformacijo samo ob vezavi pravilnega NTP glede na matrico. »Palec« je najbolj raznolika regija polimeraz. Zajema samo motiv E. Pri polimerazah je različnih dolžin in je sestavljen iz svežnja α -heliksov. Vsebuje polipeptidne sekvence, ki se izražajo v glavni tunel, se vključijo v mali žleb matrice RNA in omogočajo nameščanje tako NTP-jev kot RdRp na glavne motive »dlani«. Hkrati omogočajo premik matrične RdRp po prvi reakciji kondenzacije. »Prsti« so s »palcem« povezani preko zank, ki zapirajo konformacijo, zaobjemajo aktivno mesto oz. katalitično mesto (motiv C) polimeraz, ki se nahaja na »dlani«, vbočijo »dlan«, tako da nastane glavni (centralni) tunel določene dolžine, znotraj katerega se ob začetku podvojevanja ali prepisovanja namesti RNA. Palec in prsti igrajo pri tem ključno vlogo. Na hrbtni strani RdRp se nahaja majhen pozitivno nabit tunel (NTP tunel), ki deluje kot vstopno mesto za nukleotide in se nahaja pod kotom 110° glede na matrico. Vsebuje ohranjen motiv, specifičen za od RNA odvisne RNA polimeraze (Ferrer-Orta in sod., 2006; te Velthuis, 2014).

RNA, ki se namesti znotraj glavnega tunela, je lahko enoverižna (ssRNA virusi) ali dvoverižna (dsRNA virusi). V drugem primeru se v tunelu nahajajo dodatne regije, ki so odgovorne za ločitev obeh vijačnic in nalaganje prave matrice na aktivno mesto polimeraze. Enoverižna RNA je lahko zvita v sekundarne strukture. Dodatna regija na nekaterih polimerazah, ki se razteza čez »palec« in s »prsti« reagira preko β -lasnične zanke, omogoča odvijanje zvite RNA v linearno konformacijo. Dolžina in 3D struktura domen, prisotnih na RNA polimerazah, sta odvisni od mehanizma iniciacije podvojevanja; ali polimeraza koristi od začetnega oligonukleotida odvisen (angl. »*primer-dependent*«) ali neodvisen (angl. »*primer-independent or de novo initiation*«) mehanizem iniciacije podvojevanja. Začetni oligonukleotid pri od začetnega oligonukleotida odvisnem mehanizmu pomnoževanja je lahko bodisi oligonukleotid bodisi oligopeptid. Primer so pikornavirusi, ki izkoriščajo samo od oligopeptidov odvisen mehanizem podvojevanja (Ferrer-Orta in sod., 2006). Polimeraze, ki koristijo od začetnih oligonukleotidov odvisen mehanizem podvojevanja imajo načeloma širši tunel in bolj dostopno aktivno mesto in krajši palec. Polimeraze, ki izkoriščajo *de novo* mehanizem podvojevanja imajo načeloma manj dostopno oz. bolj zakrito aktivno mesto in daljši »palec«. Primer je flavivirusna polimeraza (Ferrer-Orta in sod., 2006). Nekatere polimeraze, predvsem tiste, ki se prevedejo kot fuzijski poliprotein, kot je pikornavirusna polimeraza, postanejo katalitično aktivne šele po proteolitski aktivaciji. Pikornavirusna polimeraza se prevede kot fuzijski protein s 3C proteazo, ki jo s cis delovanjem proteolitsko aktivira (Ferrer-Orta in sod., 2006). Nekatere polimeraze, kot je npr. polimeraza skupine Orthomyxoviridae, lahko uporabljajo tako *de novo* kot od začetnega oligonukleotida odvisen mehanizem pomnoževanja. Orthomyxoviridae uporablja od začetnega oligonukleotida odvisen mehanizem pomnoževanja pri transkripciji (te Velthuis, 2014).

Poznamo pet glavnih skupin polimeraz, ki jih označujemo z zaporednimi številkami; RdRp_1, RdRp_2, RdRp_3, RdRp_4, in RdRp_5. RdRp_1 je značilna za klade Picorna-Calici, Partiti-Picobirna, Cystoviridae, Hypoviridae, Astroviridae, Nodaviridae, Potyviridae in Nidovirales. RdRp_2 je značilna predvsem za Hepe-Virga klado (Bromoviridae, Virgaviridae, Togaviridae, Closteroviridae, Tymovirales, Endornaviridae, Benyviridae, Hepeviridae, Alphetetraviridae, Negevirus, Idaeovirus). RdRp_3 je značilna za viruse skupine Flaviviridae, Tombusviridae, Luteoviridae, Carmotetraviridae. RdRp_4 je tipična za Amalgaviridae, klado Toti-Chryso, Reoviridae, Sobemovirus, nekatere viruse iz skupine Luteoviridae, Megabirnaviridae, Barnaviridae. RdRp_5 ali reovirusna $\lambda 3$ polimeraza je tipična za nekatere Reoviruse. Poleg omenjene polimeraze pa je pri reovirusih prisotna tudi Orbi_RdRp, tipična za orbivirus, in seadorna_RdRp, tipična za seadornaviruse. Poleg reovirusov imajo lahko tudi druge virusne skupine lastno oz. zanje specifično polimerazo. Flavivirusi imajo lahko na mestu RdRp_3 prisotno Flavi_NS5 polimerazo, tipično samo za flavivirus. Birnavirusi imajo Birna_RdRp, ki je razširjena tudi pri Permutotetraviridae. Za Narnaviruse je značilna Mitovir_RdRp, mitovirusna RNA polimeraza. Leviviridae imajo zanje tipično RNA-replikazo B. Za ss(-)RNA viruse je značilno, da ima praktično vsaka klada svojo značilno polimerazo. Mono-Chu klada ima Mononeg_RNA polimerazo, ki je razširjena

tudi pri Ophioviridae. Bunyaviridae imajo Bunya_RdRp in Arenaviridae imajo Arena-RNA polimerazo. Orthomyxoviridae so edina virusna skupina s tripartitno RNA polimerazo, sestavljeno iz Flu_PB1, Flu_PB2 in Flu_PA domene. Flu_PB1 domena vsebuje sedem glavnih motivov F-E, vključno s katalitično domeno. Flu_PB_2 domena v procesu prepisovanja veže »pokrite oligomere« (angl. »*capped oligomers*«) gostiteljske pre-mRNA in Flu_PA vsebuje endonukleazno domeno, ki omenjene »pokrite oligonukleotide« v procesu prepisovanja proteolitsko procesira (Hulo in sod., 2011 – ViralZone; Finn in sod., 2016 – PFAM; Rawlings in sod., 2016 – MEROPS; Rawlings in Barrett, 1993; Marchler – Bauer in sod., 2017 – CDD, Thierry in sod., 2016; te Velthuis, 2014).

Virusne RNA polimeraze imajo visoko stopnjo napak pri vgradnji nukleotidov in načeloma nimajo 3'-5' eksonukleazne aktivnosti. Lahko pa izvedejo pirofosforolizo (angl. »*pyrophosphorolysis*«), pri čemer z reverzno reakcijo dodatka PPi (pirofosfata) nukleozid monofosfatu na istem aktivnem mestu omogoči odstranjevanje nukleotida iz novonastale verige. Nekatere polimeraze nadomestijo pirofosfat z NTP, ki jih namestijo z dodatnimi prisotnimi motivi, pri čemer nastane dinukleozid tetrafosfat (te Velthuis, 2014).

1.2.3 Razmnoževanje RNA virusov

Način razmnoževanja RNA virusov, ki je povezan z njihovo hitrostjo mutacij, kratkim generacijskim časom in posledično številčnostjo populacije, je eden ključnih dejavnikov, ki vodijo v modularnost njihove genomske arhitekture. Če razmnoževanje RNA virusov povežemo z genetskim driftom, ki ga ustvarja dinamična ekološka niša gostitelja, dobimo ključni element vzorca evolucije, ki vodi v njihovo raznolikost. Zato je razumevanje razlik v mehanizmi razmnoževanja posameznih skupin RNA virusov v povezavi s koevolucijskimi interakcijami z njihovimi gostitelji ključnega pomena za razumevanje njihove raznolikosti in poteka njihove evolucije. Mehanizem razmnoževanja RNA virusov je povezan s samo strukturo genoma virusa. Konstrukcija genoma RNA virusa namreč postavlja določene omejitve na spekter vseh teoretično možnih sistemov razmnoževanja in s tem narekuje njegov potek. Za neznan oz. nov RNA virus s podobno strukturo genoma kot kateri koli znan oz. že okarakteriziran virus lahko torej predvidimo njegove lastnosti z vidika razmnoževanja (Cann, 2012). Načeloma ima vsak RNA virus unikaten sistem podvojevanja, izgradnje, zorenja in zapuščanja celice, vendar imajo skupine virusov glede na Baltimorejev sistem klasifikacije v poteku razmnoževanja nekaj skupnih točk. Vsi mehanizmi razmnoževanja so sestavljeni iz sedmih ključnih faz: 1) prepoznavanje gostiteljske celice in pritrditev virusa, 2) vstop v gostiteljsko celico, 3) odstranjevanje virusne ovojnice (angl. »*envelope*«) in/ali kapside, 4) podvojevanje in prepisovanje, 5) sestavljanje virusnega delca, 6) zorenje virusnega delca in 7) izstop virusnega delca iz celice.

1.2.3.1 Prepoznavanje gostiteljske celice in pritrditev virusa

Prepoznavanje gostiteljske celice in pritrditev virusa vključuje biokemijsko interakcijo med celičnim receptorjem na gostiteljski celici in virusno molekulo na virusni ovojnici ali na kapsidi (če virus nima virusne ovojnice). Receptorji na celični površini so lahko proteini ali ogljikovi hidrati (polisaharidi), prisotni na glikoproteinih in glikolipidih. Nekateri virusi izkoriščajo več receptorjev na lastni površini in s tem lahko v celico vstopijo preko več mehanizmov. Kot primer proteinskih receptorjev na površini celic lahko navedemo integrine. So heterodimerni proteini, ki so sestavljeni iz ene α in ene β podenote, vsaka obstaja v več variantah (Xiong in sod., 2001). Nekateri enterovirusi in nekateri virusi iz rodu Parechovirus (Picornavirales) izkoriščajo VLA-2 integrin ($\alpha_2\beta_1$ integrin) za vstop v celico. Rotavirusi (Reoviridae) izkoriščajo integrine, kot so: $\alpha_2\beta_1$, $\alpha_4\beta_1$, $\alpha_v\beta_3$, $\alpha_x\beta_2$ (Gutierrez in sod., 2010). Drugi primeri celičnih receptorjev so imunoglobulinom podobne molekule (Rossmann in sod., 2002) s konstantnimi in variabilnimi regijami, kot so: CD4 (receptor za virus HIV), PVR (Poliovirus receptor), ICAM-1 (pikornavirusi), CAR (coxsackievirus) (Cann, 2012) in MOG (angl. »*myelin/oligodendrocyte glycoprotein*«) – glikoprotein, ki se izraža na površini oligodendrocitov (Pham-Dinh in sod., 1993). Tudi drugi proteini na površini celice, kot so celični transporterji (od natrija odvisen fosfatni transporter), regulatorji aktivacije komplemента (membranski kofaktorski protein CD46 pri flavivirusih), LDL podobni proteini (angl. »*Low density lipoprotein-like proteins*«) (nekateri virusi skupine Rhinovirus) ali aminopeptidaza N (pri koronavirusih) lahko delujejo kot molekule, ki jih prepoznajo virusi. Kot gradbena enota ogljikovih hidratov (polisaharidi), ki so pritrjeni na celične glikoproteine in glikolipide in jih virusi izkoriščajo za vstop v celico, nastopata sialna

kislina in heparan sulfat (Orthomyxoviridae, Coronaviridae, Reoviridae) (Hulo in sod., 2011 – ViralZone; Gutierrez in sod., 2010; Cann, 2012). Virusi se lahko preko antireceptorjev vežejo tudi direktno na holesterol (Otsuki in sod., 2018). Poseben primer prepoznavanja gostiteljev se odvije pri RNA bakteriofagih, ki izkoriščajo bakterijske fimbrije (pili) za prepoznavo gostitelja in pritrditev na gostitelja (skupini Cystoviridae in Leviviridae (Daehne in sod., 2005; Daugelavičius in sod., 2005).

Molekule na površini celic virusi prepoznajo s t. i. antireceptorji, ki so tako kot celični receptorji lahko ozko ali široko specifični in so večinoma glikoproteini. Primeri so 1) VP1, VP2 in VP3 (Rhv in CRPV proteini) pri pikornavirusih, 2) glikoproteini (angl. »spike glikoproteins«) pri koronavirusih (Hulo in sod., 2011 – ViralZone), 3) glikoproteini Erns, E1 in E2 pri flavivirusih oz. pestivirusih (Maurer in sod., 2004; Liszewski in sod., 1991), 4) glikoproteini E1/E2/E3 pri togavirusih (Cong in sod., 2011). Virusi, ki prepoznajo sialno kislino, imajo na površini virusne ovojnice trimer hemaglutinina in tetramer nevraminidaze (npr. virusi skupine Orthomyxoviridae) ali dimer hemaglutinina in nevraminidaze oz. esteraze (koronavirusi). Hemaglutinin na površini virusa veže sialno kislino. Nevraminidaza je esteraza, ki odcepi sialno kislino, kar sproži vstop v gostiteljsko celico. Nekateri virusi imajo na površini vezavni protein, ki ima tako hemaglutininsko kot nevraminidazno funkcijo (Paramyxoviridae). Sialna kislina ostane vezana na hemaglutinin, kar preprečuje interakcijo z drugim receptorjem (Cann, 2012, Hulo in sod., 2011 – ViralZone). dsRNA virusi imajo v mehanizmu razmnoževanja dodaten izziv, saj celotno razmnoževanje poteka v citoplazmi znotraj virusne kapside. Zato imajo načeloma dve ali tri koncentrične kapside. Ključno vlogo pri prepoznavanju celičnih receptorjev ima zunanja kapsida, ki se odstrani tekom prehoda virusa v celico. Pri rastlinskih virusih predstavlja rastlinska celična stena oviro pri prepoznavanju proteinov na plazmalemi. Zato rastlinski virusi večinoma izkoristijo integriteto celične stene za direkten prehod v celico (Cann, 2012). Kompleksna struktura bakterijskih ovojnic predstavlja dodaten izziv za bakterijske RNA viruse (Daehne in sod., 2005; Daugelavičius in sod., 2005; Hulo in sod., 2011 – ViralZone).

1.2.3.2 Vstop v gostiteljsko celico

Vstop v gostiteljsko celico se odvije v kratkem časovnem intervalu po interakciji receptorja na celični membrani in pripadajočega antireceptorja na virusu. Pri procesu se porablja ATP, generirana v gostiteljski celici, zato mora biti gostiteljska celica metabolno aktivna (Cann, 2012). Poznani so štiri primarni mehanizmi: 1) virus lahko direktno preide čez plazmalemo gostiteljske celice, 2) virus v celico preide z receptorji posredovano endocitozo, 3) virusna nukleokapsida (jedrni delec brez virusne ovojnice) v celico preide preko fuzije virusne ovojnice s celično plazmalemo in 4) virusi v celico preidejo preko oblikovanja por in injiciranja genoma direktno v citoplazmo gostiteljske celice. Virusom podobni delci brez kapside najverjetneje preidejo celično membrano z vertikalnim prenosom genoma z materinske na hčerinsko celico med delitvijo. Nekateri virusi lahko izkoriščajo več mehanizmov prehoda v gostiteljsko celico. Poseben primer vstopa v gostiteljsko celico so dsRNA virusi. Razen skupine Cystoviridae nimajo virusne ovojnice, nekatere virusne skupine (Endornaviridae, Hypoviridae) ravno tako nimajo kapside. Skupini Reoviridae in Cystoviridae imata dve ali več koncentričnih kapsid (Forzan in sod., 2007; Gutiérrez in sod., 2010; Daugelavičius in sod., 2005).

Proces **direktnega prehoda čez plazmalemo** je relativno redek proces in je še relativno neraziskan. Proces **receptorsko posredovane endocitoze** je najbolj razširjen mehanizem prehoda med virusi. Znana sta dva ločena mehanizma z receptorsko posredovano endocitozo: s klatrinom posredovane endocitoze (angl. »clathrin-mediated endocytosis«) in s kaveolinom posredovane endocitoze (angl. »caveolin-mediated endocytosis«). Proces s klatrinom posredovane endocitoze je bolj razširjen mehanizem prehoda v celico med virusi kot s kaveolinom posredovana endocitoza. Gre hkrati za enega najboljše opisanih in univerzalno razširjenih procesov v gostiteljskih celicah (Traub in Bonifacio, 2013; Maib in sod., 2017). S klatrinom posredovana endocitoza je bila dokazana za številne viruse: vezikularni stomatitis virus (angl. »Vesicular stomatitis virus«) (Mononegavirales – Rhabdoviridae), hepatitis C virus (Flaviviridae), respiratorni sincicijski virus (angl. »Respiratory syncytial virus«) (Mononegavirales – Pneumoviridae), Ebola virus (Mononegavirales – Filoviridae) in virus japonskega encefalitisa (angl. »Japanese encephalitis virus«) (Flaviviridae) (Chuang in sod., 2015). S kaveolinom posredovana endocitoza je druga najbolj razširjena oblika vstopa virusnega genoma v celice v okviru

receptorsko posredovane endocitoze. Gre za od klatrina neodvisno (angl. »clathrin-independent«) pot invaginacije membrane ob interakciji celičnega receptorja z antireceptorjem. Vezava celičnega receptorja z antireceptorjem inducira koncentriranje receptorjev v glikolipidnih raftih – s sfingolipidi in holesterolom bogatih membranskih mikrodomenah (Nabi in Le, 2003; Kiss in Botos, 2009). Nekateri RNA virusi izkoriščajo več alternativnih poti za vstop v celico, kot so virusi skupine Togaviridae. V celico lahko preidejo bodisi s klatrino posredovano endocitozo bodisi s kaveolinom posredovano endocitozo (Otsuki in sod., 2018; Cong in sod., 2011; Trinh in sod., 2018; Carvalho in sod., 2017; Sousa in sod., 2017). Poliovirus (Picornavirales) v gostiteljsko celico preide preko dveh mehanizmov z receptorji posredovane endocitoze. Prvi mehanizem vključuje s kaveolinom posredovano endocitozo (Coyne in sod., 2007; Bergelson, 2008). Drugi mehanizem je neodvisen tako od kaveolina kot od klatrina. Vezikli, ki nastanejo, so strukturno ločeni od klatrinskih in kaveolinskih veziklov (Brandenburg in sod., 2007).

Fuzija virusne ovojnice z membrano gostiteljske celice poteka izključno pri virusih, ki vsebujejo virusno ovojnico (Cann, 2012). V tem primeru je odstranjevanje virusne ovojnice povezano z vstopom v gostiteljsko celico. Pri tem ima pomembno vlogo poleg fuzijskih proteinov na virusni ovojnici, kot je protein gB pri skupini Herpesvirales (dsDNA virusi), tudi holesterol (Wudiri in sod., 2017; Yang in sod., 2017). Mehanizem je redek pri RNA virusih. Opisan je bil za skupino Paramyxoviridae (Mononegavirales) (Smith in sod., 2009).

Nekateri RNA virusi brez virusne ovojnice s proteini na celični membrani oblikujejo pore, skozi katere **injicirajo genom direktno v citoplazmo**. Eden najbolj znanih mehanizmov direktnega injiciranja genoma v gostiteljsko celico je pri pikornavirusih. Izkoriščajo imunoglobulinom podobne receptorje za prehod v gostiteljsko celico, kot so: ICAM-1, CAR (angl. »coxsackievirus-adenovirus receptor«) in PVR (angl. »poliovirus receptor«). Razlikujejo se po številu imunoglobulinskih domen. Receptor ICAM-1 ima pet (D1-D5), receptor CAR ima dve (D1, D2), receptor PV ima tri imunoglobulinske domene (D1-D3). Domena D1 vsebuje prepoznavno mesto za virus. Kapsida pikornavirusov je strukturirana iz heterotrimernih kompleksov VP1, VP2, VP3, urejenih v psevdo 3 ikozahedralno simetrijo. Na notranji strani je VP1/VP2/VP3 kompleks pritrjen na VP4, ki oblikuje desetvišjačni antiparalelni β -sodček (»ten-stranded anti-parallel β barrel«). Na genomu različnih skupin pikornavirusov se lahko nahajajo v različnih razporeditvah. Na genomu rinovirusov (angl. »human rhinovirus«) so npr. organizirane kot VP4-VP2-VP3-VP1. VP4 je prepoznana kot Pico_P1A, VP1, VP2 in VP3 pa kot tri Rhv domene. VP1, 2 in 3 so prepoznane kot Rhv ali kot CRPV domene (angl. »CRPV capsid protein«). V genomu iflavirusov se domene nahajajo v nizu VP3-VP4-VP1-VP2. VP2 je prepoznana kot CRPV domena (angl. »CRPV capsid protein«). V genomu dicistrovirusov se nahajajo na drugem ORF-ju na nizu VP1-VP4-VP2-VP3 oz. CP1-CP4-CP2-CP3. VP4 (CP4) je pri dicistrovirusih prepoznana kot Dicistro_VP4 domena, CP3 (VP3) pa kot CRPV domena. Proteini na površini pikornavirusov oblikujejo prehod za virusno RNA, ki ji sledi njen izhod iz virusa v gostiteljsko celico (Hulo in sod., 2011 – ViralZone, Rossmann in sod., 2002; Kolatkar in sod., 1999; Marchler-Bauer in sod., 2017 – CDD). Glede na biofizikalno in biokemijsko nestabilnost RNA molekule in prisotnost številnih celičnih obrambnih mehanizmov pred tujimi ribonukleinskimi kislinami in njeno dovzetnost so pikornavirusi razvili mehanizme, s katerimi je RNA stabilizirana in zaščitena pred razgrajevalnim celičnim biokemijskim aparatom (Ullmer in Semler, 2016). Skupina Leviviridae je ravno tako primer ss(+)RNA virusov, ki injicirajo genom direktno v gostitelja (Hulo in sod., 2011 – ViralZone; Daehne in sod., 2005).

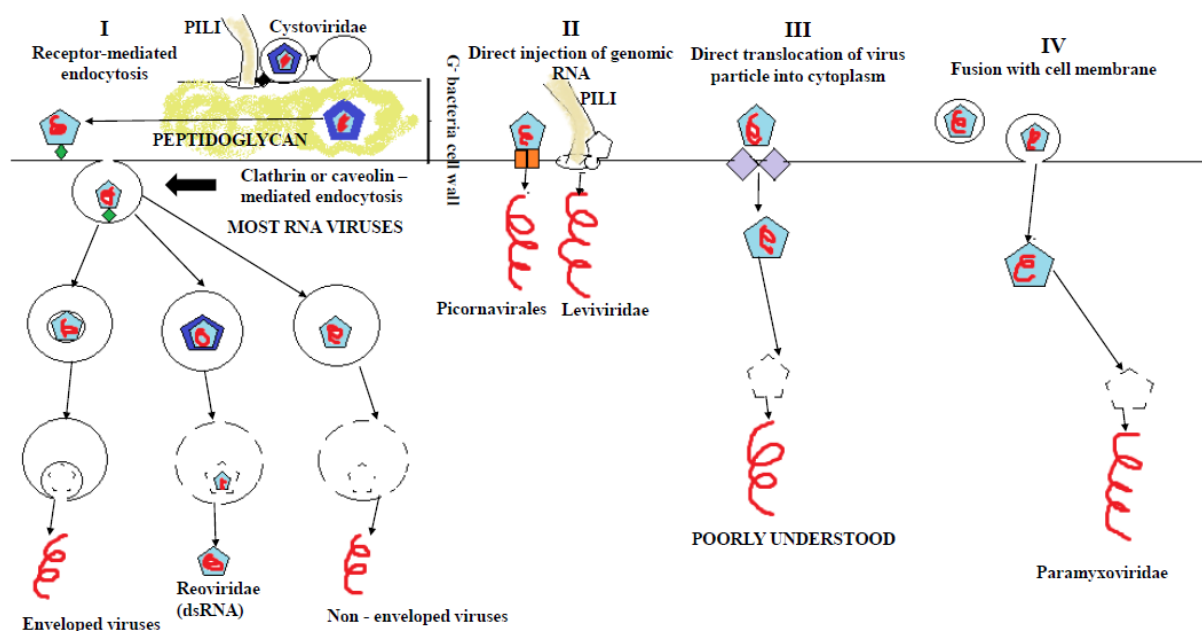
1.2.3.3 Odstranjevanje virusne ovojnice in/ali kapside

Odstranjevanje plašča in/ali virusne ovojnice je proces, ki je močno odvisen od mehanskih in biokemijskih lastnosti virusne kapside in virusne ovojnice. Virusi so fizikalno metastabilni delci. Že majhne »motnje« virusa kot sistema (temperatura, sprememba pH, sprememba zunanega tlaka, sprememba redoks stanja, proteolitična aktivnost) povzročijo odprtje oz. uničenje kapside (Greber, 2016).

Najbolj razširjena oblika vstopa v celico med virusi je receptorsko-posredovana endocitoza. Po vstopu virusa v celico se endosomi z virusnimi delci združijo z lizosomi, pri čemer pride do močnega zakisanja pH znotraj veziklov (Cann, 2012; Greber, 2016). Pri tem pozitivno nabiti oksonijevi (H3O⁺) ioni preidejo skozi virusno ovojnico in kapsido (lahko preko ionskih kanalov) in na osnovi od pH odvisnih konformacijskih sprememb reorganizirajo sekundarno strukturo virusnega genoma z vezanimi proteini (običajno

nukleoproteini). Proces je bil opisan za številne virusne skupine, kot so: Orthomyxoviridae (Greber, 2016, Banerjee in sod., 2014; Stauffer in sod., 2014), Coronaviridae (Hulo in sod., 2011 – ViralZone; Burkard in sod., 2014), Togaviridae (Hulo in sod., 2011 – ViralZone; Sousa in sod., 2017), Rhabdoviridae (Hulo in sod., 2011 – ViralZone; Carneiro in sod., 2001), Bunyaviridae (Lozach in sod., 2010), polioviruse (Brandenburg in sod., 2007) in skupino Reoviridae (Forzan in sod., 2007; Gutiérrez in sod., 2010; Hulo in sod., 2011 – ViralZone; Baker in Prasad, 2010). Orthomyxoviridae so primer z virusno ovojnico ovitih virusov, ki imajo linearno RNA sestavljeno iz šestih ali osmih segmentov. Vsak segment je zapakiran v heliksno sekundarno strukturo (ribonukleoproteine), pri kateri je RNA vezana z nukleoproteini in virusno tripartitno RdRp (PB1, PB2 in PA). Struktura je vezana v kapsidi podoben kompleks (ribonukleokapsido) s proteinom M1. Ribonukleokapsida je pri virusu obdana z M2 proteinom, hemaglutininom in nevraminidazo (ki tvorijo virusno ovojnico). Pri skupini Orthomyxoviridae prehod oksonijevih ionov skozi ionski kanal M2 povzroči konformacijske spremembe v proteinu M1, kar povzroči njegovo disociacijo z ribonukleoproteinskega kompleksa. Celotna struktura virusa vključno z virusno ovojnico se razrahlja in virus preide v metastabilno stanje. Z nadaljnjim padanjem pH v poznih endosomih pride do konformacijske spremembe hemaglutinina. Na hemaglutininu se izpostavi fuzijski peptid, ki se veže z endosomalno membrano. Hkrati protein M1 disociira in ribonukleoprotein preide v citosol (Greber, 2016, Banerjee in sod., 2014; Stauffer in sod., 2014). Orthomyxovirusi imajo tudi vezane in proste verige monoubikvitina, di-ubikvitina, in poliubikvitina (tri- in tetraubikvitina). Verige poliubikvitina se razgradijo ob povečanju koncentracije kalijevih ionov. Prosti ubikvitin deluje kot aktivator za histonsko deacetilazo HDAC6, ki se veže na dinein (angl. »dynein«) in dinaktin (angl. »dynactin«) v kompleks z aktinskimi in miozinskimi filamenti in aktivira mikrotubulski motor. Ribonukleoproteinski kompleks, vezan s filamenti, nato s pomočjo aktinskih in miozinskih filamentov v povezavi z dineinom, dinaktinom in HDAC6 disociira z mesta fuzije virusa z endosomalno membrano (Greber, 2016, Banerjee in sod., 2014; Stauffer in sod., 2014).

Pri ostalih mehanizmih vstopa virusov v celico je odstranjevanje virusne ovojnice in kapside povezano s preходом virusa v celico. Pri fuziji virusne ovojnice s plazmalemo virus odstrani virusno ovojnico že pred prihodom v celico. V citoplazmo vstopi jedrni delec oz. nukleokapsid brez virusne ovojnice. Pri injiciranju genomskega materiala v gostiteljsko celico se je virus ravno tako znebil kapside hkrati z vstopom genomskega materiala v citoplazmo. Slika 50 prikazuje povzetek mehanizmov, ki jih virusi izkoriščajo za vstop njihovega genoma v gostiteljsko celico (**Slika 1.50**).



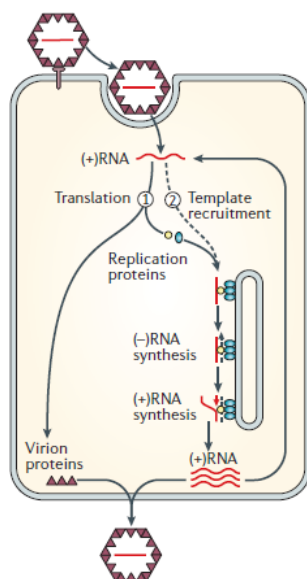
Slika 1.50: Povzetek mehanizmov vstopa virusov v gostiteljsko celico. Angleški izrazi: Receptor-mediated endocytosis – receptorsko-posredovana endocitoza, enveloped viruses – virusi z virusno ovojnico, pili – pili, peptidoglycan – peptidoglikan, clathrin or caveolin-mediated endocytosis – s klatrinom oz. kaveolinom posredovana endocitoza, most RNA viruses – večina RNA virusov, G-bacterial cell wall – celična stena gram negativnih bakterij, direct injection of genomic RNA – direktno injiciranje genomske RNA, direct translocation of virus particle into the cytoplasm – direktna translokacija virusnega delca v citoplazmo, fusion with cell membrane – fuzija s celično membrano.

1.2.3.4 Podvojevanje in prepisovanje

Po odstranitvi virusne ovojnice in/ali kapside sledi postopek pomnoževanja virusne RNA. Strategija podvojevanja in prepisovanja genomskega materiala je odvisna od narave genoma oz. matične verige RNA (Cann, 2012). Pri RNA virusih pomnoževanje in transkripcija genetskega materiala poteka znotraj t. i. virusnih tovarn (angl. »*viral factories*«). Gre za znotrajcelične oddelke (angl. »*compartments*«), ki delujejo kot zaščita virusov pred biokemijskimi obrambnimi mehanizmi gostitelja. Viroplazme so inkluzije v citoplazmi, ki jih izkoriščajo nekateri dsRNA virusi, kot so reovirusi in filovirusi (Filoviridae-Mononegavirales). Sferule so invaginacije (vpognitve) membran, ki izvirajo iz različnih celičnih organelov: peroksisomov, endoplazemskega retikla, kloroplastov, mitohondrijev, endosomov, lizosomov. Značilne so predvsem za RNA viruse, ki se pomnožujejo preko nastanka dsRNA intermediata, saj mora biti dsRNA zaščiten pred znotrajceličnim iRNA (angl. »*interference RNA*«) obrambnim mehanizmom. Tipične so za viruse Nodaviridae (mitohondriji), Narnaviridae (mitohondriji), Tymovirales (kloroplasti), Togaviridae (endosomi in lizosomi), Flaviviridae (endoplazemski retikel) in Tombusviridae (peroksisomi). Dvojni membranski vezikli so virusne tovarne, ki izvirajo iz endoplazemskega retikla in Golgijevega aparata. Tipični so za pikornaviruse, koronaviruse in arteriviruse. Tubuli (angl. »*tubes*«) so strukture, ki izvirajo iz Golgijevega aparata in so tipični za skupino Bunyavirales. Virusne tovarne se lahko oblikujejo tudi znotraj jedra. Tipične so za viruse, ki imajo mehanizem podvojevanja povezan z jedrom (Hulo in sod., 2011 – ViralZone). Načeloma ima vsak virus glede na strukturo genoma lasten mehanizem prepisovanja in podvojevanja zgodnjih in poznih ORF-jev (genov). Vseeno pa imajo tri osnovne strukture RNA molekul, ki delujejo kot matrica za nadaljnje mehanizme prepisovanja in prevajanja: 1) samo pozitivno (smerno) verigo (angl. »*positive sense RNA*«) – ss(+)RNA virusi, 2) samo negativno (protismerno) verigo (angl. »*negative sense RNA*«) – ss(-)RNA virusi ali 3) dvoverižno RNA molekulo (angl. »*double-stranded RNA*«) – dsRNA virusi. Genomska struktura RNA virusov je lahko enodelna (nesegmentirana) ali segmentirana. RNA ali segment RNA je lahko policistronska (vsebuje več ORF-jev, več start

in stop signalov) ali monocistronska (vsebuje samo en ORF, samo en start in stop signal). Vsak ORF lahko kodira en sam protein ali poliprotein, ki se nato proteolitsko procesira v strukturne in nestrukturne proteine.

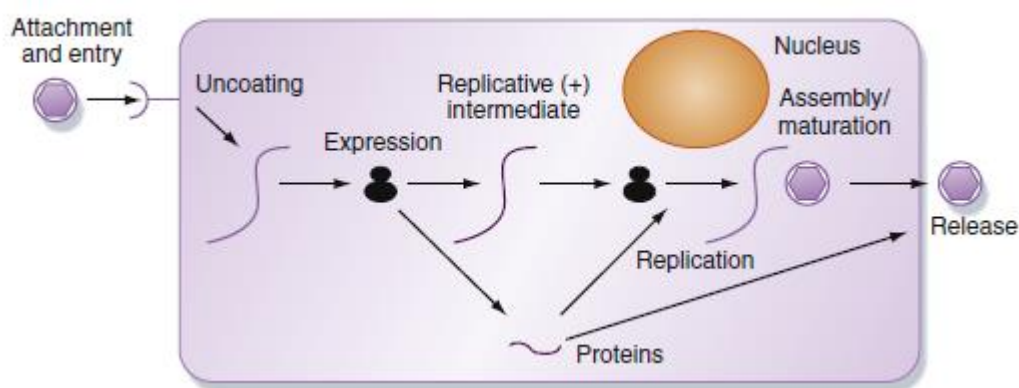
Pri ss(+)RNA virusih je v virusnih tovarnah prisotna pozitivna veriga RNA molekule, ki je »označena oz. obdana s kapo« (angl. »capped«) ali Vpg proteinom in je lahko tudi poliadenilirana. Pri določenih virusnih skupinah genomska RNA ni poliadenilirana (Nodaviridae, Permutotetraviridae, Closteroviridae, Alphatetraviridae, Carmotetraviridae in druge). Večina ss(+)RNA virusov nima genoma, kompleksiranega z RdRp. Omenjena ss(+)RNA molekula lahko služi kot matrica za prevajanje zgodnjih genov s strani ribosomov v gostiteljski celici. Mednje sodi virusna RdRp. Po prevajanju RdRp poteka podvojevanje virusnega genoma. Ponovno ss(+)RNA služi kot matrica za podvojevanje, pri čemer nastane dsRNA intermediat. Omenjen RNA intermediat (natančneje protismerna veriga) služi v nadaljevanju za prepisovanje (nastajanje) mRNA, ki je lahko identična genomski ali subgenomska. mRNA je namreč vedno v smerni (angl. »sense«) orientaciji. Po prepisovanju sledi prevajanje poznih genov. Nastale gradbene enote enkapsidirajo novonastali genom (Cann, 2012; Hulo in sod., 2011 – ViralZone). Pri flavivirusih ss(+)RNA potuje do ribosomov na zrnatem endoplazemskem retiklu, kjer poteka prevajanje poliproteina. Po prevajanju se na endoplazemskem retiklu oblikujejo membranski vezikli. V lumnu membranskih veziklov poteka podvojevanje RNA flavivirusov in kasnejše sestavljanje strukturnih in nestrukturnih enot virusa (Apte-Sengupta in sod., 2014). Slika 51 prikazuje pomnoževanje, prepisovanje in prevajanje genoma ss(+)RNA virusov (**Slika 1.51**).



Slika 1.51: Podvojevanje, prepisovanje in prevajanje ss(+)RNA virusov (Ahlquist, 2006). Angleški izrazi: Translation – prevajanje, Template recruitment – rekrutiranje matrice, replication proteins – replikacijski protein, synthesis – sinteza, virion proteins – protein viriona.

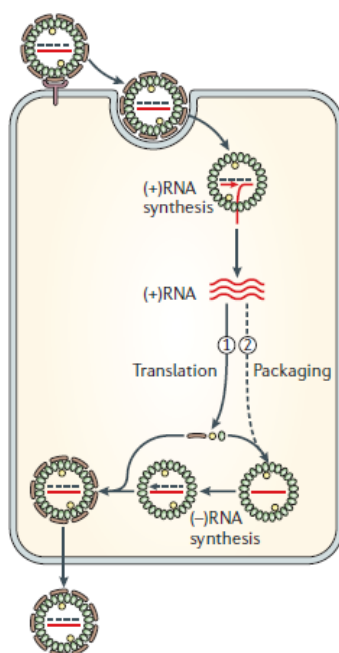
Pri ss(-)RNA virusih je v virusnih tovarnah prisotna protismerna veriga RNA, ki je običajno kompleksirana z nukleoproteini in RdRp. Je značilno označena s kapo (7-metilgvanozin) in vsebuje poli-U zaporedje (angl. »U-strech«). RdRp ss(-)RNA virusov loči dva stanja: stanje prepisovanja in stanje podvojevanja. Prepisovanje je prvi korak v ciklu podvojevanja, prepisovanja in prevajanja genoma ss(-)RNA virusov. Za ss(-)RNA viruse je značilna zaporedna (sekvenčna) transkripcija (angl. »sequential transcription«) oz. sinteza antigenoma. RdRp lahko prične sintezo antigenoma (smerna veriga RNA) na mestu +1 ali +3. V primeru, da se sinteza začne na mestu +3 (3' N-terminalnem delu genoma), nastane 20–25 nukleotidov dolgo 5'-trifosfat vodilno zaporedje RNA (Hulo in sod., 2011 – ViralZone; Harwig in sod., 2017). Nato se ustavi in ponovno prične prepisovanje na prvem start signalu. Ko RdRp naleti na stop signal, se ustavi in pomakne za en uracil nazaj, pri čemer nastane poli-A signal. Policistronske RNA molekule (z več start in stop signali) se lahko prepišejo kot več subgenomskih mRNA, pri čemer je vsaka monocistronska, metilirana (7-

metilgvanozin) in poliadenilirana. Segmentirana RNA ss(-)RNA virusov je običajno monocistronska in se prepisuje kot genomska RNA. Transkripciji sledi prevajanje z izkoriščanjem gostiteljskega biokemijskega aparata. Podvojevanje genomskega materiala se običajno prične, ko se tvori dovolj nukleoproteina, ki je običajno prvi N-terminalni ORF v genomu. V procesu podvojevanja se RdRp ponovno veže na vodilno zaporedje. Sestavljeno je iz dveh faz. Najprej nastane antigenom, ki se enkapsidira. Ko je RdRp v stanju podvojevanja, ne upošteva dodatnih start in stop signalov ORF-jev, saj mora nastati popolnoma enak komplementaren genom. V drugi fazi se cikel podvojevanja ponovi. Tokrat antigenom služi kot matrica, pri čemer nastane enak genom izvirnemu RNA genomu. Gradbene enote, ki so nastale v fazi zaporedne transkripcije in prevajanja, tokrat enkapsidirajo novonastali genom (Cann, 2012; Hulo in sod., 2011 – ViralZone). Slika 52 prikazuje podvojevanje, prepisovanje in prevajanje genoma ss(-)RNA virusov (Slika 1.52).



Slika 1.52: Podvojevanje, prepisovanje in prevajanje ss(-)RNA virusov (Cann, 2012). Angleški izrazi: Attachment and entry – vezava in vstop, uncoating – odstranjevanje virusnega plašča in/ali kapside, expression – izražanje, replicative intermediate – podvojitveni intermediat, replication – podvojevanje, proteins – protein, assembly – sestavljanje, release – sproščanje virusnega delca.

Podvojevanje dsRNA molekul poteka znotraj virusnega delca. RdRp, ki je že kompleksirana z dsRNA genomom virusa, najprej uporabi protismerno matrico za sintezo mRNA (vedno v smerni orientaciji), ki v tej fazi vsebuje kapo (7-metilgvanozin) in potuje izven virusnih delcev, kjer hkrati služi najprej kot matrica za prevajanje, pri čemer nastanejo gradbene enote virusa, med drugim tudi virusna RdRp. Pri tem se mRNA v smerni orientaciji skupaj z RdRp enkapsidira v inkluzijskih telesih (Cann, 2012; Hulo in sod., 2011 – ViralZone). Znotraj inkluzijskih teles RdRp preide v stanje podvojevanja, pri katerem smerna veriga RNA služi kot matrica za nastanek dsRNA molekul (sintezo komplementarne verige). Sledi sestavljanje virusnega delca (Mohl in Roy, 2014). Pri Orbivirusih (Reoviridae) po vstopu virusa v citoplazmo virus sestavlja 10 segmentov genomske RNA in proteini VP1 (RdRp), VP4 (angl. »capping enzyme«) s funkcijo RNA trifosfataze, gvanililtransferaze, gvanin-N7-metiltransferaze in nukleozid-2'-O-metiltransferaze, VP6 (helikaza) in proteina notranje kapside VP3 in VP7. VP1 polimeraza prepiše segmente, pri čemer nastane metilirana (7-metilgvanozin) mRNA. mRNA nato potuje izven jedrnega delca in služi kot matrica za prevajanje 11 virusnih proteinov, pri čemer nastanejo inkluzijska telesa, sestavljena predvsem iz NS2. Znotraj inkluzijskih teles nato poteka podvojevanje genoma, kar pomeni sintezo komplementarne RNA verige (Mohl in Roy, 2014). Slika 53 prikazuje pomnoževanje, prepisovanje in prevajanje genoma dsRNA virusov (Slika 1.53).



Slika 1.53: Podvojevanje, prepisovanje in prevajanje dsRNA virusov (Ahlquist, 2006). Angleški izrazi: synthesis – sinteza, translation – prevajanje, packaging – pakiranje.

Pri vseh treh tipih RNA virusov gre torej za zaporedna mehanizma prepisovanja in prevajanja, pri katerih se mora iz genomskega materiala tvoriti monocistronska mRNA v smerni (angl. »sense«) orientaciji. ss(+)RNA virusi imajo startni genom že v smerni orientaciji. Tako lahko njihova genomska RNA že služi kot matrica za prevajanje proteinov. Zato je pri njih faza prevajanja prva faza v ciklu razmnoževanja. Mehanizem, ki ga virus izkoristi pri prevajanju, je odvisen od strukture N-terminalnih in C-terminalnih regij virusov. mRNA virusov je lahko na 5' ovita s kapo (angl. »cap«) in poliadenilirana na 3', pri čemer oponaša strukturo klasične mRNA gostitelja. Strukturo kape predstavlja 7-metilgvanozin (m7G), ki je preko 5'-5' trifosfatnega mostička pripet na prvi nukleotid transkripta. Njena vloga je zaščita mRNA pred razgradnjo s 5'-eksonukleazami v procesirajočih telesih (angl. »processing bodies«) in usmerjanje izhoda mRNA iz jedra (Decroly in sod., 2011). Za nekatere skupine virusov je značilen klasičen mehanizem zaščite mRNA s kapo. Pri tem izkoriščajo bodisi svoj (Reoviridae, Bornaviridae, Rhabdoviridae, Filoviridae) bodisi gostiteljski biokemijski aparat za sintezo kape. Seleksijski pritiski gostiteljskega biokemijskega obrambnega mehanizma so prispevali k tendenci virusov po razvoju alternativnih mehanizmov zaščite mRNA molekul. Nekateri virusi ne izkoriščajo niti lastnih niti gostiteljskih sistemov za tvorbo kape, temveč jo »ukradejo« z gostiteljskih mRNA (angl. »cap-snatching«). Tipična je za skupine Arenaviridae, Bunyaviridae in Orthomyxoviridae. Vključuje tri korake: 1) virusna RdRp veže kapo gostiteljske mRNA, 2) sledi odcepljanje oligonukleotidne sekvence s kapo in 3) omenjen oligonukleotid se uporabi kot začetni oligonukleotid za sintezo virusne mRNA (Hulo in sod., 2011 – ViralZone; Decroly in sod., 2011). Pri skupini Orthomyxoviridae se na enoti PB2 nahaja domena za vezavo kape, na enoti PA RdRp pa se nahaja endonukleazna domena, odgovorna za generiranje oligonukleotida s kapo, ki izvira iz gostiteljske celice (Hulo in sod., 2011 – ViralZone; Decroly in sod., 2011). Alternativno so virusi razvili mehanizem zaščite RNA s kovalentno vezavo proteina Vpg na 5' genomske RNA. Problem pri mehanizmu prevajanja virusne genomske RNA je tendenca gostiteljevega sistema po »izklopu« globalnega prevajanja v času okužbe z virusom. Nekateri virusi so se izognili omenjenemu sistemu z zaporedji IRES (angl. »Internal Ribosomal Entry Sites«), ki so po strukturi različne in omogočajo od kape neodvisno iniciacijo translacije (Balvay in sod., 2009; Decroly in sod., 2011).

Tako kot prevajanje celične mRNA ima tudi prepisovanje virusne mRNA tri ločene faze: začetek (angl. »initiation«), podaljševanje (angl. »elongation«) in zaključek (angl. »termination«). Začetek prevajanja mRNA je lahko od kape (m7GpppN) odvisen (angl. »cap-dependent«) ali od kape neodvisen (angl. »cap-independent«). Klasičen od kape odvisen

mehanizem prevajanja izkoriščajo virusi, pri katerih je mRNA zaščitena s kapo – bodisi so 1) jo »ukradli« gostiteljem, 2) so izkoristili gostiteljski biokemijski aparat za njeno sintezo ali 3) so jo sintetizirali z lastnim biokemijskim aparatom (Deniz in sod., 2009; Hulo in sod., 2011 – ViralZone; Balvay in sod., 2009; The UniProt Consortium, 2017 – UniProt). Klasična pot od kape odvisne aktivacije prevajanja mRNA zahteva prisotnost začetnega dejavnika eIF2. Togavirusi so razvili mehanizem od kape odvisnega začetka prevajanja, ki je neodvisen od dejavnika eIF2 (Hulo in sod., 2011 – ViralZone; Toribio in Ventoso, 2010). IRES elementi lahko vodijo začetek prevajanja neodvisno od prisotnosti kape. Najbolje raziskani so pri pikornavirusih (Niepmann, 2009; Deniz in sod., 2009; Hulo in sod., 2011 – ViralZone).

1.2.3.5 Sestavljanje in zorenje virusa

Po podvojevanju, prepisovanju in prevajanju sledi sestavljanje virusa. Sestavljanje virusa vključuje zbiranje vseh strukturnih in nestrukturnih sestavnih delov virusa na specifičnem delu znotraj celice. Mesto sestavljanja virusa je odvisno od mesta podvojevanja in mehanizma, s katerim virus zapusti celico. Možna sta dva alternativna scenarija: 1) virus se v celoti sestavi znotraj celice, pri čemer lahko sestavljanje poteka znotraj inkluzijskih teles v citoplazmi ali znotraj virusnih tovarn v celičnih organelih. Pri tem je lahko zorenje virusa končano, preden zapusti celico, ali pa se začne oz. nadaljuje, ko virus zapusti celico. 2) Virus svoje gradbene enote zbira na lipidnih raftih. V tem primeru sestavljanje virusa poteka vzporedno z njegovim zapuščanjem gostiteljske celice, njegovo zorenje pa izven celice (Cann, 2012).

Virusi, ki se pomnožujejo znotraj virusnih tovarn, kot so mitohondriji, endoplazemski retikel, peroksisomi, jedro in ostali, izkoriščajo preureditve njihovih membran za sestavljanje svojih strukturnih in nestrukturnih podenot. Ob sestavljanju »brstijo« iz membran, čemur sledi zapuščanje celic z eksocitozo. Pri flavivirusih se sestavljanje odvije znotraj endoplazemskega retikla (Apte-Sengupta in sod., 2014; Griffin in sod., 2003; Ariumi in sod., 2011).

Pri virusih, pri katerih razmnoževanje poteka znotraj citoplazme, se lahko sestavljanje strukturnih in nestrukturnih komponent odvije bodisi znotraj inkluzijskih teles v citoplazmi bodisi se koncentrira na citoplazemski strani lipidnih raftov (Cann, 2012). V prvem primeru bo virus sestavljen že znotraj celice, v drugem primeru pa bo sestavljanje povezano z njegovim zapuščanjem gostiteljske celice. V primeru, da se bo virus že v celoti sestavil znotraj gostiteljskih celic, bo celico zapustil z nelitično eksocitozo ali z litičnimi mehanizmi. V primeru, da se strukturne enote virusa zbirajo na lipidnih raftih, pa bo gostiteljsko celico zapustil z brstenjem. Reovirusi vsebujejo dve ali tri koncentrične kapside (Mohl in Roy, 2014). Tako skupina Reoviridae kot tudi drugi dsRNA virusi so segmentirani in se soočajo z dodatnim problemom razporeditve segmentov med virusnimi delci, kar zahteva dodatne mehanizme sortiranja segmentov. Določeni segmentirani virusi, kot je skupina Bromoviridae, enkapsidirajo samo del segmentov, kar je omejeno z velikostjo genoma, ki se pakira znotraj delca (Sicard in sod., 2016).

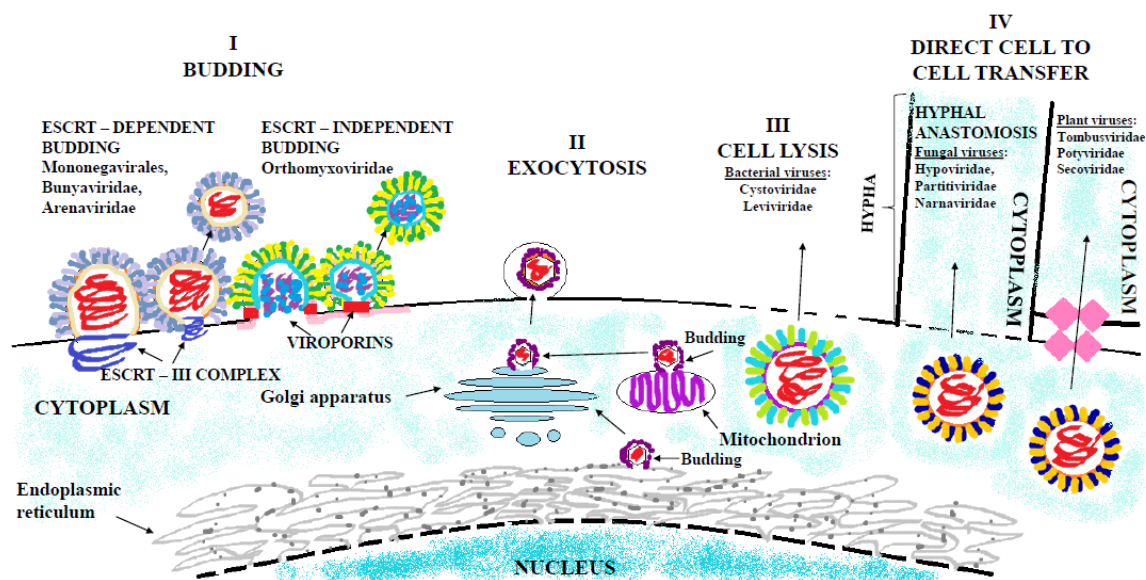
1.2.3.6 Izstop virusa iz celice

Faza zorenja virusa je lahko povezana ali ločena s fazo izstopa virusa iz celice. Mehanizmi, preko katerih virusi izstopijo iz celice, so 1) brstenje, ki je lahko od ESCRT (angl. »*endosomal sorting complex required for transport*«) odvisno ali od ESCRT neodvisno, 2) direkten prehod med celicami (hifalna anastomoza (angl. »*hyphal anastomosis*«)), 3) liza gostiteljske celice, 4) celična eksocitoza.

Od ESCRT odvisno in neodvisno brstenje sta najpogostejši obliki izhoda virusa iz celic. Tipična je za viruse s citoplazemskim mehanizmom razmnoževanja, pri katerih se gradbene enote zbirajo na lipidnih raftih. Na plazemski membrani in endosomalnih membranah se nahajajo štirje proteinski kompleksi: ESCRT-0, ESCRT-I, ESCRT-II in ESCRT-III. Pri tem naj bi ESCRT-III izvajal glavne funkcije ločevanja vezikla od gostiteljske celice. Pri virusih, ki izkoriščajo od ESCRT odvisen mehanizem brstenja, gradbene enote vsebujejo kratke L-motive (angl. »*late domain*«), ki rekrutirajo ESCR aparat, kot sta PTAP motiv (L-motiv) na p6 glikoproteinu virusa HIV-1 in L-motivi znotraj M proteinov (angl. »*matrix protein*«) skupine Mononegavirales (Irie in Harty, 2005). Na citoplazemski strani rastočih brstov virusa se kompleks ESCRT-III zbira v koncentričnih spiralah, ki krčijo vrat vezikla. Sledi disociacija ESCRT kompleksa (Hulo in sod., 2011 – ViralZone; McDonald in Martin-Serrano, 2009). Pri RNA virusih je mehanizem značilen tudi za skupine Arenaviridae, Bunyaviridae

in Paramyxoviridae (Hulo in sod., 2011 – ViralZone). Pri skupini Orthomyxoviridae so odkrili alternativni (od ESCRT neodvisen) mehanizem brstenja (Rossman in sod., 2010).

Obstaja več verzij **direktnih prehodov med celicami**. Značilen je za viruse, ki imajo kodiran MP protein (angl. »movement protein«), kot so rastlinski Tombusviridae, Potyviridae in Secoviridae (Picornavirales) (Hulo in sod., 2011 – ViralZone). Posebna oblika direktnega prenosa med celicami velja za glivne viruse – hifalna anastomoza (angl. »hyphal anastomosis«). Glivni virusi lahko izkoriščajo septe, pore v celičnih stenah znotraj hif, ki omogočajo direkten prenos virusa iz ene citoplazme v drugo. Anastomoza (združenje dveh hif) omogoča izmenjavo citoplazme in hkrati horizontalni prenos virusa iz ene citoplazme v drugo (Hulo in sod., 2011 – ViralZone). Mehanizem je tipčen za skupino Narnaviridae, Hypoviridae, Chrysoviridae in Partitiviridae. Virusi se lahko prenašajo tudi vertikalno preko spor. **Litični mehanizem** izhoda iz celic vključuje uničenje plazmaleme in celično smrt. Tipična je za bakterijske RNA viruse, kot so Leviviridae in Cystoviridae. Litične mehanizme izhoda iz celice so zabeležili tudi pri nekaterih reovirusih, kot je Bluetongue virus (Mohl in Roy., 2014). Nekateri virusi skupine Bunyaviridae, Coronaviridae, Reoviridae celico zapustijo preko Golgijevega aparata s **celično eksocitozo** (Hulo in sod., 2011). Slika 1.54 prikazuje povzetek mehanizmov izstopa virusov iz celic (Slika 1.54).

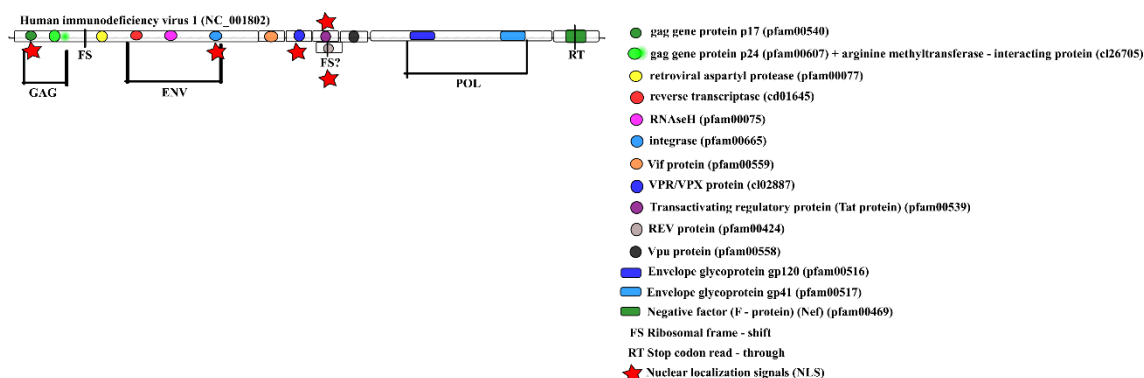


Slika 1.54: Povzetek mehanizmov izstopa virusov iz celic. Angleški izrazi: budding – brstenje, ESCRT-dependent budding – od ESCRT odvisno brstenje, ESCRT-independent budding – od ESCRT neodvisno brstenje, complex – kompleks, viroporins – viroporini, cytoplasm – citoplazma, Endoplasmic reticulum – endoplazemski retikel, Golgi apparatus – Golgijev aparat, exocytosis – eksocitoza, nucleus – jedro, cell lysis – liza celic, budding – brstenje, mitochondrion – mitohondrij, hypha – hifa, direct cell to cell transfer – direkten prehod med celicami, hyphal anastomosis – hifalna anastomoza, fungal viruses – glivni virusi, plant viruses – rastlinski virusi, cytoplasm – citoplazma.

1.2.4 Endogenizacija RNA virusov in primerjava z retroelementi

Endogenizirani virusi (angl. »endogenous viral elements«) so celotni ali deli virusnih genomov, ki so se integrirali v gostiteljski genom (Frank in Feschotte, 2017; Thézé in sod., 2014). Znani so kot virusni fosili, ki dokazujejo pretekle aktivne okužbe organizma, v čigar genom so se integrirali (Suzuki in sod., 2017). Integracija se lahko zgodi v somatskih celicah gostitelja ali spolnih celicah gostitelja. Da se virusni genom vgradi v genom gostitelja, se morajo odviti ključni koraki: 1) virus se mora pritrditi na gostiteljsko celico, 2) virus mora vstopiti v gostiteljsko celico 3) virus mora odstraniti virusno ovojnico in kapsido, 4) genom virusa mora potovati do jedra, 5) genom virusa mora preiti jedrno bariero, ki jo ustvarjajo NPC (angl. »nuclear pore complexes«) in 6) virus mora izkoristiti lasten ali celični aparat za vgradnjo genoma v gostiteljsko celico.

Retrovirusi in retrotranspozoni so najbolj poznani primeri virusov, ki uporabljajo aktivno integracijo virusnega genoma v gostiteljski genom v njihovem ciklu podvojevanja. Kot tipični primer lahko vzamemo Lentiviruse oz. virus HIV (Slika 1.55).



Slika 1.55: Genom lentivirusa HIV-1.

Genom retrovirusov je monopartitna RNA, sestavljena iz treh ključnih kodirajočih domen: 1) domene GAG, ki jo tvorita proteina p17 (protein M (angl. »matrix protein«)) in p24, 2) domene POL, ki jo tvorijo encimi reverzna transkriptaza, RNazaH in integraza in 3) domene ENV, ki tvori virusni plašč. Ključna proteina domen sta gp120 in gp41. Genom je na 5' koncu (N-terminalnem koncu) in 3' koncu (C-terminalnem koncu) ponavljajoče se regije U3-R-U5. RNA genom virusa je v kompleksu z reverzno transkriptazo, RNazaH, integrazo. Kapsido iz proteina kapside obdaja protein p17 (angl. »matrix protein«). Sledi virusna ovojnica iz dveh ključnih komponent: gp120, ki sodeluje pri vezavi CD4 in CCR5 in gp41 (Hulo in sod., 2011 – ViralZone; Marchler-Bauer in sod., 2017 – CDD), ki ob konformacijski spremembi zaradi omenjene vezave CD4/CCR5-gp120 posreduje fuzijo virusne ovojnice z membrano gostiteljskih limfocitov. Po vstopu jedrnega delca (angl. »viral core particle«) v gostiteljsko celico sledi razgradnja kapside. Sledi obratno prepisovanje virusnega genoma v komplementarno DNA (cDNA). V prvi fazi se tRNA veže na vezavno mesto za začetni oligonukleotid in generira protismerni (angl. »antisense«) fragment DNA (angl. »strong-stop DNA«), ki vsebuje ponavljajoče se regije U3-R-U5 na C-terminalnem koncu. Sledi podaljševanje protismerne DNA verige proti N-terminalnemu delu in hkrati razgradnja RNA z RNazaH. Znotraj integratne domene na centralnem delu genoma se nahaja polipurinski trakt, ki je odporen na aktivnost RNazaH. S specifično cepitvijo nastane RNA-DNA hibrid, ki sproži nastanek smiselne DNA verige. Na polipurinskemu traktu se generira začetni oligonukleotid za podaljševanje smiselne DNA verige. Pri HIV-1 se generira tudi na 5' (N-terminalnem) delu genoma in cDNA nastane kot produkt dveh polovičnih segmentov. Na polipurinskem traktu se oblikuje trovijski intermediat (angl. »DNA flap«). Nastala cDNA se s proteini p17 (angl. »matrix protein«), Vpr, integrazo, reverzno transkriptazo veže v predintegracijski kompleks, ki potuje do jedra (Sherman in Greene, 2002; Hamid in sod., 2015). Jedro sestavlja dvomembranska struktura, jedrna ovojnica, v katero so integrirane jedrne pore (NPC, angl. »nuclear pore complexes«), 125MDa strukture, ki združujejo do 100 polipeptidov (Freitas in Cunha, 2009). Med poliproteini, ki sestavljajo jedrne pore, so karioferini (angl. »karyopherins«), kot sta importin α in β . Prehod jedrnih por dovoli difuzijo delcev, manjših kot 9 nm. Večji delci morajo preiti aktivno skozi jedrne pore. Importin α , vezan z importinom β , prepozna jedrne lokalizacijske signale (angl. »nuclear localization signals«) na proteinih, namenjenih v jedro. Translokacijo proteinov, vezanih na importin α , omogoča protein Ran, ki je vezan z GTP ali GDP. Količnik RanGTP/RanGDP narašča proti jedru. RanGAP (Ran GTPase-aktivator protein) aktivira Ran s hidrolizo GTP, zato je koncentracija RanGDP večja v citoplazmi. Sledi translokacija importina s proteinom (angl. »cargo«) v jedro. V jedru RCC1 veže Ran z GTP (Freitas in Cunha, 2009; Lange in sod., 2007; Sherman in Greene, 2002). Proteini, vezani na predintegracijski kompleks, ki preidejo v jedro, morajo torej vsebovati jedrne lokalizacijske signale (angl. »nuclear localization signals«). Po vstopu predintegracijskega kompleksa v jedro sledi integracija genoma v gostiteljsko DNA.

Leta 2004 je skupina Crochu s sodelavci odkrila prvi ne-retroviralni endogeniziran fragment v genomu komarjev rodu *Aedes* (Crochu in sod., 2004). Endogenizirani fragmenti so izvirali iz RNA virusov skupine Flaviviridae. V nadaljevanju so endogenizirane genome (ali fragmente), ki izvirajo iz RNA virusov, odkrili tudi pri rastlinah, vretenčarjih, raki, glistah in členonožcih (Katzourakis in Gifford, 2010; Chiba in sod., 2011; Holmes, 2011; Horie in Tomonaga, 2011; Cui in Holmes, 2011; Feschotte in Gilbert., 2012; Fort in sod., 2012; Thézé in sod., 2014; Cotton in sod., 2016; Li in sod., 2015; Aiewsakun in Katzourakis, 2015; Shi in sod., 2016; Suzuki in sod., 2017). Tabela 1 prikazuje najdene endogenizirane genome (ali fragmente genomov), ki izvirajo iz RNA virusov pri nevretenčarjih (**Tabela 1.1**).

Tabela 1.1: Endogenizirani RNA virusi, ki jih opisuje literatura. 1 – (Shi in sod., 2016a), 2 – (Li in sod., 2015), 3 – (Cotton in sod., 2016), 4 – (Katzourakis in Gifford, 2010), 5 – (Theze in sod., 2014), 6 – (Suzuki in sod., 2017), 7 – (Crochu in sod., 2004), 8 – (Aiewsakun in Katzourakis, 2015), 9 – (Cui in Holmes, 2011), 10 – (Feschotte in Gilbert, 2012), 11 – (Fort in sod., 2012), 12 – (Holmes, 2011), črna pika – prisotnost, bela pika – odsotnost.

VIRUSNA SKUPINA	Protista	Arthropoda	Nematoda	Platyhelminthes	Crustacea	Referenca
ASTRO	○	○	○	○	○	1
BIRNA	○	○	○	○	○	1
HEPE-VIRGA	○	●	●	○	○	1, 9
HYPO	○	○	○	○	○	1
LUTEO-SOBEMO	●	●	●	○	○	1
NARNA-LEVI	●	●	○	○	○	1
BUNYA-ARENA	○	●	●	●	●	1, 2, 4, 5, 10, 12
MONO-CHU	○	●	●	●	●	1, 2, 4, 5, 8, 10, 11, 12
ORTHOMYXO	○	●	○	○	○	1, 2, 4, 10, 12
NIDO	○	●	○	○	○	1
PARTITI-PICOBIRNA	○	●	●	○	○	1, 10
PERMUTOTETRA	○	○	○	○	○	1
PICORNA-CALICI	○	●	○	○	○	1, 10
REO	○	●	○	○	○	1, 4, 8, 10, 12
TOMBUS-NODA	○	●	●	○	○	1, 3
TOTI-CHRYSO	●	●	●	○	●	1, 5, 10
QINVIRUS	●	●	●	○	○	1
YUEVIRUS	○	○	○	○	○	1
YANVIRUS	○	○	○	○	○	1
ZHAOVIRUS	○	○	○	○	○	1
WEIVIRUS	●	○	○	○	○	1
POTYVIRIDAE	○	○	○	○	○	
FLAVIVIRIDAE	○	●	○	○	○	4, 6, 7, 10, 12
OPHIOVIRIDAE	○	○	○	○	○	
CYSTOVIRIDAE	○	○	○	○	○	

Nekateri RNA virusi imajo mehanizem podvojevanja vezan na jedro gostiteljske celice. Pri omenjenih virusih je verjetnost za integracijo v genom večja. Skupina Orthomyxoviridae po prehodu v celico s klatrinom posredovano endocitozo, disociacijo kapside in virusne ovojnice v endosomih sprosti genom v kompleksu z nukleoproteinom, PA, PB1, in PB2 podenoto RdRp. Kompleks nato potuje do jedra, ki ga aktivno prečka preko importinov. Jedrni lokalizacijski signali so bili najdeni na vseh treh podenotah RdRp in nukleoproteinu. Mehanizem je torej podoben kot pri retrovirusih (Cohen in sod., 2011). Večina RNA virusov nima mehanizma podvojevanja, povezanega z jedrom. Vseeno je bilo najdenih kar nekaj virusov endogeniziranih, ki nimajo mehanizma podvojevanja povezanega z jedrom, kot je skupina Partitiviridae (Shi in sod., 2016). Mehanizem, s katerim RNA virusi, še posebej tisti, ki nimajo mehanizma podvojevanja povezanega z jedrom, preidejo skozi jedro in se vgradijo v gostiteljski genom, še ni bil pojasnjen. Virusi lahko okužujejo samo deleže se celice, pri

katerih je konstrukcija jedra nestabilna. Druga možnost je, da katera koli proteinska enota, ki je vezana z virusnim genomom, vsebuje jedrne lokalizacijske signale, ki omogočajo prehod virusa skozi jedrne pore. Kapside nekaterih virusov so dovolj majhne, da preidejo skozi jedrne pore. Nekateri virusi izkoriščajo alternativne mehanizme, pri katerih začasno destabilizirajo jedrne pore in tako preidejo v jedro (Cohen in sod., 2011).

1.3 Nevretenčarji

Vprašanje, ki je ključnega pomena za filogenomiko virusov, je, kakšna je razširjenost njihove raznolikosti med organizmi kot ekološkimi nišami in kateri organizmi predstavljajo center njihove raznolikosti. Kot že omenjeno, je hitrost mutacij virusov odvisna od natančnosti polimeraze (angl. »fidelity«), mutagenih dejavnikov in mehanizma za popravljanje mutacij. Hkrati je močno odvisna od dinamičnosti ekološke niše, v katerih se organizem nahaja. Dinamičnost ekološke niše oblikuje raven selekcijskega pritiska v določenem časovnem intervalu. Raven selekcijskega pritiska pa določa tendenco genoma k stalnemu odzivanju nanj. Večji kot je selekcijski pritisk, večja bo tendenca genoma k iskanju optimalnih rešitev za njegovo minimizacijo. Dinamičnost okolja torej opredeljuje in navzdol omejuje hitrost mutacij. Večja kot je dinamičnost okolja, večji je selekcijski pritisk in večja mora biti tista minimalna hitrost mutacij, s katerimi lahko organizmi (virusi) kompenzirajo oz. minimizirajo selekcijski pritisk. Hkrati se z večanjem hitrosti mutacij večja verjetnost škodljivih mutacij in s tem tudi letalnih mutacij. Zato je hitrost mutacij omejena z maksimalno hitrostjo mutacij, ki še ne bo ključno vplivala na preživetje celotne populacije. Hitrost mutacij in hitrost substitucij pa sta ključna dejavnika raznolikosti organizmov. Zato bomo največjo raznolikost virusov iskali pri najbolj dinamičnih gostiteljih. Členonožci (Arthropoda), še posebej skupina Insecta, veljajo za eno najbolj prilagodljivih, robustnih in raznolikih organizmov na Zemlji z največjim številom opisanih vrst. Hitro nastajanje populacij z razvitimi optimalnimi rešitvami za zmanjšanje selekcijskega pritiska jih uvršča med ene najbolj dinamičnih gostiteljev za RNA viruse. Nekateri izmed redov insektov (Diptera, Hemiptera, Polyneoptera, Odonata in Thysanoptera) so znani vektorji rastlinskih in živalskih virusov in zato predmet številnih raziskav v povezavi z raznolikostjo RNA virusov. Členonožci, specifično insekti, so pridobili oznako centra raznolikosti RNA virusov (Dudas in Obbard 2015; Engel, 2015). Shi in sod. (2016a) so pokazali raznolikost RNA virusov tudi pri glistah (Nematoda) in skupini Spiralia (Shi in sod., 2016a). S tem so nakazali, da je lahko raznolikost RNA virusov razširjena čez celotno skupino nevretenčarjev. Virusi so z gostitelji povezani v mrežo gostiteljsko-virusnih koevolucijskih vzorcev. Zato moramo za poznavanje raznolikosti RNA virusov in njihovih medsebojnih filogenetskih odnosov poznati filogenijo nevretenčarjev.

1.3.1 Skupine nevretenčarjev

Evkarionti predstavljajo raznolik spekter enoceličnih in mnogoceličnih organizmov, ki jih gradijo evkariontske celice. Zasedli so praktično vse ekološke niše na Zemlji. Filogenetsko drevo evkariontov je prešlo skozi številne preureditve (Burki, 2014; Burki in sod., 2016; Dacks in sod., 2016; Koonin in sod., 2015). Burki in sod. (2014, 2016) so poudarili glavne skupine evkariontov: Opisthokonta, Amoebozoa, Excavata, Archaeplastida s skupino SAR (imenovana po treh glavnih (pod)skupinah, ki jo sestavljajo). Skupina Opisthokonta zajema živali (Metazoa), glive (Fungi) in nekatere heterotrofne protiste. Amoebozoa vključuje v glavnem ameboidne heterotrofne protiste (kot je npr. ameba). Amoebozoa in Opisthokonta so združeni v enotno skupino »unikonti« (angl. »unikonts«). Skupina Excavata vsebuje heterotrofne protiste, ki vsebujejo hidrogenosome ali mitosome, namesto mitohondrijev (Giardia, Trichomonas), ali plastide (evglenidi – evglena). Archaeplastida je raznolika nadskupina, ki vsebuje fotosintetske organizme, uvrščene v tri glavne skupine: glavkofiti (mikroskopske alge s cianobakterijam podobnimi plastidi), rodofiti (rdeče alge) in *Viridiplantae* (zelene alge in kopenske rastline). Skupina SAR zajema tri glavne skupine: stramenopile (angl. »stramenopiles«), alveolate (angl. »alveolates«, mednje sodijo tudi dinoflagelati) in Rhizaria (Burki, 2014, Burki in sod., 2016). Katera skupina evkariontov predstavlja ali je sorodna z zadnjim skupnim prednikom evkariontov, je bilo eno izmed glavnih vprašanj številnih študij. Skupina Amoebozoa je bila predlagana kot najbližji sorodnik najmlajši skupini Opisthokonta in skupina Excavata kot najstarejša skupina in najbližji sorodnik zadnjemu skupnemu predniku evkariontov (He in sod., 2014).

Skupina Metazoa (živali) znotraj skupine Opisthokonta zajema pet glavnih skupin: Porifera (spužve), Placozoa (rod *Trichoplax*), Ctenophora (rebrače), Cnidaria (ožigalkarji), ki so starejše skupine živali, in Bilateria (dvobočno somerne oz. bilateralno – simetrične živali) (Dohrmann in Wörheide, 2017). Evolucijska starost posamezne skupine je bila ključno vprašanje številnih študij (Dohrmann in Wörheide, 2017; Dunn in sod., 2014). Tradicionalno je veljalo prepričanje, da so Porifera sestrsko klada ostalim živalim, znanim pod skupnim imenom Epitheliozoa. Vendar je bila podpora razvejišč slaba predvsem zaradi pomanjkanja genomskih podatkov najbližjih sorodnikov živali (sestrske klade Metazoa) iz širše skupine Opisthokonta (Philippe in sod., 2009). Najbližji sorodniki oz. skupni predniki, ki oblikujejo sestrsko klado s skupino živali (Metazoa), naj bi bili skupina Choanoflagellida oz. Choanoflagellata, Ichthyosporea in Filasterea. S pridobivanjem novih zaporedij najsorodnejših organizmov skupine živali in njihovih najstarejših predstavnikov, posodabljanjem prvotne vzorčne populacije v genomskih bazah podatkov za filogenetske analize in uporabo različnih analitskih modelov se je filogenija živali stalno dopolnjevala (Philippe in sod., 2009; Dunn in sod., 2014; Dohrmann in Wörheide, 2017; Sebé-Pedrós in sod., 2017). Skupine Choanoflagellida oz. Choanoflagellata, Ichthyosporea in Filasterea so skupaj z glivami (Fungi) v številnih raziskavah uporabili kot »outgroup«. Philippe in sod. (2009) so uporabili metodo CAT (Lartillot in Philippe., 2004) in v nekoreninjenem drevesu predstavili spužve kot najstarejšo skupino, ki ji po evolucijski starosti sledi Placozoa, skupini Ctenophora in Cnidaria pa uvrstili kot sestrsko skupino Bilateria. Kasneje se je zgodil preobrat. Ryan in sod. (2013) in Moroz in sod. (2014) so sekvencirali genom rebač (Ctenophora) *Mnemiopsis leidyi* in *Pleurobrachia bachei*, kar je omogočilo lažjo filogenetsko karakterizacijo skupine. V filogenetski analizi sta skupini uvrstili skupino Ctenophora kot najstarejše predstavnike živali, ki so ji sledili spužve in ožigalkarji (Ryan in sod., 2013; Moroz in sod., 2014). Pri razumevanju filogenije živali sta se oblikovali nasprotni predpostavki. Prva je predvidevala spužve kot najstarejšo skupino živali (Philippe in sod., 2009; Pisani in sod., 2015; Dohrmann in Wörheide, 2017; Simion in sod., 2017), druga je predvidevala rebrače kot najstarejšo skupino živali (Ryan in sod., 2013; Moroz in sod., 2014; Sasson in Ryan, 2017).

Naslednje vprašanje, ki je ostalo odprto znotraj živali (Metazoa), je bil evolucijski odnos skupine Xenacoelomorpha s skupino Bilateria. Nekatere skupine so jo uvrstile kot sestrsko klado skupini Bilateria, drugi raziskovalci so jih uvrstili znotraj skupine Deuterostomia (devterostomi) (Dunn in sod., 2014). Skupina Bilateria zajema dve glavni nadskupini: Deuterostomia (devterostomi) in Protostomia. Skupina Protostomia je raznolika in zajema dve glavni skupini: Ecdysozoa in širšo skupino **Spiralia**, poimenovano po spiralni simetriji delitve zigote po oploditvi. Nadaljnja filogenomska klasifikacija glavnih skupin Spiralia še vedno ni dosegla poenotenja. Literatura kot glavne skupine Spiralia navaja: Cycliophora, Dicyemida, Gnathifera (Syndremata, Gnathostomulida, Micrognathozoa, Rotifera), Gastrotricha, Platyhelminthes, Mollusca, Annelida, Nemertea, Phoronida, Bryozoa, Entoprocta, Brachiopoda in druge (Nesnidal in sod., 2013; Giribet, 2016; Lu in sod., 2017). Halanych in sod. (1995) so predlagali konsenzusno drevo, znotraj katerega so skupini Lophophorata (Bryozoa, Phoronida, Brachiopoda) in Trochozoa (Annelida, Mollusca) uvrstili v nadskupino Lophotrochozoa in definirali skupino Lophotrochozoa v ožjem smislu (sensu stricto). Skupina Lophophorata je bila poimenovana po t. i. lofofori (angl. »lophophore«), v obroč sklenjenih lovkah z migetalkami, ki predstavnikom skupine služi za namen prehranjevanja (Halanych in sod., 1995). Skupina Trochozoa je bila poimenovana po t. i. trohoforah, migetalkastih ličinkah, ki predstavljajo zgodnjo fazo njihovega razvoja (Dunn in sod., 2014). Skupina Lophophorata in skupina Brachiopoda znotraj skupine Lophophorata je bila po mnenju Halanych in sod. (1995) polifiletskega izvora. Kasnejše raziskave so zagovarjale monofilijo skupine Lophophorata (Nesnidal in sod., 2013; Lu in sod., 2017), uvrstitev ostalih skupin – Trochozoa, Cycliophora, Entoprocta, Annelida, Mollusca, Nemertea, Dicyemida in Platyhelminthes – v kontekstu skupin Deuterostomia in Ecdysozoa pa je ostala nekonsistentna in močno odvisna od uporabe nabora reprezentativnih organizmov, modela in metode za konstrukcijo filogenetskega drevesa (Nesnidal in sod., 2013; Laumer in sod., 2015; Lu in sod., 2017). Med njimi je skupina ploščatih črvov (Platyhelminthes) izjemno raznolika skupina tako parazitskih kot neparazitskih organizmov z bilateralno simetričnim telesom brez kutikole (Littlewood, 2006).

Skupina **Ecdysozoa** (kot druga večja klada znotraj skupine Protostomia) je bila prvič kot monofiletska klada definirana že v študiji Aguinaldo in sod. (1997). Zajema štiri glavne

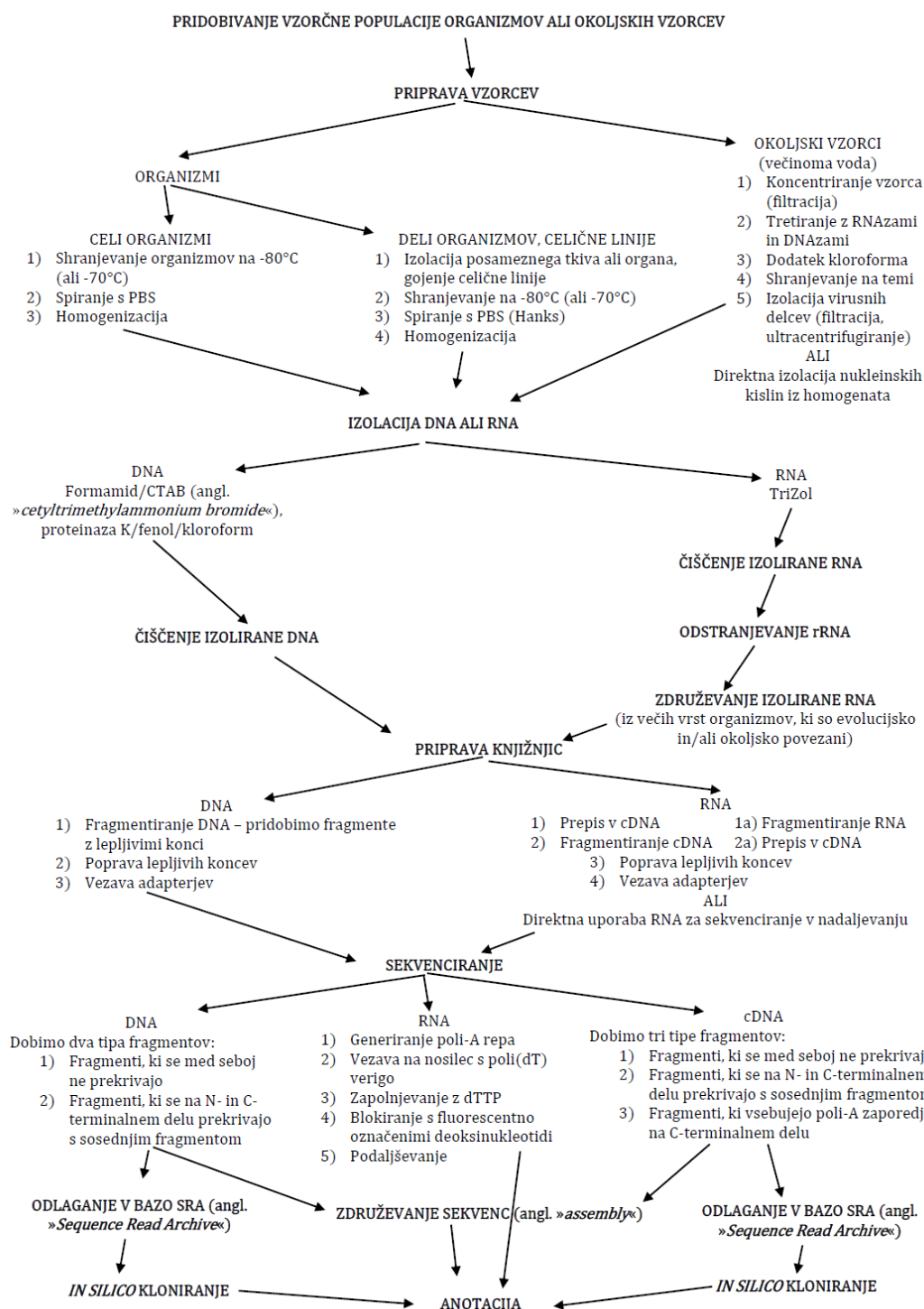
skupine: Nematoda (nematodi), Nematomorpha, Panarthropoda in Scalidophora. Skupini Nematoda in Nematomorpha so raziskave uvrstile v nadskupino Nematoida, skupino Nematoida pa s sestrsko klado Scalidophora v nadskupino Cycloneuralia (Dunn in sod., 2014). Problem pri uvrščanju glist (Nematoda) v kontekst organizmov Ecdysozoa je hitro spreminjanje njihovega genoma, ki v filogenetskih drevesih povzroči »artefakt dolgih vej« (angl. »long branch attraction«). Poveže jih s skupino Panarthropoda oz. s členonožci (Arthropoda) (Budd in Telford, 2009). Skupina Panarthropoda zajema skupine Arthropoda (členonožci), Onychophora (krempeljčarji) in Tardigrada (počasniki). Skupina Arthropoda (členonožci) zajema dve glavni skupini: Chelicerata (pajkovci) in Mandibulata (čeljustniki). Skupina Mandibulata (čeljustniki) je širša skupina in vključuje Myriapoda (stonoge) in Pancrustacea. V slednjo sodijo raki (Crustacea) in žuželke (Hexapoda). Členonožci torej zajemajo štiri glavne skupine: raki (Crustacea), pajkovci (Chelicerata), stonoge (Myriapoda) in žuželke (Hexapoda). Evolucijski odnos med štirimi skupinami, njihova monofilija, parafilija oz. polifilija in evolucijski odnos členonožcev z ostalimi skupinami so bili dolgo predmet razprave številnih filogenetskih analiz. Prvotno so skupino stonog uvrstili kot sestrsko klado žuželkam. Skupaj naj bi tvorili skupino Tracheata. Skupina Tracheata naj bi bila sestrsko skupina rakom (Crustacea) – skupaj naj bi tvorili skupino Mandibulata. Pajkovci (Chelicerata) naj bi tvorili sestrsko klado s skupino Mandibulata. Kasnejše analize so spremenile položaj žuželk v skupini členonožcev. Z raki naj bi tvorile sestrsko klado in oblikovale skupino Tetraconata (Pancrustacea) – sestrsko klado pajkovcem in stonogam. Žuželke naj bi predstavljale »kopensko vejo« skupine Pancrustacea (Budd in Telford, 2009). Skupina Crustacea zajema tri glavne skupine: Maxillopoda (Copepoda – ceponožci, Branchiura – škrgorepci in ostale skupine), Malacostraca (Amphipoda – postrance, Decapoda – deseteronožci, Isopoda – enakonožci in ostale skupine) in Branchiopoda (von Reumont in sod., 2012). Ceponožci naj bi predstavljali starejšo skupino rakov. Gre za eno najbolj raznolikih skupin rakov s 120000 opisanimi vrstami, ki dosežejo dolžino med 0,1 mm in 23 cm. Predstavljajo glavni delež zooplanktona in so vključeni v mrežo prehranjevalne verige na višjih in nižjih trofičnih nivojih (Bron in sod., 2011). Skupina Hexapoda zajema glavne skupine Collembola (skakači), Diplura (dvorepke), Protura (proturi) in Insecta (prave žuželke). Collembola (skakači) in Diplura (dvorepke) sta najstarejši skupini znotraj insektov. Skakači naseljujejo kopensko mezofavno. Gre za 0,2–3 mm dolge organizme, ki živijo v matriksu ali na površini prsti in so pomembni dekompozitorji (Czarnetzki in Tebbe, 2004). Skupina Insecta zajema starejšo linijo Monocondylia in Dicondylia. V skupino Dicondylia sodi večina žuželk, najstarejša podskupina pa je Zygentoma (ščetinorepke). Skupina Pterygota (krilate žuželke) znotraj skupine Dicondylia zajema dve glavni skupini žuželk: Neoptera in Palaeoptera. Skupina Neoptera vključuje tri glavne skupine: Holometabola (najširša skupina), Paraneoptera in Polyneoptera. Holometabola zajema glavne skupine: Coleoptera (hrošči), Diptera (dvokrilci), Hymenoptera (kožokrilci) in Lepidoptera (metulji). Skupina Polyneoptera zajema glavne redove: Dermaptera, Blattodea in Orthoptera. Skupina Paraneoptera zajema glavno skupino Hemiptera. Omenjene skupine pokrivajo več kot polovico vseh opisanih vrst organizmov (Mutanen in sod., 2010). Lepidoptera ima trenutno definiranih okrog 45 družin (Ditrysia, Yponomeutoidea, Apoditrysia, Macrolepidoptera, Obtectomera in druge), znotraj katerih se uvršča 160.000 opisanih vrst, vendar število različnih vrst ocenjujejo celo na 500.000 (Mutanen in sod., 2010; Kawahara in Breinholt, 2014). Študije na skupini Lepidoptera so bile usmerjene na posamezne družine s slabo opredeljeno filogenetsko klasifikacijo. Potekale so na bazi izolacije tako DNA kot RNA (Mutanen in sod., 2010; Kawahara in Breinholt, 2014). Skupina Hymenoptera zajema 153.000 opisanih vrst in okoli milijon izumrlih oz. neopisanih vrst (Peters in sod., 2017). Glavne skupine med njimi so Apocrita, Cephoidea, Orussoidea, Pamphilioidea, Siricoidea, Tenthredinoidea, Xiphydrioidea in Xyeloidea. Med njimi je Apocrita najmlajša in najširša skupina. Znotraj skupine Apocrita sta znani dve podskupini: Aculeata in Terebrantes. Terebrantes je evolucijsko starejša skupina kot Aculeata. Aculeata zajema kar 70.000 od 153.000 opisanih vrst Hymenoptera. Zajema skupine Apoidea (čebele, ose), Chrysidoidea, Pompiloidea, Vespoidea, Sphecoidea, Thynnoidea, Tiphiioidea, Scolioidea in Formicoidea (mravlje) kot najpomembnejše skupine. Številne raziskave niso dosegle poenotenja glede definiranja najstarejše skupine Aculeata, se pa v več raziskavah pojavlja Chrysidoidea kot najstarejša skupina (Branstetter in sod., 2017; Peters in sod., 2017).

1.4 Pregled raznolikosti in razširjenosti RNA virusov pri nevretenčarjih

1.4.1 Pregled

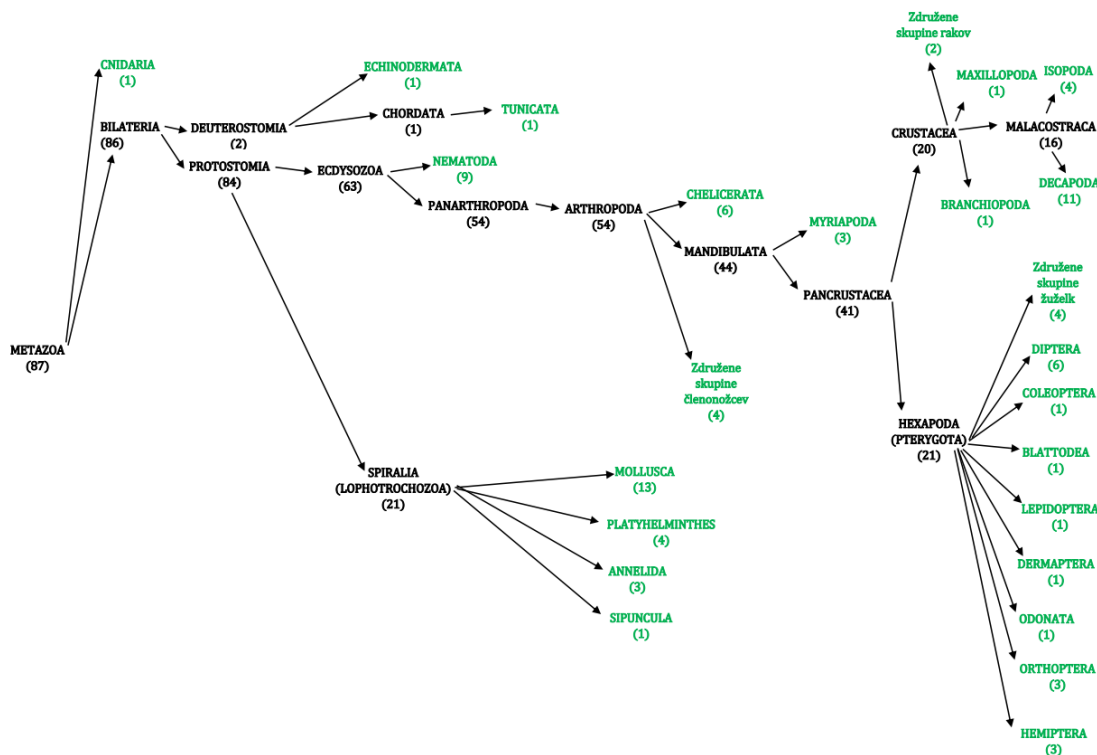
Znanje o boleznih sega veliko dlje od znanja o virusih. Znani so hieroglifi iz leta 3700 pr. n. št. o egipčanskem duhovniku, ki kaže tipične znake poliomielitisa. Na Kitajskem je okrog 1000 pr. n. št. izbruhnila epidemija črnih koz (angl. »*smallpox*«), kar je privedlo do večkrat neuspešnega prakticiranja različice cepljenja (variolacije; angl. »*variolation*«). Skorajšnja smrt Edwarda Jennerja zaradi variolacije ga je privedla do razvoja prvih cepiv proti črnim kozam. Narava infekcijskih agensov ni bila znana, dokler ni prišlo do izuma mikroskopa in tehnik mikroskopiranja, ki jih je razvil Antony van Leeuwenhoek. Izum je bil ključna podlaga za postavitev teoretičnih temeljev mehanizma prenosa bolezni oz. bakterij, takrat poimenovanih »bacili« (angl. »*germs*«) – Robert Koch in Kochovi postulati, kot tudi teoretičnih temeljev preprečevanja napredovanja bolezni – Louis Pasteur. Še vedno je veljalo prepričanje o enotni naravi infekcijskih agensov, dokler Dimitrij Ivanovski ni odkril, da po prehodu ekstrakta obolele rastline tobaka preko keramičnega filtra in nadaljnji inokulaciji zdrave rastline tobaka, se na rastlini še vedno pojavijo isti bolezenski simptomi. Odkril je, da gre za drugo obliko infekcijskih agensov, ki so manjši od bakterij (ki jih keramični filter zadrži). Do karakterizacije in poimenovanja infekcijskih agensov je prišlo šele leta 1915, ko sta Frederick Twort in Felix d'Herelle poimenovala bakteriofage kot viruse, ki napadajo bakterije. S tem sta postavila temelje virologiji. Bakteriofagi so postali modelni sistemi za študije morfologije in mehanizmov podvojevanja. Sledil je razvoj orodij za odkritje novih virusov in študijo njihove morfologije. Komarji so bili prvi prepoznani prenašalci človeških virusov. Viruse so izolirali iz insektov, vretenčarjev, rastlin in bakterij, ki so kazali znake okužbe. Razvijale so se celične kulture za gojenje virusov in sistemi (gojišča) za gojenje celičnih kultur. Morfologija in mehanizem replikacije izoliranega virusa so se raziskovali na podlagi seroloških, imunoloških metod in študij ultrastruktur (elektronska mikroskopija) (Cann, 2012). Odkrivali so se samo posamezni virusi in virologija je napredovala počasi, dokler ni prišlo do razvoja molekularne biologije, ki ga je zaznamovalo nekaj ključnih odkritij: 1) odkritje strukture DNA, 2) izum sekvenciranja z inhibitorji, ki prekinjajo verigo (angl. »*chain-terminating inhibitors*«) (Sanger in sod., 1977) in 3) izum verižne reakcije s polimerazo (PCR) (Mullis in Faloona, 1987). Izumi in odkritja so odprli vrata v klasifikacijo in karakterizacijo virusov na podlagi izolacije DNA, RNA, njunega sekvenciranja in primerjave zaporedij. Odkrivanje raznolikosti virusov in njihova karakterizacija je bilo narejena v bistveno krajšem obdobju. Vzporedno je leta 1971 ICTV (angl. »*International Committee on Taxonomy of Viruses*«) izdala svoje prvo poročilo. Odkritje novih metod sekvenciranja (tretja generacija sekvenciranja) je povzročilo premik iz virusne genetike na virusno genomiko, transkriptomiko, metagenomiko in metatranskriptomiko. Odkrivanje in karakterizacija posameznega virusa se je premaknila v odkritje in karakterizacijo populacije virusov v gostiteljskih ali okoljskih ekoloških nišah. Izraz »virus« je nadomestil izraz »virom« (Paez-Espino in sod., 2016). Tekom raziskav so vodna okolja (Angly in sod., 2006; Desnues in sod., 2008; Mizuno in sod., 2013; Filosof in sod., 2017) bistveno prevladovala nad kopenskimi okolji (Cantalupo in sod., 2011). V vodnih okoljih so bistveno prednjačile raziskave na DNA virusih, predvsem dsDNA bakteriofagih (Angly in sod., 2006; Desnues in sod., 2008; Mizuno in sod., 2013; Filosof in sod., 2017). RNA virusi so bili tako v vodnih okoljih kot na kopenskih okoljih slabše raziskani kot bakteriofagi kot modelni sistemi. Raziskave RNA virusov v kopenskih okoljih so bile nekoliko pogostejše kot raziskave RNA virusov v vodnih okoljih (Culley in sod., 2006; Rosani in Gerdol, 2017; Galinier in sod., 2017). Pri raziskavah RNA virusov v kopenskih okoljih so bile pogostejše raziskave RNA virusov v povezavi z njihovo patogenostjo pri vretenčarjih in rastlinah (Halstead, 1966; Achong, 1972; Bauer in Murphy, 1975; Virgin, 2014). Redkejšje so bile raziskave same raznolikosti in razširjenosti RNA virusov. Večina raziskav, katerih glavna tematika je bila raznolikost in razširjenost RNA virusov, je bila narejenih pri členonožcih, predvsem žuželkah in pajkovcih (Hogenhout in sod., 2008; Junglen in Drosten, 2013; Carapeta in sod., 2015; Ladner in sod., 2016; Kawakami in sod., 2016; Nouri in sod., 2015). Uporabljali so se okoljski vzorci (predvsem voda) ali vodni in kopenski organizmi. Kot vzorec se je uporabljal celoten organizem, organ, tkivo ali celična linija (Halstead, 1966; Achong, 1972; Bauer in Murphy, 1975; Angly in sod., 2006; Culley in sod., 2006; Desnues in sod., 2008; Hogenhout in sod., 2008; Cantalupo in sod., 2011; Junglen in Drosten, 2013; Mizuno in sod., 2013; Virgin, 2014; Carapeta in sod., 2015; Paez-Espino in

sod., 2016; Nouri in sod., 2015; Kawakami in sod., 2016; Ladner in sod., 2016; Rosani in Gerdol, 2017; Galinier in sod., 2017; Filosof in sod., 2017). Izolirali so bodisi celotno RNA ali DNA organizma, organa, tkiva, celične linije bodisi so RNA oz. DNA pridobili iz okoljskega vzorca. Slika 1.56 prikazuje povzetek protokolov, uporabljenih v okviru raziskav raznolikosti in razširjenosti RNA virusov (**Slika 1.56**).



Slika 1.56: Povzetek metod, uporabljenih pri identifikaciji, karakterizaciji novih RNA virusov pri raziskovanju RNA virusne diverzitete (Halstead, 1966; Achong, 1972; Bauer in Murphy, 1975; Angly in sod., 2006; Culley in sod., 2006; Desnues in sod., 2008; Hogenhout in sod., 2008; Cantalupo in sod., 2011; Junglen in Drosten, 2013; Mizuno in sod., 2013; Virgin, 2014; Carapeta in sod., 2015; Paez-Espino in sod., 2016; Nouri in sod., 2015; Kawakami in sod., 2016; Ladner in sod., 2016; Rosani in Gerdol, 2017; Galinier in sod., 2017; Philosof in sod., 2017).

Raziskave so potekale v dveh smereh. Prva smer je bila izolacija DNA ali RNA iz okoljskega vzorca, organizma, organa, tkiva ali celične linije, njeno fragmentiranje, oblikovanje knjižnic, sekvenciranje in združevanje pridobljenih kratkih zaporedij (angl. »sequence reads«). Združevanje je potekalo s pomočjo več alternativnih orodij. Za združevanje cDNA zaporedij (transkriptoma) se je večinoma uporabljalo programe Trinity, Oases, Velvet, ABySS, SeqMan (Schulz in sod., 2012; Shi in sod., 2016a). Za združevanje DNA zaporedij (genoma) organizma (ali več organizmov v primeru, da so se izolati nukleinskih kislin združevali na podlagi ekoloških ali evolucijskih povezav) pa se je uporabljalo orodje Atlas (Havlak in sod., 2004). Po združevanju zaporedij v soseske (angl. »contig«) in podaljševanju sosesk v ogrodja (angl. »scaffold«) je sledilo potrjevanje združevanja, tako da so se združena ogrodja mapirala nazaj na genom organizma, iz katerega se je genom ali transkriptom izoliral. Virusni RNA genom (ali endogeniziran genom oz. fragment genoma) se je iz celotnega transkripta izločil z iskanjem homologov proti vsem referenčnim virusnim proteinom. Po anotaciji transkriptoma je sledilo odlaganje združenega genoma ali transkriptoma v transkriptomske ali genomske baze podatkov. Druga možnost za raziskovalce je bila, da so že kratka zaporedja, pridobljena direktno po sekvenciranju, odložili v SRA (angl. »sequence read archive«) bazo podatkov. Velikost baz podatkov je z začetkom naprednejših metod sekvenciranja začela eksponentno naraščati. Zato je alternativna pot do odkritja novih RNA virusov vključevala *in silico* kloniranje in pridobivanje novih virusnih sekvenc na podlagi združevanja kratkih zaporedij iz SRA baze podatkov (Liu in sod., 2012). Ker je bilo več raziskav glede raznolikosti RNA virusov narejenih na kopenskih organizmih, členonožcih, kot znanih vektorjih rastlinskih in živalskih virusov, so postali center raznolikosti RNA virusov. Obveljalo je zavajajoče prepričanje, da je večina raznolikosti RNA virusov skoncentrirana v členonožcih, dokler nekaj raziskovalnih del ni spremenilo prepričanja in odprlo nov vpogled v razširjenost in raznolikost RNA virusov (Webster in sod., 2015; Li in sod., 2015; Webster in sod., 2016; Shi in sod., 2016a; Shi in sod., 2016b; Waldron in sod., 2017). Webster in sod., (2015, 2016) so v dveh raziskavah odkrili več kot 55 novih RNA virusov pri skupini Drosophilidae (Diptera), ki so sodili v skupine Picornavirales, Nodaviridae, Virgaviridae, Negevirus, Permutotetraviridae, Flaviviridae in Bunyavirales (Webster in sod., 2015; Webster in sod., 2016). Li in sod. (2015) so raziskali ss(-)RNA viruse pri žuželkah in raki in odkrili več kot 100 novih virusov. Definirali so novo skupino virusov, ki sodi k skupini Mononegavirales – Chuvirus (Li in sod., 2015). Shi in sodelavci so naredili najobsežnejšo analizo raznolikosti RNA virusov pri členonožcih (raki, pajkovcih, stonogah in žuželkah), glistah in skupini Lophotrochozoa (predvsem mehkužcih). Izolirali so RNA iz 220 nevretenčarskih vrst, odstranili rRNA (ribosomalno RNA), združili RNA evolucijsko povezanih organizmov, jo fragmentirali in oblikovali 87 knjižnic, ki so vsebovale RNA enega ali več organizmov. Med njimi je bila samo ena knjižnica s sekvencami najstarejših skupin Metazoa (živali) – Cnidaria, ostale so pripadale skupini Bilateria (dvobočno somerne živali). Znotraj skupine Bilateria sta bili samo dve knjižnici, ki sta pripadali skupini Deuterostomia – Echinodermata in Tunicata (Chordata), preostalih 84 pa je pripadalo skupini Protostomia. Znotraj skupine Protostomia je 21 knjižnic pripadalo skupini Spiralia (Lophotrochozoa), preostalih 63 pa skupini Ecdysozoa. V skupini Ecdysozoa je bilo 9 knjižnic od skupine Nematoda, preostalih 54 pa je bilo knjižnic skupine Panarthropoda oz. podskupine Arthropoda (členonožci). Skupina Arthropoda je zajemala 6 knjižnic pajkovcev (Chelicerata), 44 knjižnic je bilo od skupine Mandibulata, preostale štiri pa so bile kompleksne knjižnice, ki so zajemale različne predstavnike členonožcev. Znotraj skupine Mandibulata so tri knjižnice pripadale stonogam (Myriapoda), od preostalih 41 pa jih je 20 pripadalo skupinam rakov (Crustacea) in 21 žuželkam (Hexapoda). Znotraj skupine rakov sta bili dve knjižnici z mešanimi predstavniki skupine, preostale pa so v glavnem zajemale red Decapoda. Pri žuželkah (Hexapoda) je bila večina knjižnic namenjenih redu Diptera. Knjižnice so večinoma zajemale več različnih vrst organizmov (Slika 1.57).



Slika 1.57: Število metatranskriptomskih knjižnic glede na taksonomske skupine v referenci Shi in sod. (2016a).

S sekvenciranjem knjižnic so pridobili zaporedja dolžine 90–100 bp, ki so z 20-kratno pokritostjo predstavljala skupaj 6×10^{12} nukleotidov. Z združevanjem zaporedij so odkrili 1445 novih virusov. Viruse so razdelili v klade: Astro, Birna, Hepe-Virga, Hypo, Luteo-Sobemo, Narna-Levi, Bunya-Arena, Mono-Chu, Orthomyxo, Nido, Partiti-Picobirna, Permutotetra, Picorna-Calici, Reo, Tombus-Noda, Toti-Chryso. Definirali so pet novih klad: Qinivirus, Yuevirus, Yanvirus, Weivirus in Zhaovirus. Definirali so tudi nove skupine virusov, ki pripadajo določenim kladam, kot so Kelp-fly virusu podobni virusi in Nora virusu podobni virusi znotraj klade Picorna-Calici (Shi in sod., 2016a; Shi in sod., 2018a). Vseeno v omenjeno raziskavo skupina ni vključila skupine Flaviviridae. Analizo raznolikosti skupine Flaviviridae so objavili v ločeni raziskavi (Shi in sod., 2016b). Skupino Flaviviridae so postavili v ločeno kladu in definirali dve novi skupini Flavivirusov: Jingmenviruse in flavivirusom podobne viruse (angl. »*Flavi-like*«) (Shi in sod., 2016b). Raznolikost RNA virusov je dodatno dopolnila tudi skupina Waldron in sod. (2017) z raziskavo pri skupini Porifera (Waldron in sod., 2017). Omenjene raziskave so razširile sliko o raznolikosti in razširjenosti RNA virusov. Izvor in največjo raznolikost RNA virusov so premaknili s členonožcev na celotno skupino Protostomia. Dokazali so: 1) da je tekom evolucije prišlo do eksplozije raznolikosti RNA virusnih genomov, ki je bistveno večja, kot je bila prvotno prikazana ob raziskavah viroma členonožcev, 2) da se je raznolikost RNA viromov ohranila čez celotno skupino Protostomia (Webster in sod., 2015; Li in sod., 2015; Webster in sod., 2016; Shi in sod., 2016a; Shi in sod., 2016b; Waldron in sod., 2017).

1.4.2 Preostala odprta vprašanja

Shi in sod. (2016) so opravili doslej najboljše raziskavo raznolikosti RNA virusov, ki pokriva vse predstavnike živali od najstarejših skupin (Cnidaria), rakov, glist, stonog, žuželk in mehkužcev. Vseeno pa za potrditev domneve še vedno manjka analiza RNA raznolikosti pri določenih skupinah nevretenčarjev. Analizirali so samo eno vrsto ožigalkarja (Cnidaria). Drugih najstarejših skupin živali (Ctenophora, Porifera) niso analizirali. Ravno tako niso analizirali raznolikosti RNA virusov pri najstarejših skupinah žuželk (Collembola, Diplura, Zygentoma, Monocondylia). Manjkajo tudi predstavniki rakov, kot so Amphipoda in Copepoda, poleg tega, da je raznolikost pri skupini Isopoda delno pokrita. Hkrati je bila

analizirana samo ena knjižnica Lepidoptera (metulji). V raziskavo niso bili vključeni kožokrilci (Hymenoptera). Znotraj skupine Lophotrochozoa (Spiralia) so bili v glavnem zajeti mehkužci. Ostale skupine organizmov Lophotrochozoa (Spiralia), kot so ploščati črvi (Platyhelminthes), so bile slabo zastopane. Vseeno so obsežne raziskave postavile tri ključne mejnike pri raziskavi raznolikosti in razširjenosti RNA virusov pri nevretenčarjih:

- 1) Predstavili so »eksplozijo« raznolikih virusnih klad, ki se je odvila tekom evolucije skupine Protostomia, ki je bila bistveno večja kot prvotno nakazana v okviru raziskav raznolikosti in razširjenosti RNA virusov pri členonožcih.
- 2) Dokazali so, da se je raznolikost RNA viroma ohranila čez celotno skupino Protostomia.
- 3) Dokazali so, da je skupina Protostomia izvir pestrega nabora raznolikih RNA viromov.

Hkrati so z raziskavami odprli tri ključna vprašanja:

- 1) Ali je skupina Protostomia dejanski izvor »eksplozije« pestrega RNA viroma ali se je dejanska eksplozija odvila že pri predniku živali?
- 2) Kakšna je dejanska starost posameznih virusnih klad – ali najstarejše predstavnike posameznih virusnih klad najdemo pri skupini Protostomia ali pri predniku živali?
- 3) Ali se pestrost RNA viroma ohranja čez celotno skupino nevretenčarjev ali je dejansko omejena samo na skupino Protostomia?

Poglavje 2

Namen dela

2.1 Namen

Prvotne analize RNA virusov pri členonožcih so ustvarile prepričanje, da ti organizmi predstavljajo glavni vir raznolikosti RNA virusov, ki jih lahko prenašajo na vretenčarje oz. na rastline. Takšno stališče je lahko zavajajoče, saj do takrat še niso bili analizirani RNA virusi pri preostali veliki skupini nevretenčarjev, pri Lophotrochozoa (oz. Spiralia – precej velika skupina, ki obsega mehkužce in ploščate črve). Raziskave Shi in sod. (2016a, 2016b), Li in sod. (2015), Waldron in sod. (2017) in Webster in sod. (2015, 2016) so razširile znanje o raznolikosti in razširjenosti RNA virusov. Naredili so obsežne raziskave raznolikosti RNA virusov, vendar z omejenim vzorčenjem, in sicer pri skupinah Lophotrochozoa (Spiralia) ter pri rakih (Crustacea), pajkovcih (Chelicerata), stonogah (Myriapoda) in žuželkah (Insecta) znotraj skupine Ecdysozoa. Dotaknili so se tudi najstarejših skupin živali (Metazoa), vendar so se usmerili le na ožigalkarje (Cnidaria) in spužve (Porifera) samo preko analize dveh do treh vrst (Shi in sod., 2016a; Waldron in sod., 2017). Pri skupini Lophotrochozoa so se usmerili predvsem na mehkužce (Mollusca), pri rakih pa so v glavnem zajeli skupino Decapoda. **Dokazali so, 1) da je pestrost RNA virusnih genomov bistveno večja, kot je bilo razvidno iz obstoječih podatkov, 2) da je tekom evolucije skupine Protostomia prišlo do »eksplozije« RNA virusnih genomov in 3) da je pestrost RNA viroma ohranjena čez celotno skupino Protostomia** (Webster in sod., 2015; Li in sod., 2015; Webster in sod., 2016; Shi in sod., 2016a; Shi in sod., 2016b; Waldron in sod., 2017). **Center raznolikosti RNA virusov so premaknili iz členonožcev na celotno skupino Protostomia.** S tem so odprli nadaljnja vprašanja o 1) tem, ali lahko najstarejše predstavnike posameznih virusnih klad najdemo že pri predniku živali, 2) ali je bogata pestrost RNA virusnih genomov ohranjena ne samo pri skupini Protostomia, temveč čez celotno skupino nevretenčarjev in 3) ali je skupina Protostomia dejanski izvor raznolikosti RNA virusov ali pa je pravi izvor prednik živali.

Namen predlagane raziskave je torej:

- 1) **Pojasniti razširjenost ter nastanek RNA virusnih skupin, njihove evolucijske in genomske dinamike pri skupinah nevretenčarjev, ki še niso bile raziskane, in primerjati z dosedanjimi podatki iz literature.**
- 2) **Dokazati, da je raznolikost RNA virusov prisotna in ohranjena pri vseh nevretenčarjih.**
- 3) **Dokazati, da so najstarejše skupine živali vir in izvor raznolikosti RNA virusov.**

2.1.1 Cilji in utemeljitev znanstvene relevantnosti

Še vedno je slabo raziskana raznolikost RNA virusov pri najstarejših skupinah Metazoa (Ctenophora, Cnidaria in Porifera), najstarejših skupinah insektov (Collembola, Diplura, Zygentoma in Monocondylia), ploščatih črvih (Platyhelminthes), skupinah rakov Amphipoda in Copepoda in skupinah žuželk Hymenoptera in Lepidoptera. Zato bomo dopolnili znanje o raznolikosti in razširjenosti RNA virusov pri nevretenčarjih z analizo skupin: 1) ploski črvi (Platyhelminthes), 2) najstarejše skupine živali (Metazoa) – rebrače (Ctenophora), ožigalkarji (Cnidaria) in spužve (Porifera), 3) najstarejše skupine žuželk – Collembola (skakači), Diplura (dvorepke), Zygentoma in Monocondylia, 4) skupine rakov – Amphipoda (postrance) in Copepoda (ceponožci) in 5) skupine žuželk – kožekrilci (Hymenoptera) in metulji (Lepidoptera). Pridobili bomo celotne genome za različne predstavnike ss(-)RNA,

ss(+)RNA in dsRNA virusov v nevretenčarskih transkriptomskih podatkovnih bazah. Pridobili bomo tudi celotne ali okvarjene endogenizirane RNA viruse v genomskih podatkovnih bazah nevretenčarjev, kar nam bo pokazalo pretekle infekcije teh organizmov z različnimi RNA virusi. Po primerjavi z dosedanjimi podatki bomo dokazali, da je raznolikost RNA virusov prisotna pri vseh nevretenčarjih. **Končni cilj raziskave je pojasnitev razširjenosti ter nastanka novih RNA virusih skupin, njihove evolucije ter genomske dinamike RNA virusov pri vseh nevretenčarjih.**

2.1.2 Hipoteze

Nevretenčarji predstavljajo največjo skupino živali. **Analize RNA virusov so ustvarile prepričanje, da skupina Protostomia predstavlja glavni vir raznolikosti RNA virusov** (Webster in sod., 2015; Li in sod., 2015; Webster in sod., 2016; Shi in sod., 2016a; Shi in sod., 2016b; Waldron in sod., 2017). Vseeno pa je bilo vzorčenje doslej omejeno. Doslej še niso bile analizirane skupine nevretenčarjev: 1) najstarejše skupine živali (Ctenophora, Cnidaria in Porifera), 2) najstarejše skupine žuželk (Collembola, Diplura, Zygentoma in Monocondylia), 3) raki Amhipoda in Copepoda in 4) skupini Spiralia (predvsem ploščati črvi) in 5) kožekrilcev (Hymenoptera) in metuljev (Lepidoptera). **Zato je prepričanje lahko zmotno.** Trenutno je dostopnih precej transkriptomskih in genomskih podatkov za nevretenčarje, ki nam bodo omogočili sistematično in reprezentativno analizo RNA virusov pri njih. Analiza RNA virusov in endogeniziranih virusov nam bo pokazala njihovo raznolikost in razširjenost pri vseh nevretenčarjih. Pokazala nam bo evolucijski izvor virusnih klad in razkrila center raznolikosti RNA virusov. Naše hipoteze so torej naslednje:

- 1) **Predpostavljamo, da so v transkriptomih in genomih nevretenčarjev prisotni številni, doslej neznan RNA virusi, ki bodo spremenili naš pogled na raznolikost RNA virusov pri nevretenčarjih** – to bomo dokazali z iskanjem homolognih virusnih zaporedij v genomskih in transkriptomskih podatkovnih bazah nevretenčarjev.
- 2) **Predpostavljamo, da je raznolikost RNA virusov prisotna pri vseh nevretenčarjih in ohranjena čez vse nevretenčarje** – to bomo dokazali s filogenomsko analizo in analizo strukture genomov najstarejših skupin živali (Ctenophora, Cnidaria in Porifera), najstarejših skupin žuželk (Collembola, Diplura, Zygentoma in Monocondylia), rakov Amhipoda in Copepoda, skupine Spiralia (predvsem ploščatih črvov), kožekrilcev (Hymenoptera) in metuljev (Lepidoptera). Enaka pestrost strukture RNA genomov pri vseh analiziranih skupinah nevretenčarjev nam bo dokazala, da je raznolikost RNA virusov ohranjena čez vse nevretenčarje.
- 3) **Predpostavljamo, da skupina Protostomia ni dejanski izvor in center raznolikosti RNA virusov, temveč je dejanski izvor prednik živali** – to bomo dokazali s filogenetsko analizo in analizo strukturne raznolikosti RNA viroma najstarejših skupin živali. Enaka pestrost RNA viroma, kot so jo pri skupini Protostomia dokazal Shi in sod. (2016a), in odkritje najstarejših predstavnikov virusnih klad v viromu najstarejših skupin živali nam bodo potrdili domnevo.

Za potrditev hipotez bomo uporabili različne raziskovalne pristope in metode.

2.1.3 Metodologija

Podatke o raznolikosti in razširjenosti RNA virusov bomo pridobili z iterativnim iskanjem homolognih virusnih zaporedij (angl. »*iterative homology search*«) v transkriptomskih (TSA (angl. »*Transcriptome Shotgun Sequences*«)) in endogeniziranih virusih (angl. »*endogenized viral elements*«) v genomskih podatkovnih bazah (WGS (angl. »*Whole Genome Shotgun*«), neredundantno (angl. »*non-redundant*«), RefSeq_genomic, RefSeq_representative_genomes) neraziskanih oz. slabo raziskanih skupin nevretenčarjev z orodjem BLAST. Kot iskalni niz (angl. »*query*«) bomo uporabili proteinsko zaporedje RdRp in drugih proteinov že znanih klasificiranih in neklasificiranih virusov (angl. »*well established or floating genera*«) (Altschul in sod., 1990; Benson in sod., 2013; NCBI Resource Coordinators, 2016; NCBI Resource Coordinators, 2017; NCBI – <https://www.ncbi.nlm.nih.gov/>). Tako bomo pridobili celotne genome za različne predstavnike ss(-)RNA, ss(+)RNA in dsRNA virusov.

Nove RNA viruse in endogenizirane RNA viruse bomo nato klasificirali in okarakterizirali: 1) s filogenetsko analizo na novo pridobljenih RNA virusov, tako jih bomo umestili v kontekst že znane raznolikosti RNA virusov (v ustrezne virusne klade), in 2) s primerjavo strukture njihovega genoma s strukturo genoma že znanih RNA virusov, tako bomo določili razlike v strukturi genomov in prisotnosti oz. odsotnosti določenih proteinskih domen, pojav novih odprtih bralnih okvirjev (ORF-jev) in privzem omen/ORF-jev od drugih virusov (virusov, ki pripadajo drugim virusnim kladam). Pri filogenomski analizi bomo pridobljena homologna zaporedja iz genomskih in transkriptomskih podatkovnih baz nevretenčarjev prevedli (ExPASy – <https://www.expasy.org/>) ter poravnali s programom Clustal Omega (Sievers in Higgins, 2014). Filogenetsko analizo bomo nato izvedli s pomočjo programov MEGA7 in IQTree (Nguyen in sod., 2015; Kumar in sod., 2016). Pri filogenetski analizi bomo uporabili metodi »združevanja sosedov (angl. »*neighbour-joining*«) in največje verjetnosti (angl. »*maximum likelihood*«). Pri primerjavi strukture genoma na novo pridobljenih virusov s strukturo genoma bomo ravno tako virusna zaporedja prevedli (ExPASy – <https://www.expasy.org/>) in določili strukturo genoma s pomočjo programa CDD (angl. »*Conserved Domain Database*«) (Marchler-Bauer in sod., 2017 – CDD).

Po klasifikaciji in karakterizaciji bomo razširjenost na novo pridobljenih virusov med raziskanimi skupinami nevretenčarjev ovrednotili na podlagi deleža oz. prispevka (angl. »*contribution*«) vsake virusne klade k celotnemu najdenemu RNA viromu pri vsaki od analiziranih skupin nevretenčarjev. Po primerjalni analizi z dosedanjimi podatki bomo ugotovili, katere skupine RNA virusov so široko razširjene pri nevretenčarjih in katere imajo omejeno razširjenost pri določenih skupinah nevretenčarjev.

V okviru naše metodologije se bomo srečevali z različnimi izzivi.

2.2 Izzivi

Pri klasifikaciji in karakterizaciji novih RNA virusov in endogeniziranih RNA virusov se bomo srečali z izzivi, ki izvirajo iz: 1) podatkovnih baz, 2) orodij za prepoznavo domen in karakterističnih motivov (CDD; angl. »*Conserved Domain Database*«) in 3) filogenetske analize RNA virusov.

2.2.1 Izzivi, povezani s podatkovnimi bazami

Transkriptomске podatkovne baze so zbirke združenih RNA zaporedij. Skupina RNA zaporedij kot posnetek celotnega transkriptoma specifičnega organizma oz. vrste organizma (v našem primeru različne vrste nevretenčarjev) brez rRNA, vzetega v določenem časovnem okvirju in na določeni lokaciji, predstavlja RNA knjižnico omenjenega organizma (Altschul in sod., 1990; NCBI – <https://www.ncbi.nlm.nih.gov/>). Raziskovalci so lahko transkriptom določene vrste organizma vzorčili večkrat, kar pomeni, da v transkriptomskih bazah za isto vrsto organizma obstaja več knjižnic. Raziskovalci so lahko tekom postopka oblikovanja knjižnic tudi združili transkriptome več organizmov, ki so med seboj ekološko ali evolucijsko povezani (sorodni). Omenjene knjižnice so zato lahko enostavne (predstavljajo transkriptom samo enega organizma) ali sestavljene (predstavljajo transkriptom skupine organizmov). Če imamo opravka z metatranskriptomom, ne moremo določiti, kateremu od organizmov v knjižnici pripada določena RNA sekvenca (najden RNA virus). Transkriptom (transkriptomi), ki je vključen (so vključeni) v knjižnico, so lahko bili izolirani iz celotnega organizma ali samo iz dela organizma - iz organa ali tkiva organizma. Genom RNA virusa, ki je bil najden v določeni knjižnici, lahko izvira iz virusa, ki se je nahajal 1) v celici organizma, 2) v medcelični prostoru organizma, 3) na organizmu. Zato ne moremo določiti, ali virus s pripadajočim genomom, najdenim v podatkovnih bazah določene vrste organizma, dejansko aktivno okužuje organizem, je patogen za omenjen organizem ali je bil samo naključno prisoten na organizmu. Zato moramo uporabiti tudi genomske podatkovne baze. Genomske podatkovne baze so zbirke združenih popolnih ali nepopolnih zaporedij genomov organizmov (v našem primeru nevretenčarjev). Če najdemo genom virusa v genomskih podatkovnih bazah, pomeni, da se je najden genom virusa integriral v gostiteljev genom in postal endogeniziran virus. Nakažemo lahko na preteklo aktivno okužbo gostitelja z najdenim virusom. Posebej velja opozoriti, da zaradi narave endogeniziranih RNA virusov oz. narave »postopka endogenizacije RNA virusov« večinoma najdemo samo okvarjene

fragmente endogeniziranega genoma v genomih pripadajočih gostiteljev. Najdeni celotni ORF-ji genoma virusa so izjemno redki (Altschul in sod., 1990; NCBI – <https://www.ncbi.nlm.nih.gov/>).

2.2.2 Izzivi, povezani z analizo strukture genoma

Za določitev genomske strukture novih RNA virusov in endogeniziranih virusov so uporabna orodja, kot je CDD (angl. »*Conserved Domain Database*«), ki uporabljajo model strojnega učenja (angl. »*machine learning*«) za primerjavo sekvenc, ugotavljanje ohranjenih motivov, značilnih za določen protein (domeno, družino proteina), in nadaljnjo identifikacijo značilne domene znotraj ORF (npr. proteaza 3C, RdRp družine 1, helikaza naddružine 2 idr.). Pri divergentnih sekvencah z neznanimi motivi program ne bo prepoznal domen. Seveda to ne pomeni nujno, da domena ni prisotna na določenem ORF-ju (Marchler-Bauer in sod., 2017 – CDD).

2.2.3 Izzivi pri filogenetski analizi

Divergentna zaporedja so problematična tudi pri filogenetski analizi. Na podlagi primerjave sekvenc se definirajo matrike medsebojne »razdalje« med sekvencami (angl. »*distance matrix*«). Najbolj divergenta zaporedja (najbolj oddaljena zaporedja od skupine zaporedij) lahko program samo naključno poveže med seboj (angl. »*long branch attraction*«), kar pomeni, da se zaporedja po modifikaciji in izrisu novih dreves naključno pomikajo znotraj drevesa in ne moremo natančno določiti njihove povezanosti (Kumar in sod., 2016; Nguyen in sod., 2015).

2.3 Doprinos k znanosti

Naša raziskava bo dopolnila znanje o raznolikosti in razširjenosti RNA virusov pri ploščatih črvih (Platyhelminthes), kožekrilcih (Hymenoptera), metuljih (Lepidoptera), rakih (Crustacea), predvsem Amphipoda (postranice) in Copepoda (ceponožci), najstarejših skupinah žuželk (Collembola, Diplura, Zygentoma, Monocondylia) in najstarejših skupinah živali (Metazoa) – Ctenophora (rebrače), Cnidaria (ožigalkarji) in Porifera (spužve). Z novimi podatki bo naša raziskava:

- 1) Dokazala, da v genomskih in transkriptomskih podatkovnih bazah obstajajo številni doslej neznani RNA virusi.
- 2) Pojasnila, ali je raznolikost RNA virusov ohranjena čez celotno skupino nevretenčarjev ali je omejena samo pri določenih skupinah nevretenčarjev.
- 3) Obrazložila, ali »eksplozija« raznolikosti RNA virusov izvira iz skupine Protostomia ali se je pojavila že pri najstarejših skupinah živali.
- 4) Pokazala nove rezultate o starosti virusnih klad.

Rezultati naše raziskave bodo pomembni za nadaljnje sklepanje o poteku evolucije RNA virusnih skupin.

Poglavje 3

Metode dela

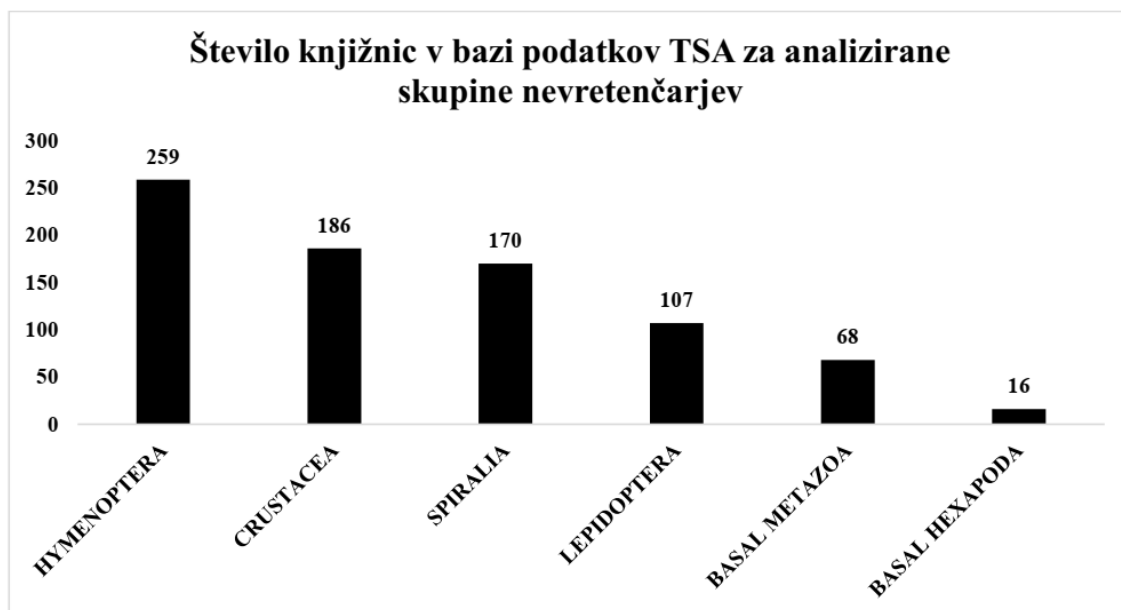
3.1 Pridobivanje podatkov

Za iskanje novih RNA virusov v transkriptomskih bazah podatkov smo potrebovali iskalni niz (angl. »*query*«), ki je predstavljal 1) proteinsko zaporedje RdRp (od RNA odvisne RNA polimeraze), 2) proteinsko zaporedje proteina, specifičnega za določeno virusno skupino, kot so SP24 za negevirus (Hepe-Virga), »Rhabdovirus spike glycoprotein« (Rhabdo_glycop; pfam00974) in »Rhabdovirus nucleocapsid protein« (Rhabdo_ncap; pfam00945) pri skupini Rhabdoviridae (Mononegavirales), in 3) proteinska zaporedja celotnih odprtih bralnih okvirjev (ORF; angl. »*Open Reading Frame*«). Proteinska zaporedja smo pridobili iz podatkovnih zbirk z bibliografsko in biološko anotacijo na strežnikih NCBI (angl. »*National Center for Biotechnology information*«) preko primarnega iskalnika Entrez (Altschul in sod., 1990; Benson in sod., 2013; Sterk in sod., 2007; NCBI Resource Coordinators, 2016; NCBI Resource Coordinators, 2017; NCBI – <https://www.ncbi.nlm.nih.gov/>). NCBI spletni strežnik ima v okviru podatkovne baze GeneBank več ločenih podskupin podatkov: 1) **DNA**: nr (angl. »*non-redundant*«), WGS (angl. »*Whole Genome Shotgun*«), HTG (angl. »*High-Throughput Genomic sequences*«), GSS (angl. »*Genomic Survey Sequences*«) 2) **RNA**: EST (angl. »*Expressed Sequence Tags*«), TSA (angl. »*Transcriptome Shotgun Assembly Database*«), HTC (»*High-Throughput cDNA*«), 3) **proteini**: nr (angl. »*non-redundant*«), 4) **DNA in RNA**: SRA (angl. »*Sequence Read Archive*«), 5) ostale: STS (angl. »*Sequence Tagged Sites*«), TPA (angl. »*Third Party Annotation*«), UNA (angl. »*Unannotated Sequences*«), SYN (angl. »*Synthetic Sequences*«), PAT (angl. »*Patented Sequences*«). Strežnik NCBI omenjene GeneBank podskupine podatkov razdeli v odseke: NCBI Gene, NCBI Nucleotide, NCBI Genome, NCBI Protein, NCBI Taxonomy, NCBI Reference Sequence Database, NCBI Viral Genomes idr. **NCBI Viral Genomes** vsebuje zbirko celotnih genomov virusov. ExPASy preko več divizij ponuja dostop do proteinskih, nukleotidnih in transkriptomskih baz podatkov in hkrati orodij za njihovo analizo: BLAST, ExPASy Translate (ExPASy – <https://www.expasy.org/>). **ViralZone** je ena izmed baz podatkov v okviru strežnika ExPASy, ki združuje nukleotidna in proteinska zaporedja virusov s povezavo na NCBI strežnik (NCBI Genome) in UniProtKB bazo podatkov (Hulo in sod., 2011 – ViralZone; SIB (Swiss Institute of Bioinformatics) – <https://www.sib.swiss/>; PIR (Protein Information Resource) – <https://pir.georgetown.edu/>). Hkrati vključuje bibliografske podatke o virusih. Podatke o proteinskih zaporedjih RdRp dobro okarakteriziranih virusov – referenčnih virusov, ki spadajo v dobro definirane in okarakterizirane klade, in RdRp sicer opisanih, vendar slabše okarakteriziranih oz. neklasificiranih virusov (angl. »*floating genera*«), smo pridobili predvsem iz baz podatkov NCBI Gene, NCBI Protein, NCBI Genome, NCBI Viral Genomes, NCBI Taxonomy in ViralZone. Iz omenjenih podatkovnih baz smo pridobili tudi proteinska zaporedja proteinov, specifičnih za posamezno virusno skupino (SP24), in proteinska zaporedja celotnih ORF-jev, ki se prepisujejo kot poliproteini.

3.2 Iterativno iskanje homolognih zaporedij (RNA virusov) v transkriptomskih bazah podatkov nevretenčarjev in endogeniziranih RNA virusov v genomskih bazah podatkov nevretenčarjev

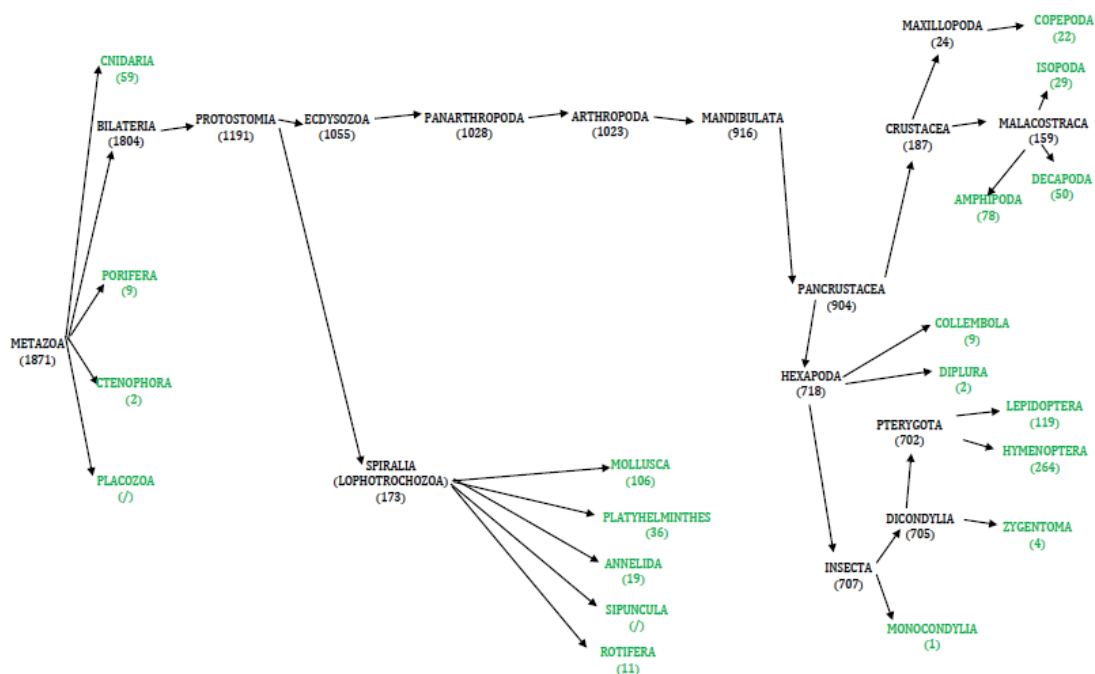
3.2.1 Iterativno iskanje homolognih zaporedij v transkriptomskih podatkovnih bazah nevretenčarjev

Kot že omenjeno, raziskovalne skupine v določenem časovnem intervalu vzorčijo celotne organizme, organe ali tkiva organizmov (v našem primeru nevretenčarjev) (Altschul in sod., 1990; Benson in sod., 2013; NCBI Resource Coordinators, 2016; NCBI Resource Coordinators, 2017; NCBI – <https://www.ncbi.nlm.nih.gov/>). Iz vzorčenih organizmov ali delov organizmov izolirajo celotno RNA, ki ji nato odstranijo ribosomalno RNA (rRNA), tako da pridobijo celoten transkriptom organizmov. Eventuelno lahko tudi združijo RNA, pridobljeno iz več organizmov, ki so evolucijsko in/ali okoljsko povezani (Shi in sod., 2016a). V nadaljevanju se oblikujejo transkriptomске knjižnice. Transkriptomске knjižnice se lahko generirajo na dva načina. Prvi način poteka s pretvorbo RNA v cDNA po dveh protokolih: 1) pretvorbo RNA v cDNA, njeno fragmentacijo na oligonukleotide z lepljivimi konci, popravo lepljivih koncev, vezavo adapterjev in sekvenciranje ali 2) fragmentacijo RNA na oligonukleotide z lepljivimi konci, pretvorbo fragmentov RNA v cDNA, popravo lepljivih koncev, vezavo adapterjev in sekvenciranje (Galinier in sod., 2017; Dheilly in sod., 2015; Griffith in sod., 2015; Liu in sod., 2013; Wang in sod., 2009). V tem primeru dobimo transkriptomске knjižnice, ki jih predstavljajo fragmenti, dolgi od 30 do 400bp (Wang in sod., 2009). Sledi sekvenciranje transkriptomskih knjižnic. Drugo obliko transkriptomskih knjižnic predstavlja kar izolirana RNA brez ribosomalne RNA. Izolirana RNA lahko izvira iz enega ali več organizmov. V tem primeru sledi direktno sekvenciranje femtomolarnih količin RNA z uporabo poli(dT) (deoksitimidin) lovilcev, ki se vežejo na poli-A rep mRNA (Ozsolak in sod., 2009). PAPI polimeraza iz *E.coli* (angl. »*Escherichia coli poly(A) polymerase* k) generira 150 nukleotidov dolg poli-A rep na RNA molekulah. Poli-A rep se nato veže na nosilce z vezanim poli(dT) (deoksitimidin) verigo. Po vezavi na lovilce se RNA zapolni z deoksitimidin trifosfatom (dTTP) in blokira z vezavo fluorescentno obarvanih VT-A, VT-C in VT-G nukleotidov. V nadaljevanju sledi sinteza (podaljševanje) komplementarne DNA verige tekom sekvenciranja (Ozsolak in sod., 2009). Po sekvenciranju se zaporedja odložijo (zavedejo) v SRA bazo podatkov ali jih združijo v pripadajoče soseske in nato v ogrodja. Združeni transkriptomi, ki pripadajo eni knjižnici (organizmu ali skupini organizmov iz določene ekološke niše, iz katerih RNA je bila izolirana v določenem časovnem okviru) se nato odložijo (zavedejo) v transkriptomsko bazo podatkov TSA (angl. »*Transcriptome Shotgun Assembly Database*«), divizijo v okviru GeneBank. Slika 58 prikazuje število knjižnic, ki so bile za posamezno skupino nevretenčarjev zavedene v podatkovni bazi TSA (angl. »*Transcriptome Shotgun Assembly*«) do 8. 1. 2018 (Slika 3.1) (Altschul in sod., 1990; Benson in sod., 2013; NCBI Resource Coordinators, 2016; NCBI Resource Coordinators, 2017; NCBI – <https://www.ncbi.nlm.nih.gov/>).



Slika 3.1: Število knjižnic v bazi podatkov TSA (angl. »Transcriptome Shotgun Assembly«) za analizirane skupine nevretenčarjev (8. 1. 2018).

Število knjižnic se praktično dnevno spreminja. Slika 59 prikazuje število knjižnic, prisotnih v transkriptomski bazi podatkov TSA (angl. »Transcriptome Shotgun Assembly«) glede na analizo, narejeno 6. 2. 2018 in 7. 2. 2018 (Slika 3.2). Vidimo lahko, da je v mesecu januarju TSA najbolj zrasla pri skupini Lepidoptera (za 12 knjižnic) in Hymenoptera (za 5 knjižnic).



Slika 3.2: Število transkriptomskih (TSA, angl. »Transcriptome Shotgun Assembly«) knjižnic pri posamezni skupini živali (Metazoa). Analiza je bila narejena 7. 2. 2018.

Največje število zavedenih transkriptomskih knjižnic smo 7. 2. 2018 našli pri skupini Hymenoptera (kožekrilci) – 264, sledile so skupine Crustacea (raki) – 187, Lophotrochozoa (polži, školjke, glavonožci, ploski črvi idr.) – 173, Lepidoptera (metulji) – 119 in najstarejših

skupin živali 70. Od 187 knjižnic pri skupini Crustacea jih je 50 pripadalo skupini Decapoda (deseteronožci), 78 skupini Amphipoda (postranice), 29 skupini Isopoda (enakonožci ali mokrice) in 22 skupini Copepoda (ceponožci). Od 173 knjižnic pri skupini Lophotrochozoa je bilo samo 36 knjižnic narejenih iz vzorcev RNA ploskih črvov (Platyhelminthes), od preostalih 137 knjižnic pa jih je 106 pripadalo mehkužcem (Mollusca), 19 skupini Annelida in 11 skupini Rotifera. Od skupno 70 knjižnic pri najstarejših skupinah živali jih je bilo 59 narejenih iz vzorcev RNA ožigalkarjev (Cnidaria), 9 iz vzorcev RNA spužev (Porifera) in samo dve iz vzorcev RNA rebrač (Ctenophora). Samo 16 knjižnic je bilo narejenih iz vzorcev RNA najstarejših skupin žuželk. Od tega 9 iz RNA skupine Collembola, dve iz RNA skupine Diplura, štiri iz RNA skupine Zygentoma in samo 1 iz RNA skupine Monocondylia (organizem *Machilis hrabei*). Več knjižnic je bilo narejenih iz RNA iste vrste organizma (sicer drug časovni okvir in druga ekološka niša). Pri skupini Hymenoptera ni bilo neenakomerne zastopanosti knjižnic, ki so pripadale posameznim vrstam skupine. Tri knjižnice so bile npr. narejene iz vzorca RNA organizma *Cephus cinctus*, vendar je načeloma vsaki vrsti organizma pripadala po ena knjižnica. Pri skupini Lepidoptera so po štiri knjižnice zavzemali organizmi *Spodoptera exigua* in *Athetis lepigone* in po tri knjižnice *Acanthopterocetes unifascia* in *Agathiphaga queenslandensis*, ostali organizmi so načeloma bili zastopani z eno knjižnico. Pri rakihi sta bila bogato zastopana rodova Proasellus (Isopoda) z 22 knjižnicami in Eulimnogammarus (Amphipoda) s 15 knjižnicami. Pri skupini Lophotrochozoa sta bila med mehkužci najbolj zastopana rodova Conus z 10 knjižnicami in Aplysia z 9 knjižnicami. Pri ploskih črvih je bil najbolj zastopan rod Schmidtea z 10 reprezentativnimi knjižnicami. Znotraj najstarejših skupin žuželk je bila skupina Collembola dokaj enakomerno zastopana s knjižnicami – ena knjižnica je načeloma pripadala eni vrsti organizma. Skupina Diplura je zajemala samo knjižnico organizma *Campodea augens*. Skupina Zygentoma je zajemala knjižnice organizmov *Thermobia domestica*, *Tricholepidion gertschi* in *Atelura formicaria*. Znotraj najstarejših skupin živali je bilo pri ožigalkarjih šest knjižnic narejenih iz RNA vzorcev rodu Hydra, po tri pa iz RNA vzorcev rodu Hydractinia, Nematostella in Porites. Pri skupini Porifera so od skupno devetih knjižnic kar štiri pripadale organizmu *Halisarca caerulea*. Pri skupini Ctenophora (rebrače) je bila samo knjižnica, narejena iz RNA vzorca organizma *Mnemiopsis leidyi* (Altschul in sod., 1990; Benson in sod., 2013; NCBI Resource Coordinators, 2016; NCBI Resource Coordinators, 2017; NCBI – <https://www.ncbi.nlm.nih.gov/>).

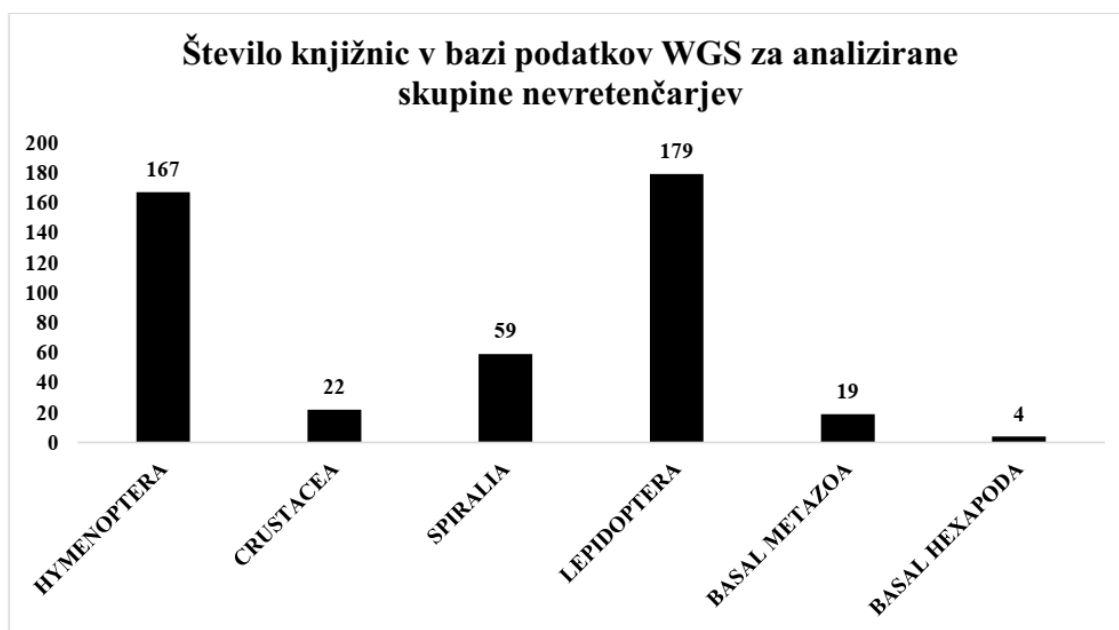
Proteinska zaporedja od RNA odvisne RNA polimeraze referenčnih virusov (dobro okarakteriziranih in taksonomsko ovrednotenih virusov) (angl. »*established viral genera*«) in slabše okarakteriziranih virusov (angl. »*floating viral genera*«) smo v prvi fazi uporabili kot iskalni niz (angl. »*query*«) za iterativno (angl. »*iterative*«) iskanje po transkriptomskih podatkovnih bazah. Analiza na nivoju proteinov je bolj informativna za bolj divergentna proteinska zaporedja. Za iterativno iskanje smo uporabili orodje BLAST preko strežnika NCBI. Orodje poišče najbolj podobna zaporedja (Pevsner, 2009; Altschul in sod., 1990; Benson in sod., 2013; NCBI Resource Coordinators, 2016; NCBI Resource Coordinators, 2017; NCBI – <https://www.ncbi.nlm.nih.gov/>).

Homologna zaporedja smo iskali v prevedenih transkriptomskih (TSA, angl. »*Transcriptome Shotgun Assembly*«) nukleotidnih bazah različnih skupin nevretenčarjev (Collembola, Diplura, Zygentoma, Monocondylia, Hymenoptera, Lepidoptera, Lophotrochozoa in Crustacea), ki so podobne našemu iskalnemu nizu. V prvi fazi so bili naši iskalni nizi zaporedja RNA virusnih RdRp, ki smo jih pridobili iz podatkovnih baz na strežnikih ExPASy in NCBI (ViralZone, NCBI Viral Genomes, NCBI Taxonomy) (ExPASy – <https://www.expasy.org/>, NCBI – <https://www.ncbi.nlm.nih.gov/>). Kot iskalne nize smo uporabili tudi celotne ORF ali zaporedja proteinov, ki so omejena samo na eno RNA virusno skupino (npr. SP24, ki je omejen samo na negeviruse). Uporabljali smo orodje tblastn. S tem smo pridobili nukleotidna zaporedja celotnih virusnih genomov ali samo dela virusnih genomov (npr. samo RdRp), ki so bili sekvencirani v okviru določene knjižnice. V drugi fazi smo naše zadetke na novo uporabili kot iskalne nize (iterativno iskanje). S tem smo povečali verjetnost, da najdemo najbolj divergentne virusne genome.

3.2.2 Iterativno iskanje homolognih zaporedij v genomskih podatkovnih bazah nevretenčarjev

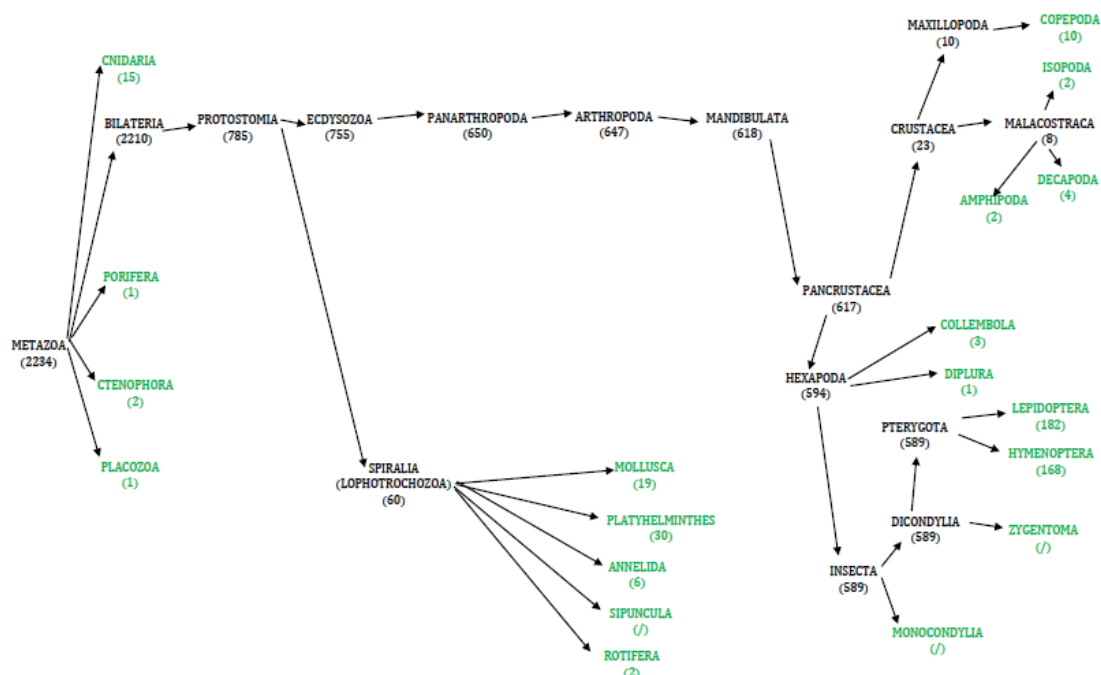
Tudi generiranje genomskih knjižnic se začne z vzorčenjem celotnega organizma, organa ali tkiva organizma (v določenem časovnem okvirju in ekološki niši). Po vzorčenju sledi

izolacija DNA, fragmentacija na oligonukleotide z lepljivimi konci, ligacija adapterjev in sekvenciranje. Po sekvenciranju pridobimo kratke odčitke genomske DNA. Odčitke lahko direktno zavedemo v bazo podatkov SRA ali jih združimo v soseske (angl. »contigs«) in v naslednjem koraku v ogrodja (angl. »scaffolds«). Združena zaporedja nepopolnih ali popolnih genomov vzorčenih organizmov se nato odložijo v genomske baze podatkov (nr (angl. »non-redundant«) – glavna divizija GeneBank ali WGS (angl. »Whole Genome Shotgun«) v okviru baze GeneBank). Popolni anotirani genomi se zavedejo v divizije RefSeq_Representative_Genomes ali RefSeq_genomic v okviru baze GeneBank. Slika 60 predstavlja število knjižnic, ki je bilo zavedenih za posamezno skupino nevretenčarjev v genomski podatkovni bazi WGS (angl. »Whole Genome Shotgun«) do 8. 1. 2018 (Altschul in sod., 1990; Benson in sod., 2013; NCBI Resource Coordinators, 2016; NCBI Resource Coordinators, 2017; NCBI – <https://www.ncbi.nlm.nih.gov/>) (Slika 3.3).



Slika 3.3: Število knjižnic v bazi podatkov WGS (angl. »Whole Genome Shotgun«) za analizirane skupine nevretenčarjev (8. 1. 2018).

Podatki oz. število knjižnic v genomskih podatkovnih bazah se posodablja bistveno počasneje kot podatki v transkriptomskih podatkovnih bazah. Slika 3.4 prikazuje število knjižnic v podatkovni bazi WGS (angl. »Whole Genome Shotgun«) za različne skupine živali (Metazoa) v analizi, narejeni 7. 2. 2018 (Slika 3.4). Do 7. 2. 2018 (v mesecu januarju) se je odložila (zavedla) še ena knjižnica skupine kožekrilcev (Hymenoptera), tri knjižnice skupine metuljev (Lepidoptera) in ena knjižnica skupine Lophotrochozoa – Platyhelminthes (ploski črvi) (Altschul in sod., 1990; Benson in sod., 2013; NCBI Resource Coordinators, 2016; NCBI Resource Coordinators, 2017; NCBI – <https://www.ncbi.nlm.nih.gov/>).



Slika 3.4: Število knjižnic v podatkovni bazi WGS (angl. »Whole Genome Shotgun«) za različne skupine živali (Metazoa). Analiza je bila narejena 7. 2. 2018 (do 17h).

Pri genomskih bazah podatkov je ena vrsta organizma načeloma zastopana z eno knjižnico. Tako kot pri transkriptomskih podatkovnih bazah je bilo do analize števila knjižnic za posamezne skupine živali (Metazoa) tudi pri genomskih podatkovnih bazah največje število knjižnic namenjeno skupini Bilateria (2210 od skupno 2234 knjižnic znotraj skupine Metazoa do 7. 2. 2018). Relativno malo knjižnic je bilo narejenih iz vzorcev najstarejših skupin živali. Znotraj najstarejših skupin živali (Metazoa) je 15 knjižnic pripadalo skupini Cnidaria, ena (*Amphimedon queenslandica*) skupini Porifera, dve skupini Ctenophora (*Mnemiopsis leidyi* in *Pleurobrachia bachei*) in ena (*Trichoplax adhaerens*) skupini Placozoa. Manj kot polovico (785) genomskih knjižnic je znotraj skupine Bilateria pripadalo skupini Protostomia. Večinski delež knjižnic znotraj skupine Protostomia je pobrala skupina Ecdysozoa (755) in znotraj skupine Ecdysozoa skupina členonožcev (Arthropoda) – 647. Znotraj skupine členonožcev je samo 23 knjižnic pripadalo rakom (Crustacea) in 594 skupini Hexapoda. Pri skupini rakov (Crustacea) sta izmed 23 knjižnic dve pripadali skupini Amphipoda (*Hyalella azteca* in *Paryphale hawaiiensis*), dve skupini Isopoda (*Ligia exotica* in *Armadillidium vulgare*), 10 skupini Copepoda in štiri skupini Decapoda (*Caridina multidentata*, *Procambarus fallax*, *Marsupenaeus japonicus* in *Penaeus monodon*). Pri skupini Hexapoda (žuželke) je bila večina knjižnic narejenih na vzorcih skupine Pterygota (589). Skupina metuljev (Lepidoptera) je zajemala 182 knjižnic, skupina kožekrilcev (Hymenoptera) pa 168. Najstarejše skupine žuželk so zajemale skupaj samo štiri knjižnice. Znotraj najstarejših skupin žuželk so samo tri knjižnice (*Orchesella cincta*, *Folsomia candida* in *Holacanthella duospinosa*) pripadale skupini Collembola in ena skupini Diplura (*Catajapyx aquilonaris*). Druga veja skupine Protostomia (Lophotrochozoa oz. Spiralia) je do 7. 2. 2018 zajemala skupno 60 knjižnic. Izmed 60 knjižnic jih je 30 pripadalo ploskim črvom (Platyhelminthes), 19 skupini mehkužcev (Mollusca), šest skupini Annelida in samo dve skupini Rotifera (*Rotaria magnacalcarata* (LJPC01) in *Adineta vaga* (CAWI02)). Do 8. 2. 2018 se je v bazo wgs (angl. »Whole Genome Shotgun«) odložilo še šest knjižnic pri skupini Rotifera (*Branchionus caliciflorus* (NTPY01) – naknadno zavedena knjižnica 7. 2. 2018 predvidoma v večernih urah, *Adineta vaga* (QENK01), *Rotaria magnacalcarata* (QENM01), *Rotaria mancurra* (QENT01) in dve knjižnici skupine *Adineta ricciae* (QEQB01 in QEQJ01) (Altschul in sod., 1990; Benson in sod., 2013; NCBI Resource Coordinators, 2016; NCBI Resource Coordinators, 2017; NCBI – <https://www.ncbi.nlm.nih.gov/>).

Proteinska zaporedja RNA virusnih RdRp in drugih proteinov, specifičnih za določeno RNA virusno skupino, smo uporabili za iskanje podobnih zaporedij v prevedenih

nukleotidnih genomskih podatkovnih bazah (WGS (angl. »*Whole Genome Shotgun*«), nr (angl. »*non-redundant*«), RefSeq_genomic in RefSeq_representative_genomes) nevretenčarjev (Collembola, Diplura, Zygentoma, Monocondylia, Hymenoptera, Lepidoptera, Spiralia (Lophotrochozoa), Crustacea in najstarejše skupine živali (Metazoa)) z orodjem tblastn. S tem smo pridobili večinoma dele virusnih genomov, ki so se integrirali v genom določenega organizma – dobili smo endogenizirane RNA viruse. V drugi fazi smo omenjene zadetke uporabili kot iskalne nize za drugo fazo iskanja (iterativno iskanje) (Altschul in sod., 1990; Benson in sod., 2013; NCBI Resource Coordinators, 2016; NCBI Resource Coordinators, 2017; NCBI – <https://www.ncbi.nlm.nih.gov/>).

3.3 Filogenetska analiza RNA virusov in endogeniziranih RNA virusov

Po pridobivanju novih genomov RNA virusov in endogeniziranih ostankov RNA virusov je sledila njihova filogenetska analiza. Nove pridobljene RNA viruse smo klasificirali v kontekst dobro okarakteriziranih in definiranih RNA virusnih skupin in slabo okarakteriziranih virusnih skupin (angl. »*floating viral genera*«) s filogenetskimi drevesi. Pred filogenetsko analizo smo RNA viruse in endogenizirane RNA viruse mogli umestiti v ustrezne virusne klade.

3.3.1 Pridobivanje podatka o virusni kladi

Pred filogenetsko analizo smo morali pridobiti podatek, v katero virusno klado (Picorna-Calici, Hepe-Virga, Tombus-Noda, Nido, Narna-Levi ...) sodi naš zadetek. Okrog 300 aminokislinskih ostankov dolgo proteinsko zaporedje od RNA odvisne RNA polimeraze (RdRp) vsakega od najdenih RNA virusov in endogeniziranih RNA virusov, ki je vključevalo katalitično domeno, smo uporabili kot iskalni niz za iskanje homologov v virusnih proteinskih podatkovnih bazah z orodjem blastp. Dobili smo seznam podobnih virusnih proteinskih zaporedij RdRp, razvrščen po padajočem odstotku podobnosti (angl. »*percent identity*«) in naraščajočo E vrednostjo. Nova zaporedja, ki smo jih dobili s homolognim iskanjem po virusnih proteinskih podatkovnih bazah, so predstavljale od RNA odvisne RNA polimeraze definiranih virusov, ki so bile zavedene v virusne podatkovne baze in ki sodijo v isto virusno klado kot iskalni niz (v našem primeru RNA virus ali endogeniziran RNA virus, ki smo ga s homolognim iskanjem po transkriptomskih in genomskih bazah podatkov našli z orodjem tblastn). Tako smo preverili, v katero virusno klado sodi vsak od najdenih RNA virusov. Dodatno potrditev o pripadnosti virusni kladi smo pridobili s primerjalno analizo genomske organizacije. Kot že omenjeno, smo s homolognim iskanjem pridobili bodisi cele genome, dele genomov bodisi samo RdRp novih RNA virusov. Pri celotnih genomih smo po prevajanju nukleotidnih zaporedij v proteinska zaporedja pridobili ORF, ki so zajemali eno ali več proteinskih domen. Proteinske domene, ki so specifične za določeno virusno klado (Rhv pri pikornavirusih, Rhabdo_N_cap pri Mononegavirales, SP24 pri negevirusih ...), so predstavljale dodatno potrditev pripadnosti predvideni virusni kladi. Podatek o virusni kladi je bil ključen podatek za nadaljnjo filogenetsko analizo (Altschul in sod., 1990; Pevsner, 2009; Benson in sod., 2013; NCBI Resource Coordinators, 2016; NCBI Resource Coordinators, 2017; NCBI – <https://www.ncbi.nlm.nih.gov/>).

3.3.2 Filogenetska analiza

Po umestitvi novih RNA virusov in endogeniziranih RNA virusov v ustrezne virusne klade je sledila filogenetska analiza. Vsaka filogenetska analiza je sestavljena iz petih osnovnih faz 1) izbora zaporedij, vključenih v filogenetsko analizo, 2) poravnave zaporedij, vključenih v filogenetsko analizo, 3) izbira najboljšega modela zamenjave aminokislin za vsako pozicijo znotraj zaporedja, 4) izris drevesa in 5) ocena drevesa.

3.3.2.1 Izbor zaporedij za filogenetsko analizo

Poleg novih RNA virusov in endogeniziranih RNA virusov smo v analizo vključili tudi klasificirane referenčne viruse in neklasificirane viruse (angl. »*floating genera*«), ki sodijo v isto virusno klado. Na podlagi že znanih klasificiranih in neklasificiranih virusov smo nove RNA viruse in endogenizirane RNA viruse umestili (integrirali) v mrežo evolucijskih odnosov z njimi. Zaporedja referenčnih virusov in neklasificiranih virusov smo izbirali tako,

da smo pokrili celotni vzorčni spekter virusne klade – od najstarejših do najmlajših predstavnikov virusne klade. V izogib redundantnosti smo iz analize izključili zaporedja, ki so več kot 60 % podobna kateremu koli od drugih zaporedij v filogenetski analizi. Hkrati smo pazili, da so bila zaporedja umeščena v isto virusno klado in niso bila manj kot 25 % podobna ostalim zaporedjem v virusni kladi, vključenim v filogenetsko analizo. V analizo smo vključili najbolj ohranjen del od RNA odvisne RNA polimeraze – 300 aminokislinskih ostankov z vključenim motivom C, ki vsebuje katalitično domeno.

3.3.2.2 Poravnava zaporedij

Prva faza v filogenetski analizi je predstavljala poravnavo homolognih zaporedij (angl. »*multiple sequence alignment*«) (Pevsner, 2009; PDB – <https://www.rcsb.org/>; Berman in sod., 2000 – Protein Data Bank (PDB); Edgar, 2004). V kontekst poravnave smo vzeli 300 aminokislinskih ostankov dolgo proteinsko zaporedje RdRp 1) naših zadetkov (RNA virusov in endogeniziranih RNA virusov), 2) definiranih in dobro okarakteriziranih virusov (referenčnih virusov), ki sodijo v isto virusno klado (angl. »*well established viral genera within the same viral clade*«), 3) definiranih, vendar slabše okarakteriziranih virusov, ki naj bi sodili v isto virusno klado (angl. »*floating viral genera within the same viral clade*«).

V okviru doktorata smo zaporedja najprej poravnali v programu MEGA7 (Kumar in sod., 2016). Ker smo ugotovili, da imajo drugi programi za izris filogenetskih dreves (IQTREE) na izbiro več statističnih modelov za substitucijo aminokislin, smo v drugi fazi uporabili progresivne metode (angl. »*progressive alignment*«) poravnave s programom Clustal Omega (Sievers in Higgins, 2014). Aminokislinska zaporedja (300 aa, vključno s katalitično domeno) najdenih RNA virusov in endogeniziranih RNA virusov, vključno z dobro definiranimi referenčnimi virusi (angl. »*established viral genera*«) in slabo okarakteriziranimi virusi (angl. »*floating genera*«), smo v obliki tekstovne datoteke naložili na spletni program in ukazali izpis poravnave v obliki (formatu) PHYLIP. Dobili smo ločeno tekstovno datoteko s poravnanimi zaporedji RdRp v formatu PHYLIP, ki smo jo nato naložili na program za izris drevesa.

3.3.2.3 Izbor optimalnega modela

Filogenetska drevesa smo najprej poskusili narisati s programom MEGA7 (Kumar in sod., 2016). V program smo naložili tekstovno datoteko zaporedij, vključenih v filogenetsko analizo, in zaporedja v programu poravnali. Programu smo ukazali, naj za poravnana zaporedja sam najde optimalni model (Models – Find best DNA / protein models). Nabor vseh možnih modelov s pripadajočimi vrednostmi testnih statistik prikaže v excelovi datoteki. Največkrat je bil kot optimalni model prepoznan LG + G + I (angl. »*LG model predicting Gamma distribution with invariant chains*«) (Pevsner, 2009; Kumar in sod., 2016). Za preverjanje optimalnega modela se uporablja glavna testna statistika AIC (angl. »*Aikake information criterion*«) in BIC (angl. »*Bayesian information criterion*«). Vrednost AIC in BIC se izračuna po enačbi:

$$AIC = -2\ln L + 2N$$

$$BIC = -2\ln L + N\ln(n_1)$$

kjer L predstavlja maksimalno verjetnost za model pri uporabi N neodvisnih parametrov in n_1 podatkov za omenjen model (Pevsner, 2009; Wit in sod., 2012; Kumar in sod., 2016). Manjša, kot je vrednost, boljši je model. Drugi modeli so še: rtREV, WAG, JTT, cpREV, mtREV idr. Program IQTREE je imel na voljo bistveno večje število vgrajenih modelov (Nguyen in sod., 2015). Programu IQTREE, ki smo ga uporabili v drugi fazi, smo ukazali, naj sam izbere optimalni verjetnostni model (Substitution model: možnost Auto). Program v »output« podatkih prikaže nabor vseh možnih modelov z izračunano BIC in AIC vrednostjo in preko vrednosti prikaže najboljši model, ki ga je uporabil pri izrisu drevesa. Modeli, ki so se v tem primeru izkazali kot optimalni, so bili popolnoma odvisni od vključenih zaporedij: Blosum62, LG+F+I+G4, PMB+I+G4 idr. (Nguyen in sod., 2015).

3.3.2.4 Izris filogenetskega drevesa

Obstajajo štiri glavne metode za izris filogenetskih dreves (Pevsner, 2009; Saitou in Nei, 1987; Gascuel in Steel, 1987; Guindon in sod., 2009; Yang in Rannala, 2012): 1) metode na osnovi razdalj (angl. »*distance-based methods*«), kot sta metoda sosedskega odnosa (angl. »*neighbour-joining*«) in neutežna metoda parnih skupin z aritmetično sredino (NPMSA) (angl. »*unweighted pair group method with arithmetic mean*«, UPGMA), 2) metoda največje varčnosti (angl. »*maximum parsimony*«), 3) metoda največje verjetnosti (angl. »*maximum likelihood*«) in 4) sklepanje po Bayesu (angl. »*Bayesian inference*«).

Najprej smo v programu MEGA7 uporabili metodi sosedskega odnosa (angl. »*neighbour-joining*«) ter metodo največje verjetnosti (angl. »*maximum likelihood*«) (Kumar in sod., 2016). Kot metodo izrisa drevesa smo uporabili NNI (angl. »*nearest neighbour interchange*«). Ker ima program IQTREE na voljo več verjetnostnih modelov za substitucijo aminokislin, smo se v drugi fazi odločili za metodo največje verjetnosti (angl. »*maximum likelihood*«) še v omenjenem programu (Nguyen in sod., 2015). Program nam na podlagi optimalnega modela poleg ML (angl. »*maximum likelihood*«) drevesa izriše tudi konsenzno drevo iz 1000 bootstrap dreves. Drevesa smo nato vizualizirali in oblikovali s programom ITOL (Letunic in Bork, 2016).

3.3.2.5 Ocena filogenetskega drevesa

Po izrisu sledi statistična ocena natančnosti dreves. Klasična metoda za oceno signifikance je »*bootstrapping*« metoda za izračun testne statistike (McGarigal in sod., 2000; Guindon in sod., 2009). Po izračunu testne statistike za signifikanco drevesa dobimo vrednost podpore (angl. »*support*«) (bootstrap podpora, aLRT podpora) za vsako vejo v drevesu v odstotkih (McGarigal in sod., 2000; Guindon in sod., 2009).

V programu MEGA7 smo uporabili testno statistiko bootstrap s 500 ponovitvami, vendar je bila zaradi manjšega števila verjetnostnih modelov substitucije, ki so bili na voljo, podpora bistveno slabša (Kumar in sod., 2016). S programom IQTREE smo pri istem naboru zaporedij dobili bistveno boljšo podporo zaradi večjega števila dostopnih verjetnostnih modelov zamenjave aminokislin. Programu smo ukazali tri testne statistike za analizo dreves: »*ultrafast bootstrap*«, »*approximate Bayes*« in aLRT. Program je za vsako vejo pokazal tri vrednosti podpore: 1) aLRT v odstotkih, 2) »*approximate Bayes*« vrednost, ki je manjša ali enaka 1 in se ne izrazi v odstotkih, in 3) »*ultrafast bootstrap podpora*« v odstotkih. Za prikaz podpore konsenznega drevesa (narejenega iz 1000 bootstrap dreves) smo uporabili »*ultrafast bootstrap podpora*« v odstotkih (Nguyen in sod., 2015).

3.4 Primerjalna analiza genomske organizacije RNA virusov pri raziskanih skupinah nevretenčarjev

Po filogenetski analizi smo naredili primerjalno analizo genomske organizacije. Nukleotidna zaporedja novih RNA virusov in endogeniziranih RNA virusov smo najprej prevedli z orodjem ExPASy Translate. Orodje prevede poljubno nukleotidno zaporedje v vseh šestih bralnih okvirjih. Pri tem imamo na voljo več genetskih kodov: standardni, ciliatni, vretenčarski ali nevretenčarski mitohondrijski in drugi. Za prevajanje smo načeloma uporabljali standardni genetski kod. Nevretenčarski mitohondrijski genetski kod smo uporabili za prevajanje nukleotidnega zaporedja RNA virusov, za katere je znano, da okužujejo mitohondrije (skupina Narnaviridae). V nadaljevanju smo v prevodu v vseh šestih bralnih okvirjih poiskali ORF-je in jih razvrstili v pravilnem vrstnem redu. ORF je kodirajoči niz aminokislin, ki je daljši od 50 aminokislin. Vsak ORF smo nato analizirali z orodjem CDD (angl. »*Conserved Domain Database*«), ki na podlagi ohranjenih motivov definira proteinske domene. Orodje je povezano s podatkovnimi bazami Pfam, SMART, COG, TIGRFAM in NCBI Protein Clusters Collection, ki poleg proteinskih zaporedij s tipičnimi motivi vsebujejo podatke o proteinskih domenah (Marchler-Bauer in sod., 2017 – CDD). Orodje CDD (angl. »*Conserved Domain Database*«) vsebuje okrog 50.000 modelov proteinskih domen (Marchler-Bauer in sod., 2017 – CDD). ORF-je, ki se jih ni dalo okarakterizirati z orodjem CDD, smo analizirali z orodjem blastp, pri katerem smo v proteinskih podatkovnih bazah poiskali najbližje homologno zaporedje. Če tudi z orodjem blastp nismo našli ustreznega homolognega zaporedja, smo kodirajoč ORF označili kot neokarakteriziran oz. nov ORF.

Poglavje 4

Rezultati

Rezultati doktorskega dela so predstavljeni v štirih tematskih sklopih: 1) analiza zastopanosti posameznih virusnih klad, 2) filogenetska analiza virusnih klad, 3) analiza endogeniziranih virusov in 4) analiza strukturne organizacije novih virusnih genomov. **Analiza zastopanosti virusnih klad** predstavlja naše ugotovitve o zastopanosti in razširjenosti vsake od virusnih klad, definiranih s strani skupine Shi in sod. (2016a) pri analiziranih skupinah nevretenčarjev. **Filogenetska analiza** predstavlja uvrstitev novih RNA virusov in endogeniziranih virusov, najdenih pri analiziranih skupinah nevretenčarjev, v kontekst trenutno znanih klasificiranih in neklasificiranih virusov. **Analiza endogeniziranih virusov** povzema naše ugotovitve o prisotnosti, odsotnosti in razširjenosti endogeniziranih virusov pri analiziranih skupinah nevretenčarjev. **Analiza strukturne organizacije novih virusnih genomov** predstavlja naše ugotovitve o genomski organizaciji novih RNA virusov in endogeniziranih RNA virusov, najdenih pri analiziranih skupinah nevretenčarjev. Grafi, ki povzemajo naše ugotovitve o zastopanosti novih RNA virusov in endogeniziranih virusov pri analiziranih skupinah nevretenčarjev, so prikazani v okviru rezultatov. Grafi, ki povzemajo ugotovitve skupin Shi in sod. (2016a), so prikazani znotraj Dodatkov (**Dodatek A**). Reprezentativni primeri filogenetskih dreves iz vsake od analiziranih skupin nevretenčarjev so prikazani znotraj rezultatov. Ostala filogenetska drevesa so prikazana znotraj Dodatkov (**Dodatek B**). Reprezentativni primeri analize strukture genoma so prikazani znotraj rezultatov. Ostale slike analiziranih struktur genomov so prikazane znotraj Dodatkov (**Dodatek C**). Zaradi slabše berljivosti slik smo vse slike znotraj tematskega sklopa analize zastopanosti virusnih klad, filogenetske analize in analize strukturne organizacije novih virusnih genomov prikazali v elektronski obliki (**CD**). Grafi analize zastopanosti virusnih klad se nahajajo znotraj istoimenske datoteke »Analiza zastopanosti virusnih klad«. Slike filogenetskih dreves se nahajajo znotraj datoteke »Filogenetska analiza« in slike, ki prikazujejo strukturno organizacijo novih virusnih genomov, se nahajajo znotraj datoteke »Analiza strukturne organizacije novih virusnih genomov«.

4.1 Analiza zastopanosti virusnih klad

4.1.1 Analiza zastopanosti posameznih virusnih klad, kot jih je definirala skupina Shi in sod. (2016a)

Zastopanost posameznih virusnih klad, kot jih je definirala skupina Shi in sod. (2016a), smo analizirali z dveh perspektiv. Prva vključuje analizo zastopanosti virusnih klad na podlagi rezultatov Shi in sod. (2016a), druga vključuje rezultate o zastopanosti posameznih virusnih klad pri skupinah nevretenčarjev, analiziranih v okviru doktorskega dela. V okviru analize zastopanosti posameznih virusnih klad so bili virusi razdeljeni na virusne klade, kot jih je definirala skupina Shi in sod. (2016a). Rezultati analiziranih skupin nevretenčarjev v okviru doktorata so bili primerjani z rezultati skupine Shi in sod. (2016a). Grafi so prikazani tudi v elektronski obliki (**CD**).

Pri analizi rezultatov skupine Shi in sod. (2016a) smo pri posameznih skupinah žuželk (Insecta), rakov (Crustacea) in Lophotrochozoa (Spiralia) najprej analizirali, katere knjižnice vsebujejo vzorce RNA predstavnikov posameznega reda (skupine). Izključili smo mešane knjižnice, narejene iz RNA vzorcev različnih predstavnikov členonožcev

(Arthropoda), različnih predstavnikov rakov (Crustacea) in različnih predstavnikov žuželk (Insecta). Pri žuželkah smo upoštevali knjižnice, narejene na nivoju redov Blattodea, Coleoptera, Diptera, Lepidoptera, Orthoptera, Dermaptera, Odonata in Hemiptera. Vsi organizmi, vključeni v nastanek iste knjižnice, so morali biti predstavniki istega reda. Pri raki smo upoštevali knjižnice, narejene na nivoju skupin Maxillopoda, Malacostraca in Branchiopoda. Izključili smo mešane knjižnice, v katerih nastanek so bili hkrati vključeni predstavniki dveh ali več skupin rakov. Pri skupini Lophotrochozoa (Spiralia) smo se osredotočili na knjižnice, narejene na predstavnikih mehkužcev (Mollusca), skupine Sipuncula in skupine Annelida. V primeru skupine Platyhelminthes, ki je v raziskavi zastopana s štirimi knjižnicami, je bilo v raziskavi Shi in sod. (2016a) najdenih premalo virusov, da bi jih vključili v analizo. V primeru skupine Lophotrochozoa skupina Shi in sod. (2016a) ni tvorila mešanih knjižnic, v katerih bi bili hkrati vključeni predstavniki različnih skupin. Hkrati smo v analizi upoštevali samo viruse, ki so bili najdeni samo v eni knjižnici. V primeru ploščatih črvov je bil samo en primer virusa, ki je bil najden samo v knjižnici SGJSC in pripada kladi Mono-Chu. Zato skupine ploskih črvov (Platyhelminthes) nismo vključili v raziskavo.

V okviru analize zastopanosti virusnih skupin pri raziskanih skupinah nevretenčarjev v okviru doktorata smo RNA viruse razdelili na virusne klade in prešteli število najdenih virusov pri posamezni skupini nevretenčarjev, ki so bili v okviru filogenetske analize uvrščeni v posamezno virusno klado. Najstarejše skupine živali smo razdelili na ožigalkarje (Cnidaria), spužve (Porifera) in rebrače (Ctenophora). Rezultati analize so prikazani samo za ožigalkarje (Cnidaria) in spužve (Porifera). V genomskih in transkriptomskih knjižnicah rebrač so bili najdeni samo trije virusi, ki so bili uvrščeni v virusne klade Mono-Chu, Orthomyxo in Picorna-Calici, zato skupine nismo vključevali v analizo. Skupino Lophotrochozoa smo razdelili na ploščate črve (Platyhelminthes), skupino Annelida, skupino Rotifera in mehkužce (Mollusca). Pri skupini kolobarnikov (Annelida) je bil najden virus uvrščen v klado Narna-Levi in virus uvrščen v klado Tombus-Noda. Dva virusa, uvrščena v kladi Narna-Levi in Tombus-Noda, sta bila najdena tudi pri skupini Rotifera. Zato tudi omenjenih skupin nismo vključili v analizo. Skupino rakov (Crustacea) smo razdelili na postranice (Amphipoda), deseteronožce (Decapoda), ceponožce (Copepoda) in enakonožce (Isopoda). Pri najstarejših skupinah žuželk smo našli pet virusov, uvrščenih v klado Hepe-Virga, en virus, uvrščen v klado Narna-Levi, en virus, uvrščen v klado Mono-Chu, dva virusa, uvrščena v klado Orthomyxo, sedem virusov, uvrščenih v klado Picorna-Calici, en virus, uvrščen v klado Reo, dva virusa, uvrščena v klado Tombus-Noda, en virus, uvrščen v klado Toti-Chriso, in en Qinvirus. Zaradi korektnosti statistične analize najstarejših skupin žuželk v omenjeno analizo nismo vključevali.

4.1.1.1 Rezultati Shi in sod. (2016a)

Shi in sod. (2016a) so oblikovali 87 knjižnic. Od tega je bilo 21 knjižnic narejenih na RNA vzorcih žuželk (Insecta), šest na RNA vzorcih pajkovcev (Chelicerata), tri na RNA vzorcih stonog (Myriapoda), 20 na RNA vzorcih rakov (Crustacea), devet na vzorcih skupine Nematoda, 21 na vzorcih skupine Lophotrochozoa, ena na vzorcih ožigalkarjev, ena na vzorcih iglokožcev (Echinodermata) in ena na vzorcih plaščarjev (Tunicata). Izmed 21 knjižnic, narejenih na vzorcih žuželk (Insecta), so štiri mešane: WHZM, CC, insectZJ in WHCCII. WHZM vsebuje RNA vzorce predstavnikov redov Siphonaptera in Hymenoptera. Knjižnica insectZJ vsebuje RNA vzorce predstavnikov redov Coleoptera, Diptera, Neuroptera in Lepidoptera. Knjižnica CC vsebuje RNA vzorce redov Diptera in Hemiptera. Knjižnica WHCCII vsebuje RNA vzorce predstavnikov redov Coleoptera, Hymenoptera, Hemiptera, Diptera, Orthoptera in Odonata. Ostale knjižnice so narejene na RNA vzorcih predstavnikov, ki pripadajo istemu redu. Izmed 20 knjižnic rakov sta dve mešani knjižnici: WLJQ in shrimp. Ostale knjižnice vsebujejo RNA vzorce predstavnikov, ki pripadajo isti skupini – večinoma pripadajo redovom Decapoda (razred Malacostraca), vključene pa so tudi skupine Isopoda (razred Malacostraca), Diplostraca (razred Branchiopoda), Sessilia (razred Maxillopoda) in nadred Hoplocarida (razred Malacostraca). Odločili smo se, da omenjene knjižnice brez mešanih knjižnic obravnavamo na nivoju razredov (Malacostraca, Maxillopoda in Branchiopoda). Skupina Lophotrochozoa (Spiralija) je zajemala 21 knjižnic, 13 z vzorci predstavnikov mehkužcev (Mollusca), štiri z vzorci ploščatih črvov (Platyhelminthes), eno z vzorci skupine Sipuncula in tri z vzorci Annelida. V tem primeru skupina ni oblikovala mešanih knjižnic. Štiri knjižnice – Arthropodmix, SHWCII, SHWC in WHCC – so vsebovale mešane RNA vzorce predstavnikov členonožcev (rakov, pajkovcev in

žuželk). V vsaki od knjižnic so Shi in sod. (2016a) določili delež virusne RNA, ki je v posameznih knjižnicah variirala od 0,5 % do 87 %. Vsaka knjižnica je vsebovala 1–20 virusnih vrst, od tega 1–6 virusnih vrst, katerih genom predstavlja več kot 0,1 % celotnega vzorca RNA posamezne knjižnice (brez rRNA, ki je bila odstranjena med postopkom izdelave knjižnic). S sekvenciranjem knjižnic in bioinformatiko analizo so našli 1445 novih RNA virusov. Virusne skupine, tako klasificirane po ICTV (angl. »International Committee on Taxonomy of viruses«) kot neklasificirane, je umestila v 16 virusnih klad: Astro, Birna, Hepe-Virga, Hypo, Luteo-Sobemo, Narna-levi, Bunya-Arena, Mono-Chu, Orthomyxo, Nido, Partiti-Picobirna, Permutotetra, Picorna-Calici, Reo, Tombus-Noda in Toti-Chryso. Znotraj vsake od klad je v filogenetski analizi pokazala oblikovanje novih virusnih skupin. Hkrati so definirali nove virusne klade: Weivirus, Qinvirus, Yuevirus, Yanvirus in Zhaovirus. Poimenovali so tudi nove oblikovane skupine pri kladi Picorna-Calici. Najdene viruse so umestili v virusne klade v kontekstu klasificiranih in neklasificiranih virusov.

Za analizo zastopanosti virusnih klad pri žuželkah (Insecta) smo izključili mešane knjižnice žuželk in mešane knjižnice členonožcev. Hkrati nismo upoštevali virusov, ki so bili najdeni v več knjižnicah hkrati. Opazimo lahko vzorce zastopanosti virusnih klad, ki so unikatni za vsak red žuželk. Nobena virusna klada ni enakomerno zastopana čez vse redove žuželk. Hkrati lahko opazimo, da imajo nekateri redovi žuželk, kot so Lepidoptera, Dermaptera in Blattodea, vsak zastopan z eno samo knjižnico, manjši nabor prisotnih virusnih klad. Pri skupini Dermaptera najdemo prisotne samo tri, pri Lepidoptera štiri in pri Blattodea samo pet virusnih klad. V knjižnicah omenjenih redov žuželk je bilo najdenih tudi manj kot deset virusov. Presenetljivo ima red Coleoptera nabor prisotnih virusnih klad, kljub temu da je zastopan samo z eno samo knjižnico. Vzrok za to je lahko visok delež virusne RNA v vsaki izmed knjižnic. Več virusne RNA v knjižnicah pomeni večjo verjetnost, da v njej najdemo pester nabor divergentnih virusov. Pri redu Coleoptera sta kladi Picorna-Calici in Partiti-Picobirna najbolj zastopani. Diptera, Hemiptera in Odonata imajo najbolj pester nabor prisotnih raznolikih virusnih klad. Red Diptera je bil v raziskavi zastopan s šestimi knjižnicami, Odonata samo z eno, Hemiptera pa s tremi knjižnicami. V knjižnicah predstavnikov redov Diptera in Odonata je bilo najdeno največje število RNA virusov v primerjavi z ostalimi redovi, in sicer pri obeh redovih okrog 100. Od tega sta kladi Mono-Chu in Bunya-Arena pri skupini Diptera nekoliko bolj zastopani, kladi Qinvirus in Nido pa samo s posameznimi virusnimi predstavniki. Glede na to, da je red Odonata zastopan samo z eno samo knjižnico, v njej pa je bilo najdenih okrog 100 virusov, pomeni, da ima knjižnica visok delež virusne RNA in v njenem okviru enega najbolj pestrih naborov raznolikih RNA virusov. Pri redu Odonata sta najbolj bogato zastopani skupini Picorna-Calici in Partiti-Picobirna. Klada Picorna-Calici je najbolj zastopana tudi pri redovih Hemiptera in Orthoptera. Če raziskavo Shi in sod. (2016a) pogledamo sumarično, opazimo, da pri žuželkah klada Picorna-Calici izstopa kot najbolj bogato zastopana RNA virusna klada, česar ne moremo opaziti pri omenjenem grafu (**Dodatek A; A.1, Slika A.1**). Dejanska slika zastopanosti virusnih klad bi se spremenila, če bi gledali na žuželke sumarično in vključili tudi mešane knjižnice žuželk (WHZM, CC, insectZ) in WHCCII). Gre za knjižnice z relativno visokim deležem virusne RNA, kar pomeni, da nosijo pester nabor raznolikih virusov, ki jih v omenjeno analizo zastopanosti virusnih klad pri žuželkah nismo vključili. Filogenetska analiza je pokazala velik delež novih najdenih RNA virusov, ki izvirajo ravno iz mešanih knjižnic žuželk. Pri žuželkah je bilo skupno najdenih okrog 460 predstavnikov vseh virusnih klad, vključno z mešanimi knjižnicami insektov in členonožcev (100 od tega je samo pikornavirusov). Če je bilo v knjižnicah, ki vsebujejo vzorce RNA organizmov, ki pripadajo enemu samemu redu žuželk (npr. samo redu Diptera, redu Hemiptera ali samo redu Odonata), najdenih okrog 290 virusov, pomeni, da je vsaj okrog 170 virusov, ki pripadajo definiranim kladam, moralo biti najdenih v mešanih knjižnicah insektov in členonožcev ali/in so bili najdeni v več knjižnicah hkrati. Zanimivo pa red Odonata izstopa, saj je bilo v eni sami knjižnici QTM najdenih kar 100 od celotnih 290 novih virusov, kar je nekoliko več kot tretjina.

Pri raki je slika podobna. Skupini Maxillopoda in Branchiopoda (raki listonožci) sta bili zastopani z eno samo knjižnico, razred Malacostraca (višji raki) pa s 17 knjižnicami. Dve knjižnici rakov sta mešani. Če štejemo tudi mešane knjižnice, je bilo vseh virusov, najdenih pri raki, okrog 430. Če štejemo knjižnice, ki vsebujejo samo predstavnike enega reda in s tem enega razreda, je bilo najdenih okrog 290, kar pomeni, da je vsaj 140 najdenih virusov izviralo iz mešanih knjižnic rakov in mešanih knjižnic členonožcev in/ali so bili najdeni v več knjižnicah hkrati. Od 290 najdenih virusov jih je bilo cca. 230 najdenih pri razredu

Malacostraca, kar je razumljivo glede na to, da je zastopan s kar 17 knjižnicami. Pri tem razredu najdemo dokaj neenakomeren vzorec zastopanosti virusnih skupin, v katerem prednjačijo klade Picorna-Calici, Narna-Levi in Tombus-Noda. Čeprav je bilo pri vsakem od razredov Maxillopoda in Branchiopoda najdenih okrog 30 RNA virusov, je več raznolikih virusnih klad prisotnih pri Maxillopoda, ki so hkrati dokaj enakomerno zastopane. Nekoliko prednjači samo skupina Narna-Levi. Skupina Branchiopoda ima najmanjši nabor raznolikih virusnih klad, prednjačijo pa Picorna-Calici in Narna-Levi (**Dodatek A; A.1, Slika A.2**).

Pri skupini Spiralia (Lophotrochozoa) ni bilo narejenih mešanih knjižnic. Če vključimo sumarično sliko (tudi viruse, ki so bili najdeni v več knjižnicah hkrati), dobimo podobno število, kot če izključimo viruse, ki so bili najdeni v več knjižnicah hkrati, in sicer okrog 330. Iz tega sledi, da so praktično vsi virusi izvirali iz ene same knjižnice. Kar 13 od 21 knjižnic izvira iz mehkužcev, štiri pripadajo ploskim črvom, tri skupini Annelida in ena skupini Sipuncula. Od skupno 330 najdenih virusov jih je bilo okrog 280 najdenih pri mehkužcih (Mollusca). Zato je razumljivo, da pri skupini mehkužcev najdemo najbolj pester nabor prisotnih skupin, pri katerih izrazito prednjači skupina Picorna-Calici, bogato pa so zastopane tudi Tombus-Noda, Narna-Levi in Weivirus. Kljub temu, da je skupina Sipuncula zastopana samo z eno knjižnico, lahko tudi pri omenjeni skupini najdemo pester nabor prisotnih virusnih klad, pri katerih izrazito prednjači skupina Picorna-Calici. Pri skupini Sipuncula je bilo najdenih okrog 20 virusov. Podobna je situacija pri skupini Annelida. Vendar gre očitno za nekoliko revnejše knjižnice z virusno RNA. V treh knjižnicah je bilo najdenih okrog 30 virusov, pri čemer po zastopanosti prednjačita virusni kladi Picorna-Calici in Narna-Levi (**Dodatek A; A.1, Slika A.3**).

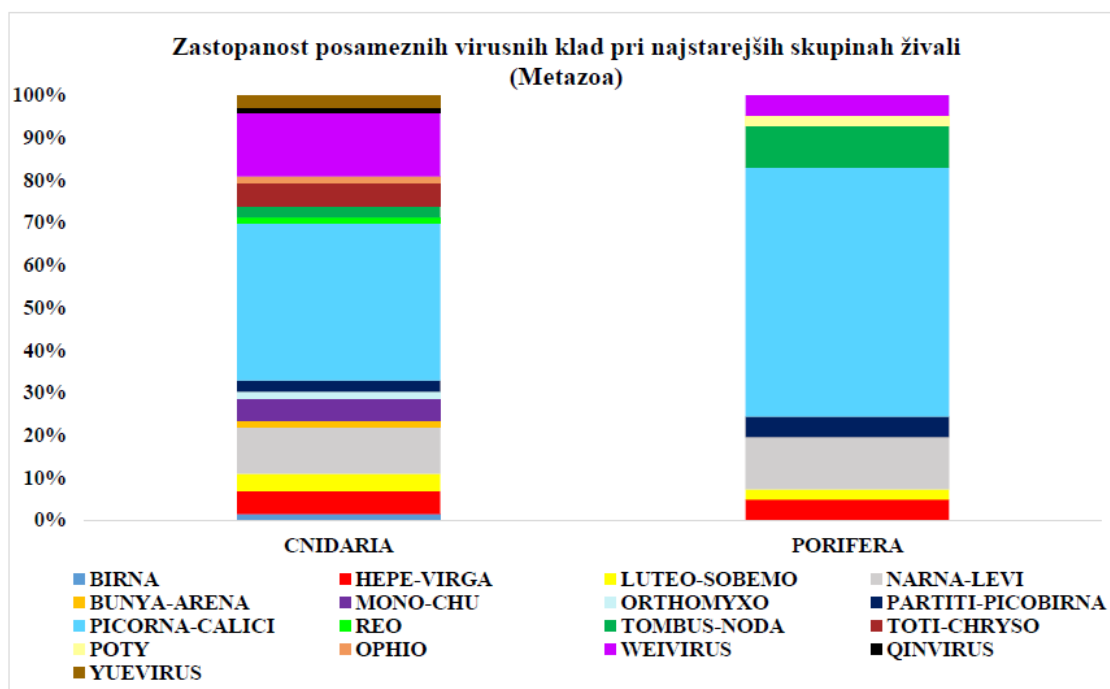
4.1.1.2 Rezultati, nastali v okviru doktorata

V okviru analize RNA virusov pri nevretenčarjih smo s homolognim iterativnim iskanjem pregledali skupno 70 transkriptomskih knjižnic (nastalih do 7. 2. 2018) najstarejših skupin živali – spužve (Porifera), ožigalkarji (Cnidaria) in rebrače (Ctenophora). Od tega je bilo 59 transkriptomskih knjižnic (do 7. 2. 2018) narejenih na vzorcih ožigalkarjev, devet na vzorcih spužev (Porifera) in dve na vzorcih rebrač (Ctenophora). Hkrati smo preverili 173 transkriptomskih knjižnic skupine Spiralia (Lophotrochozoa), od tega je bilo 106 narejenih na vzorcih mehkužcev (Mollusca), 36 na vzorcih ploščatih črvov (Platyhelminthes), 19 na vzorcih kolobarnikov (Annelida) in 11 na vzorcih Rotifera. Raki so zastopani s 187 transkriptomskimi knjižnicami, od tega jih 22 pripada skupini Copepoda (ceponožci), 29 enakonožcem (Isopoda), 50 skupini Decapoda (deseteronožci) in postranicam (Amphipoda) 78 knjižnic. Na vzorcih najstarejših skupin žuželk je bilo narejenih skupno samo 16 knjižnic, od tega 9 za skupino Collembola (skakači), dve za skupino Diplura (dvorepske), štiri za skupino Zygentoma (ščetinorepke) in ena za skupino Monocondylia. Na RNA vzorcih metuljev (Lepidoptera) je bilo narejenih 119 knjižnic, na vzorcih kožekrilcev (Hymenoptera) pa 264 transkriptomskih knjižnic. Pri tem velja poudariti, da je bilo na RNA vzorcih posameznih vrst organizmov narejenih več knjižnic, zato so nekatere vrste organizmov bolj pokrite.

V okviru genomskih knjižnic so najstarejše skupine živali zajemale skupno samo 19 knjižnic, od tega jih je 15 pripadalo ožigalkarjem (Cnidaria), ena spužvam (Porifera), dve rebračam (Ctenophora) in ena skupini Placozoa. Skupina Spiralia (Lophotrochozoa) je zajemala (do 7. 2. 2018) 60 genomskih knjižnic, od tega je 19 knjižnic pripadalo mehkužcem, 30 ploskim črvom, šest skupini Annelida in dve skupini Rotifera. Raki so v času analize knjižnic zajemali 23 knjižnic, deset od tega jih je pripadalo ceponožcem (Copepoda), dve enakonožcem (Isopoda), štiri deseteronožcem (Decapoda) in dve postranicam (Amphipoda). Najstarejše skupine žuželk so bile zastopane s štirimi genomskimi knjižnicami, od tega so tri pripadale skupini Collembola in ena skupini Diplura. Pri kožekrilcih (Hymenoptera) je bilo narejenih 168 knjižnic, na vzorcih metuljev (Lepidoptera) pa 182 genomskih knjižnic.

V okviru najstarejših skupin živali smo torej pregledali 88 transkriptomskih in genomskih knjižnic. Od tega je 74 knjižnic (83 %) pripadalo ožigalkarjem. Pri ožigalkarjih smo našli 73 novih divergentnih RNA virusov, kar pomeni v povprečju en virus na knjižnico. Skupina spužev (Porifera) je bila zastopana z desetimi knjižnicami. Skupno smo našli pri spužvah 41 virusov, kar pomeni v povprečju štirje virusi na knjižnico. Rebrače so bile zastopane samo s štirimi knjižnicami, našli smo samo tri viruse, zato jih v kasnejšo statistično analizo nismo vključili. Če gledamo zastopanost posameznih virusnih klad,

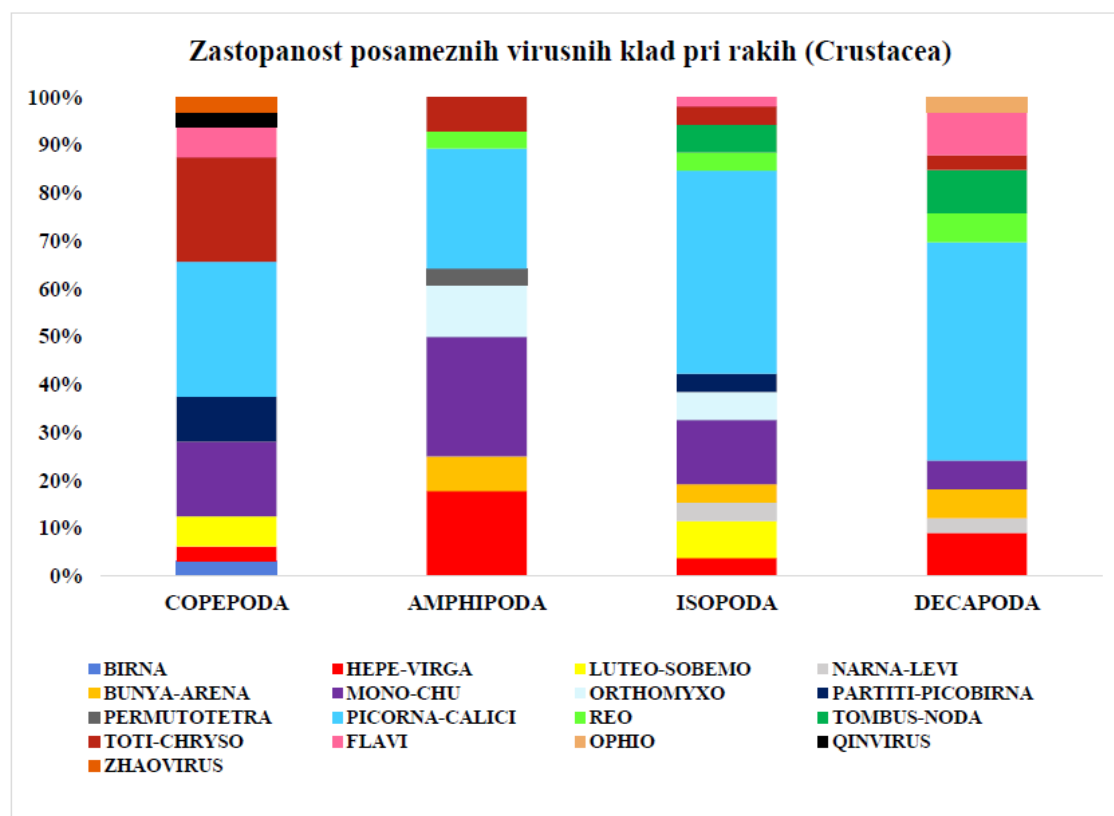
opazimo pri ožigalkarjih bistveno večji nabor prisotnih virusnih klad, kar je razumljivo glede na to, da smo pregledali sedemkrat več transkriptomskih in genomskih knjižnic kot pri spužvah (Porifera). Pri tem po zastopanosti daleč najbolj prednjači skupina Picorna-Calici, ki ji sledita skupini Narna-Levi in Weivirus. Spužve prednjačijo po bogatem številu raznolikih virusov, ki jih najdemo v knjižnicah (v povprečju štiri na knjižnico). Upoštevati pa moramo dejstvo, da so od devetih transkriptomskih knjižnic pri spužvah kar štiri pripadale organizmu *Halisarca caerulea*, kar pomeni bistveno večjo pokritost knjižnic omenjenega organizma. Pri sedemkrat manjšem številu knjižnic pa vseeno najdemo pester nabor raznolikih RNA virusov v njihovih transkriptomih, ki je sicer bistveno manjši od nabora raznolikih virusov, najdenih pri skupini ožigalkarjev (Cnidaria). Tudi pri spužvah sta bili najbolj zastopani virusni kladi Picorna-Calici in Narna-Levi. Klade Picorna-Calici, Narna-Levi, Partiti-Picobirna, Luteo-Sobemo in Hepe-Virga so zastopane v približno enakem deležu pri ožigalkarjih in pri spužvah (Porifera). Delež najdenih virusov, ki pripadajo kladi Tombus-Noda, je pri spužvah večji kot pri ožigalkarjih, nasprotno pa je delež weivirusov pri spužvah manjši kot pri ožigalkarjih. V knjižnicah spužev (Porifera) je bila najdena skupina Potyviridae, ki jih pri ožigalkarjih (Cnidaria) nismo našli. Nasprotno pa so bile v knjižnicah ožigalkarjev najdene klade Yuevirus, Qinvirus, Ophio, Toti-Chryso, Reo, Orthomyxo, Mono-Chu, Bunya-Arena in Birna, ki pri spužvah niso bile najdene. Vseeno je profil zastopanosti pri ožigalkarjih in spužvah zaradi unikatnosti težje direktno umestiti v kontekst katere koli od skupin, ki so jih analizirali Shi in sod. (2016) (Slika 4.1).



Slika 4.1: Zastopanost posameznih virusnih klad pri najstarejših skupinah živali (Metazoa).

Pri rakih smo skupno analizirali 196 genomskih in transkriptomskih knjižnic (do 7. 2. 2018). Skupina ceponožcev (Copepoda) je bila v času analize zastopana z 32 genomskimi in transkriptomskimi knjižnicami, skupina enakonožcev (Isopoda) z 31, skupina Decapoda s 54 knjižnicami in skupina postranic (Amphipoda) z 80 genomskimi in transkriptomskimi knjižnicami. Pri ceponožcih (Copepoda), postranicah (Amphipoda) in deseteronožcih (Decapoda) smo našli okrog 30 virusov (en virus na knjižnico pri ceponožcih, 0,5 virusa na knjižnico pri postranicah in 0,6 virusa na knjižnico pri deseteronožcih), medtem ko smo pri enakonožcih (Isopoda) našli okrog 50 virusov (1,7 virusa na knjižnico). Tukaj moramo poudariti, da pri enakonožcih (Isopoda) 22 transkriptomskih knjižnic od skupno 29 transkriptomskih in dveh genomskih pripada rodu *Proasellus*. Pri postranicah (Amphipoda) od skupno 78 transkriptomskih in dveh genomskih kar 15 transkriptomskih baz pripada rodu *Eulimnogammarus*. Kljub dobri pokritosti transkriptoma omenjenih rodov v podatkovnih bazah to hkrati pomeni manjšo vrstno raznolikost enakonožcev in

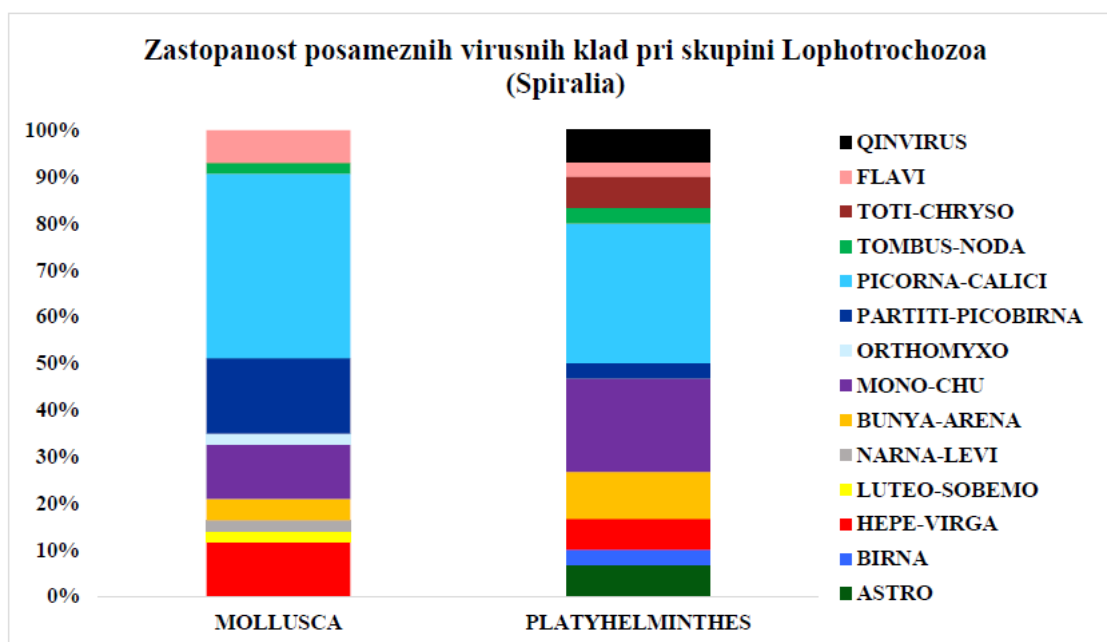
postranic, katerih vzorec transkriptoma se nahaja v podatkovnih bazah. Kljub razliki v pokritosti s strani genomskih in transkriptomskih knjižnic pa pri vseh skupinah najdemo primerljiv nabor prisotnih virusnih skupin. Klade Picorna-Calici, Mono-Chu, Hepe-Virga in Toti-Chryso so bile najdene pri vseh skupinah rakov, vendar je klada Toti-Chryso prisotna v večjem deležu pri ceponožcih (Copepoda), kladi Mono-Chu in Hepe-Virga pa pri postranicah (Amphipoda). Pri vseh skupinah rakov po zastopanosti izrazito prednjači virusna klada Picorna-Calici. V primeru skupine Copepoda ji sledita virusni kladi Mono-Chu in Toti-Chryso. V primeru skupine Amphipoda ji sledita kladi Mono-Chu in Hepe-Virga. V primeru skupine Isopoda ji sledita kladi Mono-Chu in Luteo-Sobemo in v primeru skupine Decapoda ji po približno enakomerni zastopanosti sledijo kar tri virusne klade: Flavi, Tombus-Noda in Hepe-Virga (Slika 4.2). Če naše rezultate primerjamo z analizo, ki so jo naredili Shi in sod. (2016a), opazimo pri njih bistveno večjo zastopanost skupine Narna-Levi, Partiti-Picobirna in Tombus-Noda. Poleg tega se pri skupini Shi in sod. (2016a) pojavlja bistveno večja razlika v vzorcu zastopanosti virusnih klad med posameznimi razredi. Samo razred Maxillopoda je praktično enakomerno zastopan z virusnimi kladami Tombus-Noda, Picorna-Calici, Partiti-Picobirna, Mono-Chu, Bunya-Arena in Hepe-Virga. Pri ostalih dveh razredih izrazito prednjači klada Picorna-Calici. V našem primeru se lahko opazi podoben vzorec zastopanosti virusnih klad med skupinami Amphipoda, Copepoda, Isopoda in Decapoda. V okviru naše raziskave sta kladi Hepe-Virga in Toti-Chryso bistveno bolj zastopani kot pri Shi in sod. (2016a). Poudariti moramo, da smo v raziskavo vključili petkrat večje število knjižnic kot Shi in sod. (2016a) (Slika 4.2 in Dodatek A; A.1, Slika A.2).



Slika 4.2: Zastopanost posameznih virusnih klad pri rakih (Crustacea).

Pri skupini Spiralia (Lophotrochozoa) smo skupno pregledali okrog 225 genomskih in transkriptomskih knjižnic, kar je ponovno petkrat več kot v primeru Shi in sod. (2016a). Od tega je 125 knjižnic pripadalo skupini mehkužcev (Mollusca), 66 skupini ploskih črvov (Platyhelminthes), 21 skupini Annelida in 13 skupini Rotifera. Pri skupini mehkužcev od 106 transkriptomskih knjižnic 10 transkriptomskih knjižnic pripada rodu Conus in devet rodu Aplysia. Pri skupini ploskih črvov (Platyhelminthes) od 36 transkriptomskih knjižnic jih kar 10 pripada rodu Schmidtea, kar pomeni relativno slabšo zastopanost in raznolikost vrst v transkriptomskih knjižnicah ploskih črvov (Platyhelminthes) v primerjavi z

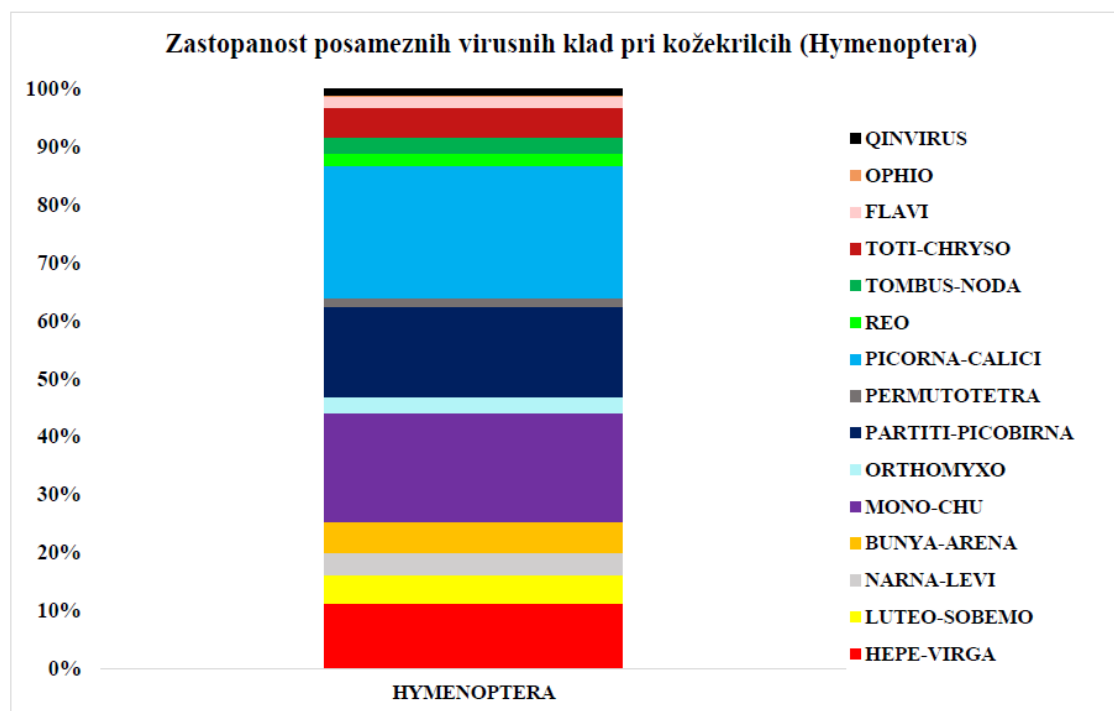
mehkužci. Pri skupini mehkužcev (Mollusca) smo našli okrog 40 novih RNA virusov, pri skupini ploskih črvov (Platyhelminthes) pa okrog 30 virusov. Pri vsaki od skupin Annelida in Rotifera smo našli en virus, ki je pripadal kladi Narna-Levi, in en virus, ki je pripadal kladi Tombus-Noda, zato skupin Annelida in Rotifera nismo vključevali v statistično analizo. Kljub dvakratni razliki med številom knjižnic, s katerimi so zastopani mehkužci in ploski črvi, najdemo v njunih vzorcih zastopanosti virusnih klad popolnoma primerljivo pester nabor prisotnih virusnih skupin. V obeh primerih po zastopanosti izrazito prednjači klada Picorna-Calici. Druge glavne klade po zastopanosti so še Mono-Chu, Partiti-Picobirna, Bunya-Arena in Hepe-Virga z rahlimi razlikami v deležu prisotnosti med mehkužci (Mollusca) in ploskimi črvi (Platyhelminthes). Kladi Partiti-Picobirna in Hepe-Virga sta bolj zastopani pri mehkužcih (Mollusca), medtem ko sta kladi Mono-Chu in Bunya-Arena bolj zastopani pri ploščatih črvih (Platyhelminthes). Skupina Shi in sod. (2016) je našla samo nekaj primerov virusov, prisotnih pri ploščatih črvih, zato jih nismo mogli vključiti v analizo za direktno primerjavo zastopanosti. Se pa vidi izrazita razlika v profilu zastopanosti virusnih klad med našimi rezultati in rezultati Shi in sod. (2016a). Tako v raziskavi Shi in sod. (2016a) kot v našem primeru pri mehkužcih (Mollusca) prednjači virusna klada Picorna-Calici. Vendar so v primeru raziskave Shi in sod. (2016a) bistveno bolj zastopane virusne klade Narna-Levi, Tombus-Noda in Luteo-Sobemo. Klade Mono-Chu pri mehkužcih v raziskavi Shi in sod. (2016a) sploh ni bilo prisotne. Naša raziskava je pokazala bistveno večjo zastopanost klad Partiti-Picobirna, Bunya-Arena in Hepe-Virga. Ravno tako je v naši raziskavi pri mehkužcih bogato zastopana skupina Mono-Chu (Slika 4.3 in Dodatek A; A.1, Slika A.3).



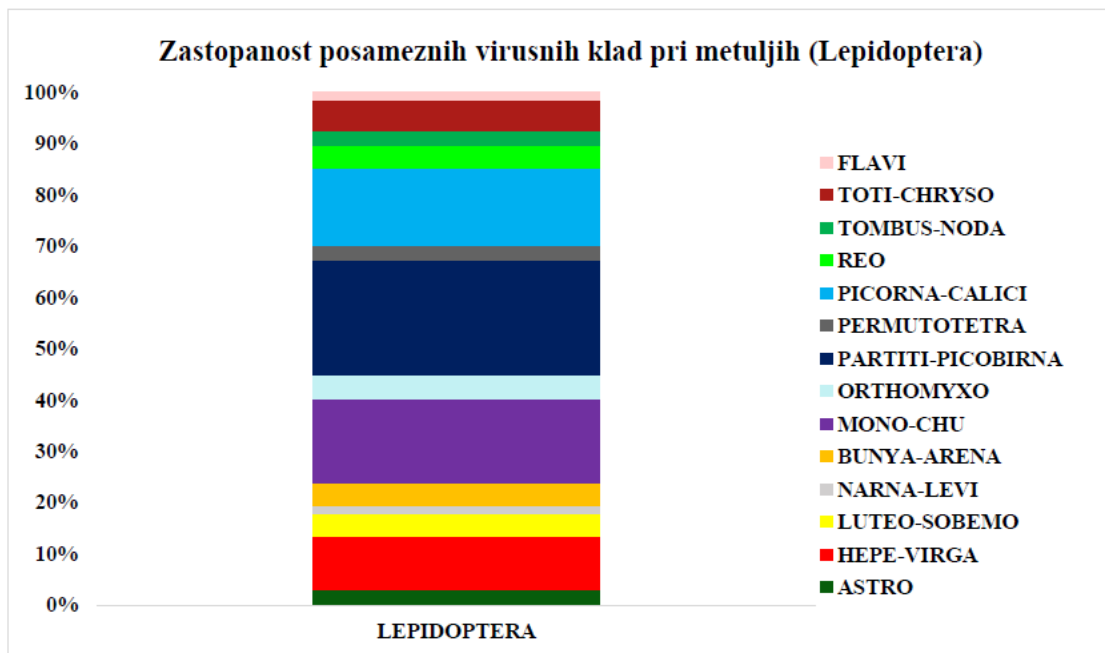
Slika 4.3: Zastopanost posameznih virusnih klad pri skupini Lophotrochozoa (Spiralia).

V okviru žuželk smo se usmerili na ključne skupine, ki so bile pri raziskavi Shi in sod. (2016a) slabo pokrite: Hymenoptera (kožekrilci), ki se v raziskavi Shi in sod. (2016a) pojavi samo v kontekstu mešanih knjižnic, Lepidoptera (metulji), ki je bila v primeru Shi in sod. (2016a) zastopana: 1) z eno samo knjižnico na nivoju reda, ki je bila revna v kontekstu zastopanosti virusnih klad in 2) v okviru mešanih knjižnic. Hkrati smo se usmerili na najstarejše skupine žuželk: Collembola (skakači), Diplura (dvorepke), Zygentoma (ščetinorepke) in Monocondylia, ki doslej še niso bile analizirane. Metulji (Lepidoptera) in kožekrilci (Hymenoptera) so bili v genomskih in transkriptomskih bazah podatkov zastopani z največjim številom knjižnic. Lepidoptera zajema skupno 301 knjižnico in Hymenoptera skupno 432 genomskih in transkriptomskih knjižnic z razliko, da je Lepidoptera bolj zastopana z genomskimi knjižnicami, Hymenoptera pa s transkriptomskimi. Obe skupini sta tudi v smislu raznolikih organizmov bogato zastopani v knjižnicah, saj ni bilo veliko RNA knjižnic narejenih na isti vrsti organizma. Pri skupini

Hymenoptera smo našli 388 virusov, medtem ko smo pri skupini Lepidoptera našli 67 virusov, kar je velika razlika glede na razliko v številu podatkovnih baz, s katerimi sta skupini zastopani. Kljub temu smo pri obeh skupinah našli izrazito raznolik in hkrati unikaten profil prisotnih virusnih klad. Pri skupini Lepidoptera so glavne virusne klade po zastopanosti: Picorna-Calici, Mono-Chu, Partiti-Picobirna in Hepe-Virga. V nekoliko manjšem deležu so prisotne tudi klade Flavi, Toti-Chryso, Tombus-Noda, Reo, Permutotetra, Orthomyxo, Bunya-Arena, Luteo-Sobemo, Narna-Levi in Astro, kar je skupaj 14 od 25 definiranih klad. V raziskavi Shi in sod. (2016a) so bile najdene prisotne samo štiri: Reo, Mono-Chu, Bunya-Arena in Luteo-Sobemo. Pri skupini Hymenoptera so glede na naše rezultate glavne RNA virusne klade po zastopanosti: Picorna-Calici, Mono-Chu, Partiti-Picobirna in Hepe-Virga, ki so iste, kot pri Lepidoptera. Poleg omenjenih smo našli prisotne tudi: Qinvirus, Ophio, Flavi, Toti-Chryso, Tombus-Noda, Reo, Permutotetra, Orthomyxo, Bunya-Arena, Narna-Levi in Luteo-Sobemo. Če povzamemo rezultate, lahko ugotovimo, da je celoten vzorec profila zastopanosti virusnih klad pri skupini Hymenoptera unikaten v smislu deležev in same prisotnosti virusnih klad. Na novo pa smo v raziskavi pri kožekrilcih (Hymenoptera) našli prisotne virusne klade Ophio in Qinvirus, ki doslej pri omenjeni skupini žuželk niso bile najdene. Profil skupine Hymenoptera glede na naše rezultate je najbolj primerljiv s profilom skupine Odonata pri raziskavi Shi in sod. (2016a) – virusno najbolj bogati kladi v raziskavi, kjer so glavne klade, ravno tako kot pri nas, Picorna-Calici, Mono-Chu in Partiti-Picobirna. Vendar je pri redu Odonata v raziskavi Shi in sod. (2016a) prisoten nekoliko večji delež skupin Tombus-Noda, Reo, Bunya-Arena in Luteo-Sobemo kot pri naših raziskavah. Pri najstarejših skupinah žuželk smo največ virusov našli prisotnih pri kladi Hepe-Virga (5) in Picorna-Calici (7). Ostale virusne klade so bile zastopane v glavnem z enim virusnim predstavnikom (Narna-Levi, Mono-Chu, Reo, Toti-Chryso in Qinvirus) ali dvema virusnima predstavnikoma (Orthomyxo in Tombus-Noda). Zato najstarejše skupine žuželk niso bile primerne za vključevanje direktno v statistično analizo (Sliki 4.4 in 4.5).



Slika 4.4: Zastopanost virusnih klad pri kožekrilcih (Hymenoptera).



Slika 4.5: Zastopanost posameznih virusnih klad pri metuljih (Lepidoptera).

Skupno smo torej našli 815 virusov, ki so spadali v 22 od 25 definiranih virusnih klad. Pri vseh skupinah nevretenčarjev je bila najbolj bogato zastopana virusna klada Picorna-Calici, zato smo se odločili za nadaljnjo obravnavo glavnih skupin klade v statistični analizi.

4.1.2 Analiza zastopanosti virusnih skupin znotraj virusne klade Picorna-Calici

Analiza zastopanosti virusnih skupin znotraj klade Picorna-Calici je bila ravno tako narejena tako na rezultatih skupine Shi in sod. (2016a) ter na raziskanih skupinah nevretenčarjev v okviru doktorata.

Pri analizi zastopanosti v okviru rezultatov Shi in sod. (2016a) smo znotraj klade Picorna-Calici upoštevali vse knjižnice, tudi mešane, da smo predstavili sumaričen rezultat vseh pikornavirusov pri rakah in žuželkah. V primeru skupine Lophotrochozoa smo se osredotočili na skupino mehkužcev (Mollusca). Pikornaviruse smo razdelili na glavne skupine, kot jih je definirala skupina Shi in sod. (2016a), in prešteli število virusov, najdenih v vseh knjižnicah žuželk (Insecta), rakov (Crustacea) in mehkužcev (Mollusca).

Pri analizi naših rezultatov v okviru doktorata smo klado Picorna-Calici razdelili na skupine, kot jih je definirala skupina Shi in sod. (2016), in prešteli število pikornavirusov, ki so bili v okviru filogenetske analize uvrščeni v posamezno skupino pikornavirusov.

4.1.2.1 Rezultati Shi in sod. (2016a)

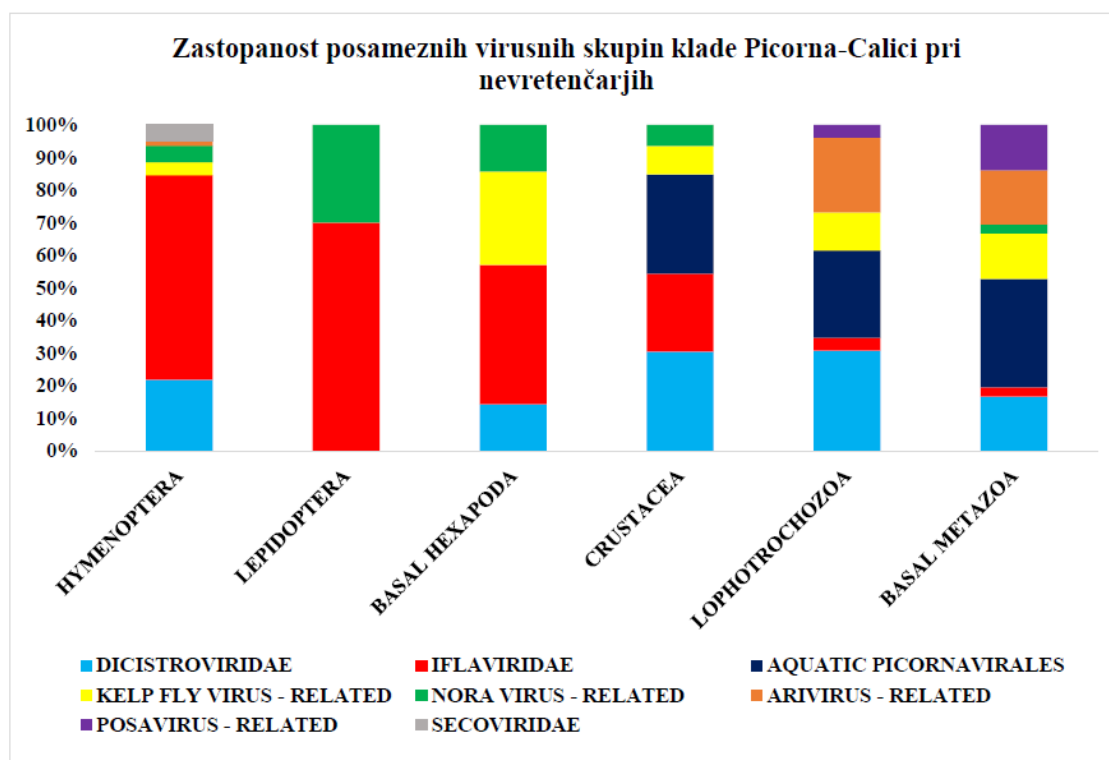
Klada Picorna-Calici je ena izmed 25 definiranih virusnih klad, ki jih je raziskovalna skupina Shi in sod. (2016a) razdelila na podskupine: Dicistroviridae, Iflaviridae, Secoviridae, Picornaviridae, »Aquatic picornavirales«, Kelp fly virusu podobne (angl. »*Kelp fly virus-related*«), Nora virusu podobne (angl. »*Nora virus-related*«) in Caliciviridae. Skupino »Aquatic picornavirales« je skupina Shi in sod. (2016a) oblikovala na podlagi združevanja po ICTV (angl. »*International Committee on Taxonomy of Viruses*«) definiranih skupin Marnaviridae, Bacillarnavirus in Labrynavirus. Skupini Kelp fly virusu podobnih (angl. »*Kelp fly virus-related*«) in Nora virusu podobnih (angl. »*Nora virus-related*«) virusov je skupina Shi in sod. (2016a) definirala na podlagi filogenetske analize, pri kateri omenjeni Kelp fly virusu podobni virusi (angl. »*Kelp fly virus-related*«) oblikujejo skupino s Kelp fly virusom in Nora virusu podobni virusi (angl. »*Nora virus-related*«) oblikujejo skupino z Nora virusom. Poleg novo definiranih skupin pikornavirusov so se v filogenetski analizi oblikovale nove skupine pikornavirusov, ki jih Shi in sod. (2016a) niso poimenovali, so pa v naši filogenetski analizi oblikovale skupino s Posavirusom kot Posavirusom podobni virusi (angl. »*Posavirus-related*«) in Arivirusom kot Arivirusom podobni virusi (angl. »*Arivirus-*

related). Za vsako od skupin pikornavirusov smo prešteli število virusov v knjižnicah žuželk (Insecta), rakov (Crustacea) in skupine Mollusca. Pri skupinah žuželk in rakov smo upoštevali tudi vse mešane knjižnice in viruse, ki so bili najdeni v več knjižnicah hkrati. Prešteli smo vse viruse, ki so bili v raziskavi Shi in sod. (2016a) uvrščeni v omenjene skupine pikornavirusov. Skupina žuželk je vključevala 21 knjižnic na nivoju reda, štiri mešane knjižnice, ki so vključevale RNA vzorce žuželk, ki pripadajo različnim redovom (insect ZJ, CC, WHCCII), in štiri mešane knjižnice, ki so vključevale RNA vzorce členonožcev (rakov, pajkovcev in žuželk; arthropodmix, SHWCII, SHWC, WHCC). Skupina rakov je vključevala 20 knjižnic na nivoju reda oz. razreda, dve mešani knjižnici, ki sta vključevali različne razrede rakov (WLJQ in shrimp), in štiri mešane knjižnice, ki so vključevale RNA vzorce členonožcev (rakov, pajkovcev in žuželk; arthropodmix, SHWCII, SHWC, WHCC). Skupina Spiralia (Lophotrochozoa) je vključevala 21 na nivoju debel mehkužcev (Mollusca), ploščatih črvov (Platyhelminthes), Annelida in Rotifera. V raziskavi se profil zastopanosti skupin pikornavirusov pri žuželkah bistveno razlikuje od profila zastopanosti pri rakah (Crustacea) in mehkužcih Mollusca. Pri skupini žuželk so bile najdene skupine Posavirusom podobnih virusov, Iflavirusov, Kelp fly virusu podobnih virusov, Nora virusu podobnih virusov in Dicistroviridae, pri čemer po deležu prisotnosti prednjačijo Iflavirusi. Sledijo jim Kelp fly virusu podobni virusi in nato skupina Dicistroviridae. Pri žuželkah ni bilo najdenih »aquatic picornavirales« in Arivirusu podobnih virusov. Pri rakah in mehkužcih, ki so morski organizmi, lahko najdemo podoben profil zastopanosti posameznih skupin pikornavirusov. V obeh primerih so prisotne skupine Posavirusu podobnih virusov, Arivirusu podobnih virusov, Nora virusu podobnih virusov, Kelp fly virusu podobnih virusov, skupine »Aquatic picornavirales« in Dicistroviridae. Dve kladi sta glavni glede zastopanosti: »Aquatic picornavirales« in Dicistroviridae, ostale so pri rakah in mehkužcih kot morskih organizmih prisotne v manjšini. Pri rakah sta nekoliko bolj zastopani skupini Nora virusu podobnih virusov in Kelp fly virusu podobnih virusov. Pri mehkužcih pa je nekoliko bolj kot pri rakah glede na raziskave Shi in sod. (2016a) zastopana skupina Posavirusu podobnih virusov. Iz filogenetske raziskave Shi in sod. (2016a) lahko zaključimo, da je skupina Iflaviridae, ki je evolucijsko relativno mlada skupina, omejena na kopenske organizme, medtem ko je skupina »Aquatic picornavirales« omejena praktično samo na morske nevretenčarje (rake, mehkužce). Ostale skupine so razširjene čez celotne nevretenčarje, čeprav so Kelp fly virusu podobni virusi nekoliko bolj pogosti pri kopenskih nevretenčarjih (**Dodatek A; A.1, Slika A.4**).

4.1.2.2 Rezultati, nastali v okviru doktorata

V filogenetski analizi smo pikornaviruse razdelili glede na skupine, kot jih je poimenovala raziskava skupine Shi in sod. (2016a) na Dicistroviridae, Iflaviridae, »Aquatic picornavirales«, Kelp fly virusu podobne (angl. »*Kelp fly virus-related*«), Nora virusu podobne (angl. »*Nora virus-related*«), Posavirusu podobne (angl. »*Posavirus-related*«), Arivirusu podobne (angl. »*Arivirus-related*«), Secoviridae in Caliciviridae. Skupine Picornaviridae kot vretenčarske skupine nismo vključevali v filogenetsko analizo. Prešteli smo vse viruse, ki so bili uvrščeni v omenjene skupine pikornavirusov. Prišli smo do drugačnih ugotovitev, kot jih je prikazala skupina Shi in sod. (2016a). Pri kožekrilcih (Hymenoptera), ki so zajemali 432 genomskih in transkriptomskih knjižnic, smo našli prisotne skupine: Secoviridae, Arivirusu podobne viruse, Kelp fly virusu podobne viruse, Nora virusu podobne viruse, Iflaviridae in Dicistroviridae. Iflaviridae in Dicistroviridae sta bili glavni kladi glede zastopanosti, ostale skupine pikornavirusov so bile najdene v manjšini. Pri kožekrilcih nismo našli Posavirusu podobnih virusov, smo pa našli skupino Secoviridae in Arivirusu podobne pikornaviruse, ki v raziskavi Shi in sod. (2016a) niso bili prisotni pri žuželkah. Pri metuljih (Lepidoptera) smo v 301 knjižnicah našli prisotni samo dve pikornavirusni skupini: Iflaviridae in Nora virusu podobni virusi. Pri najstarejši skupini žuželk, ki je skupno zajemala samo 20 genomskih in transkriptomskih knjižnic, smo našli prisotne štiri virusne klade: Nora virusu podobne viruse, Kelp fly virusu podobne viruse, Iflaviridae in Dicistroviridae, pri čemer so bili Iflaviridae in Kelp fly virusu podobni virusi nekoliko bolj zastopani kot Dicistroviridae in Nora virusu podobni virusi. Pri rakah (Crustacea) smo v 196 knjižnicah našli prisotnih šest klad: Nora virusu podobne viruse, Kelp fly virusu podobne viruse, »Aquatic picornavirales«, Iflaviridae in Dicistroviridae. Najbolj izstopajoče klade so bile pri rakah »Aquatic picornavirales«, Iflaviridae in Dicistroviridae. V nasprotju s skupino Shi in sod. (2016a), ki pri rakah ni našla prisotnih iflavirusov, je bila pri naši raziskavi ena najbolj zastopanih skupin RNA virusov v okviru

rakov. Hkrati v naši raziskavi pri rakah nismo našli Posavirusu in Arivirusu podobnih pikornavirusov. Pri skupini Lophotrochozoa smo v 225 knjižnicah našli pikornavirusne skupine: Posavirusu podobni virusi, Arivirusu podobni virusi, Kelp fly virusu podobni virusi, »Aquatic picornavirales«, Iflaviridae in Dicistroviridae, pri čemer sta bili »Aquatic picornavirales« in Dicistroviridae izstopajoči virusni kladi. Z vidika deležev je bila pri skupini Spiralia bolj zastopana klada Kelp fly virusu in Arivirusu podobnih pikornavirusov kot pri raziskavi Shi in sod. (2016a). V naši raziskavi nismo našli Posavirusu podobnih virusov pri skupini Spiralia, smo pa v nasprotju z raziskavo Shi in sod. (2016a) našli iflaviruse. Najstarejše skupine živali (Metazoa) (angl. »*basal Metazoa*«), ki so obsegale 88 genomskih in transkriptomskih knjižnic, imajo najbolj bogat profil zastopanosti pikornavirusnih skupin. Našli smo prisotnih kar sedem skupin: Posavirusu podobni virusi, Arivirusu podobni virusi, Nora virusu podobni virusi, Kelp fly virusu podobni virusi, »Aquatic picornavirales«, Iflaviridae in Dicistroviridae. Po zastopanosti najbolj bogata pikornavirusna skupina je bila »Aquatic picornavirales«, ki ji sledijo Dicistroviridae, Arivirusu podobni virusi, Kelp fly virusu podobni virusi in Dicistroviridae. Ostale virusne skupine so bile prisotne v manjšini. V splošnem lahko glede na naše rezultate trdimo, da je skupina »Aquatic picornavirales« omejena samo na vodne nevretenčarje (rake, skupino Spiralia – mehkužce, ploščate črve, skupini Annelida in Rotifera ter najstarejše skupine živali (Metazoa) – spužve (Porifera), ožigalkarje (Cnidaria) in rebrače (Ctenophora)), medtem ko so ostale skupine pikornavirusov razširjene pri vseh skupinah nevretenčarjev (tako vodnih kot kopenskih). Med njimi je tudi skupina Iflaviridae, ki smo jo našli prisotno tako pri kopenskih kot pri vodnih nevretenčarjih. Skupina je bila sicer bistveno bolj razširjena pri kopenskih nevretenčarjih (žuželkah) kot pri vodnih nevretenčarjih. Pri vodnih nevretenčarjih je bila najbolj zastopana pri rakah. Pri skupini Lophotrochozoa (Spiralia) in pri najstarejših skupinah živali je bila prisotna v manjšini. Rezultati zastopanosti iflavirusov so drugačni, kot je bilo pokazano v analizi Shi in sod. (2016a). Skupina je namreč iflaviruse našla le pri kopenskih nevretenčarjih. Glede na naše rezultate so bili dicistrovirusi enakomerno razširjeni in med glavnimi virusnimi kladami po zastopanosti tako pri kopenskih kot pri vodnih nevretenčarjih. Podobno razširjenost dicistrovirusov so prikazali tudi Shi in sod. (2016a). Noravirusu podobni virusi so sicer razširjeni pri vseh skupinah nevretenčarjev, vendar so nekoliko bolj značilni za kopenske nevretenčarje. Arivirusu podobni virusi in Posavirusu podobni virusi so bolj značilni za vodne nevretenčarje. V naših analizah smo jih našli skoraj izključno pri vodnih nevretenčarjih. Arivirusu podobne viruse smo v manjšini našli pri skupni Hymenoptera. Skupina Shi in sod. (2016a) je Posavirusu podobne viruse našla pri vseh skupinah nevretenčarjev, s tem da so bili pri rakah v manjšini, Arivirusu podobnih virusov pa ni zasledila pri žuželkah. Kelp fly virusu podobni virusi so glede na naše rezultate razširjeni tako pri kopenskih kot pri vodnih nevretenčarjih, kar je hkrati v skladu z rezultati Shi in sod. (2016a) (Slika 4.6).



Slika 4.6: Zastopanost posameznih virusnih skupin klade Picorna-Calici pri nevretenčarjih. Angleški izraz: related – podoben, npr. posavirus-related; posavirusu podobni virusi.

4.2 Endogenizirani RNA virusi v genomih nevretenčarjev

Za iskanje endogeniziranih RNA virusov je bilo v prvi fazi kot iskalni niz uporabljeno proteinsko zaporedje RdRp definiranih virusnih klad oz. definiranih virusnih skupin. Z iskalnim nizom smo iterativno iskali homologna zaporedja v genomskih podatkovnih bazah raziskovanih skupin nevretenčarjev. V okviru iterativnega iskanja smo v drugi stopnji kot iskalni niz uporabili proteinska zaporedja RdRp, ki smo jih pridobili kot zadetke v genomskih in transkriptomskih podatkovnih bazah. Ravno tako smo kot iskalni niz uporabili tudi proteinska zaporedja drugih proteinov, ki so razširjeni samo pri določeni skupini RNA virusov. Zaporedja RdRp endogeniziranih RNA virusov so bila nato uporabljena v filogenetski analizi.

Po iterativnem iskanju homolognih zaporedij proteinskemu zaporedju za RdRp za vsako od definiranih virusnih klad v prevedenih nukleotidnih genomskih podatkovnih bazah nevretenčarjev smo večinoma našli samo fragmente genomov endogeniziranih RNA virusov ali samo fragmente nukleotidnega zaporedja RdRp. Celotni ohranjeni ORF so bili redkejši (kot primer lahko navedemo endogeniziran celoten L segment pri polžu *Colubraria reticulata*). Skupno smo našli 12 virusnih klad od 25 definiranih, endogeniziranih v genomih nevretenčarjev. Tabela 4.1 prikazuje prisotnost oz. odsotnost endogeniziranih genomov RNA virusnih klad v genomih raziskanih skupin nevretenčarjev (**Tabela 4.1**). Med klade, ki jih v kontekstu endogenizacije velja bolje poudariti, so uvrščeni: Mono-Chu, Partiti-Picobirna, Toti-Chryso, Flavi in Weivirus. Odkrili smo, da je klada Mono-Chu endogenizirana pri vseh analiziranih skupinah nevretenčarjev. Našli smo tudi številne endogenizirane viruse, ki spadajo v skupino Partiti-Picobirna in Toti-Chryso, kljub temu da kot dsRNA virusi nimajo mehanizma razmnoževanja, povezanega z jedrom gostiteljske celice. V literaturi doslej še ni bilo moč zaslediti prisotnih endogeniziranih virusov klade Flaviviridae in Weivirus pri nevretenčarjih. Pri kladi Weivirus smo našli 12 weivirusov, endogeniziranih v genomih najstarejših skupin evkariontov Alveolata, in sicer v organizmu *Symbiodinium*. Gre za organizem, ki živi v sožitju s koralami (Cnidaria).

Zaporedje RdRp sicer predstavlja jedro vsakega virusa, našli pa smo endogenizirane tudi druge virusne ORF-je.

Tabela 4.1: Na novo odkriti endogenizirani RNA virusi v genomih nevretenčarjev. Črna pika pomeni prisotnost, bela pika pomeni odsotnost posameznih virusnih klad.

VIRUSNA KLADA	Hymenoptera	Lepidoptera	Najstarejše skupine žuželk	Spiralia	Crustacea	Najstarejše skupine živali
ASTRO	○	○	○	○	○	○
BIRNA	○	○	○	○	○	○
HEPE-VIRGA	○	●	●	○	○	○
HYPO	○	○	○	○	○	○
LUTEO-SOBEMO	○	○	○	○	○	○
NARNA-LEVI	●	○	○	○	○	○
BUNYA-ARENA	○	○	○	●	○	○
MONO-CHU	●	●	●	●	●	●
ORTHOMYXO	●	●	●	●	○	○
NIDO	○	○	○	○	○	○
PARTITI-PICOBIRNA	●	●	○	●	○	○
PERMUTOTETRA	○	○	○	○	○	○
PICORNA-CALICI	○	○	○	○	○	○
REO	●	○	○	○	○	○
TOMBUS-NODA	○	●	○	○	○	○
TOTI-CHRYSO	●	●	○	○	●	○
QINVIRUS	●	○	○	●	●	○
YUEVIRUS	○	○	○	○	○	○
YANVIRUS	○	○	○	○	○	○
ZHAOVIRUS	○	○	○	○	○	○
WEIVIRUS	○	○	○	○	○	○
POTYVIRIDAE	○	○	○	○	○	○
FLAVIVIRIDAE	●	○	○	○	●	○
OPHIOVIRIDAE	○	○	○	○	○	●
CYSTOVIRIDAE	○	○	○	○	○	○

Shi in sod. (2016a) so analizirali endogenizirane RNA viruse pri členonožcih (Arthropoda) in ploščatih črvih (Platyhelminthes – Spiralia). Pri členonožcih so ugotovili prisotnost endogeniziranih RNA virusov, ki spadajo v virusne klade: Hepe-Virga, Luteo-Sobemo, Narna-Levi, Bunya-Arena, Mono-Chu, Orthomyxo, Nido, Partiti-Picobirna, Picorna-Calici, Reo, Tombus-Noda, Toti-Chryso in Qinvirus. Pri ploščatih črvih so našli endogenizirani samo RNA virusni kladi Bunya-Arena in Mono-Chu (**Tabela 4.2**).

Tabela 4.2: Odkrite endogenizirane RNA virusne klade pri členonožcih in ploščatih črvih v okviru analize Shi in sod. (2016a). Črna pika pomeni prisotnost, bela pika pomeni odsotnost posameznih virusnih klad.

VIRUSNA KLADA	Arthropoda	Spiralia; Platyhelminthes
ASTRO	○	○
BIRNA	○	○
HEPE-VIRGA	●	○
HYPO	○	○
LUTEO-SOBEMO	●	○
NARNA-LEVI	●	○
BUNYA-ARENA	●	●
MONO-CHU	●	●
ORTHOMYXO	●	○
NIDO	●	○
PARTITI-PICOBIRNA	●	○
PERMUTOTETRA	○	○
PICORNA-CALICI	●	○
REO	●	○
TOMBUS-NODA	●	○
TOTI-CHRYSO	●	○
QINVIRUS	●	○
YUEVIRUS	○	○
YANVIRUS	○	○
ZHAOVIRUS	○	○
WEIVIRUS	○	○
POTYVIRIDAE	○	○
FLAVIVIRIDAE	○	○
OPHIOVIRIDAE	○	○
CYSTOVIRIDAE	○	○

Rezultati naše analize endogeniziranih RNA virusov so nekoliko drugačni od rezultatov analize Shi in sod. (2016a). Pri naši raziskavi sicer nismo odkrili prisotnih endogeniziranih RNA virusnih klad Luteo-Sobemo, Bunya-Arena, Nido in Picorna-Calici pri členonožcih (med katere sodijo raki, pajkovci, stonoge in žuželke), smo pa našli endogenizirano RNA virusno klado Flaviviridae. Pri ploščatih črvih smo poleg klad Bunya-Arena in Mono-Chu našli prisotne tudi endogenizirane RNA virusne klade: Orthomyxo, Partiti-Picobirna in Qinvirus, ki jih Shi in sod. (2016a) niso zasledili. Poudariti velja tudi, da doslej še nobena študija ni analizirala endogeniziranih RNA virusov pri najstarejših skupinah živali. Poleg klade Mono-Chu smo odkrili tudi endogenizirano RNA virusno klado Ophioviridae pri najstarejših skupinah živali. Ravno tako na novo odkriti endogenizirani Weivirusi pri simbiozih koral nakazujejo na možnost, da so endogenizirani Weivirusi dejansko povezani s koralami in so alveolati endogenizirane viruse prevzeli od koral. Gre za novo odkritje, ki doslej v literaturi še ni bilo obeleženo.

4.3 Filogenetska analiza novih RNA virusov in endogeniziranih RNA virusov pri raziskanih skupinah nevretenčarjev

Rezultati prikazujejo filogenetsko analizo novo odkritih RNA virusov v sklopu pred kratkim definiranih RNA virusnih klad (Shi in sod., 2016a). V analizo je bilo poleg proteinskega zaporedja RdRp virusov, odkritih v transkriptomskih in genomskih podatkovnih bazah nevretenčarjev, vključeno tudi proteinsko zaporedje RdRp tako klasificiranih kot neklasificiranih virusov (angl. »*floating viral genera*«). Kot glavni iskalni niz smo vzeli 300 aa dolgo proteinsko zaporedje RdRp za vsako izmed 25 definiranih virusnih klad s katalitično domeno znotraj motiva C. Omenjen del zaporedja je namreč dovolj ohranjen za ugotovitev sorodstvenih odnosov med virusi, hkrati pa dovolj divergenten za ločevanje med virusi na nivoju seva, skupine in klade. Z orodjem BLAST smo našli homologne RNA viruse v genomskih in transkriptomskih podatkovnih bazah nevretenčarjev. Novi odkriti RNA virusi so imeli lahko celotne, delne genome, lahko pa je bila znana le njihova RdRp. Našli smo torej celotne genome, delne genome ali samo RdRp RNA virusov. Dele proteinskega zaporedja RdRp smo poravnali s programom Clustal Omega (Sievers in Higgins, 2014) in s programom IQTREE (Nguyen in sod., 2015) izrisali filogenetska drevesa. Nove RNA viruse, vključene v filogenetsko analizo, smo uvrstili v posamezno virusno klado. Nove RNA viruse, ki spadajo v isto virusno klado in so bili pridobljeni iz podatkovnih baz določene skupine nevretenčarjev (kožekrilci, metulji, najstarejše skupine žuželk, raki, skupina Spiralia in najstarejše skupine živali), smo prikazali v istem filogenetskem drevesu. Večina filogenetskih dreves vsebuje večje število elementov. S tem je zmanjšana preglednost oz. berljivost slik v samem doktorskem delu. Slike filogenetskih dreves so zato priložene še v elektronski PDF verziji (CD).

4.3.1 Kopenski nevretenčarji

4.3.1.1 Kožekrilci (Hymenoptera)

Kožekrilci (Hymenoptera) so zastopani z največjim številom genomskih in transkriptomskih podatkovnih baz (432). Zajemajo podred Apocrita in več naddružin: Cephioidea, Orussoidea, Pamphilioidea, Siricoidea in druge. Podred Apocrita zajema dve glavni skupini: Aculeata in evolucijsko starejša Terebrantes. Skupina Aculeata zajema glavne naddružine: Apoidea, Chrysidoidea (evolucijsko starejša skupina znotraj skupine Aculeata), Pompiloidea, Sphecoidea in Vespoidea. Večina RNA virusov je bilo najdenih v genomskih in transkriptomskih knjižnicah skupine Apocrita. Zato smo v filogenetskem drevesu izpostavili naddružine Terebrantes in znotraj skupine Aculeata: Apoidea (čebele), Pompiloidea, Sphecoidea, Chrysidoidea. Skupino Apoidea smo označili z oranžno barvo, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno. Znotraj skupine Vespoidea smo mravlje (Formicidae) označili z rdečo, ostale predstavnike skupine Vespoidea pa s svetlo modro. RNA virusi, ki so bili najdeni v genomskih in transkriptomskih knjižnicah kožekrilcev, ki ne spadajo v skupino Apocrita, smo označili z barvo magenta. Največ RNA virusov je bilo najdenih pri mravljah (Formicidae), čebelah (Apoidea) in skupini Terebrantes. Omenjene skupine imajo torej najbolj bogato zastopanje virusnih klad. RNA virusi, ki so izvirali iz knjižnic skupine Chrysidoidea, so bili redki. V transkriptomskih in genomskih podatkovnih bazah kožekrilcev smo pridobili bodisi celotne genome RNA virusov (glede na trenutno znan interval dolžine genoma virusov, ki pripadajo določeni skupini), dele genoma virusov (npr. en segment ali en ORF, ki kodira zapis za poliprotein) bodisi samo zaporedje RdRp. Pri genomski analizi RNA virusov smo se osredotočili samo na celotne genome virusov. Pri segmentiranih virusih smo na podlagi terminalnih zaporedij poskusili pridobiti tudi zaporedja drugih segmentov.

4.3.1.1.1 Klada Bunya-Arena

Pri kladi Bunya-Arena smo našli pet RNA virusov v knjižnicah skupine Apoidea, štiri v knjižnicah Formicidae, sedem v knjižnicah Terebrantes ter enega v knjižnicah Sphecoidea. Klada Bunya-Arena je najbolj pogosta pri skupini Terebrantes. Virusi se nahajajo znotraj dveh večjih sestrskih skupin. Eno večjo skupino tvorijo skupaj s Hantavirusi, Tospovirusi, Emaravirusi in Orthobunyavirusi. Drugo večjo skupino tvorijo s Phlebovirusi in Tenuivirusi.

Nairovirusom podobni virusi oblikujejo ločeno skupino. Znotraj večje sestrške skupine, ki združuje Hantaviruse, Tospoviruse, Emaraviruse in Orthobunyaviruse, se vsi virusi združujejo znotraj iste skupine skupaj s »Seattle Prectang virusom« in »Ganda bee virusom«. Skupina, v katero se združujejo Tenuivirusu podobni virusi in Phlebovirusu podobni virusi, oblikuje dve sestrski skupini. Ena skupina predstavlja Wuhan spider virusu podobne viruse, v drugo se poleg Phlebovirusov in Tenuivirusov združujejo tudi Cumuto in Gouleako virusu podobni virusi. Naši bunyavirusi so enakomerno razporejeni med obema sestrskima skupinama. Pri skupini Shi in sod. (2016a) je situacija nekoliko drugačna. Pri skupini Shi in sod. (2016a) Nairovirusu podobni virusi oblikujejo sestrsko skupino s Phlebovirusu podobnimi virusi in skupino Arenaviridae, medtem ko Hantavirusu, Tospovirusu, Emaravirusu in Orthobunyavirusu podobni virusi oblikujejo svojo večjo skupino. V obeh primerih je drevo koreninjeno na sredini (angl. »midpoint rooting«), zato o direktni evolucijski starosti posamezne virusne skupine in naših novih bunyavirusov ne moremo govoriti (**Dodatek B; B.1, Slika B.1**).

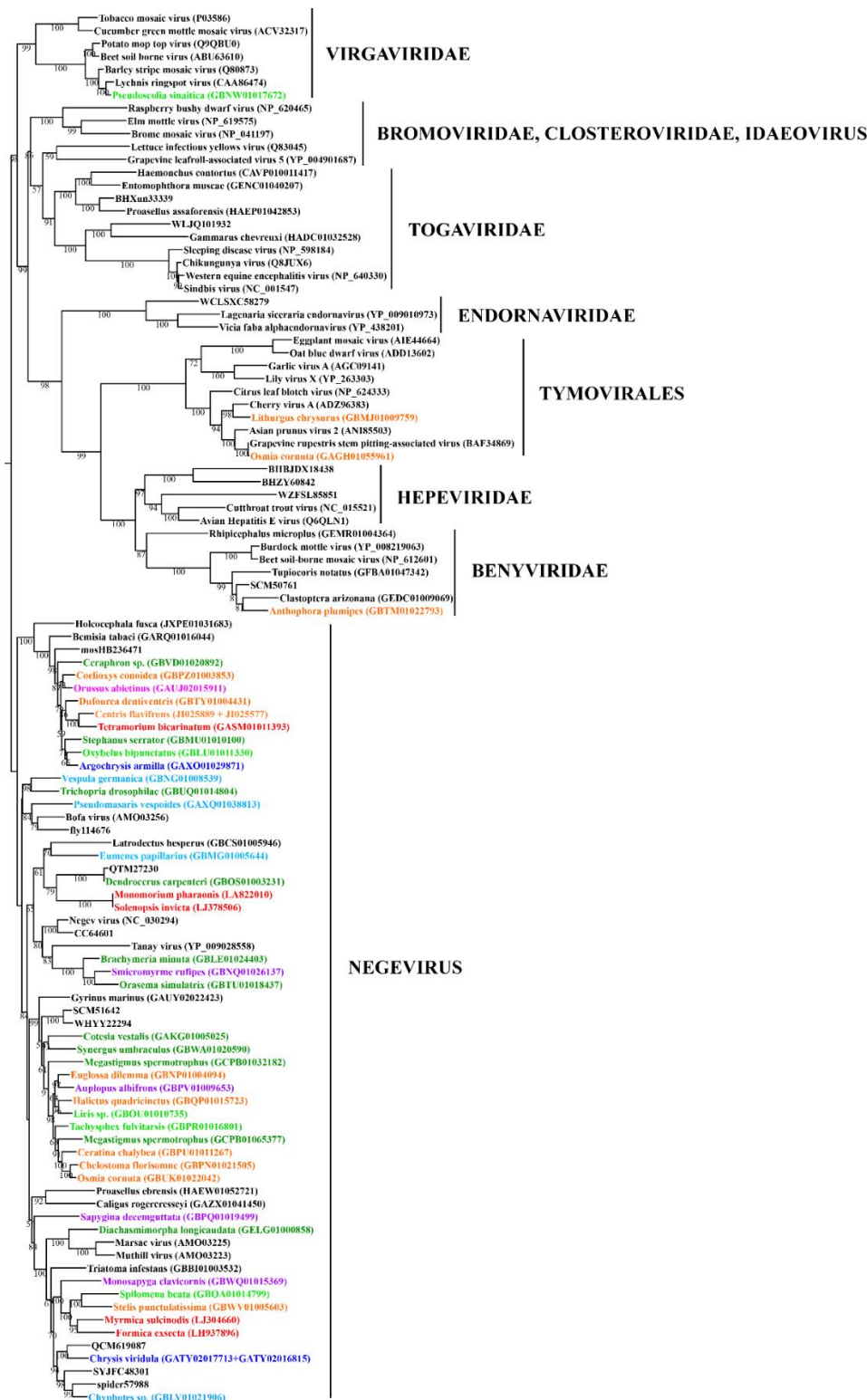
4.3.1.1.2 Klada Flavi

Klada Flavi prikazuje tri ločene skupine. Pegivirusi oblikujejo skupino s Hepacivirusi, kar kaže na njun monofiletski izvor. Pestivirusi oblikujejo skupino s flavivirusom podobnimi virusi (angl. »*Flavi-like*«), kar ravno tako nakazuje na njun monofiletski izvor. Jingmenvirusi in Flavivirusi oblikujejo tretjo monofiletsko skupino, ki je sestrška monofiletski skupini pestivirusov in Flavivirusom podobnih virusov. Vsi novi flavivirusi, ki smo jih našli med analizo genomskih in transkriptomskih podatkovnih baz kožekrilcev, so se uvrstili v skupino flavivirusom podobnih virusov. Pet flavivirusov je bilo najdenih v knjižnicah čebel (*Apoidea*). Trije od teh so bili najdeni endogenizirani v genomih čebel (*Apoidea*). Od endogeniziranih flavivirusnih genomov v genomih čebel sta bila dva v celoti ohranjena – *Ceratina calcarata* (LSNX01048950) in *Megachile rotundata* (AFJA01008853). Poleg flavivirusom podobnih virusov smo našli tudi jingmenvirus pri skupini Terebrantes (*Trichopria drosophilae*). Skupina jingmenvirusov torej tvori sestrsko skupino s flavivirusi, skupina pestivirusov tvori sestrsko skupino s flavivirusom podobnimi virusi in skupina pegivirusov tvori sestrsko skupino s hepacivirusi (**Dodatek B; B.1, Slika B.2**).

4.3.1.1.3 Klada Hepe-Virga

Klada Hepe-Virga obsega virusne družine Virgaviridae, Bromoviridae, Closteroviridae, Togaviridae, Endornaviridae, Hepeviridae, Benyviridae, rodova Idaeovirus in Negevirus ter red Tymovirales. Klada Hepe-Virga je ena največjih RNA virusnih klad s stališča raznolikosti virusnih skupin, ki se vanjo uvrščajo. V to klado smo uvrstili dvanajst RNA virusov iz transkriptomskih baz čebel (*Apoidea*), pet iz skupine Sphecoidea, pet iz mravelj, deset iz skupine Terebrantes, eno iz skupine Chrysidoidea in štiri iz skupine Pompiloidea. Velika večina novih RNA virusov, najdenih v okviru naše raziskave, se je uvrstila v skupino Negevirusov. Našli smo tudi RNA virus, ki se je uvrstil v klado Virgaviridae, dva virusa sta se uvrstila v skupino Tymovirales in RNA virus, ki se je uvrstil v skupino Benyviridae. Podobni so tudi rezultati raziskave Shi in sod. (2016a), pri kateri je bilo največ virusov klade Hepe-Virga uvrščenih v skupino Negevirus. Ta skupina je bila najbolj zastopana, pri čemer je večina virusnih genomov izvirala iz knjižnic žuželk, nekoliko manj iz knjižnic skupine Nematoda, negevirusi, ki so izvirali iz knjižnic rakov (*Crustacea*) in pajkovcev (*Chelicerata*), pa so bili v manjšini. Druga po zastopanosti je bila skupina hepevirusom podobnih virusov. Večina najdenih virusov je v tem primeru izvirala iz knjižnice rakov (*Crustacea*). Skupina Togaviridae je bila večinoma redka, našli so samo nekaj predstavnikov pri raki. Ostale skupine (Virgaviridae, Bromoviridae, Closteroviridae, Idaeovirus, Tymovirales in Endornaviridae) predstavljajo večinoma rastlinske viruse. Skupina Virgaviridae je skupina znanih rastlinskih virusov (Furovirus, Goravirus, Hordeivirus, Pecluvirus, Pomovirus, Tobamovirus, Tobravirus). V transkriptomskih bazah smo našli virgavirus, prisoten v organizmu *Pseudoscolia sinaitica*, ki sodi v skupino Sphecoidea. Gre za primer virusa, ki je po zaporedju podoben virusu »Lychnis ringspot virus«, znanemu patogenu mete (NCBI Resource Coordinators, 2016; NCBI Resource Coordinators, 2017). Pri iskanju homolognih zaporedij virgavirusne RdRp smo našli nukleotidno zaporedje rastlinskega virusa. Možna sta dva scenarija. Organizem *Pseudoscolia* je lahko vektor virusa, ki ga prenese na rastlino, čemur sledi aktivna okužba omenjene rastline. Drugi možni scenarij pa je, da je bil virus naključno prisoten na površini organizma. Organizem ga je lahko ob hranjenju naključno pobral s površine rastline. Drugi primer sta dva RNA virusa, ki sta bila najdena v

transkriptomih organizmov *Lithurgus chrysurus* in *Osmia cornuta* in sta bila umeščena v red Tymovirales. Virus, najden v organizmu *Lithurgus chrysurus*, je po zaporedju zelo podoben »Cherry virusu A«, rastlinskemu virusu skupine Betaflexiviridae-Capillovirus. Gostitelj virusa je japonska marelica (*Prunus mume*). Tudi v tem primeru sta dva možna scenarija. Organizem je lahko vektor, ki virus prenese na rastlino, ali pa je virus naključno pobral s površine rastline, na kateri se je hranil. Virus, najden v transkriptomu organizma *Osmia cornuta*, je identičen »Grapewine rupestris stem pitting-associated virus-u«, kar pomeni, da smo našli »Grapevine rupestris stem pitting-associated virus«, prisoten na organizmu *Osmia cornuta*. Virus okužuje vinsko trto, saj je bil izoliran iz kultivarja Teleki-Kober 5C (NCBI Taxonomy). Verjetno ga je pobral med hranjenjem, lahko pa je tudi vektor virusa. Virus, najden v transkriptomu organizma *Antophora plumipes*, je bil umeščen med Benyviridae. Gre za virus, ki je na nivoju zaporedja podoben »Beei soil-borne mosaic virusu«, izoliranemu iz navadne pese (*Beta vulgaris*). Tudi v tem primeru gre verjetno za naključen prenos iz rastline na površino organizma ali pa je organizem vektor omenjenega rastlinskega virusa (Slika 4.7).



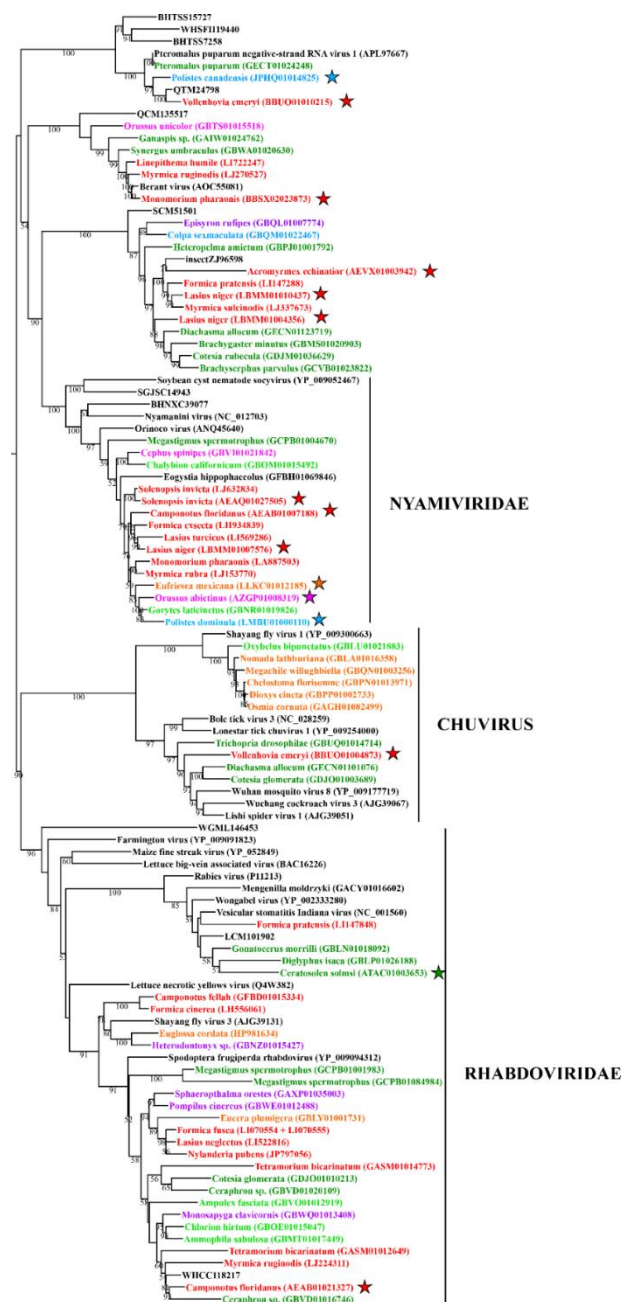
Slika 4.7: Filogenetsko drevo virusov klade **Hepe-Virga**, ki smo jih našli v transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«) v programu IQTREE (Nguyen in sod., 2015). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidioidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro.

4.3.1.14 Klada Luteo-Sobemo

Klada Luteo-Sobemo obsega pet monofiletskih skupin. Prvo tvorijo *Heterocapsa circularisquama* RNA virusu podobni virusi, drugo tvorijo *Humaita-Tubiacanga* virusu in »La Tardoie virusu« podobni virusi, tretjo tvorijo »Hubei sobemo-like virus 29« (QTM12808) podobni virusi, četrto monofiletsko skupino pa tvorijo »*Ixodes scapularis* associated virus 1« podobni virusi, peto (manjšo) monofiletsko skupino tvorijo *Carascovirusu* podobni virusi. Sedem virusov, ki spadajo v klado Luteo-Sobemo, smo našli v transkriptomih mravelj (Formicidae), tri v transkriptomih čebel (Apoidea), štiri v transkriptomih Terebrantes, dva v genomih Sphecoidea in enega v transkriptomu Chrysidoidea. Novoodkriti RNA virusi so se uvrstili v tri od petih monofiletskih skupin, vendar je najbolj bogato zastopana monofiletska skupina »*Ixodes scapularis* associated virus 1« podobnih virusov. Skupina, v katero ni bil uvrščen nobeden od naših zadetkov, so »*Heterocapsa circularisquama* RNA virusu« podobni virusi. Ugotovili smo, da je virusna klada Luteo-Sobemo enakomerno razširjena pri skupini Apocrita. Pri skupini Shi in sod. (2016a) so bili v skupino »*Heterocapsa circularisquama* RNA virusu« podobnih virusov v glavnem uvrščeni virusi iz skupine Spiralia. V skupino »*Humaita-Tubiacanga* virusu« in »*La Tardoie* virusu« podobnih virusov so bili uvrščeni virusi, ki izvirajo iz knjižnic insektov. Tudi skupina, ki jo tvorijo »Hubei sobemo-like virus 29« (QTM12808) podobni virusi, je v glavnem sestavljala predstavnike virusov iz transkriptomov insektov. Skupina »*Carascovirusu* podobnih virusov« je majhna in vključuje samo virusna predstavnika iz transkriptomov rakov in skupine Spiralia. Skupina »*Ixodes scapularis* associated virus 1« je bila najbolj bogato zastopana, podobno kot v naši raziskavi (**Dodatek B; B.1, Slika B.3**).

4.3.1.15 Klada Mono-Chu

Klada Mono-Chu jasno oblikuje šest monofiletskih skupin. Prvo predstavljajo »*Pteromalus puparum* negative strand RNA virus 1« podobni virusi, drugo predstavljajo »Berant virusu« podobni virusi, tretjo predstavljajo »Hubei diptera virus 11« (SCM51501) podobni virusi, četrto predstavlja skupina Nyamiviridae, peto Chuvirus in šesto Rhabdoviridae. Našli smo predstavnike Mono-Chu RNA virusne klade, ki so bili uvrščeni v vseh šest monofiletskih skupin. Poudariti pa velja, da so bile skupine Nyamiviridae, Rhabdoviridae in Chuvirus bistveno bolj bogato zastopane z virusnimi predstavniki kot ostale tri. Večina virusov Mono-Chu klade je bilo uvrščenih v skupino Rhabdoviridae. V transkriptomih kožekrilcev smo našli 28 virusnih predstavnikov pri mravljah (Formicidae), osem predstavnikov pri čebelah (Apoidea), 20 pri skupini Terebrantes, pet predstavnikov pri Sphecoidea in pet predstavnikov pri skupini Pompiloidea. Šest predstavnikov smo našli v knjižnicah ostalih skupin kožekrilcev (Hymenoptera). Mono-Chu je virusna klada, ki je bogato zastopana pri insektih in je enakomerno razširjena pri vseh skupinah kožekrilcev. Najbolj bogato zastopana skupina znotraj klade je skupina Rhabdoviridae. Podobni so tudi rezultati Shi in sod. (2016a), kjer po zastopanosti izrazito prednjači skupina Rhabdoviridae. Shi in sod. (2016a) so pokazali, da je klada Mono-Chu močno razširjena pri kopenskih nevretenčarjih, žuželkah (Insecta), skupini Nematoda in pajkovcih (Chelicerata) in nekoliko manj pri vodnih nevretenčarjih, predvsem pri rakah (Crustacea) (**Slika 4.8**).



Slika 4.8: Filogenetsko drevo virusov klade **Mono-Chu**, ki smo jih našli v genomih in transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«) v programu IQTREE (Nguyen in sod., 2015). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro. Endogenizirani virusi so označeni z zvezdico.

4.3.1.1.6 Klada Narna-Levi

Klada Narna-Levi oblikuje dve večji skupini: Narnaviridae in Leviviridae. Skupina Narnaviridae vsebuje tri glavne virusne skupine, ki se v filogenetski analizi kažejo kot monofiletske skupine: Narnavirus, Mitovirus in Ourmiavirus-u podobni virusi. V filogenetski analizi smo z uporabo proteinskega zaporedja narnavirusne in levivirusne RdRp kot iskalnega niza za iterativno iskanje genomskih in transkriptomskih podatkovnih baz kožekrilcev našli 15 virusov. Štiri od teh so bili najdeni v genomih mravelj (Formicidae).

Trije so bili najdeni v transkriptomih mravelj. Trije so bili najdeni v transkriptomih čebel (Apoidea), dva v transkriptomih skupine Terebrantes, eden v transkriptomih skupine Pompiloidea in eden v transkriptomih skupine Sphecoidea. Vsi virusi so bili uvrščeni v skupino Narnavirus (znotraj skupine Narnaviridae). Filogenetsko drevo, koreninjeno na sredini, skupino Leviviridae jasno opredeli kot ločeno monofiletsko skupino. Skupino mitovirusov opredeli kot sestrsko skupino skupini Leviviridae. Skupini Narnavirus in »Ourmia virusu« podobni virusi pa uvršča v dve sestrski skupini. Skupina »Ourmia virusu« podobnih virusov se loči v tri sestrške skupine: 1) Hubei narna-like virus 9 (WHBL72555) podobni virusi, 2) Ourmia melon virusu podobni in Beihai narna-like virus 20 (BWBFG40664) podobni virusi in 3) Hubei narna-like virusu 7 (WHBL72300) podobni virusi. Filogenetska analiza skupine Shi in sod. (2016a) je za razliko od naših raziskav skupine Narnaviruse, mitoviruse in Ourmiavirusu podobne viruse opredelila kot sestrške skupine, skupino Leviviridae pa uvrstila v ločeno monofiletsko skupino (Shi in sod., 2016a) (Dodatek B; B.1, Slika B.4).

4.3.1.17 Klada Tombus-Noda

Klada Tombus-Noda v naši filogenetski analizi oblikuje dve monofiletski skupini (Nodaviridae in Tombusviridae). Skupina Tombusviridae oblikuje dve sestrski skupini, prva vključuje virusu »Diaphorina citri-associated C virusu« podobne viruse, druga vključuje skupino Tombusviridae in skupino Carmotetraviridae. V transkriptomih kožekrilcev smo našli sedem tombusvirusov; tri pri Sphecoidea, enega pri mravljah (Formicidae) ter tri pri skupini Vespoidea, ki ne sodi v skupino mravelj (Formicidae). Vsi novi virusi razen enega (*Scolia hirta* (GBQJ01018513)) spadajo v skupino »Diaphorina citri-associated C virusu« podobnih virusov. V transkriptomih kožekrilcev smo našli tudi štiri nodaviruse, tri pri mravljah in enega pri skupini Sphecoidea. Novi virusi spadajo v alfanodavirusno skupino in med »Lake Sinai virusu« podobne viruse (Dodatek B; B.1, Slika B.5).

4.3.1.18 Klada Ophio

Klada Ophio v naši filogenetski analizi oblikuje dve sestrski skupini. V prvo so uvrščeni glivni »Rhizoctonia solani negative-stranded« virusu podobni virusi. V drugo so uvrščeni rastlinski »Citrus psorosis virusu« podobni virusi in tudi nov ophiovirus iz transkriptoma čebel (Apoidea). Virus smo našli v transkriptomu *Panurgus dentipes* in je podoben »Citrus psorosis virusu«, ki je rastlinski virus, izoliran iz citronalne citrone (*Citrus medica*). Lahko je šlo za primer vektorskega prenosa na rastlino ali naključnega prevzema s površine rastline kot gostitelja v času hranjenja (Dodatek B; B.1, Slika B.6).

4.3.1.19 Klada Orthomyxo

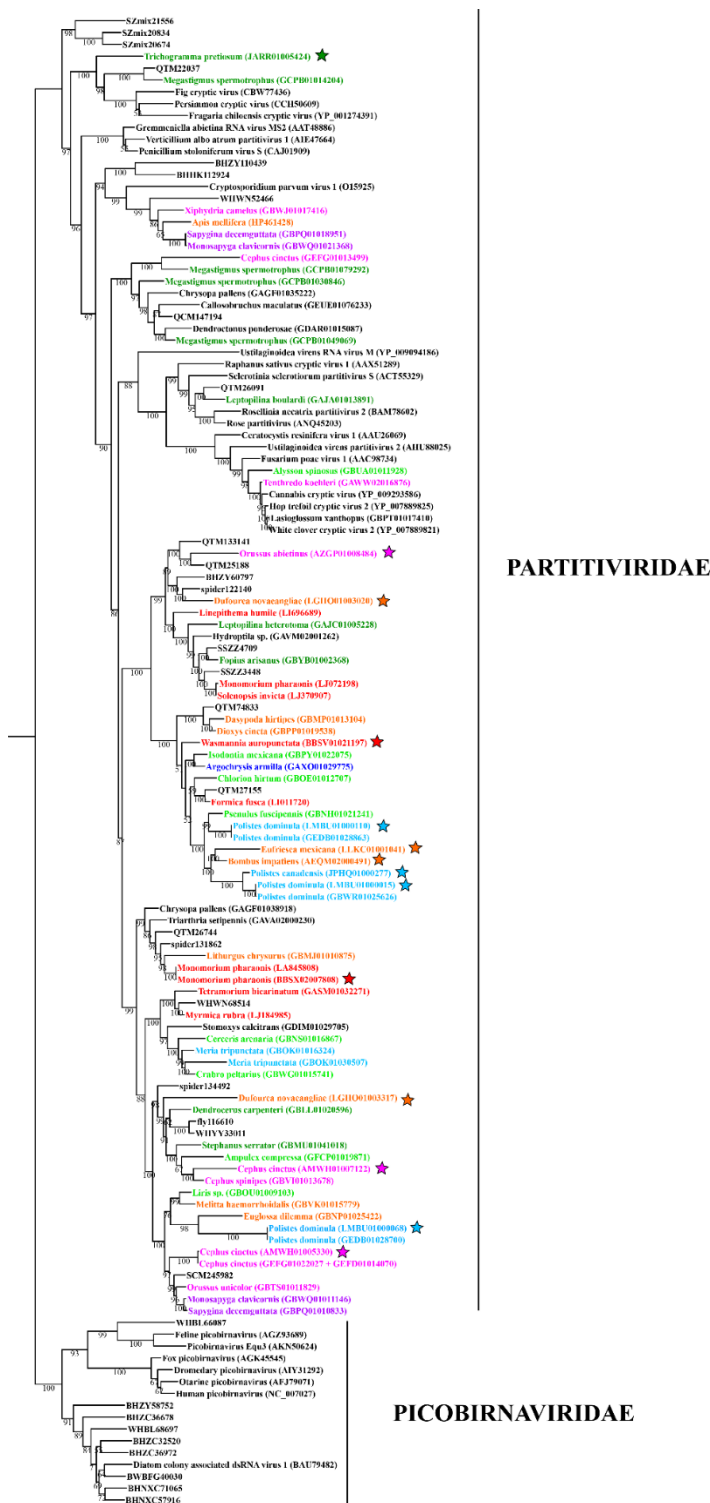
Pri kladi **Orthomyxo** smo v filogenetski analizi identificirali tri ločene skupine. Prvo skupino predstavljajo »Changping earthworm virusu 2« (CPQY18140) podobni virusi, v drugo spadajo isavirus, influenzavirusi in thogotovirusi. V tretjo, ki je v našem primeru in tudi v primeru Shi in sod. (2016a) najbolj bogato zastopana, pa so uvrščeni Quaranjavirusi oz. Quarafil virusu podobni virusi. V skupino Quarafil virusu podobnih virusov spadajo vsi novi virusi, ki jih je deset. Tri smo našli v transkriptomih skupine Terebrantes, enega v transkriptomu skupine Sphecoidea, enega v transkriptomu Pompiloidea, tri v bazah rodu Polistes (ose (Vespidae) – skupina Vespoidea, ki ne sodi v skupino mravelj (Formicidae)), od tega dva v genomih in enega v transkriptomih. Tri viruse smo našli v bazah kožekrilcev, ki ne sodijo v skupino Apocrita, od tega enega v genomih in dva v transkriptomih. Tri viruse klade Orthomyxo smo torej našli endogenizirane v genomih kožekrilcev – dva v genomih skupine os (Vespidae) in enega v genomih skupine Cephoidea. Glede na raziskave Shi in sod. (2016a) je klada Orthomyxo najbolj razširjena ravno pri žuželkah. Najbolj pogosto je zastopana skupina Quarafil virusu podobnih virusov. Orthomyxovirusi so pri pajkovcih, raki in skupini Spiralia prisotni v manjšini (Dodatek B; B.1, Slika B.7).

4.3.1.10 Klada Partiti-Picobirna

Klada Partiti-Picobirna je glede na naše raziskave in glede na raziskave Shi in sod. (2016a) bogato razširjena virusna klada tako pri žuželkah kot pri rastlinah, raki in mehkužcih. Oblikovala naj bi skupino znotraj pikornavirusom podobne naddružine, predvsem zaradi pikornavirusom podobne RdRp (Koonin in sod., 2015). Pri filogeniji klade Partiti-Picobirna

je horizontalni prenos igral pomembno vlogo, saj gre za korelacijo virusnih skupin z rastlinskimi in glivnimi gostitelji (Koonin in sod., 2015). V genomih in transkriptomih kožekrilcev smo našli 60 virusov, ki spadajo v skupino Partitiviridae. V bazah skupine Terebrantes smo našli 10 virusov, 10 v bazah čebel (Apoidea), devet v bazah mravelj, štiri pri skupini Pompiloidea, osem pri skupini Sphecoidea, enega v bazah skupine Chrysidoidea, devet pri skupini kožekrilcev, ki ne sodijo v skupino Apocrita, in devet v bazah organizmov, ki sodijo v skupino Vespoidea, vendar ne v skupino mravelj (Formicidae). Od skupno 60 novih partitivirusov smo jih kar 14 našli v genomih kožekrilcev – enega v genomih skupine Terebrantes, tri v genomih kožekrilcev, ki ne sodijo v skupino Apocrita, štiri v genomih čebel (Apoidea), dva v genomih mravelj in štiri v genomih organizmov skupine Vespoidea, ki ne sodijo v skupino mravelj. Partitivirusi so torej enakomerno razširjeni pri vseh skupinah kožekrilcev in ni bistvene razlike v zastopanosti med posameznimi skupinami kožekrilcev. Novi partitivirusi pri kožekrilcih spadajo v več virusnih skupin. Skupina virusov, najdenih v transkriptomih organizmov *Xiphydria camelus*, *Apis mellifera* in *Sapygina decemgutata* (isti virus je bil najden tudi v organizmu *Monosapyga clavicornis*), je sorodnih virusu »Cryptosporidium parvum virus 1«, izoliranemu iz parazitskega alveolata *Cryptosporidium parvum*, ki povzroča kriptosporidiozo pri človeku. V transkriptomih organizma *Leptopilina boulandi* smo našli »Hubei partiti-like virus 27« podoben oz. soroden virus (QTM26091), ki ga je skupina Shi in sod. (2016a) našla v transkriptomu Odonata. *Leptopilina boulandi* je rod parazitoidnih os, ki sodi v skupino Figitidae (NCBI Taxonomy) in napada rod vinskih mušic (*Drosophila*). V tem primeru lahko predpostavimo bodisi vektorski prenos virusa iz organizma *Leptopilina boulandi* na njegov plen bodisi prevzem od plena ob hranjenju. V genomu druge vrste parazitoidnih os smo našli endogeniziran virus, ki je soroden virusu »Hubei partiti-like virusu 11« (QTM25188) in »Hubei partiti-like virusu 10« (QTM133141). V tem primeru vemo, da je organizem gostitelj omenjenega virusa, saj frakcija genoma virusa, ki smo jo našli integrirano v gostiteljski genom, nakazuje na preteklo aktivno okužbo organizma z omenjenim virusom. V transkriptomu organizma *Tenthredo koehleri* smo našli genom virusa, soroden skupini rastlinskih virusov: »Cannabis cryptic virus«, »Hop trefoil cryptic virus 2« in »White clover cryptic virus 2«. »Cannabis cryptic virus« je rastlinski virus, izoliran iz navadne konoplje (*Cannabis sativa*). »Hop trefoil cryptic virus 2« je bil izoliran iz hmeljne meteljke (*Medicago lupulina*). »White clover cryptic virus 2« je bil izoliran iz plazeče detelje, kultivarja Lirepa (*Trifolium repens* cv. Lirepa). Omenjeni rastlinski virusi so variante istega virusa – gre za različno poimenovanje praktično istega rastlinskega virusa, ki je bil najden tudi v transkriptomu *Tenthredo koehleri*. Lahko gre za naključen prevzem ob opraševanju rastline oz. hranjenju ali pa za vektorski prenos na rastlino. Skupina oblikuje širšo skupino z virusi »Ceratocystis resinifera virus 1«, »Ustilagoidea virens partitivirus 2« in »Fusarium poae virus 1«. »Ceratocystis resinifera virus 1« je bil izoliran iz askomikotne glive *Ceratocystis resinifera* izolata LG115 (rod *Mycoascales*), »Ustilagoidea virens partitivirus 2« je bil izoliran iz askomikotne glive *Ustilagoidea virens* (rod *Hypocreales*) in »Fusarium poae virus 1« je bil izoliran iz askomikotne glive *Fusarium poae* (rod *Hypocreales*). Virus, soroden »Fusarium poae virusu 1«, je bil najden v transkriptomu *Alysson spinosus*. Verjetno gre za glivni virus, najden v transkriptomu kožekrilca iz skupine Specoidea. Omenjen virus je 47 % podoben virusu »Fusarium poae virus 1«. Virus je lahko 1) bil naključno prisoten na organizmu, 2) dejanski gostitelj omenjenega virusa je lahko gliva, ki je bila naključno prisotna na organizmu. Tretji možen scenarij je, da je šlo za parazitsko glivo, ki je bila okužena z virusom. Organizem je bil lahko plen omenjene parazitske glive, kar pomeni, da je genom omenjenega virusa dejansko povezan s parazitsko glivo in ne z organizmom *Alysson spinosus*. Glede na to, da rastlinski virusi tvorijo monofiletsko skupino z glivnimi virusi, lahko trdimo, da je v tem primeru šlo za prenos virusa iz glivnega na rastlinskega gostitelja. Proteinsko zaporedje ORF, ki kodira RdRp virusa, najdenega v transkriptomu organizma *Alysson spinosus* (GBUA01011928), je namreč kar 47 % podobno »Fusarium poae virusu 1«. Hkrati je proteinsko zaporedje ORF-ja, ki kodira zapis za RdRp virusa, najdenega v transkriptomu organizma *Tenthredo koehleri* (GAWW02016876), 46 % podobno virusu »Fusarium poae virus 1«. Verjetno je sledil premik v genetskem driftu in s tem hitrosti mutacije, kar je rezultiralo v diverzifikaciji virusnega genoma. Našli smo tri različne primere virusa, ki je bil hkrati prisoten v genomskih in transkriptomskih knjižnicah – dva različna primera istega virusa v genomskih in transkriptomskih knjižnicah organizma *Polistes dominula* in en primer istega virusa smo našli hkrati v genomskih in transkriptomskih knjižnicah organizma *Cephus cinctus*. Če najdemo isti virus v genomski in transkriptomski knjižnici istega organizma, pomeni, da je organizem dejanski gostitelj omenjenega virusa. Hkrati smo našli virus v transkriptomu

organizma, ki je integriran v genom istega organizma, kar nakazuje na preteklo aktivno okužbo organizma z danim virusom – vezava na membrano, vstop v gostitelja, podvojevanje genoma (hkrati vmes integriranje v gostiteljski genom), sestavljanje virusa in izstop virusa iz celice (**Slika 4.9**).



Slika 4.9: Filogenetsko drevo virusov klade **Partiti-Picobirna**, ki smo jih našli v genomih in transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«) v programu IQTREE (Nguyen in sod., 2015). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro. Endogenizirani virusi so označeni z zvezdico.

4.3.1.1.11 Klada Permutotetra

Klada Permutotetra je prisotna predvsem pri žuželkah (Shi in sod., 2016a). Pri rakih (Crustacea), stonogah (Myriapoda) in pajkovcih (Chelicerata) je redka. V transkriptomih kožekrilcev smo našli šest permutotetravirusov, tri pri mravljah in tri pri skupini Terebrantes. V transkriptomu organizma *Leptopilina boucardi* smo našli virus, ki je po zaporedju podoben Newfield virusu. *Leptopilina boucardi* je parazitoidna osa, ki napada vinsko mušico (*Drosophila melanogaster*). Newfield virus je bil izoliran iz vinske mušice (NCBI Taxonomy), zato lahko v tem primeru govorimo o 1) vektorskem prenosu s parazita (*Leptopilina*) na plen (*Drosophila*) ali o 2) prevzemu virusa iz plena ob aktivni okužbi (*Drosophila*). V transkriptomih organizma *Fopius arisanus* smo našli virus, soroden »Daesongdong virusu 2«, ki je bil izoliran iz komarja *Culex pipiens*. Tudi *Fopius arisanus* je parazitoidni kožekrillec, ki napada jajčeca in ličinke reda Diptera skupine Schizophora (Ceratitis, Bactrocera, Anastrepha). *Culex pipiens* sodi v isti red Diptera (družina Culicidae) (Dodatek B; B.1; Slika B.8).

4.3.1.1.12 Klada Picorna-Calici

Klada Picorna-Calici zajema dve glavni virusni skupini: Picornavirales in Caliciviridae. Picornavirales je ena največjih virusnih skupin s širokim razponom gostiteljev. Predstavlja jedro pikornavirusom podobne naddružine virusov. Virusna skupina Caliciviridae sicer ni uvrščena med pikornaviruse (Picornavirales), vendar ima z njimi skupno ohranjen S3H (angl. »superfamily 3 helicase« – helikaza superdružine 3) – Vpg-3CPro- RdRp-JRC jedrni del genoma. Lahko se povežejo s skupino Totiviridae (Koonin in sod., 2015). Skupina Shi in sod. (2016a) je klado razdelila na virusne skupine: 1) Dicistroviridae, 2) Iflaviridae, 3) Secoviridae, 4) Picornaviridae (skupine, ki so bile že prvotno identificirane v okviru ICTV (angl. »International Committee on Taxonomy of Viruses«), 5) »vodni pikornavirusi« (angl. »aquatic picornavirales«), ki poleg drugih neklasificiranih virusov (angl. »floating genera«) zajemajo ključne skupine Marnaviridae, Bacillarnavirus in Labyrnavirus, 6) Kelp fly virusu podobni virusi, 7) Nora virusu podobni virusi. V njihovi raziskavi sta se oblikovali tudi skupini Arivirusu in Posavirusu podobnih virusov, ki pa jih niso poimenovali. V transkriptomih kožekrilcev je klada Picorna-Calici ena najbolj bogatno zastopanih virusnih klad. Prisotna je pri vseh raziskanih glavnih skupinah kožekrilcev, čeprav nekoliko prednjačijo mravlje (Formicidae). V raziskavi smo našli skupno 89 novih pikornavirusov. Trije od teh spadajo v skupino Kelp fly virusu podobnih virusov, štirje v skupino Nora virusu podobnih virusov, eden v skupino Arivirusu podobnih virusov, štirje v skupino Secoviridae, 17 v skupino Dicistroviridae in 48 (več kot polovica vseh najdenih pikornavirusov pri kožekrilcih) v iflaviruse, kar kaže na dejstvo, da je skupina Iflaviridae očitno najbolj bogato zastopana pikornavirusna skupina pri kožekrilcih. V transkriptomih kožekrilcev nismo našli virusov, ki spadajo v skupino vodnih pikornavirusov (angl. »aquatic picornavirales«). Poleg omenjenih definiranih skupin v analizi 9 virusov oblikuje novo nedefinirano virusno skupino kot »Shuangao insect virusu 8« (insectZJ96360) podobni virusi. Omenjen virus oblikuje klado z virusi »Hubei picorna-like virus 51« (QTM27291), »Hubei picorna-like virus 81« (QTM27117) in »Hubei picorna-like virus 82« (spider133992). Omenjeno skupino smo torej uspeli razširiti. Trije virusi oblikujejo novo virusno skupino s »Hubei tetragnatha maxillosa virusom 3« (SSZZ2994) in »Wuhan spider virus 3« (spider133889). V raziskavi Shi in sod. (2016a) »Hubei tetragnatha maxillosa virus 3« (SSZZ2994) in »Wuhan spider virus 3« (spider133889) oblikujeta svojo sicer nedefinirano skupino, ki smo jo nekoliko obogatili z novimi RNA virusi, ki so bili vanjo uvrščeni. Pri kožekrilcih so pikornavirusi najbolj razširjeni pri mravljah (Formicidae). Našli smo kar 40 pikornavirusov v transkriptomih mravelj, šest v transkriptomih čebel (Apoidea), 18 v transkriptomih skupine Terebrantes, 15 pri skupini Sphecoidea, dve pri Pompiloidea, enega pri Chrysidoidea in štiri pri ostalih organizmih skupine Vespoidea, ki ne sodijo v skupino mravelj (Formicidae). Tri viruse smo našli v transkriptomih kožekrilcev, ki ne sodijo v skupino Apocrita. V primeru skupine Kelp fly virusov podobnih virusov smo našli identičen virus, prisoten pri organizmih *Solenopsis invicta* (rdeča ognjena mravlja) in *Monomorium pharaonis* (faraonska mravlja). Gre za virus, soroden *Solenopsis invicta* virusu 3, izoliran iz rdeče ognjene mravlje (*Solenopsis invicta*). *Solenopsis invicta* (rdeča ognjena mravlja) in *Monomorium pharaonis* (faraonska mravlja) oba sodita v skupino Solenopsidini, tako da gre za sorodna organizma. Glede na to, da je virus soroden virusu, ki je bil izoliran iz rdeče ognjene mravlje, predpostavljamo, da sta obe vrsti možna gostitelja virusa. V primeru skupine Nora virusu podobnih virusov smo našli Nora virus (izoliran iz

vinske mušice) pri organizmu *Leptopilina heterotoma*, znani parazitoidni osi. Organizem je parazit, ki inficira vinsko mušico (*Drosophila melanogaster*). Virus je lahko okuževal vinsko mušico v času, ko se je na njej pojavil parazit *Leptopilina boulandi*. Druga možnost pa je, da ga je *Leptopilina boulandi* kot vektor prenesla nanjo. Soroden virus smo našli tudi na organizmu *Fopius arisanus*, znanega parazita *Bactrocera dorsalis*, *Ceratitis capitata* in organizmov rodu *Anastrepha* sp., ki sodijo v isto skupino Schizophora kot *Drosophila melanogaster*. Tudi tukaj je možno, da je *Fopius arisanus* virus pobral v času, ko je kot parazit »okužil« gostitelja, ali pa obratno. *Fopius arisanus* je lahko virus prenesel na gostitelja kot vektor. Pri skupini rastlinskih virusov Secoviridae smo v transkriptomih organizmov *Halictus quadricinctus* in *Crabro peltarius* našli »Cherry leaf roll virusu« (izoliranemu iz navadne breze (*Betula pendula*)) identičen virus. V transkriptomu organizma *Chalybion californicum* smo našli »Tobacco ringspot virus«, ki je bil izoliran iz vinske trte kultivarja Blanc du Bois. V transkriptomu organizma *Vespa crabro* smo našli virus, po zaporedju podoben »Rice tungro spherical virusu«, izoliranemu iz riža (*Oryza sativa*). V vseh primerih gre torej za najdena nukleotidna zaporedja rastlinskim virusom skupine Secoviridae sorodnih virusov. Lahko gre za vektorski prenos virusa od žuželke na rastlino ali za prevzem virusa od rastline ob hranjenju (**Dodatek B; B.1, Slika B.9**).

4.3.1.1.13 Klada Qinvirus

Klada Qinvirus v filogenetski analizi oblikuje dve virusni skupini, Hubei qinvirus-like virus (SCM49583) podobni virusi in Beihai sesamid crab virus 4 (SCJXSX38901) podobni virusi, ki zajema med seboj dokaj divergentne viruse glede na dolžino vej filogenetskega drevesa. Našli smo štiri RNA viruse in vsi so bili uvrščeni v skupino »Hubei qinvirus-like virus« (SCM49583) podobnih virusov. Dva od tega sta bila najdena v genomih mravelj, dva pa v transkriptomih mravelj (Formicidae). Virusi, ki smo jih uvrstili v klado Qinvirus, so bili torej najdeni izključno v genomih in transkriptomih mravelj (Formicidae). Nasprotno je skupina Shi in sod. (2016) našla samo dva qinvirusa v transkriptomih žuželk. Ostali so izvirali iz transkriptomov rakov (4), glist (Nematoda) (1) in skupine Spiralia (1). Skupino qinvirusov smo torej dopolnili z naborom novih RNA virusov, ki izvirajo iz genomov in transkriptomov kožekrilcev (Hymenoptera) (**Dodatek B; B.1, Slika B.10**).

4.3.1.1.14 Klada Reoviridae

Klada Reoviridae je raznolika skupina segmentiranih virusov z 10–12 segmenti. V filogenetski analizi RNA virusov iz genomov in transkriptomov kožekrilcev klada oblikuje tri ločene skupine. V prvo se uvrščajo Aquareovirus, Orthoreovirus, Rotavirus, Phytoreovirus, Cardoreovirus, Seadornavirus, Mimoreovirus in Orbivirus, v drugo Coltivirus, Mycoreovirus in Fijivirus, v tretjo pa Cypovirusi, Oryzavirus in Dinovernavirus. Našli smo osem novih reovirusov, od tega tri v transkriptomih čebel (Apoidea), dva pri skupini Terebrantes, enega pri skupini Sphecoidea, enega pri mravljah in enega pri skupini Pompiloidea. Izmed novih reovirusov pri čebelah je bil eden najden endogeniziran v njihovem genomu. Pri skupini Shi in sod. (2016) revirusi oblikujejo podobno kot pri nas tri skupine. V prvi so skupine Cypovirus, Oryzavirus, Fijivirus, Coltivirus, Mycovirus in Orthoreovirus. V drugi so skupine Rotavirus, Seadornavirus, Phytoreovirus. V tretji sta skupini Mimoreovirus in Orbivirus. Mimoreovirus in Orbivirus oblikujeta sestrsko skupino s skupino, v kateri so skupine Rotavirus, Seadornavirus, Phytoreovirus. Filogenetsko drevo skupine Shi in sod. (2016a) reoviruse razvršča nekoliko drugače kot naša raziskava. Novi reovirusi pri naši raziskavi so bili umeščeni v vse tri glavne skupine reovirusov (**Dodatek B; B.1, Slika B.11**).

4.3.1.1.15 Klada Toti-Chryso

Klada Toti-Chryso pri kožekrilcih oblikuje dve glavni virusni skupini. V prvo so uvrščeni Totivirus, Trichomonasvirus, Leishmanivirus in Victorivirus ter Chrysovirusu podobni virusi. V drugo so uvrščeni Giardia lamblia virusu podobni virusi. Našli smo 20 virusov klade Toti-Chryso v genomih in transkriptomih kožekrilcev. Od tega je večina (10) prisotna pri čebelah – šest v genomih in štiri v transkriptomih. Štiri totiviruse smo našli endogenizirane v genomih čebel. Dva virusa smo našli v transkriptomih skupine Pompiloidea, enega v transkriptomih mravelj, dva v transkriptomih skupine Terebrantes, enega v transkriptomih skupine Sphecoidea in enega v transkriptomih skupine Vespoidea, ki ne sodijo v skupino mravelj (*Eumenes papillarius*). Večina novih totivirusov spada v virusno skupino Giardiavirusu podobnih virusov. Samo dva sta bila uvrščena v virusno skupino skupaj s

skupinami Totivirus, Trichomonasvirus, Leishmanivirus in Victorivirus in Chrysovirusu podobnih virusov. Tudi v filogenetski analizi skupine Shi in sod. (2016a) sta se oblikovali dve večji skupini, pri čemer je bila skupina Giardiavirus uvrščena oz. umeščena v ločeno virusno skupino. Prvotno znani kot glavni virusi so bili pri skupini Shi in sod. (2016) najdeni večinoma pri žuželkah, vendar tudi pri raki (Crustacea), glistah (Nematoda), pajkovcih (Chelicerata) in skupini Lophotrochozoa (v glavnem Mollusca). Ob poravnavi RdRp in filogenetski analizi, ki jo je naredila Kooninova ekipa leta 2015, so se Totivirusi povezali skupaj s Caliciviridae (Koonin in sod., 2015). Shi in sod. (2016a) so jih združili s skupino Chrysoviridae (**Dodatek B; B.1, Dodatek B.12**).

4.3.1.2 Metulji (Lepidoptera)

Metulji in vešče (Lepidoptera) so ena izmed štirih glavnih skupin žuželk s 160 000 opisanimi vrstami, pri čemer je dejansko število vrst ocenjeno na 500 000 (Kawahara in Breinholt, 2014). Preferenčno so rastlinojedi in opraševalci, ki imajo glavno vlogo pri razširjanju kritosemenk. Njihova filogenetska klasifikacija še danes ni dosegla konsenza 1) glede umestitve metuljev v kontekst ostalih skupin Lepidoptera in 2) glede uvrstitve v višje taksonomske kategorije. Skupina Lepidoptera je razdeljena na več skupin: Noctuoidea, Gracillarioidea, Yponomeutoidea, Gelechioidea, Daloceridae, Megalopygidae, Cossoidea, Pterophoridae, Tortricidae, Urodidae, Pyraloidea, Thyrididae idr. Raziskovalci so v preteklih filogenetskih analizah skupine Lepidoptera uporabili predstavnike iz vsake posamezne omenjene skupine (Kawahara in Breinholt, 2014). Skupina Lepidoptera je zajemala (do 7. 2. 2018) 301 genomsko in transkriptomsko knjižnico. Nove RNA viruse smo poiskali z iterativnim iskanjem po njihovih genomskih in transkriptomskih knjižnicah. Vsi novi RNA virusi pri metuljih so označeni z rdečo barvo.

4.3.1.2.1 Klada Bunya-Arena

Pri kladi Bunya-Arena smo opazili tri ločene skupine. Prva vključuje Hantavirus, Emaravirus, Tospovirus in Orthobunyavirus. Druga vključuje Nairovirusu podobne viruse in tretja vključuje Phlebovirusu in Tenuivirusu podobne viruse. V transkriptomih metuljev smo našli tri viruse, uvrščene v klado Bunya-Arena. En bunyavirus sodi v prvo skupino, ki vključuje Hantavirus, Emaravirus, Tospovirus in Orthobunyavirus. Soroden je Seattle prectang virusu, ki je bil izoliran iz metulja *Pasiphila rectangulata*. Dva bunyavirusa spadata v skupino Phlebovirusu in Tenuivirusu podobnih virusov. Našli smo jih v transkriptomih organizmov *Ostrinia furnacalis* in *Thyridopteryx ephemeraeformis*. Virus sta sorodna z »Wuhan spider virusom«, izoliranim iz pajkovca *Neoscona nautica* (**Dodatek B; B.2, Slika B.14**).

4.3.1.2.2 Klada Astro

Klada Astro se loči na dve skupini. V prvo so uvrščeni definirani in klasificirani astrovirusi, kot so Avastrovirus in Mamastrovirus. Drugo skupino so v filogenomski analizi prikazali Shi in sod. (2016a). Gre za »Hubei leech virusu 1« (SZmix32467) podobne viruse, ki poleg tripsinu podobne peptidazne domene (angl. »trypsin-like peptidase domain«) in RdRp vsebujejo ločen ORF, ki kodira zapis za virusni plaščni protein bodisi 1) družine peptidaza A21 (angl. »viral coat protein peptidase A21 family«) 2) bodisi domene S (angl. »viral coat protein S domain«). Virus izvira iz organizma *Whitmania pigra*, ki je kolobarnik (Annelida). Oba nova astrovirusa spadata med »Hubei leech virusu 1« (SZmix32467) podobne viruse in sta mu najbolj sorodna. Prvi izvira iz transkriptoma *Amyelosis transitela*, drugi pa iz transkriptoma *Helicoverpa assulta*. Za razliko od naše raziskave pa se v filogenetskem drevesu Shi in sod. (2016a) pri astrovirusih nista oblikovali dve ločeni skupini. Kladi Avastrovirus in Mamastrovirus oblikujeta monofiletsko skupino, ostali predstavniki, ki so jih našli v knjižnicah stonog (Myriapoda), rakov (Crustacea) in mehkužcev (Mollusca), pa oblikujejo ločene posamezne veje v drevesu, koreninjenem na sredini (angl. »midpoint rooting«). Vzrok za razlike v topologiji drevesa je boljše taksonomsko vzorčenje v drugi skupini astrovirusov (**Dodatek B; B.2, Slika B.13**).

4.3.1.2.3 Klada Flavi

Klada Flavi je pri metuljih (Lepidoptera) razdeljena na tri ločene skupine. Prvo monofiletsko skupino zajemajo Jingmenvirus in Flavivirus. Drugo zajemajo Pestivirus, Hepacivirus in Pegivirus. Tretjo skupino zajemajo Flavivirusom podobni virusi (angl. »Flavi-like«). Nov

flavivirus je bil pri naši filogenetski analizi uvrščen med Flavivirusom podobne viruse in je najbolj soroden z virusom Shayang fly virus 4, izoliranim iz muhe (*Musca domestica*). Predstavniki reda Diptera, predvsem komarji, so znani vektorji flavivirusov (**Dodatek B; B.2, Slika B.15**).

4.3.1.2.4 Klada Hepe-Virga

Klada Hepe-Virga tako kot pri kožekrilcih (Hymenoptera) tudi pri metuljih in večšah (Lepidoptera) oblikuje ločene skupine, ki vključujejo skupine Hepeviridae, Benyviridae, Virgaviridae, Togaviridae in Negevirus. V transkriptomih Lepidoptera smo našli tri virusne predstavnike, in sicer dva negevirusa v transkriptomih organizmov *Yponomeuta evonymellus* in navadnega modrina (*Polyommatus icarus*) in en virus, umeščen v skupino Benyviridae v transkriptomu organizma *Eudarcia simulatricella* (GEOF01029671). Skupina Benyviridae je skupina rastlinskih virusov. Proteinsko zaporedje RdRp virusa iz transkriptoma organizma *Eudarcia simulatricella* je 40 % podobno proteinskemu zaporedju RdRp »Burdock mottle virusa« in 38 % podobno proteinskemu zaporedju RdRp »Beet soil-borne mosaic virusa«. Zato je virus, najden v transkriptomih *Eudarcia simulatricella* (GEOF01029671), najverjetneje rastlinski virus, ki je soroden Burdock mottle virusu in Beet soil-borne mosaic virusu. Na organizem je lahko prešel naključno ob hranjenju ali pa deluje kot vektor, ki ob opraševanju rastlin na rastlino prenese virus. Našli smo tudi dva negevirusa v genomu organizma *Calycopsis cecrops* (sodi med modrine, Lycaenidae), kar pomeni preteklo aktivno okužbo vrste organizma z omenjenimi virusi. Organizem je dejanski gostitelj obeh sorodnih najdenih negevirusov. **Red Tymovirales** spada v klado Hepe-Virga, vendar smo se odločili za ločeno filogenetsko analizo dveh RdRp virusov, najdenih v transkriptomih *Athetis lepigone* (ditrisije, sovke – Noctuidae) (GARC) in *Micropterix calthella* (kalužni prametuljček) (GACP) samo v okviru reda Tymovirales. Virus, najden v transkriptomu *Athetis lepigone* (GARC), je najbolj soroden rastlinskemu virusu Citrus leaf blotch virus, izoliranemu iz pomaranče (*Citrus sinensis*) kultivarja Fengjie 72-1. Njuni RdRp sta 75 % identični. Virus, najden v transkriptomu *Micropterix calthella* (GACP), je najbolj soroden virusu »Cherry virus A«, izoliranemu iz japonske marelice (*Prunus mume*). Njuni RdRp sta 61 % identični. V obeh primerih gre za nukleotidni zaporedji genomov rastlinskih virusov, ki sta bili najdeni v dveh transkriptomih metuljev. Pri tem lahko metulja delujeta kot vektorja, ki med opraševanjem virus preneseta na rastlino ali pa sta virus naključno prevzela od rastlin med opraševanjem ali prehranjevanjem (**Dodatek B; B.2, Slika B.16 in Slika B.24**).

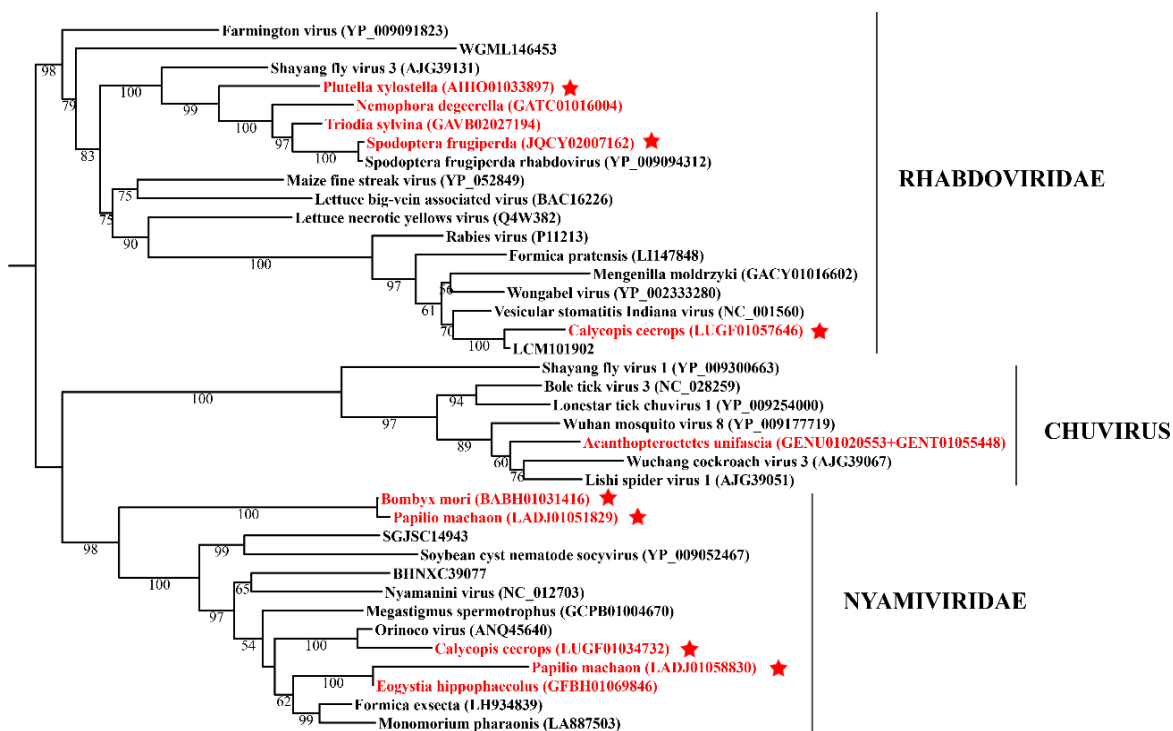
4.3.1.2.5 Klada Luteo-Sobemo

Klada Luteo-Sobemo oblikuje tri večje skupine; prvo tvorijo »Heterocapsa circularisquama RNA virusu« podobni virusi, drugo tvorijo »Humaita-Tubiacanga virusu« in »La Tardorie virusu« podobni virusi in tretjo tvorijo »Carascovirusu podobni« virusi. »Carascovirusu podobni« in »La Tardorie virusu« podobni virusi tvorijo sestrške skupine in jih lahko razdelimo na dve skupni glede na našo filogenetsko raziskavo. Naši novi RNA virusi so se uvrstili med »Humaita-Tubiacanga virusu« in »La Tardorie virusu« podobne viruse. Našli smo tri viruse v transkriptomih organizmov *Enteucha acetosae* (sodi med monotrisije, pritlikave listne zavrtače – Nepticulidae), *Dryadaula visaliella* (ditrisiji, pravi molji-Tineidae) in *Tischeria quercitella* (ditrisiji, pritlikavi listni zavrtači – Nepticulidae). V transkriptomih metuljev smo odkrili dva celotna genoma virusov, ki spadata v klado Luteo-Sobemo – *Enteucha acetosae* (GEOE01112900) in *Tischeria quercitella* (GEOU01069137) (**Dodatek B; B.2, Slika B.17**).

4.3.1.2.6 Klada Mono-Chu

Klada Mono-Chu oblikuje tri monofiletske skupine; Rhabdoviridae, Nyamiviridae in Chuvirus. Našli smo 11 novih mononegavirusov, ki so bili uvrščeni v klado Mono-Chu. Od tega smo jih našli sedem v genomih in štiri v transkriptomih. V genomih metuljev so bili trije uvrščeni v skupino rhabdovirusov in štiri v skupino Nyamiviridae. Od štirih virusov, najdenih v transkriptomih metuljev, sta bila dva uvrščena med rhabdoviruse, eden med Chuviruse in eden v Nyamiviridae. Glede na to, da fragmenti virusnih genomov ali celotni virusni genomi, ki so se integrirali v genom metuljev, kažejo na aktivne pretekle okužbe metuljev z virusi, vemo, da so vrste metuljev *Plutella xylostella* (kapusov molj), *Spodoptera frugiperda* (ditrisiji, sovke-Noctuidae), *Calycopsis cecrops* (ditrisiji, dnevni metulji – Papilionidae), *Bombyx mori* (navadna sviloprejkica) in *Papilio machaon* (ditrisiji, dnevni

metulji – Papilionidae, lastovičar) dejanski gostitelji virusov klade Mono-Chu. Da so se integrirali v virusni genom omenjenih gostiteljev, se je moral virus vezati na membrano gostitelja in preiti skozi njo, čemur je sledilo aktivno pomnoževanje genoma, vzporedno potovanje genoma skozi jedrno bariero in integriranje v genom gostitelja. Aktivna okužba pa pomeni ravno dejanski vstop virusa v gostitelja in pomnoževanje genoma v gostitelju. Ni dovolj, da se virus nahaja samo na površini celice, v medceličnini ali celo samo naključno na površini organizma. Če virus najdemo na površini organizma, v medceličnini ali na površini celice določenega organizma, še ne pomeni, da je organizem dejanski gostitelj za viruse (Slika 4.10).



Slika 4.10: Filogenetsko drevo virusov klade **Mono-Chu**, ki smo jih našli v genomih in transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo. Endogenizirani virusi so označeni z zvezdico.

4.3.1.2.7 Klada Narna-Levi

Klada Narna-Levi je v filogenetskem drevesu razdeljena na tri monofiletske skupine; Narnavirus, Mitovirus in virusi, sorodni oz. podobni Ourmiavirusu. Našli smo en sam nov virus v transkriptomu organizma *Helicoverpa assulta* (ditrisiji, sovke – Noctuidae) (GBTA01004534), ki je bil uvrščen v skupino Narnavirus. Virus *Helicoverpa assulta* (ditrisiji, sovke – Noctuidae) (GBTA01004534) je soroden glavnemu virusu »Fusarium poae narnavirus 1«, ki je bil izoliran iz askomikotne glive *Fusarium poae*. Proteinsko zaporedje RdRp virusa je 31 % podobno proteinskemu zaporedju RdRp *Fusarium poae* narnavirus 1. Genom tega novega narnavirusa kodira samo en sam nestrukturni ORF z RdRp (**Dodatek B; B.2, Slika B.18**).

4.3.1.2.8 Klada Tombus-Noda

Klada Tombus-Noda je sestavljena iz dveh skupin virusov, Nodaviridae in Tombusviridae. Znotraj skupine Tombusviridae se je uvrstila tudi skupina Carmotetraviridae. Našli smo dva virusa, ki sta bila uvrščena v skupino Tombusviridae; enega v transkriptomu organizma *Ostrinia nubilalis* (prosenca večča) (GAVD01019588) in enega v genomu organizma *Calycopis cerops* (ditrisiji, dnevni metulji – Papilionidae) (LUGF01036384). Ker je bil tombusvirus najden v genomu dnevnega metulja *Calycopis cerops*, vemo, da je omenjen

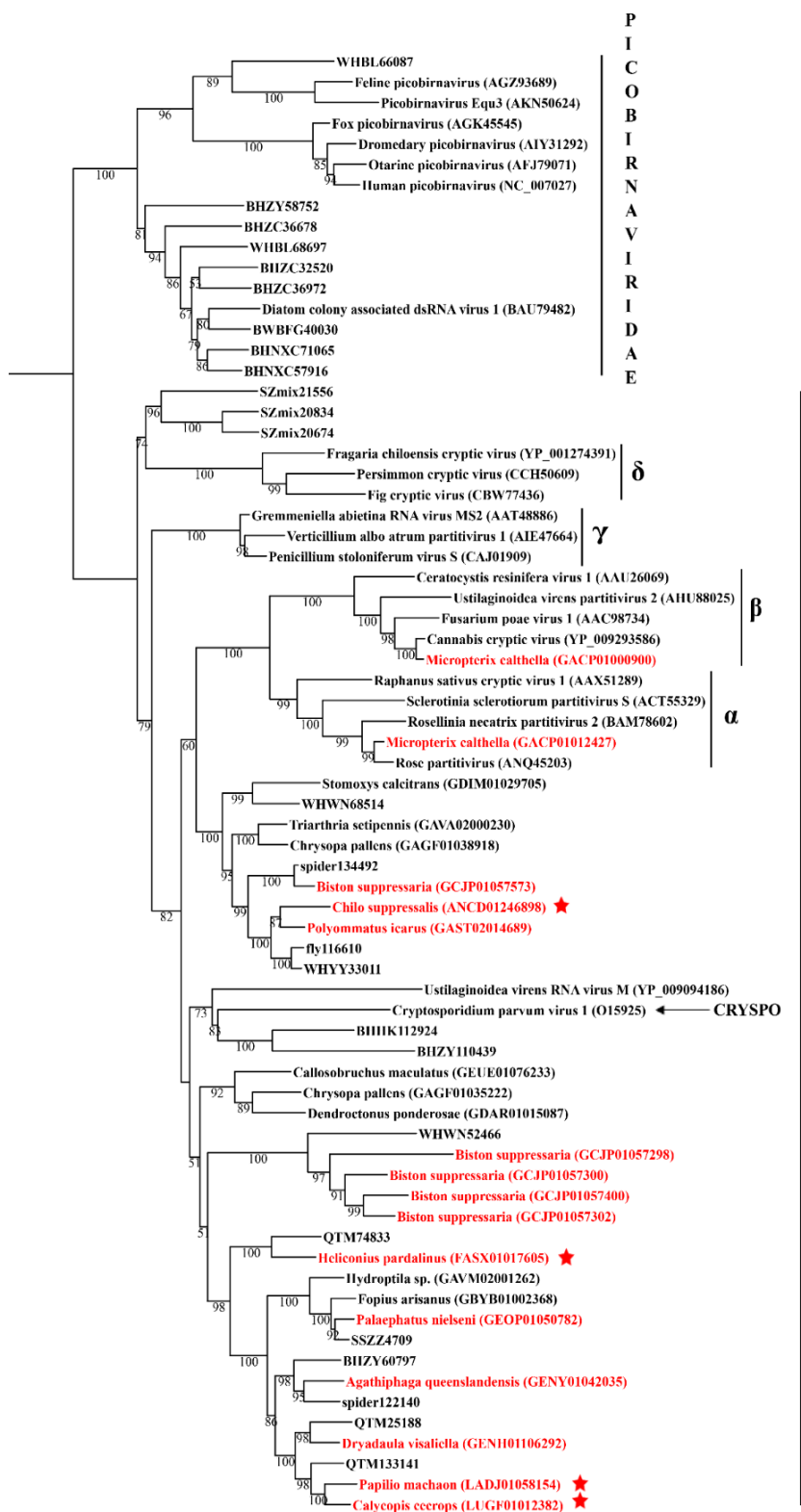
organizem dejanski gostitelj tudi za tobusviruse. Gre za dve zaporedji, ki sta sorodni »*Diaphorina citri* – associated C virusu«, segmentiranemu virusu, ki je bil izoliran iz žuželke *Diaphorina citri* (Hemiptera, enakokranci). Proteinsko zaporedje virusa, najdenega v transkriptomu *Ostrinia nubilalis* (proseni vešča) (GAVD01019588), je 59 % podobno proteinskemu zaporedju RdRp »*Diaphorina citri* – associated C virusa«, medtem ko je proteinsko zaporedje RdRp, najdeno v transkriptomu organizma *Calycomis cerops* (ditrisiji, dnevni metulji – Papilionidae) (LUGF01036384), 55 % podobno proteinskemu zaporedju RdRp »*Diaphorina citri* – associated C virusa« (Dodatek B; B.2, Slika B.19).

4.3.1.2.9 Klada Orthomyxo

Klada Orthomyxo je v filogenetski analizi ločena na dve skupini. V prvi so Isavirus, Influenzavirus in Thogotovirus. V drugi so Quarantavirusu podobni virusi. V skupino Quarantavirusu podobnih virusov spadajo trije novi virusi, dva iz transkriptomov organizmov *Zygaena fausta* (ovnič) in *Eogystia hippophaecolus* (ditrisiji, lesovrti – Cossoidea) in eden iz genoma organizma *Melita cinxia* (pikasti pisanček). Pikasti pisanček (*Melita cinxia*) je torej dejanski gostitelj virusa klade Orthomyxoviridae (APLT01048039). Vsi trije virusi so najbolj sorodni »Wuhan mosquito virus 7 (mosHB235925)« iz transkriptoma komarjev iz kraja Hubei v okviru raziskave Shi in sod. (2016a) (Dodatek B; B.2, Slika B.20).

4.3.1.2.10 Klada Partiti-Picobirna

Klada Partiti-Picobirna oblikuje dve večji ločeni skupini; Partitiviridae in Picobirnaviridae. Vsi novi RNA virusi so bili uvrščeni v skupino Partitiviridae. V transkriptomih metuljev in več smo našli 11 partitivirusov, v genomih metuljev in več pa štiri partitiviruse. Virus, najdena v transkriptomu organizma *Micropterix calthella*, spadata v skupino alfa in betapartitivirusov. Genom virusa, najdenega v transkriptomu *Micropterix calthella* (kalužni prametuljček) z referenčno številko GACP01000900, je najbolj soroden genomu »Cannabis cryptic virusa«. Proteinsko zaporedje RdRp virusa, najdenega v transkriptomu organizma *Micropterix calthella* (GACP01000900), je 81 % podoben »Cannabis cryptic virusu«. »Cannabis cryptic virus« je bil izoliran iz navadne konoplje (*Cannabis sativa*). To pomeni, da gre dejansko za genom zelo podobnega rastlinskega virusa (Cannabis cryptic virus), ki je 1) bil naključno prisoten na kalužnem prametuljčku (*Micropterix calthella*) ali 2) ga kalužni prametuljček kot opraševalec in vektor prenese na rastlino. Genom virusa iz transkriptoma *Micropterix calthella* (kalužni prametuljček) (GACP01012427) je najbolj soroden genomu »Rose partitivirus«, ki je bil izoliran iz vrtnice – kultivarja Phyllis Bide. Proteinsko zaporedje virusa iz transkriptoma *Micropterix calthella* (kalužni prametuljček) (GACP01012427) je 77 % identično rastlinskemu virusu Rose partitivirus, kar pomeni, da gre za zelo podoben virus. Virusi, najdeni v genomih organizmov *Chilo suppressalis* (ditrisiji, travniške vešče-Crambidae), *Heliconius pardalinus* (ditrisiji, dnevni metulji-Papilionoidea; pisančki Nymphalidae), *Papilio machaon* (lastovičar) in *Calycomis cerops* (ditrisiji, dnevni metulji – Papilionidae), so pokazali, da so omenjeni organizmi dejanski gostitelji za pripadajoče partitiviruse, najdene v njihovih genomih. Izmed odkritih genomov partitivirusov v transkriptomih metuljev in več smo zasledili dva celotna genoma – *Dryadula visaliella* (GENH01106292) in *Micropterix calthella* (RNA1; GACP01000900, RNA2; GACP01021597) (Slika 4.11).



Slika 4.11: Filogenetsko drevo virusov klade **Partiti-Picobirna**, ki smo jih našli v genomih in transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo. Endogenizirani virusi so označeni z zvezdico.

4.3.1.2.11 Klada Permutotetra

Klado Permutotetra lahko razdelimo na dve sestrski skupini s po dvema podskupinama. V prvi so Niehaus virusu podobni virusi, v drugi so Thosea asigna virusu podobni virusi, v tretji so Newfield virusu podobni virusi in v četrti so Drosophila A virusu in Daeseongdong virusu 2 podobni virusi. V transkriptomih metuljev smo našli dva virusa, enega v transkriptomu organizma *Eogystia hippophaecolus* – (GFBH01042206) in drugega v transkriptomu organizma *Athetis lepigone* – (GARC01021157). Glede na našo filogenetsko analizo sta najbolj sorodna »Niehaus virusu« (Dodatek B; B.2, Slika B.21).

4.3.1.2.12 Klada Picorna-Calici

Klada Picorna-Calici je razdeljena na skupine Kelp fly virusu podobni virusi, Posavirusu podobni virusi, Arivirusu podobni virusi, Nora virusu podobni virusi, »aquatic picornavirales«, Dicistroviridae in Iflavididae. Virusi, uvrščeni v klado Picorna-Calici, ki smo jih našli v transkriptomih metuljev in več (Lepidoptera), spadajo med Iflavididae ali Nora virusu podobne viruse. Našli smo deset pikornavirusov, sedem je bilo iflavirusov in trije nora virusi (Slika 4.12).

4.3. Filogenetska analiza novih RNA virusov in endogeniziranih RNA virusov pri raziskanih skupinah nevretenčarjev

113



Slika 4.12: Filogenetsko drevo virusov klade **Picorna-Calici**, ki smo jih našli v transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo.

4.3.1.2.13 Klada Reo

Klada Reo je klada raznolikih segmentiranih (10–12 segmentov) virusov. V transkriptomih metuljev smo našli tri reoviruse; enega v transkriptomu *Thaumetopoea pityocampa* (pinijev sprevodni prelec) (GBZB01001266), enega v transkriptomu *Athetis lepigone* (ditrisiji, sovke – Noctuidae) (GARB01026632) in enega v transkriptomu *Helicoverpa armigera* (južna plodovrtka) (GBDM01000458). Reovirus iz organizma *Thaumetopoea pityocampa* je najbolj soroden »Rice dwarf virusu«, izoliranemu iz lisičjega repa (*Alopecurus aequalis*), vendar je proteinsko zaporedje njegove RdRp samo 27 % identično omenjenemu virusu. Znotraj klade Reo smo v transkriptomih metuljev našli tudi »Thaumetopoea pityocampa cypovirus 5« (ki je bil izoliran iz pinijevega sprevodnega prelca – *Thaumetopoea pityocampa*) v transkriptomu južne plodovrtke (*Helicoverpa armigera*). Gre za isti virus kot »Heliothis armigera cypovirus 5«, ki je bil izoliran iz metulja *Heliothis armigera* (ditrisija, sovke – Noctuidae). V tem primeru lahko trdimo, da so metulji *Thaumetopoea pityocampa*, *Helicoverpa armigera* in *Heliothis armigera* dejanski gostitelji omenjenega virusa, saj vsi sodijo v naddružino sovk (Noctuoidea) (**Dodatek B; B.2, Slika B.23**).

4.3.1.2.14 Klada Toti-Chryso

Klada Toti-Chryso zajema dve glavni skupini, prvo tvorijo »Giardiavirusu podobni virusi«, drugo tvorijo Victorivirus, Leishmanivirus, Trichomonasvirus, Totivirus in »Chrysovirusu podobni virusi«. Našli smo tri totiviruse v transkriptomih metuljev; enega pri *Bombyx mori* (navadna sviloprejkica) (GBJR) in dva pri *Helicoverpa assulta* (ditistriji, sovke – Noctuidae) (GBTA). Totivirus iz navadne sviloprejkice je najbolj soroden kožekrilcu *Camponotus yamaokai*. Njuni RdRp sta 43 % identični. Hkrati smo našli totivirus, integriran v genom *Calycopis cerops*, kar pomeni, da je omenjen organizem dejanski gostitelj virusa, saj endogenizacija nakazuje na pretekle aktivne okužbe (**Dodatek B; B.2, Slika B.22**).

4.3.1.3 Najstarejše skupine žuželk: skakači (Collembola), dvorepke (Diplura), ščetinorepke (Zygentoma), Monocondylia

Pregledali smo 16 transkriptomskih in štiri genomske knjižnice najstarejših skupin žuželk oz. t. i. prazuželk. Med najstarejše skupine žuželk sodijo skakači (Collembola), dvorepke (Diplura) in znotraj pravih žuželk (Insecta) 1) skupina Monocondylia in 2) skupina ščetinorepk (Zygentoma) znotraj skupine Dicondylia. Gre za skupino nekrlatih žuželk (Apterygota), ki še niso bile analizirane glede raznolikosti RNA virusov. Krilate žuželke (Pterygota) so druga glavna skupina pravih žuželk (Insecta) znotraj skupine Dicondylia. V filogenetskih drevesih smo skupino skakačev (Collembola) označili z rdečo, skupino dvorepk (Diplura) z vijolično, skupino ščetinorepk (Zygentoma) z modro in skupino Monocondylia z zeleno barvo.

4.3.1.3.1 Klada Mono-Chu

Klada Mono-Chu je podobno kot pri kožekrilcih, metuljih in veččah razdeljena na tri monofiletske skupine: Rhabdoviridae, Nyamiviridae in Chuvirus. V genomu prazuželke – dvorepke (Diplura) *Catajapox aquilonaris* smo zasledili virus, uvrščen v skupino Chuvirusov znotraj klade Mono-Chu. Catajapox je dejanski gostitelj omenjenega virusa. Genom je najbolj soroden zaporedju »Shayang fly virus 1«, vendar gre za divergentno zaporedje. Zaporedji RdRp sta okrog 30 % identični (**Dodatek B; B.3, Slika B.25**).

4.3.1.3.2 Klada Narna-Levi

Pri kladi **Narna-Levi** nismo našli levivirusov, zato smo naredili filogenetsko drevo samo za skupino Narnaviridae. Drevo se loči na tri večje skupine: Narnavirus, Mitovirus in Ourmiavirusu podobne viruse. V transkriptomu skakača (Collembola) *Tetradontophora bielensis* smo našli narnavirus, ki je najbolj podoben »Beihai narna-like virusu 23« (BHZY57139). Zaporedji RdRp obeh virusov sta 40 % identični (**Dodatek B; B.3, Slika B.26**).

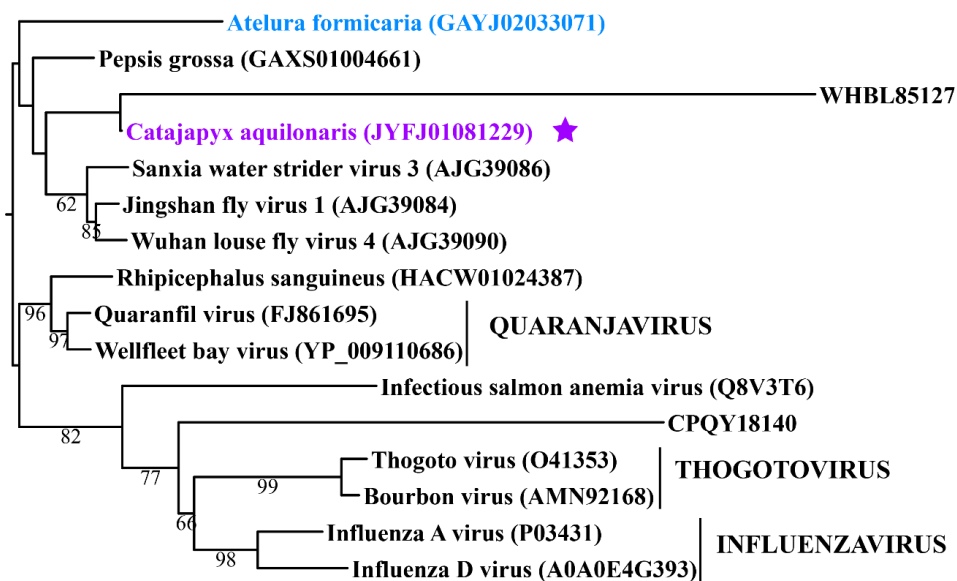
4.3.1.3.3 Klada Hepe-Virga

V okviru klade **Hepe-Virga** smo v genomih in transkriptomih prazuželk našli samo negevirus. Zato smo se odločili za podrobnejšo filogenetsko analizo v okviru negevirusov.

Tri negevirusi smo našli v transkriptomih skakačev (*Collembola*), enega v transkriptomu dvorepke (*Diplura*) *Campodea augens* (GAYN) in enega v genomu *Catajapyx aquilonaris* (JYFJ). Filogenetsko drevo negevirusov razdeli v štiri ločene skupine: 1) Negev virusu in Loreto virusu podobne viruse (skupina I), 2) Tanay virusu in Santana virusu podobne viruse (skupina II), 3) Bofa virusu in Boutonnet virusu podobne viruse (skupina III) in 4) Muthill in Marsac virusu podobne viruse (skupina IV). Razdelitev je skladna z razlikami v genomski strukturi. Virus iz transkriptoma skakača *Tetrodontophora bielanensis* je bil uvrščen v Muthill in Marsac virusu podobne viruse (skupina IV). Vendar gre za divergentno proteinsko zaporedje RdRp, ki je samo 39 % podobno Negev virusu in 34 % podobno proteinskemu zaporedju RdRp Muthill virusa. Virus, najdena v genomu *Catajapyx aquilonaris* in transkriptomu *Campodea augens*, sta bila uvrščena med Bofa virusu podobne viruse (skupina III). Proteinsko zaporedje RdRp virusa iz transkriptoma *Campodea augens* je 49 % podobno proteinskemu zaporedju RdRp Boutonnet virusa. Dve omenjeni zaporedji iz transkriptomov skakačev *Anurida maritima* in *Sminthurus viridis* nista bili uvrščeni v nobeno od skupin Negevirusov. Proteinsko zaporedje RdRp virusa, najdenega v transkriptomu skakača *Anurida maritima* – (GAUE02021637), je 46 % identično zaporedju Negevirusa. Proteinsko zaporedje RdRp virusa iz transkriptoma skakača *Sminthurus viridis* – (GATZ02022882) je 39 % identično zaporedju Negevirusa (**Dodatek B; B.3, Slika B.28**).

4.3.1.3.4 Klada Orthomyxo

Klada Orthomyxo pri prauželkah oblikuje tri ločene skupine. Prvo zajemajo Thogotovirus in Influenzavirus, drugo skupino zajemajo Quarantavirusu podobni virusi in tretjo skupino zajemajo Sanxia water strider virus 3 podobni virusi. V transkriptomu ščetinorepke (*Zygentoma*) male mravlje ribice (*Atelura formicaria*) in v genomu dvorepke (*Diplura*) *Catajapyx aquilonaris* smo našli virusa, ki sta bila uvrščena med Sanxia water strider virus 3 podobne viruse. Zaporedje RdRp virusa, najdenega v transkriptomu male mravlje ribice (*Atelura formicaria*), je le 31 % identično proteinskemu zaporedju RdRp virusa »Sanxia water strider virusu 3« (**Slika 4.13**).

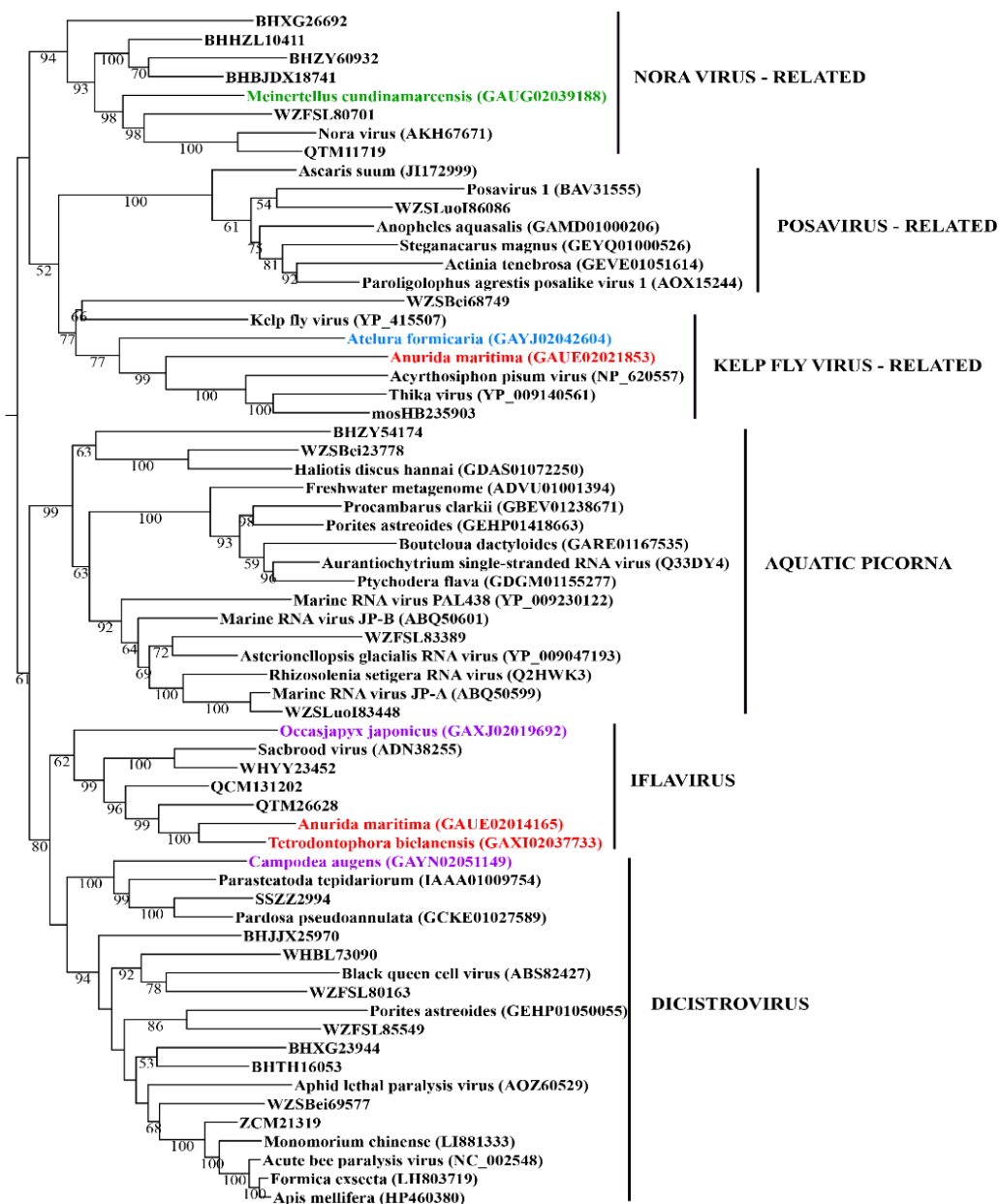


Slika 4.13: Filogenetsko drevo virusov klade **Orthomyxo**, ki smo jih našli v genomih in transkriptomih najstarejših skupin žuželk. Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Skakači (*Collembola*) so označeni z rdečo, dvorepke (*Diplura*) z vijolično, ščetinorepke (*Zygentoma*) z modro in skupina *Monocondylia* z zeleno barvo. Endogenizirani RNA virusi so označeni z zvezdico.

4.3.1.3.5 Klada Picorna-Calici

Klada Picorna-Calici pri prauželkah zajema Nora virusu podobne viruse, Posavirusu podobne viruse, Kelp fly virusu podobne viruse, vodne pikornaviruse (angl. »aquatic picornavirales«) ter skupini Iflaviridae in Dicistroviridae. V transkriptomih prauželk smo

našli sedem virusov uvrščenih v klado Picorna-Calici. En pikornavirus smo našli v transkriptomu *Meinertellus cundinamarcensis* (Monocondylia), spada med Nora virusu podobne viruse in je najbolj soroden z Wenzhou picorna-like virusom 47 (WZFSL80701). Proteinski zaporedji RdRp sta 34 % identični. Tri pikornaviruse smo našli v transkriptomih skakačev (Collembola). Virus iz transkriptoma obmorskega skakača – *Anurida maritima* (GAUE02021853) spada med Kelp-Fly virusu podobne viruse. Proteinsko zaporedje RdRp je 27 % podobno proteinskemu zaporedju Kilifi virusa. Virus iz transkriptomov obmorskega skakača – *Anurida maritima* (GAUE02014165) in orjaškega skakača – *Tetrodontophora bielanensis* (GAXI02037733) sta bila uvrščena med iflaviruse. Proteinsko zaporedje RdRp virusa, najdeno v transkriptomu obmorskega skakača – *Anurida maritima* (GAUE02014165), je 37 % identično »Hubei picorna-like virusu 31« (QTM26628). Proteinsko zaporedje RdRp virusa, najdeno v transkriptomu orjaškega skakača – *Tetrodontophora bielanensis* (GAXI02037733), je 45 % identično »Hubei picorna-like virusu 31« (QTM26628). Našli smo tudi pikornavirus v transkriptomu male mravlje ribice (*Atelura formicaria*) (GAYJ02042604). Proteinsko zaporedje njegovega RdRp je 26 % identično z Arivirusom 2 oz. okrog 25 % identično s katerim koli pikornavirusom. Našli smo tudi dva virusa v transkriptomih dvorepk. Virus iz transkriptoma dvorepke *Occasjapyx japonicus* (GAXJ02019692) je bil uvrščen med iflaviruse. Proteinsko zaporedje njegove RdRp je samo 28 % identično s Sacbrood virusom. Virus, najden v transkriptomu dvrepke *Campodea augens* (GAYN02051149), je bil uvrščen med dicistroviruse. Proteinsko zaporedje njegove RdRp je 37 % identično najbližjim sorodnikom »Hubei tetragnatha maxillosa virus 3« (SSZZ2994) (Slika 4.14).



Slika 4.14: Filogenetsko drevo virusov klade **Picorna-Calici**, ki smo jih našli v transkriptomih najstarejših skupin žuželk. Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Skakači (Collembola) so označeni z rdečo, dvorepke (Diplura) z vijolično, ščetinorepke (Zygentoma) z modro in skupina Monocondylia z zeleno barvo.

4.3.1.3.6 Klada Qinvirus

Klada Qinvirus zajema dve skupini: 1) Xinzhou nematode virus 3 (XZSJSC65579) podobni virusi in 2) Hubei qinvirus-like virus 1 (SCM49583) podobni virusi. Virus, najden v transkriptomu obmorskega skakača – *Anurida maritima* (GAUE02014037), je bil uvrščen v skupino Hubei qinvirus-like virus 1 (SCM49583) podobnih virusov. Proteinsko zaporedje njegove RdRp je 31 % podobno proteinskemu zaporedju RdRp »Hubei qinvirus-like virusa 1« (SCM49583) (**Dodatek B; B.3, Slika B.27**).

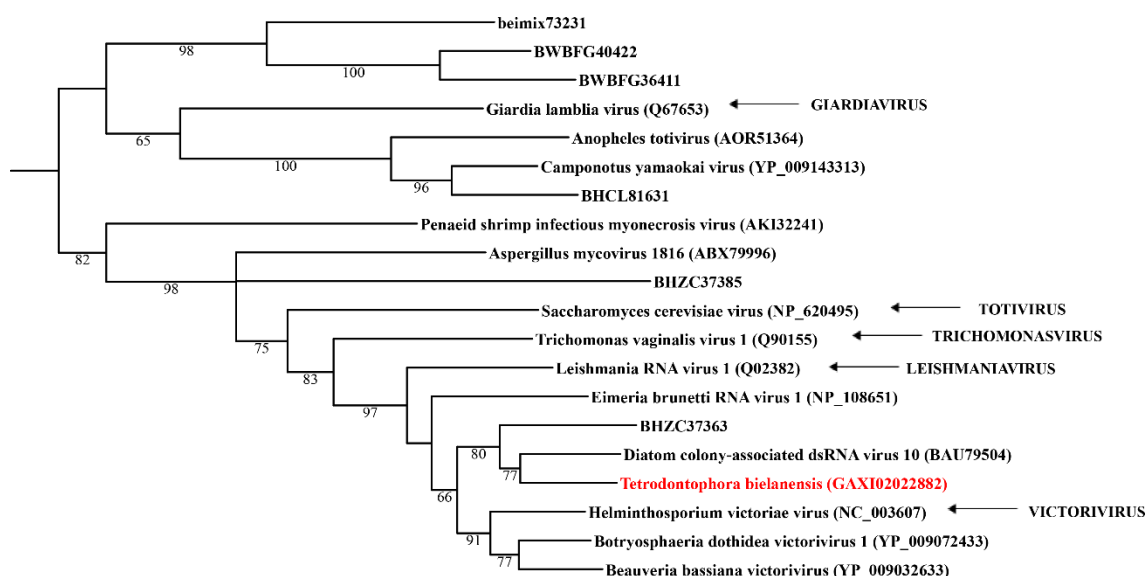
4.3.1.3.7 Klada Tombus-Noda

Klada Tombus-Noda zajema dve ključni skupini: 1) Diaphorina citri-associated C virusu podobni virusi in 2) druge po ICTV (angl. »international committee on taxonom of viruses«) definirane skupine tombusvirusov; Zeavirus, Umbravirus, Tombusvirus, Macanavirus, Avenavirus, Dianthovirus, Betanecrovirus, Aureusvirus, Gallantivirus, Alphanecrovirus,

Panicovirus, Alphacarmovirus. V transkriptomu ščetinorepk (*Zygentoma*) male mravlje ribice (*Atelura formicaria*) in *Tricholepidion gertschi* smo našli dva virusa, ki spadata v skupino »Diaphorina citri-associated C virusu« podobnih virusov. Najbližji sorodnik virusa iz transkriptoma male mravlje ribice (*Atelura formicaria*) (GAYJ02032904) je »Diaphorina citri-associated C virus«. Proteinski zaporedji njunih RdRp sta 52 % identični. Proteinsko zaporedje RdRp virusa, najdenega v transkriptomu *Tricholepidion gertschi* (GAS002037726), je divergentno zaporedje in je 24–27 % identično kateremu koli trenutno znanemu tombusvirusu (**Dodatek B; B.3, Slika B.29**).

4.3.1.3.8 Klada Toti-Chryso

Klada Toti-Chryso tudi tvori dve monofiletski skupini. V prvo so vključeni Giardiavirusu podobni virusi, v drugo pa Totivirus, Trichomonasvirus, Leishmanivirus in Victorivirus. V transkriptomih pražuželk smo našli en sam totivirus, in sicer v transkriptomu orjaškega skakača – *Tetrodontophora bielanensis* (GAXI02022882). Najbližji sorodnik virusa je »Diatom colony-associated dsRNA virus 10«. Proteinski zaporedji njunih RdRp sta 39 % identični (**Slika 4.15**).



Slika 4.15: Filogenetsko drevo virusa klade **Toti-Chryso**, ki smo jih našli v transkriptomih najstarejših skupin žuželk. Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Skakači (Collembola) so označeni z rdečo, dvorepke (Diplura) z vijolično, ščetinorepke (*Zygentoma*) z modro in skupina Monocondylia z zeleno barvo.

4.3.1.3.9 Klada Reo

V transkriptomu organizma *Anurida maritima* smo našli prisoten virus, ki je bil uvrščen v klado Reoviridae. Proteinsko zaporedje RdRp je sorodno Eyach virusu. Zaporedji sta si 39 % identični (**Dodatek B; B.3, Slika B.30**).

4.3.2 Vodni nevretenčarji

4.3.2.1 Raki (Crustacea)

Skupino rakov lahko razdelimo na več razredov: Branchiopoda (raki listonožci), Malacostraca (višji raki), Maxillopoda, Remipedia in Cephalocarida. Med višje rake (Malacostraca) prištevamo tri ključne skupine: Decapoda (desteronožci), Amphipoda (postranice) in Isopoda (enakonožci). Pregledali smo skupno 196 podatkovnih baz rakov – 23 genomskih in 187 transkriptomskih. Znotraj skupine Maxillopoda je bila za naše raziskave ključna skupina Copepoda (ceponožci). V filogenetski analizi smo skupino deseteronožcev (Decapoda) označili z vijolično, skupino postranic (Amphipoda) z modro, skupino enakonožcev (Isopoda) z rdečo in skupino ceponožcev (Copepoda) z zeleno barvo. V transkriptomu skupine Branchiopoda smo našli en sam primer virusa (flavivirus pri

Daphnia sp.), označili smo ga z rjavo barvo. Ravno tako smo našli samo en virus v transkriptomu skupine Branchiura (Maxillopoda) (*Argulus* sp., Picorna-Calici), označili smo ga z magenta barvo.

4.3.2.1.1 Klada Bunya-Arena

Klada Bunya-Arena pri rakih oblikuje tri monofiletske skupine. Prvo predstavljajo Feraviridae, Jonviridae, Peribunyaviridae, Hantaviridae, Tospoviridae in Fimoviridae, drugo tvori skupina Nairoviridae podobnih virusov in tretjo tvori skupina Phenuiviridae podobnih virusov. V transkriptomih rakov smo našli šest bunyavirusov. Dva smo našli v transkriptomu deseteronožcev (Decapoda), dva v transkriptomu postranic (Amphipoda) in dva v transkriptomu enakonožcev (Isopoda). Virus, najden v transkriptomu deseteronožca signalnega raka – *Pacifastacus leniusculus* (GBYW01039086), je bil najbolj soroden zaporedju Whenzhou shrimp virusu 2 (shrimp6186). Proteinski zaporedji njunih RdRp sta 51 % identični. Virus, najden v transkriptomu postranice *Talitrus saltator* (GDUJ01083694), je bil najbolj soroden Wuhan pillworm virus 1 (WHSFII20185). Proteinski zaporedji njunih RdRp sta 49 % identični. Virus, najden v transkriptomu postranice *Hyalelloopsis stebbingi* (GEQJ01053822), je divergenten virus. Proteinsko zaporedje RdRp virusa je 34 % identično proteinskemu zaporedju RdRp Gouleako virusa. Virus, najdena v transkriptomih enakonožcev *Bragasellus peltatus* (HAEN01054030) in *Proasellus coxalis* (HAET01052131), sta sorodna zaporedju »Beihai bunya-like virus 1« (BHTSS17606). Proteinsko zaporedje RdRp virusa, najdenega v transkriptomu *Bragasellus peltatus* (HAEN01054030), je 53 % identično »Beihai bunya-like virus 1« (BHTSS17606). Proteinsko zaporedje RdRp virusa, najdenega v transkriptomu *Proasellus coxalis* (HAET01052131), je 54 % identično »Beihai bunya-like virus 1« (BHTSS17606). Pri skupini Shi in sod. je bilo največ virusov najdenih v knjižnicah rakov, uvrščenih med Nairovirusom podobne viruse (**Dodatek B; B.4, Slika B.31**).

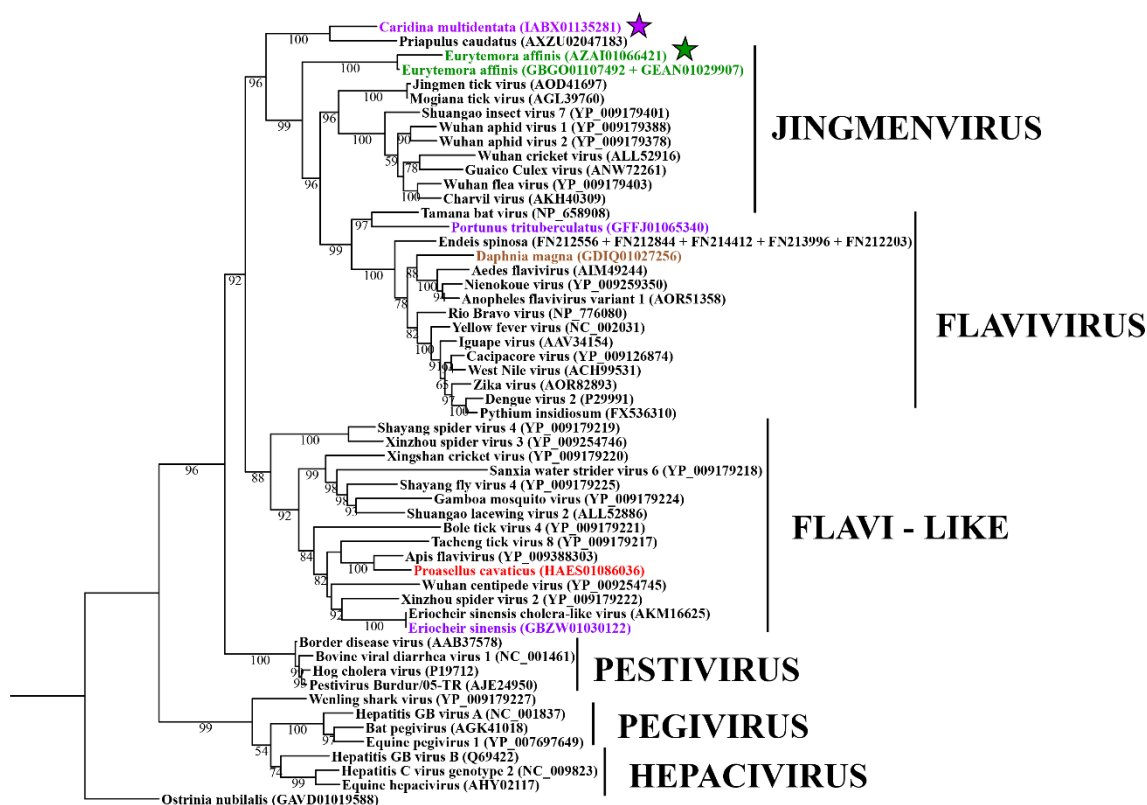
4.3.2.1.2 Klada Birna

Klada Birna oblikuje dve ločeni skupini: 1) »Drosophila X virusu« podobni virusi (Entomobirnavirusi) in 2) Aquabirnavirus, Avibirnavirus in Blosnavirus. V transkriptomu ceponožca (Copepoda) *Eurytemora affinis* (GBGO01030594) smo našli virus, ki spada med Entomobirnavirusu podobne viruse. Proteinsko zaporedje RdRp je 46 % identično proteinskemu zaporedju Drosophila X virusu, izoliranemu iz *Drosophila melanogaster* (**Dodatek B; B.4, Slika B.32**).

4.3.2.1.3 Klada Flavi

Klada Flavi je ločena na šest skupin: Jingmenvirus, Flavivirus, Flavivirusus podobni virusi (angl. »Flavi-like«) (Shi in sod., 2016b), Pestivirus, Hepacivirus in Pegivirus. Filogenetsko drevo flavivirusov, najdenih v transkriptomih rakov, smo koreninili pri proteinskemu zaporedju RdRp, prevedenemu iz virusnega genoma, najdenega v transkriptomu *Ostrinia nubilalis* (GAVD01019588). Gre za virus, katerega proteinsko zaporedje RdRp je 52 % identično proteinskemu zaporedju RdRp virusu »Diaphorina citri associated C virusu«, ki sodi med tobusviruse (Tombusviridae). Ker smo drevo vključili v kontekst najbližjih sorodnikov skupine Flaviviridae, ki so Tombusviridae, lahko sklepamo o evolucijski starosti posamezne skupine flavivirusov. Glede na naše filogenetsko drevo sta Flavivirus in Jingmenvirus, ki oblikujeta sestrski kladi, najmlajši monofiletski skupini Flavivirusov. Evolucijsko starejša skupina so Flavivirusom podobni virusi. Od flavivirusom podobnih virusov sta najbolj divergentni skupini pestivirusi in skupina, ki tvori dve sestrski kladi – Pegivirus in Hepacivirus. V genomih in transkriptomih rakov smo našli sedem virusov. Trije flavivirusi so bili odkriti v genomih in transkriptomih deseteronožcev (Decapoda), od tega eden v genomu *Caridina multidentata* (IABX01135281) in dva v transkriptomih *Portunus tritoberculatus* (GFFJ01065340) in *Eriocheir sinensis* (GBZW01030122). Virus, odkrit v genomu *Caridina multidentata* (IABX01135281), verjetno predstavlja novo, evolucijsko starejšo obliko skupine Flaviviridae in Jingmenvirusa, ki se ni uvrstila med flavivirusom podobne viruse. Glede na to, da je bil virus odkrit v genomu, sklepamo, da je *Caridina multidentata* dejanski gostitelj tega flavirusa. Poleg omenjenega se je celoten genom virusa ohranil v genomu raka. Virus *Caridina multidentata* (IABX01135281) je eden izmed redkih RNA virusov, katerih endogeniziran genom smo odkrili v celoti ohranjen. Najbližji sorodnik virusa, odkritega v transkriptomu *Portunus tritoberculatus* (GFFJ01065340), je »Tamana bat virus«. Njuni RdRp sta 48 % identični. Dva virusa sta bila odkrita pri

ceponožcu (Copepoda) *Eurytemora affinis*, eden v genomu (AZAI01066421) in eden v transkriptomu (GBGO01107492 + GEAN01029907). Zaporedji njunih RdRp sta si med seboj 100 % identični. Tako kot virus, najden v genomu *Caridina multidentata* (IABX01135281), predstavljata virusa iz *Eurytemora affinis* (AZAI01066421) in *Eurytemora affinis* (GBGO01107492 + GEAN01029907) starejša virusa, kot sta skupini Flavivirus in Jingmenvirus oz. predstavljata starejšo, predhodno obliko, ki ni bila uvrščena med flavivirusom podobne viruse. RdRp virusa, najdenega v genomu *Eurytemora affinis* (AZAI01066421), je 29 % identičen virusu rumene mrzlice (skupina Flavivirus) oz. 30 % identičen proteinskim zaporedjem RdRp katerega koli trenutno znanega flavivirusa. Virus, najden v transkriptomu *Daphnia magna* (GDIQ01027256), spada med flaviviruse (skupina Flavivirus). Najbližji sorodniki so »Nienokoue virus«, »Aedes flavivirus« in »Anopheles flavivirus«, ki so razširjeni pri komarjih (*Culicidae* sp., *Aedes* sp., *Anopheles* sp.). RdRp virusa, najdenega v transkriptomu *Daphnia magna* (GDIQ01027256), je okrog 50 % identičen najbližjim sorodnikom oz. 51 % identičen »Nienokoue virusu«. Virus, najden v transkriptomu enakonožca *Proasellus cavaticus* (HAES01086036), spada med flavivirusom podobne viruse. Najbolj soroden je »Apis flavivirus«, ki je bil izoliran iz čebele *Apis mellifera*. Njuni RdRp sta 50 % identični (Slika 4.16).



Slika 4.16: Filogenetsko drevo virusov klade **Flavi**, odkritih v genomih in transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Skupina rakov enakonožcev (Isopoda) je označena z rdečo, skupina ceponožcev (Copepoda) je označena z zeleno, skupina rakov postranic (Amphipoda) je označena z modro, skupina deseteronožcev (Decapoda) je označena z vijolično, skupina Branchiopoda (*Daphnia*) je označena z rjavim odtenkom. Endogenizirani virusi so označeni z zvezdico.

4.3.2.14 Klada Hepe-Virga

Klada Hepe-Virga vključuje skupine: 1) Endornaviridae, 2) Tymovirales, 3) hepevirusom podobni virusi (angl. »hepe-like«), 4) Hepeviridae, 5) Benyviridae, 6) Bromoviridae, Closteroviridae, Idaeovirus, 7) Togaviridae, 8) Virgaviridae in 9) Negevirus (Shi in sod., 2016a). Izmed omenjenih skupin so skupine Endornaviridae dsRNA, Benyviridae, Bromoviridae, Closteroviridae, Idaeovirus, Virgaviridae in Tymovirales razširjene med

glivami in rastlinami. V genomih in transkriptomih rakov smo našli 11 virusov, uvrščenih v klado Hepe-Virga. Pet virusov smo našli v transkriptomih postranic (Amphipoda). Štiri spadajo med hepevirusom podobne viruse, eden pa v skupino Togaviridae. Tri viruse smo odkrili v transkriptomih deseteronožcev (Decapoda). Od teh eden spada v skupino hepevirusom podobnih virusov, dva pa v skupino Negevirusov. Dva virusa smo našli v transkriptomih enakonožcev (Isopoda). Od teh eden spada med togaviruse, drugi pa med negeviruse. Odkrili smo tudi virus v transkriptomih ceponožca *Caligus rogercresseyi*, ki spada med negeviruse. Od skupno 11 virusov jih pet spada med hepevirusom podobne viruse, dva med togaviruse in štiri med Negeviruse. Hepevirusom podobni virusi so torej najbolj pogosta virusna skupina pri rakih (Crustacea), sledijo ji Negevirusi. Podobni so tudi rezultati skupine Shi in sod. (2016a). Od skupno 25 virusov virusne klade Hepe-Virga, najdenih v transkriptomih rakov, je bilo 19 virusov uvrščenih v skupino podobnih virusov, dva med negeviruse, dva med togaviruse in dva med endornaviruse. Negevirusi in Togavirusi sta bili najbolj bogato razširjeni virusni skupini v okviru klade Hepe-Virga pri žuželkah. Pri togavirusih je virus, odkrit v transkriptomu *Proasellus assaforensis*, najbolj soroden »Beihai charybdis crab virusu 1« (BHXun33339), virus, ki je bil odkrit v transkriptomu *Gammarus cheuvreuxi* – (GFCV01053886), pa je najbolj soroden »Wenling toga-like virusu« (WLJQ101932). Tako »Beihai charybdis crab virus 1« (BHXun33339) kot tudi »Wenling toga-like virus« (WLJQ101932) izvirata iz transkriptomov rakov. Pri hepevirusu podobnih virusih je virus, odkrit v transkriptomu *Eulimnogammarus violaceus*, najbolj soroden »Wenzhou shrimp virusu 3« (shrimp12824), virus, najden v transkriptomu *Pontastacus leptodactylus* (GAFS01001332), pa je najbolj podoben virusu »Beihai hepe-like virus 4« (BHZY60842, YYSZX17852), virusu, ki je bil najden hkrati v transkriptomu rakov deseteronožcev in mehkužcev. Pri skupini Decapoda gre za skupino rakov, ki so naselili kopno, sladke in slane vode. Knjižnica YYSZX je bila narejena iz transkriptoma raka rodu *Portunus sp.*, plavajočih rakovic, ki živijo tako v morju kot v celinskih vodah. Verjetno je šlo za naključni stik raka z mehkužcem v istem habitatu, pri čemer je prišlo do prenosa virusa »Beihai hepe-like virus 4«. Drugi primer hepevirusom podobnih virusov smo našli v transkriptomu *Omatogammarus flavus* (GEQS01054466). Virus, najdena v transkriptomih deseteronožcev *Eriocheir sinensis* in *Callinectes sapidus*, sta najbolj podobna virusu »Sanxia atyid shrimp virus 1« (SXXX35475), ravno tako izolirana iz transkriptoma raka, tokrat rodu *Atyidae sp.* (sladkovodne kozice) v okviru skupine Shi in sod. (2016a). Uvrščena sta bila v skupino Negevirusov. Praktično vsi virusi, ki smo jih odkrili v transkriptomih, so najbolj sorodni virusom, katerih genomi so bili izolirani iz transkriptomov rakov v okviru skupine Shi in sod. (2016a) (**Dodatek B; B.4, B.33**).

4.3.2.15 Klada Luteo-Sobemo

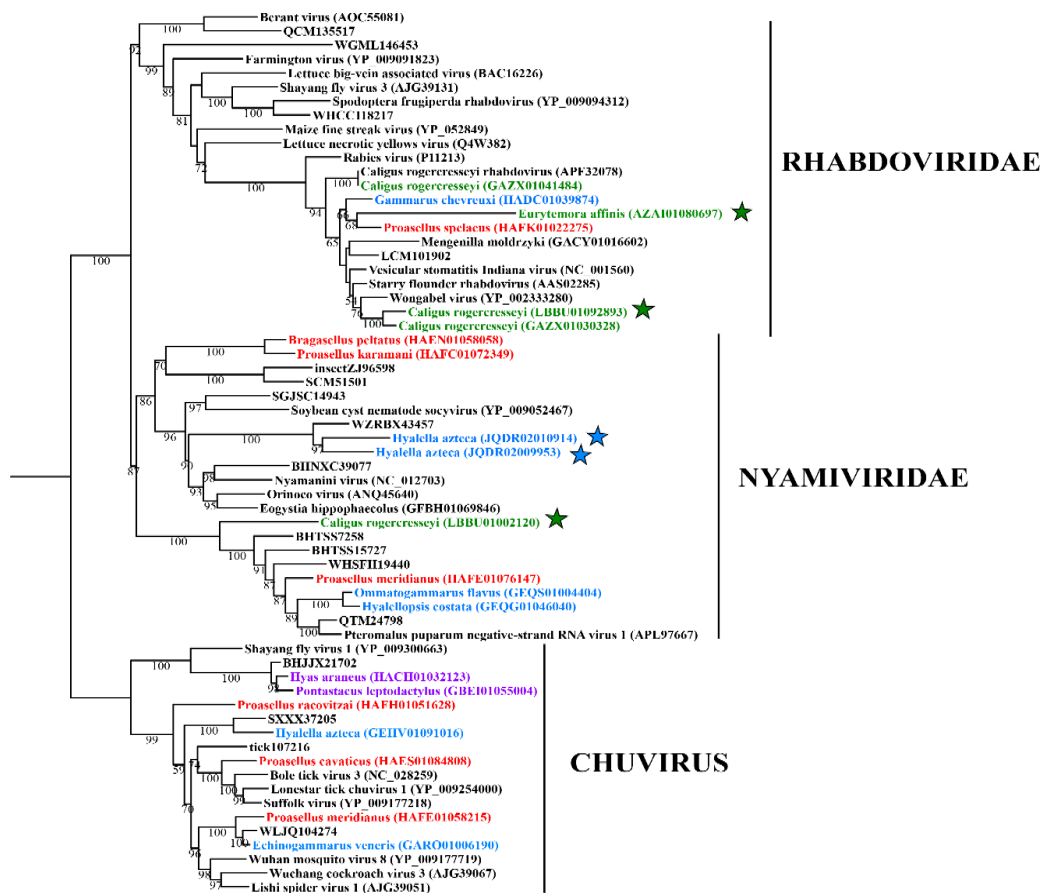
Klada Luteo-Sobemo je pri rakih razdeljena na štiri monofiletske skupine. V prvo sodijo »Carascovirusu podobni virusi«, v drugo sodijo »Ixodes scapularis associated virus 1« podobni virusi, v tretjo »Humaita-Tubiacanga virusu« podobni virusi in v četrto »Wenling sobemo-like virusu 2« (WLJQ95081) podobni virusi. V transkriptomih rakov smo odkrili šest virusov, ki so bili uvrščeni v klado Luteo-Sobemo, od tega sta bila dva virusa odkrita v transkriptomih ceponožcev (Copepoda) in dva v transkriptomih enakonožcev (Isopoda). Virus iz transkriptoma *Proasellus parvulus* (majceni vodni osliček) (HAFG01087285) je najbolj soroden virusu, ki je bil izoliran iz transkriptoma kačjih pastirjev (Odonata) – »Hubei sobemo-like virus 45« (QTM25680). Zaporedji RdRp sta 45 % identični. Virus, odkrit v transkriptomu raka ceponožca *Eurytemora affinis* (GBGO01020472), je najbolj soroden virusu »Sanxia sobemo-like virus 3« (SXSSP2531), izoliranemu iz transkriptoma reda Hemiptera – vodnih drsalcev. Zaporedji RdRp sta 42 % identični. Virus, izolirani iz transkriptomov enakonožcev *Proasellus escolai* (HAEV01103964), *Proasellus parvulus* (HAFG01115646) in *Bragasellus peltatus* (HAEN01054361), so najbolj sorodni Wenling sobemo-like virusu 2 (WLJQ95081), ki je bil v raziskavi Shi in sod. (2016a) odkrit v mešani transkriptomski knjižnici rakov WLJQ. Virus *Proasellus escolai* (HAEV01103964), *Proasellus parvulus* (HAFG01115646) in *Bragasellus peltatus* (HAEN01054361) oblikujejo svojo podskupino v monofiletski skupini »Wenling sobemo-like virusu 2« (WLJQ95081) podobnih virusov (**Dodatek B; B.4, Slika B.34**).

4.3.2.16 Klada Mono-Chu

Klada Mono-Chu je pri rakih (Crustacea) razdeljena na tri skupine: Rhabdoviridae, Nyamiviridae in Chuvirus. V genomih in transkriptomih rakov smo odkrili 21 genomov RNA

virusov. Od teh je bilo sedem pri postranicah (Amphipoda) (pet v transkriptomih, dva v genomih), sedem v transkriptomih enakonožcev (Isopoda), pet pri ceponožcih (Copepoda) (tri v genomih, dva v transkriptomih) in dva v transkriptomih deseteronožcev (Decapoda). Od virusov, ki so bili odkriti v postranicah (Amphipoda), eden spada v skupino Rhabdoviridae, štirje v skupino Nyamiviridae in dva v skupino Chuvirusov. Med štirimi virusi pri postranicah (Amphipoda), ki spadajo v skupino Nyamiviridae, sta bila dva odkrita v genomu organizma *Hyaella azteca* (JQDR02010914) in (JQDR02009953). Proteinski zaporedji njunih RdRp sta med seboj 46 % identični, hkrati pa sta najbolj sorodni virusu, izoliranemu iz transkriptoma rakov deseteronožcev (Decapoda) rodu *Charybdis*–Wenzhou crab virus 1 (WZRBX43457). Proteinski zaporedji RdRp virusa, odkritega v genomu *Hyaella azteca* (JQDR02010914) in Wenzhou crab virusa 1 (WZRBX43457) sta 46 % identični. Virus, odkrit v transkriptomih *Ommatogammarus flavus* (GEQS01004404) in *Hyalellopsis costata* (GEQG01046040), sta najbolj sorodna s »Pteromalus puparum negative-stranded RNA virus 1« (izoliranim iz kožekrilca *Pteromalus puparum*) in virusom, izoliranim iz knjižnice QTM – z raznolikimi RNA virusi najbolj bogato knjižnico v okviru raziskave Shi in sod. (2016a) – »Hubei rhabdo-like virus 6« (QTM24798). Proteinski zaporedji RdRp virusov *Ommatogammarus flavus* (GEQS01004404) in *Hyalellopsis costata* (GEQG01046040) sta 78 % identični. Proteinsko zaporedje RdRp virusa *Ommatogammarus flavus* (GEQS01004404) je s proteinskim zaporedjem RdRp »Pteromalus puparum negative-stranded RNA virus 1« 48 % identično. Med virusi, odkritimi v transkriptomih postranic (Amphipoda) in uvrščenimi v skupino Chuvirus, je bil en genom virusa odkrit v transkriptomu *Hyaella azteca* (GEHV01091016) in en virus v transkriptomu *Echinogammarus veneris* (GAR001006190). Proteinsko zaporedje RdRp virusa *Hyaella azteca* (GEHV01091016) je najbolj sorodno proteinskemu zaporedju RdRp »Sanxia atyid shrimp virus 4« (SXXX37205), izoliranemu iz deseteronožca *Atyidae* sp. Proteinski zaporedji RdRp sta 55 % identični. Proteinsko zaporedje RdRp virusa *Echinogammarus veneris* (GAR001006190) je najbolj sorodno proteinskemu zaporedju RdRp »Beihai chuvirus-like virus 2« (WLJQ104274) iz mešane transkriptomske knjižnice rakov, narejene v raziskavi Shi in sod. (2016a). Zaporedji sta 68 % identični. Med genomi virusov, odkritimi v genomih in transkriptomih ceponožcev (Copepoda), štirje spadajo v skupino Rhabdoviridae in eden v skupino Nyamiviridae. Med štirimi virusi, odkritimi v knjižnicah ceponožcev in uvrščenimi v skupino Rhabdoviridae, sta bila dva genoma virusov odkrita v genomih organizmov *Eurytemora affinis* (AZAI01080697) in *Caligus rogercresseyi* (LBBU01092893). Proteinsko zaporedje virusa *Caligus rogercresseyi* (LBBU01092893) je 71 % identično proteinskemu zaporedju RdRp virusa, odkritega v transkriptomu iste vrste organizma – *Caligus rogercresseyi* (GAZX01030328). Zaporedji sta najbolj sorodni »Wongabel virusu«, izoliranemu iz *Culicoides austropalpalis* (Diptera). Proteinsko zaporedje RdRp »Wongabel virusa« je 62 % identično proteinskemu zaporedju virusa *Caligus rogercresseyi* (LBBU01092893). Zaporedje virusa, odkritega v genomu organizma *Eurytemora affinis* (AZAI01080697), se grupira skupaj z virusoma *Gammarus chevreuxi* (HADC01039874) in *Proasellus spelaeus* (HAFK01001022275). Gre za skupino divergentnih virusov, ki so sicer sorodni »Wongabel virusu« in »Vesicular stomatitis Indiana virusu«, vendar oblikujejo lastno skupino. Virus iz genoma ceponožca *Caligus rogercresseyi* (LBBU01002120) se poveže z »Pteromalus puparum negative-stranded RNA virusom 1«, vendar gre za divergenten genom. Med sedmimi genomi virusov, odkritih v transkriptomih enakonožcev (Isopoda), so vsi virusi divergentni. Virus *Proasellus spelaeus* (HAFK01001022275) se grupira skupaj z *Eurytemora affinis* (AZAI01080697) in *Gammarus chevreuxi* (HADC01039874). Proteinski zaporedji RdRp virusov *Bragasellus peltatus* (HAEN01058058) in *Proasellus karamani* (HAFK01072349) sta med seboj 66 % identični. Najbolj sorodni sta virusom »Shuangao Fly Virus 2« (insectZJ96598), izoliranim iz mešane transkriptomske knjižnice žuželk, in »Hubei diptera virus 11« (SCM51501), izoliranim iz transkriptomske knjižnice žuželk rodu Diptera v okviru raziskave Shi in sod. (2016a). Podobnost proteinskih zaporedij RdRp med *Bragasellus peltatus* (HAEN01058058) in Shuangao Fly Virus 2 (insectZJ96598) je 32 %. Proteinsko zaporedje RdRp chuvirusa *Proasellus cavaticus* (HAES01084808) je 55 % identično Bole tick virusu 3, izoliranemu iz *Hyalomma asiaticum* (Chelicerata). Proteinsko zaporedje RdRp *Proasellus meridianus* (HAFE01058215) se grupira skupaj z zaporedjema »Beihai chuvirus-like virus 2« (WLJQ104274) in *Echinogammarus veneris* (GAR001006190). Genoma virusov, najdenih v transkriptomih deseteronožcev *Hyas araneus* (HACH01032123) in *Pontastacus leptodactylus* (GBEI01055004), sta se uvrstila med chuviruse in se grupirata skupaj z »Beihai hermit crab virus 3« (BHJX21702), izoliranim v okviru raziskav Shi in sod. (2016a)

iz transkriptoma rakov deseteronožcev *Anomura sp.* Proteinski zaporedji RdRp *Hyas araneus* (HACH01032123) in *Pontastacus leptodactylus* (GBEI01055004) sta med seboj 77 % identični. Proteinski zaporedji RdRp *Hyas araneus* (HACH01032123) in »Beihai hermit crab virus 3« (BHJX21702) sta med seboj 75 % identični (Slika 4.17).



Slika 4.17: Filogenetsko drevo virusov klade **Mono-Chu**, ki smo jih našli v genomih in transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) je z zeleno, postranice (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično. Endogenizirani virusi so označeni z zvezdico.

4.3.2.1.7 Klada Narna-Levi

Klada Narna-Levi oblikuje dve monofiletski skupini: Narnaviridae in Leviviridae. Skupina Narnaviridae oblikuje tri ključne podskupine Narnavirus, Mitovirus in ourmiavirusu podobni virusi. V transkriptomih rakov enakonožcev smo našli genoma dveh narnavirusov *Proasellus karamani* (HAF01092190) in (HAF01105261). Virus *Proasellus karamani* (HAF01092190) je soroden »Phytophthora infestans RNA virus 4«. Monopartitna genomska RNA kodira en sam ORF z RdRp. »Phytophthora infestans RNA virus 4« je bil izoliran iz oomicete (Stramenopiles) *Phytophthora infestans*, ki so jo pridobili iz okuženega paradižnika. Njuni zaporedji RdRp sta 39 % identični. Virus *Proasellus karamani* (HAF01105261) je soroden virusu »Hubei narna-like virus 12« (SZmix21586), odkritemu v transkriptomski knjižnici mehkužcev v raziskavi Shi in sod. (2016a). Gre za eno izmed novih klad, ki se grupirajo skupaj z Ourmiavirusu podobnimi virusi (Shi in sod., 2016a). Levivirus, ki smo ga našli v transkriptomu deseteronožca *Hyas araneus* (HACH01002411), je soroden virusu »Wenling levi-like virus 1« (WLJQ101898), izoliranemu iz mešane transkriptomske knjižnice rakov pri raziskavi Shi in sod. (2016a). Glede na rezultate Shi in sod. (2016a) so virusi, ki sodijo v nove skupine virusov (Shi in sod., 2016a), najbolj razširjene pri rakih in mehkužcih, Narnavirusu podobni virusi pri rakih in žuželkah,

mitovirusi pri glivah in levivirusi pri rakah in mehkužcih. Zato ne preseneča, da levivirusov pri žuželkah in pražuželkah nismo našli (**Dodatek B; B.4, Slika B.35**).

4.3.2.18 Klada Tombus-Noda

Klada Tombus-Noda zajema dve večji skupini: Nodaviridae in Tombusviridae. Vsaka je razdeljena na več podskupin. Družina Tombusviridae zajema tudi virusno družino Carmotetraviridae. Našli smo genome treh virusov klade Tombus-Noda v transkriptomih deseteronožcev in tri viruse v transkriptomih enakonožcev. Med virusi, najdenimi v transkriptomih deseteronožcev (Decapoda), je bil en virus uvrščen v skupino Tombusviridae in dva v skupino Nodaviridae. Med virusi, najdenimi v transkriptomih enakonožcev (Isopoda), je bil ravno tako eden uvrščen v skupino Tombusviridae in dva v skupino Nodaviridae. Tombusvirus, ki smo ga našli pri deseteronožcu *Callinectes sapidus* (GEID), je soroden Gallantivirusom, Macanavirusom in Alphanecrovirusom, kot so »Gallinsoga mosaic virus« (Gallantivirus), »Furcraea necrotic streak virus« (Macanavirus) in »Tobacco necrosis virus A«, izolirani iz tobaka (*Nicotiana tabacum*) in agavovk (*Furcraea*). Tombusvirus, ki smo ga našli v transkriptomu enakonožca (Isopoda), je najbolj soroden »Xinzhou nematode virus 7« (XZSJSC65582), virusu, najdenemu v transkriptomu glist (Nematoda). Nodavirus, najden v transkriptomu deseteronožca morski pajek – *Hyas araneus* (HACH01051331), je 85 % identičen glede na proteinsko zaporedje RdRp »Shuangao noda-like virusu 1« (insectZJ94204), izoliranemu iz mešane transkriptomske knjižnice, ki vsebuje RNA vzorce različnih redov žuželk v okviru raziskave Shi in sod. (2016a). V tem primeru se odpre vprašanje o prenosu virusa s kopenskega vretenčarja na vodnega vretenčarja in o stičišču njunih ekoloških niš. Nodavirus, ki smo ga našli v transkriptomski knjižnici *Scylla paramamosain* (GEUT01000022) (medvedji raki – Scyllaridae), je identičen virusu »Wenzhou tombus-like virus 18« (beimix29339), izoliranemu iz mehkužcev. Do prenosa virusa je lahko prišlo v okviru skupne ekološke niše omenjenih organizmov. Virus, izoliran iz transkriptoma enakonožca *Proasellus ortizi* (HAFF01024124), je soroden virusoma, odkritima v transkriptomih mehkužcev; 1) Wenzhou noda-like virus 5 (WZFSL73727) in 2) »Wenzhou noda-like virus 4« (WZFSL75166). Virus *Proasellus meridianus* (HAFE01129314) je najbolj soroden Pariacoto virusu glede na naše filogenetsko drevo. Glede na raziskave Shi in sod. (2016a) nobena od skupin nevretenčarjev ne izstopa v razširjenosti klade Tombus-Noda. Klada je praktično enakomerno razširjena pri žuželkah, rakah in mehkužcih, nekoliko manj pri pajkovcih, redka pa je pri glistah (Nematoda) (**Dodatek B; B.4, Slika B.36**).

4.3.2.19 Kladi Ophio in Qin

Skupina Shi in sod. (2016a) je v okviru ss(-)RNA virusov prikazala grafično filogenetsko drevo, ki ponazarja sorodstvene odnose vseh ss(-) RNA virusov med seboj. Pri iskanju novih RNA virusov pri rakah smo našli dva zelo divergentna virusa, za katera smo vedeli, da spadata v gručo virusnih skupin Qinvirus, Ophioviridae, Yuevirus, Orthomyxoviridae in Mononegavirales. Da bi dokazali njuno filogenetsko umestitev, smo nova RNA virusa prikazali v filogenetskem drevesu z virusnimi skupinami: Qinvirus, Ophioviridae, Yuevirus, Orthomyxoviridae in Mononegavirales. En virus smo našli v genomu ceponožca *Eurytemora affinis* (AZAI), drugi v transkriptomu deseteronožca *Carcinus maeans* (GBXE). Virus, odkrit v genomu *Eurytemora affinis* (AZAI), spada med qinviruse, virus, odkrit v genomu *Carcinus maeans* (GBXE), pa spada v virusno skupino Ophioviridae (**Dodatek B; B.4, Slika B.37**).

4.3.2.10 Klada Orthomyxo

Klada Orthomyxo je razdeljena na tri monofiletske skupine: 1) Isavirus, Thogotovirus in Influenzavirus, 2) Quaranjavirus in 3) nova skupina, ki jo je v filogenetskem drevesu pokazala tudi skupina Shi in sod. (2016a) – Changping earthworm virus 2 (CPQY18140) podobni virusi. V transkriptomih rakov enakonožcev smo odkrili tri viruse. Dva, ki sta si podobna, sta bila uvrščena v skupino »Influenza«, »Thogoto« in »Isavirusu« podobnih virusov. Eden se je uvrstil med »Changping earthworm virusu 2« (CPQY18140) podobne viruse. V transkriptomih postranic (Amphipoda) smo odkrili tri orthomyxoviruse. Med njimi je bil virus, najden v transkriptomu *Micruropus wahlui*, uvrščen v skupino Influenza, Thogoto in Isavirusu podobnih virusov. Virus, najden v transkriptomu *Pachyschelus branchialis*, je bil uvrščen med quaranjaviruse, in virus, najden v transkriptomu *Gammarus chevreuxi*, je bil uvrščen med Changping earthworm virus 2 (CPQY18140) podobne viruse. Glede na raziskavo Shi in sod. (2016a) so orthomyxovirusi v večinskem deležu razširjeni pri

žuželkah. Redki so pri mehkužcih in pajkovcih. Pri rakih je skupina odkrila samo en virus (»Beihai orthomyxo-like virus 1« – BHTSS17872) (**Dodatek B; B.4, Slika B.38**).

4.3.2.1.11 Klada Partiti-Picobirna

V transkriptomih rakov smo našli pet virusov, ki spadajo v klado **Partiti-Picobirna**. Med njimi so trije virusi iz transkriptomov rakov ceponožcev in dva iz rakov enakonožcev. Virus iz *Tigriopus californicus* (JW510638) in *Tigriopus japonicus* (GCHA01047740) sta sorodna skupini »Raphanus sativus cryptic virus 1« podobnih virusov (izoliran iz redkvice, *Raphanus sativus*), »Sclerotinia sclerotium partitivirus S«, »Rosellinia necatrix partitivirus 2« in »Rose partitivirus«. Skupina Shi in sod. (2016a) je isti virus odkrila tudi v transkriptomu rakov deseteronožcev *Procambarus sp.* »Sclerotinia sclerotium partitivirus S« je glivni virus, izoliran iz rastlinske patogene glive *Sclerotinia sclerotiorum*. »Rosellinia necatrix partitivirus 2« je bil izoliran iz rastlinske patogene glive *Rosellinia necatrix*. Rose partitivirus je rastlinski virus, ki je bil izoliran iz vrtnice kultivarja Phyllis Bide (NCBI taxonomy). Gre torej za skupini rastlinskih in glivnih virusov. Glive, iz katerih so bili izolirani glivni partitivirusi, so rastlinske patogene glive. Povezujejo se skupaj z rastlinskimi partitivirusi. Že Koonin in sod. (2015) so omenjali pomembno vlogo horizontalnega prenosa virusov med gostitelji pri partitivirusih, konkretnije horizontalnega prenosa iz glivnega patogena na rastlino, pri čemer je seveda dotični virus prešel skozi premik v hitrosti mutacije in substitucije. Virus iz *Tigriopus californicus* (JW510638) in *Tigriopus japonicus* (GCHA01047740) sta glede na zaporedje RdRp 50 % identična med seboj, z rastlinskimi in glivnimi virusi pa v rangi od 34 % (*Raphanus sativus cryptic virus 1*) do 35 % (*Rose partitivirus*). Skupina znanih rastlinskih virusov je identična v podobnem rangi (33 %–35 %). Ostali virusi so sorodni partitivirusom, odkritim v transkriptomih žuželk (QTM – Odonata), pajkov (spider) in rakov (BHTSS), narejenih v okviru raziskave Shi in sod. (2016a) (**Dodatek B; B.4, Slika B.39**).

4.3.2.1.12 Klada Permutotetra

Klada Permutotetra ima štiri monofiletske skupine: 1) »Niehaus virusu« podobni virusi, 2) »Thosea asigna virusu« podobni virusi, 3) »Newfield virusu« podobni virusi in 4) »Drosophila A virusu« podobni virusi. »Niehaus virusu« podobni virusi tvorijo sestrsko skupino s »Thosea asigna virusu« podobnimi virusi. »Newfield virusu« podobni virusi tvorijo sestrsko klado z »Drosophila A virusu« podobnimi virusi. Referenčni klasificirani virusi so glede na dolžino vej filogenetskega drevesa divergentni in imajo nizko medsebojno podobnost zaporedij. V transkriptomu raka *Talitrus saltator* smo našli genom permutotetravirusa, ki spada v skupino »Thosea asigna virusu«– podobnih virusov (**Dodatek B; B.4, Slika B.40**).

4.3.2.1.13 Klada Picorna-Calici

V transkriptomih rakov smo našli 54 genomov virusov, ki spadajo v klado **Picorna-Calici**. Od tega smo našli sedem pikornavirusov v transkriptomih postranic (Amphipoda), devet v transkriptomih ceponožcev (Copepoda), 22 v transkriptomih enakonožcev (Isopoda), 15 v transkriptomih deseteronožcev in enega v transkriptomu skupine Branchiura, ki smo ga označili z magenta barvo. Izmed 54 zadetkov jih 14 spada med »aquatic picornavirales«, 14 v skupino Dicistroviridae, 11 v skupino Iflaviridae, tri v skupino Nora virusu podobnih virusov in štiri v skupino Kelp fly virusu podobnih virusov. Sedem pikornavirusov v transkriptomih *Procambarus sp.*, *Pontastacus sp.*, *Cherax sp.*, *Caligus sp.* in *Lepeophtheirus sp.* je oblikovalo ločeno skupino. Kot »outgroup« skupine pikornavirusov smo uporabili skupino Caliciviridae. Ker je filogenetsko drevo preko »outgroupa« orientirano v času, poleg sorodstvenih odnosov med virusi, prikazuje tudi evolucijsko starost filogenetskih skupin pikornavirusov. Pri tem so Kelp fly virusu podobni virusi najstarejša skupina pikornavirusov, sledijo Nora virusu podobni virusi in posavirusu podobni virusi (kot sestrška klada), nato skupina rastlinskih virusov Secoviridae, Iflaviridae in gruča Dicistroviridae in »aquatic picornavirales kot gruča najmlajših sestrskih skupin. Skupina Shi in sod. (2016a) je uporabila isto skupino kot »outgroup«, vendar je pri njih evolucijska starost skupin razporejena drugače. V nizu si sledijo: Kelp fly virusu podobni virusi, Aquatic picornavirales, Posavirusu podobni virusi, sledi niz Dicistroviridae, Iflaviridae, Noravirusu podobni virusi in Arivirusu podobni virusi, ki tvorijo sestrsko klado, in skupina Picornaviridae kot najmlajša skupina. Pri skupini Shi in sod. (2016a) so bile skupine Dicistroviridae, »aquatic picornavirales«, Arivirusu podobni virusi, Kelp fly virusu podobni

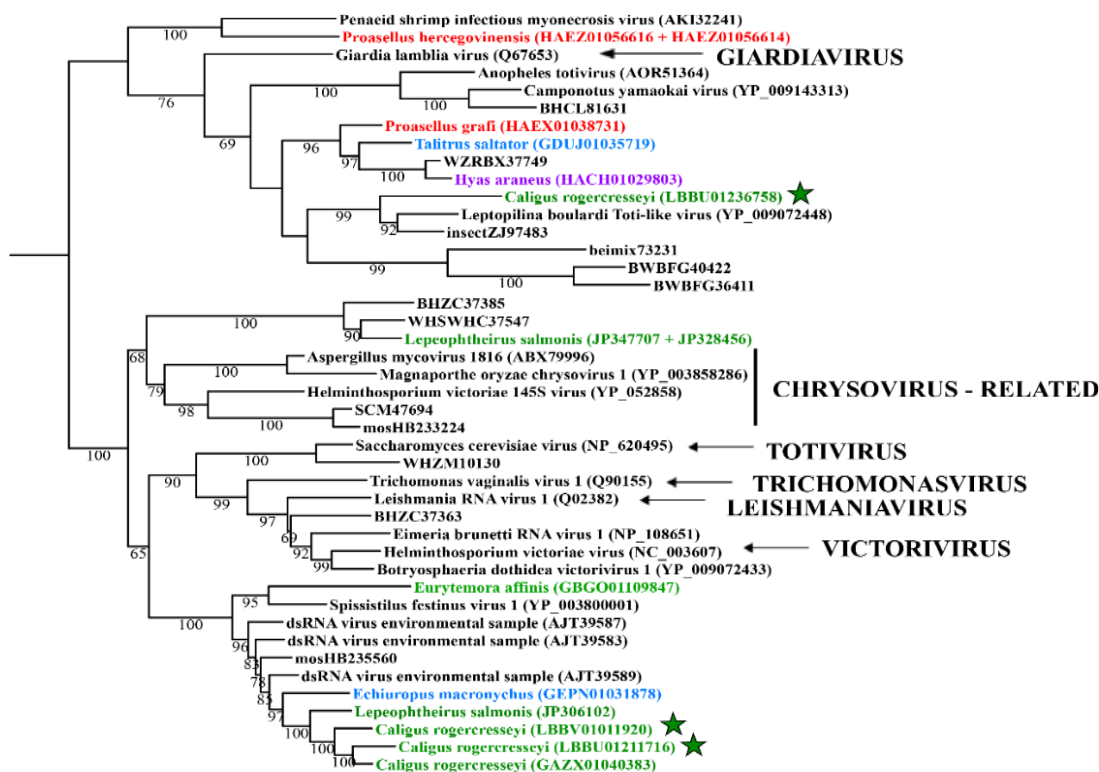
virusi in Nora virusu podobni virusi bogato zastopane pri rakih. Vendar v transkriptomih rakov praktično niso našli iflavirusov. Pri njih je bila zastopana skoraj izključno pri žuželkah. Mi smo našli 11 iflavirusov v transkriptomih rakov 1) enakonožcev; *Proasellus sp.*, *Bragasellus sp.*, 2) postranic: *Talitrus sp.* in *Pachyschesis sp.*, 3) deseteronožci: *Pontastacus sp.* in *Cherax sp.* ter 3) *Argulus sp.* (Branchiura). Iflavirusi so bili najbolj bogato prisotni pri enakonožcih. Večina iflavirusov, najdenih v transkriptomih rakov, se poveže skupaj z »Nassonia vitripennis virusom«, izoliranim iz parazitoidne ose *Nassonia vitripennis*. Našli smo tudi genom virusa v transkriptomu *Procambarus clarkii* (GBEV01028478), ki je starejši od skupine »aquatic picornavirales«, Arivirusu podobnih virusov, Iflaviridae, Dicistroviridae in Secoviridae. Od genomske strukture celotnih pikornavirusov trije virusi izstopajo glede na prisotnost novih domen, ki niso pogosto prisotne pri pikornavirusih (Dodatek B; B.4, Slika B.41).

4.3.2.1.14 Klada Reo

Klada Reo oblikuje tri ločene monofiletske skupine. V transkriptomih deseteronožcev smo našli genom dveh virusov, v transkriptomih enakonožcev dveh virusov in v transkriptomih postranic genom enega virusa. Virus iz *Eriocheir sinensis* (HAAX01020223) je soroden *Callinectes sapidus* reovirusu 1, ki je bil izoliran iz deseteronožca *Callinectes sapidus*. Proteinski zaporedji RdRp sta 83 % identični. Zaporedji, najdeni v transkriptomih *Echinogammarus veneris* (GAR0) in *Proasellus racovitzai* (HAFH), sta divergentni zaporedji, ki sta se povezali s Cimodo virusom, vendar gre za raven 25 % identičnosti zaporedij RdRp (privlak dolgih vej – LBA, angl. »Long Branch Attraction«) (Dodatek B; B.4, Slika B.42).

4.3.2.1.15 Klada Toti-Chryso

V transkriptomih in genomih rakov smo skupno našli 12 virusov, ki spadajo v **klado Toti-Chryso**. Od tega sedem pri ceponožcih (Copepoda), dva pri enakonožcih (Isopoda), dva pri postranicah (Amphipoda) in enega pri deseteronožcih (Decapoda). Pri ceponožcih smo našli tri totiviruse v genomu *Caligus rogercresseyi* in štiri v transkriptomih *Eurytemora sp.*, *Caligus sp.* in *Lepeophtheirus sp.* Dva izmed njih – totivirus iz transkriptoma *Eurytemora affinis* (GBGO) in genoma *Caligus rogercresseyi* (LBBU01236758) – sta sorodna virusom, izoliranim iz žuželk *Spissistilus festinus* (Hemiptera) in *Leptopilina boulandi* (Hymenoptera) (Slika 4.18).



Slika 4.18: Filogenetsko drevo virusov klade **Toti-Chryso**, ki smo jih našli v genomih in transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postranice (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično. Endogenizirani virusi so označeni z zvezdico.

4.3.2.1.16 Klada Zhaovirus

Klada Zhaovirus, ki so jo definirali Shi in sod. (2016a), je v raziskavi zajemala tri viruse, izolirane iz rakov enakonožcev (BHTSS) in deseteronožcev. V naši analizi smo našli virus tudi v transkriptomu ceponožca (Copepoda) *Lepeophtheirus salmonis* (Dodatek B; B.4, Slika B.43).

4.3.2.2 Skupina Spiralia: mehkužci (Mollusca), ploščati črvi (Platyhelminthes), kolobarniki (Annelida) in kotačniki (Rotifera)

V okviru skupine Spiralia smo pregledali 225 genomskih in transkriptomskih knjižnic. Genome virusov smo našli v genomih in transkriptomih skupin kolobarnikov (Annelida), Rotifera, mehkužcev (Mollusca) in ploščatih črvov (Platyhelminthes). Mehkužce (Mollusca) smo označili z rdečo barvo, skupino ploščatih črvov (Platyhelminthes) z vijolično, kolobarnike (Annelida) z zeleno barvo in kotačnike (Rotifera) z modro barvo. Po profilu raznolikosti RNA virusnih klad in po številčnosti virusov v transkriptomih in genomih so izstopali mehkužci (Mollusca). Pri tem velja poudariti, da od 225 genomskih in transkriptomskih knjižnic kar 125 (56 %) pripada mehkužcem. Po raznolikosti in številčnosti prisotnih virusnih skupin so sledili ploščati črvi, ki zajemajo 66 od 225 (29 %) genomskih in transkriptomskih knjižnic. Pri kolobarnikih (Annelida) in kotačnikih (Rotifera) smo našli samo posamezne primere. Hkrati njihove knjižnice predstavljajo samo 5–10 % vseh genomskih in transkriptomskih knjižnic skupine Spiralia.

4.3.2.2.1 Klada Astro

Klada Astro vsebuje dve ločeni monofiletski skupini. Prvo predstavljajo klasificirani avastro in mamastrovirusi, druga je neklasificirana in je bila prvič prikazana v okviru skupine Shi in sod. (2016a). V transkriptomih ploščatih črvov smo našli dva divergentna astrovirusa, in sicer pri *Bdelloura candida* (GFKB01038875 + GFKB01027725) in *Pomphorhynchus laevis*

(AM918867). Proteinski zaporedji RdRp omenjenih astrovirusov sta 53 % identični. Glede na naše vzorčenje sta najbolj sorodni salmovirusu, vendar sta 24–30 % identični kateremu koli astrovirusu. Pri virusih *Bdelloura candida* (GFKB01038875 + GFKB01027725) in *Pomphorhynchus laevis* (AM918867) nismo vedeli, ali sodita v skupino astrovirusov ali astrovirusom soroden klaster virusov, kot sta kladi Luteo-Sobemo in Picorna-Calici. Zato smo slednji skupini vključili v filogenetsko drevo (**Dodatek B; B.5, Slika B.44**).

4.3.2.2 Klada Birna

Klada Birna vsebuje dve monofiletski skupini. V eni se nahajajo klasificirani birnavirusi: Aquabirnavirus, Avibirnavirus, Entomobirnavirus in Blosnavirus. Drugo skupino je v filogenetski analizi prva prikazala skupina Shi in sod. (2016a), gre za Tellina virusu 1 podobne viruse. V transkriptomu ploščatega črva *Procotyla fluviatilis* (GAKZ) smo odkrili divergenten birnavirus, ki spada v skupino Tellina virusu 1 podobnih virusov (**Dodatek B; B.5, Slika B.45**).

4.3.2.3 Klada Bunya-Arena

Klada Bunya-Arena vsebuje tri skupine. Prvi so Nairovirusu podobni virusi, v drugi so skupine: Hantavirus, Orthobunyavirus, Tospovirus in Emaravirus. Tretjo predstavljajo Phlebovirusu in Tenuivirusu podobni virusi. Odkrili smo dva virusa v transkriptomih mehkužcev in tri v transkriptomih ploščatih črvov. Proteinski zaporedji RdRp iz mehkužcev *Elysia timida* (GBRM01106997) in *Mytillus galloprovincialis* (GAEN01011486) sta sorodni Puumala virusu, sesalskem virusu, izoliranemu iz gozdne voluharice (*Myodes glareolus*). *Elysia timida* (GBRM01106997) in *Mytillus galloprovincialis* (GAEN01011486) sta divergentni zaporedji, ki sta 25–30 % identični virusom skupine Bunyavirales. Virus, ki smo ju našli v transkriptomih ploščatih črvov *Procotyla fluviatilis* (GAKZ01048260) in *Schmidtea polychroa* (GBZN01006195), sta med seboj glede na zaporedje RdRp 56 % identična. V drevesu sta se povezala z »Wuhan spider virusom«. Proteinsko zaporedje RdRp »Wuhan spider virusa« je 41 % identično s proteinskim zaporedjem RdRp virusa *Procotyla fluviatilis* (GAKZ01048260). Virus pri *Trichobilharzia regenti* (GDKR01011304) je soroden z virusom »Hubei blood fluke virus 1« (HBXXC56361), izoliranim iz transkriptomov ploščatih črvov v okviru raziskave Shi in sod. (2016a) (**Dodatek B; B.5, Slika B.46**).

4.3.2.4 Klada Flavi

Klada Flavi obsega skupine: Flavivirus, Hepacivirus, Pegivirus, Pestivirus, flavivirusom podobni virusi in jingmenvirusi. Skupine Hepacivirus, Pegivirus in Pestivirus so bile prvotno označene kot skupine sesalskih virusov, dokler Shi in sod. (2016b) niso dokazali razširjenosti skupine Hepacivirusov pri morskem psu. Skupina jingmenvirusov oblikuje sestrsko klado s Flavivirusi. Obe sta evolucijsko mladi (Shi in sod., 2016b), vendar so se jingmenvirusi med diverzifikacijo začeli segmentirati. Prvotno je bila znana njihova razširjenost pri sesalcih (primatih) in žuželkah, kot so komarji, klopi, bolhe. Skupina flavivirusom podobnih virusov naj bi predstavljala starejšo obliko oz. prednika skupin Flavivirusov in Jingmenvirusov. So bistveno širše razširjeni kot ostale skupine Flaviviridae (Shi in sod., 2016b). Pri skupini Spiralia smo našli štiri divergentne flaviviruse – tri v knjižnicah mehkužcev – *Watasenia sp.*, *Idiosepius sp.*, *Elysia sp.* In enega pri skupini ploščatih črvov – *Procotyla fluviatilis*. Virusi iz transkriptomov *Procotyla fluviatilis*, *Watasenia sp.*, *Idiosepius sp.* so bili uvrščeni med flaviviruse. Virus iz *Elysia sp.* je bil uvrščen med flavivirusom podobne viruse (**Dodatek B; B.5, Slika B.47**).

4.3.2.5 Klada Hepe-Virga

Klado Hepe-Virga pri skupini Spiralia smo predstavili na treh ločenih filogenetskih drevesih. Prvo je vključevalo skupino Hepeviridae. Skupina oblikuje dve podskupini. Prva predstavlja klasificirane hepeviruse, kot sta ortohepevirus in piscihepevirus, drugo podskupino so odkrili Shi in sod. (2016a) in predstavlja Hepevirusom podobne viruse. Hepevirusi so bili do pred kratkim znani le pri vretenčarjih – ribah in sesalcih. Shi in sod. (2016a) so našli novo podskupino hepevirusom podobnih virusov, ki so pogosti pri raki, redki pa pri žuželkah in mehkužcih. V transkriptomu *Elliptio complanata* (GAHW01000160) smo odkrili divergenten hepevirus, ki je glede na zaporedje RdRp 28–30 % identičen z ostalimi virusi skupine Hepeviridae. Drugo filogenetsko drevo je vključevalo skupino negevirusov. Po filogenetski analizi negevirusov smo opazili, da vsebujejo štiri skupine: 1) Negev virusu

podobni virusi, 2) Tanay virusu podobni virusi, 3) Bofa virusu podobni virusi in 4) »Muthill virusu podobni« virusi. V transkriptomih organizmov skupine Spiralia smo našli pet negevirusov – tri v transkriptomih mehkužcev *Elysia sp.* in dva v transkriptomih ploščatih črvov *Schmidtea mediterranea* in *Fasciola hepatica*. Virusi iz *Elysia sp.* in *Schmidtea mediterranea* so oblikovali ločeno skupino virusov. Med seboj so samo 30 % identični, ravno tako tudi z ostalimi negevirusi. Glavna katalitična domena RdRp v motivu C ima zaporedje AGDD ali GGDD, v primeru virusa, najdenega v transkriptomu *Schmidtea mediterranea*, pa ima LGDD. Virus iz *Fasciola hepatica* (GEVX) spada v podskupino Negev virusu podobnih virusov. Tretje filogenetsko drevo obsega skupine Togaviridae, Virgaviridae in Negevirus. Virus iz *Elysia timida* (GBRM01100701) je divergenten virus, ki je bil uvrščen med togaviruse in je 27 % identičen z ostalimi trenutno znanimi virusi klade Hepe-Virga (**Dodatek B; B.5, Slika B.49, Slika B.51 in Slika B.57**).

4.3.2.2.6 Klada Luteo-Sobemo

Klada Luteo-Sobemo obsega tri skupine: 1) Carascovirusu podobni virusi, 2) Wenzhou sobemo-like virus 3 (mosZJ35256) podobni virusi in 3) La Tardorie virusu podobni virusi. V transkriptomu sipe *Euprymna scolopes* smo našli divergenten luteovirus, ki spada med Carascovirusu podobne viruse (**Dodatek B; B.5, Slika B.48**).

4.3.2.2.7 Klada Narna-Levi

Klada Narna-Levi obsega dve virusni skupini: Narnaviridae in Leviviridae. Skupina Narnaviridae vsebuje tri monofiletske skupine: Narnavirus, Mitovirus, Ourmiavirusu podobni virusi. V genomih in transkriptomih skupine Spiralia nismo našli virusov, ki bi bili iz skupine Leviviridae. Zato smo se pri kladi Narna-Levi pri skupini Spiralia osredotočili na skupino Narnaviridae. Odkrili smo tri genome virusov, ki spadajo v družino Narnaviridae – enega v transkriptomu kotačnika *Rotaria tardigrada* (GDRK01026552), drugega v transkriptomu kolobarnika *Enchytraeus crypticus* (GALF01016517) in enega v transkriptomu školjke *Alasmodonta varicosa* (HP641214). Virus iz *Rotaria tardigrada* (GDRK01026552) spada med mitovirusom podobne viruse. Soroden je s »Hubei narna-like virus 22« (SZmix20730). Z njim je glede na zaporedje RdRp 39 % identičen. Virus iz *Enchytraeus crypticus* (GALF01016517) je soroden »Alternaria arborescens mitovirusu 1«, glivnemu virusu, izoliranemu iz rastlinske patogene glive *Alternaria arborescens*. Proteinsko zaporedje RdRp virusa iz *Alasmodonta varicosa* (HP641214) je 66 % identično zaporedju »Hubei narna-like virusa 2« (WHBL67117) (**Dodatek B; B.5, Slika B.50**).

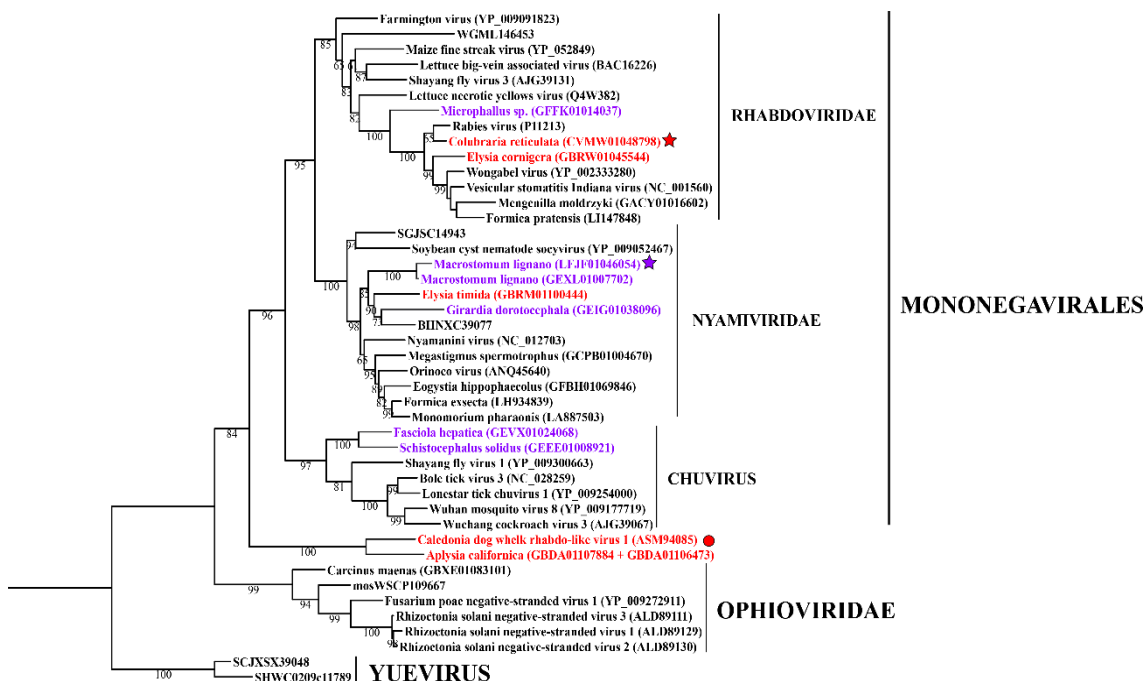
4.3.2.2.8 Klada Tombus-Noda

V okviru klade **Tombus-Noda** smo naredili dve ločeni filogenetski drevesi – eno za družino Nodaviridae in eno za družino Tombusviridae. Družina Nodaviridae oblikuje pet ločenih podskupin: 1) »Beihai noda-like virus 20« (BHZY60039) podobni virusi, 2) Betanodavirusu podobni virusi, 3) »Hubei noda-like virusu 2« (WHBL70552) podobni virusi in 4) »Le Blanc« in »Santeuil« nodavirusu podobni virusi, ki oblikujejo še dve gruči. V transkriptomih skupine Spiralia smo odkrili dva nodavirusa, in sicer v transkriptomu ploščatega črva *Opisthorchis felinus* (GBJA01011878) in kotačnika *Rotaria magnacalcarata* (GDRE01024231). Virus iz *Opisthorchis felinus* (GBJA01011878) je soroden »Pariacoto virusu«, ki je razširjen med žuželkami. Proteinski zaporedji njunih RdRp sta 35 % identični. Zaporedje virusa je 26–39 % identično z ostalimi trenutno znanimi nodavirusi. Virus iz *Rotaria magnacalcarata* (GDRE01024231) spada v skupino »Le Blanc« in »Santeuil nodavirusu« podobnih virusov, vendar je tvoril ločeno podskupino z »Beihai noda-like virus 5« (BHZC36307). Proteinski zaporedji RdRp virusov *Rotaria magnacalcarata* (GDRE01024231) in »Beihai noda-like virus 5« (BHZC36307) sta 34 % identični. Skupina Tombusviridae tvori dve ločeni podskupini. Prvo predstavljajo novo pridobljena virusna zaporedja v okviru raziskave Shi in sod. (2016a). Drugo podskupino predstavljajo klasificirani tombusvirusi: Tombusvirus, Umbravirus, Gallantivirus, Macanavirus, Zeavirus, Aureusvirus, Alphanecrovirus, Betanecrovirus, Machlomovirus, Avenavirus, Dianthovirus, Panicovirus, Alphacarmovirus. Odkrili smo tombusvirusa v transkriptomu kolobarnika *Ophryotrocha diadema* (GENQ01001917) in v transkriptomu mehkužca *Elysia cornigera* (GBRW01030946). Virus *Elysia cornigera* (GBRW01030946) je soroden virusu »Beihai tombus-like virus 7« (BHWZXX15571) in »Hubei tombus-like virus 31« (WHBL72828). Oba tombusvirusa se grupirata skupaj s klasificiranimi virusi skupine Tombusviridae

(Tombusvirus, Umbravirus, Gallantivirus, Macanavirus idr.) (Dodatek B; B.5, Slika B.52 in Slika B.55).

4.3.2.2.9 Klada Mono-Chu

Klada Mono-Chu zajema tri skupine: Rhabdoviridae, Nyamiviridae in Chuviruse. V genomih in transkriptomih ploščatih črvov smo našli šest genomov virusov. Virus iz *Microphallus sp.* (GFFK01014037) spada v skupino Rhabdoviridae. Poveže se skupaj z virusom stekline, Wongabel virusom (izolirani iz komarja *Culicoides austropalpalis*) in Vesicular stomatitis Indiana virusom (izoliranim iz goveda). Poveže se torej skupaj z virusi, ki so razširjeni pri sesalcih in komarjih – znanih vektorjih sesalskih virusov. Proteinsko zaporedje RdRp virusa iz *Microphallus sp.* (GFFK01014037) je z virusom stekline samo 37 % identično. V transkriptomu organizma *Macrostomum lignano* (GEXL01007702) smo odkrili genom virusa, ki se poveže skupaj z »Beihai rhabdo-like virusom 3« (BHNXC39077). Proteinski zaporedji njunih RdRp sta 43 % identični. V genomu *Macrostomum lignano* (LFJF01046054) smo odkrili zaporedje, ki je 85 % identično zaporedju *Macrostomum lignano* (GEXL01007702). Ploščati črv *Macrostomum lignano* je zato dejanski gostitelj virusa. Virus iz *Macrostomum lignano* (GEXL01007702) in »Beihai rhabdo-like virus 3« (BHNXC39077) sta se povezala tudi skupaj z virusom, ki smo ga odkrili v transkriptomu *Girardia dorocephala* (GEIG01038096). Proteinsko zaporedje RdRp virusa *Girardia dorocephala* (GEIG01038096) je 37 % identično proteinskemu zaporedju RdRp Beihai rhabdo-like virusa 3 (BHNXC39077) in 34 % identično proteinskemu zaporedju RdRp virusa *Macrostomum lignano* (LFJF01046054). Omenjena skupina virusov spada v skupino Nyamiviridae. Nyamivirusi so bili prvotno skupina virusov, razširjena izključno med žuželkami. Pokazali smo, da so Nyamivirusi bogato razširjeni tudi med ploščatimi črvi, pri čemer jih tudi dejansko okužujejo. Pokazali smo namreč, da nyamivirusom podobna skupina virusov, ki se je razširila tudi med vodnimi nevretenčarji, aktivno okužuje *Macrostomum lignano*. Pri nyamivirusih so namreč oblikovane tri sestrške podskupine – virusi, ki so razširjeni pri žuželkah, virusi, ki so razširjeni pri mehkužcih in ploščatih črvih (vodnih nevretenčarjih), in virusi, ki so razširjeni pri glistah in ploščatih črvih (»Soybean cyst nematode socyvirus« in »Wenzhou tapeworm virus 1« (SGJSC14943)). V transkriptomu organizmov *Fasciola hepatica* (GEVX) in *Schistocephalus solidus* (GEEE) smo odkrili tudi dva chuvirusa, ki se povežeta skupaj s »Shayang fly virusom 1«, »Bole tick virusom 3«, »Lonestar tick chuvirusom 1«, »Wuchan mosquito virusom 8« in »Wuchang cockroach virusom 3«. Ti virusi so razširjeni pri žuželkah; klopah, komarjih in ščurkih. Proteinski zaporedji njunih RdRp sta med seboj 37 % identični, z ostalimi chuvirusi pa sta identični od 25–30 %. V transkriptomu mehkužca *Elysia cornigera* (GBRW01045544) smo odkrili genom virusa, ki se poveže skupaj z *Microphallus sp.* (GFFK01014037). Proteinsko zaporedje njegove RdRp je 45–50 % identično z virusom stekline. Hkrati je proteinsko zaporedje RdRp 47 % identično s proteinskim zaporedjem RdRp virusa iz genoma *Colubraria reticulata* (CVMW01048798) – organizma, ki je dejanski gostitelj za omenjeni virus (CVMW01048798). V transkriptomu mehkužca *Elysia cornigera* (GBRM01100444) smo odkrili nyamivirus, ki klastira skupaj z virusom *Girardia dorocephala* (GEIG01038096), *Macrostomum lignano* (GEXL01007702), *Macrostomum lignano* (LFJF01046054) in Beihai rhabdo-like virusom 3 (BHNXC39077) – klastrom nyamivirusov, razširjenih med mehkužci in ploščatimi črvi. Poleg omenjenih virusov smo našli divergenten virus v transkriptomu mehkužca *Aplysia californica*, ki se je v drevesu povezal z virusom »Caledonia dog whelk rhabdo-like virusom 1«. Omenjen virus je bil v raziskavi Waldron in sod. (2017) izoliran iz škrlatnega voleka (angl. »dog whelk«), kot pove že samo ime virusa. Proteinski zaporedji RdRp sta med seboj samo 23–30 % identični, okrog 25 % pa sta identični z ostalimi predstavniki skupine Mononegavirales. Pokazali smo, da sta virusa Caledonia dog whelk rhabdo-like virus 1 in virus, odkrit v transkriptomu organizma *Aplysia californica*, verjetno starejša predstavnika skupine Mononegavirales. Hkrati smo pokazali, da so mehkužci, kot je *Colubraria reticulata*, dejanski gostitelji rhabdovirusov, medtem ko so ploščati črvi, kot je *Macrostomum lignano*, dejanski gostitelji virusov skupine Nyamiviridae (Slika 4.19).



Slika 4.19: Filogenetsko drevo virusov klade **Mono-Chu**, ki smo jih odkrili v genomih in transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo. Endogenizirani virusi so označeni z zvezdico. Virusi, odkriti v referenci Waldron in sod. (2017), so označeni s krogcem.

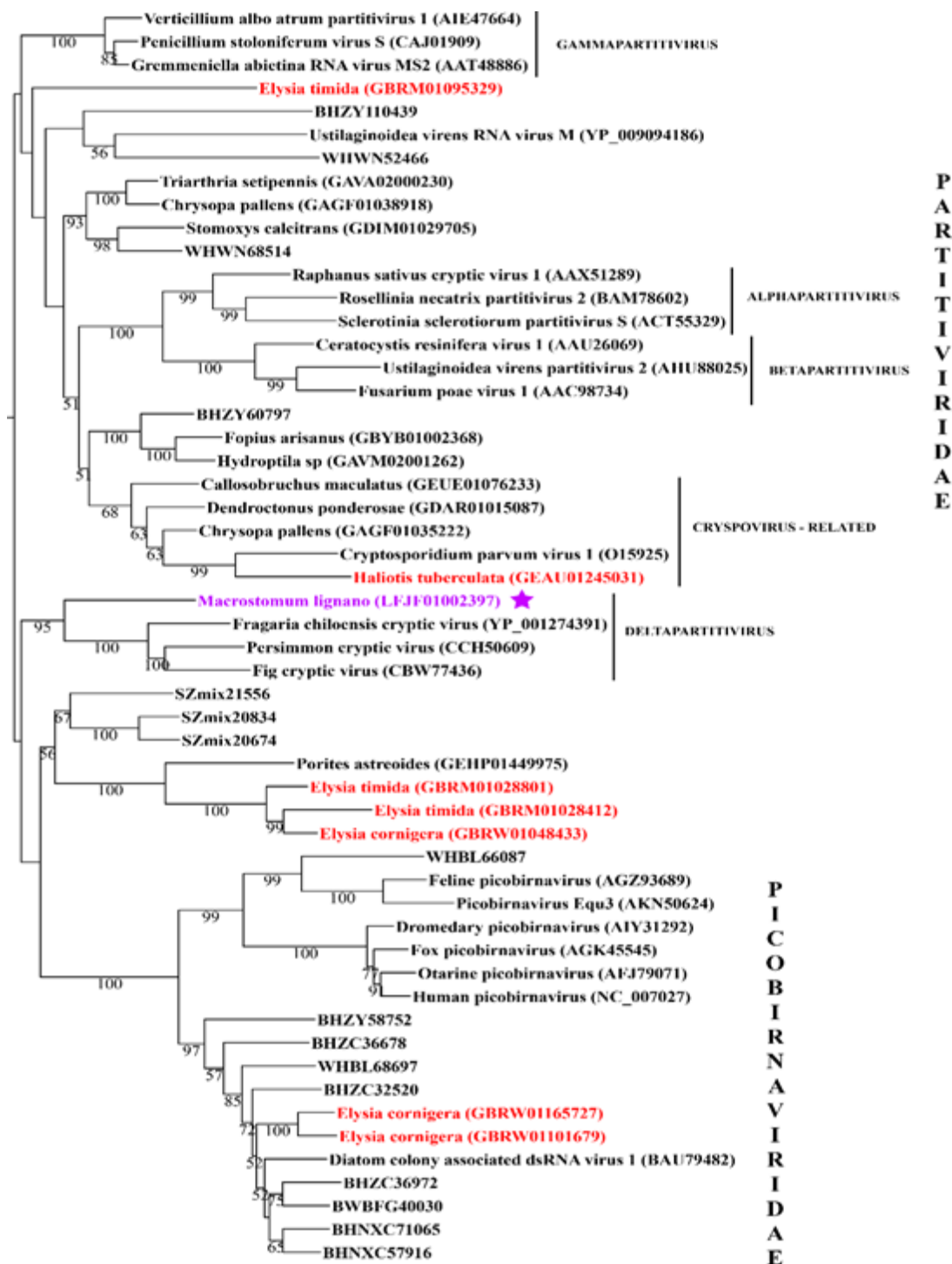
4.3.2.2.10 Klada Orthomyxo

Klada Orthomyxo tvori tri ločene skupine: 1) klaster »virusov gripe« in »thogotovirusov«, 2) klaster »quarjanvirusov« in 3) novo skupino virusov, ki je bila odkrita v raziskavi Shi in sod. (2016a). V genomu morskega polža *Colubraria reticulata* smo odkrili delni genom virusa, poveže se skupaj z virusi gripe in »thogotovirusi« (**Dodatek B; B.5, Slika B.53**).

4.3.2.2.11 Klada Partiti-Picobirna

Klada Partiti-Picobirna obsega družini Partitiviridae in Picobirnaviridae. V knjižnicah skupine Spiralia smo našli genome osmih virusov, ki spadajo v klado Partiti-Picobirna. Sedem od teh smo odkrili v transkriptomih mehkužcev (Mollusca) in enega v genomu ploščatega črva *Macrostromum lignano* (LFJF01002397). Proteinsko zaporedje RdRp virusa *Macrostromum lignano* (LFJF01002397) se poveže skupaj z deltapartitivirusi. Proteinsko zaporedje RdRp je okrog 30 % identično deltapartitivirusom. V transkriptomih mehkužcev smo našli viruse, ki spadajo med partitiviruse, in tudi nove picobirnaviruse. Virus iz *Elysia timida* (GBRM01095329) je divergenten partitivirus, ki je glede na proteinsko zaporedje RdRp 25–30 % identičen z ostalimi trenutno znanimi partitivirusi. Virus iz *Halotis tuberculata* (GEAU01245031) se je v filogenetski analizi povezal s »Cryptosporidium parvum virus 1« v skupino crispovirusov. »Cryptosporidium parvum virus 1« je bil izoliran iz alveolata *Cryptosporidium parvum*, ki sodi med najstarejše skupine evkariontov. Njuni proteinski zaporedji RdRp sta 38 % identični. Trije virusi, izolirani iz transkriptoma *Elysia sp.*, tvorijo lastno skupino z virusom *Porites astreoides* (GEHP01449975), vendar predstavljajo sestrski klaster virusom »Hubei partiti-like virus 55« (SZmix21556), »Hubei partiti-like virus 54« (SZmix20834) in »Hubei partiti-like virus 53« (SZmix20674). Skupini sta si med seboj 30 % identični. Dva virusa iz *Elysia sp.* spadata v skupino Picobirnaviridae. Družina tvori dve podskupini. Prvo predstavljajo vretenčarske pikobirnaviruse – »Feline picobirnavirus«, »Human picobirnavirus«, »Human picobirnavirus« idr. Drugo predstavljajo »Diatom colony-associated dsRNA virusu 1« podobni virusi in nevretenčarski virusi skupine Picobirnaviridae, ki so jih opisali Shi in sod. (2016a). Oba virusa iz *Elysia cornigera* (GBRW01165272) in (GBRW01101679) spadata v skupino nevretenčarskih

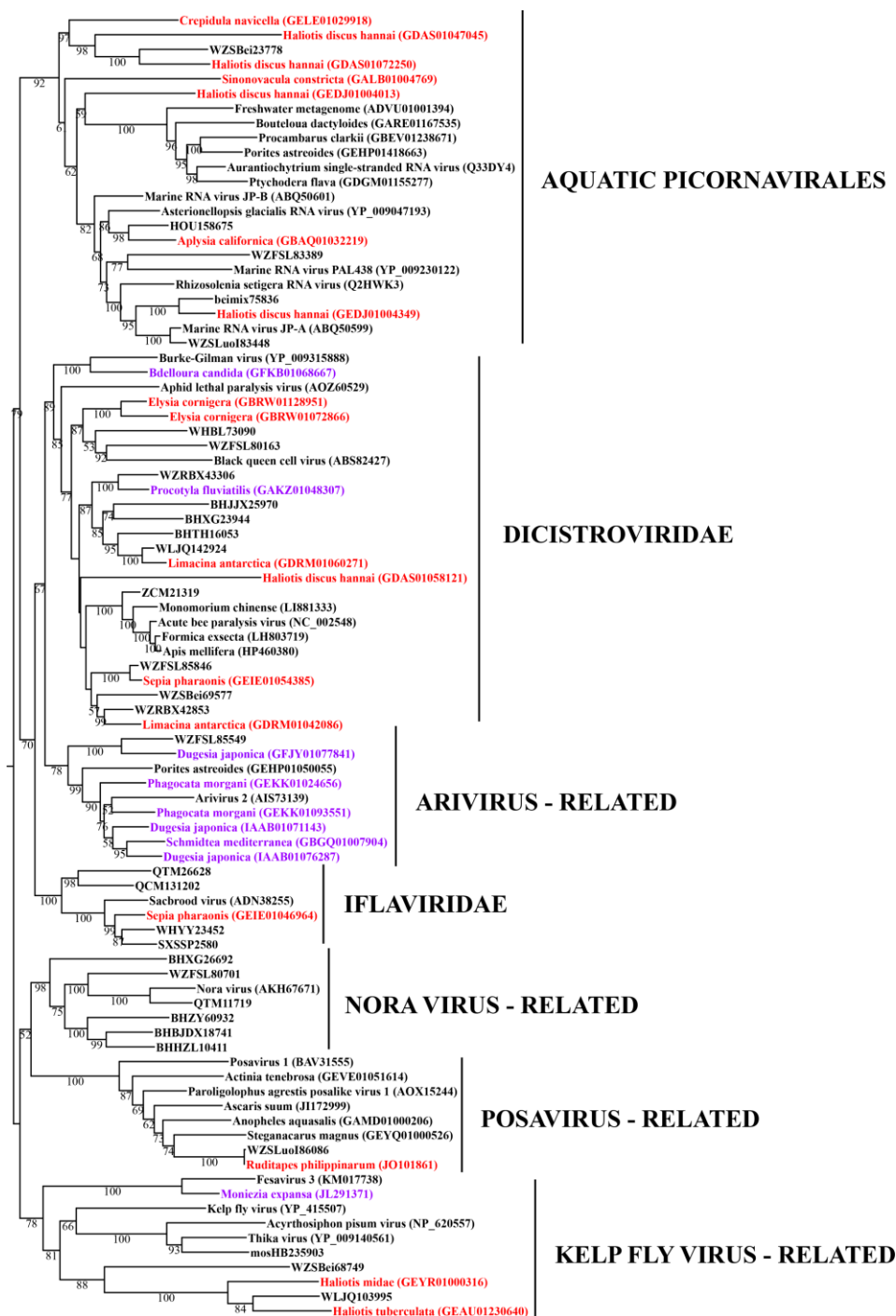
virusov v družini Picobirnaviridae. Skupina Shi in sod. (2016) ni odkrila vretenčarskim picobirnavirusom podobnih virusov, ki bi bili razširjeni pri nevretenčarjih. Našli pa so novo skupino vretenčarskim pikobirnavirusom podobnih virusov, ki so precej pogosti pri rakah in mehkužcih, ne najdemo pa jih pri kopenskih nevretenčarjih (Slika 4.20).



Slika 4.20: Filogenetsko drevo virusov klade **Partiti-Picobirna**, ki smo jih odkrili v genomih in transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo. Endogenizirani virusi so označeni z zvezdico.

4.3.2.2.12 Klada Picorna-Calici

V transkriptomih skupine Spiralia smo odkrili 26 virusov, ki spadajo v **klado Picorna-Calici**. Devet genomov pikornavirusov smo odkrili pri ploščatih črvih, 17 mehkužcih. Med genomi pikornavirusov pri ploščatih črvih dva spadata v skupino Dicistroviridae, eden v skupino Kelp fly virusu podobnih virusov, preostalih šest pa v skupino arivirusu podobnih virusov. Med arivirusom podobne viruse spadajo samo virusi, ki smo jih odkrili pri ploščatih črvih (*Dugesia sp.*, *Schmidtea sp.*, in *Phagocata sp.*). Skupina šestih virusov je 40 % identična z Arivirusom 2, med seboj so od 30 % do 50 % identični. Proteinskemu zaporedju *Dugesia japonica* (IAAB01076287) je zaporedje RdRp virusa *Schmidtea mediterranea* (GBGQ01007904) 51 % identično, zaporedje RdRp virusa *Dugesia japonica* (IAAB01071143) 52 % identično, zaporedje RdRp virusa *Phagocata morgani* (GEKK01093551) 48 % identično, zaporedje RdRp virusa *Phagocata morgani* (GEKK01024656) 45 % identično in proteinsko zaporedje RdRp *Dugesia japonica* (GFJY01077841) 30 % identično. Med dicistrovirusi pri skupini Spiralia je virus iz *Bdelloura candida* soroden z »Burke-Gilman virusom«, virus iz ploščatega črva *Procotyla fluviatilis* pa je soroden »Wenzhou shrimp virusu 4« (WZRBX43306). Virus iz transkriptoma *Moniezia expansa* je soroden »Fesavirusu 3«. V okviru filogenetske raziskave smo dosedanje podatke o raznolikosti pikornavirusov pri skupini Spiralia bistveno obogatili s podatki o raznolikosti pikornavirusov pri ploščatih črvih. Našli smo številne pikornaviruse pri skupini ploščatih črvov. Večina se jih je uvrstila med Arivirusom podobne viruse. Med 17 virusi, najdenimi v transkriptomih mehkužcev, se jih je sedem uvrstilo med »aquatic picornavirales«, šest med dicistroviruse, eden med iflaviruse, eden med posavirusom podobne viruse in dva med Kelp fly virusu podobne viruse (**Slika 4.21**).



Slika 4.21: Filogenetsko drevo virusov klade **Picorna-Calici**, ki smo jih našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«) v programu IQTREE (Nguyen in sod., 2015). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo.

4.3.2.2.13 Klada Qinvirus

Skupina Shi in sod. (2016a) je odkrila **klado Qinvirus** pri rakah, žuželkah, glistah in mehkužcih. V okviru naše filogenetske analize smo našli dva divergentna Qinvirusa v genomih ploščatih črvov *Macrostomum lignano* in *Echinostoma caproni*. Tako smo dokazali, da so qinvirusi razširjeni tudi pri ploščatih črvih oz. čez celotno skupino Spiralia (**Dodatek B; B.5, Slika B.54**).

4.3.2.2.14 Klada Toti-Chryso

Klada Toti-Chryso obsega dve skupini. V prvo spadajo Totivirus, Trichomonasvirus, Victorivirus in Leishmanivirus, v drugo sodijo Giardiavirusu podobni virusi. V drugo skupino spadata dva totivirusa, ki smo ju odkrili pri ploščatih črvih *Schmidtea mediterranea* in *Phagocata morgani*. Virus sta sorodna »Wenzhou toti-like virusu 2« (beimix73231) iz mehkužcev v okviru raziskave (Shi in sod., 2016a). Proteinski zaporedji njunih RdRp sta 43 % identični. Zaporedje RdRp virusa iz *Schmidtea mediterranea* je 28 % identično s proteinskim zaporedjem RdRp »Wenzhou toti-like virusa 2« (beimix73231) (**Dodatek B; B.5, Slika B.56**).

4.3.2.3 Najstarejše skupine živali: ožigalkarji (Cnidaria), spužve (Porifera) in rebrače (Ctenophora)

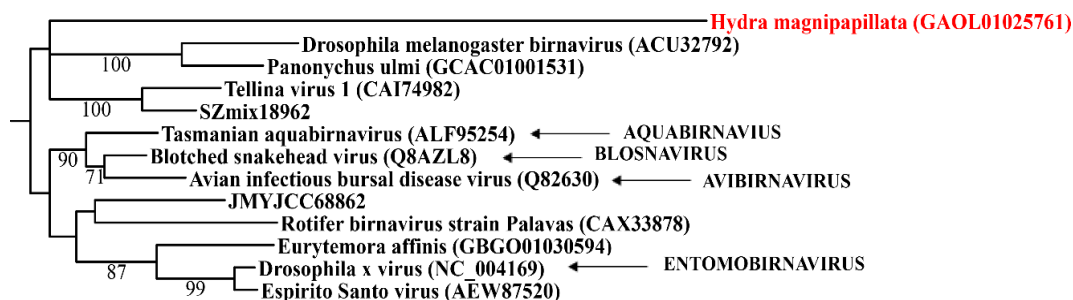
Najstarejše skupine živali zajemajo poleg organizmov Placozoa ožigalkarje (Cnidaria), spužve (Porifera) in rebrače (Ctenophora). Glede na ostale skupine nevretenčarjev (kožekrilci, metulji, raki, skupina Spiralia) so poleg najstarejših skupin žuželk (pražuželk) ena najslabše raziskanih skupin v genomskih in transkriptomskih podatkovnih bazah. V podatkovnih bazah obstaja 88 genomskih in transkriptomskih knjižnic za najstarejše skupine živali, 74 (84 %) jih pripada ožigalkarjem (Cnidaria). Tukaj moramo upoštevati tudi dejstvo, da niso vse transkriptomске knjižnice enako bogate z virusnimi zaporedji. V okviru naše filogenetske analize smo ožigalkarje (Cnidaria) označili z rdečo barvo, spužve (Porifera) z zeleno in rebrače (Ctenophora) z modro barvo.

4.3.2.3.1 Klada Bunya-Arena

Klada Bunya obsega tri ločene skupine: 1) Hantavirus, Emaravirus, Tospovirus in Orthobunyavirus, 2) Tenuvirusu in Phlebovirusu podobni virusi in 3) Nairovirusu podobni virusi. Prvo skupino (Hantavirus, Emaravirus, Tospovirus in Orthobunyavirus) predstavlja skupina rastlinskih in sesalskih virusov. Znotraj skupine so si zaporedja med seboj podobna (pri sesalskih virusih 60–80 %), skupine Hantavirus, Emaravirus, Tospovirus in Orthobunyavirus pa so si med seboj podobne le med 25–30 %. Druga skupina, ki obsega Tenuvirusu in Phlebovirusu podobne viruse, je širše razširjena med členonožci (klopi, komarji), rastlinami (riž) in sesalci (človekom). Tretja skupina (Nairovirusu podobni virusi) je razširjena pri nevretenčarjih – klopah, mehkužcih, pajkovcih, kolobarnikih in strunarjih (Chordata). V transkriptomu rebrače *Mnemiopsis leidyi* (GFAT) smo odkrili virus, ki smo ga sestavili iz številnih manjših zaporedij (GFAT01115010 + GFAT01138919 + GFAT01007784 + GFAT01017202 + GFAT01031184 + GFAT01090215 + GFAT01127526). V filogenetskem drevesu je ta virus označen z modrim pravokotnikom. Virus spada v skupino Phlebovirusu in Tenuvirusu podobnih virusov. Zaporedje virusa s katalitično domeno TNDD je 30–35 % identično s trenutno znanimi virusi skupine Bunyavirales. V drevesu se je povezalo skupaj s »Hubei blood fluke virusom 1« (HBXXC56361) (Shi in sod., 2016a) iz transkriptoma ploščatih črvov in »Beihai bunya-like virusom 2« (BHZC35426) iz transkriptoma mehkužcev (Mollusca) (Shi in sod., 2016a). V transkriptomu ožigalkarja *Rhopilema esculentum* smo našli nukleotidno zaporedje virusa (GEMS01099264), ki je glede na proteinsko zaporedje RdRp ranga 30 % identičnosti (pri slabi E-vrednosti) z ostalimi trenutno znanimi virusi skupine Bunyavirales (**Dodatek B; B.6, Slika B.58**).

4.3.2.3.2 Klada Birna

Klada Birna obsega dve ločeni skupini. V prvi so *Drosophila melanogaster* birnavirusi, ki so razširjeni pri žuželkah, pajkovcih (*Panonychus ulmi*) in mehkužcih (*Tellina tenuis*). V drugo skupino spadajo klasificirani birnavirusi, Aquabirnavirus, Avibirnavirus, Entomobirnavirus in Blosovirus, ki so razširjeni predvsem pri vretenčarjih, razen entomobirnavirusa, ki je razširjen pri žuželkah. V transkriptomu ožigalkarja *Hydra magnipapillata* smo našli birnavirus, ki spada med nevretenčarske birnaviruse (**Slika 4.22**).



Slika 4.22: Filogenetsko drevo klade Birna. Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) z modro barvo.

4.3.2.3.3 Klada Hepe-Virga

Pri kladi Hepe-Virga smo našli samo viruse, ki spadajo v virusno družino Hepeviridae in red Tymovirales, zato smo izrisali ločeni filogenetski drevesi za skupini Hepeviridae in Tymovirales. Pri družini Hepeviridae se jasno ločita dve skupini. Prvo predstavljajo Hepevirusu podobni virusi, ki so jih odkrili Shi in sod. (2016a) in so razširjeni predvsem pri raki in mehkužcih. Drugo predstavljajo klasificirani hepevirusi, ki so razširjeni predvsem pri vretenčarjih (ribe in sesalci). V filogenetski analizi hepevirusov pri najstarejših živalih smo pokazali, da so hepevirusi razširjeni tudi pri najstarejših skupinah živali. Že Waldron in sod. (2017) so pokazali prisotnost virusne skupine Hepeviridae oz. Hepevirusom podobnih virusov pri spužvi *Halichondria panicea*. V transkriptomih ožigalkarjev *Craterolophus convolvulus* in *Corynactis australis* smo našli virusa, ki spadata v skupino hepevirusom podobnih virusov, razširjenih pri raki in mehkužcih. V transkriptomu ožigalkarja *Anthopleura elegantissima* smo našli virus, soroden sesalskemu »Bastrovirus-like virus«. V transkriptomih spužev smo našli dva hepevirusa, eden se je uvrstil v skupino hepevirusom podobnih virusov, drugi iz transkriptoma *Leucosolenia sp.* spada v skupino klasificiranih hepevirusov in je soroden »Wenzhou hepe-like virusu 2« (WZFSL85851) iz transkriptoma mehkužcev (Shi in sod., 2016a). Red Tymovirales obsega rastlinske in nekaj glivnih virusov. V transkriptomu spužve *Porites asteroides* smo našli »Garlic virusu A« soroden viruse. »Garlic virus A« je bil izoliran iz česna (*Allium sativum*). Njuni RdRp sta 44 % identični (Dodatek B; B.6, Slika B.59 in Slika B.70).

4.3.2.3.4 Klada Luteo-Sobemo

Klada Luteo-Sobemo obsega dve ločeni skupini. Obe skupini sta razširjeni pri žuželkah (dvokrilci, kožekrilci, komarji) in mehkužcih. V okviru naše filogenetske analize smo pokazali, da je klada prisotna tudi pri ožigalkarjih in spužvah. V transkriptomih *Aurelia aurita* in *Nematostela sp.* smo našli isti virus, katerega proteinsko zaporedje RdRp je 33 % identično »Humaita-Tubiacanga virusu«. Virus »Caledonia beadlet anemone sobemo-like virus 1« in »Barns Ness breadcrumb sponge sobemo-like virus 2« sta se umestila v klaster skupaj s »Carascovirusom« (Dodatek B; B.6, Slika B.60).

4.3.2.3.5 Klada Narna-Levi

Klada Narna-Levi obsega dve ločeni virusni družini – Narnaviridae in Leviviridae. Filogenetski drevesi smo naredili za vsako virusno družino posebej. Družina Narnaviridae je razdeljena na tri skupine (rodove) – Narnavirus, Mitovirus in »Ourmiavirusu podobni virusi«. V okviru filogenetske analize smo dokazali razširjenost skupine Narnaviridae tudi pri najstarejših skupinah živali. V transkriptomih ožigalkarjev smo odkrili sedem virusov, med njimi *Hydractinia symbiolongicarpus* (GCHW01014342+GCHW01012278), *Calvadosia cruximelitensis* (HAHC01099958), *Lucernaria quadricornis* (HAHD01054073) in *Myxobolus cerebalis* (GBKL01052498). Waldron in sod. (2017) so pri spužvi *Halichondria panicea* odkrili nove viruse, ki so jih poimenovali »Barns Ness breadcrumb sponge narna-like virus« (Waldron in sod., 2017). V filogenetskem drevesu smo izpostavili pet virusov; »Barns Ness breadcrumb sponge narna-like virusi« 2, 3, 5, 8 in 7. Gre za divergentna zaporedja, ki so v rangu neke od 30–40% identični ostalim znanim virusom družine Narnaviridae. Družina Leviviridae kot skupina RNA bakteriofagov naj bi predstavljala evolucijskemu predniku skupine Narnaviridae podobno skupino. So najbližji

sorodniki družine Narnaviridae. Shi in sod. (2016a) so dokazali njihovo bogato razširjenost tudi pri raki in mehkužcih. Vendar so levivirusi bakterijskega izvora. Našli so bakterijske viruse iz mikrobioma nevretenčarjev. V transkriptomu korale *Porites australiensis* smo našli levivirus, ki je glede na proteinsko zaporedje RdRp 53 % identičen virusu iz transkriptoma pajkovca – »Beihai levi-like virus 30« (HOU142311) (**Dodatek B; B.6, Slika B.61 in B.74**).

4.3.2.3.6 Klada Tombus-Noda

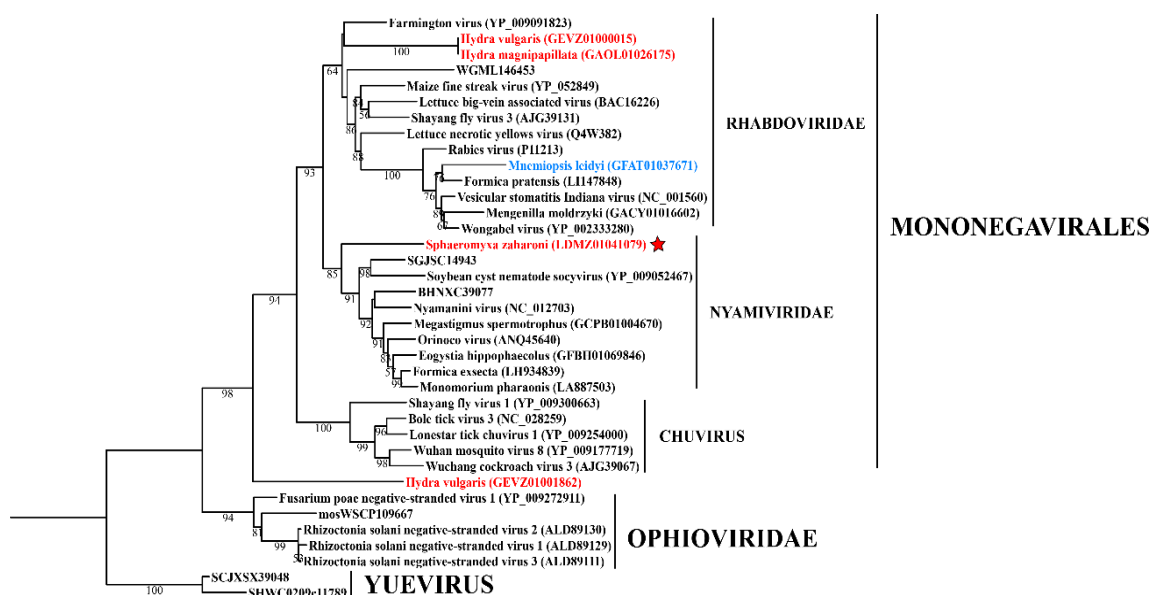
Pri **kladi Tombus-Noda** smo naredili filogenetska devesa za družini Nodaviridae in Tombusviridae. Družina Nodaviridae vsebuje tri ločene skupine. Prva predstavlja Nodamura virusu sorodne viruse. Gre za skupino virusov, ki so razširjeni predvsem pri žuželkah. Drugo skupino predstavljajo virusi, ki so razširjeni pri žuželkah ribah. Tretja skupina »Le Blanc nodavirus« podobnih virusov naj bi bila razširjena pri glistah (Nematoda). **Družina Nodaviridae** naj bi bila razširjena predvsem pri žuželkah, glistah in ribah. Shi in sod. (2016a) so dokazali bogato razširjenost skupine pri vodnih nevretenčarjih – raki in mehkužcih. V okviru naše raziskave smo dokazali, da so nodavirusi prisotni tudi pri najstarejših skupinah živali. V transkriptomu spužve *Oscarella sp.* smo našli divergenten nodavirus, ki se je povezal skupaj z »Nodamura virusom«, »Newington virusom« in »Flock house virusom«. **Družina Tombusviridae** se loči na dve skupini. Prva obsega »Diaphorina citri associated C virusu« podobne viruse. Diaphorina citri associated C virus je segmentiran virus, ki je bil izoliran iz žuželke *Diaphorina citri* (angl. »Asian citrus psyllid«). V drugo spadajo klasificirani rastlinski tombusvirusi (Tombusvirus, Zeavirus, Machlomovirus, Panicovirus idr.). Shi in sod. (2016a) so pokazali, da so tombusvirusi razširjeni ne samo pri žuželkah, temveč tudi pri glistah, pajkovcih, najbolj bogato pa pri vodnih nevretenčarjih – raki in mehkužcih. Waldron in sod. (2017) so pokazali, da so tombusvirusi razširjeni tudi pri spužvi *Halichondria panicea* in so jih poimenovali kot »Barns Ness breadcrumb sponge tombus-like virus 1-3«. Omenjene viruse smo vključili v filogenetsko analizo in analizo strukture njihovega genoma. V transkriptomih ožigalkarjev *Porites astreoides* in *Craterolophus convolvulus* (HAGZ01003353) smo odkrili divergentna tombusvirusa (**Dodatek B; B.6, Slika B.62 in Slika B.72**).

4.3.2.3.7 Klada Ophio

Klada Ophioviridae obsega dve ločeni skupini. V prvi so zajeti rastlinski ophiovirusi. Drugo skupino predstavljajo glivni virusi. Za filogenetsko analizo smo vzeli proteinsko zaporedje RdRp virusov, izoliranih iz gliv *Fusarium poae* in *Rhizoctonia solani*, znanih rastlinskih patogenih gliv. Tako kot pri skupini Partitiviridae naj bi pri evoluciji skupine Ophioviridae horizontalni prenos virusov med gostitelji igral pomembno vlogo. V genomu ožigalkarja *Sphaeromyxa zaharoni* (LDMZ01009652) smo našli endogeniziran ophiovirus. Ker je bilo zaporedje najdeno v genomu organizma, pomeni, da je *Sphaeromyxa zaharoni* dejanski gostitelj omenjenega virusa. Dokazali smo torej, da so gostitelji ophiovirusov, ne samo glive in rastline, temveč tudi najstarejše skupine živali. Proteinsko zaporedje RdRp virusa je le 25 % identično z ostalimi ophiovirusi. Glede na novo sliko razširjenosti družine Ophioviridae in njeni homologiji s klado Mononegavirales lahko sklepamo o njunem skupnem predniku (**Dodatek B; B.6, Slika B.63**).

4.3.2.3.8 Klada Mono-Chu

Glede na naša sklepanja in rezultate Shi in sod. (2016a) o skupnem izvoru klad Mono-Chu Ophio in Yuevirus smo v filogenetsko analizo **klade Mono-Chu** pri najstarejših skupinah živali vključili tudi skupini Ophioviridae in Yuevirus. Drevo smo koreninili pri kladi Yuevirus – najstarejši skupini. Opazimo tri ločene skupine z monofiletskim izvorom. Klada Mono-Chu oblikuje tri skupine. V okviru klade Mono-Chu smo v transkriptomih ožigalkarja *Hydra sp.* našli tri genome virusov. Eden izmed njih (najden v dveh različnih vrstah *Hydra vulgaris* in *Hydra magnapapillata*) je rhabdovirus. Virus, iz transkriptoma *Hydra vulgaris* – (GEVZ01001862) predstavlja enega izmed najstarejših predstavnikov skupine Mono-Chu. Proteinsko zaporedje njegove RdRp je 23–26 % identično z ostalimi predstavniki klade Mono-Chu. Gre za najdaljši doslej odkrit mononegavirus (Rhabdovirus). Poleg omenjenega smo našli tudi divergenten genom nyamivirusa, integriran v genom *Sphaeromyxa zaharoni*, kar pomeni preteklo aktivno okužbo organizma z nyamivirusom (**Slika 4.23**).



Slika 4.23: Filogenetsko drevo virusov klade **Mono-Chu**, ki smo jih našli v genomih in transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo. Endogenizirani virusi so označeni z zvezdico.

4.3.2.3.9 Klada Orthomyxo

Klada Orthomyxo pri najstarejši skupini živali tvori tri skupine. Prva obsega ISA virus, viruse gripe in togotoviruse. Z izjemo »Sinu virusa«, ki je bil izoliran iz komarja, in »Infectious salmon anemia virusa«, ki je bil izoliran iz ribe (atlantskega lososa), gre za skupino sesalskih virusov. Druga skupina obsega quaranjaviruse (Li in sod. 2015). Tretja skupina zajema orthomyxoviruse vodnih nevretenčarjev (Shi in sod., 2016a). Našli smo dva zelo divergentna orthomyxovirusa – enega v transkriptomu rebrače *Mncmiopsis leidy* (GFAT01014782) in drugega v transkriptomu ožigalkarja *Hydra vulgaris* (HAAC01043982). Proteinsko zaporedje RdRp virusa, najdenega v transkriptomu *Mncmiopsis leidy*, je 26 % identično proteinskemu zaporedju RdRp »Barns Ness dogwhelk orthomyxo-like virus 1«, ki ga prikaže kot edini zadetek ob iskanju homolognih zaporedij z orodjem blastp. Proteinsko zaporedje virusa RdRp virusa, najdenega v transkriptomu *Hydra vulgaris*, je samo 24 % identično proteinskemu zaporedju RdRp »Hubei orthomyxo-like virus 5« (WHBL85127) (Dodatek B; B.6, Slika B.64).

4.3.2.3.10 Klada Partiti-Picobirna

Klada Partiti-Picobirna obsega dve ločeni RNA virusni družini – Partitiviridae in Picobirnaviridae. V transkriptomih najstarejših skupin živali smo odkrili tri partitiviruse, enega v transkriptomu spužve *Leucosolenia* in dva v transkriptomih ožigalkarjev *Madracis auretenra* in *Porites astreoides*. Waldron in sod. (2017) so odkrili tudi divergenten picobirnavirus v transkriptomu spužve *Halichondria panicea* (»Barns Ness breadcrumb sponge picobirna-like virus 1«). Proteinsko zaporedje RdRp virusa, najdenega v transkriptomu korale *Porites astreoides* (GEHP01449975), je 21–26 % identično proteinskemu zaporedju RdRp katerega koli trenutno znanega partitivirusa. Zaporedje, ki smo ga našli v transkriptomu spužve *Leucosolenia sp.*, spada v skupino deltapartitivirusov, vendar gre za divergentno zaporedje. Proteinsko zaporedje RdRp »Barns ness breadcrumb sponge picobirna-like virus 1« je 25–30 % identično z ostalimi trenutno znanimi picobirnavirusi (Dodatek B; B.6, Slika B.65).

4.3.2.3.11 Klada Picorna-Calici

Klada Picorna-Calici zajema več skupin: »aquatic picornavirales«, Dicistroviridae, Iflaviridae, Caliciviridae, Kelp fly virusu podobni virusi, Arivirusu podobni virusi, Nora virusu podobni virusi, Posavirusu podobni virusi in Caliciviridae. Poleg omenjenih smo v

naši filogenetski analizi odkrili tudi nove skupine. V transkriptomih najstarejših skupin živali smo našli 51 pikornavirusov. Od tega so tri identificirali Waldron in sod. (2017) v transkriptomu spužve. Ostalih 48 smo našli v okviru naše analize. 21 smo odkrili v transkriptomih spužev, 26 v transkriptomih ožigalkarjev in enega v transkriptomih rebrač. Izmed 48 odkritih genomov virusov jih je bilo 12 (25 %) uvrščenih v skupino »aquatic picornavirales«, 6 v skupino Dicistroviridae (12,5 %), 6 v skupino arivirusom - podobnih virusov (12,5%), eden v skupino Iflaviridae, eden v skupino Nora virusu podobnih virusov, pet (10 %) v skupino posavirusu podobnih virusov in pet v skupino Klep fly virusu podobnih virusov (10 %). Kar 12 (25 %) pikornavirusov spada v nove skupine. Ena izmed njih se poveže skupaj z virusom »Hubei tetragynathia maxillosa virus 3« (SSZZ2994). Pri analizi Shi in sod. (2016a) je skupino, ki vsebuje omenjen virus, povežalo z družino Secoviridae (**Dodatek B; B.6, Slika B.66**).

4.3.2.3.12 Klada Reo

V transkriptomih najstarejših skupin živali smo odkrili tudi genom virusa, ki spada v **klado Reo**. Reoviridae je široko razširjena družina in je prisotna pri nevretenčarjih, vretenčarjih, glivah in rastlinah. Virus smo odkrili v transkriptomu ožigalkarja *Craterolophus convolvulus* in se poveže skupaj z »Eyach virusom«, »Colorado tick fever virusom« in glivnimi reovirusi (Mycoreovirus) (**Dodatek B; B.6, Slika B.69**).

4.3.2.3.13 Klada Toti-Chryso

Klada Toti-Chryso obsega dve ločeni skupini. Prva vsebuje Gardiavirusu podobne viruse in je širše razširjena med bičkarji (*Giardia lamblia*), žuželkami (komarji, mravlje – *Camponotus yamaokai*) in raki. Druga skupina obsega klasificirane glivne totiviruse in totiviruse, ki okužujejo enoceličarje. Shi in sod. (2016a) so pokazali, da je skupina »Gardiavirusu podobnih virusov« močno razširjena pri žužkah, pajkovcih, glistah in mehkužcih. Druga skupina klasificiranih Totivirusov, Chrysovirusov, Leishmaniavirusov in Victorivirusov je poleg gliv in bazalnih evkariontov razširjena tudi pri žužkah. V transkriptomih ožigalkarjev smo našli štiri totiviruse, ki se povežejo skupaj z »Beihai razor shell virus 4« (BH2C37385). Pri Shi in sod. (2016a) se omenjeni virus poveže z »Diatom colony associated dsRNA virus 17« (**Dodatek B; B.6, Slika B.73**).

4.3.2.3.14 Klada Poty

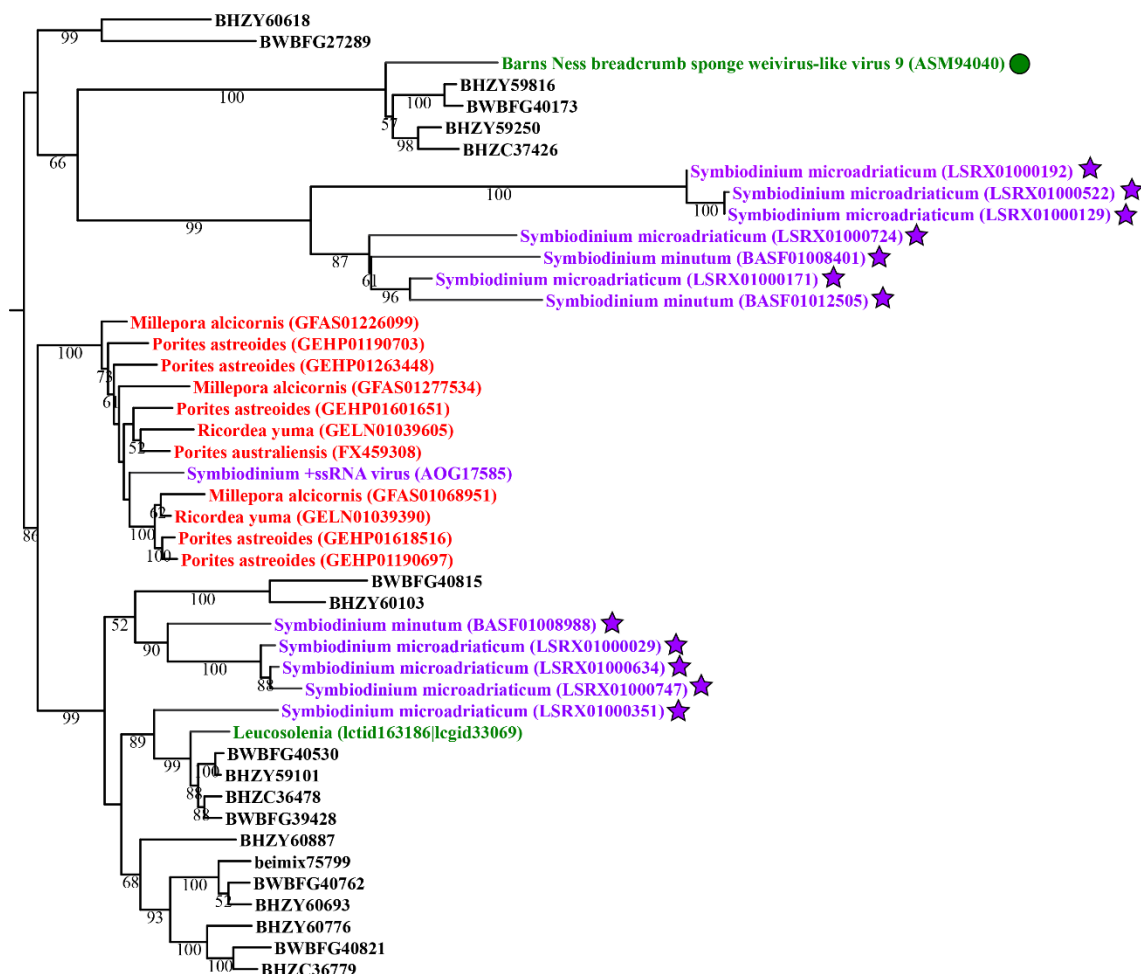
Družina Potyviridae je največja skupina rastlinskih virusov. V okviru filogenetske analize smo v transkriptomu spužve *Leucosolenia* sp. našli virus, ki spada v skupini Bymovirusa in Macluravirusa. Gre za divergenten virus, katerega proteinsko zaporedje RdRp je 30 % identično zaporedju katerega koli Potyvirusa. Dokazali smo, da je skupina Potyviridae razširjena tudi pri najstarejših skupinah živali (Metazoa) (**Dodatek B; B.6, Slika B.68**).

4.3.2.3.15 Klada Qinvirus

V transkriptomu ožigalkarja *Alatina alata* smo našli virus, ki spada v **klado Qinvirus**. Shi in sod. (2016a) so odkrili qinviruse v transkriptomih žuželk, rakov in mehkužcev. Mi smo ga odkrili tudi pri najstarejših skupinah živali. Qinvirus iz *Alatina alata* (GEUJ01027564) ima ohranjeno strukturo genoma qinvirusov (**Dodatek B; B.6, Slika B.67**).

4.3.2.3.16 Klada Weivirus

Klada Weivirus je bila prvič definirana pri analizi Shi in sod. (2016a) kot klada, ki je razširjena izključno pri mehkužcih. V naši raziskavi smo našli 25 weivirusov. 11 od teh v transkriptomih ožigalkarjev, 12 pa v genomih in enega v transkriptomu alveolata *Symbiodinium*. Našli smo tudi weivirus v transkriptomu spužve *Leucosolenia* sp. *Symbiodinium* je alveolat, sodi med starejše skupine evkariontov in je simbiot koral. Zaradi narave sožitja ne moremo trditi, da so genomi virusov, ki smo jih našli v genomu *Symbiodinium*, dejansko povezani z gostiteljem ali pa jih je *Symbiodinium* sp. prevzel od koral. Weivirusi jasno oblikujejo tri ločene skupine. Waldron in sod. (2017) so odkrili weivirus pri spužvi *Halichondria panicea* – »Barns Ness breadcrumb sponge weivirus-like virus 9« (**Slika 4.24**).



Slika 4.24: Filogenetsko drevo virusov klade **Weivirus**, ki smo jih našli v genomih in transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo. Endogenizirani RNA virusi so označeni z zvezdico. Virusi iz Waldron in sod. (2017) so označeni s krogci.

4.3.2.3.17 Klada Yuevirus

Klada Yuevirus je evlucijsko starejša od klade Ophio. Odkrili so jo Shi in sod. (2016a). Gre za najmanjšo klado trenutno znanih virusov, ki je doslej zajemala samo dva predstavnika. Genom je sestavljen iz dveh segmentov. Prvi kodira zapis za RdRp. Doslej je bila klada prisotna le v transkriptomih rakov. V okviru študije smo odkrili yuevirus v dveh transkriptomih *Aurelia aurita* in *Nematostella vectensis*, zaporedji sta 100 % identični. Zaporedje RdRp virusa je 50 % identično zaporedju virusa »Beihai sesarmid crab virus 3« (SCJXSX39048) (Dodatek B; B.6, Slika B.71).

4.4 Strukturna raznolikost genomov RNA virusov pri raziskanih skupinah nevretenčarjev

V okviru analize strukturne raznolikosti RNA virusov smo primerjali genomsko organizacijo novih RNA virusov s trenutno znanimi klasificiranimi in neklasificiranimi virusi, ki spadajo v isto virusno klado. Nukleotidna zaporedja novih RNA virusnih genomov, pridobljena v genomskih in transkriptomskih podatkovnih bazah, smo prevedli in analizirali proteinske domene v vsakem ORF z orodjem CDD (Marchler-Bauer in sod., 2017). Slike so priložene tudi v elektronski PDF verziji (CD).

4.4.1 Kopenski nevretenčarji

4.4.1.1 Kožekrilci (Hymenoptera)

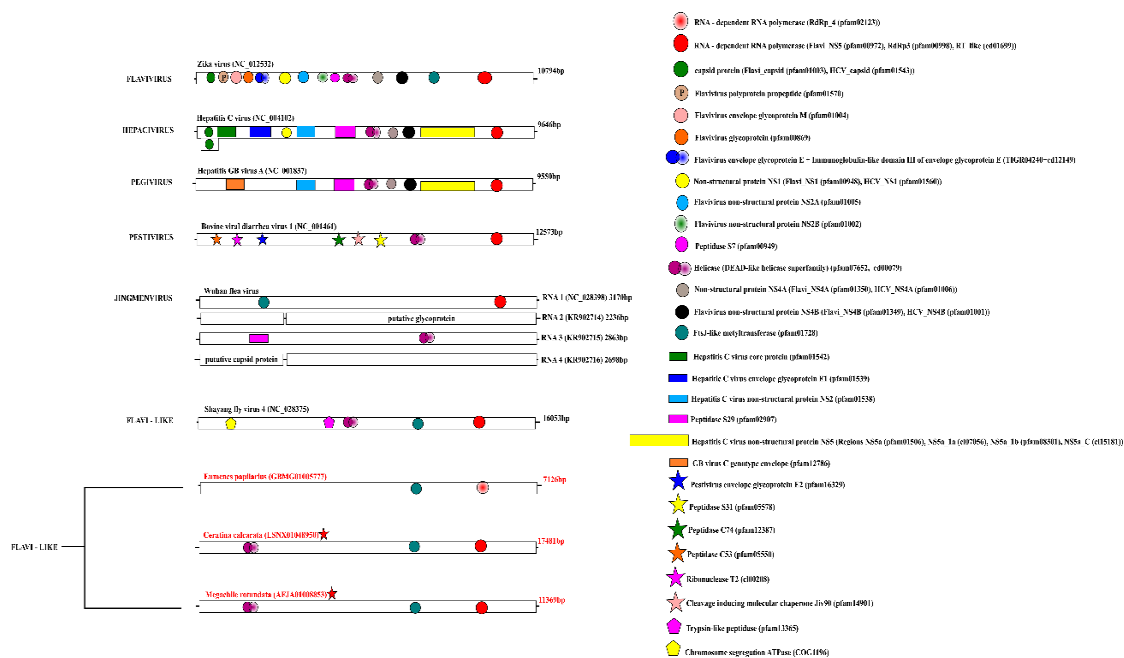
V iskanju novih RNA virusnih genomov pri kožekrilcih smo pregledali 432 genomskih in transkriptomskih podatkovnih baz. Celotne genome smo našli za predstavnike 13 virusnih klad: **Bunya-Arena**, **Flavi**, **Hepe-Virga**, **Luteo-Sobemo**, **Mono-Chu**, **Narna-Levi**, **Tombus-Noda**, **Orthomyxo**, **Partiti-Picobirna**, **Permutotetra**, **Picorna-Calici**, **Qinvirus** in **Toti-Chyryo**.

4.4.1.1.1 Klada Bunya-Arena

Pri skupini kožekrilcev smo v transkriptomih klade **Bunya-Arena** našli samo L segmente virusov klade Bunya-Arena, ki kodirajo zapis za Bunyavirusno specifično RdRp. Genomska struktura L segmenta se od trenutno znanih virusov razlikuje pri dveh virusih – *Cotesia glomerata* (GDJ001034815), kjer je viden dodaten ORF in *Formica truncorum* (LI396531), kjer je viden premik v bralnem okvirju (angl. »frameshift«) znotraj gena za RdRp (**Dodatek C; C.1, Slika C.1**).

4.4.1.1.2 Klada Flavi

Pri kladi **Flavi** je genomska struktura flavivirusov, ki smo jih našli pri kožekrilcih, sestavljena iz monocistronske RNA, ki kodira zapis za en poliprotein. Pri vseh treh flavivirusih je ohranjena helikaza naddružine II (SF2), »FtsJ podobna metiltransferaza« in RdRp. Ostale domene, najdene pri skupini Flaviviridae, niso bile prepoznane pri novih flavivirusih (**Slika 4.25**).



Slika 4.25: Genomska organizacija virusov klade **Flavi**, ki smo jih našli pri kožekrilcih (Hymenoptera). Endogenizirana RNA virusa sta označena z zvezdico. Novi RNA virusi so označeni z rdečo pisavo.

4.4.1.1.3 Klada Hepe-Virga

Znotraj klade **Hepe-Virga**, smo našli celotne genome pripadnikov skupine Negevirus. Tukaj je večina razlik v dodatnih ORF-jih, ki vsebujejo proteine brez doslej znanih proteinskih domen oz. znanih homologov. Izpostavila bi dva primera. Prvi vključuje virus iz čebele *Centris flavifrons* (JI025889 + JI025577). Genom je enako dolg kot tipičen negevirus (nad 9000 bp) in vsebuje samo nestrukturni ORF z virusno metiltransferazo, »FtsJ metiltransferazo«, helikazo in RdRp. Primer virusa iz transkriptoma mravlje *Formica execta* (LH937896) podobno vsebuje samo nestrukturni poliprotein s helikazo in RdRp. Iz ostalih struktur in struktur referenčnih negevirusov je razvidno, da je tekom evolucije

prišlo do dodajanja novih ORF-jev v genomu na policistronske mRNA (**Dodatek C; C.1, Slika C.2**).

4.4.1.1.4 Klada Luteo-Sobemo

Znotraj klade Luteo-Sobemo so bili genomi virusov, ki smo jih našli, strukturno podobni virusom, ki jih je odkrila skupina Shi in sod. (2016a). Peptidazna in RdRp domena sta bili na ločenih ORF-jih. Po dolžini so bili nekoliko krajši od tipičnih virusov klade Luteo-Sobemo (**Dodatek C; C.1, Slika C.3**).

4.4.1.1.5 Klada Mono-Chu

Pri kladi Mono-Chu je večina razlik v genomske strukturi novih RNA virusov, ki spadajo v klado Mono-Chu v nedefiniranih domenah, za katere ne poznamo biološke funkcije (**Dodatek C; C.1, Slika C.4**).

4.4.1.1.6 Klada Narna-Levi

Pri kladi Narna-Levi smo v okviru analize strukture genomov narnavirusov pri kožekrilcih našli štiri primere virusov z dodatnimi ORF-ji, ki se berejo v nasprotni smeri od ORF-ja z RdRp (angl. »ambisense«) – *Formica pratensis* (LJ195079), *Linepithema humile* (LI705470), *Fopius arisanus* (GBYB01012090) in *Ceratina chalybea* (GBPU01010966) (**Dodatek C; C.1, Slika C.5**).

4.4.1.1.7 Klada Tombus-Noda

Klada Tombus-Noda se deli na dve klasificirani virusni družini: Tombusviridae in Nodaviridae. Pri skupini Tombusviridae smo analizirali genomske strukture štirih tombusvirusov. Gre za genome, ki so pol krajši od povprečne dolžine genoma tombusvirusov in so sestavljeni iz dveh ORF-jev. Prvi ni okarakteriziran, drugi pa kodira protein RdRp. Razlika je predvsem v tem, da se RdRp nahaja na ločenem ORF, razen v primeru *Brachycistis timberlakei* (GAZU01004632), ko smo zaznali premik bralnega okvirja med slabo okarakteriziranim ORF in ORF, ki kodira zapis za RdRp.

Pri nodavirusih smo odkrili tri nesegmentirane viruse. Prvi iz transkriptoma *Chalybion californicum*) ima dva ORF-ja. Prvi ORF ni okarakteriziran (nima ustreznega homolognega zaporedja), drugi ORF vključuje RdRp. Drugi nodavirus iz *Linepithema humile* (LI681900) ima dva ORF na monopartitni RNA. Prvi kodira virusno metiltransferazo, drugi pa RdRp. Tretjega smo našli v transkriptomu *Stigmatomma oregonense* (GAXR01015452). Njegova ORF-ja se prekrivata. Prvi kodira RdRp, drugi pa RNA vezavni protein B2. Glede na dolžino zaporedja lahko sklepamo, da gre za prvi segment genoma. Ni še namreč znanega primera nodavirusa, katerega genom bi vključeval samo RdRp in protein B2 (**Dodatek C; C.1, Slika C.6**).

4.4.1.1.8 Klada Orthomyxo

Genom klade Orthomyxo vključuje šest ali osem segmentov. Naši zadetki so bili po genomske strukturi podobni »Quaranjafil virusu« s šestimi segmenti – prvi trije kodirajo tripartitno RdRp, peti kodira glikoprotein (**Dodatek C; C.1, Slika C.7**).

4.4.1.1.9 Klada Partiti-Picobirna

Klada Partiti-Picobirna obsega virusni skupini Partitiviridae in Picobirnaviridae. V okviru analize smo pri kožekrilcih našli dva primera celotnih genomov partitivirusov. Eden je nesegmentiran, pri katerem prvi ORF kodira RdRp, drugi ni okarakteriziran. Drugi genom ima klasično strukturo partitivirusov. Našli smo tudi druge primere partitivirusov, vendar samo segmente z RdRp. Med evolucijo je torej prišlo do segmentacije ali pa do zlepljanja fragmentov (**Slika 4.26**).

Arena, Hepe-Virga, Luteo-Sobemo, Mono-Chu, Narna-Levi, Tombus-Noda, Orthomyxo, Partiti-Picobirna, Permutotetra, Picorna-Calici in Toti-Chryso.

4.4.1.2.1 Klada Bunya-Arena

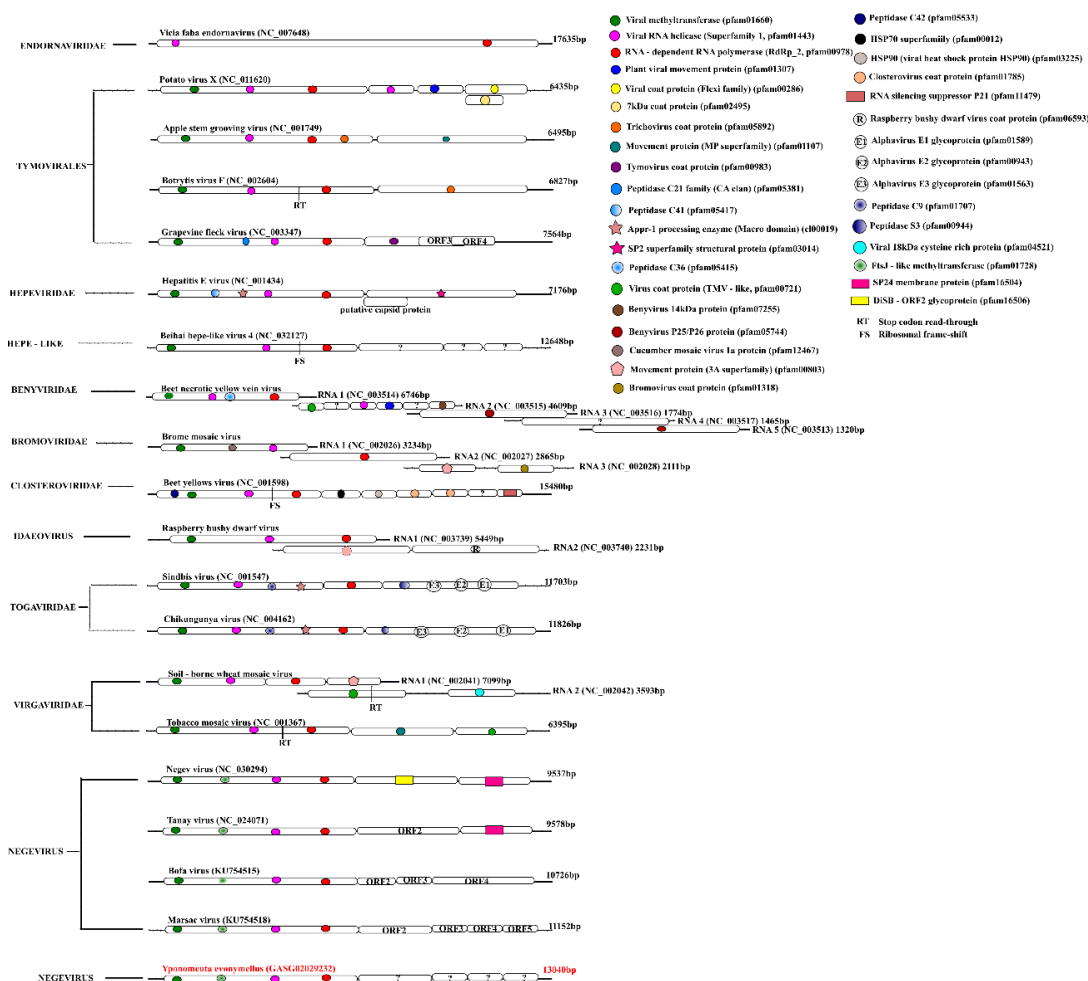
Znotraj klade **Bunya-Arena** smo pri metuljih in veččah našli samo L segmente bunyavirusov v okviru analize raznolikosti genomske strukture (**Dodatek C; C.2, Slika C.13**).

4.4.1.2.2 Klada Astro

Pri kladi **Astro** smo pri metuljih metulja *Amyelosis transitella* odkrili genom astrovirusa, ki je po strukturi podoben »Hubei leech virusu 1« (SZmix32467). Tako kot pri permutotetravirusih je pri evoluciji astrovirusov očitno veliko vlogo igral horizontalni prenos genov od skupine Alphetetraviridae (Hepe-Virga klada), Nidovirales in ostalih rastlinskih virusov (**Dodatek C; C.2, Slika C.12**).

4.4.1.2.3 Klada Hepe-Virga

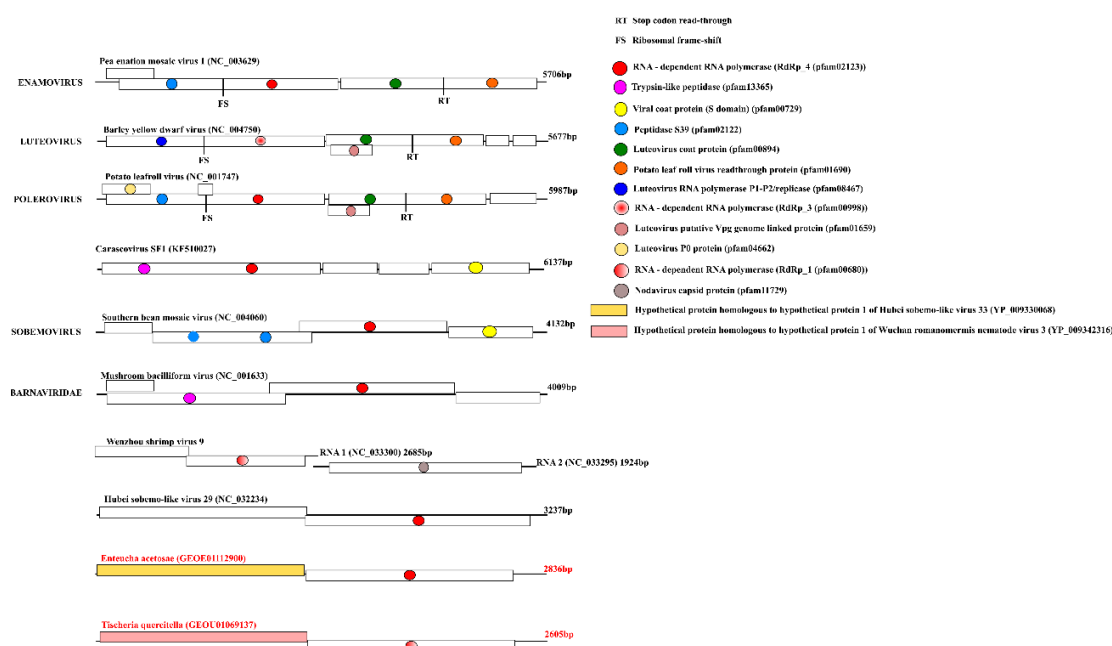
Znotraj klade **Hepe-Virga** smo pri metuljih in veččah (Lepidoptera) odkrili celoten genom negevirusa – *Yponomeuta evonymellus* (GASG02029232), ki je dolg 13040 bp, kar je več od povprečne dolžine genoma negevirusov (10kb). Strukturno je podoben »Marsac virusu« s petimi odprtimi bralnimi okvirji (ORF). Nestrukturni poliprotein negevirusov kodirajo proteini: FtsJ metiltransferaza, helikaza naddružine 1 in RdRp_2 (RdRp družine 2). Ostali ORF-ji so hipotetični, v proteinskih podatkovnih bazah ne obstaja funkcionalno okarakteriziran proteinski homolog (**Slika 4.27**).



Slika 4.27: Genomska organizacija virusov klade **Hepe-Virga**, ki smo jih našli pri metuljih in veččah (Lepidoptera). Novi RNA virusi so označeni z rdečo pisavo.

4.4.1.2.4 Klada Luteo-Sobemo

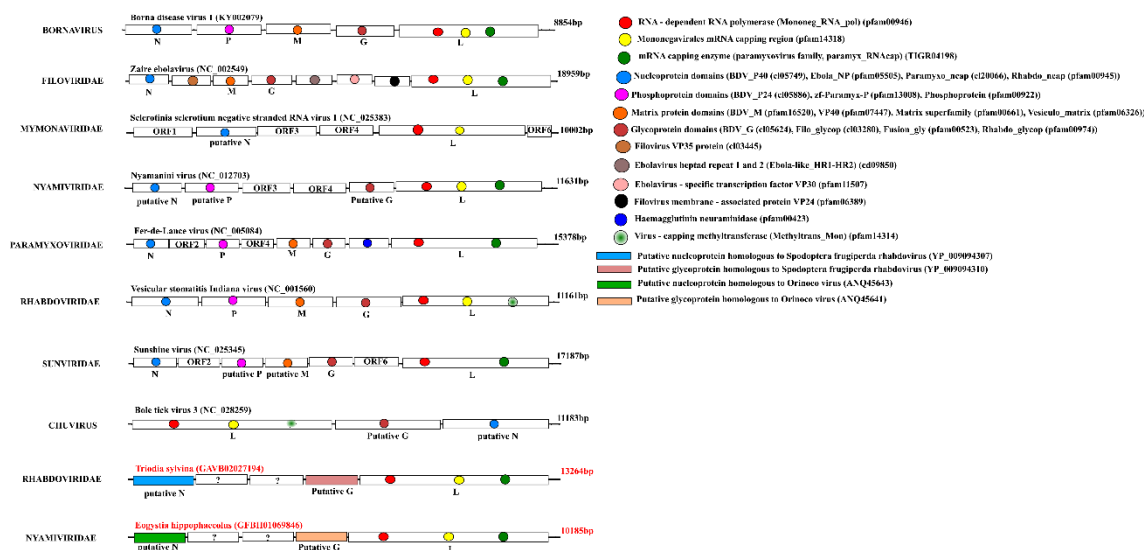
Pri kladi **Luteo-Sobemo** smo pri metuljih odkrili dva celotna genoma virusov pri *Enteucha acetosae* (GEOE01112900) in *Tischeria quercitella* (GEOU01069137). Ta genoma virusov sta nekoliko krajša od povprečne dolžine genoma klade Luteo-Sobemo. Strukturno sta podobna virusu »Hubei sobemo-like virus 29 (QTM12808)«. Monopartitna bicistronska RNA v vseh treh primerih genomov zajema dva ločena ORF. Prvi je najverjetneje strukturni in je slabše okarakteriziran. Drugi je nestrukturni in kodira RdRp. ORF1 virusa *Enteucha acetosae* (GEOE01112900) je homologen hipotetičnemu proteinu 1 »Hubei sobemo-like virusa 33«. ORF1 virusa *Tischeria quercitella* (GEOU01069137) je homologen hipotetičnemu proteinu 1 virusa »Wuchan romanomermis nematode virus 3«. Omenjena proteina je odkrila skupina Shi in sod. (2016a). Funkcionalno nista okarakterizirana (Slika 4.28).



Slika 4.28: Genomska organizacija virusov klade **Luteo-Sobemo**, ki smo jih našli pri metuljih in veččah (Lepidoptera).

4.4.1.2.5 Klada Mono-Chu

Pri metuljih smo znotraj klade **Mono-Chu** odkrili dva celotna genoma virusov pri *Triodia sylvina* (GAVB02027194) in *Eogystia hippophaecolus* (GFBH01069846). Njuna genomska organizacija ne odstopa od referenčnih virusov iz skupine Mononegavirales. Monopartitna RNA je sestavljena iz petih ORF-jev. Prvi ORF kodira nukleoprotein, drugi in tretji sta slabo okarakterizirana, četrti kodira glikoprotein in zadnji L poliprotein (L segment). Nukleoprotein in glikoprotein sta v primeru virusa *Triodia sylvina* (GAVB02027194) homologna »Spodoptera frugiperda rhabdovirusu«, v primeru virusa *Eogystia hippophaecolus* (GFBH01069846) pa »Orinoco virusu« (Slika 4.29).



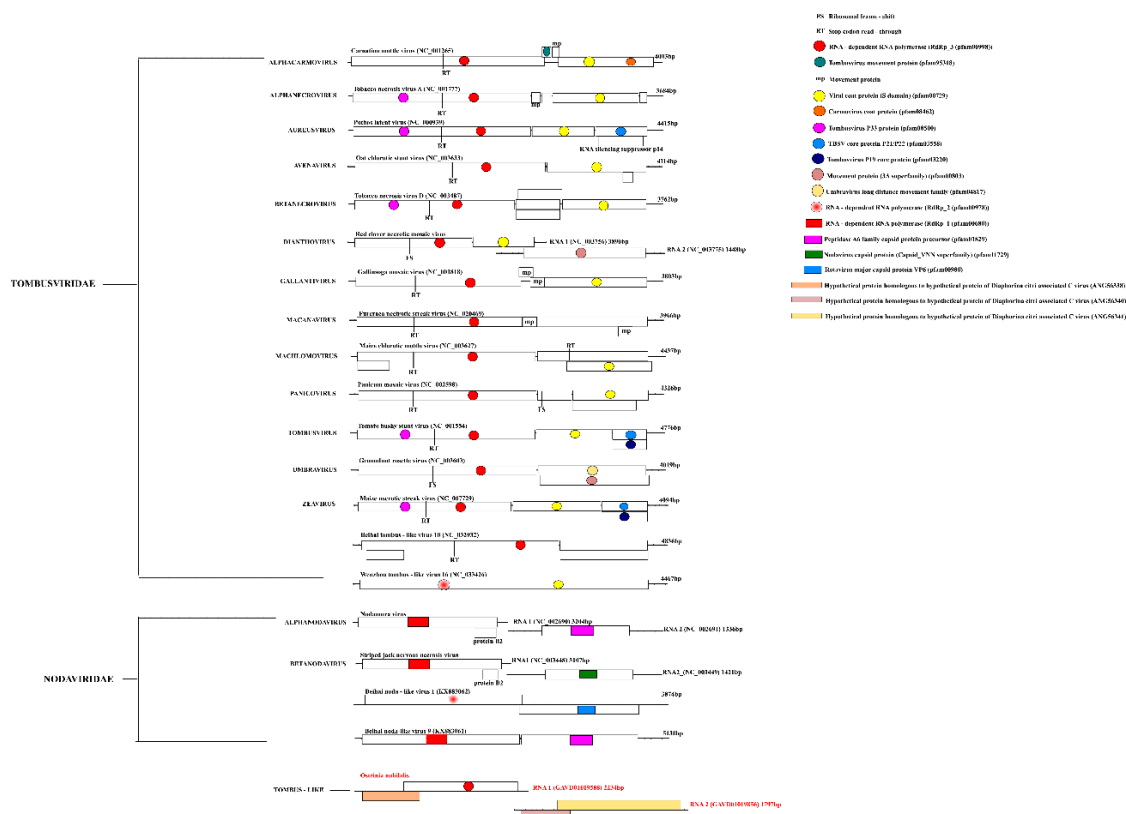
Slika 4.29: Genomska organizacija virusov klade **Mono-Chu**, ki smo jih našli pri metuljih in večšah (Lepidoptera).

4.4.1.2.6 Klada Narna-Levi

V okviru klade **Narna-Levi** pri metuljih je zanimiv narnavirus iz *Helicoverpa assulta* (ditrisiji, sovke – Noctuidae) (GBTA01004534). Genomska struktura tega virusa kodira samo en nestrukturni ORF z RdRp (**Dodatek C; C.2, Slika C.15**).

4.4.1.2.7 Klada Tombus-Noda

V okviru analize genomske strukture klade **Tombus-Noda** pri metuljih je zanimiv tombusvirus iz *Ostrinia nubilalis* (proseni večša) (GAVD01019588). Proteinsko zaporedje RdRp zadetka je sorodno »Diaphorina citri-associated C virusu«. »Diaphorina citri-associated C virus« je segmentiran tombusvirus. Sestavljen je iz bipartitne genomske RNA. Virus *Ostrinia nubilalis* (GAVD01019588) je tudi bipartiten – sestavljen iz dveh segmentov. Vsak segment je bicistronski, kar pomeni, da kodira dva ORF. ORF-ja, kodirana na vsakem od segmentov, se prekrivata. Drugi ORF na prvem segmentu kodira RdRp, ostali so homologni hipotetičnim proteinom »Diaphorina citri-associated C virus« (**Slika 4.30**).



Slika 4.30: Genomska organizacija virusov klade **Tombus-Noda**, ki smo jih našli pri metuljih in veščah (Lepidoptera).

4.4.1.2.8 Klada Orthomyxo

Znotraj **klade Orthomyxo** smo v transkriptomu metulja *Eogystia hippophaecilus* odkrili del genoma orthomyxovirusa. Orthomyxovirus je po genomski strukturi in proteinskih domenah podoben oz. soroden »quaranjavirusu«, vendar nismo uspeli najti drugega in šestega segmenta (**Dodatek C; C.2, Slika C.14**).

4.4.1.2.9 Klada Partiti-Picobirna

Pri analizi celotnih genomov virusov iz **klade Partiti-Picobirna** pri metuljih sta zanimiva dva partitivirusa. Genom partitivirusa iz metulja *Micropteryx calthella* (RNA1; GACP01000900, RNA2; GACP01021597) je segmentiran in ima strukturo genoma tipičnega partitivirusa. Genom virusa *Dryadulla visaliella* (GENH01106292) pa se razlikuje glede na genom referenčnih partitivirusov in genom novih virusov, ki jih je odkrila skupina Shi in sod. (2016a). Dolžina genoma je podobna vsoti dolžine vseh segmentov »Wuhan milipede virus 4«, torej okrog 6 kb. RNA kodira dva ORF-ja. Prvi ORF je daljši in nestrukturni. Zajema dve proteinski domeni – endonukleaza naddružine DDE in RdRp. Drugi ORF ni funkcionalno okarakteriziran. Vmes je daljša nekodirajoča regija (**Dodatek C; C.2, Slika C.17**).

4.4.1.2.10 Klada Permutotetra

Znotraj klade *Permutotetra* pri metuljih smo našli genoma virusov iz *Eogystia hippophaecolus* (GFBH01042206) in *Athetis lepigone* (GARC01021157). Genomski strukturi se nekoliko razlikujeta od referenčnih permutotetravirusov, njuna dolžina pa je tipična za permutotetraviruse. V obeh primerih je skoraj čez celotno dolžino genoma kodirana RdRp. Znotraj genoma se pri *Eogystia hippophaecolus* (GFBH01042206) nahajata drugi in tretji ORF. Drugi kodira virusni protein plašča (domena S) (angl. »*viral coat protein; S domain*«), tretji ni funkcionalno okarakteriziran. Pri *Athetis lepigone* (GARC01021157) se znotraj ORF z RdRp nahajata dva ORF-ja, med katerima je stop kodon, in se najverjetneje prevajata kot en poliprotein. CDD je prepoznala samo virusni protein plašča (domena S) (angl. »*viral coat protein; S domain*«) pred stop kodonom (**Dodatek C; C.2, Slika C.16**).

4.4.1.2.11 Klada Picorna-Calici

Večina virusov znotraj **klade Picorna-Calici** pri metuljih spada med iflaviruse. Vsi so kodirani z enim samim poliproteinom. Glavna razlika med referenčno strukturo genoma iflavirusa in iflavirusov, ki smo jih odkrili pri metuljih, je v dodatni Rhv domeni, razen v primeru iflavirusa *Zygaena fausta* (GAYB02008221), ki ima strukturo, podobno »Beihai picorna-like virusu 87«, ki je bil uvrščen med dicistroviruse. V tem primeru nista na N-terminalnem delu poliproteina Rhv domeni, temveč je helikaza naddružine 3, ki ji sledi peptidaza C3, RdRp₁, dve Rhv domeni in CRPV domena (**Dodatek C; C.2, Slika C.18**).

4.4.1.2.12 Klada Toti-Chryso

Genom totivirusa iz metulja *Helicoverpa assulta* (GBTA01007552), predstavnika **klade Toti-Chryso**, ima podobno dolžino kot referenčni genomi. Vendar se njegova struktura razlikuje od strukture referenčnih genomov. Monopartitna RNA je kodirana z enim ORF, ki vsebuje dve proteinski domeni; Rho faktor transkripcije in terminacije (angl. »*transcription termination factor Rho*«) in RdRp (**Dodatek C; C.2, Slika C.19**).

4.4.1.3 Najstarejše skupine žuželk: skakači (Collembola), dvorepke (Diplura), ščetinorepke (Zygentoma), Monocondylia

Pregledali smo 16 transkriptomov in štiri genome najstarejših skupin žuželk oz. t. i. pražuželk. Našli smo celotne genome predstavnikov klad: **Hepe-Virga, Narna-Levi, Tombus-Noda, Orthomyxo, Picorna-Calici** in **Qinvirus**.

4.4.1.3.1 Klada Narna-Levi

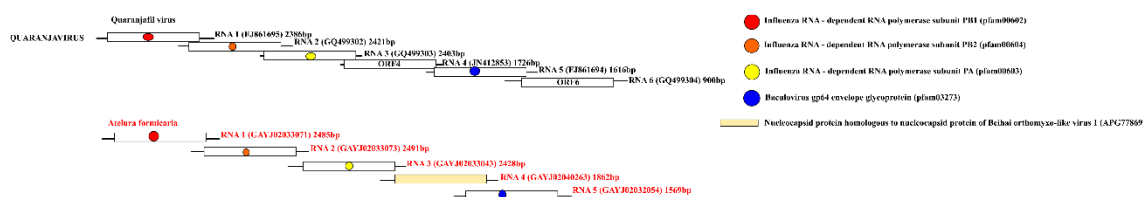
Med analizo **klade Narna-Levi** pri najstarejših skupinah žuželk smo našli genom narnavirusa pri *Tetrodontophora bielanensis* (GAXI). Genom tega narnavirusa ima tako kot referenčni en sam ORF, ki kodira RdRp (**Dodatek C; C.3, Slika C.20**).

4.4.1.3.2 Klada Hepe-Virga

V okviru **klade Hepe-Virga** velja pri najstarejših skupinah žuželk izpostaviti dva genoma pri *Anurida maritima* (GAUE02021637) in pri *Sminthurus viridis* (GATZ02022882). Genoma obeh negevirusov sta celotna. Njuna genomska struktura nekoliko odstopa od referenčne genomske strukture Negevirusov. Glavni nestrukturni poliprotein, kodiran s prvim ORF, zajema tri glavne proteinske domene, značilne za alfavirusno naddružino: virusno metiltransferazo, helikazo naddružine I in RdRp₂. Vendar v obeh primerih za virusno metiltransferazo (angl. »*viral methyltransferase*«) manjka FtsJ metiltransferaza. Pri negevirusu iz *Anurida maritima* (GAUE02021637) se pojavi dodatno odstopanje. Za glavnim nestrukturnim ORF, ki kodira poliprotein z RdRp pri referenčnih sevih, sledijo strukturni ORF, ki so pri »Bofa virusu podobnih« negevirusih in »Marsac virusu podobnih« negevirusih funkcionalno slabo okarakterizirani. Pri Negev in Tanay virusu podobnih virusih pa strukturni del genoma zajemata dva karakteristična ORF. Prvi strukturni ORF pri Negev virusu kodira DiSB-ORF2 glikoprotein, drugi strukturni ORF kodira protein SP24. Pri Tanay virusu je prvi strukturni ORF slabo funkcionalno okarakteriziran. Pri virusu iz *Anurida maritima* (GAUE02021637), katerega genom je podobno dolg kot genom povprečnega negevirusa, prvi strukturni ORF za glavnim jedrnim nestrukturnim ORF (ki kodira jedrni poliprotein) kodira protein SP24 (**Dodatek C; C.3, Slika C.21**).

4.4.1.3.3 Klada Orthomyxo

Znotraj **klade Orthomyxo** smo pri najstarejših skupinah žuželk odkrili genom orthomyxovirusa pri *Atelura formicaria*, ki je vseboval pet od šestih segmentov. Virus je strukturno podoben quaranjavirusom (**Slika 4.31**).



Slika 4.31: Genomska organizacija virusov klade **Orthomyxo**, ki smo jih našli pri najstarejših skupinah žuželk (Collembola, Diplura, Zygentoma, Monocondylia). Novi RNA virusi so označeni z rdečo pisavo.

4.4.1.3.4 Klada Picorna-Calici

Pri analizi genomske strukture klade **Picorna-Calici** pri najstarejših skupinah žuželk smo našli dva divergentna genoma: pikornavirus iz *Anurida maritima* (GAUE02021853) in pikornavirus iz *Occasjapyx japonicus* (GAXJ02019692). Genomska struktura virusa iz *Anurida maritima* (GAUE02021853) nekoliko odstopa od strukture trenutno znanih kelp fly virusu podobnih virusov. Znotraj monopartitne RNA z enim kodirajočim ORF smo odkrili dve helikazni domeni, nismo pa zaznali peptidazne domene. V genoski strukturi virusa *Occasjapyx japonicus* (GAXJ02019692) pa smo odkrili dodatno Rhv domeno (**Dodatek C; C.3, Slika C.24**).

4.4.1.3.5 Klada Qinvirus

Pri kladi **Qinvirus** genomska struktura qinvirusa iz *Anurida maritima* (GAUE02014037) ne odstopa od strukture ostalih trenutno znanih qinvirusov. Bipartitna RNA je monocistronska. Prvi segment kodira ORF s polimerazo. Drugi segment ni funkcionalno okarakteriziran, vendar je homologen »hipotetičnemu proteinu« »Wuhan insect virusa 15« (**Dodatek C; C.3, Slika C.23**).

4.4.1.3.6 Klada Tombus-Noda

Pri analizi genomske strukture virusov klade **Tombus-Noda**, ki smo jih našli pri najstarejših skupinah žuželk, sta zanimiva virusa iz *Atelura formicaria* (GAYJ) in iz *Tricholepidion gertschi* (GASO). Pri obeh genomih gre najverjetneje za segmentirana virusa, ki sta po genoski strukturi podobna »Diaphorina cotri-associated C virusu«. Vendar smo odkrili samo en segment. Na segmentu sta dva ORF, prvi ni funkcionalno okarakteriziran, drugi pa kodira RdRp (**Dodatek C; C.3, C.22**).

4.4.2 Vodni nevretenčarji

4.4.2.1 Raki (Crustacea)

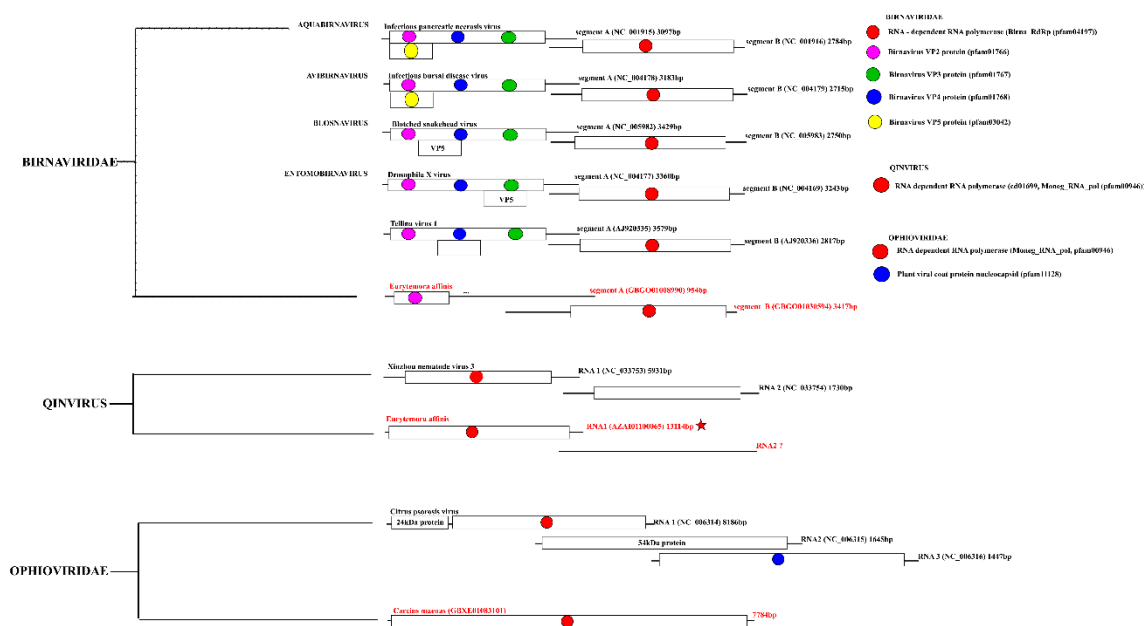
V okviru analize smo pregledali 196 genomov in transkriptomov. Našli smo celotne genome predstavnikov 14 virusnih klad: **Birna, Qinvirus, Ophio, Bunya-Arena, Flavi, Hepe-Virga, Luteo-Sobemo, Mono-Chu, Narna-Levi, Orthomyxo, Partiti-Picobirna, Permutotetra, Picorna-Calici** in **Toti-Chryso**.

4.4.2.1.1 Klada Bunya-Arena

V okviru klade **Bunya-Arena** smo pri raki odkrili samo celotne L segmente genoma virusov, uvrščenih v klado Bunya-Arena. Eden izmed njih – *Hyalellopsis stebbingi* (GEQJ01053822) – ima poleg bunyavirusno specifične RdRp tudi bunyavirusni L protein (**Dodatek C; C.4, Slika C.25**).

4.4.2.1.2 Klade Birna, Qinvirus in Ophio

Pri analizi genomov klade **Birna** pri raki smo odkrili dva segmenta birnavirusa iz *Eurytemora affinis* (GBG001030594). Segment B je kompleten, medtem ko smo odkrili samo fragment segmenta A, ki je vključeval Birnavirusni VP2 protein. Pri **Qinvirusih** smo odkrili samo prvi segment qinvirusa iz *Eurytemora affinis* (AZA101100065) z RdRp. Pri **Ophiovirusih** smo našli samo en segment pri ophiovirusu iz rakovice *Carcinus maenas* (GBXE01083101) (**Slika 4.32**).



Slika 4.32: Genomska organizacija virusov klad **Birna**, **Qinvirus** in **Ophioviridae**, ki smo jih našli pri rakih (Crustacea).

4.4.2.1.3 Klada Flavi

Pri genomski analizi flavivirusov (**klada Flavi**) pri rakih smo našli sedem virusov. Trije virusi imajo celotne genome: genom iz *Caridina multidentata* (IABX01135281), genom iz *Eriocheir sinensis* (GBZW01030122) in genom iz *Proasellus cavaticus* (HAES01086036). Genomska struktura virusa iz *Caridina multidentata* (IABX01135281) je divergentna glede na genomsko strukturo referenčnih virusov iz družine Flaviviridae. Genom kodira poliprotein z dvema proteinskima domenama – helikazo naddružine II (DEAD/DEAH box helikaza) in RdRp. Znotraj helikaze smo opazili premik bralnega okvirja. CDD prepozna Flavi NS5 polimerazo, ki je samo 26–28 % identična kateremu koli trenutno znanemu flavivirusu. Genom virusa iz *Eriocheir sinensis* (GBZW01030122) je divergenten glede na genom ostalih flavivirusov. Sestavljen je iz monopartitne RNA, ki vsebuje 2 ORF. Prvi ORF kodira flavivirusni glikoprotein (angl. »flavivirus glycoprotein«) (pfam00869) in glikoprotein E (TIGR04240), drugi ORF kodira helikazo DEAD/DEAH box, FtsJ metiltransferazo in RdRp. Genom virusa iz *Proasellus cavaticus* (HAES01086036) kodira en poliprotein z DEAD/DEAH box helikazo, FtsJ metiltransferazo in RdRp. Struktura je sorodna genomski strukturi flavivirusom podobnih virusov, kot je Shayang fly virus 4. Genomska strukturi se razlikujeta samo v N-terminalnih proteinskih domenah. Virus iz *Proasellus cavaticus* (HAES01086036) vsebuje domeno »reticulocyte-binding/rhoptry protein«, medtem ko »Shayang fly virus« vsebuje »Chromosome segregation ATPase« in tripsinu podobno peptidazo (**Dodatek C; C.4, Slika C.26**).

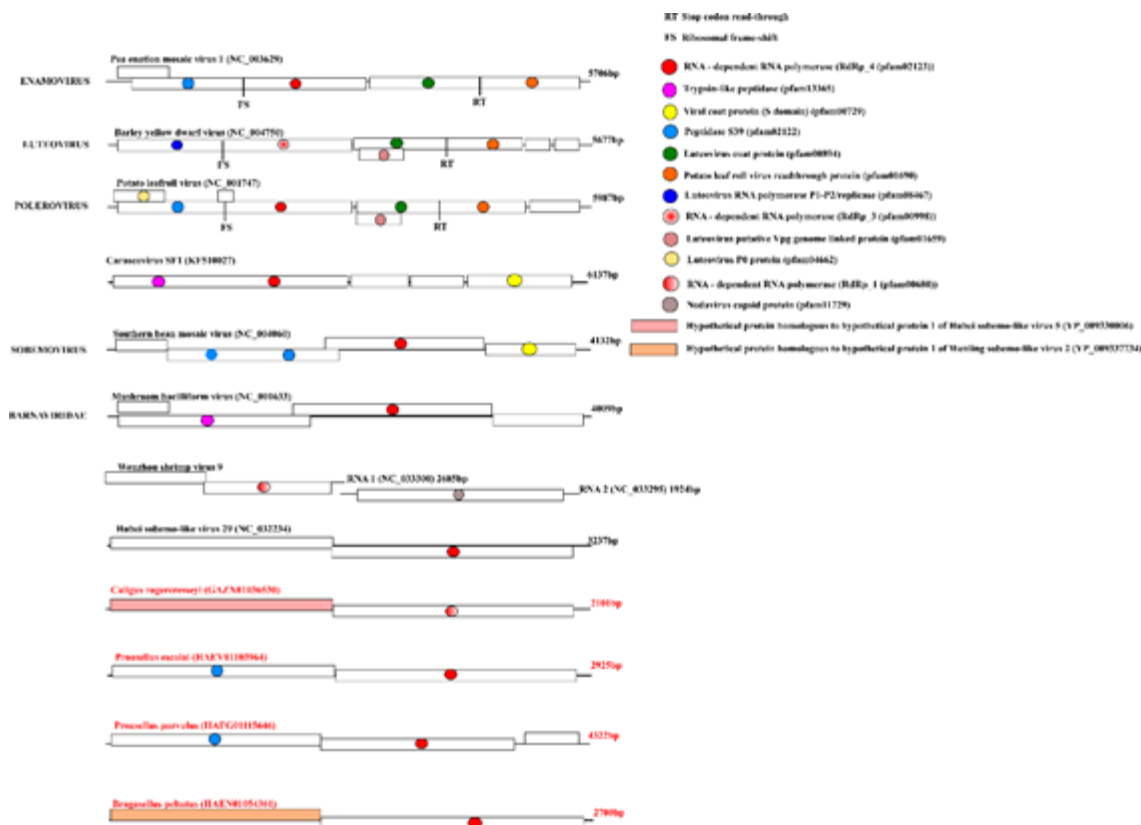
4.4.2.1.4 Klada Hepe-Virga

Pri kladi Hepe-Virga pri rakih smo našli genome 11 virusov, ki spadajo v **klado Hepe-Virga**. V okviru analize genoma klade Hepe-Virga velja izpostaviti naslednje viruse: *Gammarus cheuvreuxi* (GFCV01053886), ki spada med togaviruse, *Pontastacus leptodactylus* (GAFS01001332) in *Omatogammarus flavus* (GEQS01054466), ki spadata med hepevirusom podobne viruse in *Eriocheir sinensis* (GBZW01014116), ki spada med Negevirus. Genomska struktura togavirusa iz *Gammarus cheuvreuxi* (GFCV01053886) najverjetneje ni popolna. Njena dolžina je namreč okrog 7 kb, genom tipičnega togavirusa pa je dolg okrog 11–12 kb. Macro domene in peptidaze C9, ki je značilna za togaviruse, nismo našli. Hepevirus iz *Pontastacus leptodactylus* (GAFS01001332) ima strukturo, ki spominja na strukturo negevirusov (Marsac virusu podobnih virusov). Sestavljajo ga štirje funkcionalno neokarakterizirani ORF in glavni nestrukturni ORF, ki kodira virusno metiltransferazo, helikazo in RdRp. Macro domene in peptidaze C41 nismo našli. Hepevirusom podobne viruse s strukturo, podobno omenjeni strukturi odkritega virusa, so

v genomski analizi prikazali že Shi in sod. (2016a). Glede na podobno strukturo genoma bi pričakovali, da bodo oblikovali sestrsko skupino z negevirusi. Virus iz *Omatogammarus flavus* (GEQS01054466) ima znotraj nestrukturnega ORF-ja s helikazo in RdRp prisotno tudi NADAR (angl. »*E.coli swarming motility protein*«) domeno. Temu ORF sledijo funkcionalno neokarakterizirani ORF. Virus iz *Erichoeir sinesis* (GBZW01014116) ima v nestrukturnem ORF z RdRp kodirano poleg metiltransferaze in helikaze tudi Macro in NADAR domeno. Drugi ORF je neokarakteriziran, tretji pa kodira za negevirus specifičen protein SP24. Celotna struktura virusa torej spominja na strukturo Tanay virusa, vendar ima namesto FtsJ metiltransferazne domene, NADAR in Macro domeni. Macro domena je razširjena pri togavirusih in hepevirusih (Dodatek C; C.4, Slika C.27).

4.4.2.1.5 Klada Luteo-Sobemo

Pri rakih smo odkrili genome šestih virusov, ki spadajo v **klado Luteo-Sobemo**. Zanimivi so štirje primeri celotnih genomov. *Proasellus parvulus* (HAFG01115646) ima dolžino genoma okrog 4 kb. Dolžina je podobna dolžini genoma sobemovirusov in je nekoliko krajša od dolžine genoma luteovirusov (5–6 kb). Verjetno gre za celoten genom. Sestavljen je iz treh ORF. Prvi kodira peptidazo S39, drugi kodira RdRp, tretji pa ni funkcionalno okarakteriziran. Genomi sobemo-like virusov iz *Bragasellus peltatus* (HAEN01054361), *Caligus rogercresseyi* (GAZX01036530) in *Proasellus escolai* (HAEV01103964) na isti RNA kodirajo dva ORF. Prvi ni okarakteriziran, drugi kodira RdRp. Glede na dolžino RNA okrog 2–3 kb v primerjavi z dolžino vsakega od segmentov pri virusih klade Luteo-Sobemo v okviru (Shi in sod., 2016a) – »Wuhan house centipede virus« (WHYY19909); segment 1; 3 kb – lahko sklepamo, da gre pri omenjenih virusih najverjetneje za prvi segment bipartitne RNA (Slika 4.33).



Slika 4.33: Genomska organizacija virusov klade **Luteo-Sobemo**, ki smo jih našli pri rakih (Crustacea).

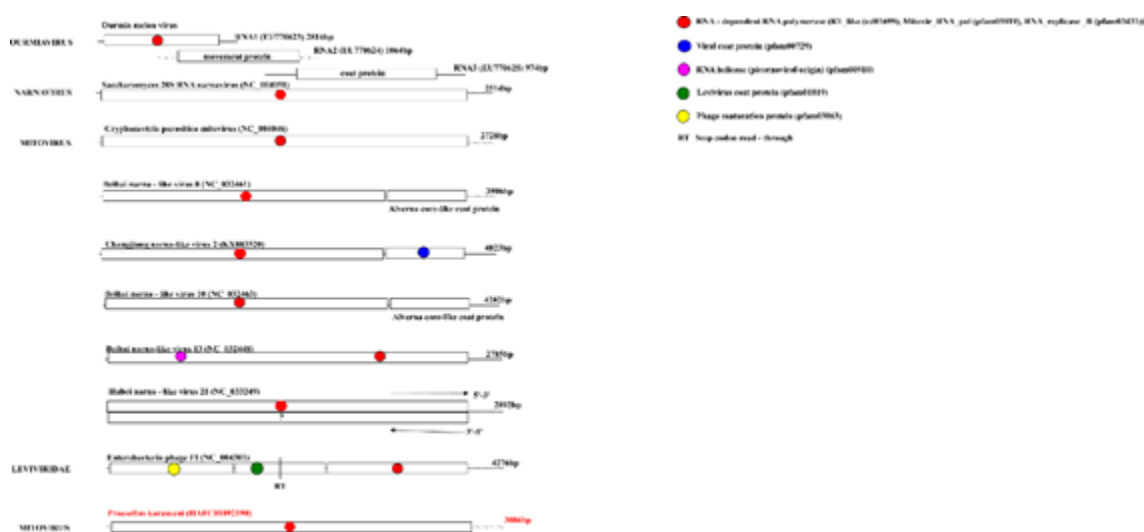
4.4.2.1.6 Klada Mono-Chu

V okviru **klade Mono-Chu** pri rakih smo našli celotne genome mononegavirusov pri *Caligus rogercresseyi* (GAZX01041484), *Gammarus chevreuxi* (HADC01039874), *Proasellus spelaeus* (HAFK01022275), *Proasellus karamani* (HAFK01072349), *Proasellus meridianus*

(HAFE01076147), *Hyalellopsis costata* (GEQG01046040) in *Proasellus cavaticus* (HAES01084808). Njihova genomska organizacija ne odstopa bistveno od tipičnih genomske strukture referenčnih virusov iz klade Mono-Chu, razen v funkcionalno neokarakteriziranih ORF. Virus iz *Hyas araneus* (HACH01032123) spada med chuviruse, vendar nima tipične strukture chuvirusov. Prvi segment kodira glikoprotein (angl. »putative glycoprotein«), drugi ni funkcionalno okarakteriziran, tretji je nukleoprotein in zadnji je L segment. Chuvirusi imajo tipično tri segmente, prvi je L segment, drugi je glikoprotein in tretji je nukleoprotein. Virus iz *Hyas araneus* (HACH01032123) ima strukturo genoma podobno družini Nyamiviridae, vendar tretji ORF namesto glikoproteina kodira nukleoprotein. Glikoprotein je kodiran s prvim ORF (Dodatek C; C.4, Slika C.29).

4.4.2.1.7 Klada Narna-Levi

Genomske strukture celotnih genomov klade Narna-Levi ne odstopajo od trenutno znanih genomske strukture narnavirusov iz literature. Mitovirus iz enakonožca *Proasellus karamani* (HAF01092190) kodira samo nestrukturni protein RdRp (Slika 4.34).



Slika 4.34: Genomska organizacija virusov klade Narna-Levi, ki smo jih našli pri rakih (Crustacea).

4.4.2.1.8 Klada Orthomyxo

Pri kladi Orthomyxo smo poleg Flu_PB1 segmenta pri *Micruropus wahlui* in *Pachyschesis branchialis* našli še Flu_PA segmenta. Pri *Pachyschesis branchialis* smo našli tudi četrti segment, homologen Hubei orthomyxo-like virusu 2. PB1 segment virusa iz *Gammarus chevreuxi* vsebuje še domeno FKBP-tip peptidil-prolil cis-trans izomeraza (angl. »FKBP-type peptidyl-prolyl cis-trans isomerase«) (Dodatek C; C.4, Slika C.28)

4.4.2.1.9 Klada Partiti-Picobirna

Pri kladi Partiti-Picobirna ima partitivirus iz *Proasellus escolai* (HAEV01085407) genom, velik 2,3 kb, ki ga predstavlja monopartitna RNA, ki kodira dva ORF. Prvi kodira RdRp, drugi ni funkcionalno okarakteriziran. Virus iz *Proasellus cavaticus* (HAES01044667) ima genom, podoben referenčnim partitivirusom (Dodatek C; C.4, Slika C.30).

4.4.2.1.10 Klada Permutotetra

V okviru klade Permutotetra smo pri raku *Talitrus saltator* odkrili genom permutotetravirusa. Monopartitna RNA je kodirana z dvema ORF-jema. Prvi kodira RdRp, drugi kodira nodavirusni protein kapside (angl. »nodavirus capsid protein«) (Dodatek C; C.4, Slika C.31).

4.4.2.1.11 Klada Picorna-Calici

Pri kladi **Picorna-Calici** od genomskih struktur celotnih pikornavirusov trije virusi izstopajo glede na prisotnost novih domen, ki niso tipično prisotne pri pikornavirusih. Genom virusa iz *Macrobrachium nipponese* (GCVG01031551) je kodiran z monopartitno RNA, ki vsebuje 1 sam poliprotein. Omenjen ORF vsebuje proteinske domene: helikaza naddružine 3 – peptidaza C3 – RdRp_1 in »RING finger superfamily domain«. Genom virusa iz *Eucyclops serrulatus* (GARW01000621) je kodiran z enim ORF, ki vključuje proteinske domene: dve helikazni domeni, peptidaza C37, RdRp – Bunyavirusni nestrukturni protein NS-S (angl. »bunyavirus nonstructural protein NS-S«), ki je tipičen za bunyaviruse, in »calicivirusni protein plašča« (angl. »calicivirus coat protein«). V tem primeru je lahko prišlo do horizontalnega prenosa gena za bunyavirusni protein NS-S od bunyavirusov na pikornaviruse. Genom virusa iz *Macropereiopus parvus* (GEQM01001122) je ravno tako kodiran z enim ORF, ki vsebuje proteinske domene v nizu helikaza-peptidaza C37-RdRp-domena NADAR (»*E.coli swarming motility protein*«) – Rhv domena. Domena NADAR je prisotna pri skupini Nidovirales. Shi in sod. (2016a) so jo odkrili tudi pri nekaterih pikornavirusih (**Dodatek C; C.4, Slika C.32**).

4.4.2.1.12 Klada Toti-Chryso

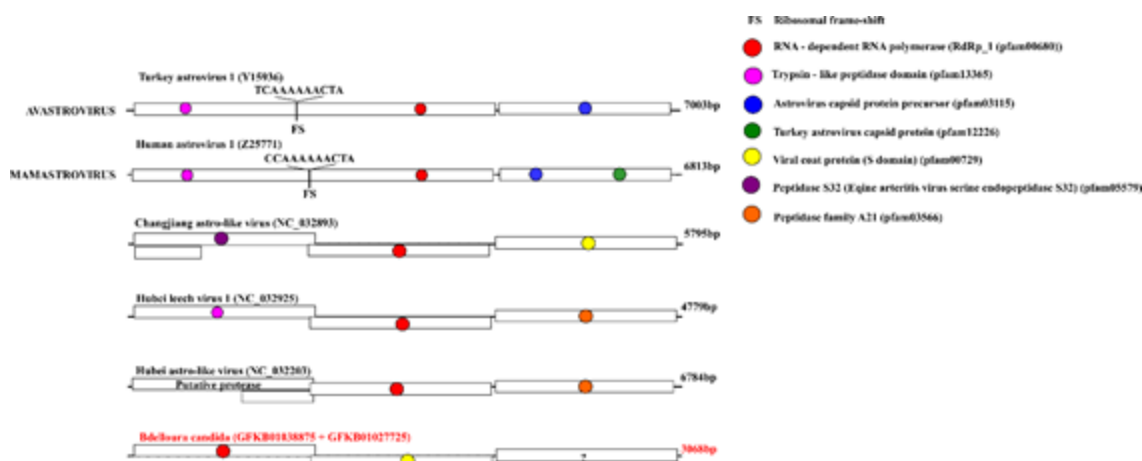
Pri kladi Toti-Chryso so celotni genomi totivirusov, ki smo jih odkrili pri raki *Talitrus saltator*, *Hyas araneus* in *Eurytemora affinis*, kodirani z dvema ORF-jema; prvi ni funkcionalno okarakteriziran, drugi kodira RdRp. Pri virusu *Proasellus grafi* (HAEX01038731) se poleg omenjenih ORF-jev na N-terminalnem koncu pojavljajo dodatni neokarakterizirani ORF-ji (**Dodatek C; C.4, Slika C.33**).

4.4.2.2 Skupina Spiralia: mehkužci (Mollusca), ploščati črvi (Platyhelminthes), kolobarniki (Annelida) in kotačniki (Rotifera)

V okviru skupine Spiralia smo pregledali 225 genomov in transkriptomov. Odkrili smo celotne genome predstavnikov 12 virusnih klad: **Astro, Birna, Bunya-Arena, Flavi, Hepe-Virga, Luteo-Sobemo, Mono-Chu, Narna-Levi, Tombus-Noda, Partiti-Picobirna, Picorna-Calici, Toti-Chryso**.

4.4.2.2.1 Klada Astro

Pri kladi **Astro** smo odkrili en primer celotnega genoma virusa. Struktura celotnega genoma virusa *Bdelloura candida* (GFKB01038875 + GFKB01027725) je sestavljena iz monopartitne RNA, ki kodira 3 odprte bralne okvirje (ORF). Prvi (na N-terminalnem koncu) kodira RdRp, drugi kodira virusni protein plašča (S domena) (angl. »viral coat protein; S domain«), tretji ni funkcionalno okarakteriziran. Referenčni oz. trenutno znani astrovirusi so sestavljeni iz treh ORF. Prvi ORF kodira peptidazo (tripsinu podobna peptidaza ali peptidaza S32), drugi kodira RdRp in tretji kodira virusni protein plašča, ki je lahko rastlinski virusni protein plašča (S domena) (angl. »viral coat protein; S domain«) ali protein plašča, tipičen za astroviruse; »*Turkey astrovirus capsid protein*« ali »*astrovirus capsid protein precursor*«. Astrovirus, ki smo ga odkrili v transkriptomu ploščatega črva *Bdelloura candida*, ima RdRp kodirano v prvem ORF (**Slika 4.35**).



Slika 4.35: Genomska organizacija virusov klade **Astro**, ki smo jih našli pri skupini *Spiralia*. Novi RNA virusi so označeni z rdečo pisavo.

4.4.2.2.2 Klada Birna

Pri **kladi Birna** smo našli delni genom, ki kodira segment B birnavirusa *Procotyla fluvialis* (GAKZ). Drugi segment ni bil prisoten v transkriptomih omenjene vrste (**Dodatek C; C.5, Slika C.34**).

4.4.2.2.3 Klada Bunya-Arena

Tudi v primeru virusov **klade Bunya-Arena** smo pri skupini *Spiralia* našli samo delne genome, ki kodirajo L protein. Virus iz *Trichobilharzia regenti* (GDKR01011304) ima na L segmentu poleg RdRp domene kodiran tudi za Bunyaviruse značilen L protein (**Dodatek C; C.5, Slika C.35**).

4.4.2.2.4 Klada Flavi

Pri **kladi Flavi** smo pri skupini *Spiralia* odkrili celoten genom flavivirusa pri sipi *Idiosepius notoides* (GFNE01070033). Genom virusa je divergenten glede na genome trenutno znanih flavivirusov. Monopartitna RNA, ki kodira en poliprotein, zajema niz proteinskih domen: flavivirusni glikoprotein (angl. »flavivirus glycoprotein«) – protein DinF – helikaza naddružine 1 (DEAH/DEAD box) – RdRp (**Dodatek C; C.5, Slika C.36**).

4.4.2.2.5 Klada Hepe-Virga

Pri **kladi Hepe-Virga** smo našli šest primerov celotnih genomov virusov. Zanimivi so trije virusi: virus iz *Elliptio complananta* (GAHW01000160) (Hepeviridae), virus iz *Fasciola hepatica* (GEVX01001241) (Negevirus) in virus iz *Elysia timida* (GBRM01100701) (Togaviridae). Nestrupturni poliprotein virusa *Elliptio complananta* (GAHW01000160) nima peptidaze C41 in Macro domene, temveč samo virusno metiltransferazo, helikazo in RdRp. Znotraj genomske RNA so vidni dodatni neokarakterizirani ORF. Genom virusa iz *Fasciola hepatica* (GEVX01001241) kodira dva ORF-ja – glavni nestrupturni ORF in dodaten funkcionalno neokarakteriziran ORF. Glavni nestrupturni ORF vsebuje katalitične domene, značilne za negeviruse, v tipičnem vrstnem redu: virusna metiltransferaza (angl. »viral methyltransferase«) – FtsJ metiltransferaza – helikaza – RdRp. Pri negevirusih iz *E. cornigera* (GBRW01169136), *E. cornigera* (GBRW01075309) in *E. timida* (GBRM01098777) v glavnem nestrupturnem poliproteinu ni bilo FtsJ metiltransferaze, pojavi pa se premik bralnega okvirja (angl. »ribosomal frameshift«) in »stop codon read-through«. Prisotni so dodatni neokarakterizirani ORF. Virus iz *Elysia timida* (GBRM01100701) spada med togaviruse, vendar nima tipične genomske strukture. Glavnemu nestrupturnemu ORF-ju sledijo trije funkcionalno neokarakterizirani ORF. Nestrupturni ORF kodira virusno metiltransferazo, helikazo, FtsJ metiltransferazo in RdRp. FtsJ metiltransferaza je tipična za negeviruse in flaviviruse. Pri negevirusih se tipično nahaja na N-terminalnem koncu (oz. pred) helikazo in ne na C-terminalnem koncu helikaze tako kot v našem primeru. V filogenetski analizi smo z dodajanjem virgavirusov in negevirusov želeli dokazati, da omenjen virus ne sodi v drugo virusno skupino (**Dodatek C; C.5, Slika C.37**).

4.4.2.2.6 Klada Luteo-Sobemo

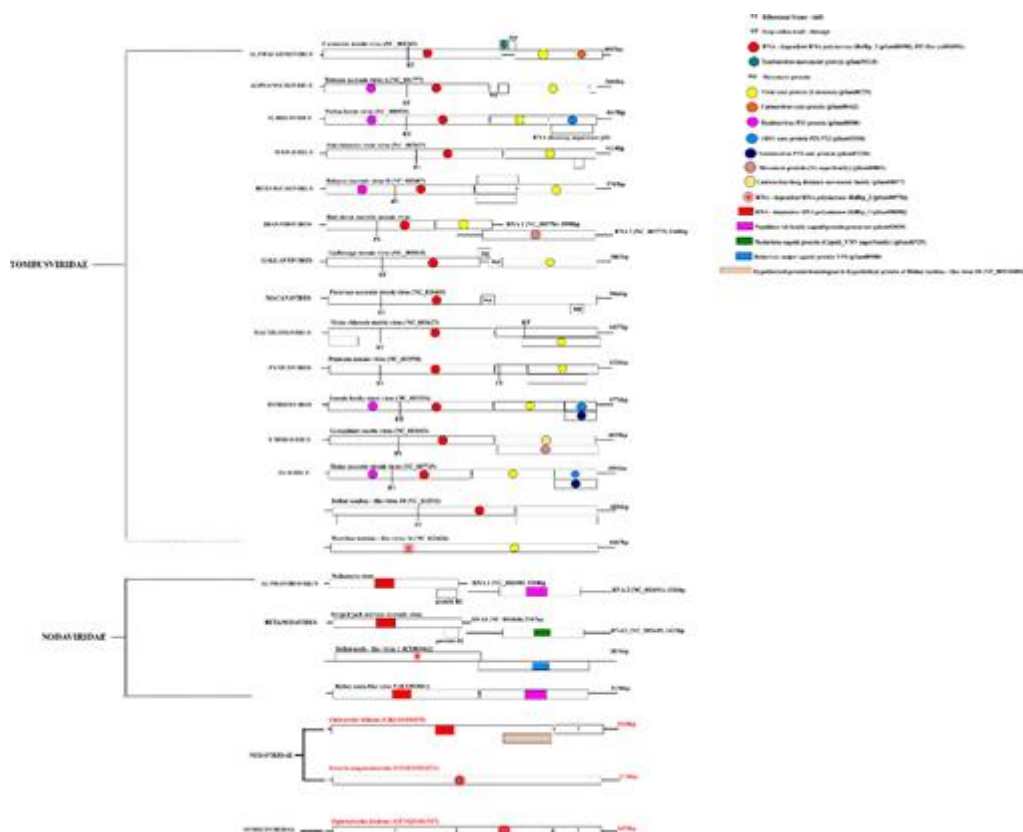
Pri kladi **Luteo-Sobemo** smo pri sipi *Euprymna scolopes* odkrili divergenten luteovirus, ki se je uvrstil med Carascovirusu podobne viruse. Genom virusa kodira dva ORF. Prvi kodira tripsinu podobno peptidazno domeno (angl. »*Trypsin-like peptidase*«), drugi ORF kodira RdRp_1. Genom je dolg okrog 3 kb in je malo krajši od tipičnih luteovirusov (**Dodatek C; C.5, Slika C.38**).

4.4.2.2.7 Klada Narna-Levi

Genomi narnavirusov klade **Narna-Levi** pri skupini Spiralia kodirajo samo en ORF z RdRp, razen v primeru *Alasmidonta varicosa* (HP641214), ki spada med ourmiavirusom podobne viruse. V tem primeru lahko glede na dolžino najdenega zaporedja (2116 bp) sklepamo, da smo našli samo prvi segment virusa, ne pa tudi ostalih dveh za ourmiaviruse značilnih segmentov (**Dodatek C; C.5, Slika C.40**).

4.4.2.2.8 Klada Tombus-Noda

Pri kladi **Tombus-Noda** smo našli tri genomske strukture. Genom nodavirusa iz *Opisthorchis felineus* (GBJA01011878) ne odstopa bistveno od strukture trenutno znanih nodavirusov, razen z vidika neokarakteriziranih ORF. Genom nodavirusa iz kotačnika *Rotaria magnacalcarata* (GDRE01024231) je podobne dolžine kot RNA1 alfa- in betanodavirusa, vendar kodira RdRp družine 3 (RdRp_3). Najverjetneje smo našli samo prvi segment virusa. Tombusvirus iz *Ophryotrocha diadema* (GENQ01001917) je soroden virusu »Wenzhou tombus-like virus 10« (WZFSL57346). Genom kodira tri ORF, od katerih dva nista okarakterizirana, drugi ORF pa kodira RdRp (**Slika 4.36**).



Slika 4.36: Genomska organizacija virusov klade **Tombus-Noda**, ki smo jih našli pri skupini Spiralia.

4.4.2.2.9 Klada Mono-Chu

Pri kladi **Mono-Chu** smo našli dva primera celotnih genomov. Genoma virusov iz *Girardia dorocephala* (GEIG01038096) in iz *Elysia cornigera* (GBRW01045544) ne odstopata od tipične genomske organizacije nyamivirusov in rhabdovirusov (**Dodatek C; C.5, Slika C.39**).

4.4.2.2.10 Klada Partiti-Picobirna

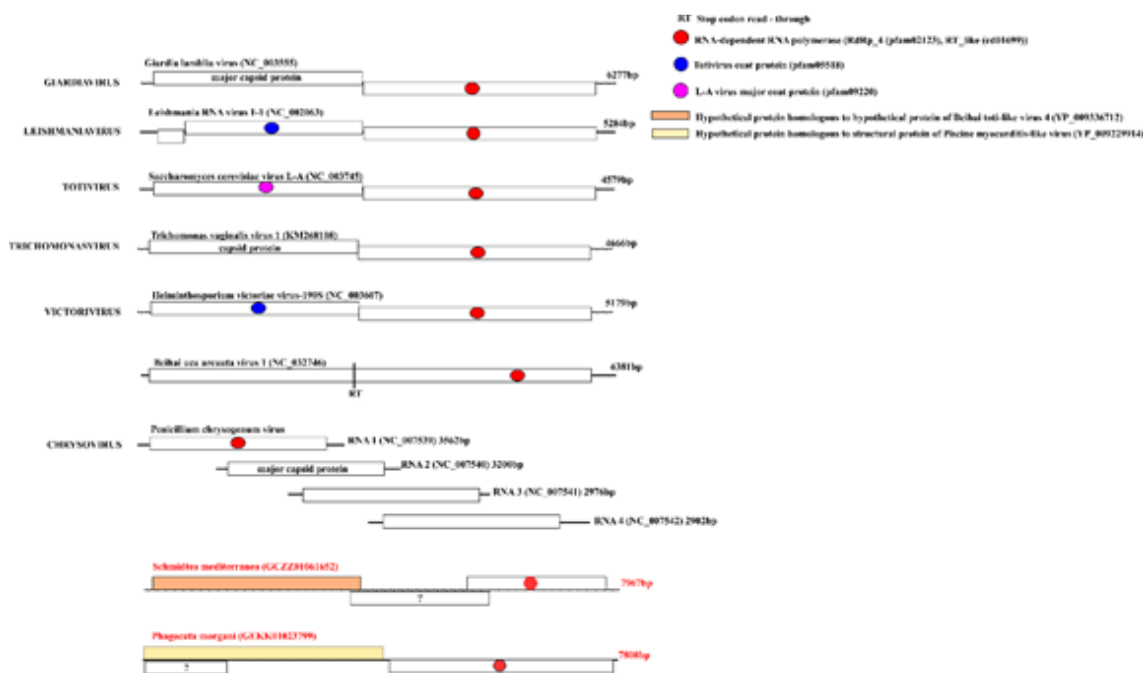
Pri kladi **Parititi-Picobirna** smo našli pet primerov celotnih genomov virusov – tri iz družine Partitiviridae in dva iz družine Picobirnaviridae. Pri partitivirih iz *Elysia timida* (GBRM01095329), *E. timida* (GBRM01028801) in *E. timida* (GBRM01028412) smo odkrili samo prvi segment, ki kodira RdRp. Pikobirnavirusa iz *Elysia cornigera* (GBRW01165272) in (GBRW01101679) imata genom, ki je nesegmentiran in zajema dva ORF – prvi ni okarakteriziran, drugi kodira RdRp (**Dodatek C; C.5, Slika C.42**).

4.4.2.2.11 Klada Picorna-Calici

Pri kladi **Picorna-Calici** se genomska struktura na novo odkritih pikornavirusov ne razlikuje bistveno od trenutno znanih pikornavirusov. Shi in sod. (2016a) so odkrili skupino dicistrovirusov, ki so kodirani samo z enim ORF. ORF kodira poliprotein z nizom domen: helikaza – peptidaza C3 – RdRp_1 – Rhv – Rhv-CRPV. Dva izmed dicistrovirusov iz *E. cornigera* (GBRW01072866) in (GBRW01128951) vsebujeta za Nora viruse značilno HAM1 domeno (cd00515). Celotni genomi nekaterih novih arivirusov, kot so *Phagocata morgani* (GEKK01024656), imajo identično strukturo kot dicistrovirusi odkriti v raziskavi Shi in sod. (2016a): celoten genom je kodiran z enim ORF. Proteinske domene si sledijo v nizu; helikaza – peptidaza C3 – RdRp_1 – Rhv – Rhv – CRPV. Enaka struktura genoma je vidna pri virusu iz *Haliotis discus hanna* (GDAS01047045), ki spada med »aquatic picornavirales«, in virusu *Ruditapes philipinarum* (JO101861), ki spada med posaviruse podobne viruse. Arivirusi imajo lahko tudi strukturo tipičnega dicistrovirusa s helikazo, peptidazo C3 in RdRp na prvem nestrukturiranem ORF in dvema Rhv domenama, vključno s CRPV domeno na strukturalnem ORF. Primer je virus, ki smo ga našli pri *Dugesia japonica* (IAAB01076287) (**Dodatek C; C.5, Slika C.41**).

4.4.2.2.12 Klada Toti-Chryso

Pri kladi **Toti-Chryso** smo našli dva celotna genoma. Genom totivirusa iz *Schmidtea mediterranea* (GCZZ01061652) kodira tri ORF, prva dva nista okarakterizirana, tretji pa kodira RdRp. Genom *Phagocata morgani* (GEKK01023799) kodira dva glavna ORF; prvi ni okarakteriziran, drugi kodira RdRp. Znotraj prvega ORF je v drugem bralnem okvirju kodiran še en neokarakteriziran ORF (**Slika 4.37**).



Slika 4.37: Genomska organizacija virusov klade **Toti-Chryso**, ki smo jih našli pri skupini Spiralia.

4.4.2.3 Najstarejše skupine živali: ožigalkarji (Cnidaria), spužve (Porifera) in rebrače (Ctenophora)

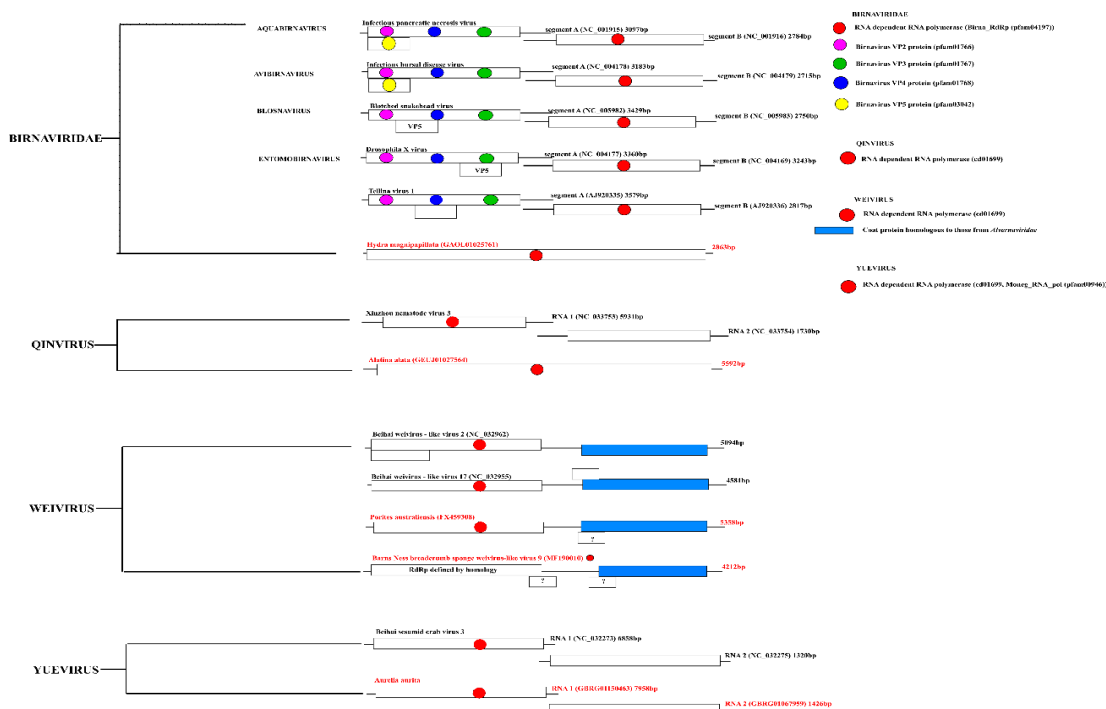
Najstarejše skupine živali zajemajo poleg organizmov Placozoa tudi skupine ožigalkarjev (Cnidaria), spužev (Porifera) in rebrač (Ctenophora). V 88 genomih in transkriptomih smo našli celotne genome predstavnikov 14 virusnih klad: **Birna**, **Qinvirus**, **Weivirus**, **Yuevirus**, **Bunya-Arena**, **Hepe-Virga**, **Luteo-Sobemo**, **Mono-Chu**, **Narna-Levi**, **Tombus-Noda**, **Partiti-Picobirna**, **Picorna**, **Reo**, **Toti-Chryso**.

4.4.2.3.1 Klada Bunya-Arena

Pri kladi **Bunya-Arena** smo odkrili samo L segment arenavirusa pri *Rhopilema esculentum* (GEMS01099264), ki pa je za razliko od ostalih Bunyavirusov kodiran z dvema odprtima bralnima okvirjema (angl. »open reading frame«; ORF). Prvi je funkcionalno neokarakteriziran, drugi kodira RdRp (**Dodatek C; C.6, Slika C.43**).

4.4.2.3.2 Klade Birna, Qinvirus, Weivirus in Yuevirus

Pri kladi **Birna** smo pri ožigalkarju *Hydra magnipapillata* odkrili segment B birnavirusa. Njegovo zaporedje RdRp je le 23–25 % identično kateremu koli trenutno znanemu birnavirusu. Segment B je dolg 2863 bp, kar je tipična dolžina segmenta B birnavirusov. Pri kladi **Qinvirus** ima virus iz *Alatina alata* (GEUJ01027564) tipično strukturo qinvirusov. Pri kladi **Weivirus** so Waldron in sod. (2017) odkrili weivirus v spužvi *Halichondria panicea* – »Barns Ness breadcrumb sponge weivirus-like virus 9«. Pri analizi genomov weivirusov pa smo našli celoten genom weirusa *Porites australiensis* (FX459308). Oba navedena virusa imata tipično strukturo weivirusov. Genomsko RNA sestavljata dva ORF, med katerima je daljša nekodirajoča regija. Prvi ORF (N-terminalni konec) je nestrukturni ORF, ki kodira RdRp. Drugi ORF je strukturni in kodira protein plašča, ki je homologen kapsidi iz genoma družine *Alvernaviridae*. Pri kladi **Yuevirus** je zaporedje virusa iz *Aurelia aurita* (GBRG01150463 + GBRG01067959), tako kot pri ostalih yuevirusih, sestavljeno iz dveh segmentov. Vsak segment predstavlja monocistronska RNA, ki kodira en ORF. Prvi segment kodira RdRp, drugi segment ni funkcionalno okarakteriziran (**Slika 4.38**).



Slika 4.38: Genomska organizacija virusov klad **Birna**, **Qinvirus**, **Weivirus** in **Yuevirus**, ki smo jih našli pri najstarejših skupinah živali (Porifera, Cnidaria, Ctenophora). Novi RNA virusi so označeni z rdečo pisavo.

4.4.2.3.3 Klada Hepe-Virga

V okviru klade Hepe-Virga smo našli genom timovirusa iz korale *Porites asteroides* (GEHP01110756), ki spada v skupino Tymovirales, in dveh hepevirusov, »Barns Ness breadcrumb sponge hepe-like virus 3« in *Craterolophus convolvulus* (HAGZ01026599), ki spadata v družino Hepeviridae. Timovirus iz *Porites asteroides* (GEHP01110756) ima nekoliko drugačno strukturo glede na viruse iz reda Tymovirales. Glavnemu nestrukturnemu ORF, ki kodira proteinske domene – virusno metiltransferazo – helikazo – RdRp_2 – sledijo trije strukturni ORF. Prva dva nista okarakterizirana, tretji kodira virusni protein plašča družine Flexi (angl. »*viral coat protein; Flexi family*«), ki je tipičen za red Tymovirales. Virus »Barns Ness breadcrumb sponge hepe-like virus 3« je sestavljen iz glavnega nestrukturnega ORF s proteinskimi domenami – virusna metiltransferaza – helikaza – RdRp_2 in štirih funkcionalno neokarakteriziranih ORF. Hepevirus iz *Craterolophus convolvulus* (HAGZ01026599) ima strukturo hepevirusom podobnih virusov, kot je npr. Beihai hepe-like virus 4. Glavnemu nestrukturnemu ORF, ki kodira poliprotein, s proteinskimi domenami – virusna metiltransferaza (angl. »*viral methyltransferase*«) – helikaza – RdRp_2. Vendar se na N-terminalnem koncu helikaze nahaja tudi domena NADAR. Proti 3' genomskega zaporedja sledijo trije neokarakterizirani ORF (Dodatek C; C.6, Slika C.44).

4.4.2.3.4 Klada Luteo-Sobemo

V okviru klade Luteo-Sobemo smo našli dva celotna genoma sobemovirusov. Sobemovirusa »Caledonia beadlet anemone sobemo-like virus 1« in *Aurelia aurita* (GBRG01058654) imata monopartitno genomsko RNA, ki kodira 2 ORF. Prvi je pri virusu »Caledonia beadlet anemone sobemo-like virus 1« homologen hipotetičnemu proteinu 1 »Beihai sobemo-like virusa 17«, pri virusu *Aurelia aurita* (GBRG01058654) pa je homologen hipotetičnemu proteinu 1 »Wuhan romanomermis nematode virusa 3«. Drugi ORF pri obeh primerih kodira RdRp_1. Njuna dolžina je nekoliko krajša od povprečne dolžine genoma sobemovirusov (Dodatek C; C.6, Slika C.45).

4.4.2.3.5 Klada Narna-Levi

Pri kladi Narna-Levi smo našli osem celotnih genomov. Narnavirusi iz *Hydractinia symbiolongicarpus* (GCHW01014342+GCHW01012278), *Calvadosia cruximelitisensis* (HAHC01099958), *Lucernaria quadricornis* (HAHD01054073) in *Myxobolus cerebralis* (GBKL01052498) imajo tipično genomsko strukturo narnavirusov. Genomsko RNA kodira en ORF z RdRp proteinsko domeno. Viruse skupine Narnaviridae so odkrili tudi pri spužvi *Halichondria panicea* (Waldron in sod., 2017). Poimenovali so jih »Barns Ness breadcrumb sponge narna-like virus«. »Barns Ness breadcrumb sponge narna-like virus 8« znotraj standardnega ORF z RdRp kodira še štiri prekrivajoče ORF, ki kodirajo, se prepisujejo in prevajajo v obratni smeri (3'–5'). »Barns Ness breadcrumb sponge narna-like virus 2« ima znotraj standardnega ORF z RdRp še dva ORF-ja; enega kodiranega v isti smeri kot glavni ORF z RdRp (5'–3') in enega kodiranega v obratni smeri (3'–5'). Virus »Barns Ness breadcrumb sponge narna-like virus« 3 in 7 sledita tipični strukturi narnavirusov (Dodatek C; C.6, Slika C.46).

4.4.2.3.6 Klada Tombus-Noda

V okviru klade Tombus-Noda smo našli štiri celotne genome, ki spadajo med tombusviruse. Genom virusov »Barns Ness breadcrumb sponge tombus-like virus 1, 2 in 3« kodirata dva glavna ORF. Na 5' je glavni nestrukturni ORF z RdRp in bodisi stop kodonom (»stop codon readthrough«) bodisi premikom bralnega okvirja (angl. »*stop codon readthrough*«) na 5' RdRp proteinske domene. Znotraj glavnega nestrukturnega ORF se pri »Barns Ness breadcrumb sponge tombus-like virusu 1« in »Barns Ness breadcrumb sponge tombus-like virusu 2« nahajata domnevna dodatna ORF. Na 3' koncu nestrukturnega ORF se nahaja dodaten daljši neokarakteriziran ORF. Virus *Craterolophus convolvulus* (HAGZ01003353) kodira glavni nestrukturni protein z RdRp. 5' glede na RdRp je viden premik bralnega okvirja (angl. »*ribosomal frameshift*«), na 3' koncu glavnega ORF sta kodirana dva ORF. Eden kodira virusni protein plašča (domena S) (angl. »*viral coat protein; S domain*«), drugi ni okarakteriziran (Dodatek C; C.6, Slika C.47).

IVSWG – PGVYTVVESY. Ob primerjavi obeh zaporedij vidimo relativno ohranjenost v bloku (Slika 4.40).

```

Hubei      MLK-----FIILIL
HADN01010937.1 CVATETELSPFKEEEVIEKTEVEEILELSQMOSQSEKFAFALALYLLFSNPFCLIF
                :                                     *: **:

Hubei      I-----GRSYSF-----TLLTRIV-----NGEPSTD-----
HADN01010937.1 IAMSCWQLGREWFYEPFLAPLYVKLIERVVYLVCASQLEFDLIQGVLKSTVFDARIREM
                *      *: : :      *: *:      : :  **

Hubei      -YTHPLFYQGTPICTAVA-----
HADN01010937.1 QYRRVFTKGRDIGKIVGMIVGSSLALYGLVRLFKSFMNSPDKAQAFQGRVEAPPQNFV
                * : *: :*. * . *.

Hubei      -----VSPTT-----AIT
HADN01010937.1 GSGDPWARIPHVPAFYLELNPSTYSPQELRTYCQSLVNVKVNAGCMNGIDVGDGYVLT
                :. *: *                      :.*

Hubei      AAHCTH---GYTNQFYISDLRVTRFVTHPSFE-----PLTLANDISLLFASEFQFYVG
HADN01010937.1 MAHVFNNGAVGRIHPLQAGMCTVTSGDHVSFKVDDSRVVLKSRFDLALVYVPEVVHGKG
                ** : * : : . ** :* **:      *: *: :*: :*. . *

Hubei      LYSLPPLNSPVTVVGYGSTCYKCVGTSGVQSAVLEIKICPQRFVAPNTLCAGLDDGS
HADN01010937.1 LFFFIPTTSQCVAQCVLGDA--ELVTPSNIKSAA---PVCASAYYTEPST-----DPSW
                *: **: . *:      :. : * .*: :*.      :. : . *. * .

Hubei      K---DSCQGDSGGPLFYEGKLLAIVSW-GYMCAQPLTPGVYTVVESYQNFIQSQHSSE
HADN01010937.1 KGLFPTEAGDCGSPLIGRFSRWFGIVGLHGTISCTNL----TTVASYCEDI-----
                * : *. *. *: . . : :*. * :.. *      *. * *: *

Hubei      SQSVSSPISQQYNHCLAFHLVILYLYH
HADN01010937.1 SQTEINLLKSKFK----THVKMQTKRI
                **: . :. : : :      ** :

```

Slika 4.40: Primerjava Tripsinu podobnih serinskih proteaz (angl. »Trypsin-like serine protease«) med *Nematostella vectensis* (HADN01010937) in »Hubei virga-like virusom 23« (Katoh in Standley, 2013 – MAFFT).

4.4.2.3.10 Klada Reo

Pri **kladi Reo** smo našli delni genom Reovirusa pri ožigalkarju *Craterolophus convolvulus*. Poleg segmenta z RdRp smo našli še segmenta, ki kodirata proteina VP2 in VP4 (**Dodatek C; C.6, Slika C.50**).

4.4.2.3.11 Klada Toti-Chryso

Genom totivirusa pri *Myxobolus cerebralis* (GBKL01005516) kodira samo en ORF z RdRp in je daljši od trenutno znanih totivirusov. Dolžina zaporedja je 7240 bp. Našli smo celoten genom, ki je za razliko od ostalih totivirusov kodiran z enim ORF. Ostale domnevne proteinske domene na omenjenem virusu niso okarakterizirane (**Dodatek C; C.6, Slika C.49**).

Poglavje 5

Razprava

V okviru raziskav raznolikosti RNA virusov so prednjačile raziskave rastlinskih in vretenčarskih RNA virusov v povezavi z boleznimi, ki jih povzročajo. Raziskave RNA virusov pri nevretenčarjih so bile povezane z njihovo raznolikostjo pri členonožcih kot znanih vektorjih rastlinskih in živalskih virusov. Zato so bili členonožci postavljeni kot center raznolikosti in razširjenosti RNA virusov (Halstead, 1966; Achong, 1972; Bauer in Murphy, 1975; Angly in sod., 2006; Culley in sod., 2006; Desnues in sod., 2008; Hogenhout in sod., 2008; Cantalupo in sod., 2011; Junglen in Drosten, 2013; Mizuno in sod., 2013; Virgin, 2014; Carapeta in sod., 2015; Paez-Espino in sod., 2016; Nouri in sod., 2015; Kawakami in sod., 2016; Ladner in sod., 2016; Rosani in Gerdol., 2017; Galinier in sod., 2017; Filosof in sod., 2017).

Pretekle filogenetske analize ss(+)RNA virusov so glede na skupne karakteristike genomskih struktur in zaporedja RdRp razdelila RNA viruse na tri večje skupine: 1) virusne skupine, podobne flavivirusom (angl. »*flavivirus-like*«), 2) virusne skupine, podobne pikornavirusom, oz. pikornavirusna naddružina (angl. »*picornavirus-like*«), in 3) virusne skupine, podobne alfavirusom (angl. »*alphavirus-like*«) ter 4) ostale posamezne skupine, kot so: Narna-, Levi-, Ourmiavirus, Cystoviridae, Reoviridae, Nidovirales in ss(-) RNA virusi. Skupne karakteristike so bazirale na podlagi zaporedja RdRp in strukture genoma z značilnimi geni oz. domenami (Koonin in Dolja, 2014; Koonin in sod., 2015).

Znanje o RNA virusnih naddružinah so dopolnile raziskave Webster in sod. (2015), Li in sod. (2015), Webster in sod. (2016), Shi in sod. (2016a), Shi in sod. (2016b) in Waldron in sod. (2017). Z rezultati njihovih raziskav se je pojavila nova slika o raznolikosti in razširjenosti RNA virusov. Raznolikost RNA virusov naj ne bi bila omejena samo na členonožce, ampak naj bi bila razširjena čez celotno skupino nevretenčarjev. Webster in sod. (2015), Li in sod. (2015), Webster in sod. (2016), Shi in sod. (2016a), Shi in sod. (2016b) in Waldron in sod. (2017) so naredili obsežne raziskave raznolikosti in razširjenosti RNA virusov pri nevretenčarjih. Center raznolikosti RNA virusov so premaknili s členonožcev na celotno skupino Protostomia. Vendar so se pri skupini žuželk usmerili predvsem na redove Coleoptera, Diptera, Odonata. Pri najstarejših skupinah živali so doslej odkrili viruse predvsem pri ožigalkarjih (Cnidaria), število odkritih virusov pa je bilo majhno. Pri rakihi so se usmerili predvsem na deseteronožce (Decapoda, nekoliko manj na isopoda), pri skupini Lophotrochozoa predvsem na mehkužce (Mollusca). Nevretenčarske skupine kožekrilci (Hymenoptera), metulji (Lepidoptera), raki (Amphipoda in Copepoda), skupina Lophotrochozoa (predvsem ploščati črvi – Platyhelminthes), najstarejše skupine žuželk in najstarejše skupine živali so le slabo raziskane oz. neraziskane s stališča raznolikosti RNA virusov. Namen naše analize je bil:

- 1) **Dokazati, da je raznolikost RNA virusov prisotna pri vseh nevretenčarjih.**
- 2) **Pojasniti razširjenost, raznolikost in nastanek novih RNA virusnih skupin, njihovo evolucijo in genomsko dinamiko pri skupinah nevretenčarjev, ki doslej še niso bile raziskane.**
- 3) **Dokazati, da so najstarejše skupine živali center in izvor raznolikosti RNA virusov.**

Razprava je osredotočena na razširjenost, raznolikost in genomsko dinamiko novih RNA virusov in endogeniziranih virusov pri vsaki od raziskanih skupin nevretenčarjev. Endogenizirani RNA virusi so obravnavani ločeno.

5.1 Kožekrilci (Hymenoptera)

Najobsežnejšo študijo v okviru analize raznolikosti in razširjenosti RNA virusov pri žuželkah so opravili Shi in sod. (2016a). Izmed skupno 87 knjižnic je bilo 21 transkriptomskih knjižnic narejenih na vzorcih RNA žuželk. Vendar Shi in sod. (2016a) v svoji študiji niso analizirali kožekrilcev kot pomembnega reda žuželk. **Kožekrilce (Hymenoptera) so vključili samo v mešani knjižnici WHZM in WHCCII. Obe knjižnici sta revni z vidika raznolikosti in števila odkritih RNA virusov.** Analizirali smo 432 genomskih in transkriptomskih knjižnic kožekrilcev (Slika 4.4). Kožekrilci imajo največjo zbirko genomov in transkriptomov knjižnic izmed vseh skupin organizmov. Odkrili smo 388 novih RNA virusov in s tem bistveno **obogatili sliko o razširjenosti RNA virusov pri žuželkah.** Z analizo RNA virusov pri kožekrilcih smo odkrili prisotnost 15 od 25 definiranih virusnih klad: Qinivirus, Ophio, Flavi, Toti-Chryso, Tombus-Noda, Reo, Picorna-Calici, Permutotetra, Partiti-Picobirna, Orthomyxo, Mono-Chu, Bunya-Arena, Narna-Levi, Luteo-Sobemo in Hepe-Virga. Ugotovili smo, da je klada Picorna-Calici najbolj zastopana. Sledijo ji Partiti-Picobirna, Mono-Chu in Hepe-Virga, ostale virusne klade so redkejšje. Znotraj klade Picorna-Calici smo dokazali prisotnost šestih od osmih virusnih skupin: Iflaviridae, Dicistroviridae, Secoviridae, Kelp fly virusu podobni virusi, Nora virusu podobni virusi in Arivirusu podobni virusi. Najbolj pogosta je bila skupina Iflaviridae (Slika 4.6). Shi in sod. (2016a) so analizirali 21 transkriptomskih knjižnic, ki so pripadale devetim redovom žuželk, in našli 460 virusov, ki so pripadali 17 od 25 virusnim kladam. Mi smo analizirali 20-krat več knjižnic, ki so bile narejene iz vzorcev enega samega reda, in odkrili dokaj podobno število novih RNA virusov (388), ki so pripadali podobnemu številu virusnih klad (15/25). **Naše vzorčenje pa je bilo bistveno bolj reprezentativno.**

Filogenetska analiza novih RNA virusov in analiza raznolikosti genomskih struktur sta ključnega pomena za pridobitev podatkov o sami raznolikosti RNA virusov pri določeni skupini organizmov. **Shi in sod. (2016a) v okviru svoje študije kožekrilcev niso analizirali.** Pri razlagi naših rezultatov analize raznolikosti in razširjenosti RNA virusov pri kožekrilcih pride do izraza pomembna lastnost transkriptomskih knjižnic. Virus je lahko v času vzorčenja prisoten v celici organizma, v medceličnici tkiva organizma ali je naključno prisoten na organizmu kot »artefakt«. Vsak organizem v specifični ekološki niši je udeleženec v mreži ekosistema, torej je z ostalimi organizmi v ekosistemu v določenem odnosu. **Odnos med organizmi v ekosistemu se je pri analizi raznolikosti in razširjenosti RNA virusov izkazal kot pomemben dejavnik mehanizma evolucije njihovih pripadajočih viromov, kar je lepo vidno v okviru analize RNA virusov pri kožekrilcih.** Najpogostejši obliki odnosov med dvema organizmoma sta simbioza in parazitizem. **Kot prvi primer parazitizma bi lahko navedli odnos med parazitoidnimi osami in njihovimi gostitelji, ki je pomembna gonilna sila pri oblikovanju strukture in evolucije njunih pripadajočih viromov.** Odkrili smo namreč številne primere virusov v transkriptomih parazitoidnih os, ki so identični ali sorodni virusom, ki okužujejo njihove gostitelje. Pri kladi Tombus-Noda lahko navedemo primer virusa, najdenega v transkriptomu parazitoidne ose *Podalonia hirsuta*, ki napada gosenice metuljev in je sorodna »Diaphorina citri-associated C virusu«, izoliranemu iz *Diaphorina citri* (Hemiptera, Psyllodea) (Dodatek B; B.1, Slika B.5). Pri kladi Partiti-Picobirna smo v transkriptomu parazitoidne ose *Leptopilina boulardi* našli »Hubei partiti-like virus 27« podoben oz. soroden virus (QTM26091), ki so ga Shi in sod. (2016a) odkrili v knjižnici žuželk skupine Odonata. *Leptopilina boulardi* je vrsta parazitoidnih os, ki napada vinske mušice (*Drosophila melanogaster*) (Slika 4.9). Pri kladi Permutotetra smo našli dva primera, kjer parazitizem med parazitoidnimi osami in pripadajočimi gostitelji igra ključno vlogo v evoluciji klade. Odkrili smo primer »Newfileld virusu - podobnega virusa« v transkriptomu parazitoidne ose *Leptopilina boulardi*. V transkriptomskih knjižnicah organizma *Fopius arisanus* smo našli virus, soroden »Daesongdong virusu 2«, ki je bil izoliran iz komarja *Culex pipiens* (Dodatek B; B.1, Slika B.8). Pri kladi Picorna-Calici smo našli Nora virus, ki je bil izoliran iz vinske mušice, prisoten pri vrsti *Leptopilina heterotoma*, znani parazitoidni osi. Soroden virus smo našli tudi pri vrsti *Fopius arisanus* (Dodatek B; B.1, Slika B.9). Pri vseh navedenih primerih ne moremo sklepati o dejanskem gostitelju virusov. Dejanski gostitelji virusov so lahko parazitoidne ose, ki so virus horizontalno prenesle na gostiteljsko žuželko tekom evolucije virusne klade. Ali pa je bila pot obratna, torej je parazitoidna osa prevzela virus od gostiteljske žuželke. Kot drugi primer bi lahko navedli **parazitizem med rastlinskimi patogenimi glivami in njihovimi rastlinskimi gostitelji.** Pri kladi Partiti-Picobirna (Slika 4.9) smo v transkriptomu vrste *Tenthredo koehlerii* našli genom virusa, ki je soroden

rastlinskemu virusu »Cannabis cryptic virus«. Ti partitivirusi so grupirani z glivnimi virusi, kot je »Fusarium poae virus 1«. Glede na to, da rastlinski z glivnimi virusi tvorijo podskupino, lahko sklepamo, da je tekom evolucije klade Partiti-Picobirna prišlo do horizontalnega prenosa virusa med glivnim in rastlinskim gostiteljem. Smeri prenosa pa ne moremo natančno določiti, saj je drevo koreninjeno na sredini in s tem ni postavljeno v kontekst evlucijskega časa. Lahko je prišlo do prenosa prvotno glivnega virusa na rastlinskega gostitelja in njegovo nadaljnjo diverzifikacijo, pri čemer je virus pridobil dodatne gene za preživetje v rastlinah. Obratno pa je lahko prišlo do prenosa prvotno rastlinskega virusa na glive. Rezultat je premik v njegovi hitrosti mutacij. Parazitski odnos med rastlinskimi patogenimi glivami je v tem primeru za evolucijo klade igral ključno vlogo. Virus, soroden »Fusarium poae virusu 1«, je bil najden v tudi transkriptomu ose *Alysson spinosus*. V tem primeru je lahko *Alysson spinosus* dejanski gostitelj virusa. Verjetnejši scenarij pa je, da igra vmesni člen v parazitskem odnosu med rastlinsko patogeno glivo in rastlino, s katero se *Alysson spinosus* prehranjuje. Virus, odkrit v transkriptomu *Alysson spinosus*, je lahko dejansko povezan ali z rastlino, s katero se prehranjuje, ali z glivo, ki je za omenjeno rastlino patogen. Tretji primer odnosa med organizmi je vrsta »plenilstva«, tipična za kožekrilce, ki podpira njihovo pomembno vlogo kot vektorji rastlinskih virusov. Njihov mehanizem prehranjevanja z rastlinskimi tkivi in produkti (nektar) in vloga kot pomembni opraševalci rastlin je glavni dejavnik horizontalnega prenosa rastlinskih virusov nanje in z njih na druge rastline. Pri analizi raznolikosti RNA virusov pri kožekrilcih smo namreč v transkriptomih kožekrilcev našli nukleotidna zaporedja rastlinskih virusov ali rastlinskim virusom sorodnih virusov. Pri kladi Bunya-Arena (Dodatek B; B.1, Slika B.1) smo v transkriptomu rdeče ognjene mravlje (*Solenopsis invicta*) našli nukleotidno zaporedje rastlinskemu virusu »Ramu stunt virusu podobnega virusa« (izoliranega iz sladkornega trsa). Pri kladi Hepe-Virga (Slika 4.7) smo v transkriptomu ose *Pseudoscolia sinaitica* (Sphecoidea) našli rastlinski virgavirus. Gre za virus, ki je soroden virusu »*Lychnis ringspot virus*«, znanemu patogenu mete. Drugi primer sta virusa, najdena v vrstah *Lithurgus chrysurus* in *Osmia cornuta*, ki sta bila uvrščena med rastlinske viruse skupine Tymovirales. Virus, najden v organizmu *Lithurgus chrysurus*, je soroden rastlinskemu »Cherry virusu A«, virusu izoliranemu iz japonske marelice (*Prunus mume*). Virus, najden v transkriptomu vrste *Osmia cornuta*, je identičen rastlinskemu »Grapewine rupestris stem pitting-associated virus-u«, virusu, ki okužuje vinsko trto. Virus iz transkriptoma vrste *Antophora plumipes* spada med Benyviridae in je soroden rastlinskemu »Beei soil-borne mosaic virusu«, izoliranemu iz navadne pese (*Beta vulgaris*). Pri kladi Ophio (Dodatek B; B.1, Slika B.6) smo v transkriptomu vrste *Panurgus dentipes* našli rastlinski virus, soroden rastlinskemu virusu »Citrus psorosis virusu«, izoliranemu iz citronalne citrone (*Citrus medica*). Pri kladi Picorna-Calici (Dodatek B; B.1, Slika B.9) smo odkrili predstavnika rastlinskih virusov iz družine Secoviridae v transkriptomih vrst *Halictus quadricinctus* in *Crabro peltarius*. Našli smo »Cherry leaf roll virus«, ki je bil izoliran iz navadne breze (*Betula pendula*). V transkriptomu vrste *Chalybion californicum* smo našli »Tobacco ringspot virus«, ki je bil izoliran iz vinske trte kultivarja Blanc du Bois. V transkriptomu ose *Vespa crabro* smo našli virus, soroden »Rice tungro spherical virusu«, izoliranemu iz riža (*Oryza sativa*). Pri vseh navedenih primerih sta možna dva scenarija. Kožekrilec, na katerem je bil najden rastlinski virus, je vektor za omenjen rastlinski virus ali pa je bil virus na njem naključno prisoten v času vzorčenja, ker ga je od rastline prevzel med hranjenjem. Shi in sod. (2016a) so v njihovi študiji pokazali pomen redov Odonata, Blattodea, Hemiptera, Diptera, Coleoptera, Orthoptera in Dermaptera pri horizontalnem prenosu RNA virusov med gostitelji ali med patogenom, kot je patogena gliva, in njegovim gostiteljem. Odkrili so primere rastlinskih Virgavirusov, Bromovirusov, skupine Tymovirales, glivnih virusov skupine Hypoviridae, glivnih virusov skupine Narnaviridae, rastlinskih virusnih pripadnikov skupine Rhabdoviridae, glivnih in rastlinskih virusov klade Partiti-Picobirna in druge primere rastlinskih in glivnih virusov oz. sorodnih virusov, ki so jih odkrili v transkriptomih analiziranih žuželk. V okviru doktorskega dela smo sliko o pomenu žuželk za horizontalni prenos RNA virusov med gostitelji in s tem za evolucijo RNA virusov dopolnili tudi s kožekrilci (Hymenoptera). Pokazali smo, da tudi kožekrilci igrajo pomembno vlogo pri prenosu RNA virusov med elementi biosfere. S tem smo dokazali njihov pomen v pretoku genetskih informacij in evlucijskem poteku sprememb v strukturi genoma RNA virusnih klad.

Za nadaljnja sklepanja o dejanskih genomskih strukturah in poteku izmenjave genomskega materiala znotraj vsake od virusnih klad je pomembna evlucijska starost »gostiteljev«. Skupina kožekrilcev (Hymenoptera) je relativno evlucijsko mlada skupina

žuželk. Velika večina novih RNA virusov, ki smo jih odkrili v genomih in transkriptomih kožekrilcev, se poveže skupaj s trenutno znanimi virusi, ki spadajo v isto virusno klad. Kratka dolžina vej v filogenetskih drevesih vsake od virusnih klad pri kožekrilcih (**Dodatek B.1**) kaže na visoko podobnost novih RNA virusov s trenutno znanimi RNA virusi iz iste virusne klade. Večina novih RNA virusov, ki smo jih našli pri kožekrilcih in so bili uvrščeni v določeno virusno klad glede na zaporedje, se ne loči od trenutno znanih predstavnikov iste virusne klade. Zato na podlagi filogenetske analize novih RNA virusov pri kožekrilcih ne moremo sklepati o dejanski evolucijski starosti vsake od definiranih virusnih klad. So pa ugotovitve o diverzifikaciji RNA virusov pri evolucijsko mlajši skupini kožekrilcev pomembne za nadaljnje sklepanje o faktorjih, ki so vplivali na samo diverzifikacijo. Možna sta namreč dva scenarija. Po prvem scenariju bi lahko naključna mutacija in diverzifikacija virusa izzvala potrebo po menjavi gostitelja. Verjetnejši pa je obratni scenarij. Glede na to, da lahko najdemo isti virus prisoten na dveh različnih vrstah organizmov, je verjetnejša razlaga, da je naključen prehod na drugega gostitelja (npr. z rastline na kožekrilca, s patogene glive na rastlino ali s patogene glive na kožekrilca) izzval spremembo v genetskem driftu, premik v hitrosti mutacij, potrebo po prilagoditvi na drugo okolje in diverzifikacijo. Dejstvo, da smo pri skupini kožekrilcev odkrili mlade RNA viruse, se kaže tudi v strukturi njihovih genomov. Večina genomov novih RNA virusov, najdenih v genomih in transkriptomih kožekrilcev, je podobna genomski strukturi trenutno znanih pripadnikov virusnih klad. Pri kladi **Narna-Levi** smo tako kot Shi in sod. (2016a) odkrili primere narnavirusov z dodatnimi ORF-ji v nasprotnih bralnih okvirjih (**Dodatek C; C.1, Slika C.5**). Večino sprememb predstavljajo nedefinirani ORF, za katere ne obstaja ustrezno karakterizirano homologno zaporedje v proteinskih podatkovnih bazah in za katere ne poznamo biološke funkcije, kot lahko opazimo pri kladi **Mono-Chu** (**Dodatek C; C.1, Slika C.4**). Vseeno pa smo v okviru analize genomskih struktur odkrili primere, ki se nekoliko razlikujejo od trenutno znanih genomskih struktur. Pri kladi **Picorna-Calici** se pojavljajo dodatne domene, kot so RING, podtip domene cinkovih prstov s C3HC4 aminokislinskim motivom in dvema cinkovima kationoma, peptidaza C39 pri virusu iz *Mischocyttarus flavitarsis* (GAXM01039151) in dodatne Rhv domene (**Dodatek C; C.1, Slika C.9**). Genomi nekaterih novih odkritih predstavnikov virusnih klad so pri kožekrilcih nekoliko krajši od tipičnih oz. trenutno znanih predstavnikov iste virusne klade, kot na primer lahko vidimo pri kladah **Tombus-Noda** (**Dodatek C; C.1, Slika C.6**) in **Luteo-Sobemo** (**Dodatek C; C.1, Slika C.3**). Krajši genom lahko odraža evolucijsko mlajše stanje virusov.

V okviru analize raznolikosti RNA virusov pri kožekrilcih smo pokazali, da tudi kožekrilci predstavljajo bogat rezervoar raznolikih RNA virusov in prisotnost širokega spektra virusnih klad. Pokazali smo, da je raznolikost RNA virusov razširjena čez vse redove žuželk. Novo pridobljeni podatki o zastopanosti in raznolikosti novih RNA virusov pri kožekrilcih so pomembni za nadaljnje sestavljanje slike o 1) raznolikosti RNA virusov pri žuželah, 2) razširjenosti RNA virusov pri žuželah in 3) o poteku izmenjave genomskega materiala znotraj vsake od definiranih virusnih klad in med definiranimi virusnimi kladami tekom evolucije.

5.2 Metulji (Lepidoptera)

Shi in sod. (2016a) so v okviru analize raznolikosti RNA virusov pri metuljih na vzorcih metuljev naredili eno samo knjižnico (LCM). Analiza RNA virusov v tej knjižnici je pokazala prisotnost samo enega predstavnika klade **Luteo-Sobemo**, enega predstavnika klade **Bunya-Arena**, enega predstavnika klade **Mono-Chu** in tri predstavnike klade **Reo** (**Dodatek A; A.1 Slika A.1**). Ta knjižnica je bila omejena z vidika številčnosti in raznolikosti prisotnih RNA virusnih klad. V okviru doktorskega dela smo analizirali 301 genomov in transkriptomov metuljev in odkrili 67 novih RNA virusov, ki so spadali v 14 od 25 definiranih RNA virusnih klad: **Flavi**, **Toti-Chryso**, **Tombus-Noda**, **Reo**, **Picorna-Calici**, **Permutotetra**, **Partiti-Picobirna**, **Orthomyxo**, **Mono-Chu**, **Bunya-Arena**, **Narna-Levi**, **Luteo-Sobemo**, **Hepe-Virga** in **Astro**. Klade **Flavi**, **Toti-Chryso**, **Tombus-Noda**, **Reo**, **Picorna-Calici**, **Permutotetra**, **Partiti-Picobirna**, **Orthomyxo**, **Mono-Chu**, **Bunya-Arena**, **Narna-Levi**, **Luteo-Sobemo** in **Hepe-Virga** so prisotne tako pri metuljih kot pri kožekrilcih in kažejo podobne deleže zastopanosti pri obeh skupinah. Klade **Picorna-Calici**, **Partiti-Picobirna** in **Mono-Chu** so najbolj pogoste virusne klade, sledi **Hepe-Virga**. Klade **Bunya-Arena**, **Luteo-Sobemo**, **Orthomyxo** in **Reo** kažejo podoben delež zastopanosti tako pri kožekrilcih kot tudi pri metuljih (3–5 %). Ostale virusne klade so redke tako pri metuljih kot pri kožekrilcih. Predstavnike klade **Astro** smo odkrili samo pri metuljih, predstavnike klad **Ophio** in **Qinvirus** pa samo pri kožekrilcih. Klada

Astro je v okviru žuželk omejena predvsem na metulje, kladi **Ophio** in **Qinvirus** pa na kožekrilce (Sliki 4.4 in 4.5). Pri Shi in sod. (2016a) je bilo tako kot pri kožekrilcih tudi pri metuljih vzorčenje omejeno. Vzorci transkriptomov metuljev so bili predmet samo ene enostavne knjižnice (na nivoju enega reda žuželk) in ene mešane knjižnice. Nobena od knjižnic ni bila bogata z raznolikimi RNA virusi. Skupno so odkrili samo okrog 30 RNA virusov. V okviru doktorata smo sliko o raznolikosti prisotnih virusnih klad pri žuželkah obogatili poleg kožekrilcev tudi z analizo raznolikosti prisotnih klad pri metuljih z bistveno bolj reprezentativnim vzorčenjem. Kljub primerljivemu številu analiziranih knjižnic med kožekrilci (432) in metulji (301) pa smo pri metuljih našli šestkrat manjše število novih RNA virusov. Eden izmed možnih vzrokov je lahko večji prispevek genomskih podatkovnih baz glede na transkriptomske pri metuljih. Lahko pa so metulji manj bogat rezervoar RNA virusov kot kožekrilci.

Zastopanost in številčnost virusnih klad pri metuljih v primerjavi s kožekrilci je pomembna za razumevanje vzorca razširjenosti virusnih klad pri žuželkah. Gre namreč za dva izmed štirih redov žuželk z največ znanimi različnimi vrstami: Coleoptera, Diptera, Hymenoptera in Lepidoptera. Z analizo prisotnosti RNA virusnih klad hkrati dobimo vpogled v spekter raznolikosti prisotnih virusnih klad pri žuželkah. Za celoten vpogled pa potrebujemo še informacije o 1) njihovi filogenetski klasifikaciji in 2) o raznolikosti prisotnih genomskih struktur. Tako kot pri žuželkah tudi pri metuljih pride do izraza značilnost transkriptomskih knjižnic, da ne poznamo vira virusne genske RNA. Virusna genska RNA iz transkriptoma organizma lahko izvira iz virusa, ki je bil naključno prisoten na organizmu, se je nahajal v medceličnici »tkiva« organizma ali je bil dejansko prisoten v celici v času aktivne okužbe organizma. Ne glede na vir RNA virusov pa smo med filogenetsko analizo ugotovili, **da je tako kot pri kožekrilcih tudi pri metuljih odnos med organizmi v ekosistemu ključen pri razširjanju in pretoku njihovih RNA viromov.** Prva oblika odnosa, ki je vplivala na pretok virusov med elementi biosfere, je parazitizem. **Pri tem so bili metulji najverjetneje vmesni člen.** Eden izmed primerov parazitizma, ki je igral pomembno vlogo pri pretoku virusov med elementi biosfere, je **parazitizem med rastlinskimi patogenimi glivami in njihovimi rastlinski gostitelji.** Pomen odnosa za evolucijo RNA virusov se lepo vidi v primerih klad Narna-Levi in Partiti-Picobirna. Obe skupini vsebujeta predstavnike rastlinskih in glivnih virusov. V ospredju je funkcija rastlinske parazitoidne glive in njenega pripadajočega rastlinskega gostitelja. Metulji pri tem lahko igrajo funkcijo stranskega člena kot opraševalci ali kot rastlinojedci. Med prehranjevanjem je lahko metulj 1) naključno pobral glivni virus in ga prenesel na naslednjo rastlino, 2) naključno pobral rastlinsko patogeno glivo, ki jo okužuje glivni virus, in jo prenesel na rastlinskega gostitelja. Navedemo lahko primer virusa pri kladi Narna-Levi (Dodatek B; B.2, Slika B.18) – *Helicoverpa assulta* (ditrisiji, sovke – Noctuidae) (GBTA01004534), ki je soroden glivnemu virusu »Fusarium poae narnavirus 1«, ki je bil izoliran iz askomikotne glive *Fusarium poae*, rastlinske patogene glive. Drugi možni scenarij je, da je metulj pobral rastlinski virus in prešel v stik s pripadajočo rastlinsko patogeno glivo, ki je virus prevzela. Navedemo lahko primere pri kladi Partiti-Picobirna (Slika 4.11). Genom virusa iz *Micropterix calthella* (GACP01000900) je 81 % podoben rastlinskemu partitivirusu »Cannabis cryptic virusu«. Proteinsko zaporedje RdRp »Fusarium poae virusa 1« je 47 % podobno »Cannabis cryptic virusu«. Virus *Micropterix calthella* (GACP01000900) je hkrati 54 % podoben glivnemu virusu *Fusarium poae* virusu 1. Drugi primer je genom virusa *Micropterix calthella* (GACP01012427), ki je 77 % identičen rastlinskemu virusu »Rose partitivirus« in hkrati 54 % identičen glivnemu virusu »Rosellinia necatrix partitivirus 2«. V obeh primerih gre za tipičen primer horizontalnega prenosa virusa med rastlinskim in glivnim gostiteljem, pri čemer je glivni gostitelj hkrati rastlinska patogena gliva. **Pri drugem tipu odnosov, ki je posebna oblika »plenilstva«, pa imajo metulji vlogo glavnega igralca.** Tako kot pri kožekrilcih smo namreč v transkriptomih metuljev našli genome ali dele genomov rastlinskih virusov. Pri kladi Hepe-Virga (Dodatek B; B.2, Slika B.16) je virus iz *Eudarcia simulatricella* (GEOF01029671) najverjetneje rastlinski virus, ki je soroden rastlinskima virusoma »Burdock mottle virusu« in »Beet soil-borne mosaic virusu«. Na organizem je lahko prešel naključno ob hranjenju ali pa deluje kot vektor, ki ob opraševanju rastlin na rastlino prenese virus. Virus skupine **Tymovirales**, najden v transkriptomu *Micropterix calthella* (GACP), je najverjetneje rastlinski virus, soroden rastlinskemu virusu »Cherry virus A« (Dodatek B; B.2, Slika B.24). Shi in sod. (2016a) so s prikazom prisotnosti številnih rastlinskih in glivnih virusom sorodnih virusov v transkriptomih žuželk redov Coleoptera, Diptera in Hemiptera izpostavili njihovo ključno vlogo v prenosu virusov. Redovi Coleoptera, Hemiptera in Diptera so sicer že od prej znani vektorji rastlinskih in živalskih virusov (Roossinck, 2015).

V okviru doktorata smo poleg kožekrilcev poudarili še vlogo metuljev kot opraševalcev, ki igrajo pomembno vlogo kot vektorji rastlinskih virusov in s tem ključno vlogo v evoluciji njihovih genomov.

Za nadaljnja sklepanja o poteku izmenjave genetskega materiala med viromi metuljev in ostalimi organizmi in o poteku evolucijskih sprememb njihovih genomov pa potrebujemo informacije o genomski strukturi RNA virusov metuljev v povezavi z njihovo filogenetsko karakterizacijo. Pri tem se je izkazala informacija o evolucijski starosti organizmov, pri katerih smo našli prisotne RNA viruse, kot zanimiva. Tako kot kožekrilci so tudi metulji evolucijsko mlajša skupina. Shi in sod. (2016a) so pri žuželkah prikazali nepopolno sliko, saj niso analizirali kožekrilcev in metuljev in njihove pripadajoče raznolikosti RNA virusov, ki bi jo povezali z raznolikostjo genomskih struktur in filogenetsko klasifikacijo. **Tako kot pri kožekrilcih se tudi pri metuljih večina novih odkritih RNA virusov ne razlikuje bistveno od drugih trenutno znanih RNA virusov iste virusne klade oz. v filogenetskem drevesu kažejo na pozno oz. nedavno diverzifikacijo.** Večina vej novih RNA virusov, najdenih pri metuljih, je kratkih, kar kaže na bližnje sorodstvene odnose s trenutno znanimi virusi iste virusne klade. **Najdemo pa tudi primere nekaterih nekoliko daljših vej, kot na primer pri kladah Flavi (Dodatek B; B.2, Slika B.15), Hepe-Virga (Dodatek B; B.2, Slika B.24) in Mono-Chu (Slika 4.10). Ugotovitve o raznolikosti genomskih struktur sledijo ugotovitvam na podlagi filogenetske analize.** Večina genomskih struktur novih RNA virusov vsake od virusnih klad se ne razlikuje od genomskih struktur trenutno znanih RNA virusov iste virusne klade. Vseeno pa smo med analizo naleteli na določene primere virusnih genomov, uvrščenih v specifično virusno klado, ki se nekoliko razlikujejo od nabora tipičnih virusnih genomov, ki so značilni za to virusno klado. Ene izmed pogostih razlik so v dolžini genoma. Pri kladi **Hepe-Virga (Slika 4.27)** smo v transkriptomu metuljev odkrili negevirus pri *Yponomeuta evonymellus* (GASG02029232), ki je daljši od povprečnega negevirusa (13040bp). Struktura sicer spominja na strukturo Marsac virusa, vendar glavne razlike tvorijo štirje neokarakterizirani ORF. Nasprotno so nekateri genomi nekoliko krajši od povprečne dolžine, tipične za pripadajočo virusno klado. Pri kladi **Luteo-Sobemo (Slika 4.28)** sta genoma sobemovirusov iz *Enteucha acetosae* (GEOE01112900) in iz *Tischeria quercitella* (GEOU01069137) krajša od povprečne dolžine genoma klade Luteo-Sobemo. Tako kot smo ugotovili pri kožekrilcih, lahko krajši genom odraža evolucijsko mlajše stanje. Pri kladi **Permutotetra (Dodatek C; C.2, Slika C.16)** smo naleteli na dva primera genomov iz *Eogystia hippophaecolus* (GFBH01042206) in iz *Athetis lepigone* (GARC01021157), ki se razlikujeta od referenčnih permutotetravirusov. Tipični permutotetravirusi kodirajo dva ali tri ORF – glavni nestrukturni ORF, ki kodira RdRp, in strukturni ORF, ki kodira protein plašča, ki je lahko naddružine peptidaze A21, protein SP24 ali rastlinski virusni protein plašča (angl. »viral coat protein, S domain«). Pri permutotetravirusih iz *Eogystia hippophaecolus* (GFBH01042206) in iz *Athetis lepigone* (GARC01021157) se strukturni protein (virusni protein plašča) nahaja znotraj glavnega nestrukturnega ORF-ja, ki kodira RdRp. Prepisuje se v drugem bralnem okvirju kot glavni nestrukturni protein. Pri *Athetis lepigone* – (GARC01021157) se znotraj strukturnega ORF nahaja stop kodon. Pri kladi **Picornia-Calici (Dodatek C; C.2, Slika C.18)** se pri iflavirusih iz *Yponomeuta evonymellus* (GASG020225545) in iz *Athetis lepigone* (GARD01006938) nahaja dodatna Rhv domena. Z analizo genomskih struktur v povezavi s filogenetsko analizo smo ugotovili, da se večina genomskih struktur ne razlikuje od trenutno znanih genomskih struktur virusnih klad. Tudi v filogenetski analizi kratka dolžina vej večine novih RNA virusov nakazuje na nedavno diverzifikacijo in podobnost z drugimi trenutno znanimi predstavniki virusnih klad. Vseeno pa smo pri metuljih našli več divergentnejših predstavnikov RNA virusov kot pri kožekrilcih. Tako smo obogatili sliko o raznolikosti genomskih struktur RNA virusov glede na trenutno sliko, v kateri so manjkale informacije o kožekrilcih in metuljih (Shi in sod., 2016a).

Z analizo raznolikosti RNA virusov pri metuljih smo ugotovili, da so bogat vir raznolikih RNA virusov, vendar smo jih odkrili kljub analizi primerljivega števila genomskih in transkriptomskih knjižnic nekoliko manj kot pri kožekrilcih. Ugotovitve o raznolikosti RNA virusov pri kožekrilcih in metuljih so pomembne za celostno sliko o raznolikosti genomskih struktur RNA virusov pri žuželkah v povezavi z njihovo filogenetsko karakterizacijo.

5.3 Najstarejše skupine žuželk: skakači (Collembola), dvorepke (Diplura), ščetinorepke (Zygentoma), Monocondylia

Za celosten vpogled v 1) zastopanost virusnih klad, 2) profil raznolikosti genomskih struktur in 3) filogenetsko sliko posameznih virusnih klad potrebujemo še informacije o profilu raznolikosti RNA virusov pri najstarejših skupinah žuželk oz. pri pražuželkah (angl. »*basal Hexapoda*«), kot so skakači (Collembola), dvorepke (Diplura), ščetinorepke (Zygentoma) in Monocondylia. Najstarejše skupine žuželk predstavljajo starejšo kopensko vejo skupine Pancrustacea (Budd in Telford, 2009). Kot pomembni dekompozitorji, ki naseljujejo matriks prsti, predstavljajo potencialno zalogo raznolikih RNA virusov. **Raznolikost RNA virusov pri najstarejših skupinah žuželk (Shi in sod., 2016a) doslej še ni bila raziskana.** V okviru doktorata smo analizirali 16 transkriptomov in štiri genome najstarejših skupin žuželk. Največ knjižnic (tri genomske in devet transkriptomskih) je bilo narejenih na vzorcih skakačev. V primerjavi s kožekrilci in metulji knjižnice najstarejših skupin žuželk trenutno še vedno niso obsežne. Našli smo posamezne primere novih RNA virusov pri kladah: **Mono-Chu** (en predstavnik) (Dodatek B; B.3, Slika B.25), **Narna-Levi** (en predstavnik) (Dodatek B; B.3, Slika B.26), **Hepe-Virga** (pet predstavnikov) (Dodatek B; B.3, Slika B.28), **Orthomyxo** (en predstavnik) (Slika 4.13), **Picorna-Calici** (sedem predstavnikov) (Slika 4.14), **Qinvirus** (en predstavnik) (Dodatek B; B.3, Slika B.27), **Tombus-Noda** (dva predstavnika) (Dodatek B; B.3, Slika B.29), **Toti-Chryso** (en predstavnik) (Slika 4.15). Skupno smo odkrili 19 RNA virusov pri najstarejših skupinah živali, ki so spadali v osem od 25 RNA virusnih klad. **Torej smo kljub omejenemu številu genomov in transkriptomov pokazali bogato raznolikost RNA virusnih klad tudi pri najstarejših skupinah žuželk.** Relativno majhno število dostopnih genomov in transkriptomov pa je najverjetnejši vzrok, da smo našli **manjšo pestrost RNA virusnih klad pri najstarejših žuželkah.** Evolucijsko mlajše skupine žuželk so zastopane s 15 do 20-krat večjim naborom genomskih in transkriptomskih knjižnic. Pri najstarejših skupinah žuželk sta bili najbolj zastopani **klada Hepe-Virga** (Dodatek B; B.3, Slika B.28), pri kateri smo odkrili izključno negevirusne, in **klada Picorna-Calici**. Znotraj klade Picorna-Calici (Slika 4.14) smo odkrili predstavnike Nora virusu podobnih virusov, Kelp fly virusu podobnih virusov, iflaviruse in dicistroviruse. Izmed sedmih pikornavirusov trije spadajo med iflaviruse. **Iflavirusi so torej tako kot pri kožekrilcih in metuljih tudi pri najstarejših skupinah žuželk najbolj zastopana skupina. Z analizo genomov in transkriptomov najstarejših skupin žuželk smo obogatili in razširili sliko o raznolikosti RNA virusov pri žuželkah in posledično tudi pri nevretenčarjih.** Rezultati analize RNA virusov pri najstarejših skupinah žuželk se ujemajo tako z vidika filogenetske analize kot z vidika analize genomskih struktur.

Vsi odkriti predstavniki RNA virusov pri kladah Mono-Chu, Narna-Levi, Hepe-Virga, Orthomyxo, Picorna-Calici, Qinvirus, Toti-Chryso in Tombus-Noda so divergentni glede na trenutno znane predstavnike posamezne virusne klade. Pri **kladi Mono-Chu** (Dodatek B; B.3, Slika B.25) je endogeniziran virus, najden v genomu *Catajapyx aquilonaris*, samo 30 % identičen najbližjemu sorodniku »Shayang fly virusu 1«. Pri kladi Narna-Levi je virus iz *Tetrodontophora bielanensis* (GAXI) 40 % soroden »Beihai narna-like virusu 23« (BHZY57139). Pri **kladi Hepe-Virga** (Dodatek B; B.3, Slika B.28) smo odkrili samo predstavnike negevirusov, ki so manj kot 50 % identični kateremu koli trenutno znanemu negevirusu. Pri **kladi Orthomyxo** je virus, ki smo ga našli pri mali mravlji ribici (*Atelura formicaria*), 31 % identičen »Sanxia water strider virusu 3« (Slika 4.13). Pri **kladi Picorna-Calici** (Slika 4.14) smo odkrili primere virusov, ki so manj kot 40 % identični najbližjim sorodnikom. Pri **kladi Qinvirus** (Dodatek B; B.3, Slika B.27) je virus iz *Anurida maritima* (GAUE02014037) 31 % podoben proteinskemu zaporedju RdRp »Hubei qinvirus-like virusa 1« (SCM49583). Pri **kladi Tombus-Noda** (Dodatek B; B.3, Slika B.29) je virus *Tricholepidion gertschi* (GASO02037726) 24–27 % identičen kateremu koli trenutno znanemu tombusvirusu. Pri **kladi Toti-Chryso** (Slika 4.15) je virus iz *Tetrodontophora bielanensis* (GAXI02022882) 39 % identičen »Diatom colony-associated dsRNA virusu 10«.

Z vidika strukture genomov smo pri najstarejši skupini žuželk odkrili primere genomov virusov, ki odstopajo od trenutno znanih virusov iste klade. Pri kladi Hepe-Virga (Dodatek C; C.3, Slika C.21) genomska struktura virusov iz *Anurida maritima* (GAUE02021637) in iz *Sminthurus viridis* (GATZ02022882) nekoliko odstopa od genomske strukture referenčnih negevirusov. Glavni nestrukturni protein sicer kodira virusno metiltransferazo, helikazo naddružine I in RdRp_2. Vendar ni prisotne FtsJ metiltransferaze. Pri negevirusu *Anurida*

maritima (GAUE02021637) glavnemu nestrukturnemu poliproteinu sledi samo en strukturni ORF, ki kodira protein SP24. Pri **kladi Picorna-Calici (Dodatek C; C.3, C.24)** genomska struktura pikornavirusa iz *Anurida maritima* (GAUE02021853) nekoliko odstopa od strukture trenutno znanih Kelp fly virusu podobnih virusov. Monopartitna RNA kodira dve helikazni domeni, ne kodira pa peptidazne domene. Pri pikornavirusu iz *Occasjapyyx japonicus* (GAXJ02019692) smo odkrili dodatno Rhv domeno. **Rezultati filogenetske analize RNA virusov pri najstarejših skupinah žuželk skupaj z analizo njihovih genomov predstavljajo pomembno podlago za nadaljnja sklepanja o povezavi med evolucijsko starostjo skupine žuželk in naborom najdenih divergentnih RNA virusov.**

Filogenetska drevesa RNA virusov pri pražuželkah (angl. »*basal Hexapoda*«) smo koreninili na sredini. Zato ne moremo dolžine posameznih vej povezati z evolucijsko starostjo posameznega virusa oz. skupine virusov. Sklepamo lahko samo o sorodstvenih odnosih med virusi v skupini in med posameznimi skupinami znotraj virusne klade. **Zanimivo je, da smo pri najstarejših skupinah žuželk našli najbolj divergentne predstavnike virusov v okviru žuželk (Hexapoda).** Bili so le 20–40 % identični kateremu koli predstavniku posamezne virusne klade. **Evolucijska starost skupine žuželk je povezana z divergentnostjo najdenih virusov v njenih genomskih in transkriptomskih podatkovnih bazah.** Starejša kot je določena taksonomska skupina, manj podobno bo zaporedje RdRp virusov, ki so prisotni v njihovih genomskih in transkriptomskih knjižnicah z RdRp katerega koli definirane virusa, ki spada v isto virusno kladu. Virusi so namreč z žuželkami povezani v mrežo koevolucije gostitelj-virus.

V okviru filogenetske analize RNA virusov pri najstarejših skupinah žuželk smo **odkrili pester nabor novih RNA virusov. S tem smo razširili analizo RNA virusov na celotne žuželke.** Odkriti RNA virusi so bili precej divergentni glede na trenutno znane RNA viruse iz iste virusne klade. **S tem smo pokazali, da so vse žuželke bogat vir raznolikih RNA virusov.** Odkrili smo, da je pri žuželkah skupaj prisotnih 19 od 25 virusnih klad. Pri žuželkah so odsotne virusne klade Weivirus, Yanvirus, Yuevirus, Zhaovirus, Potyviridae in Cystoviridae. Klade Hepe-Virga, Narna-Levi, Mono-Chu, Orthomyxo, Picorna-Calici, Tombus-Noda in Toti-Chryso smo našli prisotne pri vseh skupinah žuželk. Prvotno so bile z žuželkami povezane samo klade: Birna, Reo, Nido, Picorna-Calici, Hepe-Virga, Permutotetra, Bunya-Arena in Mono-Chu. Če združimo naše analize s predhodnimi analizami, je klada Birna do sedaj omejena na žuželke reda Diptera (najdena je bila tudi pri glistah in pri skupini Lophotrochozoa). Skupina Nido je znotraj žuželk ravno tako omejena na red Diptera, našli pa so jo prisotno tudi pri rakah, glistah in skupini Lophotrochozoa (Shi in sod., 2016a). Ostale virusne klade so širše razširjene pri žuželkah, pri čemer je najbolj zastopana klada Picorna-Calici, znotraj klade Picorna-Calici pa izrazito izstopajo iflavirusi (Hulo in sod., 2011 – ViralZone; Shi in sod., 2016a). **Shi in sod. (2016a) so odkrili prisotnost 17 od 25 virusnih klad pri žuželkah. V naši raziskavi smo odkrili, da sta pri žuželkah prisotni tudi kladi Astro in Ophio. S tem smo povečali nabor ugotovljenih prisotnih virusnih klad pri žuželkah s 17 na 19.** Pogoste klade pri žuželkah so: Picorna-Calici, Partiti-Picobirna, Mono-Chu, Partiti-Picobirna, Tombus-Noda, Toti-Chryso, Hepe-Virga. Nekoliko redkejša so: Reo, Narna-Levi, Luteo-Sobemo, Permutotetra, Bunya-Arena, Orthomyxo in Flavi. Redke klade so: Astro, Birna, Qinivirus, Ophio, Hypo in Nido. Klade, ki so odsotne pri žuželkah, so: Yanvirus, Zhaovirus, Potyviridae, Weivirus, Yuevirus in Cysto.

5.4 Raki (Crustacea)

Rake sestavljajo tri glavne skupine: **Maxillopoda** (Copepoda – ceponožci, Branchiura – škrgorepci in ostale skupine), **Malacostraca** (Amphipoda – postrance, Decapoda – deseteronožci, Isopoda – enakonožci in ostale skupine) in **Branchiopoda** (Budd in Telford, 2009; von Reumont in sod., 2012). Shi in sod. (2016a) so izvedli najobširnejšo analizo raznolikosti RNA virusov pri rakah. Iz RNA vzorcev rakov so pripravili 20 knjižnic, ki so pripadale trem glavnim skupinam rakov. Na vzorcih skupine Branchiopoda in Maxillopoda je bila narejena ena sama knjižnica. Na vzorcih skupine Malacostraca je bilo narejenih 16 knjižnic. Od tega je bilo 11 knjižnic narejenih na vzorcih deseteronožcev in štiri pri skupini enakonožcev. Pri vseh skupinah rakov so odkrili bogato pestrost prisotnih virusnih klad (**Dodatek A; A.1 Slika A.2**). **Odkrili so več kot 420 novih RNA virusov, ki so pripadali 19 od 25 virusnih klad.** Vsaka od glavnih skupin rakov je vsebovala RNA virom z unikatnim profilom prisotnih virusnih klad. Pri skupinah Branchiopoda in Malacostraca je virusna klada Picorna-Calici najbolj zastopana virusna klada. Pri skupini Maxillopoda imajo klade Picorna-

Calici, Mono-Chu, Hepe-Virga, Partiti-Picobirna, Narna-Levi in Bunya-Arena približno enako zastopanost, medtem ko so klade Toti-Chryso, Tombus-Noda in Luteo-Sobemo bolj redko zastopane. Pri skupini Malacostraca, ki je bila bolj reprezentativno vzorčena, po zastopanosti izrazito izstopajoči kladi Picorna-Calici sledita kladi Narna-Levi in Tombus-Noda, ostale klade pa so redkejše. Skupina Branchiopoda je pokazala prisotnost samo šestih virusnih klad, med katerimi prednjačijo Picorna-Calici, Narna-Levi in Bunya-Arena. **Vseeno Shi in sod. (2016a in 2016b) niso prikazali popolne slike raznolikosti RNA virusov pri rakihi.** Nobena knjižnica namreč ni bila narejena pri skupinah ceponožcev (Copepoda) in postranic (Amphipoda). Skupina ceponožcev je evolucijsko starejša in raznolika skupina rakov z vidika življenjskih ciklov in preferenc habitata. Dosežejo dolžino od 0,1 mm do 23 cm, poznamo pa okrog 120000 različnih vrst. So dominantni predstavniki zooplanktona in predstavljajo največji delež živalske biomase na Zemlji. Vključeni so v mrežo prehranjevalnih verig, pri kateri plenijo mikroorganizme in so glavni vir hrane višjih trofičnih nivojev (Bron in sod., 2011). Zato predstavljajo pomembno bogato zalogo raznolikih RNA virusov. **V okviru doktorata smo sliko o raznolikosti RNA virusov pri rakihi dopolnili še z odkritji raznolikih RNA virusov pri postranicah in ceponožcih (Slika 4.2).** Pregledali smo 187 transkriptomov in 23 genomov štirih skupin rakov: deseteronožcev, ceponožcev, enakonožcev in postranic. Od tega je kar 10 genomov in 22 transkriptomov pripadalo ceponožcem (Copepoda), dva genoma in 78 transkriptomov pa postranicam (Amphipoda). **Pri tem smo odkrili 145 novih divergentnih RNA virusov, ki so spadali v skupno 17 od 25 definiranih virusnih klad.** Pri vsaki od analiziranih skupin smo tako kot Shi in sod. (2016a) (Dodatek A; A.1, Slika A.2) zasledili unikaten profil zastopanosti RNA virusov. Pri vseh skupinah je najbolj bogato prisotna klada Picorna-Calici. Profil zastopanosti ostalih virusnih klad je unikaten za vsako od analiziranih skupin rakov. Pri ceponožcih po zastopanosti skupini Picorna-Calici sledi Mono-Chu, Toti-Chryso in Partiti-Picobirna. Pri postranicah kladi Picorna-Calici sledi Mono-Chu, Hepe-Virga in Orthomyxo. Pri enakonožcih sledi Mono-Chu, pri deseteronožcih pa sledijo Flavi, Tombus-Noda in Hepe-Virga. Ostale skupine so prisotne v manjšini (Slika 4.2). Kljub relativno omejenemu številu transkriptomov in genomov pri ceponožcih (32) smo odkrili dokaj pester nabor RNA virusnih klad (10/25). Pri postranicah, ki so zastopane z nekoliko večjim številom baz (80), pa smo odkrili osem od 25 RNA virusnih klad, kar pomeni nekoliko revnejši spekter virusnih genomov pri postranicah. Z analizo genomov in transkriptomov rakov smo ugotovili, da sta pri rakihi prisotni tudi virusni kladi Birna (Dodatek B; B.4, Slika B.32) in Ophio (Dodatek B; B.4, Slika B.37), ki doslej še nista bili odkriti pri rakihi (Shi in sod., 2016a, Shi in sod., 2016b). **S tem smo razširili nabor prisotnih virusnih klad z 19 na 21 od 25 definiranih.** Pri rakihi so pogoste klade: Picorna-Calici, Toti-Chryso, Mono-Chu, Narna-Levi, Hepe-Virga, Partiti-Picobirna in Tombus-Noda. Klade Astro, Birna, Luteo-Sobemo, Bunya-Arena, Orthomyxo, Nido, Permutotetra, Reo, Toti-Chryso, Qinvirus, Zhaovirus, Yuevirus in Yanvirus so pri rakihi redko zastopane. **Samo klade Hypo, Weivirus, Cysto in Potyviridae niso bile odkrite pri rakihi.** Znotraj najbolj zastopane skupine pri rakihi, klade Picorna-Calici, so Shi in sod. (2016a) odkrili šest od skupno sedmih skupin: vodni pikornavirusi, dicistrovirusi, Posavirusu podobni virusi, Arivirusu podobni virusi in Nora virusu podobni virusi (Dodatek A; A.2; Slika A.4). Izpostavili so bogato prisotnost vodnih pikornavirusov (angl. »aquatic picornavirales«) in dicistrovirusov kot prevladujočih skupin. V okviru doktorata smo našli šest pikornavirusnih skupin: Nora virusu podobni virusi, Kelp fly virusu podobni virusi, vodni pikornavirusi, dicistrovirusi in iflavirusi (Slika 4.6). Nismo pa odkrili RNA virusov, ki bi bili uvrščeni med Posavirusu podobne viruse in Arivirusu podobne viruse, vendar smo odkrili, da je tudi skupina Kelp fly virusu podobnih virusov prisotna pri rakihi. **Presenetljivo smo odkrili, da je poleg vodnih pikornavirusov in dicistrovirusov pri rakihi bogato zastopana tudi skupina iflavirusov.** Prvotno so bili iflavirusi povezani predvsem z žuželkami. Prikazali smo celostno sliko o prisotnosti RNA virusnih klad pri rakihi, ki je bila doslej nepopolna zaradi omejenega vzorčenja pri filogenomskih analizah (Shi in sod., 2016a).

V okviru analize prisotnosti virusnih klad smo ugotovili prisotnost skoraj vseh virusnih klad, razen štirih. **Za nadaljnjo celostno sliko o prisotnih raznolikih RNA virusih pri rakihi smo potrebovali vpogled v njihovo filogenetsko karakterizacijo in njihove genomske strukture.**

V okviru filogenetske analize smo našli številne RNA viruse, ki so glede na proteinsko zaporedje RdRp divergentni glede na kateri koli trenutno znan RNA virus, ki spada v isto virusno klado. Proteinska zaporedja njihovih RdRp so od 20–40 % identična kateremu koli trenutno znanemu RNA virusu iste virusne klade. Pri kladi Bunya-Arena (Dodatek B; B.4, Slika B.31) smo odkrili virus iz *Hyalalopsis stebbingi* (GEQJ01053822), katerega

proteinsko zaporedje je 34 % identično RdRp najbližjega sorodnika – »Gouleaku virusa«. Pri **kladi Birna (Dodatek B; B.4, Slika B.32)** je virus iz *Eurytemora affinis* (GBG001030594) glede na proteinsko zaporedje RdRp 46 % identičen proteinskemu zaporedju »Drosophila X virusa«. Pri **kladi Flavi (Slika 4.16)** je zaporedje predstavnika *Eurytemora affinis* (AZAI01066421) samo 29 % identično z virusom rumene mrzlice (skupina Flavivirus) oz. 30 % identično s proteinskim zaporedjem RdRp katerega koli trenutno znanega flavirusa. Flavi NS5 polimeraza virusa iz *Caridina multidentata* (IABX01135281) je samo 26–28 % identična kateremu koli trenutno znanemu flavivirusu oz. 26 % identična jingmenvirusu. Pri **kladi Luteo-Sobemo (Dodatek B; B.4, Slika B.34)** je virus iz *Proasellus parvulus* (HAFG01087285) 45 % identičen najbližjemu sorodniku »Hubei sobemo-like virus 45 (QTM25680)«. Virus, odkrit v transkriptomu raka ceponožca *Eurytemora affinis* (GBG001020472), je 42 % identičen virusu »Sanxia sobemo-like virus 3« (SXSSP2531). Pri **kladi Mono-Chu (Slika 4.17)** je proteinsko zaporedje RdRp virusa iz *Hyaella azteca* (JQDR02010914) 46 % identično zaporedju »Wenzhou crab virusa 1« (WZRBX43457). Virus iz *Ommatogammarus flavus* (GEQS01004404) je 48 % identičen »Pteromalus puparum negative-stranded RNA virusu 1« in 49 % identičen »Hubei rhabdo-like virusu 6« (QTM24798). Pri **kladi Narna-Levi (Dodatek B; B.4, Slika B.35)** je virus iz *Proasellus karamani* (HAFG01092190) 39 % identičen »Phytophthora infestans RNA virusu 4«. Pri **kladi Reo (Dodatek B; B.4, Slika B.42)** smo odkrili divergentni zaporedji v transkriptomih *Echinogammarus veneris* (GAR0) in *Proasellus racovitzai* (HAFH), ki sta 25 % identični s Cimodo virusom. Pri **kladi Partiti-Picobirna (Dodatek B; B.4, Slika B.39)** se virusa iz *Tigriopus californicus* (JW510638) in *Tigriopus japonicus* (GCHA01047740) povežeta skupaj s skupino rastlinskih in glivnih virusov. 34 % identična sta z »Raphanus sativus cryptic virusom 1«. **Hkrati s primeri novih divergentnih RNA virusov smo v okviru filogenetske analize odkrili nekatere evolucijsko starejše predstavnike virusnih klad. Klada Flavi (Slika 4.16)** je ločena na skupine: jingmenvirus, Flavivirus, Flavivirusu podobni virusi (angl. »Flavi-like«) (Shi in sod., 2016b), Pestivirus, Hepacivirus in Pegivirus. Filogenetsko drevo flavivirusov je bilo koreninjeno glede na proteinsko zaporedje RdRp tombusvirusa *Ostrinia nubilalis* (GAVD01019588), kar pomeni, da poleg sorodstvenih odnosov kaže tudi evolucijsko starost posamezne virusne skupine. Naši rezultati kažejo na to, da Flavivirus in jingmenvirus oblikujeta sestrski kladi in sta najmlajši monofiletski skupini. Flavivirusom podobni virusi predstavljajo evolucijsko starejšo skupino. Vendar moramo tukaj za predstavitev širšo interpretacijo rezultatov vključiti tudi samo razširjenost flavivirusov. Skupina jingmenvirusov je bila izolirana iz klopotov (»Jingmen tick virus«, »Mogiana tick virus«), listnih uš (Hemiptera, Aphioidea) (»Wuhan aphid virus 1«, »Wuhan aphid virus 22), komarjev (*Aedes sp.*, *Culex sp.*) (Guaico culex virus) in bolh (»Wuhan flea virus«). Flavivirusi so skupina, za katero je znano, da je bila izolirana iz komarjev (*Aedes sp.*, *Anopheles sp.*) (»Aedes flavivirus«, »Anopheles flavivirus«) ali sesalcev (»Yellow fever virus« (virus rumene mrzlice) iz črnega vriskača (*Alouatta caraya*), »Zika virus« iz človeka, »Dengue virus« (virus denga vročice) iz človeka). Gre torej za sestrski, kladi ki sta razširjeni pri ozki skupini žuželk in pri evolucijsko mladi skupini sesalcev. Pri tem flavivirusi in jingmenvirus, razširjeni pri komarjih, kot vektorji prenašajo viruse na ozko skupino sesalskih gostiteljev. Evolucijsko starejša skupina flavivirusom podobnih virusov, ki so jo definirali Shi in sod. (2016b), je razširjena pri več nevretenčarjih. Poleg klopotov (»Bole tick virus 4«, »Tacheng tick virus 8«), črčkov (Xinshan cricket virus), komarjev (Gamboa mosquito virus), muh (Shayang fly virus 4), vodnih drsalcev (Sanxia water strider virus 6) in čebel (*Apis flavivirus*) je bila odkrita tudi pri pajkovcih (Chelicerata) (»Shayang spider virus 4«, »Xinzhou spider virus 3«, »Xinzhou spider virus 2«), stonogah (»Wuhan centipede virus«), rakih (»Eriocheir sinensis cholera-like virus«) in glistah (Nematoda). »Soybean cyst nematode virus 5«, ki spada v skupino flavivirusom podobnih virusov, je bil izoliran iz gliste *Heterodera glycines*, ki napada rastline (Shi in sod., 2016b). Skupine pegivirusov, pestivirusov in hepacivirusov so evolucijsko nekoliko starejše skupine kot flavivirusom podobni virusi glede na naše rezultate. **Zanimivo, da so glede na filogenetsko drevo v okviru naše analize ravno evolucijsko najstarejše skupine hepacivirusov, pegivirusov in pestivirusov skoraj izključno sesalski virusi.** Shi in sod. (2016b) so odkrili morskega psa kot gostitelja »Wenling shark virusa«, uvrščenega med hepaciviruse. Proteinsko zaporedje njegove RdRp je bilo bolj divergentno v primerjavi z zaporedji ostalih hepacivirusov. Družina Flaviviridae je skupina virusov, ki tvori ločene skupine, kar pomeni, da so virusi znotraj skupine zelo podobni, virusi med skupinami pa se izrazito razlikujejo tako na nivoju zaporedja kot na nivoju strukture genoma. Koonin in sod. (2015) je Flaviviruse skupaj s Tombusvirusi uvrstil v ločeno naddružino flavivirusom podobnih virusov. Glede na naše rezultate in rezultate Shi

in sod. (2016b) imata skupini Flaviviridae in Jingmenvirus skupnega prednika. Nekateri Jingmenvirusi so segmentirani virusi, zato je bila segmentacija pomemben proces ob diverzifikaciji Flavivirusov od Jingmenvirusov. Flavivirusom podobni virusi tvorijo monofiletsko skupino virusov, ki so bolj divergentni, jasno pa se ločijo tudi od ostalih skupin. Ta skupina ima širši nabor gostiteljev, kot je bilo prvotno objavljeno. Genomi nekatih flavivirusom podobnih virusov so bistveno daljši glede na skupino Flaviviridae in so skoraj primerljivi s skupino Nidovirales. Raznolikost skupine flavivirusom podobnih virusov in njihov položaj v filogenetskem drevesu kaže na možnost, da predstavljajo predniško obliko flavivirusov, iz katerih so se razvili evolucijsko mlajši predstavniki skupine Flaviviridae in Jingmenvirusi (Shi in sod. 2016b). **Vseeno pa je rezultat filogenetske analize v nasprotju s pričakovanji, da so virusi, ki so razširjeni v ozki skupini evolucijsko mladih vretenčarjev, evolucijsko starejši kot flavivirusom podobni virusi, ki so razširjeni pri bistveno širši in evolucijsko starejši skupini nevretenčarjev.** Hkrati je zanimivo, da so evolucijske veje od skupnih prednikov hepaci, pegi in pestivirusov do zadnjega skupnega prednika hepacivirusov, zadnjega skupnega prednika pegivirusov in zadnjega skupnega prednika pestivirusov dolge, obdobje diverzifikacije posamezne skupine pa kratko (skupine med seboj izredno raznolike, posamezni člani skupine pa so si med seboj zelo podobni). V filogenetski analizi smo poleg omenjenih skupin flavivirusov odkrili predstavnika flavivirusov, ki se ni uvrstil v nobeno od trenutno znanih skupin flavivirusov – virus *Caridina multidentata* (IABX01135281). **Glede na dolžino in položaj veje lahko sklepamo, da gre za zametek nove skupine flavivirusov, ki predstavlja stanje evolucijskega prednika skupine Flaviviridae in Jingmenvirusov, hkrati pa tvori sestrsko klado s flavivirusom podobnimi virusi.** Filogenetska analiza RNA virusov pri raki predstavlja osnovo za proučevanje mehanizmov horizontalnega prenosa virusov med elementi biosfere. **V okviru filogenetske študije RNA virusov pri raki smo ugotovili, da so nekateri RNA virusi, ki smo jih našli pri raki, praktično identični ali sorodni z virusi, ki so bili odkriti pri kopenskih organizmih, predvsem kopenskih nevretenčarjih in rastlinah.** Tukaj se pojavi vprašanje o mehanizmi prenosa virusov med kopenskimi in vodnimi nevretenčarji. Pri kladi Flavi (Slika 4.16) smo našli viruse v transkriptomih rakov *Daphnia magna* in *Proasellus cavaticus* (vodni nevretenčarji), ki so sorodni virusom, razširjenim pri komarjih in čebelah. Poleg njih smo odkrili virus v transkriptomu deseteronožca *Portunus trituberculatus*, ki je soroden virusu »Tamana bat virusu«, izoliranemu iz sesalca. Pri kladi Tombus-Noda (Dodatek B; B.4, Slika B.36) je nodavirus iz *Hya araneus* (HACH01051331) 85 % identičen glede na proteinsko zaporedje RdRp »Shuangao noda-like virusu 1« (insectZJ94204). Virus je bil odkrit v mešani transkriptomski knjižnici žuželk (Shi in sod., 2016a). Pri kladi Partiti-Picobirna (Dodatek B; B.4, Slika B.39) so Shi in sod. (2016a) ravno tako odkrili »Raphanus sativus cryptic virus 1« (iz redkvice) v transkriptomu raka deseteronožca *Procambarus sp.*

V okviru filogenetske študije smo našli številne primere divergentnih RNA virusov, ki glede na proteinsko zaporedje RdRp odstopajo od trenutno znanih RNA virusov iste virusne klade. Filogenetska pestrost RNA virusov je glede na divergentnost zaporedij RdRp podobna kot pri najstarejših skupinah žuželk. V obeh primerih smo našli bistveno več virusov, ki so 20–40 % identični s katerim koli virusom iste virusne klade. **Za razlago mehanizma spreminjanja genomskega materiala pri RNA viromu rakov so pomembne ugotovitve o genomski strukturi raznolikosti.** V okviru doktorata smo našli številne primere virusov, ki odstopajo tudi v strukturi genomov glede na trenutno znane RNA viruse določene virusne klade. Odstopajo bodisi na podlagi 1) dodatnih prisotnih proteinskih domenah, 2) na podlagi dolžine genoma bodisi 3) na podlagi dodatnih neokarakteriziranih ORF. Pri kladi Flavi (Dodatek C; C.4, Slika C.26) je virus iz *Caridina multidentata* (IABX01135281) eden izmed redkih RNA virusov s celotnim virusnim genomom. Genom kodira poliprotein z dvema proteinskima domenama; helikazo naddružine 1 (DEAD/DEAH box helikaza) in RdRp. Znotraj helikaze je prišlo do premika bralnega okvirja. Ravno tako je genom virusa iz *Eriocheir sinensis* (GBZW01030122) divergenten glede na genome ostalih flavivirusov. Njegova monopartitna RNA kodira poliprotein z dvema ORF. Prvi kodira flavivirusni glikoprotein (angl. »flavivirus glycoprotein«) (pfam00869) in glikoprotein E (TIGR04240), drugi ORF kodira helikazo DEAD/DEAH box, FtsJ metiltransferazo in RdRp. Virus iz *Proasellus cavaticus* (HAES01086036) ima strukturo, podobno »Shayang fly virusu 4«, vsebuje pa še dodatno domeno »reticulocyte-binding/rhoptry protein«. Pri kladi Hepe-Virga (Dodatek C; C.4, Slika C.27) ima virus iz *Omatogammarus flavus* (GEQS01054466) prisotno tudi domeno NADAR (angl. »E.coli swarming motility protein«). Virus iz *Eriocheir sinensis* (GBZW01014116) kodira poleg NADAR domene tudi Macro domeno, tipično za toga in hepeviruse. Pri kladi Luteo-Sobemo (Slika 4.33) je genom virusa iz *Proasellus parvulus*

(HAFG01115646) kodiran s tremi ORF-ji; tretji ni funkcionalno okarakteriziran. Tipično so luteovirusi kodirani s štirimi (ali več) okarakteriziranimi ORF-ji. Med prvima dvema je »frameshift« mutacija. Stop kodon med drugima dvema delno prepozna tRNA (angl. »stop codon readthrough«). Ostali ORF-ji se prekrivajo z glavnima ORF-jema in so neokarakterizirani. Shi in sod. (2016a) so odkrili primere segmentiranih luteovirusov, ki na prvem segmentu kodirajo dva ORF-ja. Prvi je neokarakteriziran, drugi ORF je nestruktur in kodira RdRp. Drugi segment je strukturni in kodira virusni protein plašča (angl. »viral coat protein, S domain«). Genom virusa iz *Proasellus parvulus* (HAFG01115646) je nekoliko krajši od tipičnega genoma luteovirusa. Od trenutno znanih luteovirusov ga loči še dodaten neokarakteriziran ORF. Podobno genomsko strukturo imajo tudi virusi iz *Brasellus peltatus* (HAEN01054361), *Caligus rogercresceyi* (GAZX01036530) in *Proasellus escolai* (HAEV01103964). Glede na povprečno dolžino genoma luteovirusov in glede na strukturo segmentiranih luteovirusov (»Wuhan house centipede virus« (WHYY19909); segment 1; 3 kb), ki so jih odkrili Shi in sod. (2016a), gre verjetno za segmentirane predstavnike. Vendar pa nismo našli drugega segmenta. Pri **kladi Mono-Chu (Dodatek C; C.4, Slika C.29)** izstopa primer chuvirusa iz *Hya araneus* (HACH01032123), pri katerem so domene monopartitne RNA v obratnem vrstnem redu kot pri referenčnih chuvirusih. Prvi segment je glikoprotein (angl. »putative glycoprotein«), drugi segment je funkcionalno slabo okarakteriziran, tretji segment je nukleoprotein in zadnji segment monopartitne RNA je L segment (RdRp). Pri ostalih primerih je razlika v glavnem v neokarakteriziranih ORF. Pri **kladi Picorna-Calici (Dodatek C; C.4, Slika C.32)** se pri nekaterih predstavnikih pikornavirusov, pri katerih smo odkrili celoten genom, pojavljajo dodatne domene, kot je »RING finger superfamily domain« pri pikornavirusu iz *Macrobrachium nipponense* (GCVG01031551), dodatna helikazna domena in peptidaza C37 in Bunyavirusni nestruktur in protein NS-S (angl. »bunyavirus nonstructural protien NS-S«) pri virusu iz *Eucyclops serrulatus* (GARW01000621) in domena NADAR pri virusu iz *Macroporeiopus parvus* (GEQM01001122). Pri **kladi Toti-Chryso (Dodatek C; C.4, Slika C.33)** se ravno tako pri virusu iz *Proasellus grafi* (HAEX01038731) pojavljajo dodatni neokarakterizirani ORF-ji.

Analiza strukture genomov je dodatno potrdila rezultate naše filogenetske analize. Pri skupini rakov smo dokazali pester nabor raznolikih virusnih genomov. Z dodatnim vzorčenjem raznolikosti RNA virusov ne samo pri deseteronožcih in enakonožcih, temveč tudi pri ceponožcih in postranicah, smo prikazali celotno sliko o kompleksnosti RNA viroma pri raki. Dokazali smo, da so ne samo žuželke na splošno, temveč tudi raki bogat rezervoar raznolikih RNA virusov. Potrdili smo, da je celotna skupina Pancrustacea pester rezervoar raznolikih RNA virusov. Bogat vir raznolikih RNA virusnih genomov, ki so jih v preteklih raziskavah dokazali za najmlajše skupine žuželk, je obstajal že pri evlucijsko starejših raki.

5.5 Skupina Spiralia: mehkužci (Mollusca), ploščati črvi (Platyhelminthes), kolobarniki (Annelida) in kotačniki (Rotifera)

Raki in žuželke (Pancrustacea) so skupina členonožcev, ki skupaj z glistami tvori skupino Ecdysozoa (Dunn in sod., 2014). S filogenetskimi analizami RNA virusov pri raki in žuželkah smo dokazali prisotnost raznolikih genomov pri skupini Pancrustacea. Shi in sod. (2016a) so v študijo raznolikosti RNA virusov vključili tudi starejše skupine Ecdysozoa, kot so gliste (Nematoda). S tem so pokazali, da je celotna skupina Ecdysozoa bogat vir raznolikih RNA virusov. Skupina Ecdysozoa tvori sestrsko klado s skupino Spiralia. Skupaj definirajo skupino Protostomia (protostomi). Da bi dokazali, da celotna skupina Protostomia, ne samo Ecdysozoa, vsebuje pester nabor raznolikih RNA virusov, so Shi in sod. (2016a) izvedli najboljše analizo raznolikosti RNA virusov še pri drugi veji skupine Protostomia – skupini Lophotrochozoa. Iz vzorcev RNA organizmov skupine Lophotrochozoa so naredili 21 knjižnic. Pri tem jih je 13 pripadalo mehkužcem, štiri ploščatim črvom, tri kolobarnikom (Annelida) in ena skupini Sipuncula. Skupno so našli okrog 330 novih RNA virusov, ki so pripadali 18 od 25 virusnih klad (Dodatek A; A.1, Slika A.3). Pri skupini niso bile odkrite prisotne klade: Permutotetra, Reo, Yuevirus, Ophio, Poty, Flavi in Cysto. Ugotovili so unikaten profil prisotnih virusnih klad tako pri mehkužcih kot pri kolobarnikih in skupini Sipuncula. Pri vseh treh skupinah je bila ponovno najbolj zastopana virusna klada Picorna-Calici. Zastopanost ostalih virusnih klad je bila unikatna za posamezne skupine znotraj organizmov Lophotrochozoa. Pri mehkužcih so kladi Picorna-Calici sledile klade Tombus-

Noda in Narna-Levi. Pri kolobarnikih (Annelida) sta sledili Narna-Levi in Partiti-Picobirna. Pri skupini Sipuncula sta sledili Narna-Levi in Tombus-Noda. Znotraj klade Picorna-Calici so ugotovili izrazito bogato zastopanost vodnih pikornavirusov (angl. »aquatic picornavirales«), sledita ji Posavirusu podobni virusi in dicistrovirusi. Arivirusu, Noravirusu in Kelp fly virusu podobni virusi so bili redki pri skupini Lophotrochozoa (Dodatek A; A.2 Slika A.4). Shi in sod. (2016a) so z obsežno analizo RNA virusov pri skupini Lophotrochozoa postavili celotno skupino Protostomia kot center raznolikosti RNA virusov. Vendar skupina Shi in sod. (2016a) ni prikazala celostne slike o raznolikosti RNA virusov pri skupini Spiralia. Vzrok je bil namreč omejeno vzorčenje. Kljub temu, da so analizirali štiri knjižnice ploščatih črvov, so v njihovih transkriptomih odkrili samo dva predstavnika klade Bunya-Arena, enega predstavnika klade Mono-Chu in enega predstavnika klade Reo. **Vzorčenje ploščatih črvov je bilo torej revno.** Ploščati črvi (Platyhelminthes) so izjemno raznolika skupina organizmov. Njihovo bilateralno simetrično telo ni prekrito s kutikulo. Nekatere skupine so obligatorni paraziti, medtem ko so se druge skupine adaptirale na kopensko ali vodno življenje (Littlewood, 2006). **Zaradi njihovega načina prehranjevanja in njihove fiziologije bi lahko ploščati črvi predstavljali pomemben vir raznolikih RNA virusov.** V okviru doktorata smo analizirali 225 genomskih in transkriptomskih knjižnic. Od tega jih je 125 pripadalo mehkužcem in 66 ploščatim črvom. Našli smo več kot 70 novih RNA virusov. Od tega okrog 30 pri ploščatih črvih in okrog 40 pri mehkužcih (Slika 4.3). Pri mehkužcih smo odkrili 10 od 25 RNA virusnih klad, pri ploščatih črvih pa 11 od 25. Tako mehkužci kot ploščati črvi imajo unikaten virom, pri katerem je najbolj zastopana klada Picorna-Calici. Pri mehkužcih sledijo klade Partiti-Picobirna, Hepe-Virga in Mono-Chu, pri ploščatih črvih pa sledita kladi Mono-Chu in Bunya-Arena. Ostale klade so redkeje prisotne pri vsaki izmed skupin. Za skupino Spiralia smo ugotovili prisotnost 14 od 25 virusnih klad. **Izmed klad, ki smo jih našli pri skupini Spiralia, sta bili kladi Reo in Flavi prvič odkriti v okviru naše raziskave.** Znotraj klade Picorna-Calici smo odkrili prisotnost šestih od osmih pikornavirusnih skupin, sicer v nekoliko drugačni zastopanosti kot Shi in sod. (2016a). Nobeden od naših zadetkov se ni uvrstil med Kelp fly virusu podobne viruse. **V nasprotju s Shi in sod. (2016a) pa smo dokazali prisotnost iflavirusov tudi pri skupini Spiralia (Slika 4.6).** V okviru doktorata smo izvedli celotno vzorčenje raznolikih RNA virusov pri skupini Spiralia, s tem da smo analizirali obsežen nabor genomov in transkriptomov ploščatih črvov. S tem smo v nasprotju s Shi in sod. (2016a) prikazali celotno sliko prisotnosti virusnih klad pri skupini Spiralia. Ob primerjavi naših podatkov z literaturo smo ugotovili, da virom skupine Lophotrochozoa obsega 20 od 25 definiranih virusnih klad, kar je primerljivo z njeno sestrsko klado Ecdysozoa. Pogoste klade so: Picorna-Calici, Mono-Chu, Partiti-Picobirna, Hepe-Virga, Narna-Levi in Tombus-Noda. Nekoliko redkejša klade so: Flavi, Toti-Chryso, Orthomyxo, Bunya-Arena, Luteo-Sobemo in Weivirus. Redke klade so: Astro, Birna, Hypo, Nido, Qinvirus, Yanvirus in Zhaovirus. Klade, ki so odsotne pri skupini Lophotrochozoa, so: Permutotetra, Ophio, Poty in Cysto.

Za nadaljnje sklepanje o celotnem naboru divergentnih RNA virusov pri skupini Spiralia in o poteku evolucijskih sprememb njihovih genomskih struktur smo izvedli 1) **filogenetsko analizo** in 2) **primerjalno analizo genomskih struktur novih RNA virusov** pri skupini Spiralia. Ugotovitve so bile pomembne za nadaljnje preverjanje ugotovitev Shi in sod. (2016a), da je skupina Protostomia dejansko center raznolikosti RNA virusov.

Tako kot pri rakih smo v okviru filogenetske analize tudi pri skupini Spiralia našli številne divergentne RNA viruse, ki so od 20–40 % identični kateremu koli virusu iste virusne klade. Pri kladi Astro (Dodatek B; B.5, Slika B.44) smo našli primer divergentna astrovirusa iz *Bdelloura candida* (GFKB01038875 + GFKB01027725) in iz *Pomphorhynchus laevis* (AM918867), ki sta manj kot 30 % identična kateremu koli astrovirusu. Pri kladi Birna (Dodatek B; B.5, Slika B.45) smo v transkriptomu ploščatega črva *Procotyla fluviatilis* (GAKZ) našli divergenten birnavirus. Pri kladi Bunya-Arena (Dodatek B; B.5, Slika B.46) je zaporedje RdRp Wuhan spider virusa 41 % identično proteinskemu zaporedju RdRp virusa iz *Procotyla fluviatilis* (GAKZ01048260). Pri kladi Flavi (Dodatek B; B.5, Slika B.47) smo našli štiri divergentne flaviviruse. Pri kladi Hepe-Virga (Dodatek B; B.5, Slika B.49, Slika B.57 in Slika B.51) smo odkrili divergenten hepevirus pri *Elliptio complanata* (GAHW01000160), ki je 28–30 % identičen z ostalimi hepevirusi. Odkrili smo tudi divergentne negevirus in divergenten togavirus pri *Elysia timida* (GBRM01100701), ki je v rang 27 % identičnosti z ostalimi togavirusi. Pri kladi Luteo-Sobemo (Dodatek B; B.5, B.48) smo našli divergenten luteovirus, ki spada med »Carascovirusu podobne viruse«. Pri kladi Narna-Levi (Dodatek B; B.5, Slika B.50) velja izpostaviti virus iz *Rotaria tardigrada*

(GDRK01026552), ki je 39 % identičen virusu »Hubei narna-like virus 22« (SZmix20730). Pri **kladi Tombus-Noda (Dodatek B; B.5, Slika B.52)** smo našli divergenten nodavirus pri *Opisthorchis felinus* (GBJA01011878), ki je okrog 30 % identičen ostalim trenutno znanim nodavirusom. Tudi pri **kladi Mono-Chu (Slika 4.19)** smo odkrili nekaj divergentnih virusov, ki so v rangi 30 % identični z ostalimi virusi klade Mono-Chu, kot so *Microphallus sp.* (GFFK01014037), *Girardia dorotocephala* (GEIG01038096), *Fasciola hepatica* (GEVX) in *Schistocephalus solidus* (GEEE). Pri **kladi Partiti-Picobirna (Slika 4.20)** velja izpostaviti divergentna virusa pri *Macrostromum lignano* (LFJF01002397) in pri *Elysia timida* (GBRM01095329). Pri **kladi Toti-Chryso (Dodatek B; B.5, B.56)** je proteinsko zaporedje RdRp virusa, ki smo ga našli v transkriptomu *Schmidtea mediterranea*, 28 % identično s proteinskim zaporedjem RdRp »Wenzhou toti-like virusa 2« (beimix73231). **Nekateri izmed odkritih divergentnih predstavnikov predstavljajo hkrati starejše predstavnike virusnih klad.** Izpostaviti velja primer klade Mono-Chu (Slika 4.19), kjer smo v transkriptomu *Aplysia californica* našli divergenten virus, ki se je v drevesu povezal z virusom »Caledonia dog whelk rhabdo-like virusom 1«. Virus »Caledonia dog whelk rhabdo-like virus 1« je bil v raziskavi Waldron in sod. (2017) izoliran iz škrlatnega voleka (angl. »dog whelk«). Proteinski zaporedji sta med seboj manj kot 30 % identični, hkrati pa v filogenetskem drevesu zavzemata položaj najstarejših predstavnikov klade Mono-Chu.

Rezultati analize raznolikosti genomskih struktur se ujemajo z rezultati filogenetske analize novih RNA virusov pri skupini Spiralia. Večina genomskih struktur novih RNA virusov se bistveno ne razlikuje od genomskih struktur trenutno znanih RNA virusov iste virusne klade. **Pojavljajo pa se številni primeri spremenjenih genomskih struktur** glede 1) dodatnih proteinskih domen, 2) zamenjanega vrstnega reda proteinskih domen ali 3) prisotnosti funkcionalno neokarakteriziranih ORF. Pri **kladi Astro (Slika 4.35)** genom virusa iz *Bdelloura candida* (GFKB01038875 + GFKB01027725) kodirajo tri ORF-ji v zaporedju RdRp-virusni protein plašča (S domena) – ORF3 (ki je funkcionalno neokarakteriziran). Referenčni astrovirusi imajo ORF kodirane v nizu peptidaza – RdRp – virusni protein plašča. Pri kladi Flavi divergenten genom virusa iz *Idiosepius notoides* (GFNE01070033) kodira niz: flavivirusni glikoprotein (angl. »flavivirus glycoprotein«) – protein DinF – helikaza naddružine 1 (DEAH/DEAD box) – RdRp. **Pri kladi Hepe-Virga (Dodatek C; C.5, Slika C.37)** glavni poliprotein hepevirusa iz *Elliptio complananta* (GAHW01000160) nima C41 in Macro domene, genom pa tvorijo tudi dodatni neokarakterizirani ORF. Togavirus iz *Elysia timida* (GBRM01100701) kodira glavni nestrukturni ORF, ki mu sledijo trije funkcionalno neokarakterizirani ORF – struktura, ki je tipična za negevirusne. Glavni nestrukturni ORF pri togavirusu zajema niz domen virusna metiltransferaza-helikaza-FtsJ metiltransferaza, ki je pri negevirusih tipično na N-teminalnem delu nestrukturnega poliproteina. Pri skupini Spiralia najdemo podobne genomske strukture **klade Luteo-Sobemo (Dodatek C; C.5, Slika C.38)** kot pri rakih. Genom virusa, najdenega v transkriptomu *Euprymna scolopes*, je kodiran z dvema ORF. Prvi kodira tripsinu podobno peptidazno domeno (angl. »Trypsin-like peptidase«), drugi ORF kodira RdRp_1. Glede na dolžino genoma 3 kb sklepamo, da je tako kot pri luteovirusu »Xingshan nematode virus 5« (XSNXC27943) sestavljen še iz drugega strukturnega segmenta, ki pa ga nismo odkrili. Pri **kladi Toti-Chryso (Slika 4.37)** glavne razlike v strukturah genomov tvorijo neokarakterizirani ORF.

Z reprezentativnim vzorčenjem transkriptomov pri analizi raznolikosti in razširjenosti RNA virusov smo v analizo vključili vse najpomembnejše skupine Spiralia, med njimi tudi ploščate črve, ki do sedaj še niso bili analizirani s strani raznolikosti prisotnega viroma. **Tako z vidika prisotnosti virusnih klad kot z vidika raznolikosti virusnih genomov smo potrdili ugotovitve skupine Shi in sod. (2016a), da celotna skupina Protostomia predstavlja bogat vir raznolikih RNA virusov. Vseeno pa se odpira nadaljnje vprašanje, ali bogat nabor raznolikih genomskih struktur vsake od virusnih klad izvira pri skupini Protostomia ali je njihov izvor v resnici starejši in je nastal že pri najstarejših skupinah živali.** Pojavlja se vprašanje, kje se je tekom koevolucije virusov in gostiteljev odvila »eksplozija« diverzifikacije virusnih klad – pri skupini Protostomia ali že pri najstarejših skupinah živali. **Ugotovitve o raznolikosti RNA virusov pri skupini Protostomia sprožijo nadaljnji val sklepov o dejanski starosti vsake od virusnih klad.** Teoretični možnosti bi lahko bili dve: najstarejše predstavnike virusnih klad lahko najdemo pri skupini Protostomia, lahko pa najdemo evolucijsko najstarejše predstavnike virusnih klad pri predniku živali. V drugem primeru se je evolucija virusne klade začela skupaj z nastankom prednika živali. **V vsakem primeru nam nadaljnje ugotovitve v okviru doktorata o raznolikosti in razširjenosti RNA virusov pri najstarejših skupinah živali lahko podajo odgovor na omenjena vprašanja.**

5.6 Najstarejše skupine živali: ožigalkarji (Cnidaria), spužve (Porifera) in rebrače (Ctenophora)

Le malo študij se je ukvarjalo z raznolikostjo in razširjenostjo RNA virusov pri najstarejših skupinah živali. Dve izmed njih sta Shi in sod. (2016a) in Waldron in sod. (2017). Shi in sod. (2016a) so izolirali RNA iz morske vetrnice *Bolocera sp.*, pri čemer so ustvarili eno samo knjižnico. Našli so okrog 20 virusov, pri čemer je bila večina skoraj identična RNA virusom mehkužcev in rakov. En virus je pripadal kladi Hepe-Virga, eden Partiti-Picobirna, 12 Picorna-Calici in šest kladi Tombus-Noda. Raziskava je pokazala zelo omejeno zastopanost samo štirih od 25 definiranih virusnih klad pri najstarejših skupinah živali – najbolj zastopana je bila Picorna-Calici. Waldron in sod. (2017) so izolirali RNA iz spužve *Halichondria panicea* (angl. »breadcrumb sponge«) in *Actinia equina* (angl. »beadlet sea anemone«) ter odkrili okrog 60 novih RNA virusov. Njihovi rezultati so pokazali bistveno večjo raznolikost in razširjenost RNA virusov pri najstarejših skupinah živali. RNA virom organizmov *Halichondria panicea* (angl. »breadcrumb sponge«) in *Actinia equina* (angl. »beadlet sea anemone«) je zajemal virusne klade Hepe-Virga (4 virusi), Picorna-Calici (15 virusov), Tombus-Noda (9 virusov), Mono-Chu (2 virusa), Narna-Levi (10 virusov), Partiti-Picobirna (2 virusa), Luteo-Sobemo (3 virusi) in Weivirus (13 virusov). Raziskavi sta nakazali, da so pri najstarejših skupinah živali prisotne raznolike virusne klade; skupno sta odkrili več kot 80 RNA virusov. **Vendar je večina odkritih virusov 80 %-100 % identičnih z RNA virusi, ki so izvirali iz mehkužcev in rakov (ostalih vodnih nevretenčarjev).** Poleg omenjenega gre za vzorčenje samo treh organizmov iz najstarejših skupin živali, kar ne predstavlja reprezentativnega vzorca, ki bi pokazal celotno sliko o raznolikosti in razširjenosti RNA virusov pri nevretenčarjih.

V okviru doktorskega dela smo analizirali 89 genomov in transkriptomov najstarejših skupin živali. Največ (74) genomov in transkriptomov je pripadalo ožigalkarjem in 10 spužvam. Poudariti moramo, da so bile izmed 10 knjižnic spužev kar štiri narejene iz vrste *Halisarca caerulea*, kar nakazuje na slabo vzorčenje transkriptomov spužev. Našli smo 114 novih RNA virusov, ki so pripadali kar 17 od 25 definiranih virusnih klad: Birna, Partiti-Picobirna, Toti-Chryso, Reo, Mono-Chu, Bunya-Arena, Orthomyxo, Ophio, Yuevirus, Qinvirus, Hepe-Virga, Luteo-Sobemo, Narna-Levi, Nido, Picorna-Calici, Tombus-Noda in Weivirus (**Slika 4.1**). Tako kot Shi in sod. (2016a) in Waldron in sod. (2017) smo ugotovili, da je **pri spužvah in ožigalkarjih najbolj zastopana skupina Picorna-Calici**. Skupina Weivirus je nekoliko bolj bogato prisotna pri ožigalkarjih, skupina Tombus-Noda pa pri spužvah. Po naših rezultatih so skupine Yuevirus, Qinvirus, Ophio, Toti-Chryso, Reo, Orthomyxo, Bunya-Arena in Birna redke skupine pri najstarejših skupinah živali in omejene samo na ožigalkarje, skupina Poty pa je omejena na spužve. Skupine Potyviridae nismo zasledili pri nobeni od drugih analiziranih skupin nevretenčarjev, torej je očitno omejena samo na ožigalkarje. Nasprotno pa evolucijsko mlajše skupine flavivirusov nismo zasledili pri najstarejših skupinah živali. Ožigalkarji kažejo bistveno bolj bogato prisotnost raznolikih virusnih skupin (17/25 virusnih klad) v primerjavi z ožigalkarji (8/25 virusnih klad). Vzrok za to je lahko večja zastopanost z genomske in transkriptomske knjižnicami oz. bolj reprezentativno vzorčenje. Znotraj klade Picorna-Calici smo pri najstarejših skupinah živali našli prisotnih sedem od osmih skupin pikornavirusov (**Slika 4.6**). Pri ožigalkarjih so bile prisotne vse pikornavirusne skupine, razen rastlinskih Secoviridae – »aquatic picornavirales«, Dicistro, Ifla, Nora virusom podobni virusi, Posavirusom podobni virusi, Arivirusom podobni virusi, Kelp fly virusom podobni virusi. Poleg omenjenega sta se v filogenetski analizi pojavili dve novi skupini, ki prvotno nista bili definirani. Pri spužvah so prisotni predstavniki skupin »aquatic picornavirales«, Dicistrovirus, Kelp fly virusu podobni virusi, Arivirusu podobni virusi in predstavniki, ki pripadajo dvema novima virusnima skupinama.

Odkritje bogatih in raznolikih RNA viromov pri najstarejših skupinah živali je ključno za nadaljnjo razlago **o izvoru, prvem pojavu za najstarejše skupine živali tipične virusne skupine (virusne klade) in mehanizmu diverzifikacije virusnih klad**. Glede na rezultate predhodnih filogenetskih analiz se sklepa, da so spužve najstarejša skupina živali. Presenetljivo so v naši filogenetski raziskavi spužve pokazale nekoliko revnejši virom kot ožigalkarji. Glede na trenutno slabo zastopanost spužev v genomske in transkriptomske podatkovnih bazah je to lahko najverjetnejši vzrok za revnejšo sliko njihovega RNA viroma. Za razlago nastanka prvotnega živalskega RNA viroma obstajata dva možna scenarija. Prvi vključuje začetno

omejeno raznolikost RNA viroma pri predniku živali in kasnejšo diverzifikacijo RNA viroma vzporedno s pojavom novih skupin živali (ožigalkarji, rebrače, dvobočno somerne živali (Bilateria)). Nasprotno bi se lahko predpostavila zgodnja diverzifikacija RNA virusov že v začetnih stopnjah diverzifikacije skupin živali (že pri predniku vseh živali), vendar trenutni rezultati ne nakazujejo na veliko verjetnost slednjega scenarija. Za končni sklep o dejanskem poteku diverzifikacije živalskega RNA viroma bi potrebovali večje, bolj bogato vzorčenje transkriptomov spužev. Poseben primer je klada Picorna-Calici. Skupina Shi in sod. (2016a) je dokazala prisotnost vseh osmih skupin Picorna-Calici klade pri skupini Protostomia. V okviru doktorskega dela smo dokazali prisotnost sedmih od osmih skupin pikornavirusov tudi pri najstarejših skupinah živali, kar dokazuje eksplozijo raznolikosti pikornavirusov in njeno zgodnjo diverzifikacijo že pri najstarejših skupinah živali.

Predhodne filogenetske analize so pokazale bogato raznolikost genomov in razširjenost RNA virusov pri skupini Protostomia (Shi in sod., 2016a). Naše filogenetske analize RNA viromov pri živalih so pokazale nove rezultate z vidika 1) **starosti virusnih klad**, 2) **filogenetske klasifikacije** in 3) **genomske strukture**.

Predhodne raziskave so pokazale, da so RNA virusi razširjeni čez celotno skupino Protostomia. V okviru doktorskega dela smo dokazali njihovo prisotnost tudi pri najstarejših skupinah živali. Nabor gostiteljev je torej pri večini virusnih klad s tem razširjen, hkrati pa lahko sklepamo o dejanski starosti RNA virusnih skupin. Prisotnost virusnih skupin (med njimi tudi Picorna-Calici) pri najstarejših skupinah živali dokazuje, da je **evolucijska starost večine virusnih skupin bistveno starejša, kot je bilo prvotno ugotovljeno**.

Večina novih RNA virusov pri najstarejših skupinah živali se ne poveže skupaj s trenutno znanimi RNA virusi, temveč bodisi 1) tvorijo samostojno vejo bodisi 2) tvorijo novo skupino skupaj z ostalimi novimi RNA virusi, odkritimi v okviru doktorskega dela (LBA, angl. »*long branch attraction*«, privlak dolgih vej). Omenjeni RNA virusi oz. proteinska zaporedja RdRp omenjenih virusov so samo 20–30 % identična s trenutno znanimi RNA virusi. To kaže na **zgodnji nastanek novih potencialnih virusnih skupin znotraj virusnih klad oz. zgodnjo diverzifikacijo virusnih skupin**. Nekateri izmed divergentnih predstavnikov RNA virusnih klad pri najstarejših skupinah živali so glede na položaj filogenetskega drevesa tudi **najstarejši predstavniki virusnih klad in hkrati dokazujejo izvor virusnih skupin pri najstarejših skupinah živali**. Tipična primera sta odkrita najstarejša predstavnika pri **kladah Mono-Chu iz *Hydra vulgaris* (GEVZ01001862) (Slika 4.23) in Picorna-Calici iz *Anthopleura elegatissima* (GBYC01009384) (Dodatek B; B.6, Slika B.66)** Najstarejši predstavniki virusnih klad, ki smo jih našli pri najstarejših skupinah živali, omogočajo ključen vpogled v zgodnjo evolucijo nevretenčarskih RNA virusov.

V okviru analize genomov RNA virusov pri najstarejših skupinah živali smo ugotovili, da je pri večini novih predstavnikov RNA virusnih klad genomska struktura podobna genomski strukturi trenutno znanih predstavnikov iste virusne klade. Našli pa smo **primere novih RNA virusnih genomskih struktur, ki se razlikujejo od tipičnih predstavnikov pri skupini Spiralia in pri členonožcih**. Pri kladi Mono-Chu večino razlik v genomski organizaciji predstavlja število ORF, ki so v glavnem strukturno neokarakterizirani. Pri predstavnikih nekaterih virusnih skupin se **pojavi tudi dodatne domene, ki niso tipične za virusno kladu**. Pri kladi Hepe-Virga (Dodatek C; C.6, Slika C.44) je prvi primer virus iz *Craterolophus convolvulus* (HAGZ01026509), ki na glavnem nestrukturnem poliproteinu kodira tudi domeno NADAR (angl. »*E. coli swarming motility protein*«). Drugi primer je virus iz *Porites asteroides* (GEHP01110756), ki na tretjem strukturnem ORF kodira virusni protein plašča družine Flexi (angl. »*viral coat protein; Flexi family*«), ki je tipičen za red Tymovirales. Pri kladi Picorna-Calici (Slika 4.41) smo odkrili primere virusov z 1) dodatnimi Rhv domenami (*Millepora alcicornis* (GFAS012252765), *Halyclystus auricola* (HAHA01054928+HAHA01054926) in *Halyclystus sanjuanensis* (HAHB01071359)), 2) dodatno RING domeno (*Halyclystus auricola* (HAHA01047705)), 3) zamenjavo tipične peptidaze C3 s tripsinu podobno serinsko proteazo (»*trypsin-like serine protease*«) (*Nematostella vectensis* (HADN01010937)). Pri najstarejših skupinah živali smo odkrili tudi primere novih RNA virusov z **nekoliko drugačno strukturno organizacijo genomov**. Pri kladi Picorna-Calici (Slika 4.41) je bila pri virusu iz *Millepora alcicornis* (GFAS01286708) RdRp kodirana na ločenem ORF v drugem bralnem okvirju znotraj glavnega ORF. Presenetljivo so **genomi nekaterih predstavnikov RNA virusov, najdenih v transkriptomih najstarejših skupin živali, bistveno daljši v primerjavi s tipično dolžino genoma, značilno za dotično virusno kladu**. Pri kladi Mono-Chu (Slika 4.39) na primer rabdovirus iz *Hydra*

vulgaris – (GEVZ01001862) z dolžino 21067 bp predstavlja najdaljši doslej odkrit mononegavirus. Genomi virusov, najdenih v transkriptomih najstarejših skupin živali, ki so daljši od tipične dolžine genomov, lahko predstavljajo izvirno stanje prednika virusne klade, kar pomeni, da je tekom evolucije lahko prišlo do skrajšanja virusnega genoma. Raznolikost genomskih struktur je ključna za nadaljnjo interpretacijo poteka diverzifikacije vsake od virusnih klad. Celoten spekter raznolikosti genomskih struktur za vsako od RNA virusnih klad, ki so jih Shi in sod. (2016a) odkrili pri evlucijsko mlajših skupinah nevretenčarjev, se pojavlja tudi pri najstarejših skupinah živali (z izjemo posameznih domen). **Iz tega lahko sklepamo, da je bogata diverzifikacija virusnih klad nastala že v predniku živali. Pojav lahko najbolj vidimo pri najbolj zastopani virusni kladi Picorna-Calici V okviru klade Picorna-Calici lahko vidimo pojav »Big Bang«, ali eksplozija raznolikih genomskih struktur že pri predstavniki pikornavirusov, povezanih z najstarejšimi skupinami živali. Eksplozija raznolikosti genomskih struktur pikornavirusov se nato ohrani pri vseh skupinah nevretenčarjev.**

Razlike in podobnosti v strukturi organizaciji genomov predstavnikov sorodnih virusnih klad lahko v nekaterih primerih pri najstarejših skupinah živali povežemo s filogenetsko klasifikacijo v zgodbo o najverjetnejšem poteku prenosov posameznih virusnih genov in sprememb v virusnih genomih. Tipični primer je **klada Mono-Chu (Slika 4.23)**. Shi in sod. (2016a) so dokazali njeno sorodnost s klado Yuevirus. V filogenetsko analizo klade Mono-Chu smo torej vključili tudi kladi Yuevirus in Ophioviridae. Yuevirus je skupina segmentiranih virusov z dvema segmentoma monocistronske RNA. Vsak segment nosi en odprt bralni okvir (ORF, angl. »open reading frame«). Prvi segment je nestrukturiran in vsebuje ORF, ki kodira zapis za RdRp z ohranjenim katalitičnim motivom SDD. Gre za divergentno obliko RdRp z neopredeljeno družino. Skupina Ophioviridae je evlucijsko mlajša od skupine Yuevirus. Glivni ofiovirusi imajo relativno enostavno genomsko strukturo, 7,8 kb dolg genom, sestavljen iz enega glavnega odprtega bralnega okvirja, ki je homologen L segmentu, značilnemu za skupino Mononegavirales. Pri glivnih ofiovirusih je poleg mononeg_RdRp z ohranjenim katalitičnim motivom SDD vidna tudi šibka domena mononeg_mRNA_cap (angl. »*Mononegavirales mRNA-capping region V*«), ki ni vidna pri rastlinskih ofiovirusih. Rastlinski ofiovirusi so segmentirani, poleg glavnega segmenta, homolognega L segmentu pri skupini Mononegavirales oz. pri kladi Mono-Chu, vsebujejo še dva ali tri segmente. Eden izmed njih vsebuje odprt bralni okvir, ki kodira protein plašča (angl. »*Plant viral coat protein nucleocapsid*«) (pfam11128) (Marchler-Bauer in sod., 2017 – CDD). Mononegavirusi so široko razširjena skupina RNA virusov med vretenčarji, nevretenčarji in rastlinami. Najstarejša skupina znotraj skupine Mononegavirales naj bi bili Chuvirusi z monopartitno RNA, ki je načeloma sestavljena iz treh ORF. Za red Mononegavirales je značilen L segment (mononegavirusna RNA z »mRNA capping enzyme« in »mononegavirales mRNA capping region« se nahaja na N – terminalnem delu genoma). Pri skupinah Nyamiviridae in Rhabdoviridae, ki oblikujeta sestrski skupini, evlucijsko mlajši od Chuvirusov, se L segment nahaja na C-terminalnem delu. Tekom evolucije je lahko prišlo do zlepljanja segmentov. Možno je, da sta se zgodila dva neodvisna pojava. Pri evoluciji Chuvirusov so se ostali segmenti lahko dodajali na C-terminalnem koncu L segmenta. Pri evoluciji prednika Nyamivirusov in Rhabdovirusov pa so se dodajali na N-terminalnem koncu L segmenta. Skupina, v katero se je vključil »Lettuce big-vein associated varicosavirus« (BAC16226) (varicosavirus), je evlucijsko starejša skupina Mononegavirales, ima dva segmenta – za mononegaviruse značilen L segment in ločen segment policistronske RNA, na katerem se poleg drugih proteinov nahaja protein plašča (angl. »*Plant viral coat protein nucleocapsid*«), homologen proteinu plašča skupine Ophioviridae. Lahko ga je prevzel od skupine Ophioviridae. »Lettuce big-vein associated varicosavirus« (BAC16226) je edini mononegavirus, pri katerem je protein plašča (angl. »*Plant viral coat protein nucleocapsid*«) ohranjen. Poleg omenjenega virusa so tudi nekateri drugi primeri mononegavirusov »ostali« segmentirani.

Shi in sod. (2016a) so premaknili center raznolikosti RNA virusov od členonožcev na skupino Protostomia. Z analizo raznolikosti RNA virusov pri najstarejših skupinah živali smo center in vir raznolikosti premaknili od skupine Protostomia na najstarejše skupine živali. Pokazali smo, da so najstarejše skupine živali pomemben rezervoar številnih divergentnih RNA virusov, ki ponujajo vpogled v izvor in evolucijo genomske organizacije RNA virusov. Najstarejše skupine živali predstavljajo pomemben rezervoar virusne genetske raznolikosti in igrata ključno vlogo v evoluciji RNA virusov. Pokazali smo tudi, da se je »eksplozija« raznolikih RNA virusov vsake od virusnih klad pravzaprav pojavila skupaj s prednikom živali

in da se je evolucija vsake od virusnih klad s pretokom genomskega materiala začela že v predniku živali.

5.7 Nevretenčarji

Z obsežno analizo raznolikosti RNA virusov pri žuželah, raki, pajkovcih, stonogah, glistah in skupini Spiralia so Shi in sod. (2016a) premaknili glavni vir raznolikosti RNA virusov od členonožcev na celotno skupino Protostomia. Vseeno je bilo njihovo vzorčenje omejeno in ni bilo reprezentativno. Pri vsaki od analiziranih skupin nevretenčarjev so izključili pomembne podskupine. Pri žuželah niso analizirali kožekrilcev, metuljev in najstarejših skupin žuželk (pražuželk). Pri raki niso analizirali ceponožcev, ene najpomembnejših zooplanktonskih biomas na Zemlji, in postranic. Pri skupini Spiralia so se v glavnem osredotočili na mehkužce, medtem ko ploščati črvi kot pomemben vir raznolikih RNA virusov niso bili reprezentativno raziskani z vidika RNA viroma. Vseeno so Shi in sod. (2016a) odkrili bogat nabor raznolikih RNA virusnih genomov pri skupini Protostomia. Odkrili so, da je raznolikost RNA virusnih genomov pri analiziranih nevretenčarjih bistveno večja, kot je bila prvotno definirana pri členonožcih. Ugotovili so, da je pri evoluciji skupine Protostomia prišlo do »eksplozije« raznolikih RNA viromov, ki je vodila v pestrost prisotnih virusnih klad pri celotni skupini Protostomia. **Raziskave Shi in sod. (2016a) so postavile tri ključna vprašanja:** 1) o izvoru eksplozije raznolikih genomov oz. ali je skupina Protostomia dejanski izvor eksplozije raznolikosti RNA virusnih klad ali se je »Big Bang« raznolikih virusnih genomov odvil že ob koevoluciji najstarejših skupin živali, 2) kakšna je dejanska starost posameznih RNA virusnih klad oz. ali najstarejše predstavnike virusnih klad najdemo pri skupini Protostomia ali najdemo še starejše predstavnike že pri predniku živali in 3) ali je pestra raznolikost RNA viromov omejena na skupino Protostomia ali je ohranjena čez vse nevretenčarje. Za odgovor na ta vprašanja so bile potrebne dodatne analize raznolikosti RNA viroma pri najstarejših skupinah živali in ostalih neraziskanih skupinah nevretenčarjev. Pridobljeni rezultati so nam podali odgovore na hipoteze, postavljene v okviru doktorata:

- 1) Da je raznolikost RNA viroma ohranjena čez celotno skupino nevretenčarjev in ni omejena samo na skupino Protostomia.
- 2) Da so najstarejše skupine živali vir raznolikih živalskih RNA viromov.

V okviru doktorata smo v nasprotju s Shi in sod. (2016a) reprezentativno vzorčili skupine nevretenčarjev, ki so različne evolijske starosti – od najstarejših predstavnikov živali do žuželk. Skupno smo analizirali 453 genomskih in 832 transkriptomskih knjižnic (skupno 1285 knjižnic) in odkrili 815 novih RNA virusov, ki spadajo v 22 od 25 definiranih klad. Tabela 5.1 prikazuje trenutno stanje o prisotnosti in odsotnosti RNA virusnih klad pri posameznih skupinah nevretenčarjev (Tabela 5.1). **Združili smo podatke iz literature z rezultati naših analiz (Shi in sod., 2016a; Shi in sod., 2016b, Waldron in sod., 2017; Hulo in sod., 2011 – ViralZone).** Pri žuželah je bila že prvotno iz literature znana prisotnost klad: Birna, Hepe-Virga, Luteo-Sobemo, Narna-Levi, Bunya-Arena, Mono-Chu, Orthomyxo, Nido, Partiti-Picobirna, Permutotetra, Picorna-Calici, Reo, Tombus-Noda, Toti-Chryso, Qivirus, in Ophio, torej 17/25 klad. **V okviru doktorata smo z analizo RNA viromov pri kožekrilcih, metuljih in najstarejših skupinah živali dokazali tudi prisotnost klad Astro pri metuljih in Ophio pri kožekrilcih, ki prvotno nista bili odkrili pri žuželah, in s tem povečali število prisotnih klad na 19/25.** Pri skupini rakov so bili iz literature znani podatki o prisotnosti 19/25 klad: Astro, Hepe-Virga, Luteo-Sobemo, Narna-Levi, Bunya-Arena, Mono-Chu, Orthomyxo, Nido, Partiti-Picobirna, Permutotetra, Picorna-Calici, Reo, Tombus-Noda, Toti-Chryso, Yuevirus, Qivirus, Zhaovirus, Yanvirus in Flaviviridae. **V okviru doktorata smo z analizo RNA viromov pri deseteronožcih, enakonožcih, postranicah in ceponožcih odkrili prisotnost klad Birna pri ceponožcu in Ophio pri deseteronožcu. S tem smo povečali število prisotnih klad na 21/25.** Pri skupini Spiralia je bila prvotno iz literature znana prisotnost klad: Astro, Birna, Hepe-Virga, Hypo, Luteo-Sobemo, Narna-Levi, Bunya-Arena, Mono-Chu, Orthomyxo, Nido, Partiti-Picobirna, Reo, Picorna-Calici, Tombus-Noda, Toti-Chryso, Qivirus, Zhaovirus, Yanvirus, Weivirus, torej 19/25 klad. **V okviru doktorata smo pri skupini Spiralia odkrili še prisotnost klade Flavi in s tem povečali število na 20/25 virusnih klad.** Pri najstarejših skupinah živali je bila iz literature znana prisotnost 8/25 virusnih klad: Hepe-Virga, Luteo-Sobemo, Narna-Levi, Mono-Chu, Partiti-Picobirna, Picorna-Calici, Tombus-Noda in Weivirus. **V okviru doktorata smo z analizo najstarejših skupin živali**

ugotovili še prisotnost skupin: Birna, Bunya-Arena, Orthomyxo, Reo, Toti-Chryso, Yuevirus, Qinvirus, Ophio in Poty. S tem smo število prisotnih virusnih klad povečali na 17/25. V ločeni analizi nidovirusov pri najstarejših skupinah živali smo našli tudi nidovirus, prisoten pri sladkovodnem trdoživu (rezultati niso prikazani v okviru doktorskega dela), in s tem povečali število prisotnih virusnih klad pri najstarejših skupinah živali na 18/25. S primerjavo naših podatkov s podatki iz literature smo tako ugotovili, da je pri skupini nevretenčarjev prisotnih 24 od 25 definiranih virusnih klad. Samo klada Cysto doslej ni bila odkrita pri nobeni od raziskanih nevretenčarskih skupin. Nevretenčarske skupine različnih evlucijskih starosti torej kažejo na podobno pestrost prisotnih virusnih klad (19 ali 21 od 25 definiranih). S tem smo pokazali, da se je pestrost prisotnih RNA virusnih klad enakomerno ohranila čez celotno skupino nevretenčarjev tekom evolucije. Poudariti pa moramo, da niso vse virusne klade enakomerno razširjene pri vseh skupinah nevretenčarjev. Pri vseh skupinah nevretenčarjev, od najstarejših skupin živali do žuželk, smo našli klade: Hepe-Virga, Hypo, Luteo-Sobemo, Narna-Levi, Bunya-Arena, Mono-Chu, Orthomyxo, Partiti-picobirna, Picorna-Calici, Permutotetra, Tombus-Noda, Toti-Chryso, Qinvirus in Flaviviridae. Pri tem je bila izrazito najbolj zastopana klada Picorna-Calici. Pogoste klade pri nevretenčarjih so še: Mono-Chu, Partiti-Picobirna, Hepe-Virga, Narna-Levi in Tombus-Noda. Vse omenjene klade so bogato prisotne tako pri vodnih kot pri kopenskih nevretenčarjih. Nekoliko redkejšje klade so Flavi, Toti-Chryso, Bunya-Arena, Luteo-Sobemo, Orthomyxo, Reo in Permutotetra. Skupine Flavi, Permutotetra, Reo, Orthomyxo in Toti-Chryso so pogostejše pri kopenskih kot pri vodnih nevretenčarjih, skupini Bunya-Arena in Luteo-Sobemo sta približno enako pogosti tako pri kopenskih kot pri vodnih nevretenčarjih. Redke klade pri nevretenčarjih so: Nido, Astro, Birna, Hypo, Qinvirus, Yanvirus, Weivirus, Zhaovirus, Yuevirus, Ophio ter Poty. Skupine Yanvirus, Zhaovirus, Weivirus, Yuevirus in Poty so omejene samo na vodne nevretenčarje. Kladi Zhaovirus in Yanvirus sta prisotni le pri raki in skupini Spiralia. Klada Yuevirus je omejena na rake in najstarejše skupine živali, klada Weivirus pa je omejena na vodne nevretenčarje, in sicer na skupino Spiralia in na najstarejše skupine živali (korale). Klado Poty smo našli samo pri najstarejših skupinah živali. Z rezultati prisotnosti in odsotnosti RNA virusnih klad smo potrdili, da se je pestra raznolikost RNA virusnih klad ohranila pri vseh skupinah nevretenčarjev (**Tabela 5.1**). Znotraj najpogostejše klade Picorna-Calici smo z združevanjem podatkov iz literature z našimi rezultati ugotovili, da so najpogostejše skupine pikornavirusov vodni pikornavirusi, iflavirusi in dicistrovirusi. Vodni pikornavirusi (angl. »aquatic picornavirales«) so razširjeni izključno pri vodnih nevretenčarjih, kjer so tudi izrazito prevladujoča skupina. Iflavirusi so razširjeni pri vseh skupinah nevretenčarjev, vendar so pogostejši pri kopenskih nevretenčarjih, kjer so tudi prevladujoča skupina (**Dodatek A; A.2, Slika A.4**). Dicistrovirusi so pogosta skupina tako pri kopenskih kot tudi pri vodnih nevretenčarjih. Ostale skupine so redkejšje. Skupini Nora virusu podobnih virusov in Kelp fly virusu podobnih virusov sta razširjeni pri vodnih in pri kopenskih nevretenčarjih, medtem ko sta skupini Posavirusu podobnih virusov in Arivirusu podobnih virusov omejeni samo na vodne nevretenčarje. Poudariti velja, da smo pri najstarejši skupini živali odkrili največjo pestrost prisotnih pikornavirusnih skupin, pravi »Big Bang« pikornavirusnih skupin (7 od 8), ki se je nato ohranila tudi pri ostalih skupinah nevretenčarjev. Evlucijsko najstarejša skupina pikornavirusov naj bi bili Kelp fly virusu podobni virusi, sledijo Posa in Nora virusu podobni virusi, iflavirusi, Arivirusu podobni virusi, Dicistrovirusi, vodni pikornavirusi pa se kažejo kot najmlajša skupina. Prisotnost praktično vseh skupin pikornavirusov pri najstarejših skupinah živali torej kaže na »eksplozijo« pikornavirusnih skupin že pri predniku živali, ki se je ohranila čez vse nevretenčarje.

Tabela 5.1: Prisotnost in odsotnost virusnih klad pri žuželkah, raki, skupini Spiralia in najstarejših skupinah živali. Združili smo podatke iz literature z rezultati naših analiz. Črna pika – prisotnost znana že iz literature, kljukica – prisotnost, ki je bila na novo ugotovljena v okviru naših raziskav, bela pika – odsotnost virusne klade.

VIRUSNA KLADA	INSECTA	CRUSTACEA	SPIRALIA	NAJSTAREJŠE SKUPINE ŽIVALI	NEVRETEČARJI
ASTRO	✓	•	•	○	•
BIRNA	•	✓	•	✓	•
HEPE-VIRGA	•	•	•	•	•
HYP0	•	○	•	○	•
LUTEO-SOBEMO	•	•	•	•	•
NARNA-LEVI	•	•	•	•	•
BUNYA-ARENA	•	•	•	✓	•
MONO-CHU	•	•	•	•	•
ORTHOMYXO	•	•	•	✓	•
NIDO	•	•	•	○	•
PARTITI- PICOBIRNA	•	•	•	•	•
PERMUTOTETRA	•	•	○	○	•
PICORNA-CALICI	•	•	•	•	•
REO	•	•	•	✓	•
TOMBUS-NODA	•	•	•	•	•
TOTI-CHRYSO	•	•	•	✓	•
YUEVIRUS	○	•	○	✓	•
QINVIRUS	•	•	•	✓	•
ZHAOVIRUS	○	•	•	○	•
YANVIRUS	○	•	•	○	•
WEIVIRUS	○	○	•	•	•
OPHIO	✓	✓	○	✓	•
FLAVI	•	•	✓	○	•
POTY	○	○	○	✓	✓
CYSTO	○	○	○	○	○
VSOTA	19/25	21/25	20/25	17/25	24/25

Pestrost prisotnih RNA virusnih klad se je torej ohranila pri vseh skupinah nevretenčarjev. Ugotovitev še dodatno podprejo nadaljnje ugotovitve raznolikosti genomskih struktur. **Tako pri kladi Picorna-Calici kot tudi pri ostalih RNA virusnih kladah najdemo širok nabor raznolikih genomskih struktur že pri najstarejših skupinah živali, ki se ohrani tudi pri ostalih skupinah nevretenčarjev** – skupini Spiralia, raki in žuželkah. Pri nobeni od navedenih skupin nevretenčarjev znotraj nobene virusne klade ni viden upad v pestrosti virusnih klad. Najlepše pa je pojav viden pri kladi Picorna-Calici. Pri tej kladi se že pri najstarejših skupinah živali pojavi širok nabor genomskih struktur z različnim številom odprtih bralnih okvirjev in različnim zaporedjem prisotnih proteinskih domen, ki se ohrani tudi pri evolucijsko mlajših skupinah živali. **Tako smo dokazali, da se je raznolikost genomskih struktur in s tem raznolikost RNA virusov ohranila pri vseh skupinah nevretenčarjev in ni omejena samo na skupino Protostomia.**

Z ugotovitvijo o dejanski starosti virusnih klad in o tem, kje je dejanski vir raznolikosti RNA virusov, smo rezultate o raznolikosti genomskih struktur združili s filogenetsko analizo. **Rezultati filogenetske analize kažejo, da smo pri evolucijsko starejših skupinah nevretenčarjev odkrili bolj divergenten nabor raznolikih virusnih klad. Najbolj divergentne**

predstavnik smo odkrili pri najstarejših skupinah živali. Hkrati smo pri najstarejših skupinah živali odkrili najstarejše predstavnike virusnih klad. S tem smo postavili izvor raznolikih genomskih struktur RNA virusov v časovni kontekst. Ugotovili smo, da je dejanska starost virusnih klad bistveno starejša, kot je bilo prvotno nakazano. Prvi predstavniki virusnih klad se namreč pojavijo že pri predniku živali. Ker so se pester genomske strukture pri vseh virusnih kladah pojavile že pri najstarejših skupinah živali, lahko torej sklepamo, da se je eksplozija genomskih struktur pri vseh virusnih kladah odvila že pri predniku živali in se nato ohranila tekom evolucije živali. **Pestrost RNA virusnih genomov torej izvira iz RNA viroma prednika živali in se je ohranila tekom evolucije živali. Najstarejše skupine živali so torej izvor in center raznolikosti RNA virusov, ki se je ohranila čez celotno skupino nevretenčarjev.**

Potrdili smo torej vse naše hipoteze:

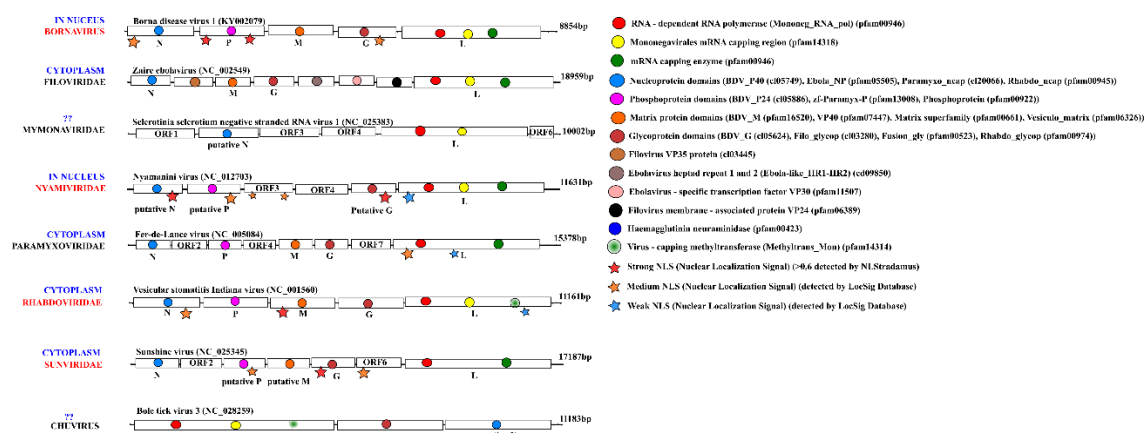
- 1) V transkriptomih in genomih nevretenčarjev smo odkrili bogato zbirko raznolikih RNA virusnih genomov.
- 2) Najstarejše skupine živali so center in izvor raznolikosti RNA virusov.
- 3) Pestra raznolikost RNA virusov se je ohranila med evolucijo pri vseh nevretenčarjih.

5.8 Endogenizirani RNA virusi v genomih nevretenčarjev

Kljub ugotovitvam o pestrosti RNA viroma pri nevretenčarjih pa zaradi narave transkriptomskih knjižnic ne moremo sklepati o dejanskih gostiteljih virusnih klad. Virus je bil v trenutku vzorčenja lahko prisoten 1) znotraj celice gostitelja, 2) v medcelični ali pa 3) na površini organizma (Crochu in sod., 2004; Holmes, 2011). Da bi izvor pestrosti RNA virusnih klad lahko povezali z izmenjavo genomskega materiala med virusi in prehodom virusa med elemente RNA viroma, potrebujemo podatke o dejanskih gostiteljih virusov. Prisotnost virusnih zaporedij v genomu gostitelja pomeni, da se je RNA virus med evolucijo integriral v genom gostitelja, kar pomeni, da je virus aktivno okuževal gostitelja. Prisotnost virusnega zaporedja v genomu določenega organizma nam torej pove, da je organizem gostitelj RNA virusa. Endogenizirani RNA virusi so fosili, ki kažejo na pretekle okužbe (Crochu in sod., 2004; Holmes, 2011). Naša analiza nevretenčarskih genomov je pokazala prisotnost številnih endogeniziranih zaporedij. Seveda pa so redka v primerjavi s številom in raznolikostjo RNA virusov, ki smo jih odkrili v transkriptomih. Endogenizacija je namreč redek pojav. Da virusno zaporedje postane endogenizirano, mora: 1) preiti skozi membrano celice, 2) potovati do jedra, 3) preiti jedrno bariero in 4) izkoristiti gostiteljev biokemijski aparat za integracijo v gostitelja (Crochu in sod., 2004; Holmes, 2011; Katzourakis in Gifford, 2010). Retrovirusi že kodirajo biokemijski aparat in integracijo v genom izkoriščajo v okviru mehanizma podvojevanja. RNA virusi pa ne kodirajo biokemijskega aparata za integracijo v gostiteljski genom, hkrati pri večini primerov njihov mehanizem podvojevanja ni povezan z jedrom. Zato je pojav pri večini RNA virusnih klad redek. Vseeno smo opazili pogostejšo integracijo v genom pri določenih virusnih kladah.

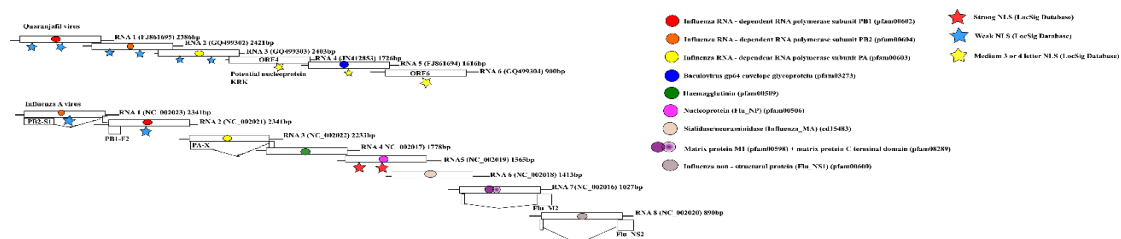
Klado **Mono-Chu** smo našli endogenizirano pri vseh raziskanih skupinah nevretenčarjev, in sicer: dva pri skupini Spiralia, enega pri najstarejših skupinah insektov, enega pri najstarejših skupinah živali, sedem pri metuljih, petnajst pri kožekrilcih in pet pri raki, skupno torej 31. Hkrati je Mono-Chu virusna klada, pri kateri smo našli največje število endogeniziranih RNA virusov. Nekatere virusne skupine znotraj Mono-Chu imajo mehanizem razmnoževanja vezan na jedro, kot sta Bornavirus in Nyamiviridae (**Slika 5.1**). Za omenjene skupine je verjetnost za endogenizacijo virusnega genoma večja kot pri skupinah klade Mono-Chu, ki imajo mehanizem razmnoževanja povezan s citoplazmo, kot so Filoviridae, Paramyxoviridae, Rhabdoviridae in Sunviridae. Kot že omenjeno, genom retrovirusov (kot je npr. virus HIV) v jedro vstopi v obliki predintegracijskega kompleksa (PIC, angl. »*preintegration complex*«), v katerem je genom virusa vezan z glavnimi proteini: integrazo, reverzno transkriptazo, proteinom matriksa (angl. »*matrix protein*«) in proteinom Vpr. Ključno vlogo pri vstopu predintegracijskega kompleksa v jedro imajo jedrni lokalizacijski signali na proteinih, vezanih z genomom retrovirusa (Sherman in Greene, 2002). Zato smo tudi za klado Mono-Chu preko programa NLStradamus in LocSig Database (LocSigDb) na vseh kodiranih proteinih poiskali morebitno prisotnost jedrnih lokalizacijskih signalov (NLS, angl. »*Nuclear Localization Signals*«) (Negi in sod., 2015; Nguyen in sod., 2009). Jedrne lokalizacijske signale smo odkrili s programom NLStradamus, v katerem je prednastavljena meja predikcije 0,6. Mejo predikcije smo poskusili višati do

0,9 in določili, če nam pri višanju še vedno zazna isti NLS. V določenih primerih smo NLS našli samo z LocSigDB. Slika 103 prikazuje mesto najdenih NLS. Rdeče zvezdice prikazujejo močne jedrne lokalizacijske signale z vrednostjo predikcije nad 0,6 v programu NLStradamus (Slika 5.1). Oranžne in modre zvezdice označujejo srednje močne in šibke NLS, ki so bili zaznani s programom LocSig Database. Rdeče obarvane virusne skupine: Bornaviridae, Nyamiviridae, Rhabdoviridae in Sunviridae vsebujejo močne jedrne lokalizacijske signale. Vsem virusnim skupinam klade Mono-Chu je skupna ribonukleokapsida – virusni genom v kompleksu z RdRp, fosfoproteinom in nukleoproteinom. Ribonukleokapsido obdaja virusna ovojnica iz matriks proteina (angl. »*matrix protein*«) in glikoproteina. Skupini Bornaviridae in Nyamiviridae imata mehanizem razmnoževanja vezan na jedro. V obeh primerih smo našli jedrne lokalizacijske signale na nukleoproteinih, ki so direktno vezani na genom. Pri skupini Bornaviridae smo našli jedrni lokalizacijski signal tudi na fosfoproteinu, ki je ravno tako vezan na virusni genom, kar kaže na njegovo potencialno vlogo v kompleksu z virusnim genomom pri prehodu skozi jedro. Virusi, ki nimajo mehanizma razmnoževanja povezanega z jedrom, lahko v jedro pridejo z 1) mehanizmi, ki destabilizirajo NPC (angl. »*Nuclear Pore Complexes*«), 2) če imajo dovolj majhen premer, lahko celi pridejo skozi jedrno poro, 3) okužujejo lahko samo celice, ki se delijo in imajo NPC nestabilen, 4) proteini kapside lahko vsebujejo NLS (angl. »*Nuclear Localization Signals*«), ki igra vlogo pri prehodu virusa skozi jedrne pore (Cohen in sod., 2011). Rhabdoviridae in Sunviridae pridejo v celico z receptorsko posredovano endocitozo, ki ji sledi odstranjevanje virusne ovojnice iz matriks proteina (angl. »*matrix protein*«) in glikoproteina. V citoplazmo preide ribonukleokapsida – kompleks virusnega genoma z RdRp, fosfoproteinom in nukleoproteinom. Sledi pomnoževanje virusne RNA v citoplazmi. Kljub temu, da nimata mehanizma podvojevanja, vezanega direktno z jedrom, smo našli močne jedrne lokalizacijske signale na matriks proteinu (Rhabdoviridae), glikoproteinu (Sunviridae) in zmrno močne na nukleoproteinu (Rhabdoviridae) in fosfoproteinu (Sunviridae). Verjetno lahko ribonukleokapsida v določenih primerih s pomočjo citoskeleta potuje do NPC, kjer NLS na nukleoproteinu in fosfoproteinu sodelujejo pri vstopu ribonukleokapside v jedro.



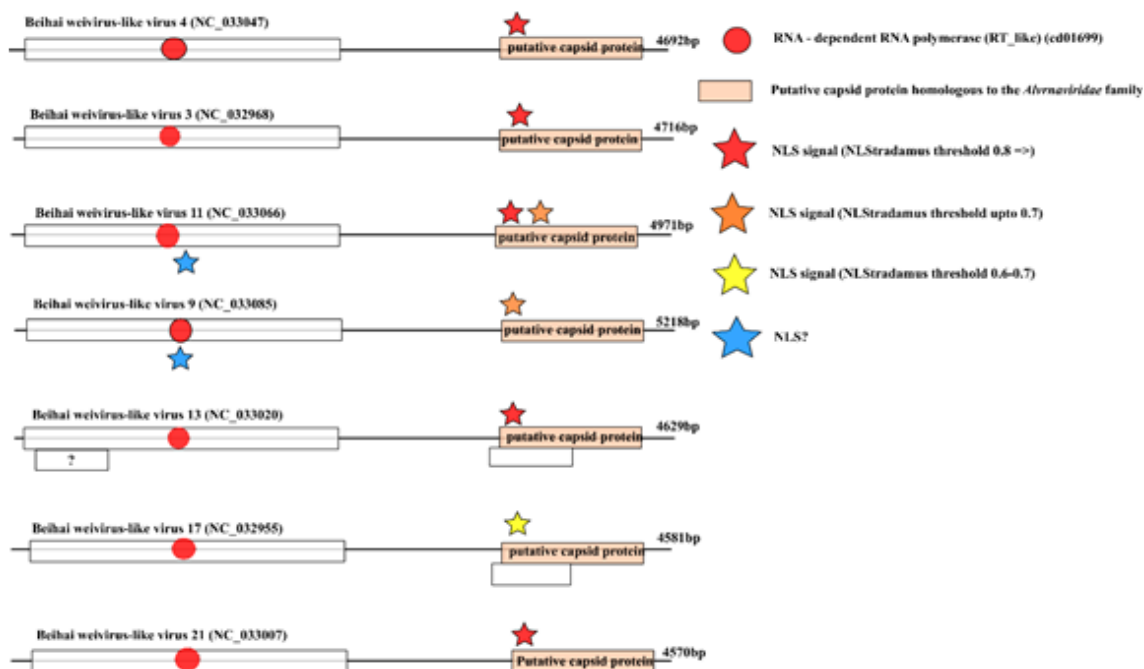
Slika 5.1: Jedrni lokalizacijski signali pri virusni kladi Mono-Chu.

Virusna klada **Orthomyxoviridae** ima najboljše raziskan mehanizem podvojevanja, ki je povezan z jedrom (Greber, 2016). Po prehodu virusa v citoplazmo z receptorsko posredovano endocitozo in odstranjevanjem virusnega plašča ribonukleokapsida potuje do jedra s pomočjo citoskeleta, kjer jedrni lokalizacijski signali na nukleoproteinih igrajo ključno vlogo pri prepoznavi s strani importinov in vstopu virusnega genoma v jedro gostiteljske celice (Slika 5.2). Našli smo endogeniziran po en virus klade Orthomyxoviridae pri skupini Spiralia, najstarejših skupinah žuželk in metuljih ter tri pri kožekrilcih.



Slika 5.2: Jedrni lokalizacijski signali pri skupini Orhtomyxoviridae.

Klada Weivirus ima genom zgrajen iz dveh odprtih bralnih okvirjev (ORF-jev). ORF1 je nestrukturni in kodira zapis za RdRp, ORF2 je strukturni ORF, ki kodira zapis za protein kapside, ki je homologen s kapsidnim proteinom skupine Alvernaviridae. Našli smo relativno močne NLS na N-terminalnem delu kapsidnega proteina pri vseh analiziranih primerih weivirusov s programom NLStradamus (Slika 5.3). Verjetno kapsidni protein igra vlogo pri prepoznavi s strani importinov in prehodu virusa v jedro. Našli smo namreč 12 endogeniziranih weivirusov pri organizmih iz rodu *Symbiodinium* (najstarejše skupine evkariontov, Alveolata). Organizem je znan simbiot koral, ki sodijo med ožigalkarje. Glede na naravo odnosa sožitja je težko povezati, ali so virusi, najdeni v genomu *Symbiodinium*, dejansko povezani z njegovim genomom ali so bolj povezani z genomom ožigalkarjev, s katerimi so v sožitju (Slika 4.24).



Slika 5.3: Jedrni lokalizacijski signali pri virusni kladi Weivirus.

Klada **Flaviviridae** je zanimiva, ker je edina RNA virusna klada, kjer smo našli ohranjene celotne endogenizirane genome pri rakih (Crustacea) in kožekrilcih (Hymenoptera). Presenetljivo smo našli relativno veliko število endogeniziranih genomov pri dsRNA virusnih kladah Partiti-Picobirna in Toti-Chryso. Zanje je značilno, da genom pride v stik s citoplazmo gostiteljske celice po prepisu dsRNA v mRNA z virusno RdRp znotraj kapside virusa, ki se nahaja v citoplazmi. mRNA namreč po prepisu potuje izven genoma v citoplazmo gostiteljske celice. Pri kladi **Partiti-Picobirna** smo našli en endogeniziran virus pri skupini Spiralia, štiri pri metuljih in 14 pri kožekrilcih. Pri kladi Toti-Chryso smo našli en endogeniziran virus pri metuljih, šest pri kožekrilcih in tri pri rakih.

Poglavje 6

Zaključki

Nevretenčarji predstavljajo eno najbolj dinamičnih ekoloških niš za RNA viruse. Pojavila se je hipoteza, da so nevretenčarji center raznolikosti in evolucije RNA virusov, vendar so prvotne raziskave pokazale le delno sliko o raznolikosti in razširjenosti RNA virusov pri nevretenčarjih.

Skupina Shi in sod. (2016a) je izvedla eno najobsežnejših študij raznolikosti RNA virusov. Raziskala je raznolikost in razširjenost RNA virusov, prisotnih v 87 knjižnicah, in postavila skupino Protostomia kot center raznolikosti RNA virusov. Manjkajo pa podatki o raznolikosti in razširjenosti virusov v 1) žuželkah reda Hymenoptera (kožekrilci) in Lepidoptera (metulji in vešče), 2) najstarejših skupinah žuželk, 3) rakih (Crustacea), predvsem v postranicah (Amphipoda), enakonožcih (Isopoda) in ceponožcih (Copepoda), 4) skupini Spiralia, predvsem v ploščatih črvih (Platyhelminthes) in 5) najstarejših skupinah živali (Metazoa), v spužvah (Porifera), ožigalkarjih (Cnidaria) in rebračah (Ctenophora). Da bi dobili bolj kompletno sliko o raznolikosti in razširjenosti RNA virusov, smo analizirali raznolikost in razširjenost RNA virusov pri omenjenih skupinah. Raznolikost in razširjenost RNA virusov smo raziskali z 1) iskanjem novih RNA virusov in endogeniziranih RNA virusov v transkriptomskih in genomskih podatkovnih bazah nevretenčarjev, 2) filogenetsko analizo novih in endogeniziranih RNA virusov, 3) primerjavo genomskih struktur novih in endogeniziranih RNA virusov glede na trenutno znane klasificirane in neklasificirane viruse iz iste skupine ter 4) statistično analizo prisotnosti posameznih virusnih klad pri raziskanih skupinah nevretenčarjev.

V okviru filogenetske analize RNA virusov smo prišli do naslednjih zaključkov:

Pri kožekrilcih (Hymenoptera) smo našli posamezne primere horizontalnih prenosov virusov s parazitoidnih os na njihove gostitelje (klada Picorna-Calici, Tombus-Noda), z rastlin na njihove vektorje (klada Hepe-Virga, Partiti-Picobirna, Picorna-Calici) ali z gliv na domnevne vektorje (klada Partiti-Picobirna). Našli smo namreč 1) iz žuželk izolirane viruse, ki so sorodni (ali skoraj identični) virusom, ki so prisotni v transkriptomih parazitoidnih os, 2) rastlinskim virusom sorodne viruse v transkriptomih kožekrilcev in 3) glivnim virusom sorodne viruse v transkriptomih kožekrilcev. V primeru prenosa rastlinskih virusov na kožekrilce sta možna dva scenarija. Virus je lahko naključno prisoten na kožekrilcu in je dejansko povezan z rastlino, s katero se hrani, ali pa je omenjeni kožekrilce vektor rastlinskega virusa. V primeru prenosa glivnih virusov na kožekrilce so možni trije scenariji. Virus je lahko naključno prisoten na organizmu. Druga možnost je, da je gliva rastlinski patogen in je virus dejansko povezan z rastlino, ki jo gliva okužuje. Tretja možnost je, da je kožekrilce vektor glivnih virusov in hkrati glavni akter za prenos glivnih virusov na rastlinske gostitelje (**Slika 4.7, Slika 4.9, Dodatek B; B.1 - Slika B.9 in Slika B.5**)

Pri metuljih in veščah (Lepidoptera) smo v njihovih transkriptomih našli primere rastlinskim virusom sorodnih virusov (klada Hepe-Virga, Partiti-Picobirna). Virus iz *Micropteryx calthella* (GACP01000900), ki spada v klado Partiti-Picobirna, je 81 % identičen rastlinskemu »Cannabis cryptic virus«. Pri kladi Partiti-Picobirna je znana ključna vloga horizontalnega prenosa virusov z glivnih na rastlinske gostitelje. Metulji so lahko v tem primeru delovali kot vektorji. V transkriptomu metuljev smo našli tudi primer zelo divergentnega reovirusa, ki je 27 % identičen ostalim trenutno znanim reovirusom (**Slika 4.11 in Dodatek B, B.2 - Slika B.23**).

Pri najstarejših skupinah žuželk smo našli več primerov divergentnih virusov, ki so v rangi 20–40 % podobni kateremu koli trenutno znanemu virusu. Negevirusa iz *Anurida maritima* (GAUE02021637) in iz *Sminthurus viridis* (GATZ02022882) sta primera starejših predstavnikov skupine Negevirusov (**Dodatek B; B.3, Slika B.28**).

Tudi **pri rakih (Crustacea)** smo našli več divergentnih virusov (klada Bunya-Arena, Narna-Levi, Partiti-Picobirna). Posebej velja izpostaviti klado Picorna-Calici, kjer smo našli v transkriptomih rakov primere virusov, ki tvorijo novo skupino, kar pomeni, da se povežejo skupaj s katerim koli trenutno znanim pikornavirusom. Našli smo tudi bistveno večje število iflavirusov v transkriptomih rakov kot skupina Shi in sod. (2016a). Našli smo tudi starejše predstavnike pikornavirusov, kot je na primer *Procambarus clarkii* (GBEV01028478) (**Dodatek B; B.4, Slika B.41**).

Pri skupini Spiralia smo našli več divergentnih virusov (klada Astro, Birna, klada Hepe-Virga, Orthomyxo, Picorna-Calici) in primere virusov, ki znotraj virusne skupine tvorijo novo podskupino (Negevirusi znotraj klade Hepe-Virga, klada Partiti-Picobirna). Poleg omenjenega smo pri kladi Mono-Chu našli enega izmed najstarejših predstavnikov virusne skupine mononegavirusov pri *Aplysia californica* (GBDA01107884+GBDA01106473) (**Dodatek B; B.5, Slika B.44, Slika B.45, Slika B.49, B.51, B.57, Slika B.53 in Slika 4.21**).

Pri najstarejših skupinah živali (Metazoa) smo našli največ divergentnih virusov, ki so le 20–40 % podobni kateremu koli trenutno znanemu virusu iz iste virusne klade – glede na proteinsko zaporedje RdRp. Našli smo tudi primere virusov, ki znotraj virusne skupine tvorijo novo podskupino, kar pomeni, da se ne povežejo skupaj s katerim koli od trenutno znanih virusov (klada Picorna-Calici). Poleg omenjenega smo pri najstarejši skupini živali našli največ najstarejših predstavnikov virusnih klad (klada Birna, klada Mono-Chu, klada Orthomyxo, klada Picorna-Calici). Klado Yuevirus smo našli tudi pri najstarejših skupinah živali. Posebej velja izpostaviti klado Weivirus, kjer smo našli weiviruse pri koralah, ki oblikujejo novo podskupino. Ravno tako smo našli precej endogeniziranih weivirusov v alveolatu *Symbiodinium* (simbiontu koral), ki tvorijo dve novi podskupini. Shi in sod. (2016a) so weiviruse odkrili izključno pri mehkužcih, yueviruse pa izključno pri rakih (**Slika 4.22, Slika 4.23, Slika 4.24 in Dodatek B; B.6, Slika B.64 in Slika B.66**).

Prisotnost najbolj divergentnih in hkrati najstarejših predstavnikov RNA virusov pri najstarejši skupini živali kaže na ugotovitev, da so virusne klade evolucijsko starejše, kot je bilo prvotno nakazano. Prvi predstavniki virusnih klad so se namreč pojavili pri najstarejši skupini živali.

V okviru **primerjave genomske strukture** smo prišli do naslednjih ugotovitev:

Večina novih RNA virusov, odkritih v okviru naše raziskave, ima genomsko strukturo virusov, ki so jo odkrili Shi in sod. (2016a). Pojavijo pa se posamezni primeri virusov z drugimi dodatnimi domenami: 1) »RING finger« domena, tripsinu podobna serinska proteaza in peptidaza C39, dodatne Rhv domene pri kladi Picorna-Calici in »Reticulocyte-binding rhoptry protein« pri kladi Flavi. Izpostaviti velja tudi primer najdaljšega najdenega mononegavirusa *Hydra vulgaris* (GEVZ01001862) z dolžino genoma 21 kb. Pikornavirus iz koral *Millepora alcicornis* (GFAS01286708) ima divergentno strukturo, saj kodira RdRp na ločenem ORF v drugem bralnem okvirju, ki se nahaja znotraj glavnega ORF, ki kodira poliprotein (**Dodatek C; Dodatki C.1-6**).

Ugotovili smo, da se je raznolikost genomskih struktur pojavila pri najstarejših skupinah živali in ohranila pri vseh nevretenčarjih. Glede na to, da so se najstarejši predstavniki virusnih klad pojavili že pri predniku živali, lahko sklepamo, da raznolikost genomov virusov izvira iz prednika živali. Najstarejše skupine živali so torej center in izvor raznolikosti RNA virusov.

V okviru **analize endogeniziranih virusov** smo prišli do naslednjih ugotovitev:

V genomih analiziranih skupin nevretenčarjev smo našli endogenizirane predstavnike 12 virusnih klad. Klada Mono-Chu je endogenizirana pri vseh analiziranih skupinah nevretenčarjev. Našli smo presenetljivo veliko endogeniziranih partitivivirusov, ki se načeloma razmnožujejo v citoplazmi znotraj lastne kapside. Našli smo tudi 12 endogeniziranih weivirusov v simbiontu koral *Symbiodinium*. Vendar ne moremo trditi, da so dejansko povezani z alveolatom. Lahko so povezani s koralami, saj je organizem simbiont koral.

V okviru **analize razširjenosti virusnih klad** smo prišli do naslednjih ugotovitev:

Našli smo 815 virusov, ki spadajo v 22 od 25 virusnih klad. Klade Mono-Chu, Partiti-Picobirna in Picorna-Calici so bile najbolj zastopane virusne klade pri analiziranih skupinah nevretenčarjev. Klade Hepe-Virga, Narna-Levi, Mono-Chu, Orthomyxo, Partiti-Picobirna, Picorna-Calici, Tombus-Noda in Toti-Chryso so prisotne pri vseh analiziranih skupinah nevretenčarjev (**Slike 4.1-4.5**). Znotraj klade Picorna-Calici so bili iflavirusi prisotni pri vseh analiziranih skupinah nevretenčarjev, vendar so bili bolj zastopani pri kopenskih nevretenčarjih. Skupina »aquatic picornavirales« je bila zastopana izključno pri vodnih nevretenčarjih. Skupina Dicistroviridae je bila bogato zastopana tako pri kopenskih kot pri vodnih nevretenčarjih. Skupina Nora virusu podobnih virusov je bila pogosteje zastopana pri kopenskih kot pri vodnih nevretenčarjih, obratno pa so bili Posavirusu in Arivirusu podobni pikornavirusi bolj pogosti pri vodnih nevretenčarjih (**Slika 4.6**). **S primerjavo naše analize s podatki iz literature smo ugotovili, da je pri nevretenčarjih prisotnih 24 od 25 definiranih RNA virusnih klad (Tabela 5.1). Samo klada Cysto ni prisotna pri nobeni skupini nevretenčarjev. Pestrost prisotnih virusnih klad se je ohranila pri vseh nevretenčarjih.**

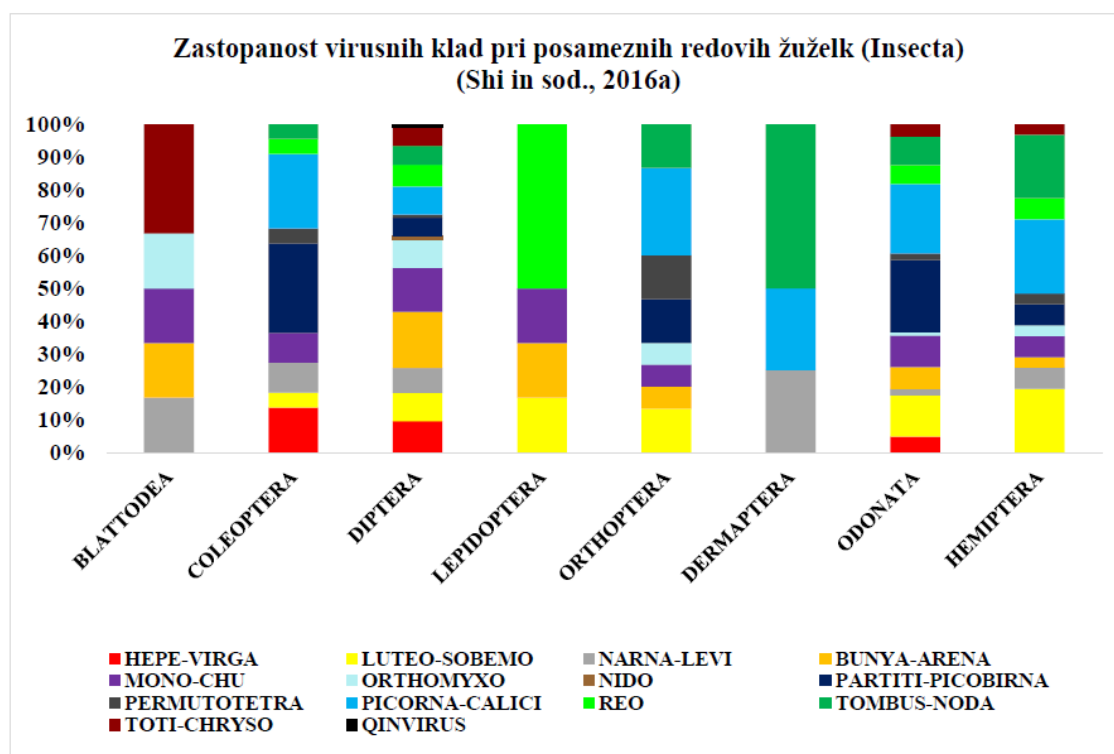
Potrdili smo vse naše hipoteze:

- 1) V genomih in transkriptomih nevretenčarjev je prisotna bogata zaloga RNA virusnih zaporedij.
- 2) Najstarejše skupine živali predstavljajo center in izvor raznolikosti RNA virusov.
- 3) Raznolikost RNA virusov je ohranjena čez celotno skupino nevretenčarjev.

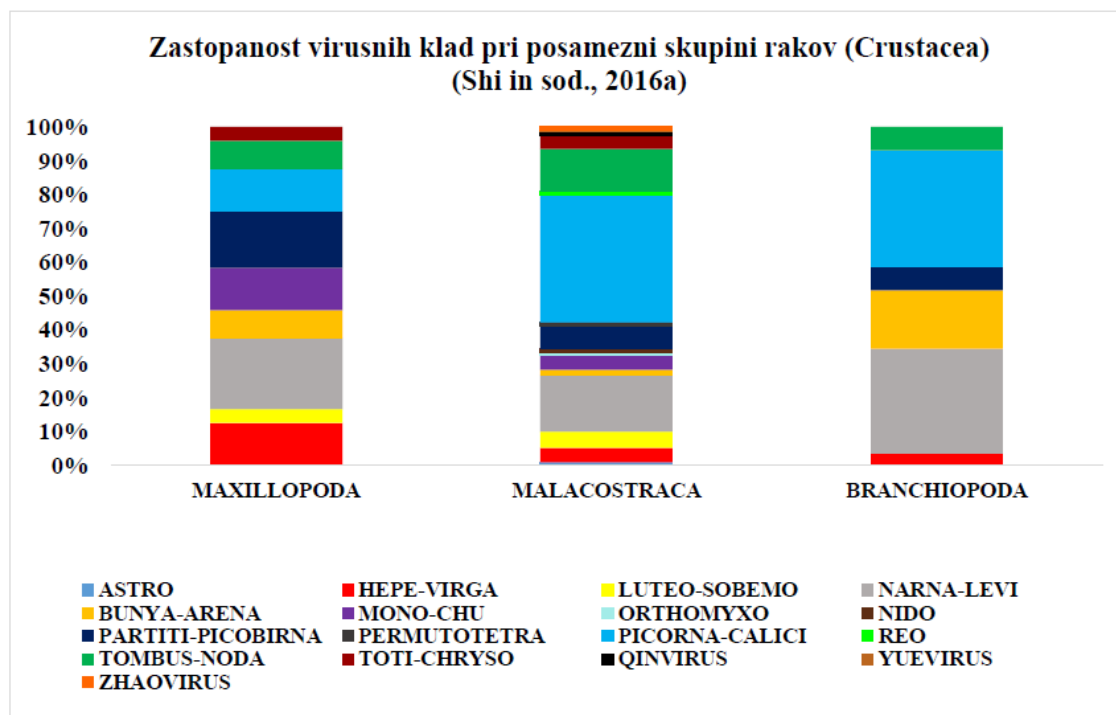
Dodatek A

Analiza zastopanosti virusnih klad

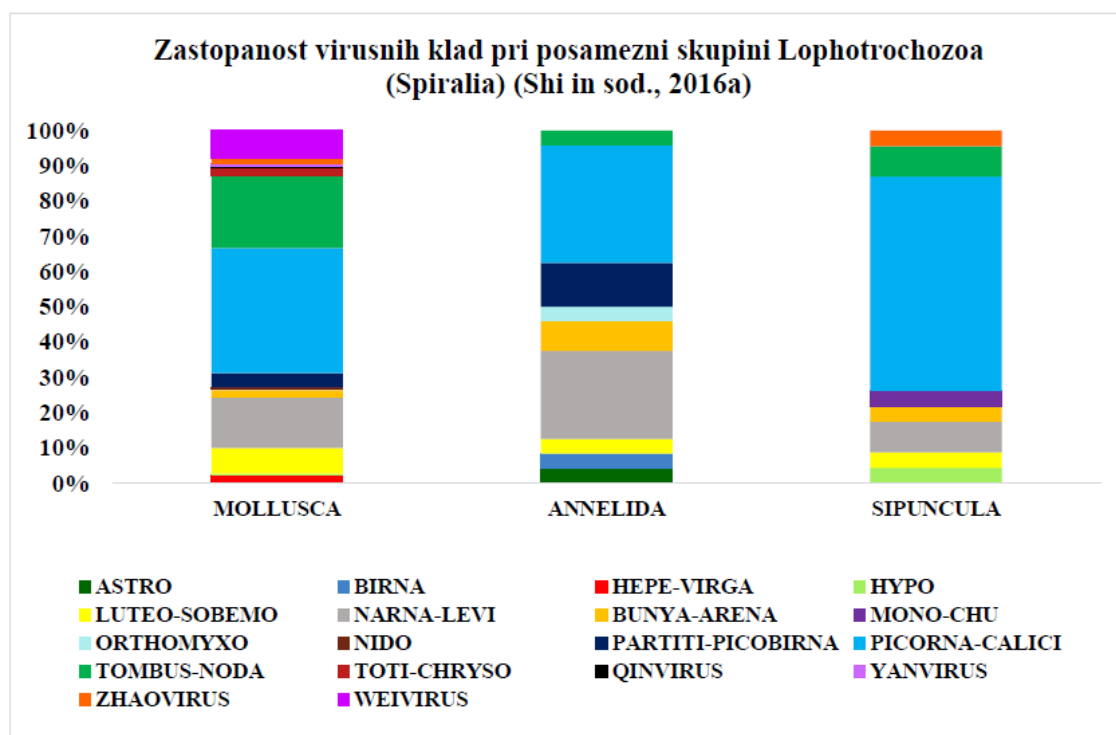
A.1 Analiza zastopanosti posameznih virusnih klad na podlagi rezultatov Shi in sod. (2016a)



Slika A.1: Zastopanost virusnih klad pri posameznih redovih žuželk (Insecta) (Shi in sod., 2016a).

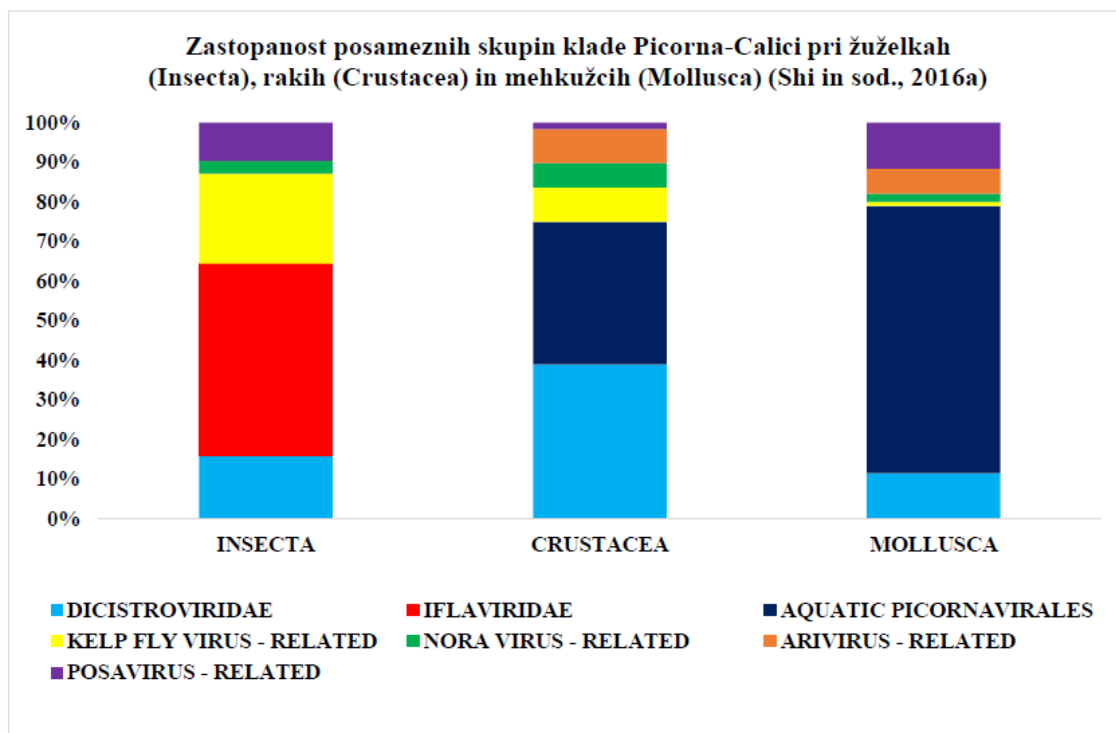


Slika A.2: Zastopanost virusnih klad pri posamezni skupini rakov (Crustacea) (Shi in sod., 2016a).



Slika A.3: Zastopanost virusnih klad pri posamezni skupini Spiralia (Lophotrochozoa) (Shi in sod., 2016a).

A.2 Analiza zastopanosti virusnih skupin znotraj klade Picorna-Calici na podlagi rezultatov Shi in sod. (2016a)

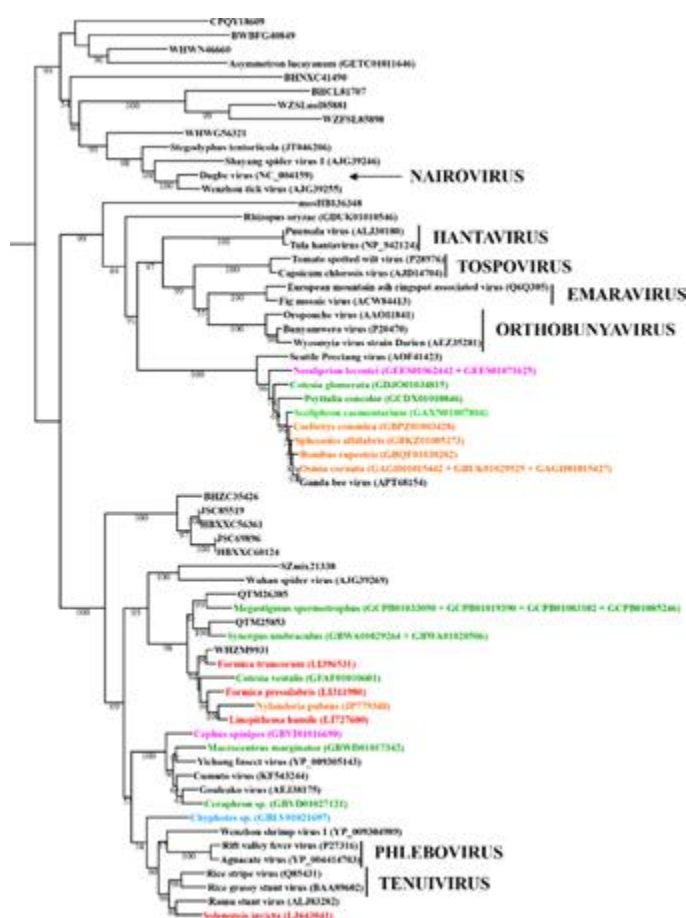


Slika A.4: Zastopanost posameznih virusnih skupin klade Picorna-Calici pri žuželkah (Insecta), rakih (Crustacea) in mehkužcih (Mollusca) (Shi in sod., 2016a). Angleški izraz: related – podoben, npr. posavirus-related; posavirusu podobni virusi.

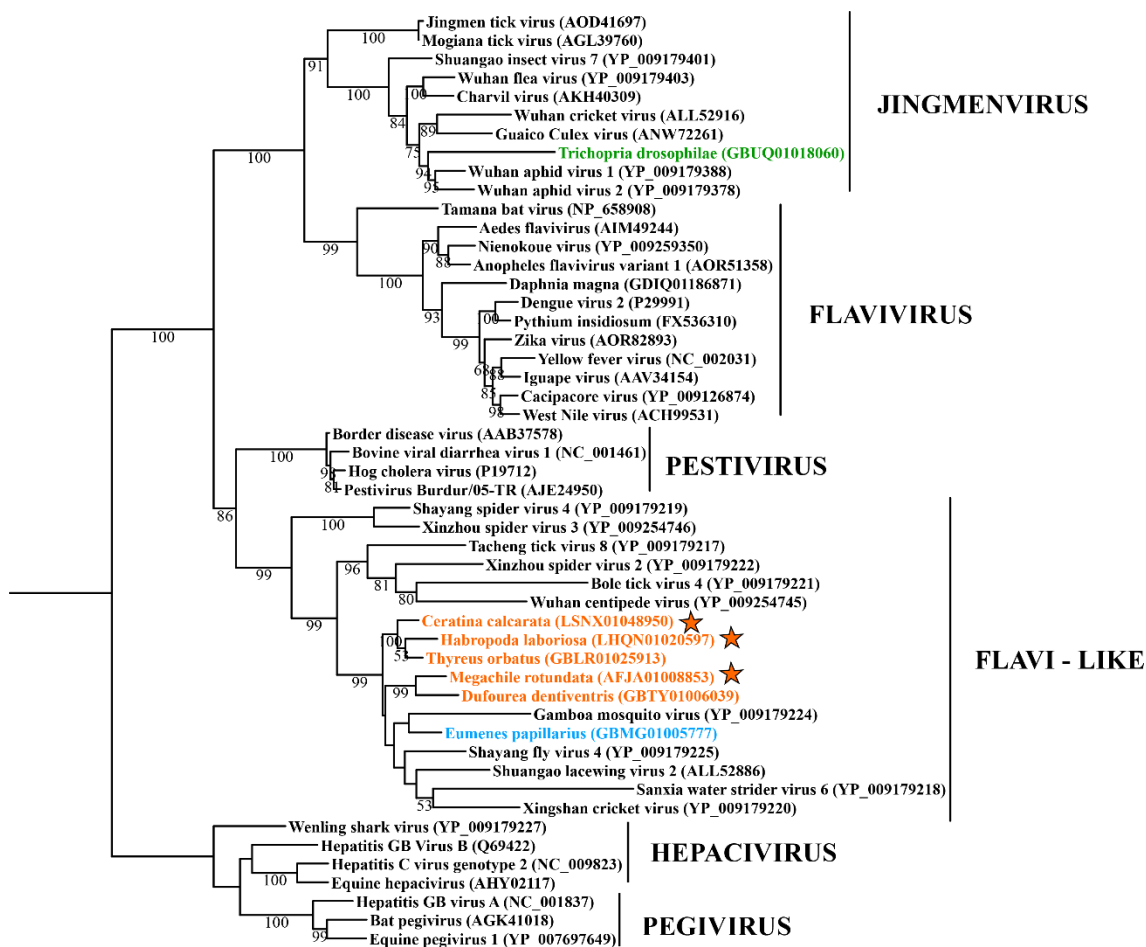
Dodatek B

Filogenetska analiza

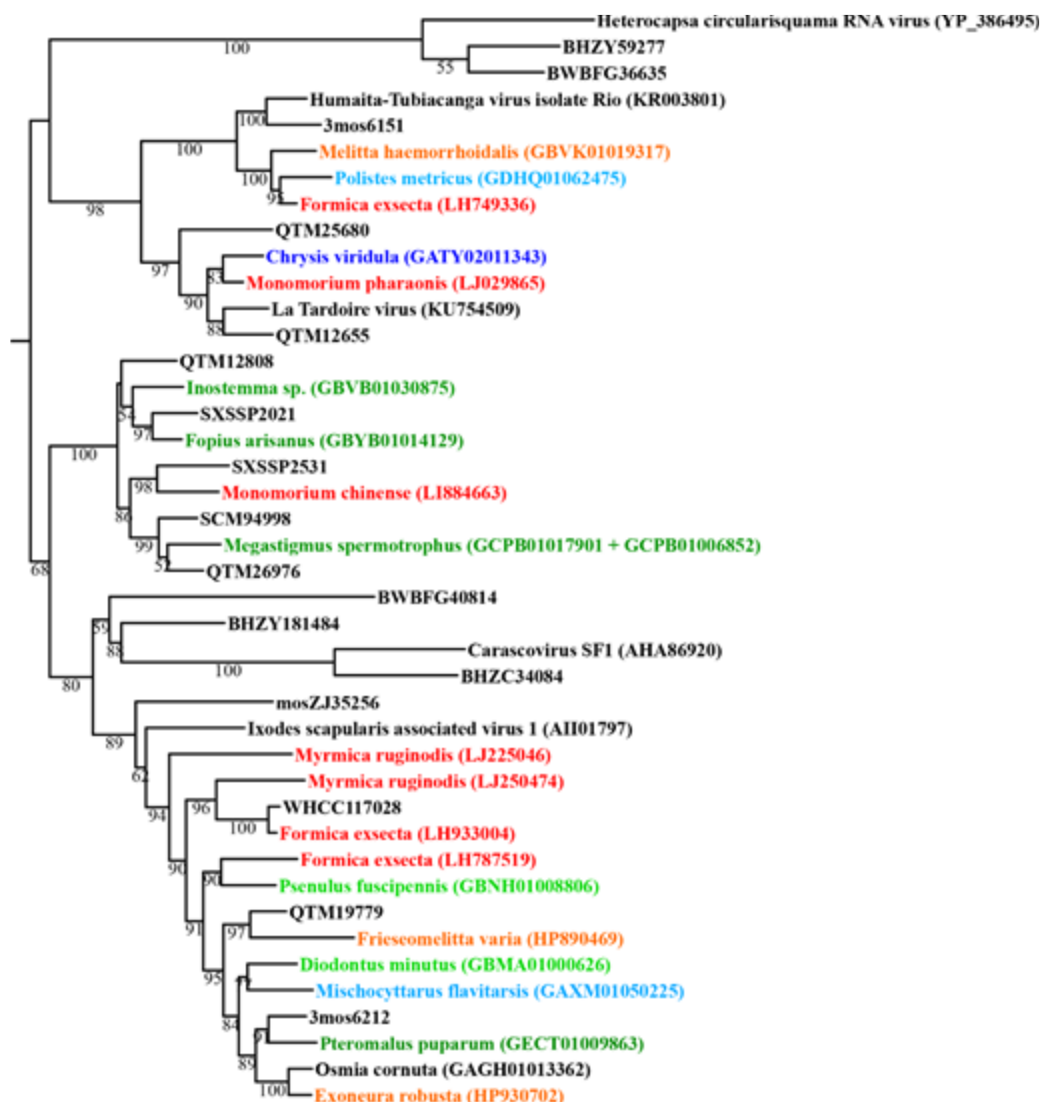
B.1 Kožekrilci (Hymenoptera)



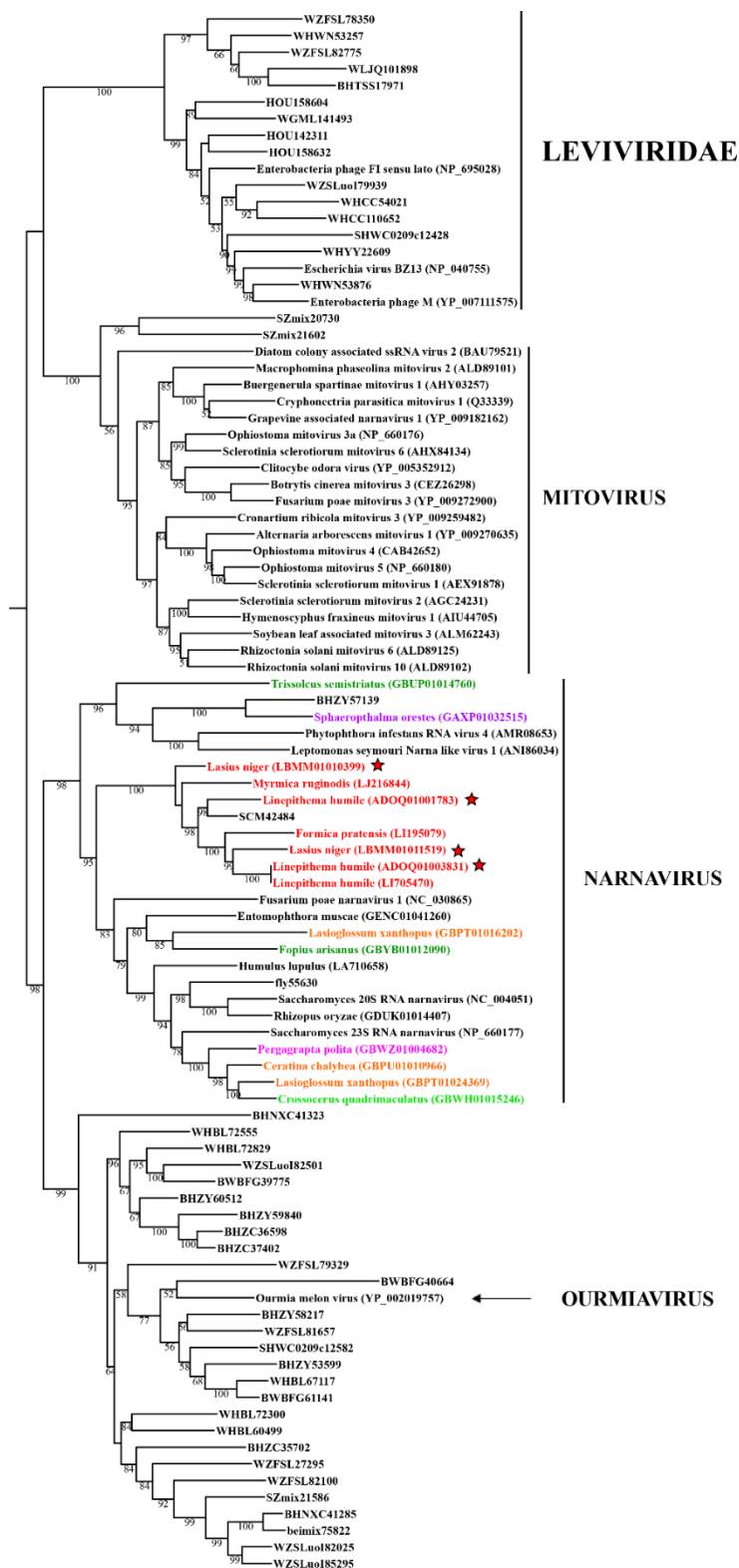
Slika B.1: Filogenetsko drevo virusov klade **Bunya-Arena**, ki smo jih našli v transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro.



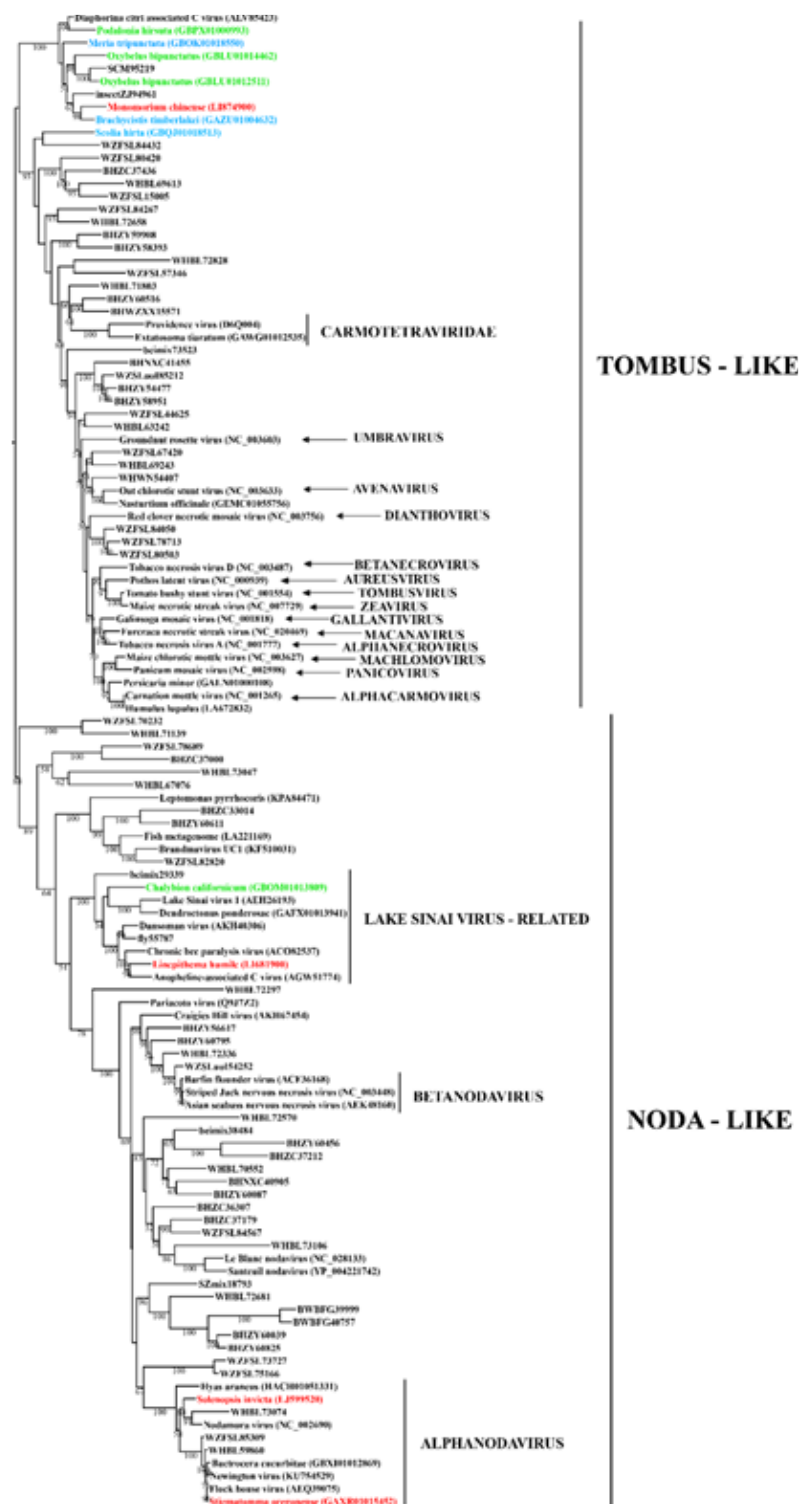
Slika B.2: Filogenetsko drevo virusov klade **Flavi**, ki smo jih našli v genomih in transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro. Endogenizirani virusi so označeni z zvezdico.



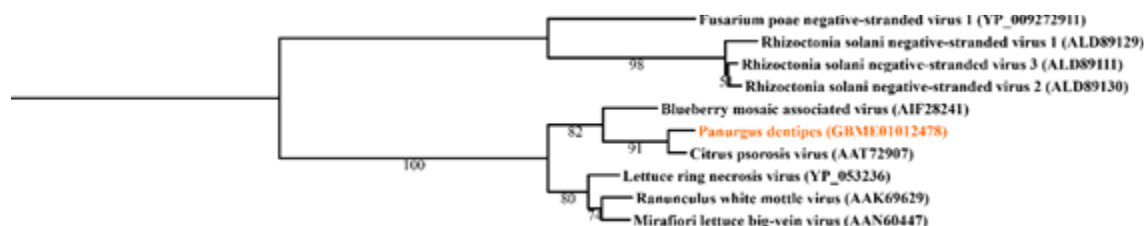
Slika B.3: Filogenetsko drevo virusov klade **Luteo-Sobemo**, ki smo jih našli v transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro.



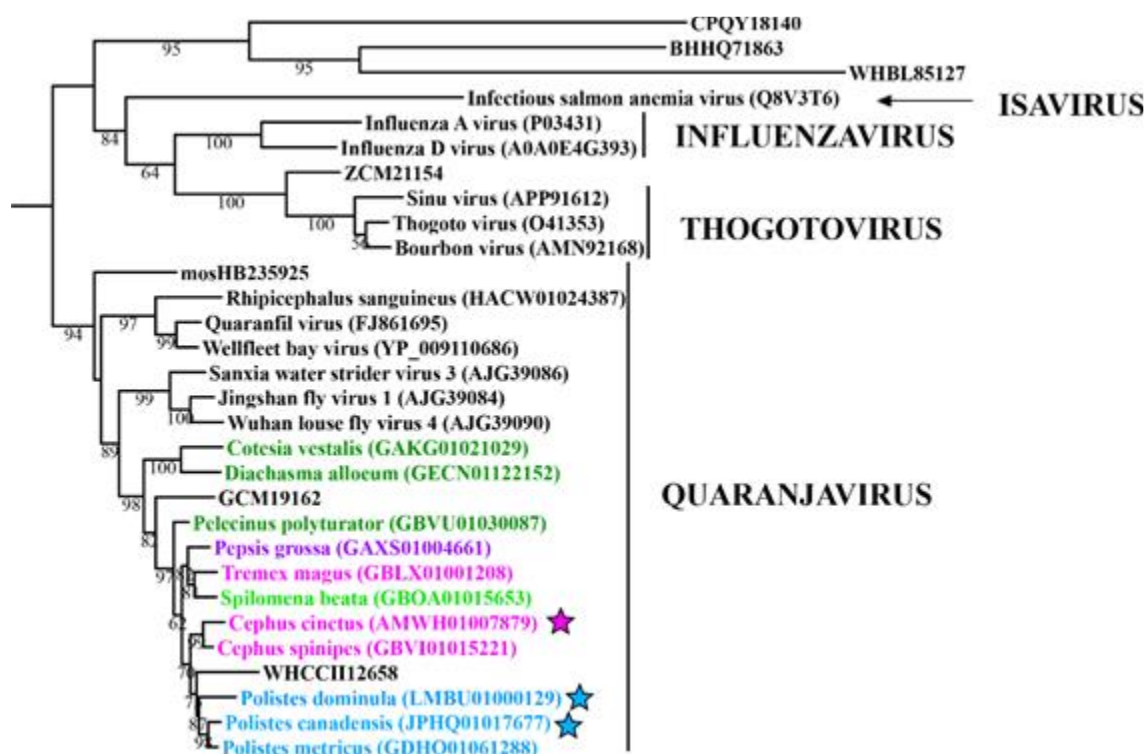
Slika B.4: Filogenetsko drevo virusov klade **Narna-Levi**, ki smo jih našli v genomih in transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro. Endogenizirani virusi so označeni z zvezdico.



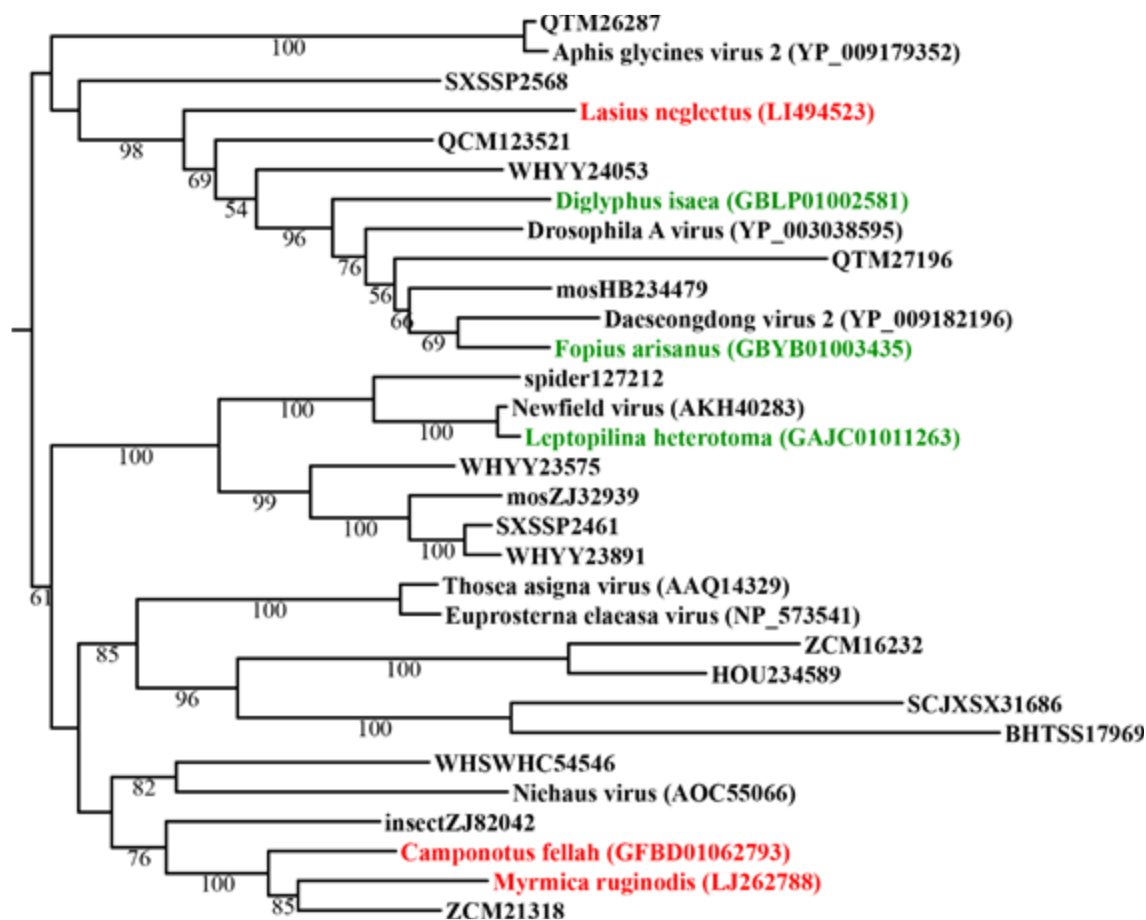
Slika B.5: Filogenetsko drevo virusov klade **Tombus-Noda**, ki smo jih našli v transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro.



Slika B.6: Filogenetsko drevo virusa klade **Ophio**, ki smo ga našli v transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro.



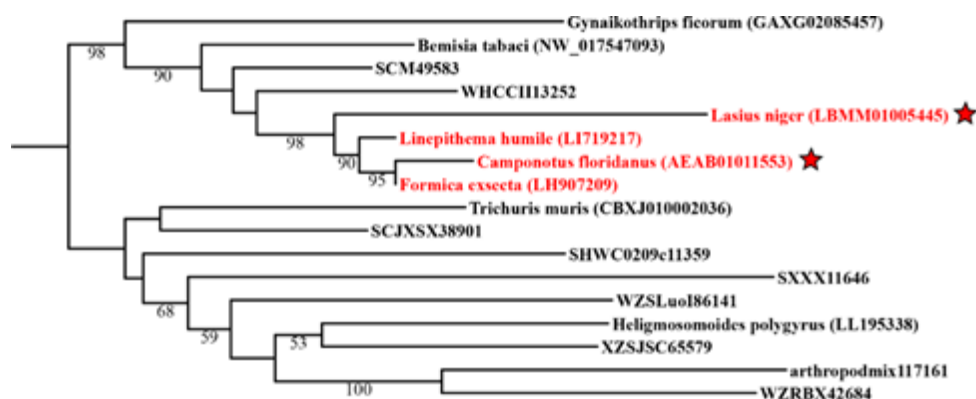
Slika B.7: Filogenetsko drevo virusov klade **Orthomyxo**, ki smo jih našli v genomih in transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro. Endogenizirani virusi so označeni z zvezdico.



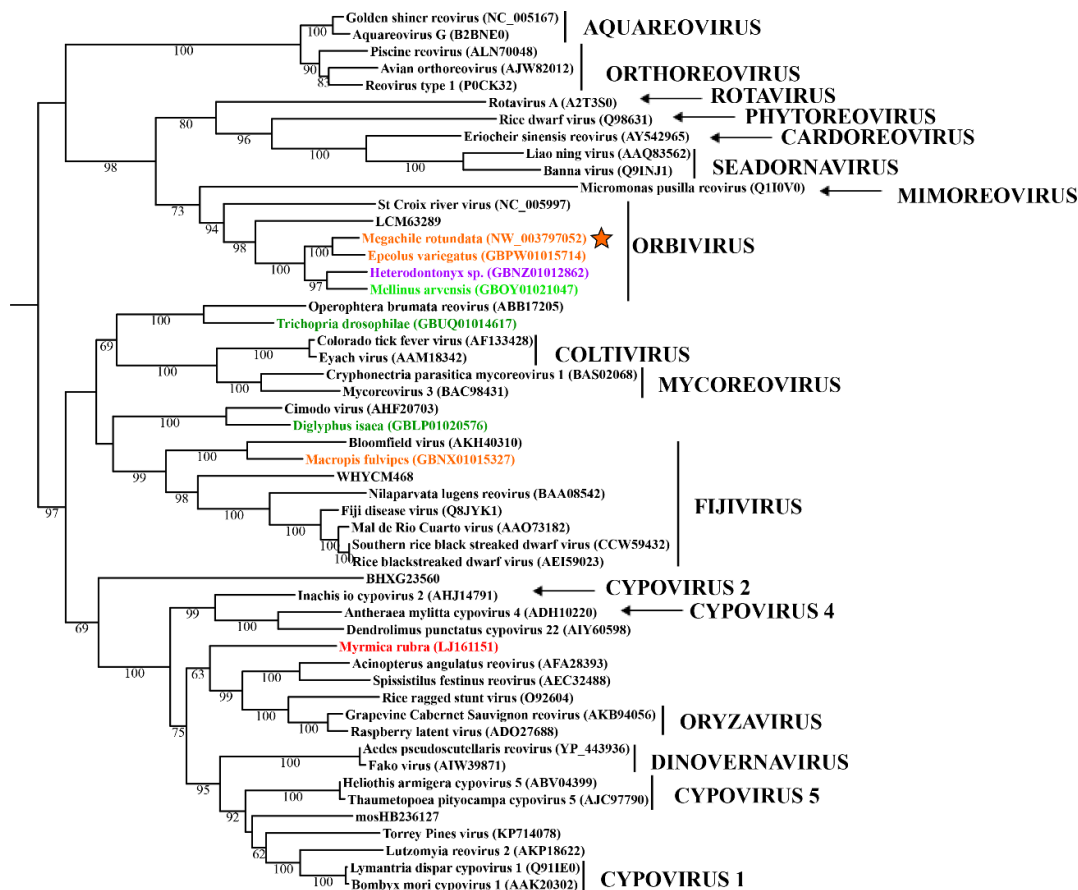
Slika B.8: Filogenetsko drevo virusov klade **Permutotetra**, ki smo jih našli v transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro.



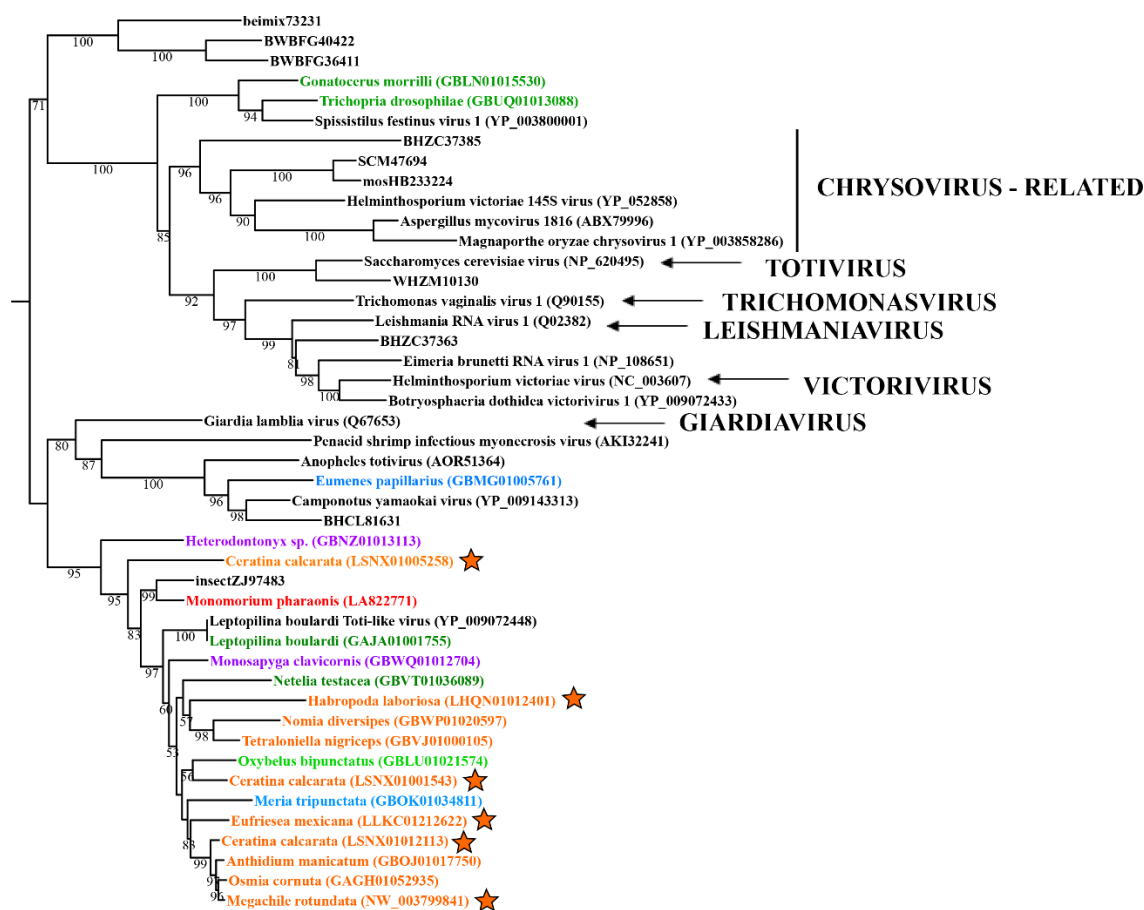
Slika B.9: Filogenetsko drevo virusov klade **Picorna-Calici**, ki smo jih našli v transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro.



Slika B.10: Filogenetsko drevo virusov klade **Qinivirus**, ki smo jih našli v genomih in transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro. Endogenizirani virusi so označeni z zvezdico.

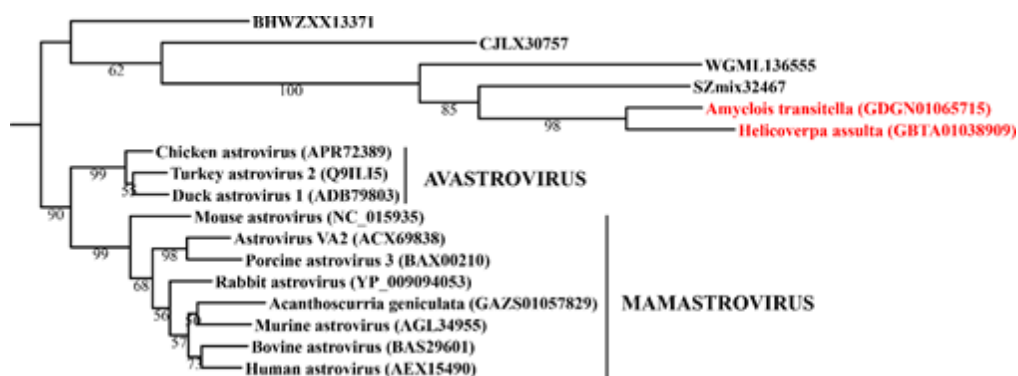


Slika B.11: Filogenetsko drevo virusov klade **Reo**, ki smo jih našli v genomih in transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro. Endogenizirani virusi so označeni z zvezdico.

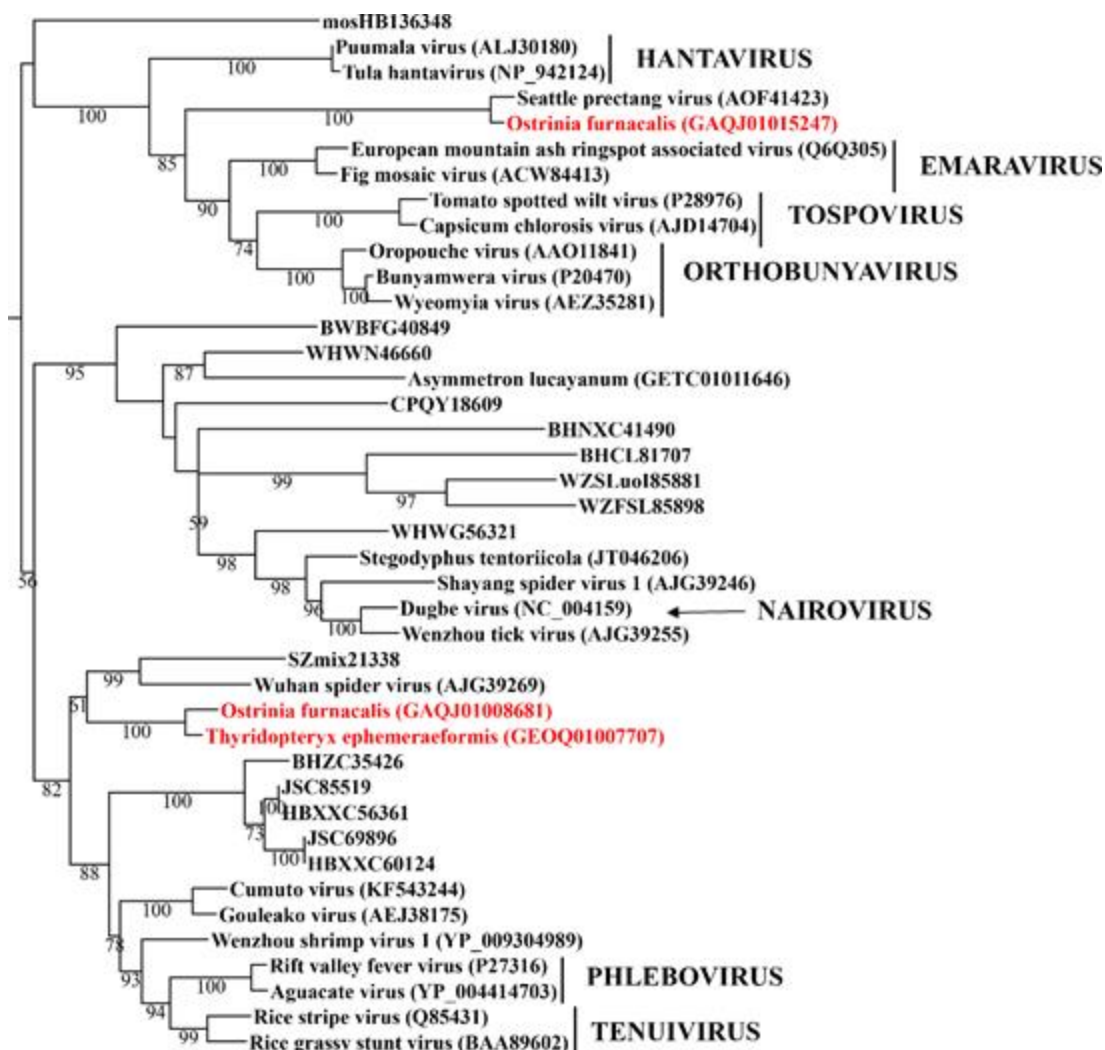


Slika B.12: Filogenetsko drevo virusov klade **Toti-Chryso**, ki smo jih našli v genomih in transkriptomih kožekrilcev (Hymenoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupina Apoidea je označena z oranžno, Pompiloidea z vijolično, Sphecoidea s svetlo zeleno, Chrysidoidea s temno modro, Terebrantes s temno zeleno, Formicidae z rdečo in ostali organizmi Vespoidea (ki niso Formicidae) s svetlo modro. Endogenizirani virusi so označeni z zvezdico.

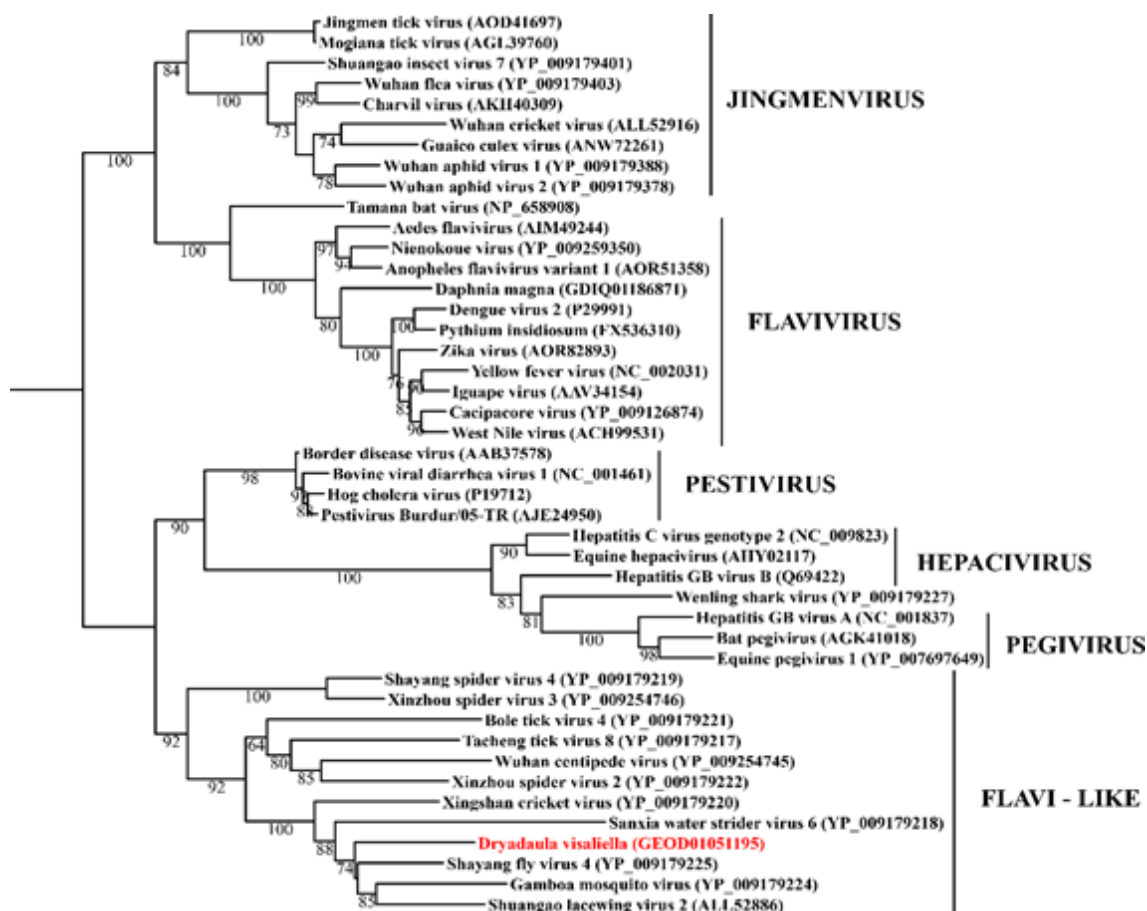
B.2 Metulji (Lepidoptera)



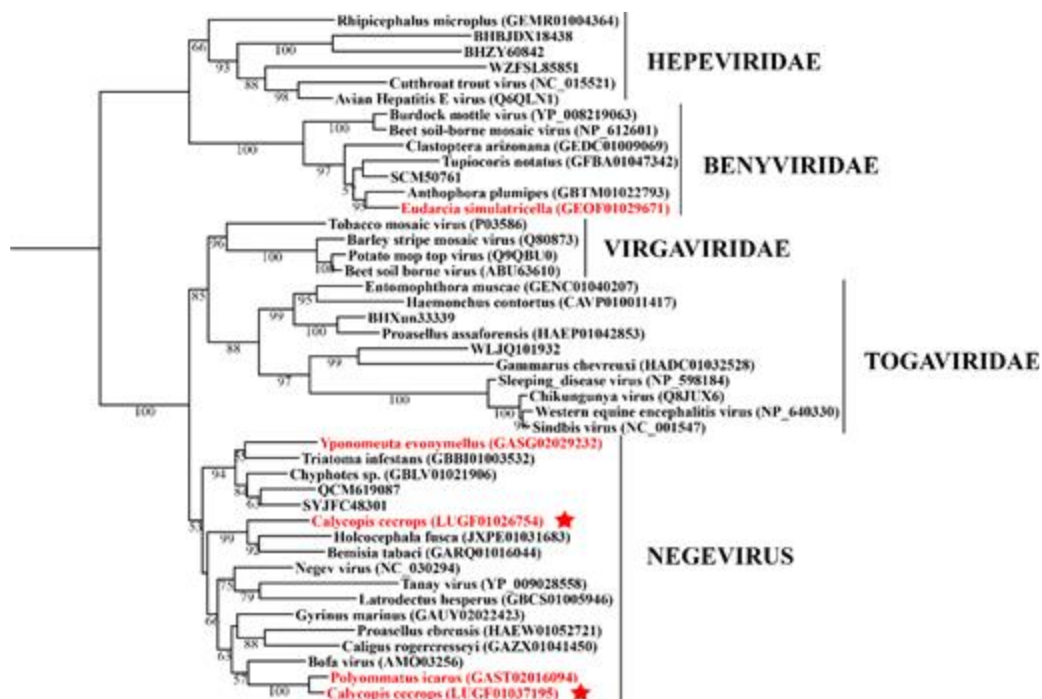
Slika B.13: Filogenetsko drevo virusov klade **Astro**, ki smo jih našli v transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo.



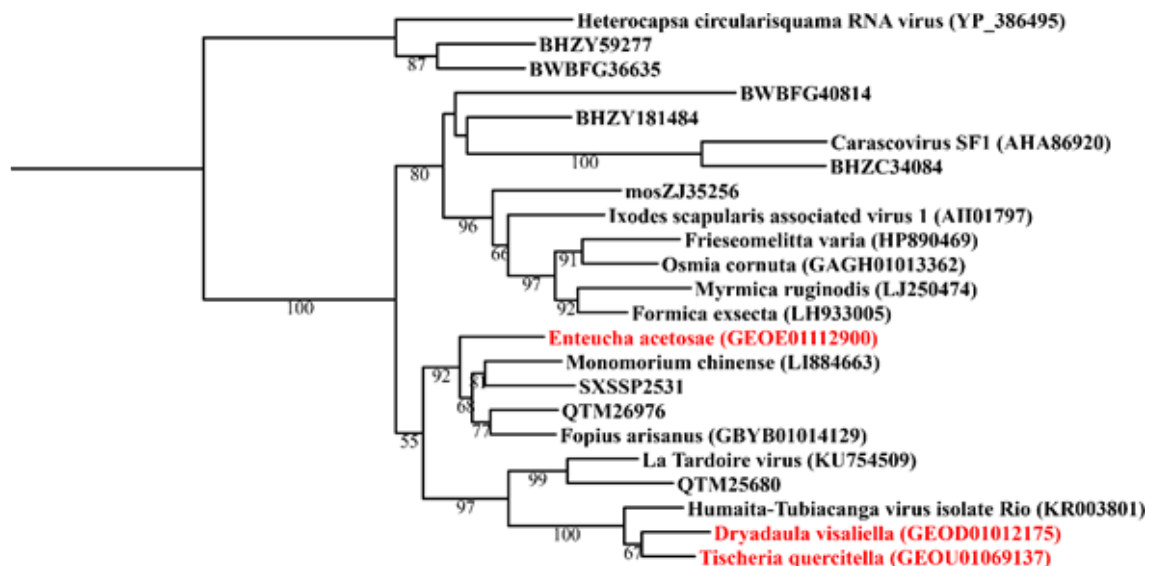
Slika B.14: Filogenetsko drevo virusov klade **Bunya-Arena**, ki smo jih našli v transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupino več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo.



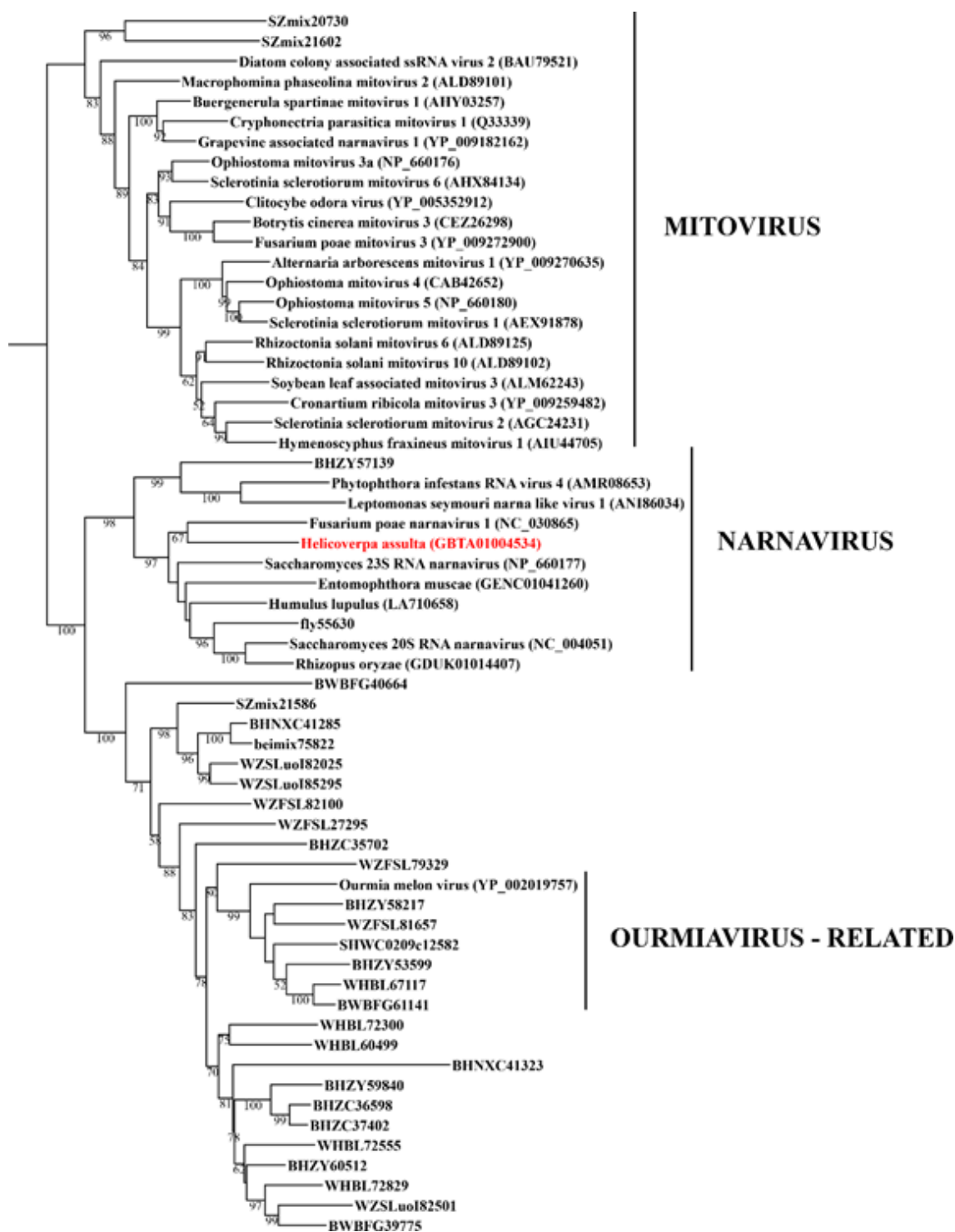
Slika B.15: Filogenetsko drevo virusa klade **Flavi**, ki smo ga odkrili v transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo.



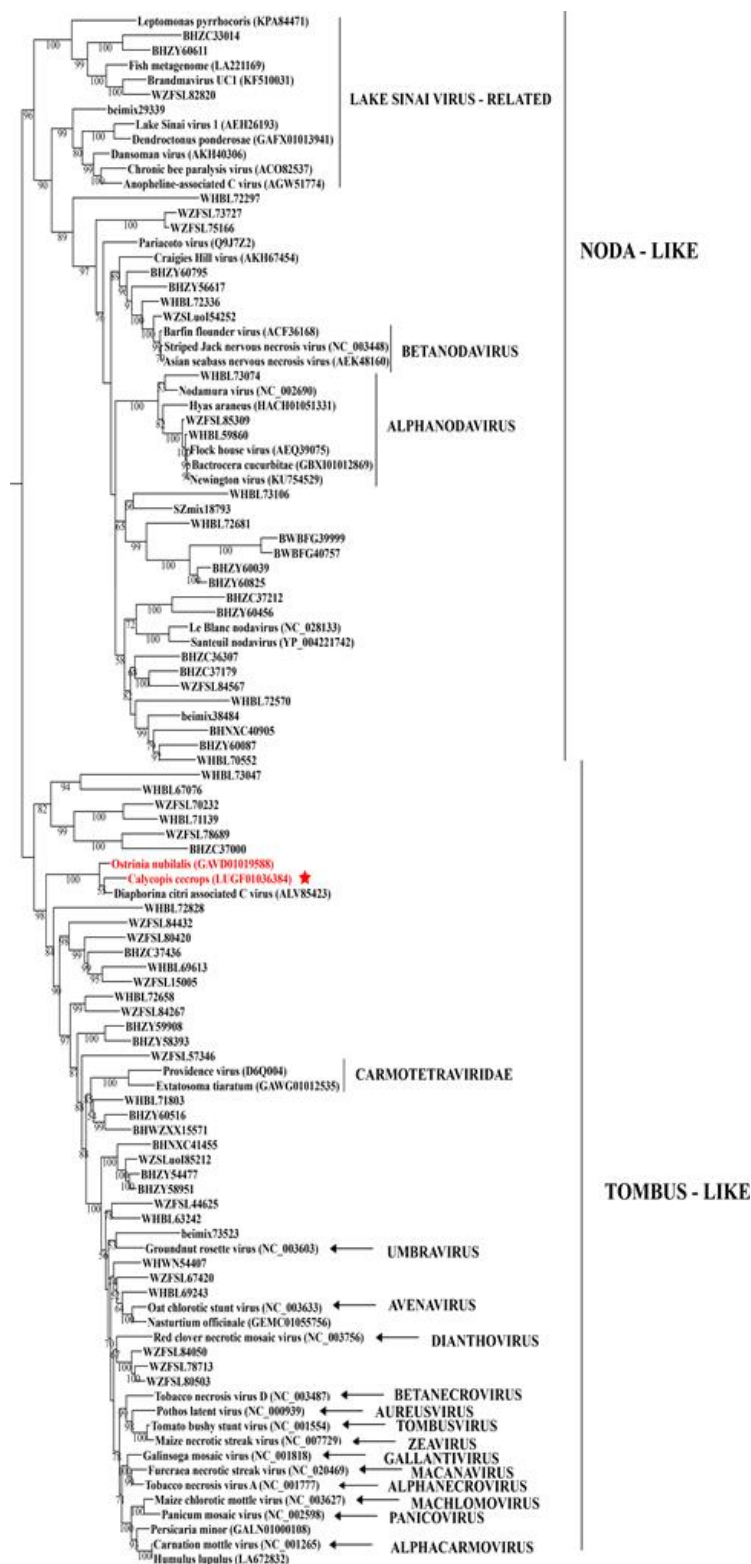
Slika B.16: Filogenetsko drevo virusov klade **Hepe-Virga**, ki smo jih našli v genomih in transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupino več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo. Endogenizirani virusi so označeni z zvezdico.



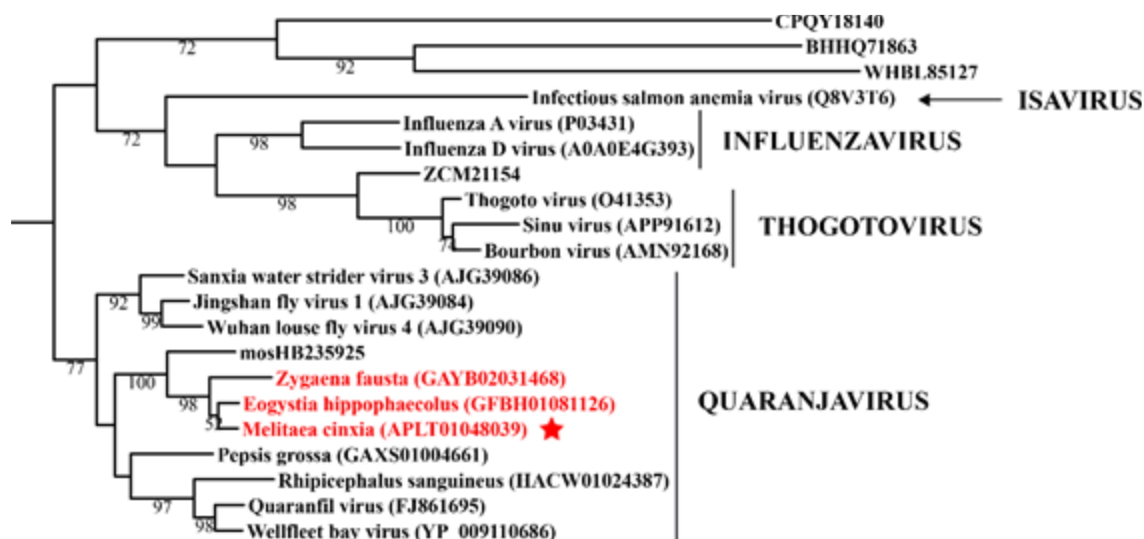
Slika B.17: Filogenetsko drevo virusov klade **Luteo-Sobemo**, ki smo jih našli v transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupino več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo.



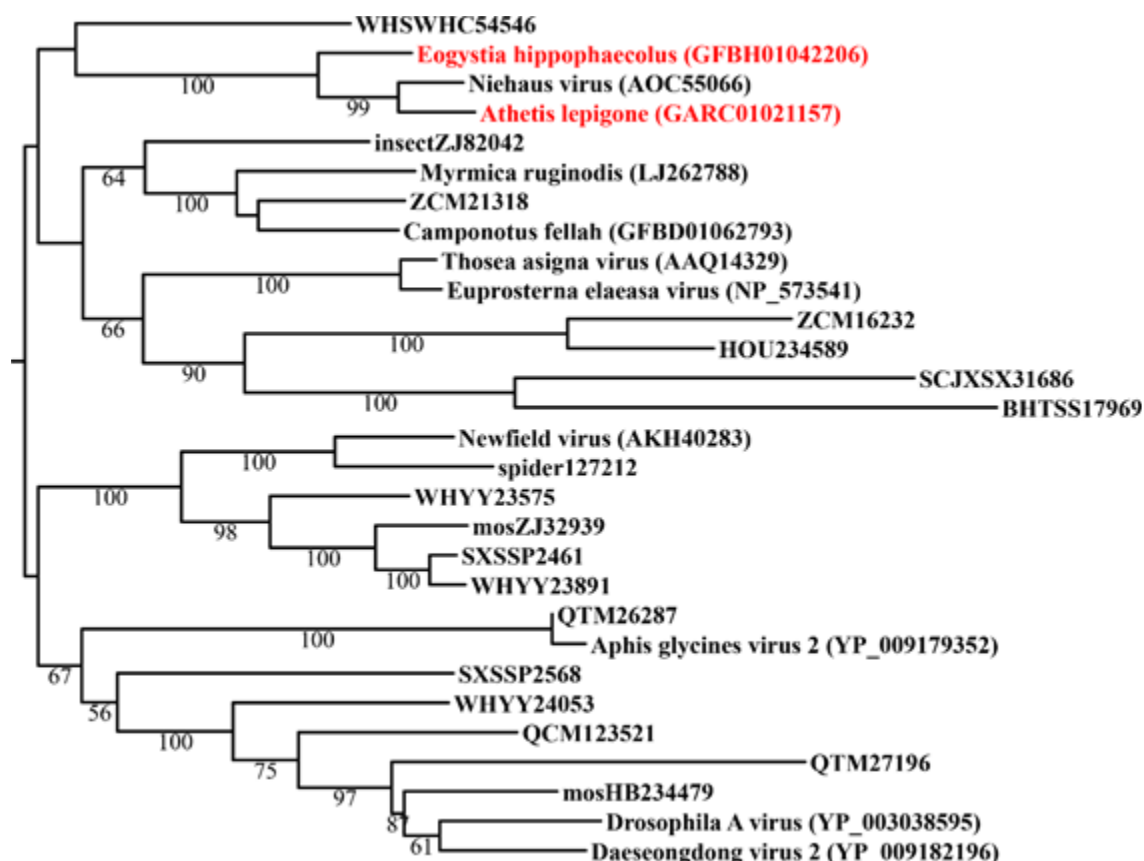
Slika B.18: Filogenetsko drevo virusa družine **Narnaviridae** (klada **Narna-Levi**), ki smo ga odkrili v transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupino več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo.



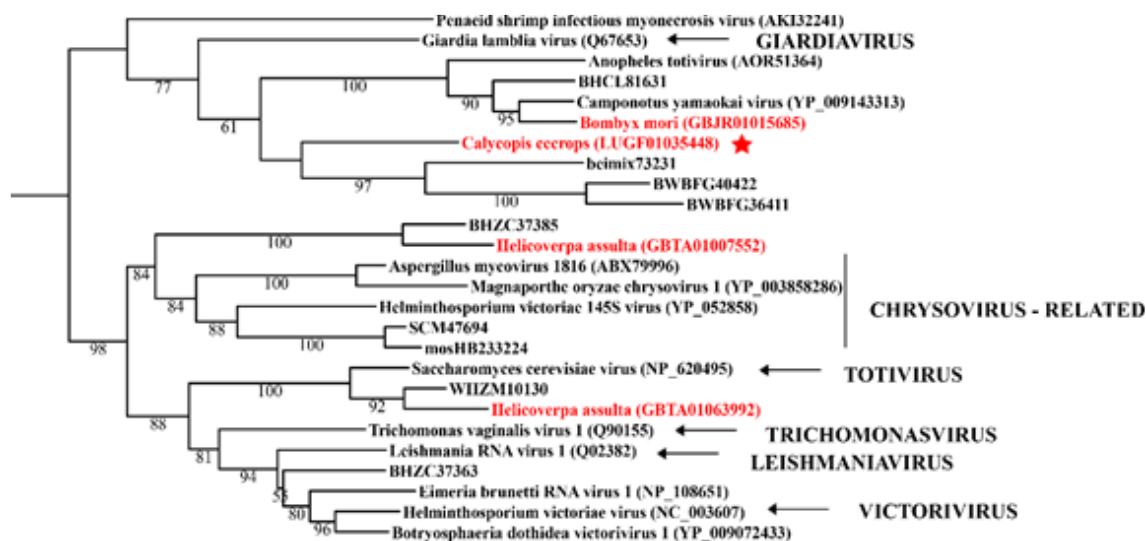
Slika B.19: Filogenetsko drevo virusov klade **Tombus-Noda**, ki smo jih našli v genomih in transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«) v programu IQTREE (Nguyen in sod., 2015). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo. Endogenizirani virusi so označeni z zvezdico.



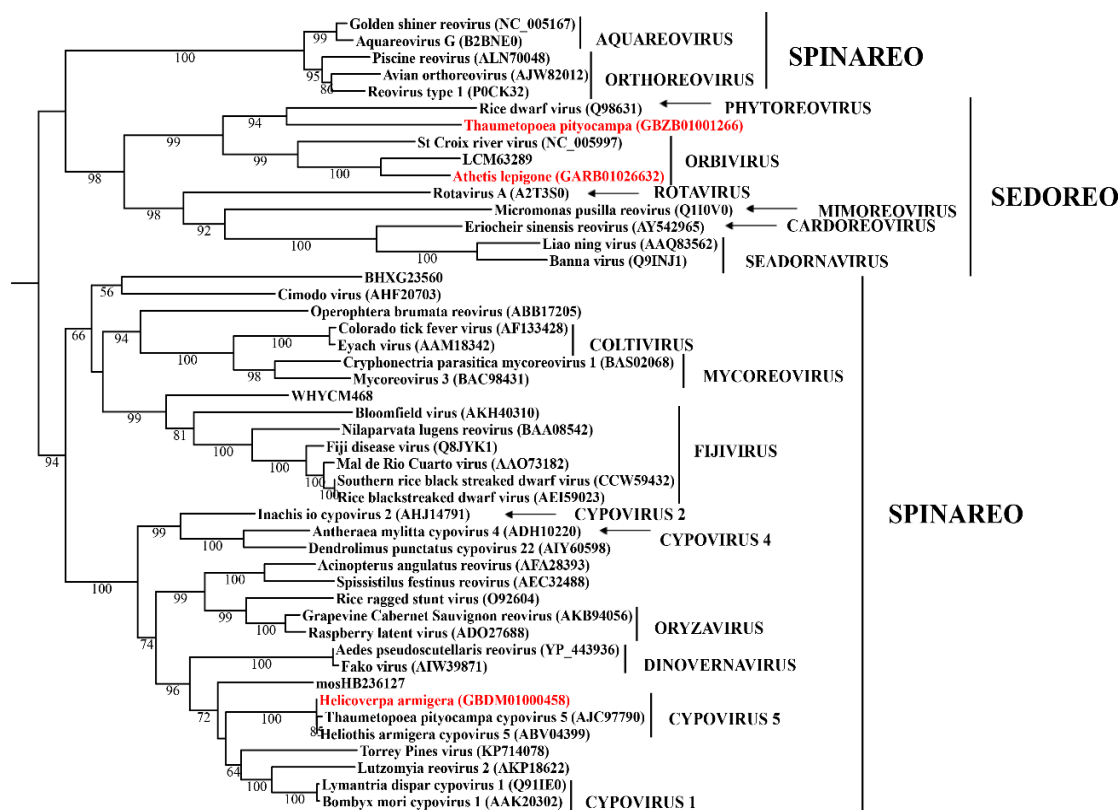
Slika B.20: Filogenetsko drevo virusov klade **Orthomyxo**, ki smo jih našli v genomih in transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo. Endogenizirani virusi so označeni z zvezdico.



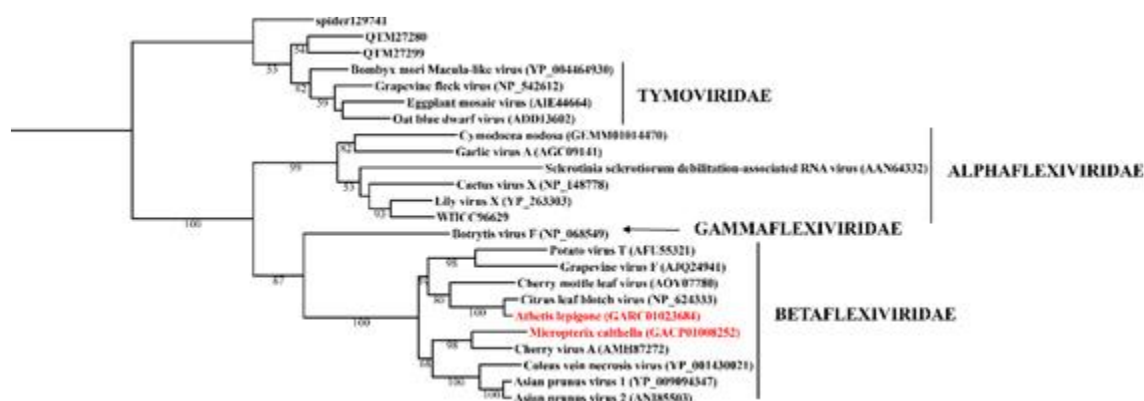
Slika B.21: Filogenetsko drevo virusov klade **Permutotetra**, ki smo jih našli v transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo.



Slika B.22: Filogenetsko drevo virusov klade **Toti-Chryso**, ki smo jih našli v genomih in transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo. Endogenizirani virusi so označeni z zvezdico.

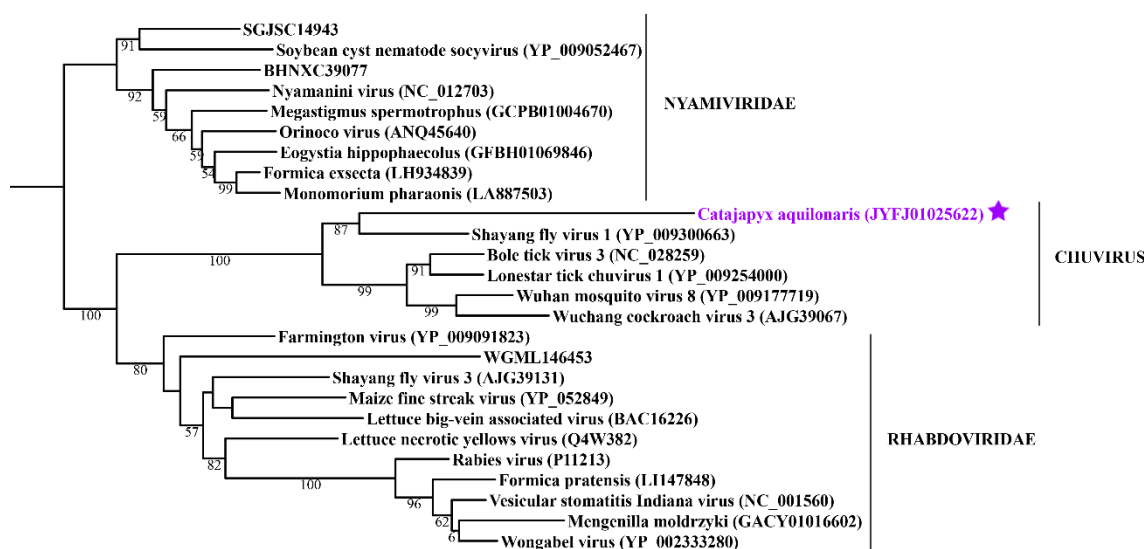


Slika B.23: Filogenetsko drevo virusov klade **Reo**, ki smo jih našli v transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo.

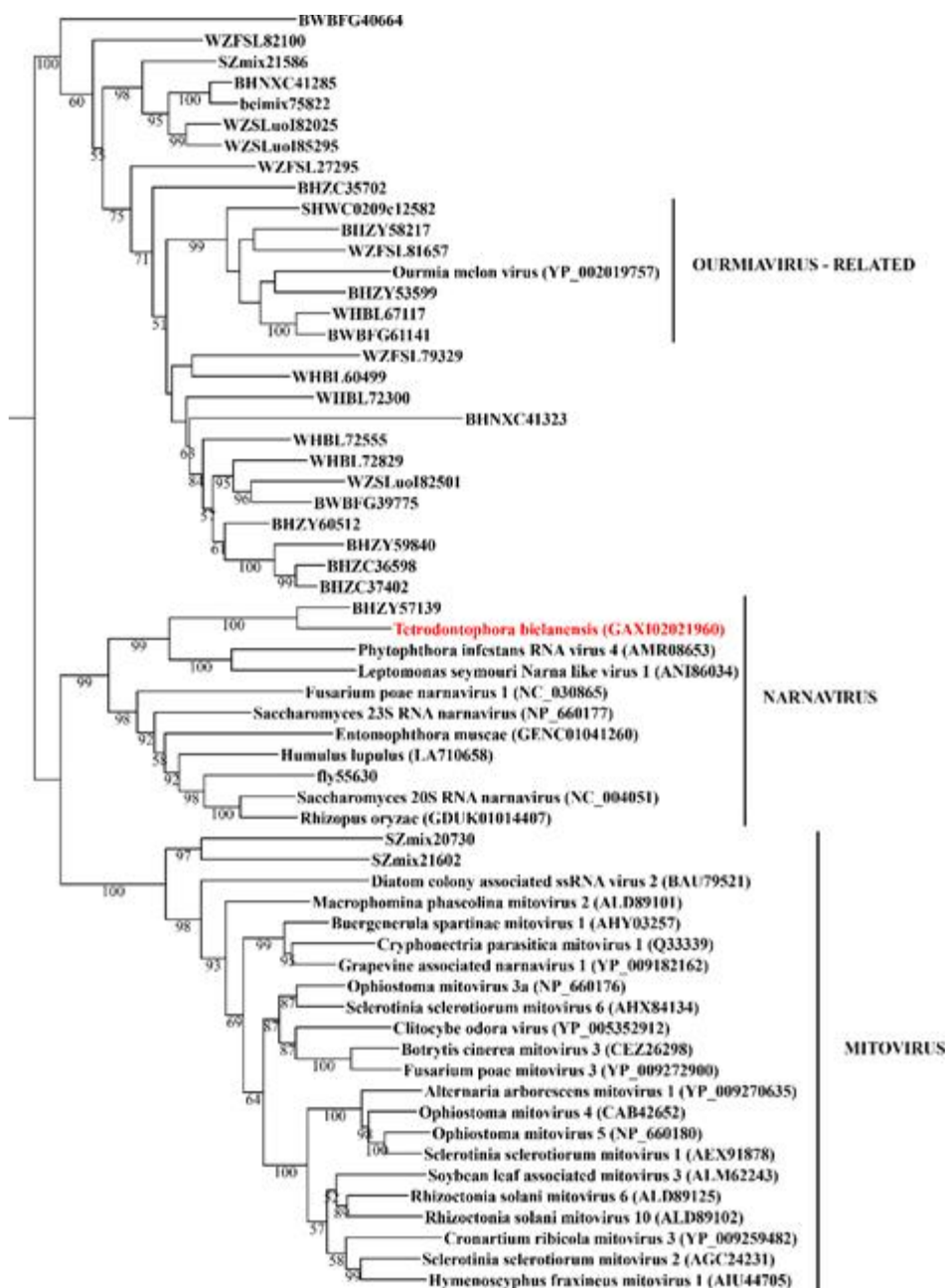


Slika B.24: Filogenetsko drevo virusov redu **Tymovirales** (klada **Hepe-Virga**), ki smo jih odkrili v transkriptomih več in metuljev (Lepidoptera). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skupine več in metuljev nismo ločevali na podskupine, zato so vsi odkriti virusi označeni z rdečo.

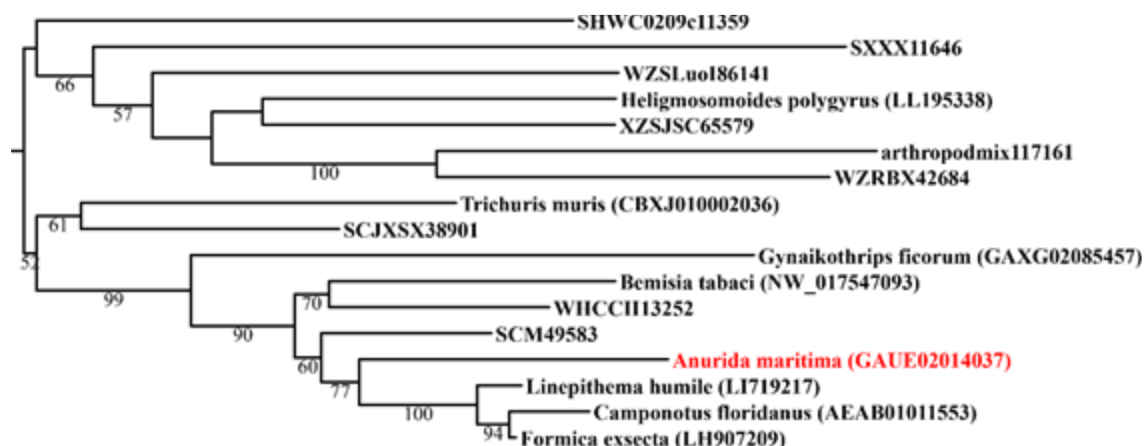
B.3 Najstarejše skupine žuželk: skakači (Collembola), dvorepke (Diplura), ščetinorepke (Zygentoma), Monocondylia



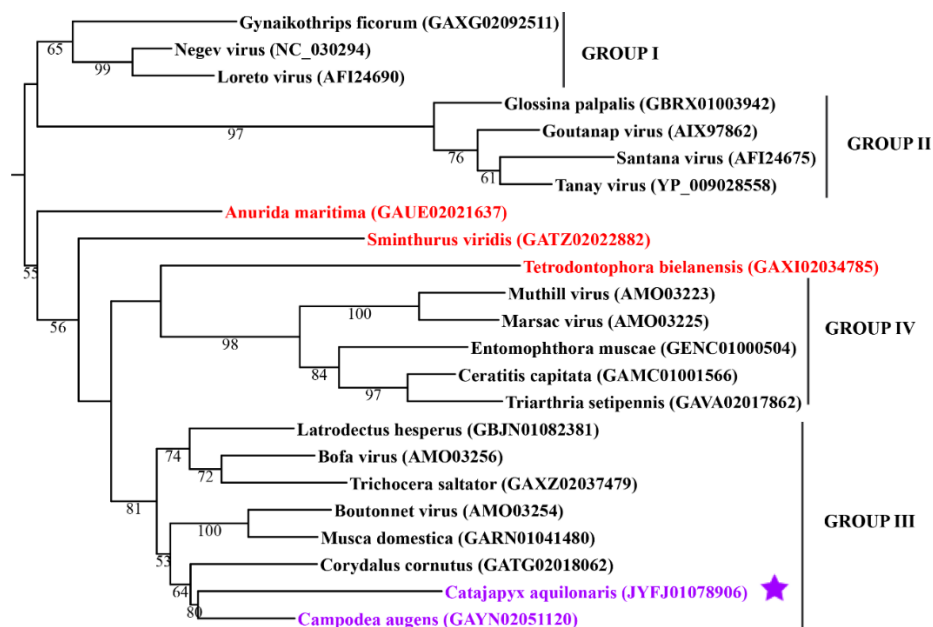
Slika B.25: Filogenetsko drevo virusa klade **Mono-Chu**, ki smo ga našli v genomu najstarejših skupin žuželk. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skakači (Collembola) so označeni z rdečo, dvorepke (Diplura) z vijolično, ščetinorepke (Zygentoma) z modro in skupina Monocondylia z zeleno barvo. Endogeniziran RNA virus je označen z zvezdico.



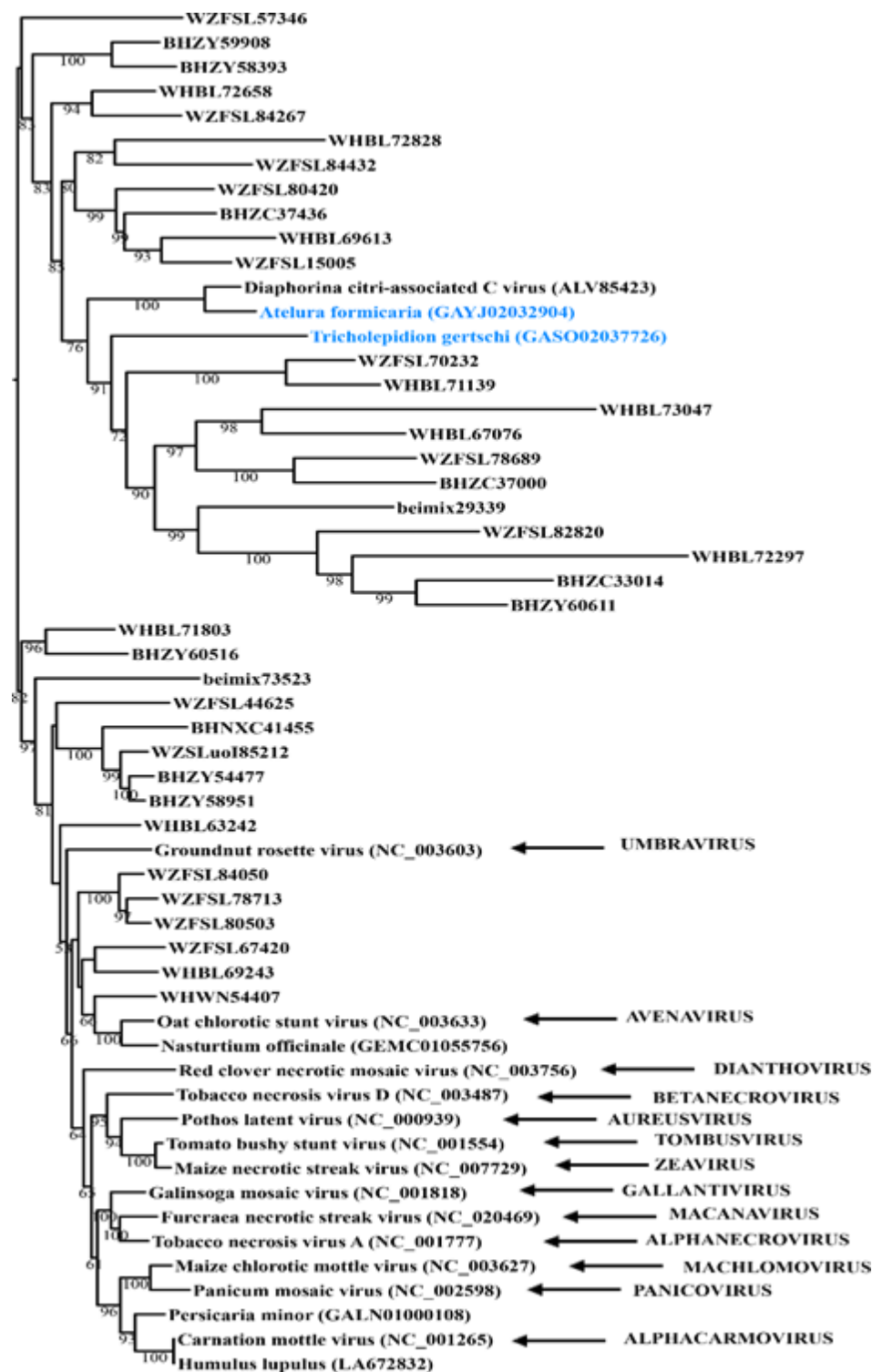
Slika B.26: Filogenetsko drevo virusa družine **Narnaviridae** (klade **Narna-Levi**), ki smo ga odkrili v transkriptomih najstarejših skupin žuželk. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skakači (Collembola) so označeni z rdečo, dvorepke (Diplura) z vijolično, ščetinorepke (Zygentoma) z modro in skupina Monocondylia z zeleno barvo.



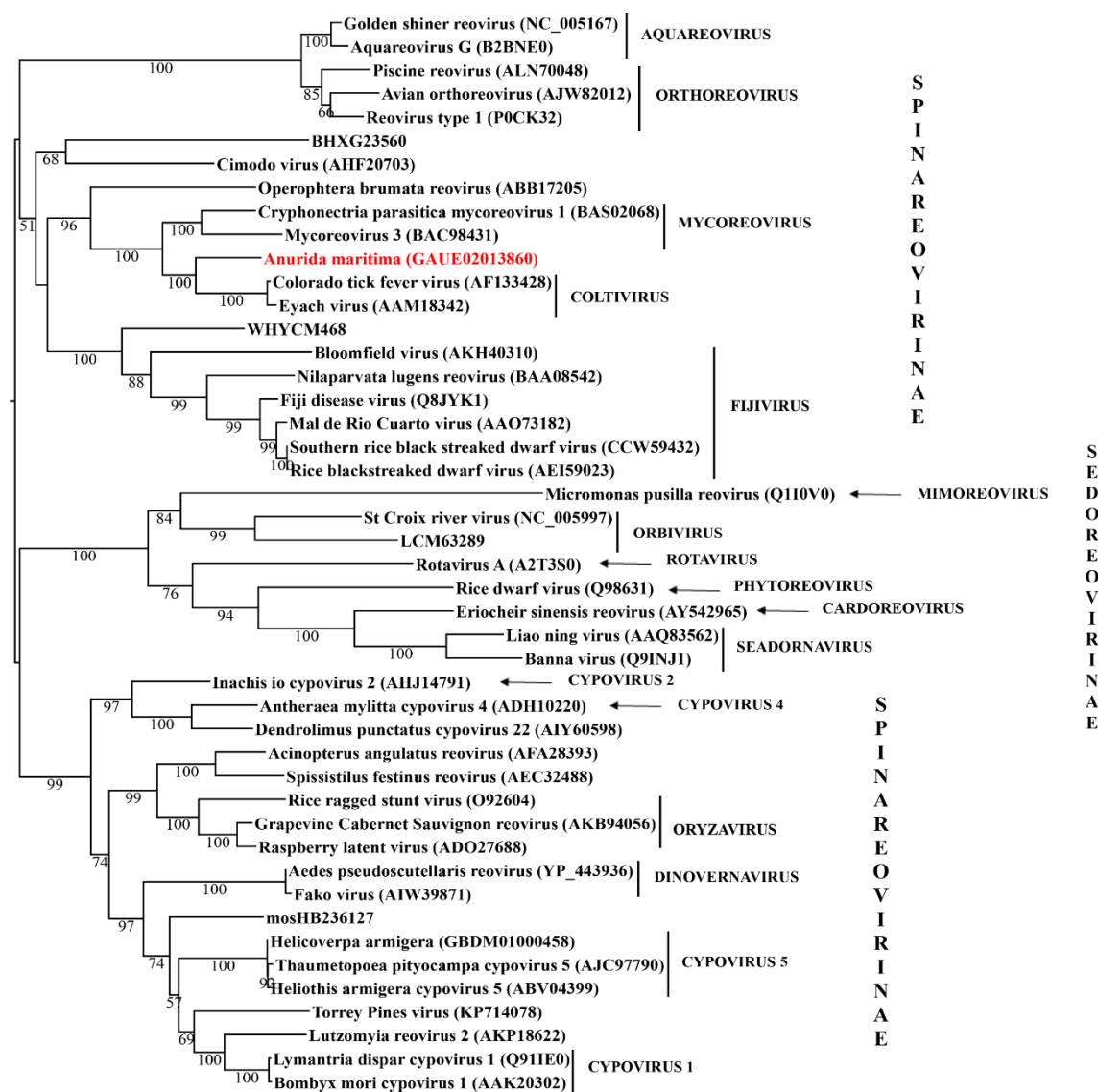
Slika B.27: Filogenetsko drevo virusa klade **Qinvirus**, ki smo ga našli v transkriptomih najstarejših skupin žuželk. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skakači (Collembola) so označeni z rdečo, dvorepke (Diplura) z vijolično, ščetinorepke (Zygentoma) z modro in skupina Monocondylia z zeleno barvo.



Slika B.28: Filogenetsko drevo virusov skupine **Negevirus** (klada **Hepe-Virga**), ki smo jih našli v genomih in transkriptomih najstarejših skupin žuželk. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skakači (Collembola) so označeni z rdečo, dvorepke (Diplura) z vijolično, ščetinorepke (Zygentoma) z modro in skupina Monocondylia z zeleno barvo. Endogenizirani RNA virusi so označeni z zvezdico. Angleški izraz: group – skupina.

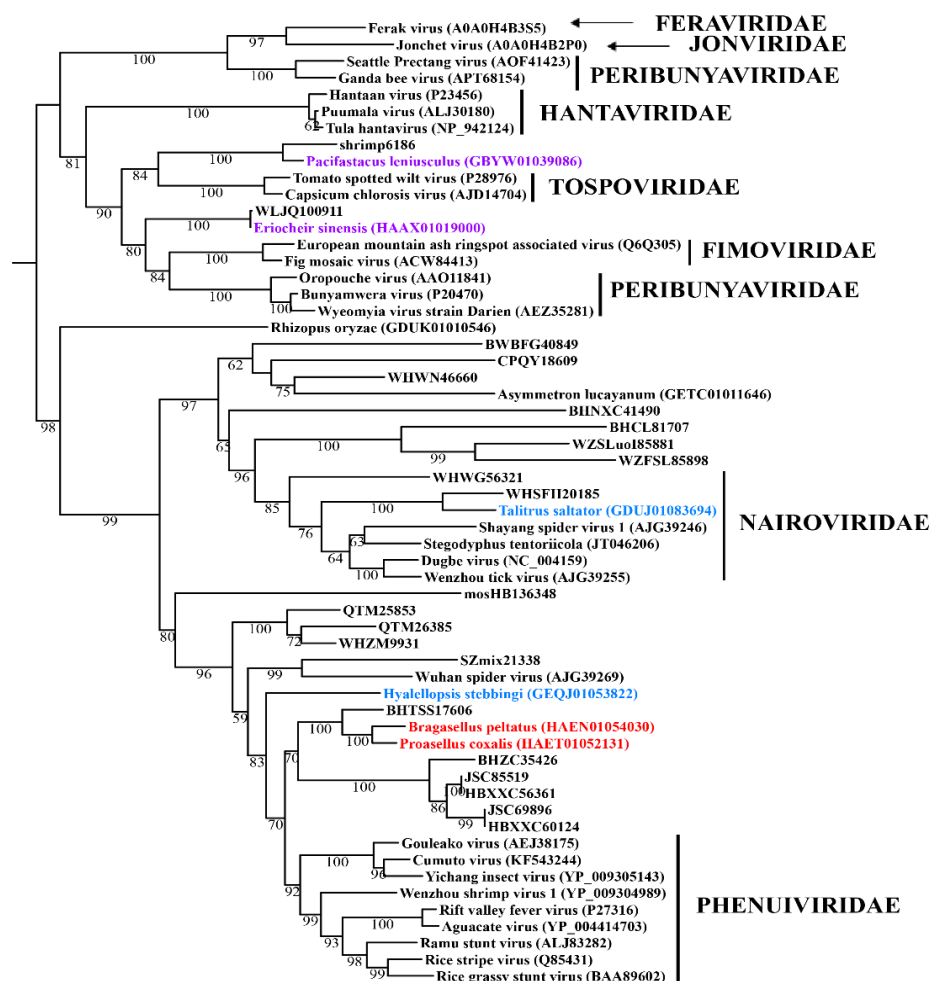


Slika B.29: Filogenetsko drevo virusov družine **Tombusviridae** (klada **Tombus-Noda**), ki smo jih našli v transkriptomih najstarejših skupin žuželk. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skakači (Collembola) so označeni z rdečo, dvorepke (Diplura) z vijolično, ščetinorepke (Zygentoma) z modro in skupina Monocondylia z zeleno barvo.

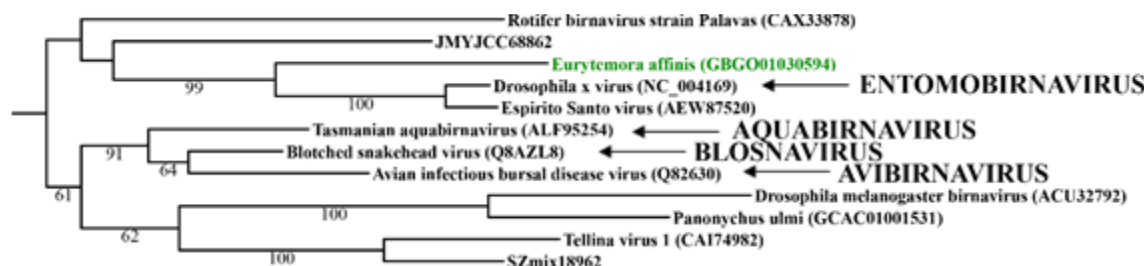


Slika B.30: Filogenetsko drevo virusov klade **Reo**, ki smo jih našli v transkriptomih najstarejših skupin žuželk. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Skakači (Collembola) so označeni z rdečo, dvorepke (Diplura) z vijolično, ščetinorepke (Zygentoma) z modro in skupina Monocondylia z zeleno barvo.

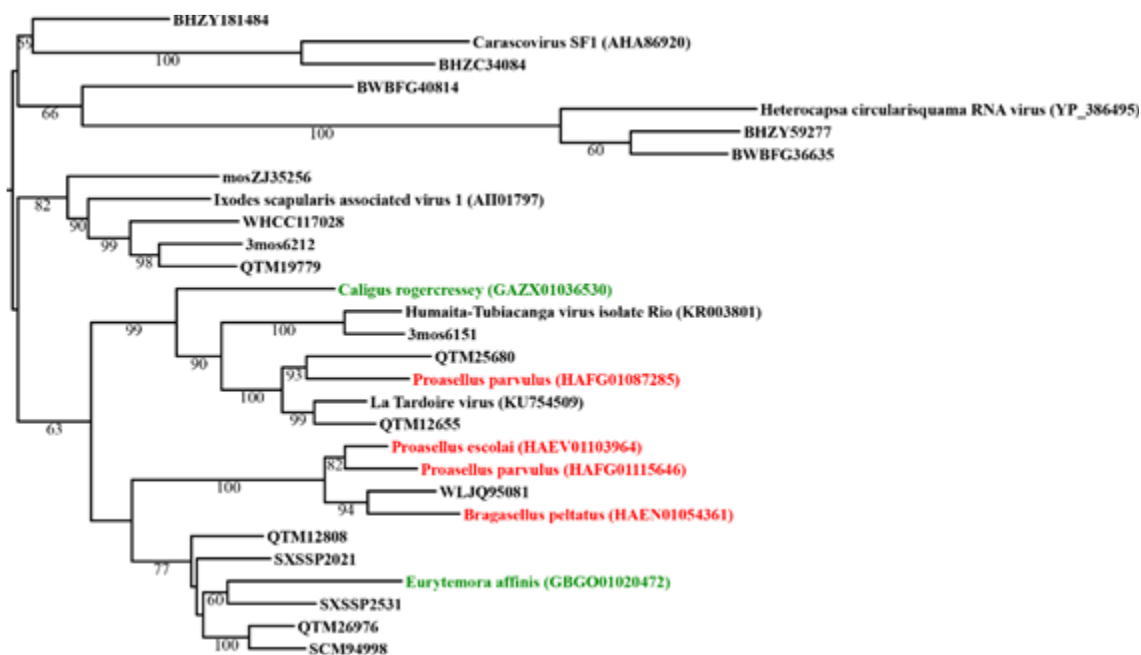
B.4 Raki (Crustacea)



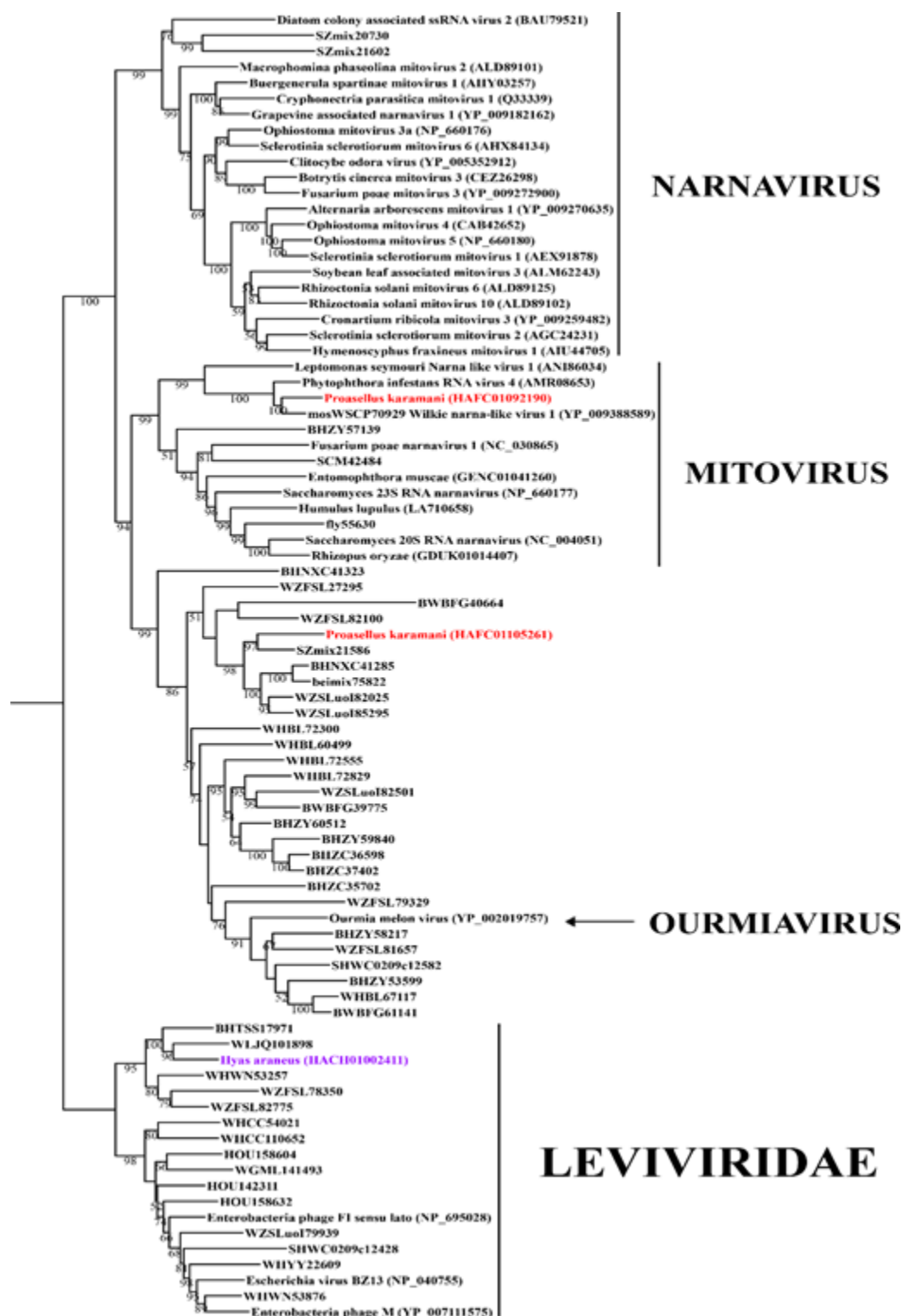
Slika B.31: Filogenetsko drevo virusov klade **Bunya-Arena**, ki smo jih odkrili v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postranice (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično.



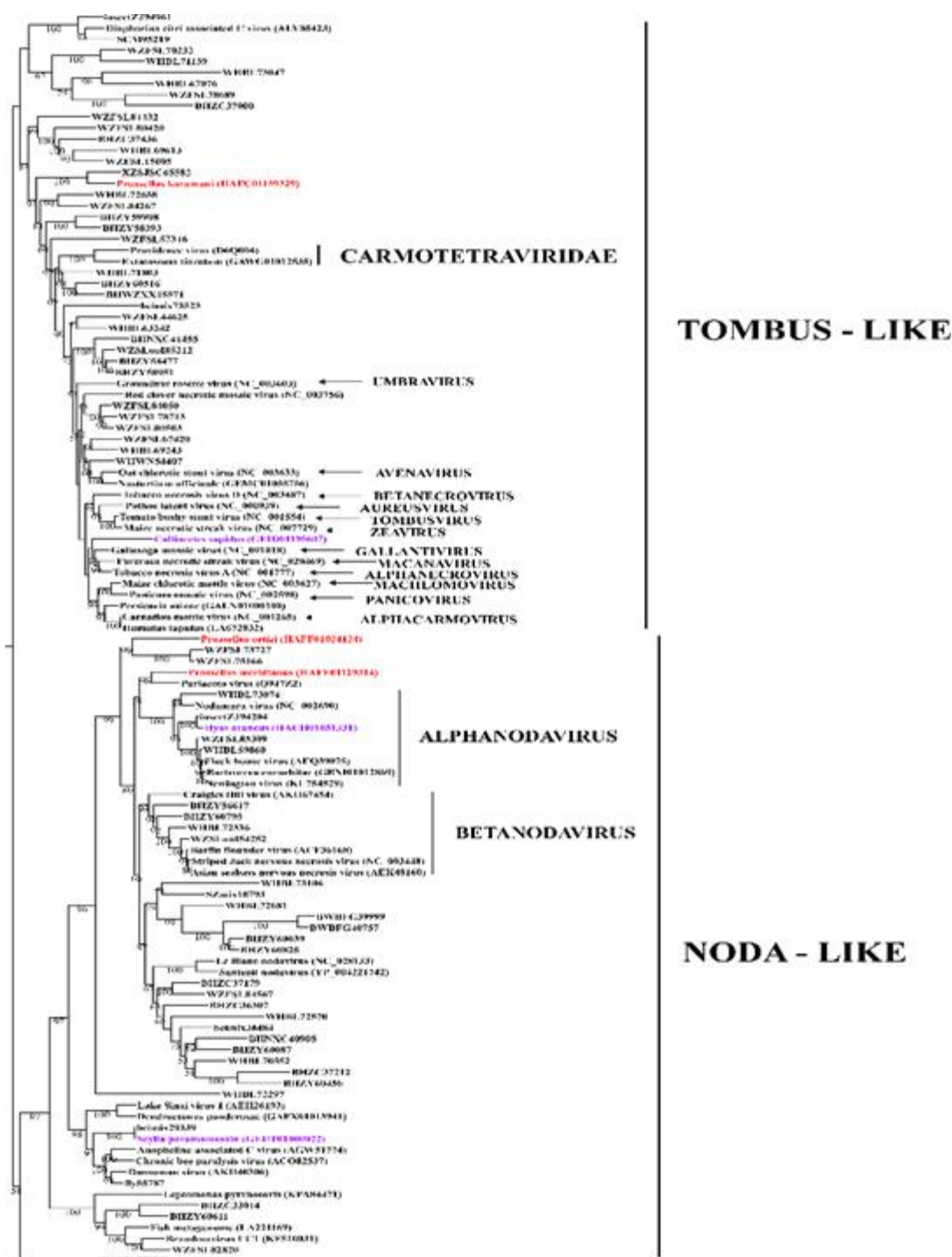
Slika B.32: Filogenetsko drevo virusa klade **Birna**, ki smo ga našli v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postranice (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično.



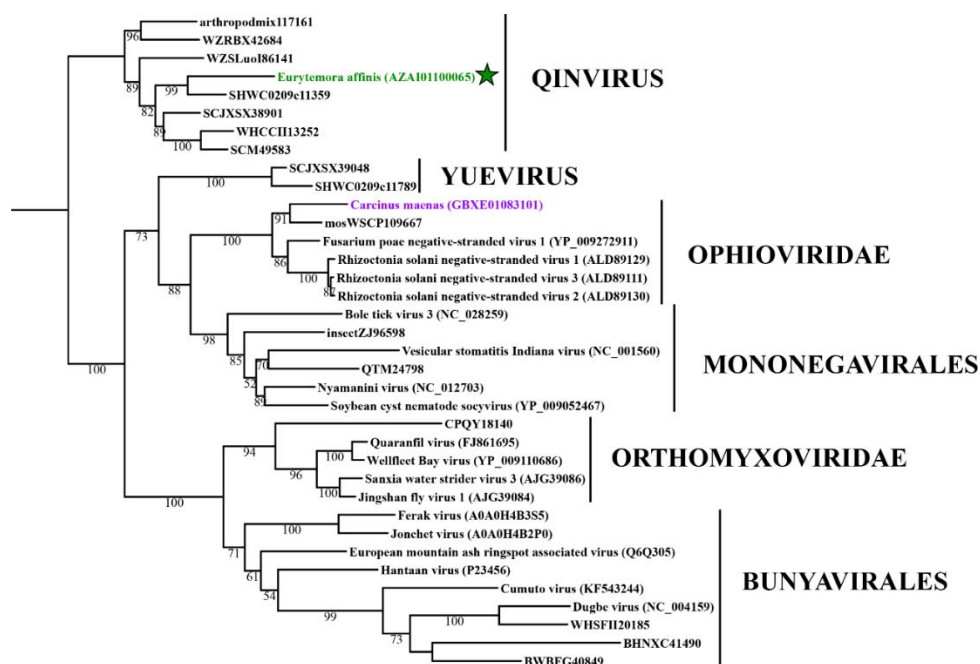
Slika B.34: Filogenetsko drevo virusov klade **Luteo-Sobemo**, ki smo jih našli v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postranice (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično.



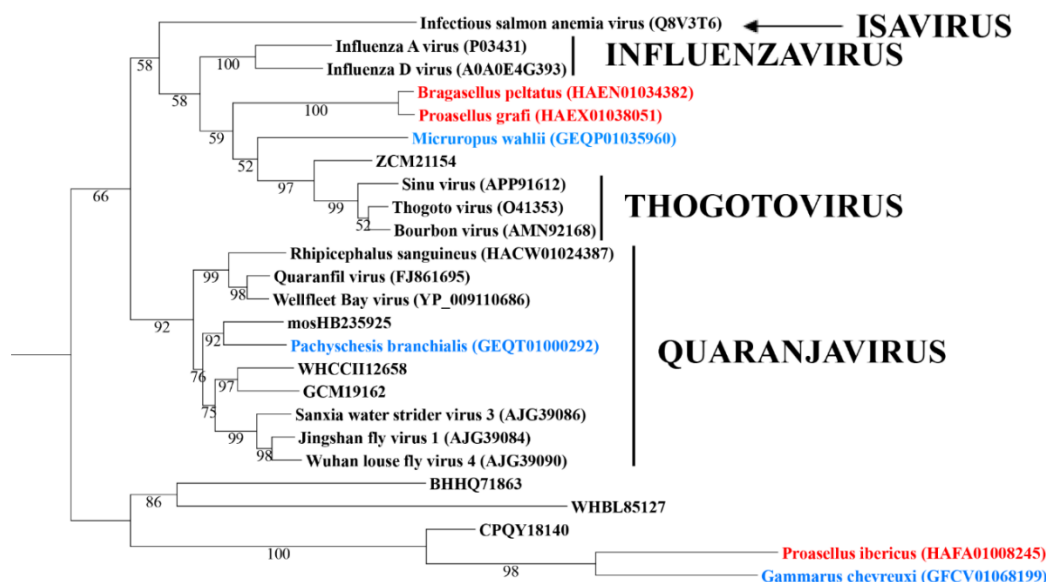
Slika B.35: Filogenetsko drevo virusov klade **Narna-Levi**, ki smo jih našli v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postrance (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično.



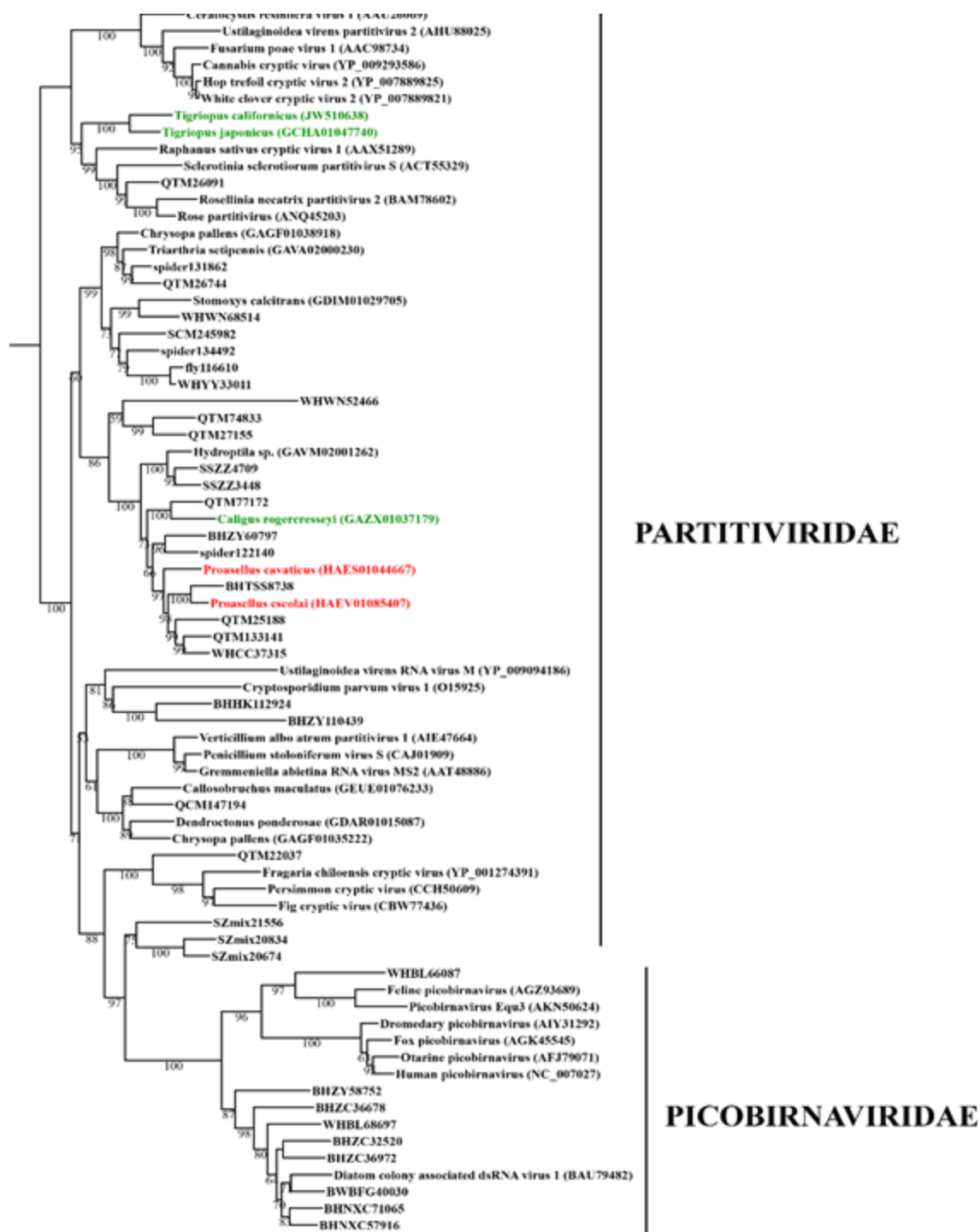
Slika B.36: Filogenetsko drevo virusov klade **Tombus-Noda**, ki smo jih odkrili v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postranice (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično.



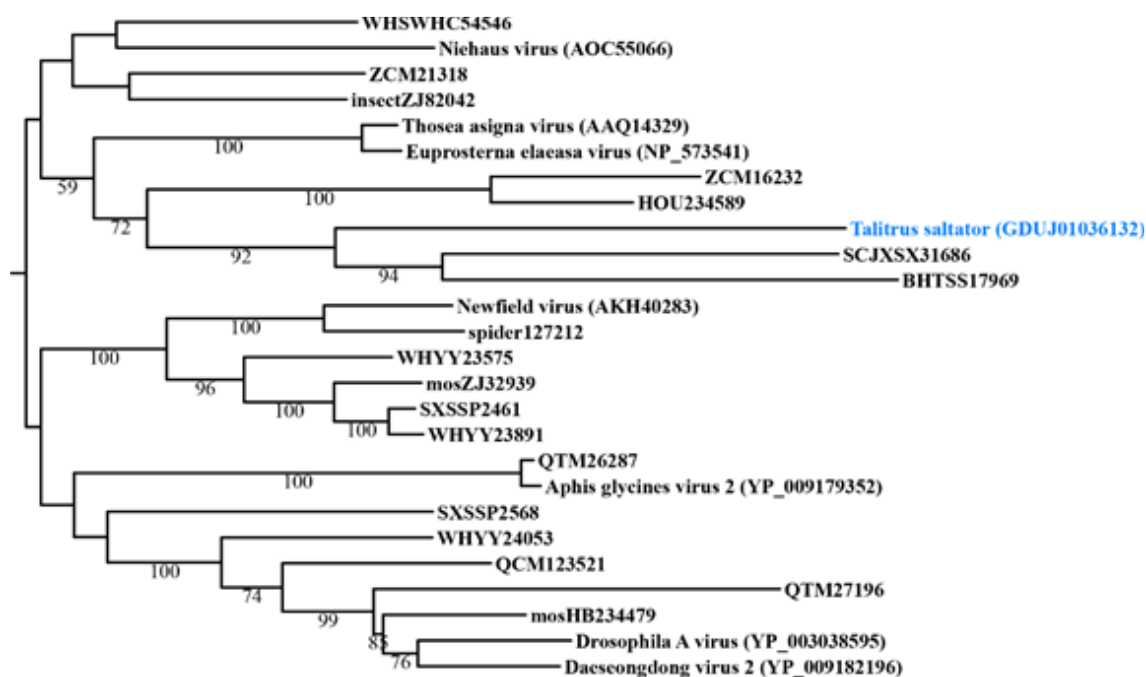
Slika B.37: Filogenetsko drevo virusov klad Ophio in Qin (skupine Qinvirus, Yuevirus, Ophioviridae, Mononegavirales, Orthomyxoviridae in Bunyavirales), ki smo jih našli v genomih in transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postranice (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično. Endogenizirani virusi so označeni z zvezdico.



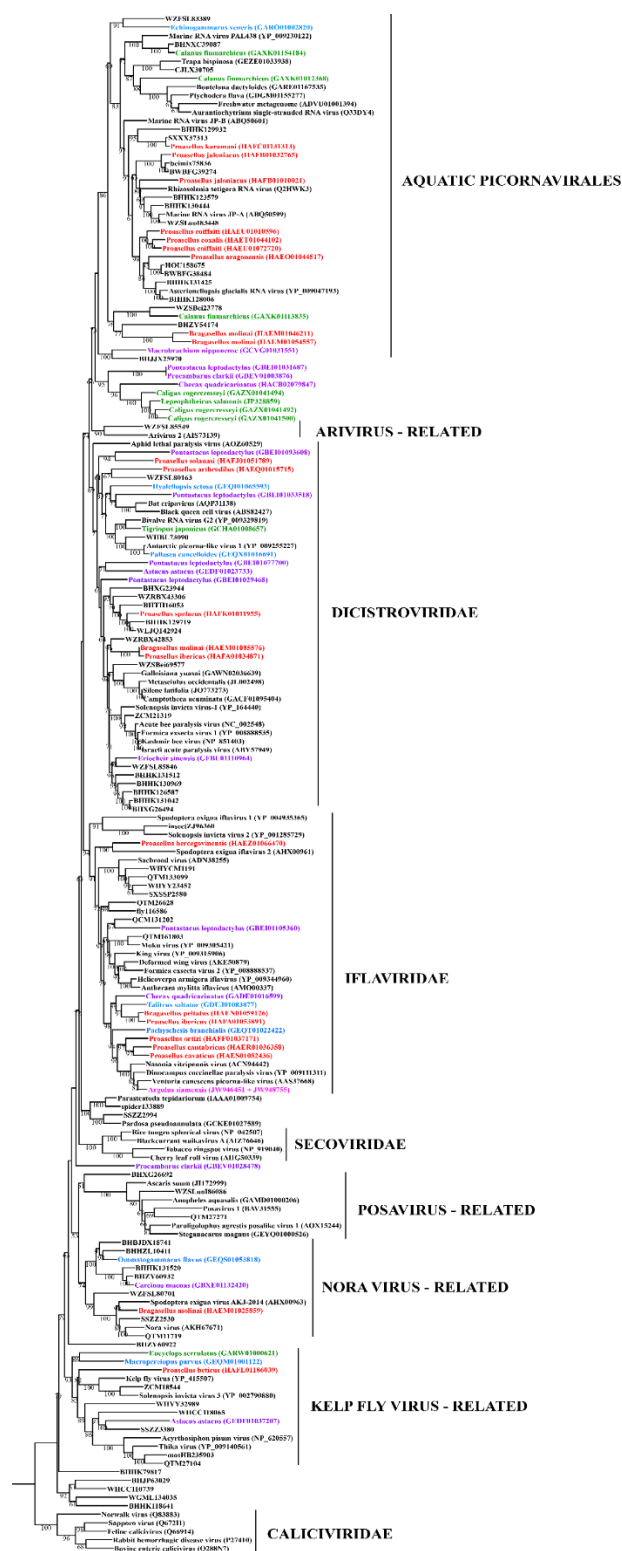
Slika B.38: Filogenetsko drevo virusov klade Orthomyxo, ki smo jih odkrili v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postranice (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično.



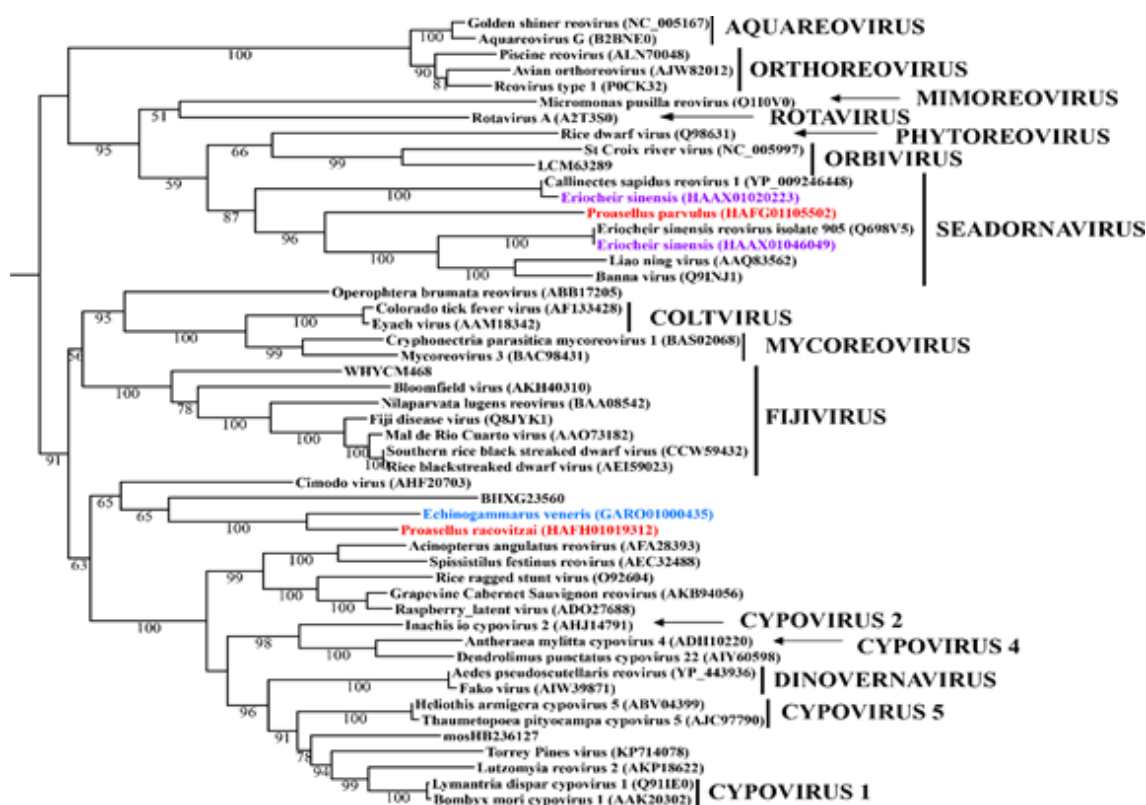
Slika B.39: Filogenetsko drevo virusov klade **Partiti-Picobirna**, ki smo jih našli v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postranice (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično.



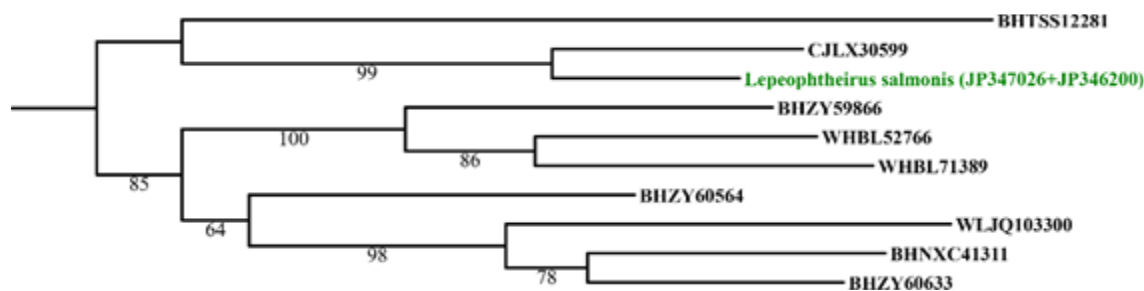
Slika B.40: Filogenetsko drevo virusa klade **Permutotetra**, ki smo ga odkrili v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postranice (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično.



Slika B.41: Filogenetsko drevo virusov klade **Picorna-Calici**, ki smo jih našli v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postrance (Amphipoda) z modro, deseteronožci (Decapoda) z vijolično in skupina Branchiura (Argulus) z magenta barvnim odtenkom.

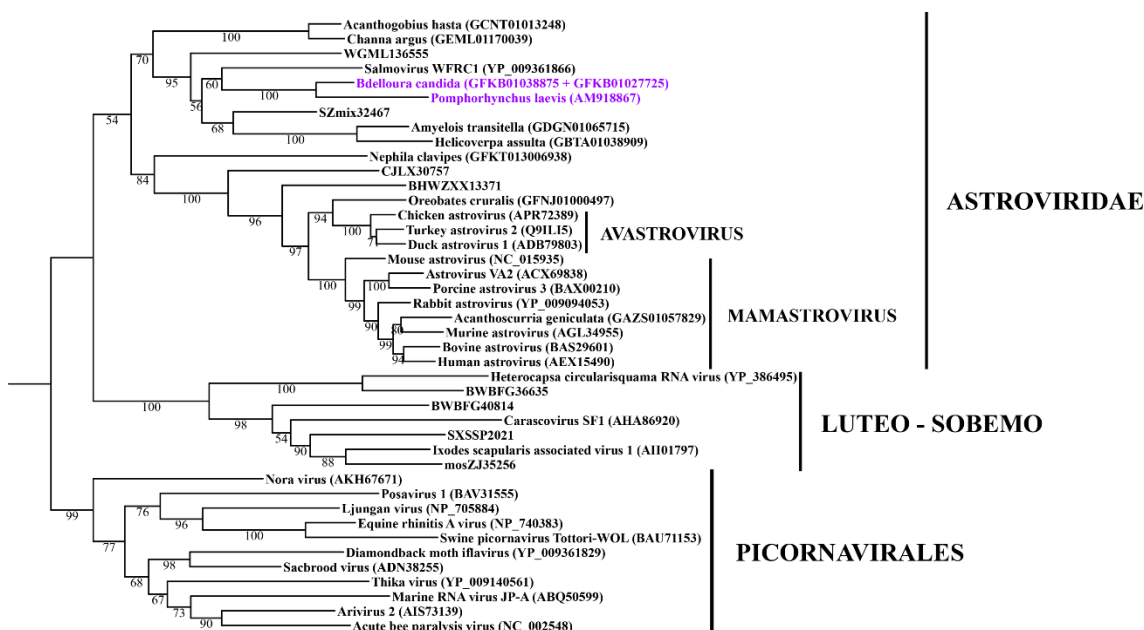


Slika B.42: Filogenetsko drevo virusov klade **Reo**, ki smo jih našli v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«) v programu IQTREE (Nguyen in sod., 2015). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postranice (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično.

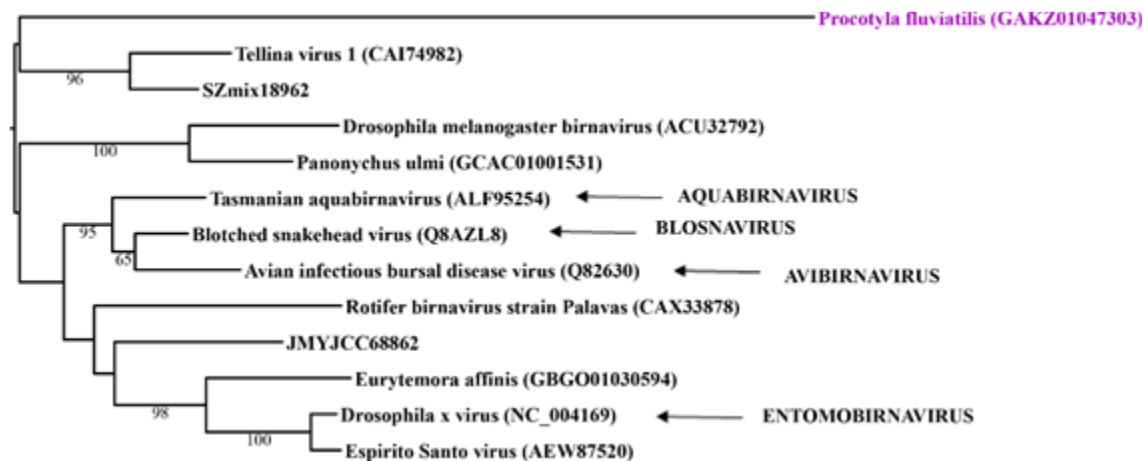


Slika B.43: Filogenetsko drevo virusa klade **Zhaovirus**, ki smo ga odkrili v transkriptomih rakov (Crustacea). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Enakonožci (Isopoda) so označeni z rdečo, ceponožci (Copepoda) z zeleno, postranice (Amphipoda) z modro in deseteronožci (Decapoda) z vijolično.

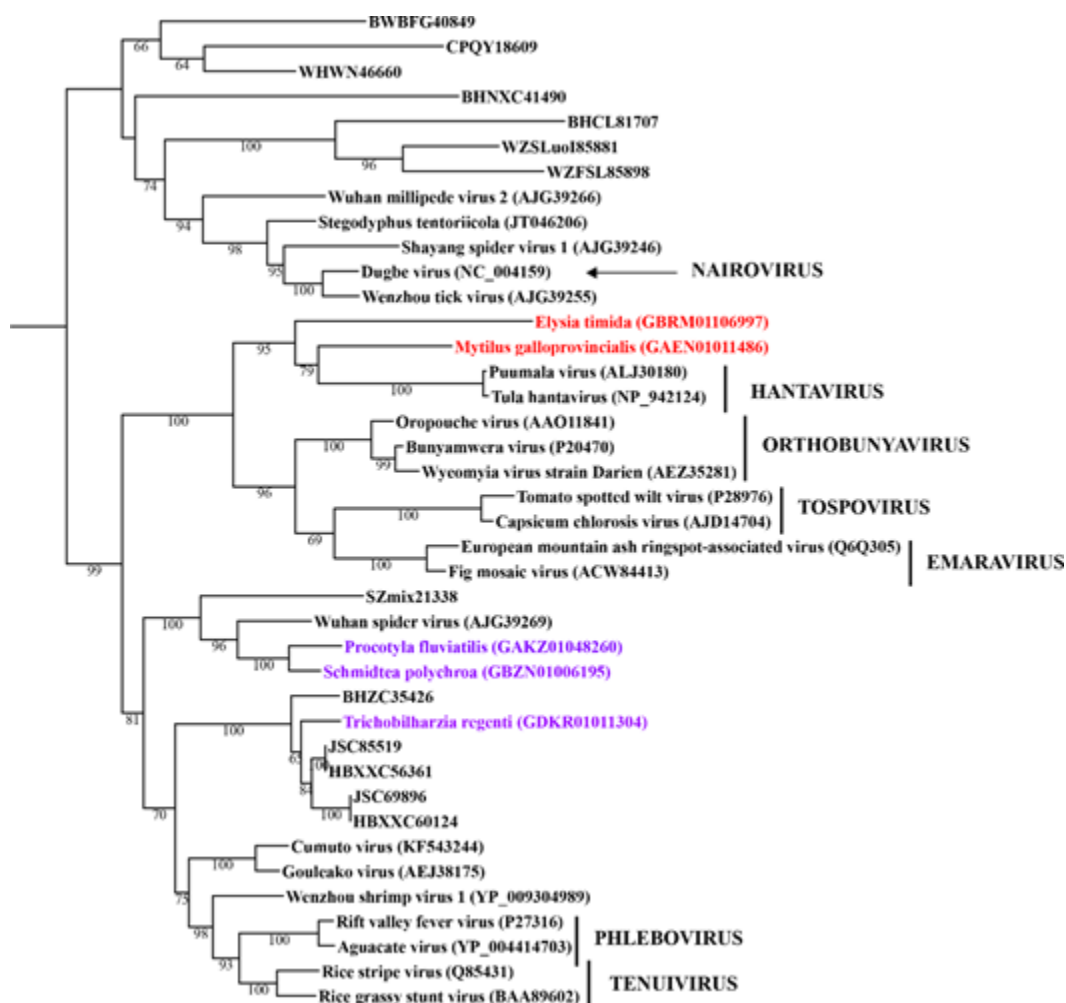
B.5 Skupina Spiralia: mehkužci (Mollusca), ploščati črvi (Platyhelminthes), kolobarniki (Annelida) in kotačniki (Rotifera)



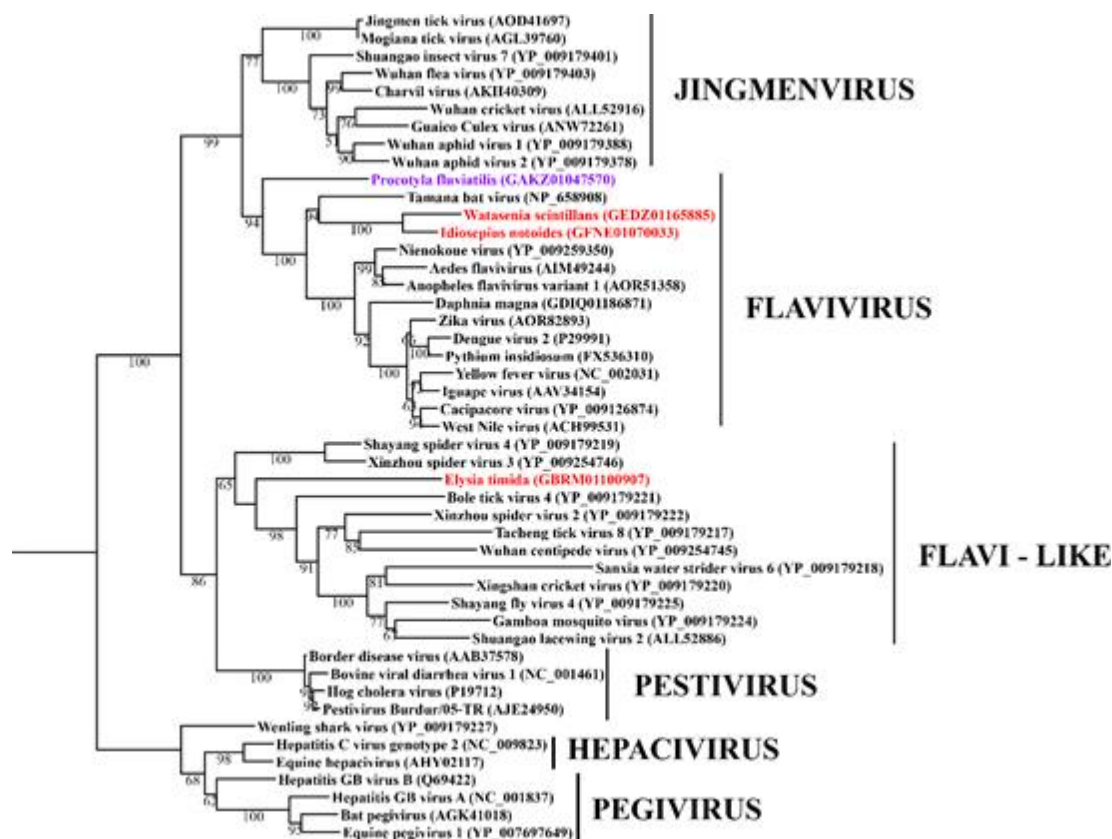
Slika B.44: Filogenetsko drevo virusov klade **Astro**, ki smo jih odkrili v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo.



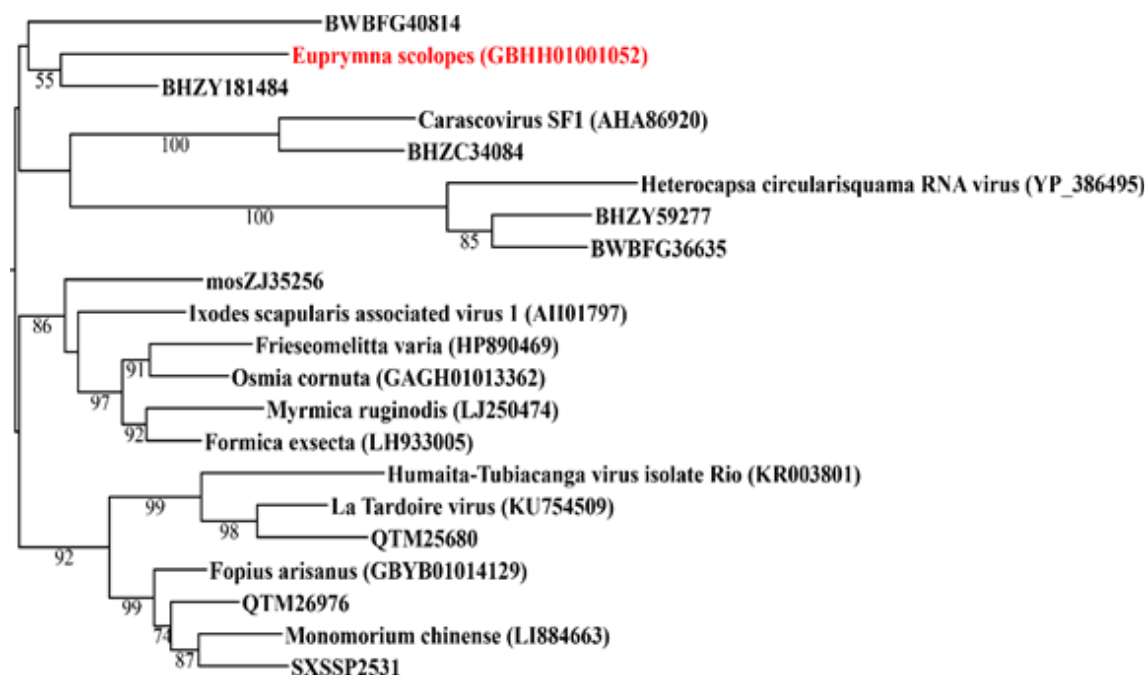
Slika B.45: Filogenetsko drevo virusa klade **Birna**, ki smo jih našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »maximum likelihood«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo.



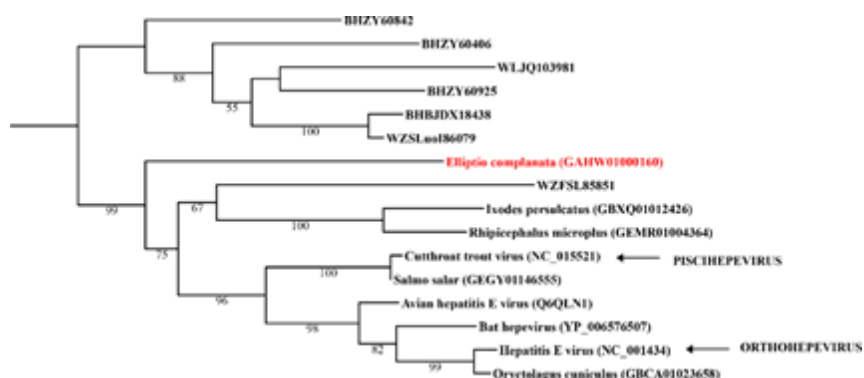
Slika B.46: Filogenetsko drevo virusov klade **Bunya-Arena**, ki smo jih našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo.



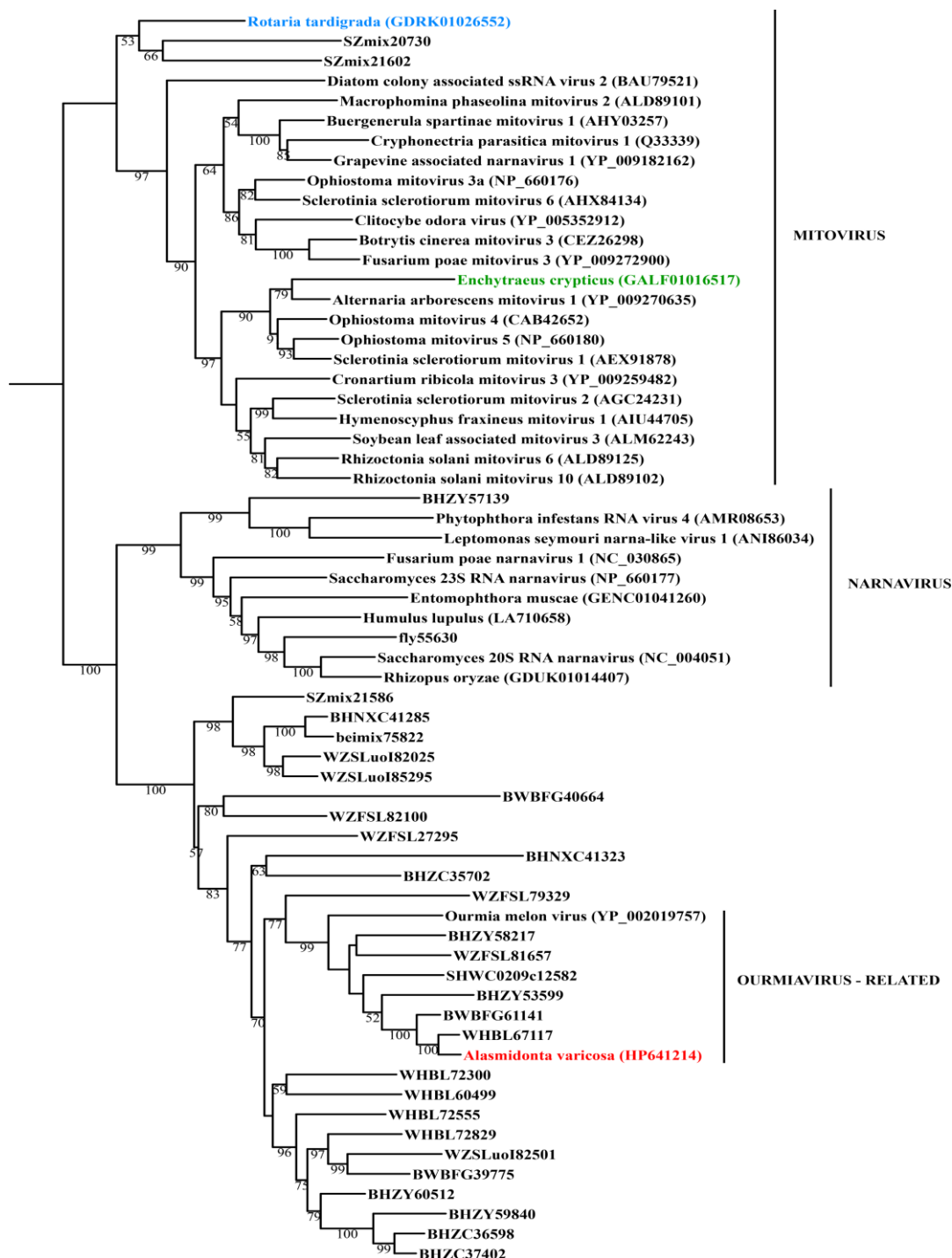
Slika B.47: Filogenetsko drevo virusov klade **Flavi**, ki smo jih našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo.



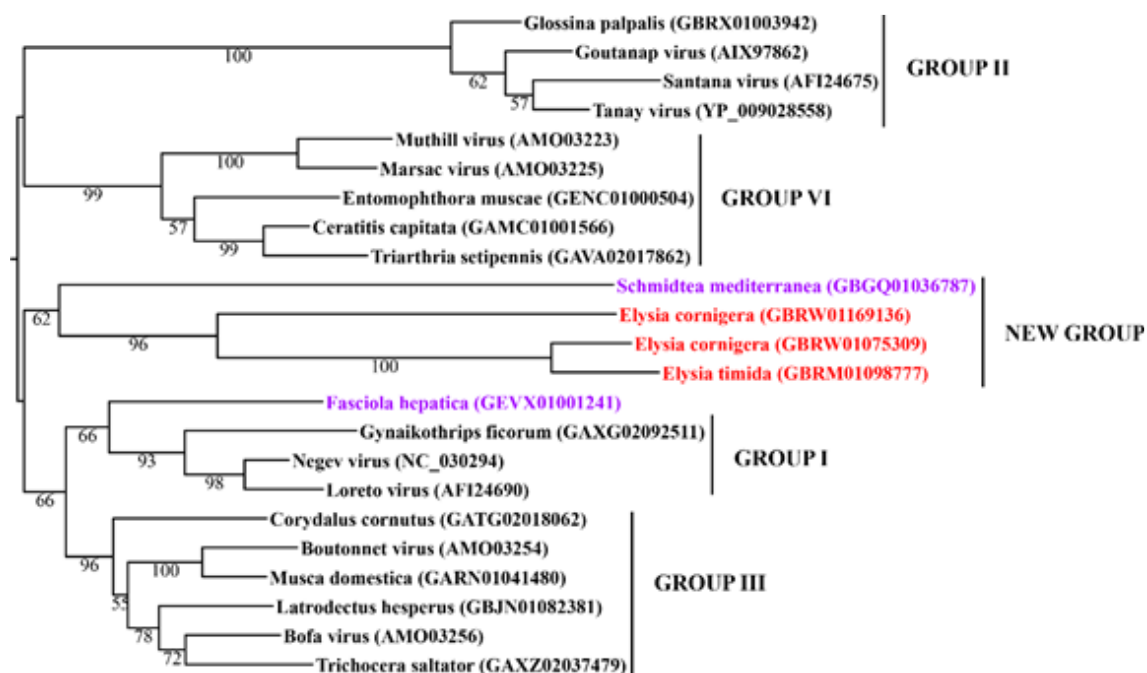
Slika B.48: Filogenetsko drevo virusa klade **Luteo-Sobemo**, ki smo ga našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«) v programu IQTREE (Nguyen in sod., 2015). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo.



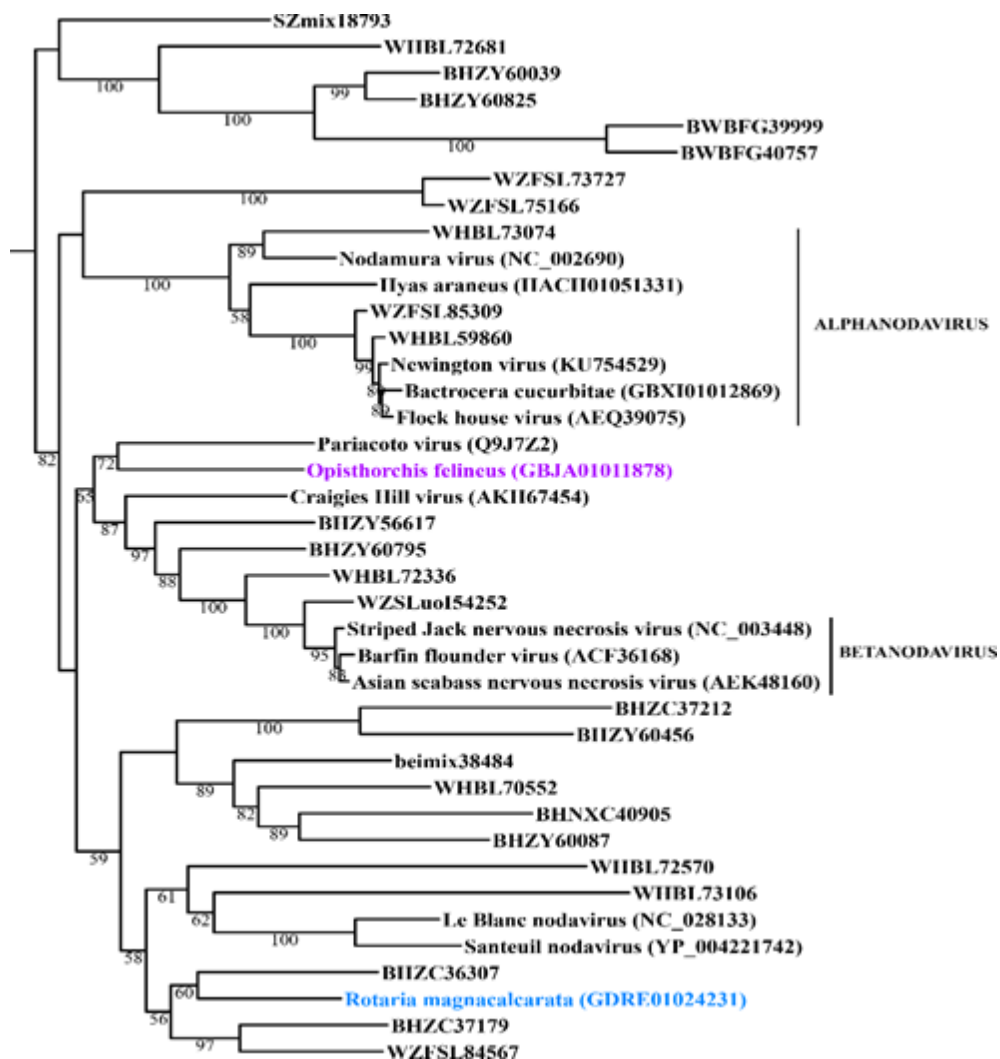
Slika B.49: Filogenetsko drevo virusa družine **Hepeviridae** (klada **Hepe-Virga**), ki smo ga našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«) v programu IQTREE (Nguyen in sod., 2015). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, skupina kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo.



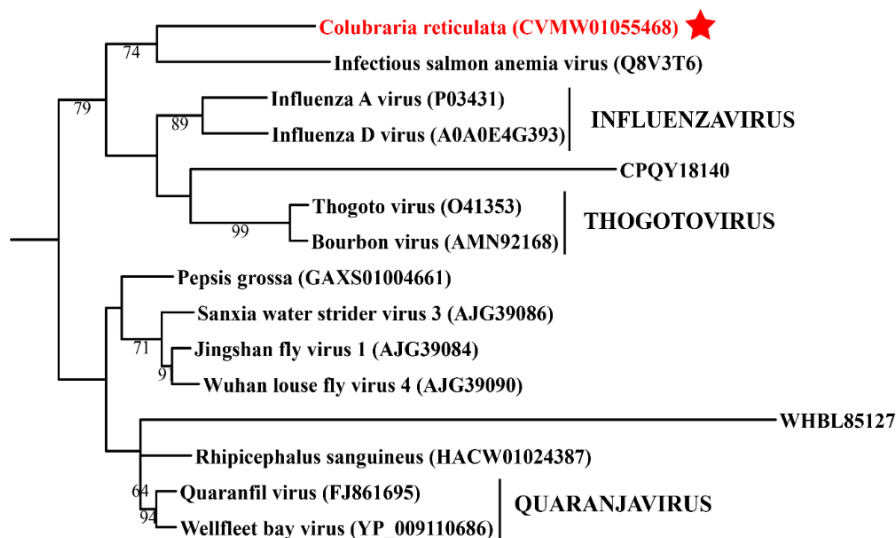
Slika B.50: Filogenetsko drevo virusov družine **Narnaviridae** (klada **Narna-levi**), ki smo jih našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo.



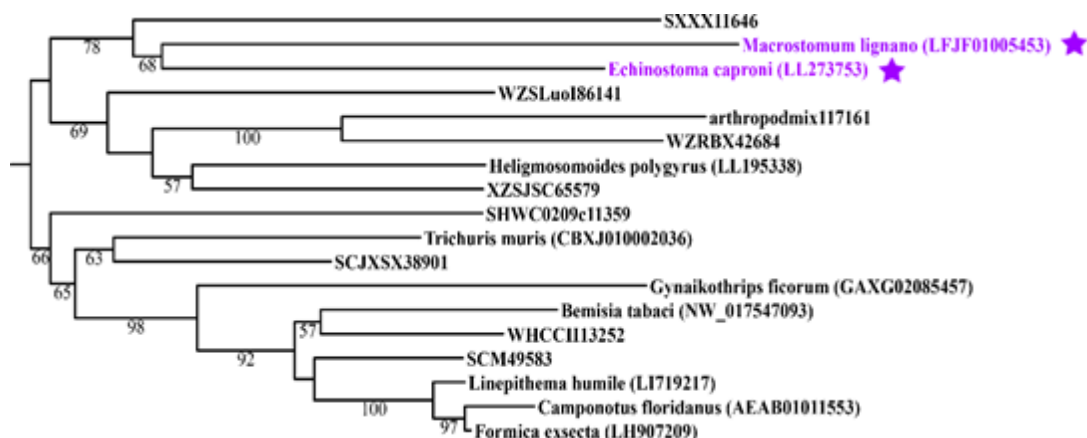
Slika B.51: Filogenetsko drevo virusov skupine **Negevirus** (klada **Hepe-Virga**), ki smo jih našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, skupina kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo. Angleška izraza: new – nova; group: skupina.



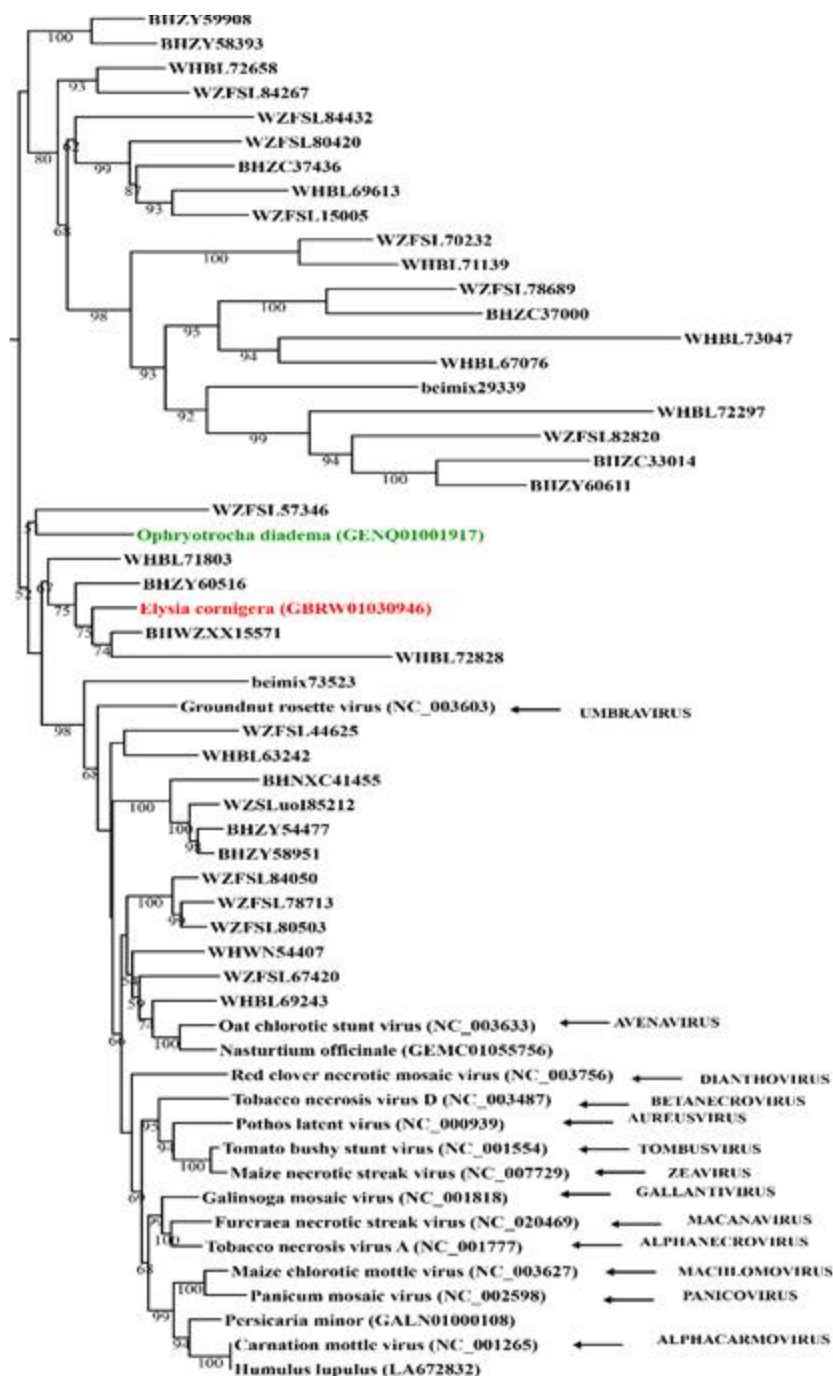
Slika B.52: Filogenetsko drevo virusov družine **Nodaviridae** (klada **Tombus-Noda**), ki smo jih našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, skupina kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo.



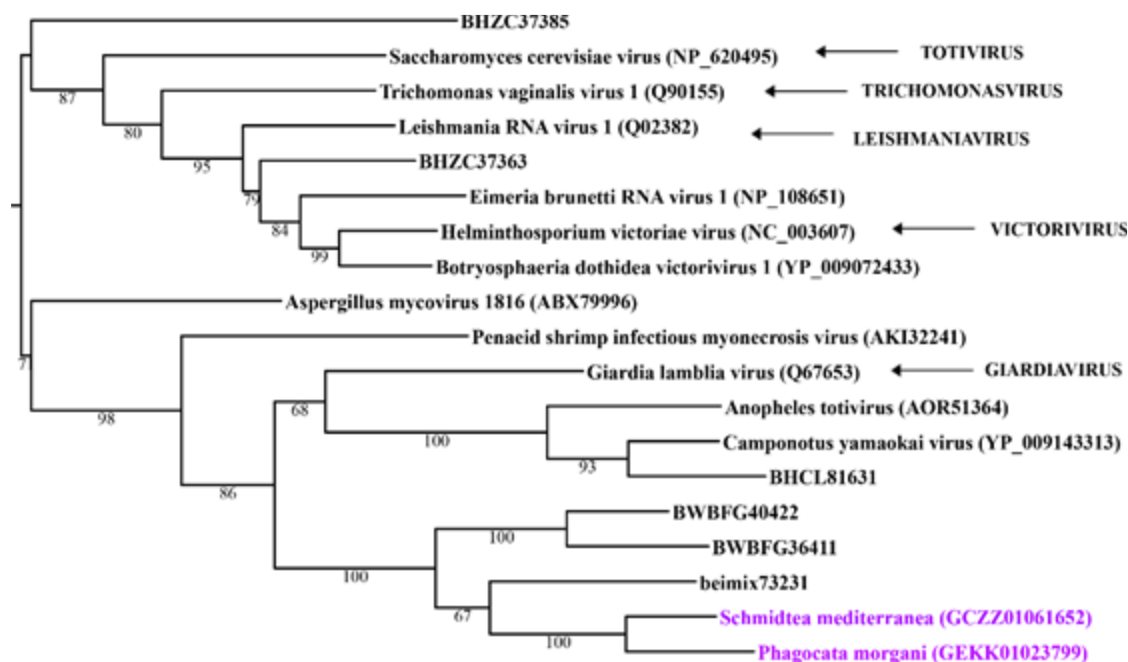
Slika B.53: Filogenetsko drevo virusa klade **Orthomyxo**, ki smo jih našli v genomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo. Endogenizirani virusi so označeni z zvezdico.



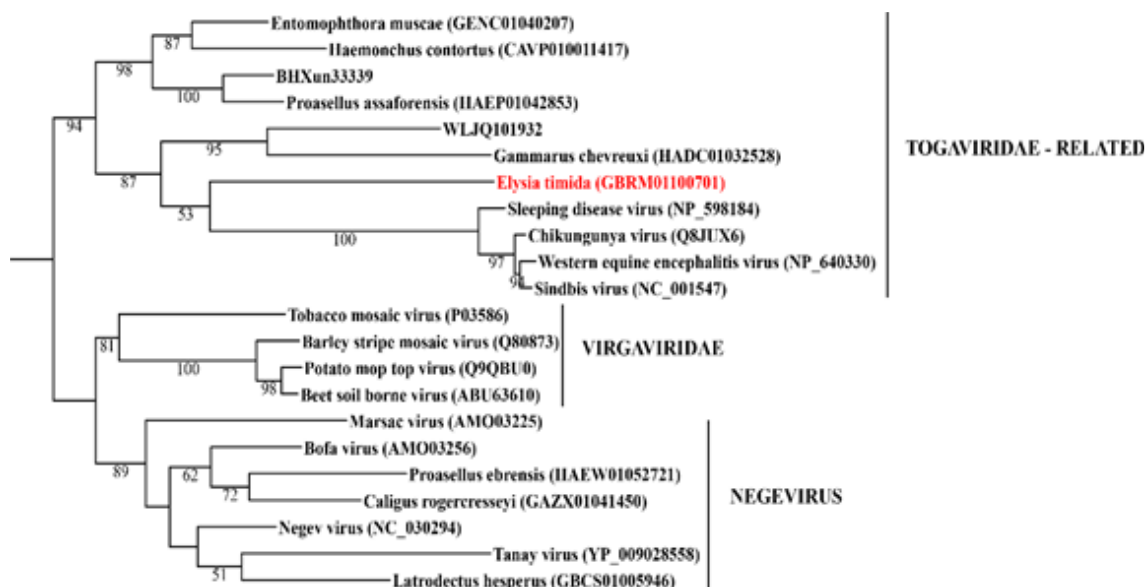
Slika B.54: Filogenetsko drevo virusov klade **Qinvirus**, ki smo ga našli v genomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo. Endogeniziran virus je označen z zvezdico.



Slika B.55: Filogenetsko drevo virusov družine **Tombusviridae** (klada **Tombus-Noda**), ki smo jih našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«) v programu IQTREE (Nguyen in sod., 2015). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo.

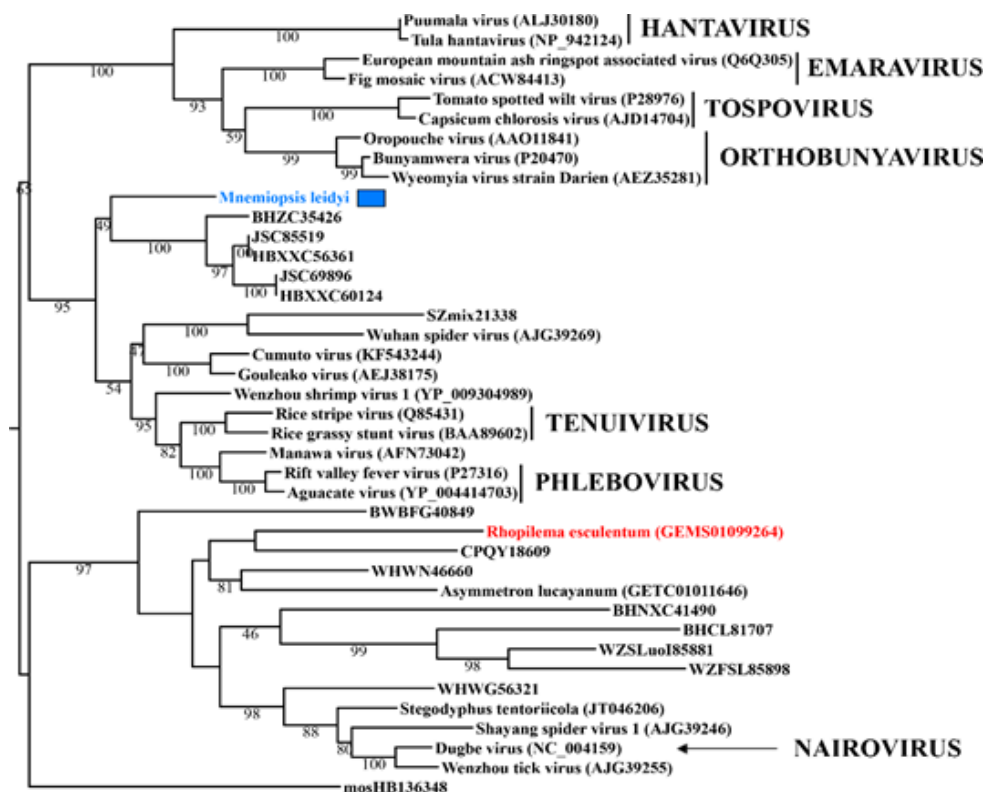


Slika B.56: Filogenetsko drevo virusov klade **Toti-Chryso**, ki smo jih našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo.

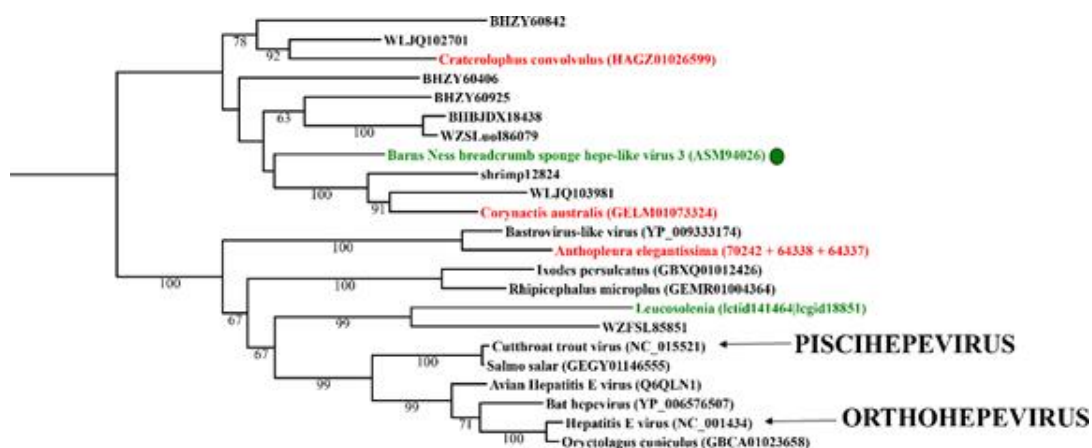


Slika B.57: Filogenetsko drevo virusa družine **Togaviridae** (klada **Hepe-Virga**), ki smo ga našli v transkriptomih skupine Spiralia. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Mehkužci (Mollusca) so označeni z rdečo, ploščati črvi (Platyhelminthes) z vijolično, kolobarniki (Annelida) z zeleno in kotačniki (Rotifera) z modro barvo.

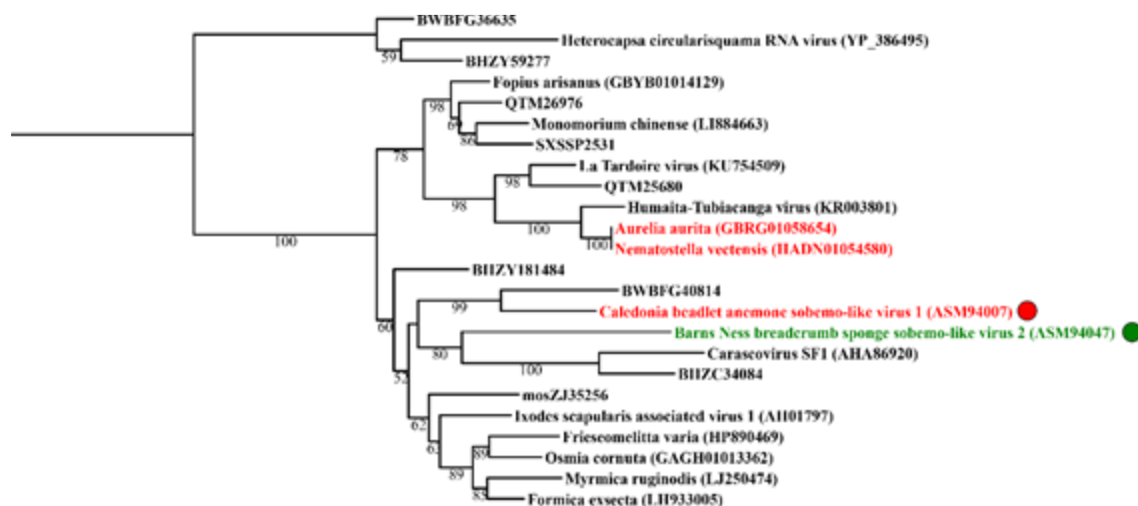
B.6 Najstarejše skupine živali: ožigalkarji (Cnidaria), spužve (Porifera) in rebrače (Ctenophora)



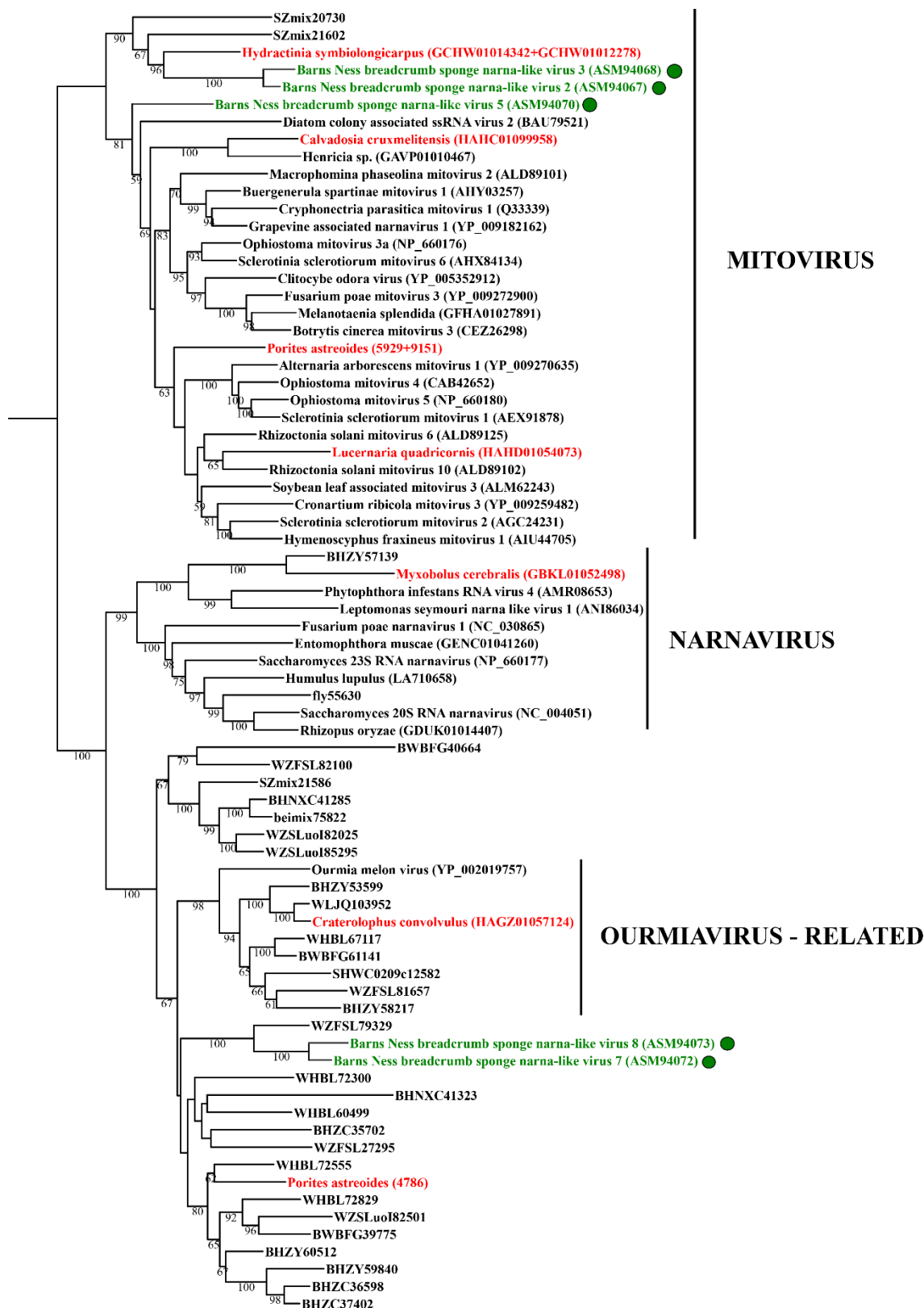
Slika B.58: Filogenetsko drevo virusov klade **Bunya-Arena**, ki smo jih našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo. Virus, ki smo ga našli v transkriptomu *Mnemiopsis leidyi* (GFAT01115010 + GFAT01138919 + GFAT01007784 + GFAT01017202 + GFAT01031184 + GFAT01090215 + GFAT01127526), smo označili s pravokotnikom.



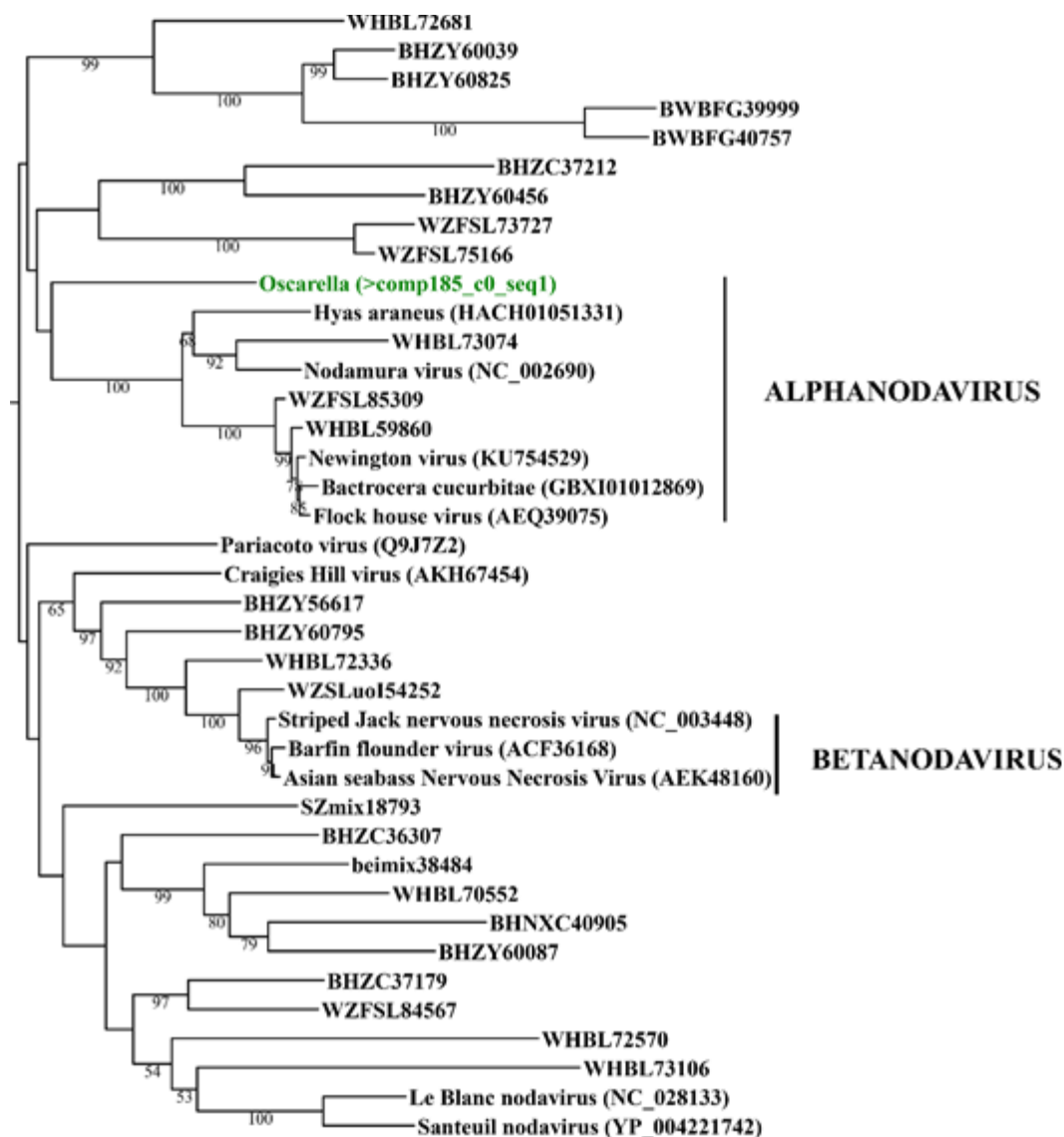
Slika B.59: Filogenetsko drevo virusov družine **Hepeviridae** (klada **Hepe-Virga**), ki smo jih odkrili v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo.



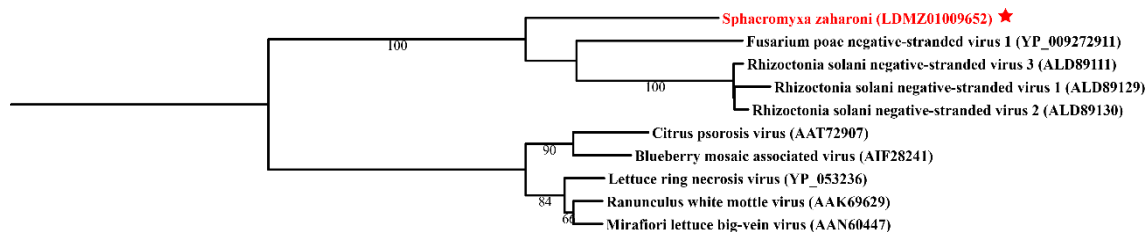
Slika B.60: Filogenetsko drevo virusov klade **Luteo-Sobemo**, ki smo jih našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo. Virusi iz Waldron in sod. (2017) so označeni s krogi.



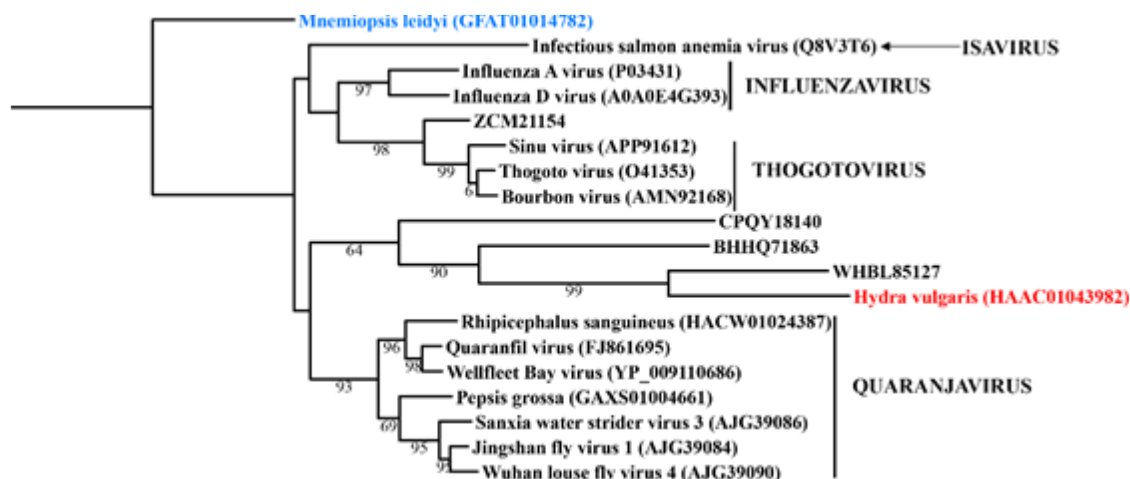
Slika B.61: Filogenetsko drevo virusov družine **Narnaviridae** (klada **Narna-Levi**), ki smo jih našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo. Virusi iz Waldron in sod. (2017) so označeni s krogci.



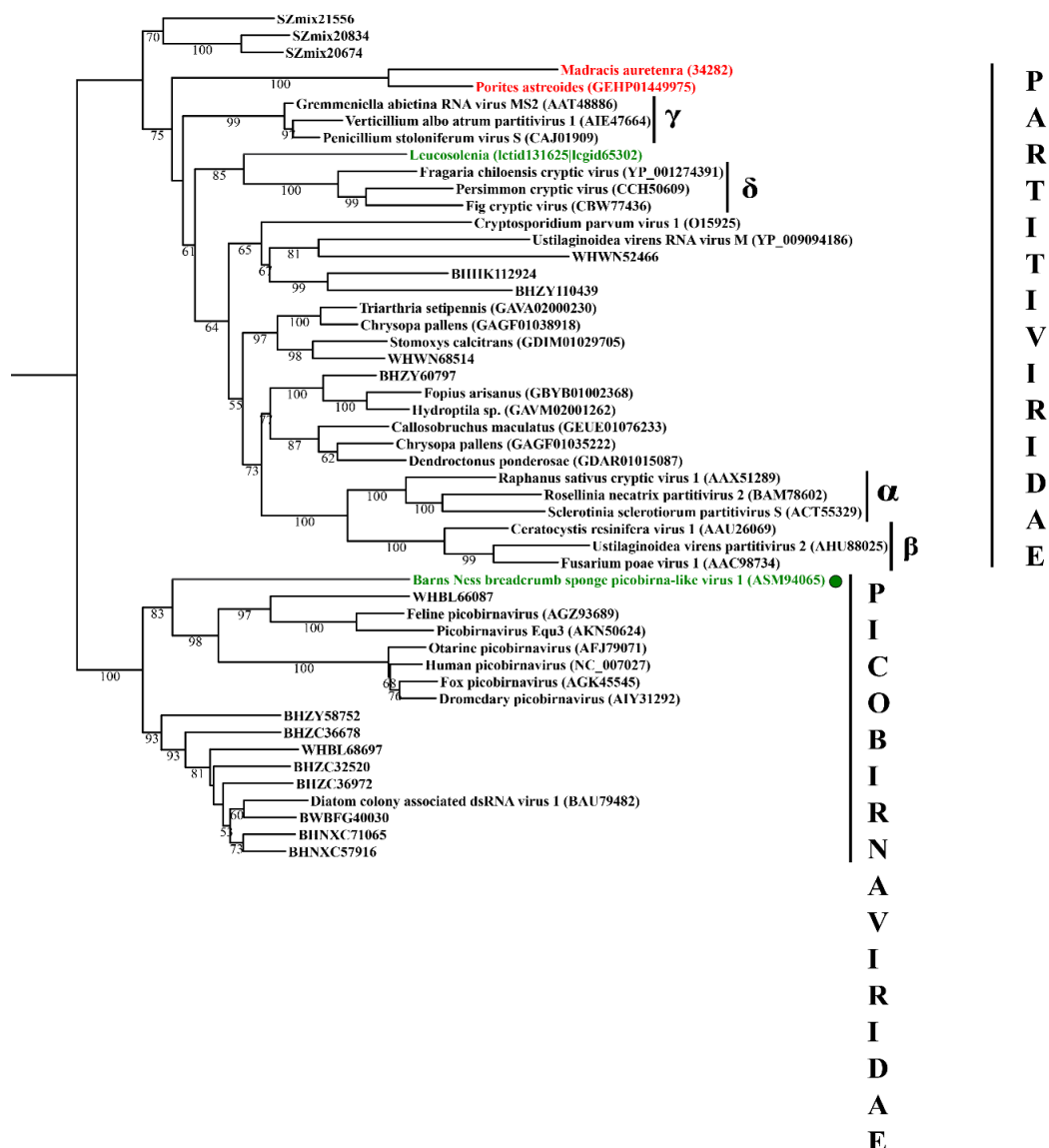
Slika B.62: Filogenetsko drevo virusa družine **Nodaviridae** (klada **Tombus-Noda**), ki smo ga našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo.



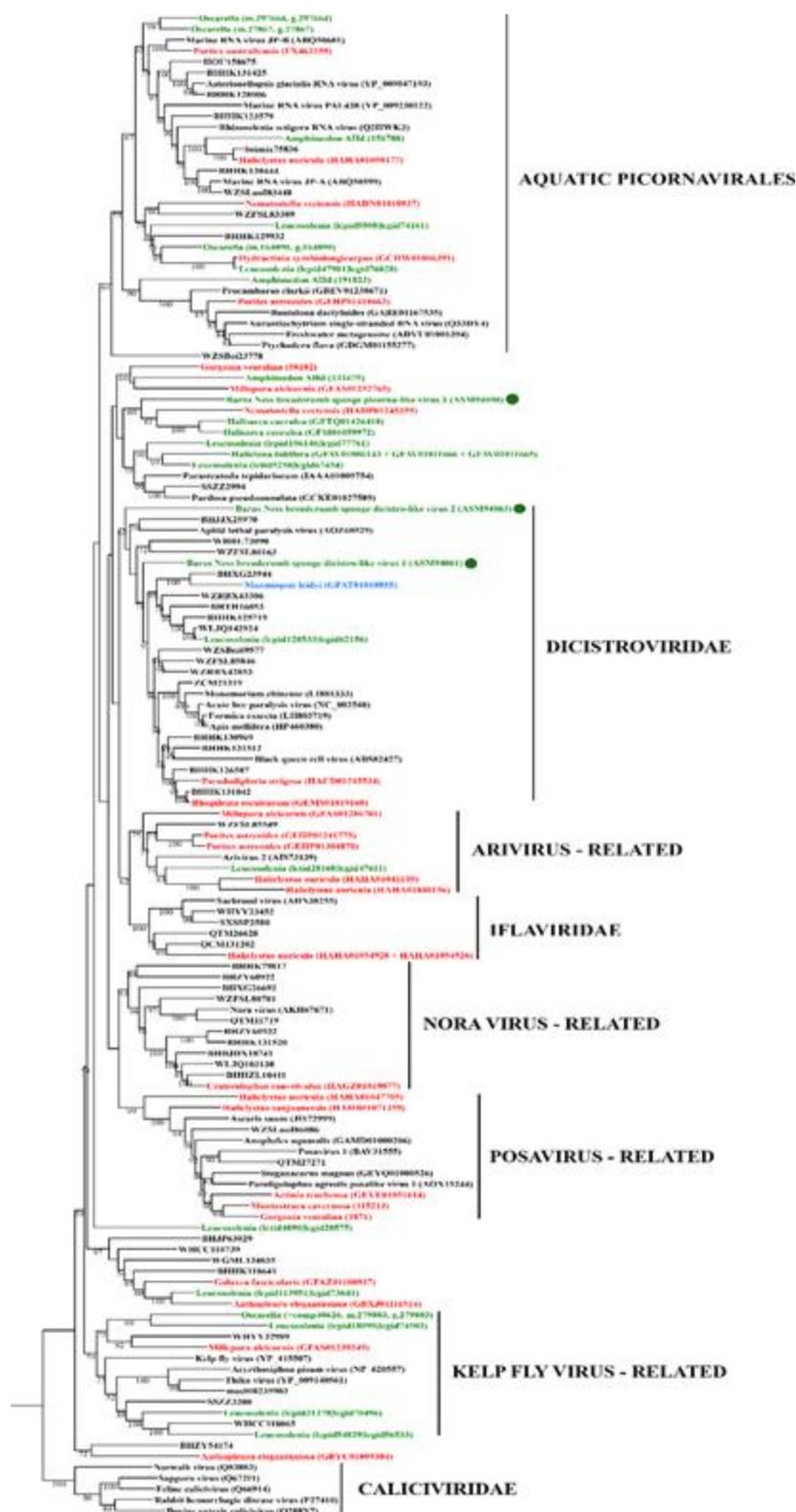
Slika B.63: Filogenetsko drevo virusa klade **Ophioviridae**, ki smo ga našli v genomu najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo. Endogeniziran virus je označen z zvezdico.



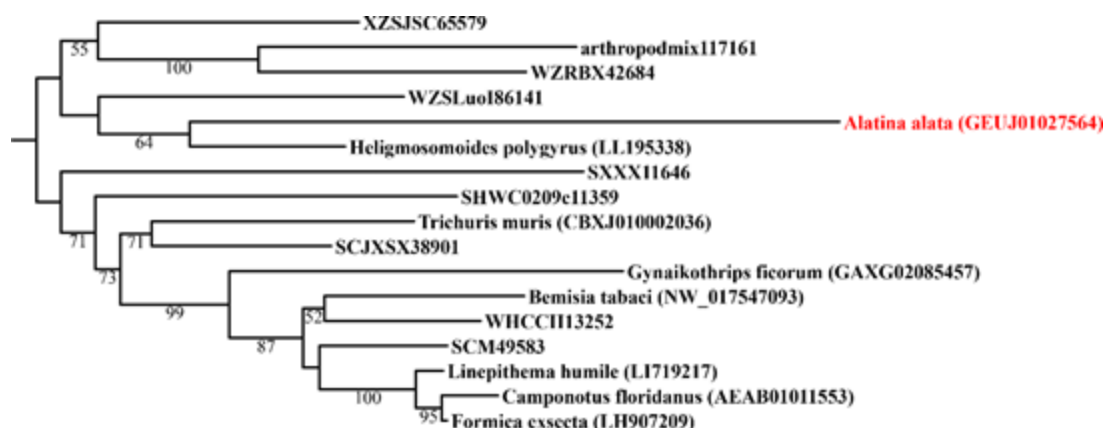
Slika B.64: Filogenetsko drevo virusov klade **Orthomyxo**, ki smo jih našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo.



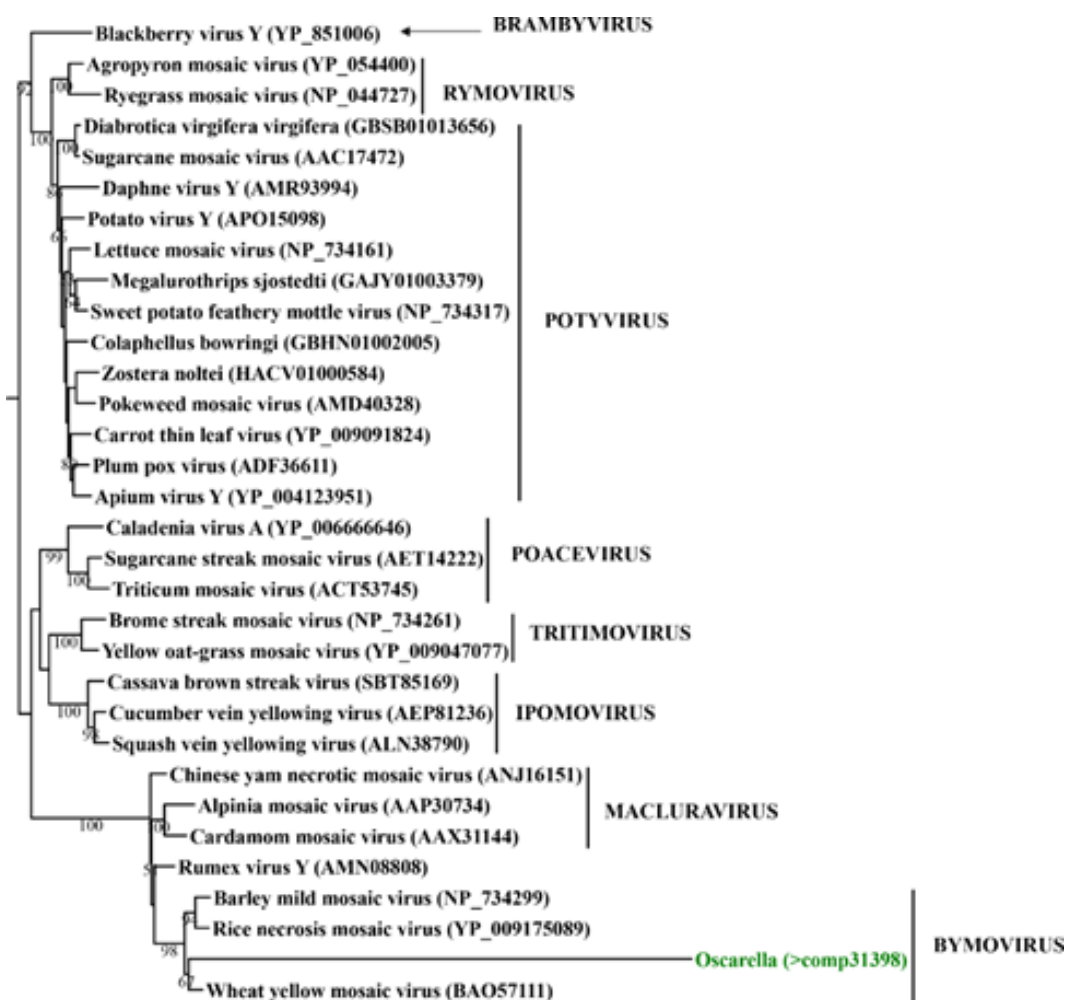
Slika B.65: Filogenetsko drevo virusov klade **Partiti-Picobirna**, ki smo jih našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo. Virusi iz Waldron in sod. (2017) so označeni s krogci.



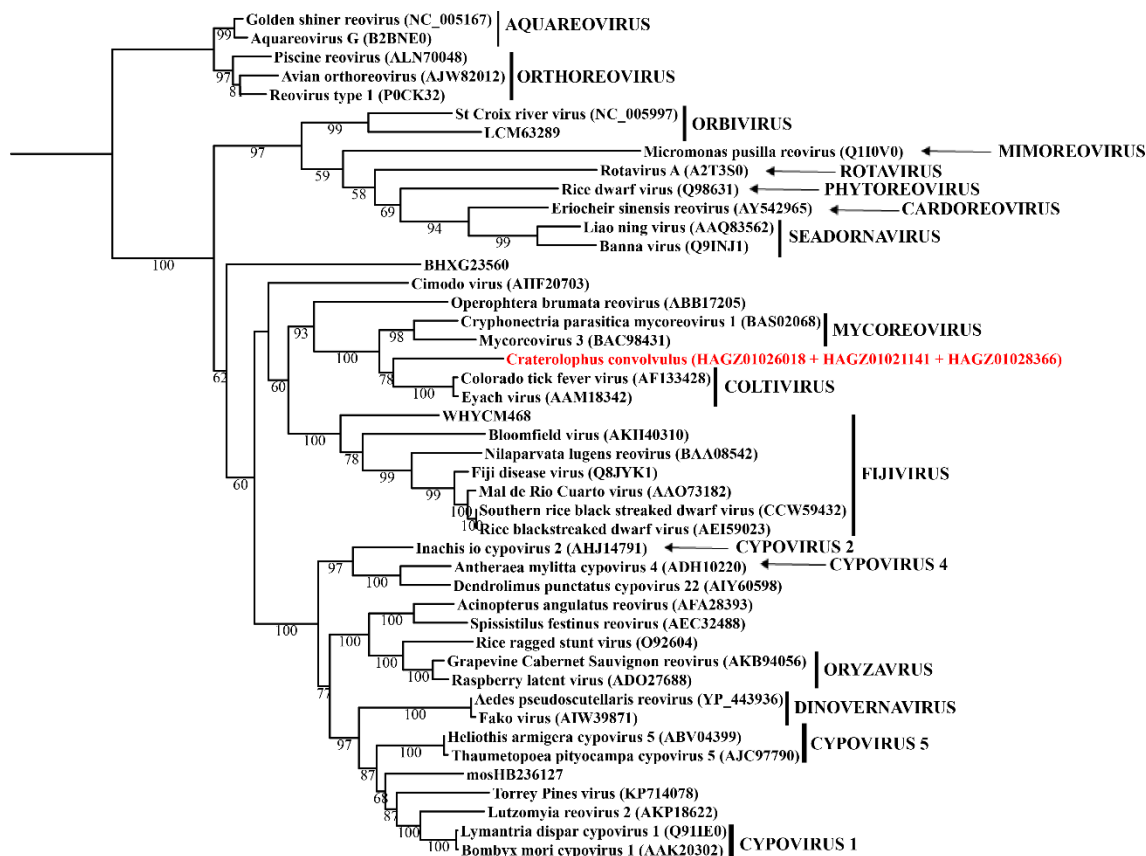
Slika B.66: Filogenetsko drevo virusov klade **Picorna-Calici**, ki smo jih našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo. Virusi iz Waldron in sod. (2017) so označeni s krogci.



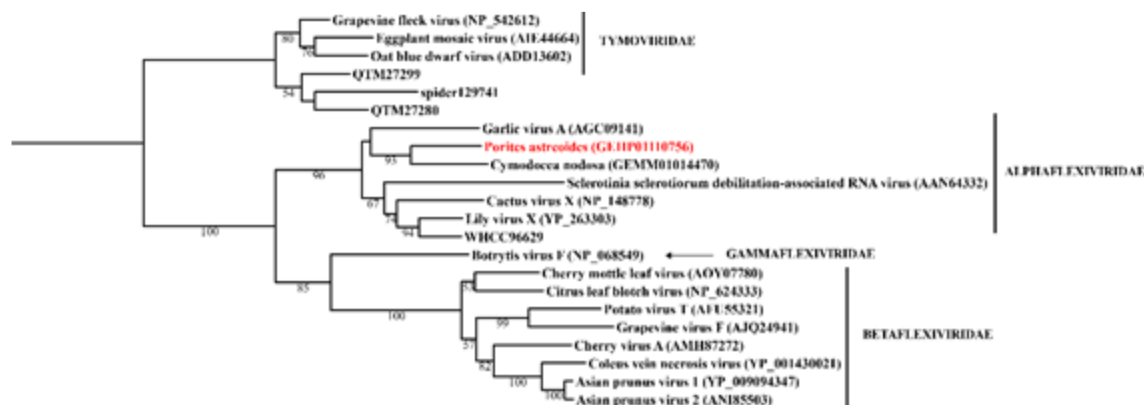
Slika B.67: Filogenetsko drevo virusa klade **Qinvirus**, ki smo ga našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo.



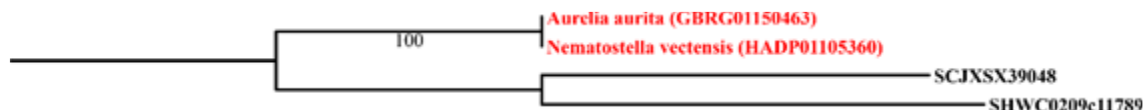
Slika B.68: Filogenetsko drevo virusa klade **Potyviridae**, ki smo ga našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo.



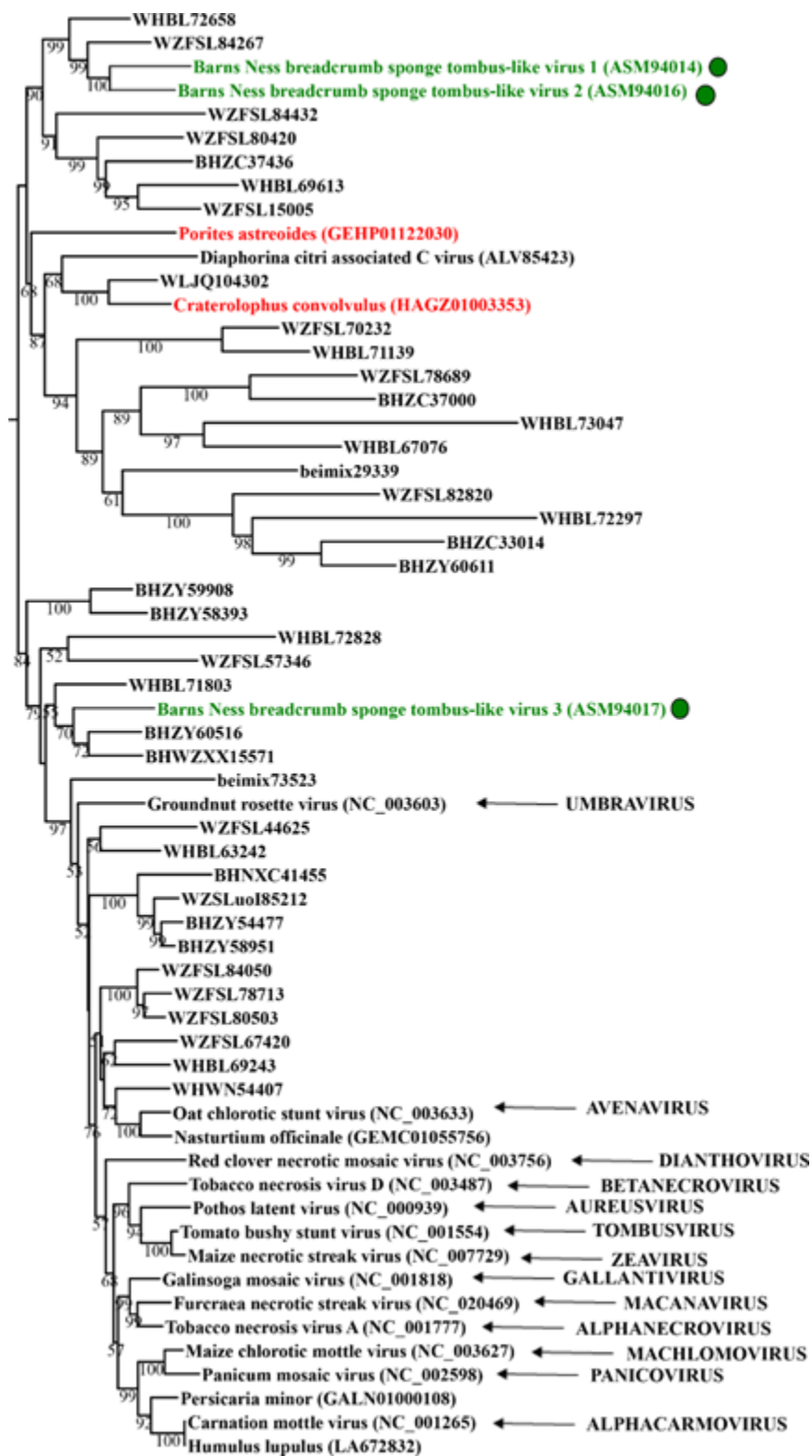
Slika B.69: Filogenetsko drevo virusa klade **Reo**, ki smo ga našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo.



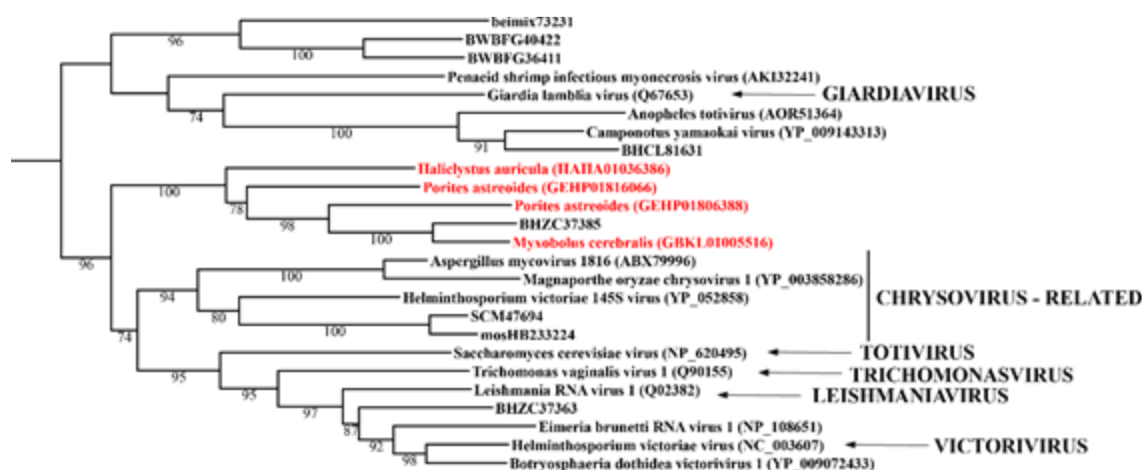
Slika B.70: Filogenetsko drevo virusa redu **Tymovirales** (klada **Hepe-Virga**), ki smo ga našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo.



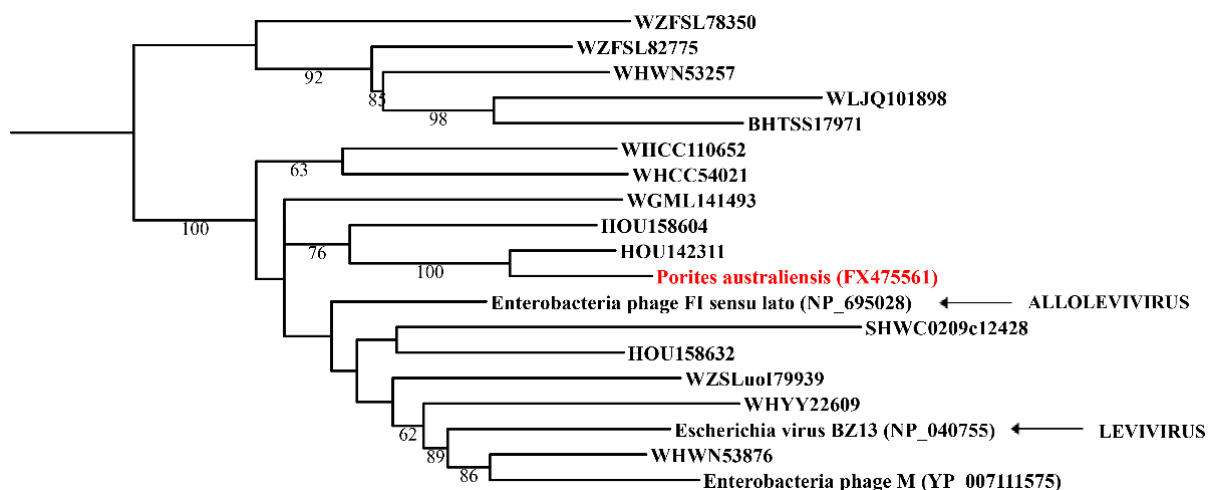
Slika B.71: Filogenetsko drevo virusov klade **Yuevirus**, ki smo jih našli v transkriptomih najstarejših skupin živali (Porifera, Cnidaria, Ctenophora). Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo.



Slika B.72: Filogenetsko drevo virusov družine **Tombusviridae** (klada **Tombus-Noda**), ki smo jih našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo. Virusi iz Waldron in sod. (2017) so označeni s krogi.



Slika B.73: Filogenetsko drevo virusov klade **Toti-Chryso**, ki smo jih našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo.

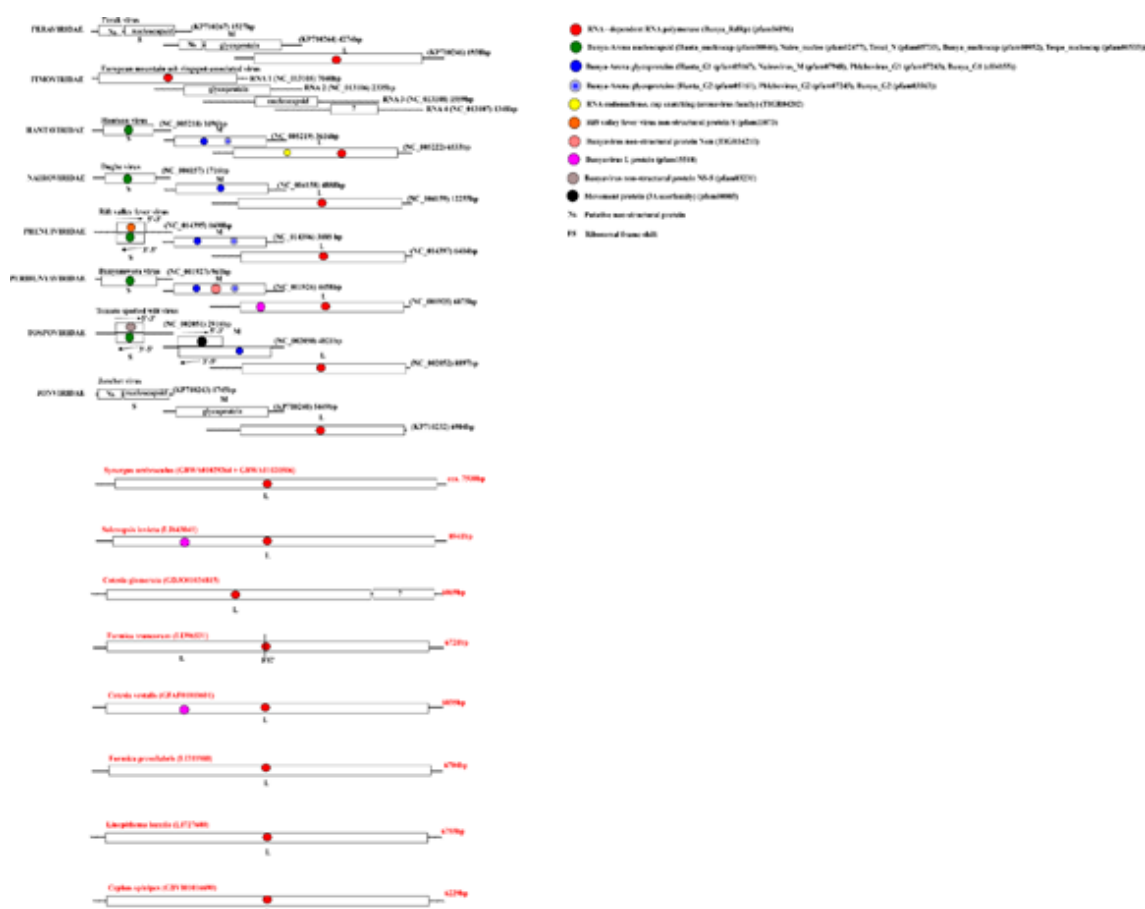


Slika B.74: Filogenetsko drevo virusov družine Leviviridae (klada **Narna-Levi**), ki smo jih našli v transkriptomih najstarejših skupin živali. Drevo je bilo narisano z metodo največje verjetnosti (angl. »*maximum likelihood*«). Ožigalkarji (Cnidaria) so označeni z rdečo, spužve (Porifera) z zeleno, rebrače (Ctenophora) pa z modro barvo.

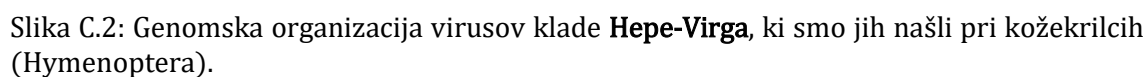
Dodatek C

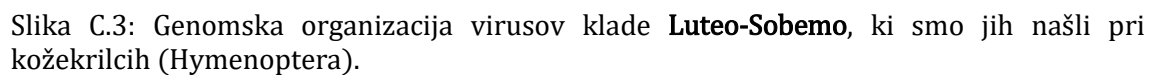
Analiza organizacije novih virusnih genomov

C.1 Kožekrilci (Hymenoptera)



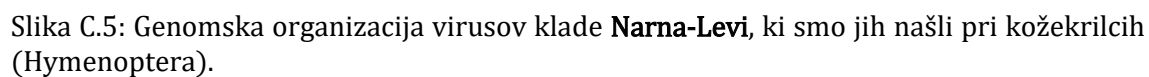
Slika C.1: Genomska organizacija virusov klade **Bunya-Arena**, ki smo jih našli pri kožekrilcih (Hymenoptera). Novi RNA virusi so označeni z rdečo pisavo.

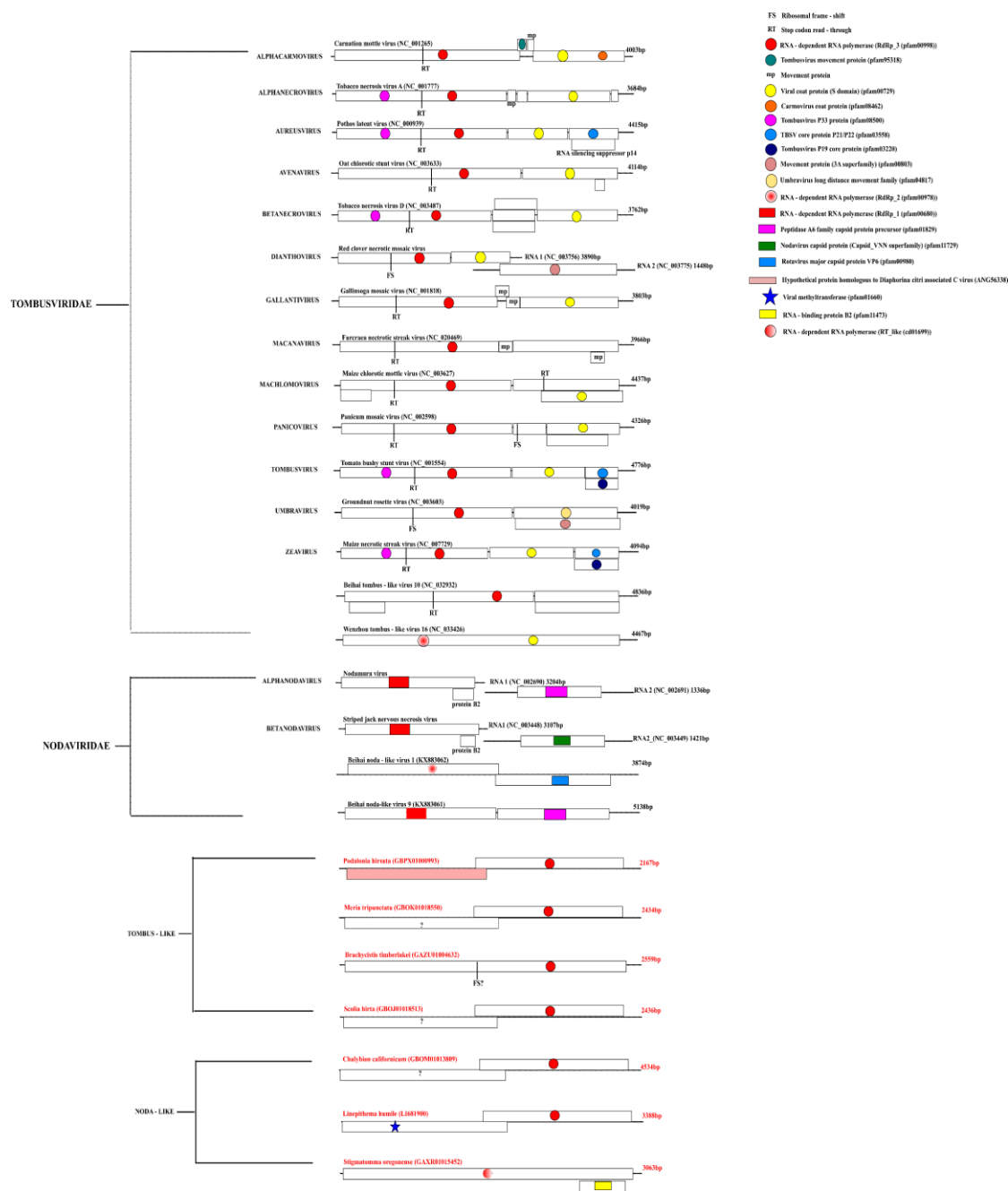






Slika C.4: Genomska organizacija virusov klade **Mono-Chu**, ki smo jih našli pri kožekrilcih (Hymenoptera).

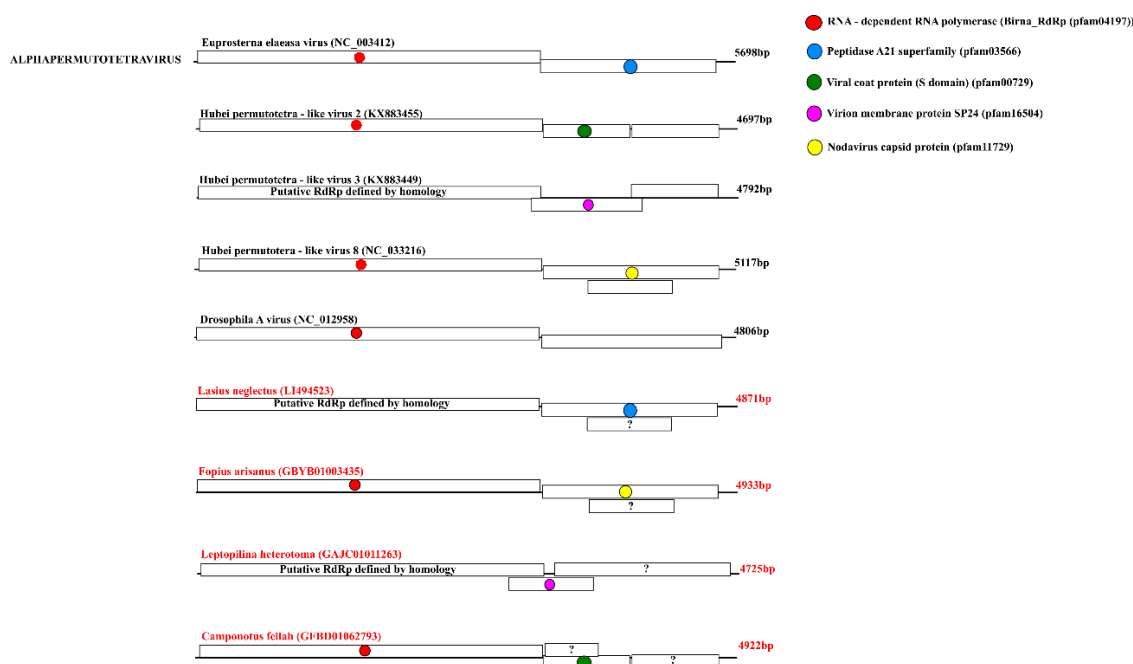




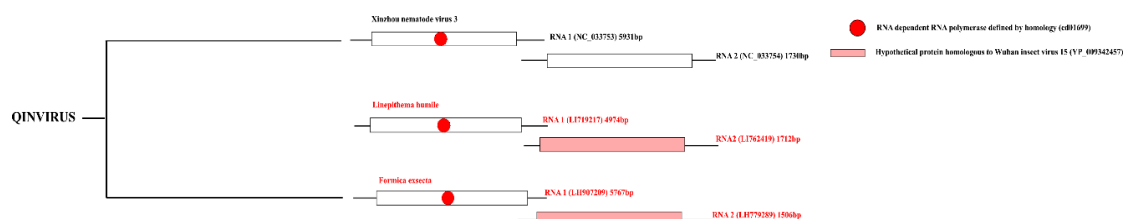
Slika C.6: Genomska organizacija virusov klade **Tombus-Noda**, ki smo jih našli pri kožekrilcih (Hymenoptera).



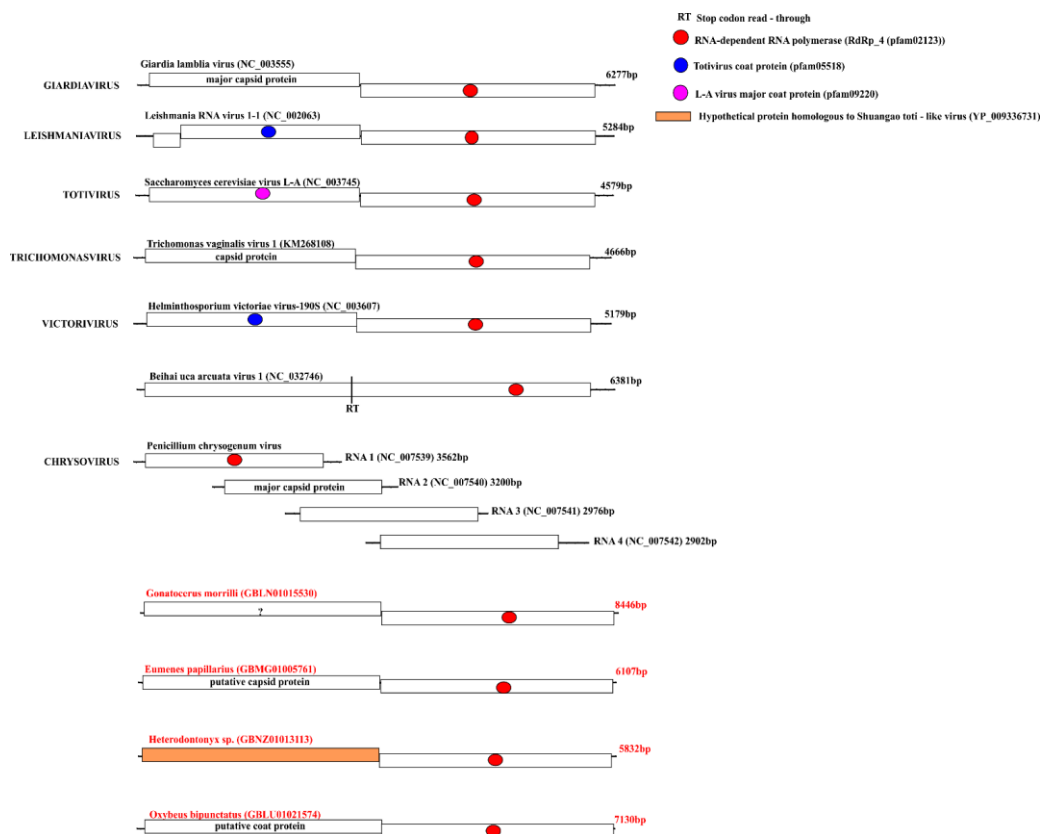
Slika C.7: Genomska organizacija virusov klade **Orthomyxo**, ki smo jih našli pri kožekrilcih (Hymenoptera).



Slika C.8: Genomska organizacija virusov klade **Permutotetra**, ki smo jih našli pri kožekrilcih (Hymenoptera).

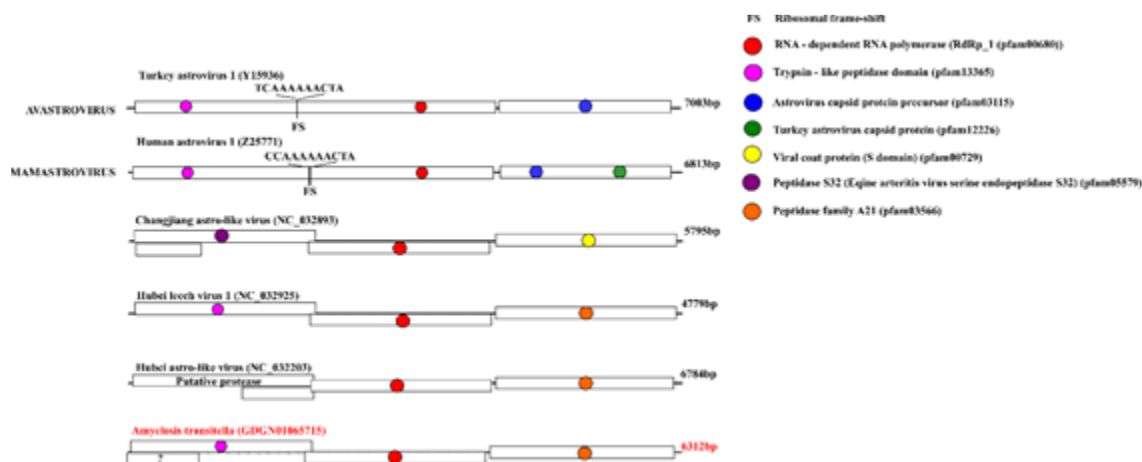


Slika C.10: Genomska organizacija virusov klade **Qinvirus**, ki smo jih našli pri kožekrilcih (Hymenoptera).

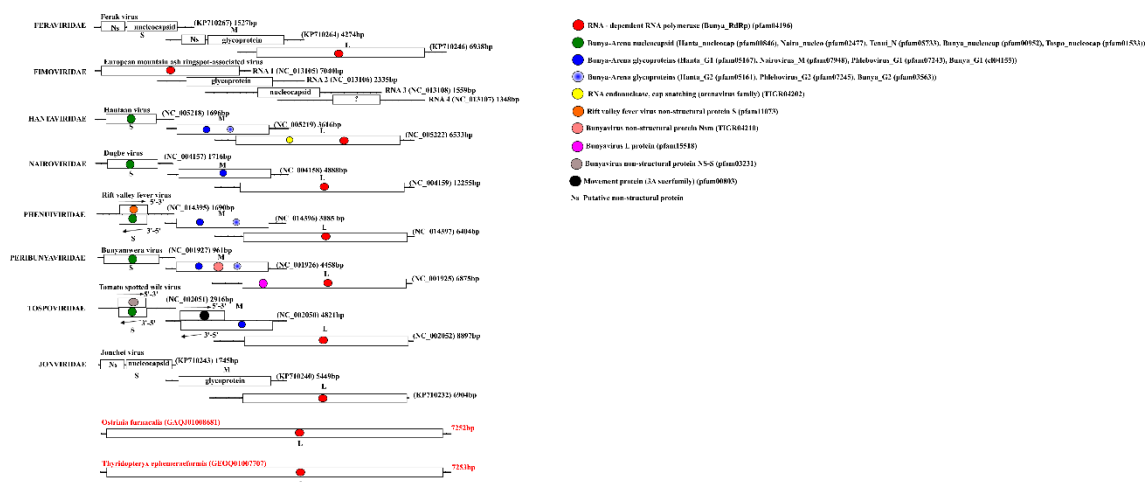


Slika C.11: Genomska organizacija virusov klade **Toti-Chryso**, ki smo jih našli pri kožekrilcih (Hymenoptera).

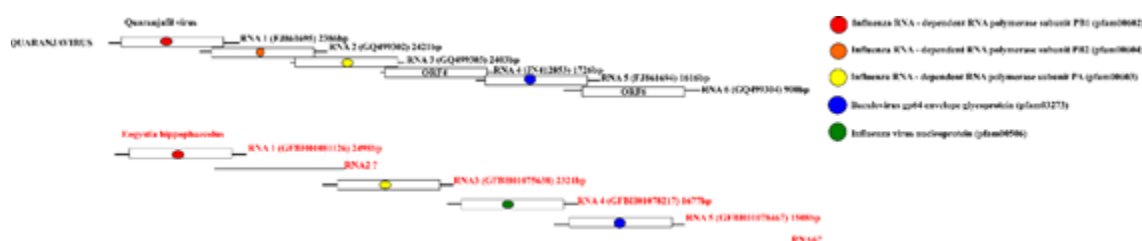
C.2 Metulji (Lepidoptera)



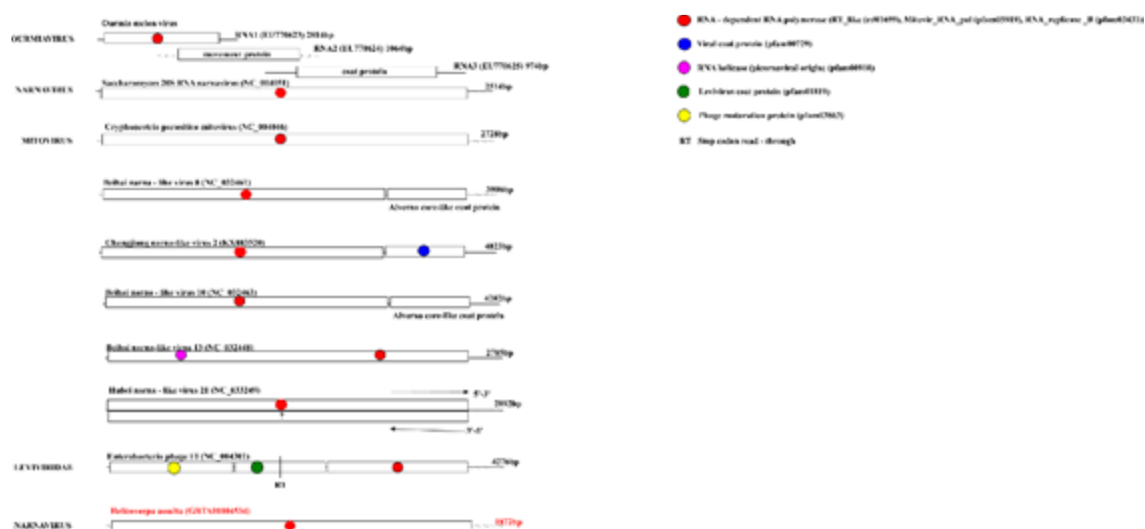
Slika C.12: Genomska organizacija virusov klade **Astro**, ki smo jih našli pri metuljih in veščah (Lepidoptera). Novi RNA virusi so označeni z rdečo pisavo.



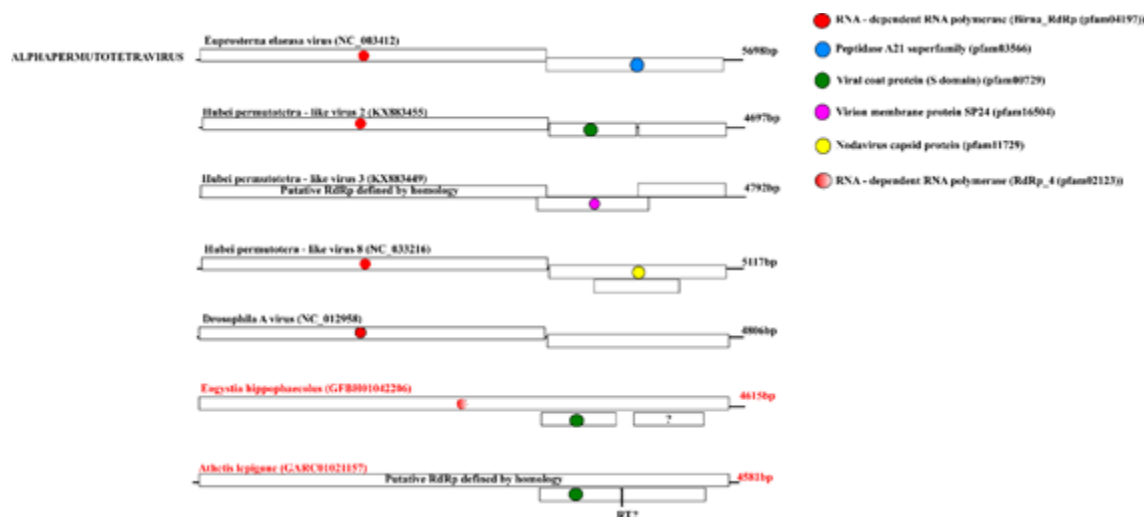
Slika C.13: Genomska organizacija virusov klade **Bunya-Arena**, ki smo jih našli pri metuljih in veščah (Lepidoptera).



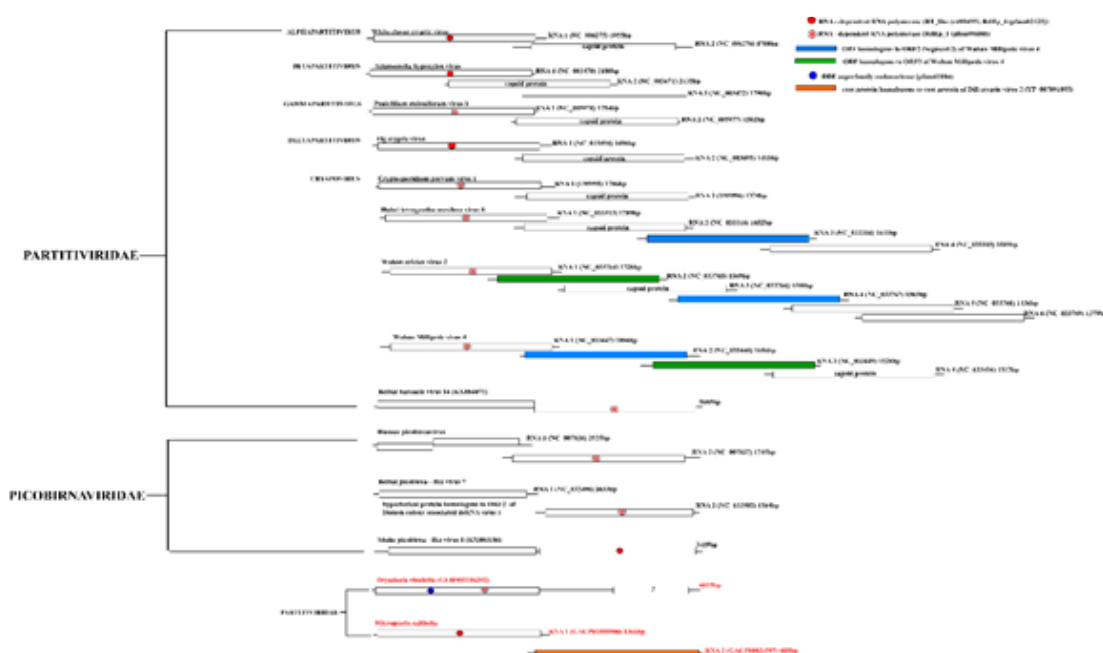
Slika C.14: Genomska organizacija virusov klade **Orthomyxo**, ki smo jih našli pri metuljih in veščah (Lepidoptera).



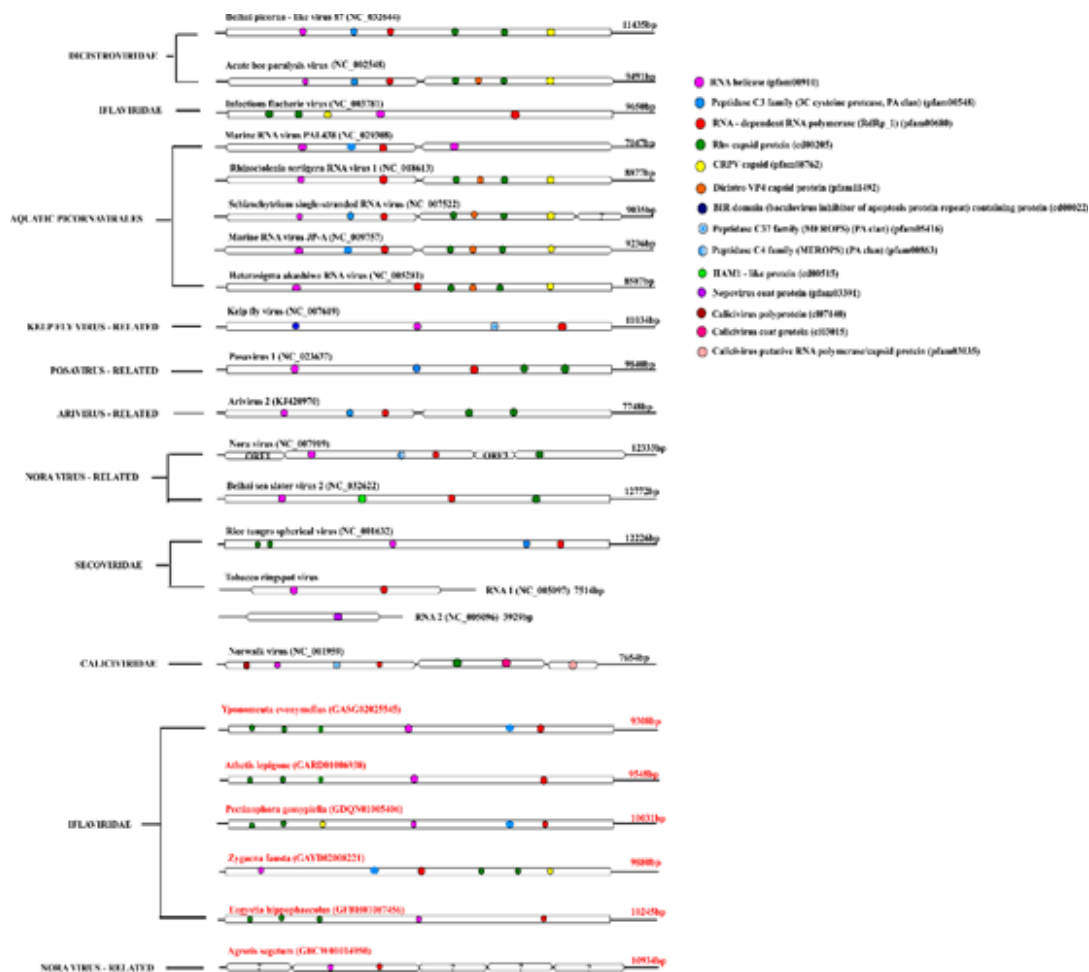
Slika C.15: Genomska organizacija virusov klade **Narna-Levi**, ki smo jih našli pri metuljih in veččah (Lepidoptera).



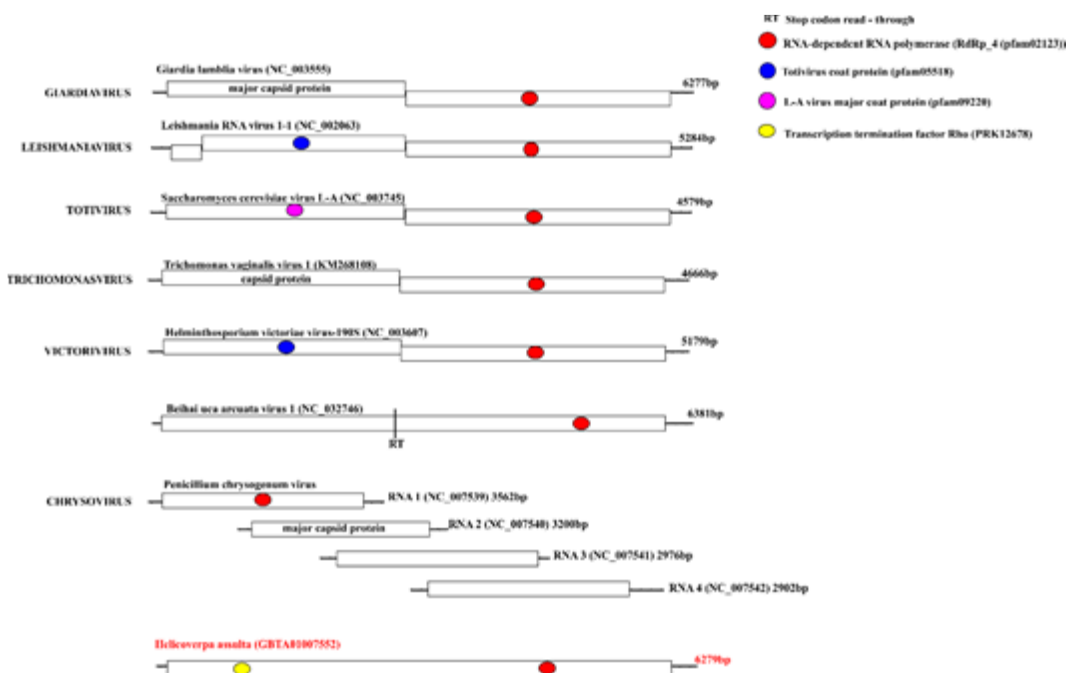
Slika C.16: Genomska organizacija virusov klade **Permutotetra**, ki smo jih našli pri metuljih in veččah (Lepidoptera).



Slika C.17: Genomska organizacija virusov klade **Partiti-Picobirna**, ki smo jih našli pri metuljih in veččah (Lepidoptera).

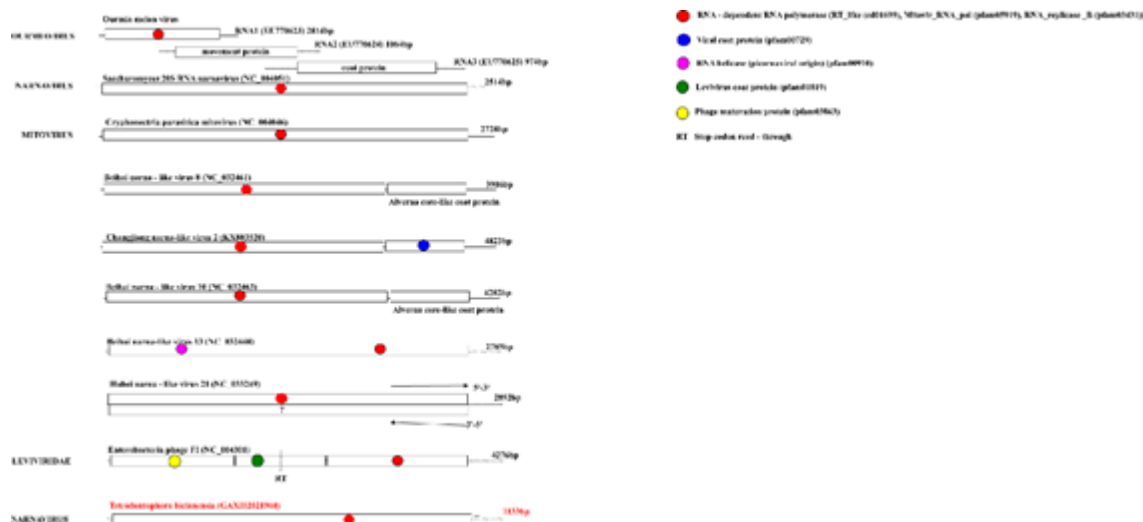


Slika C.18: Genomska organizacija virusov klade **Picornia-Calici**, ki smo jih našli pri metuljih in veščah (Lepidoptera).



Slika C.19: Genomska organizacija virusov klade **Toti-Chryso**, ki smo jih našli pri metuljih in veččah (Lepidoptera).

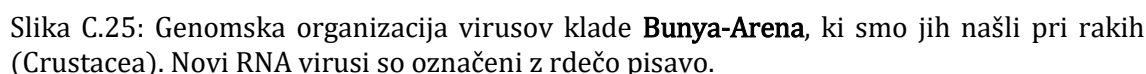
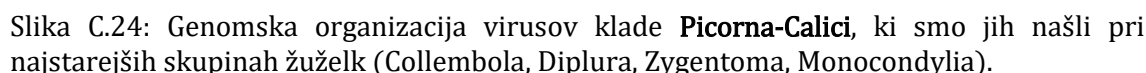
C.3 Najstarejše skupine žuželk: skakači (Collembola), dvorepke (Diplura), ščetinorepke (Zygentoma), Monocondylia

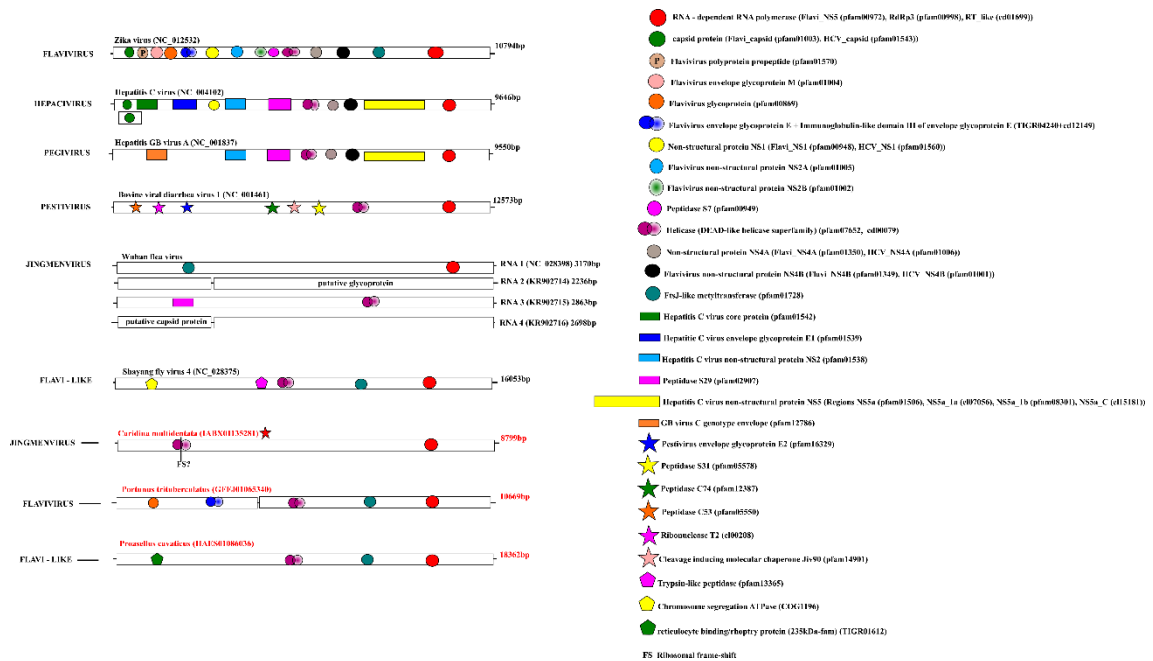


Slika C.20: Genomska organizacija virusov klade **Narna-Levi**, ki smo jih našli pri najstarejših skupinah žuželk (Collembola, Diplura, Zygentoma, Monocondylia). Novi RNA virusi so označeni z rdečo pisavo.



Slika C.21: Genomska organizacija virusov klade **Hepe-Virga**, ki smo jih našli pri najstarejših skupinah žuželk (Collembola, Diplura, Zygentoma in Monocondylia).





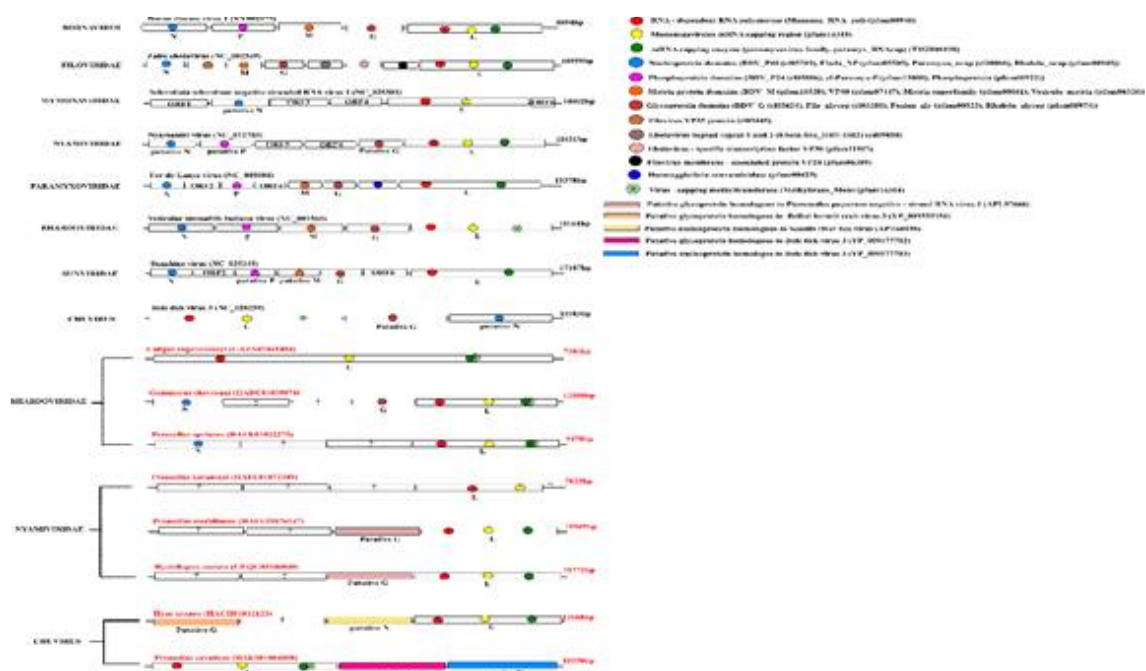
Slika C.26: Genomska organizacija virusov klade **Flavi**, ki smo jih našli pri rakih (Crustacea). Endogeniziran virus je označen z zvezdico.



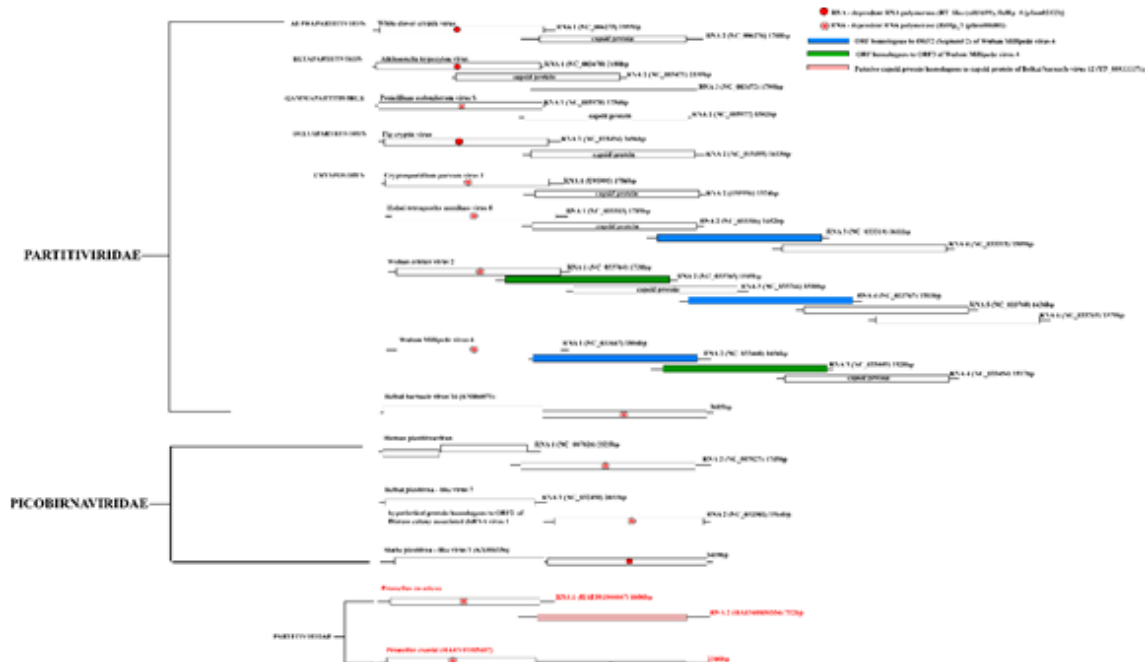
Slika C.27: Genomska organizacija virusov klade **Hepe-Virga**, ki smo jih našli pri rakih (Crustacea).



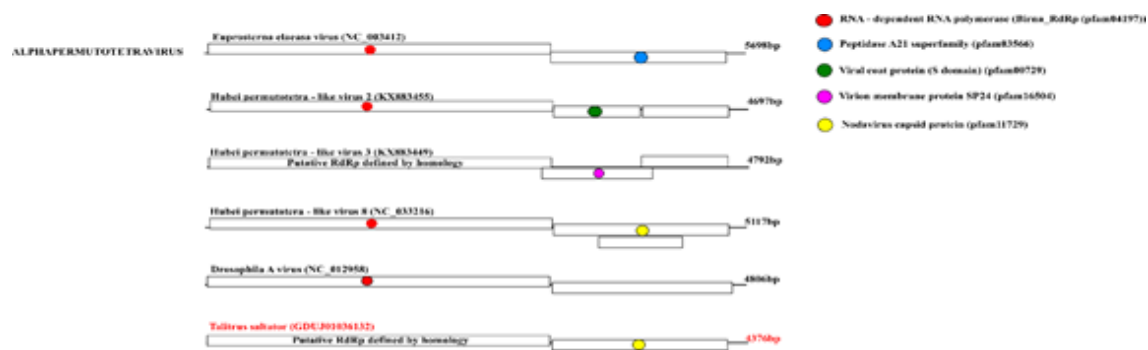
Slika C.28: Genomska organizacija virusov klade **Orthomyxo**, ki smo jih našli pri rakih (Crustacea).



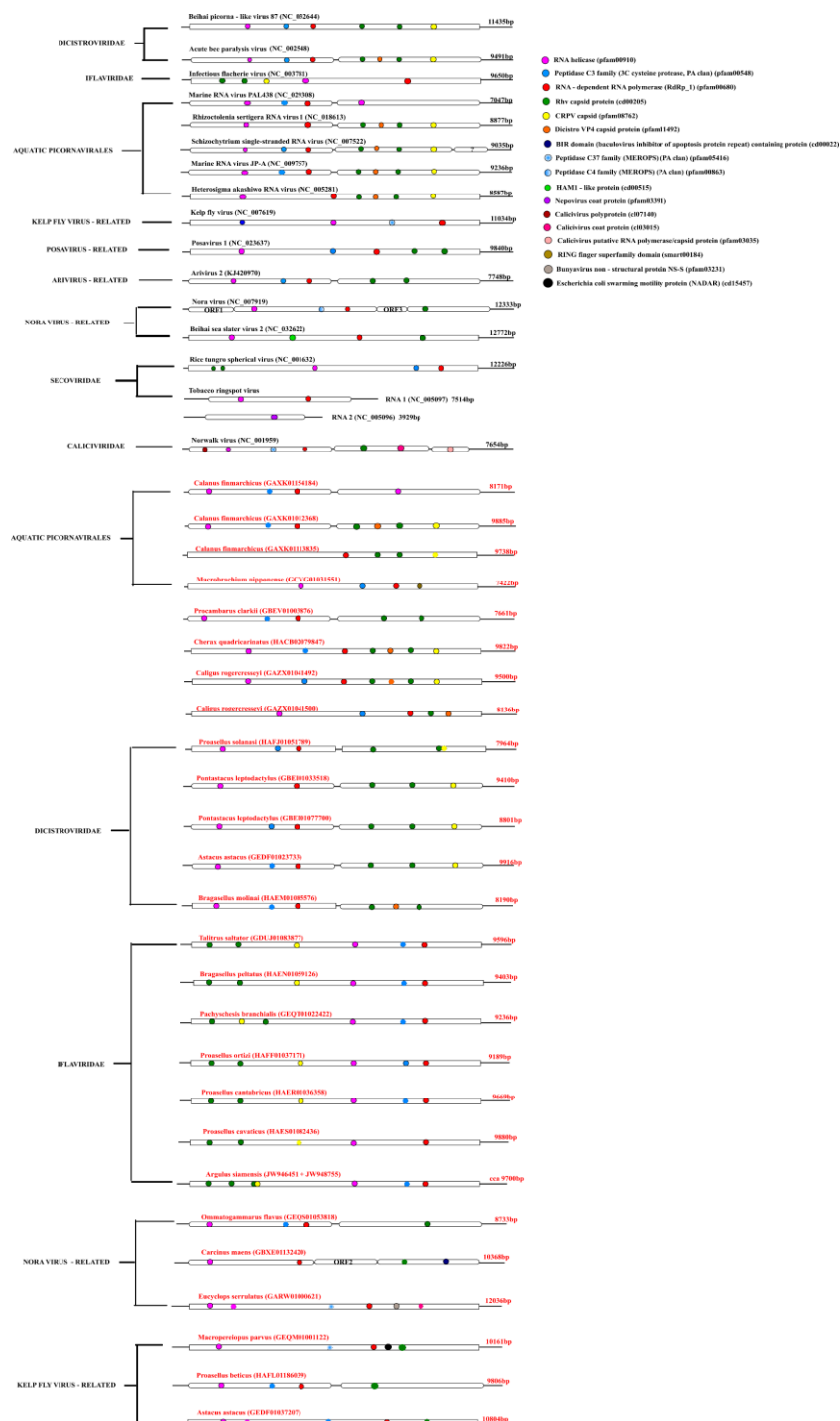
Slika C.29: Genomska organizacija virusov klade **Mono-Chu**, ki smo jih našli pri raki (Crustacea).



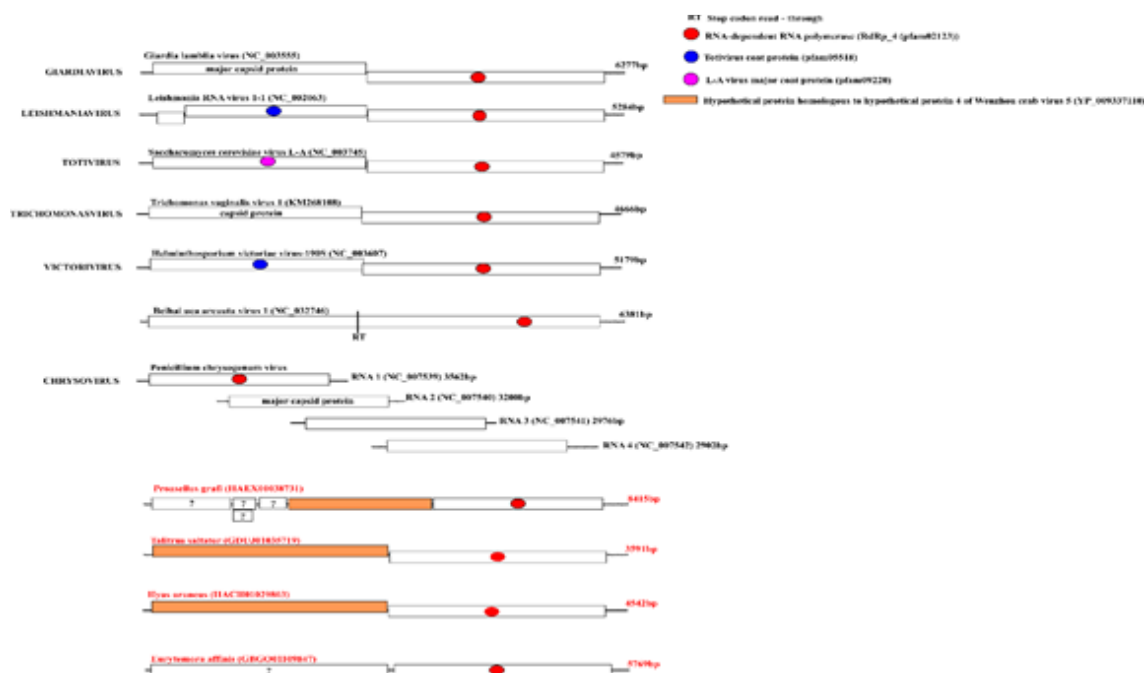
Slika C.30: Genomska organizacija virusov klade **Partiti-Picobirna**, ki smo jih našli pri raki (Crustacea).



Slika C.31: Genomska organizacija virusov klade **Permutotetra**, ki smo jih našli pri rakih (Crustacea).

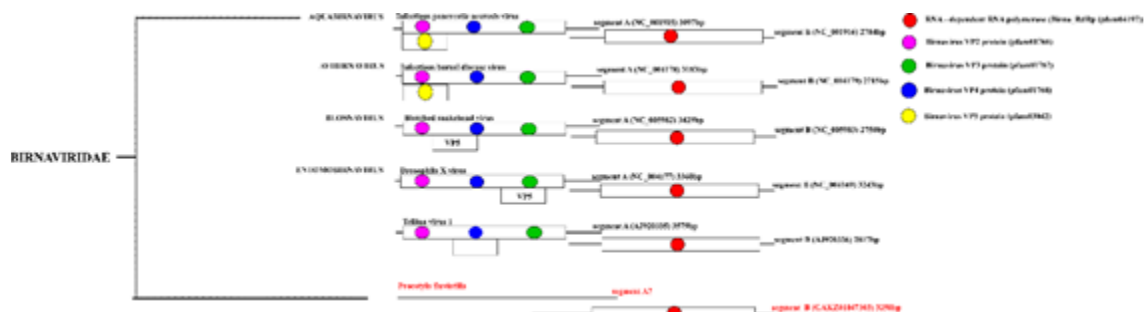


Slika C.32: Genomska organizacija virusov klade **Picornna-Calici**, ki smo jih našli pri rakih (Crustacea).



Slika C.33: Genomska organizacija virusov klade **Toti-Chryso**, ki smo jih našli pri rakih (Crustacea).

C.5 Skupina Spiralia: mehkužci (Mollusca), ploščati črvi (Platyhelminthes), kolobarniki (Annelida) in kotačniki (Rotifera)



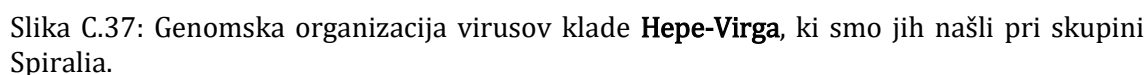
Slika C.34: Genomska organizacija virusov klade **Birna**, ki smo jih našli pri skupini Spiralia. Novi RNA virusi so označeni z rdečo pisavo.

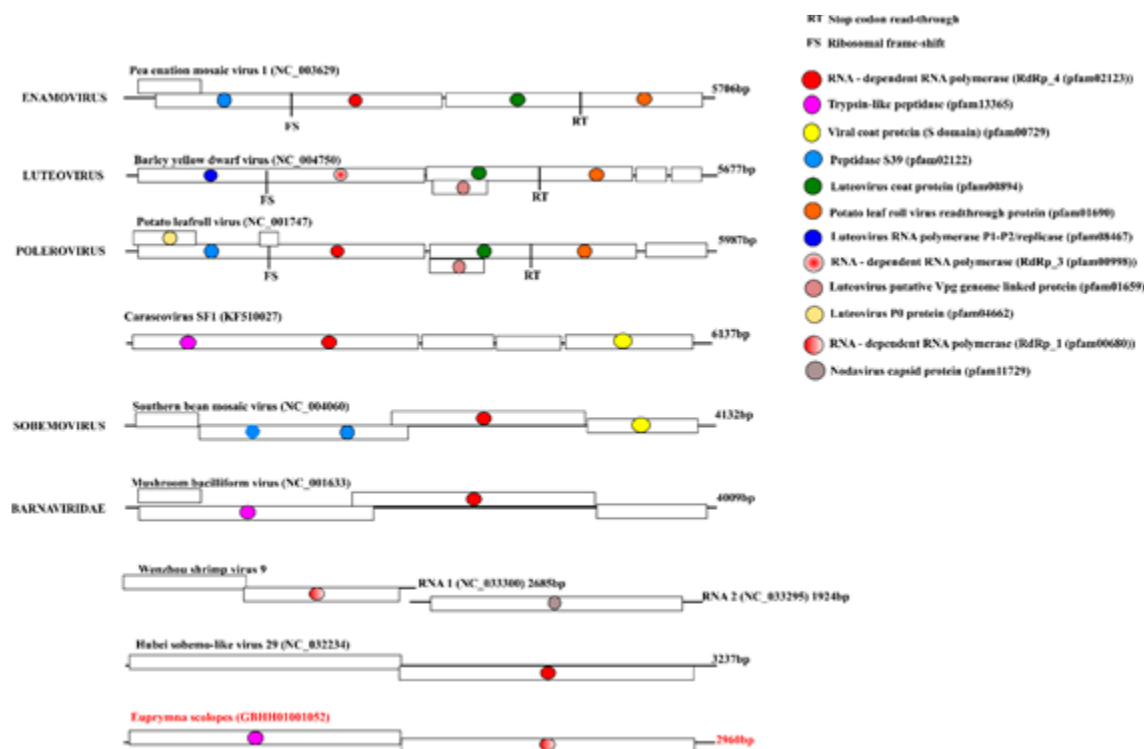
271



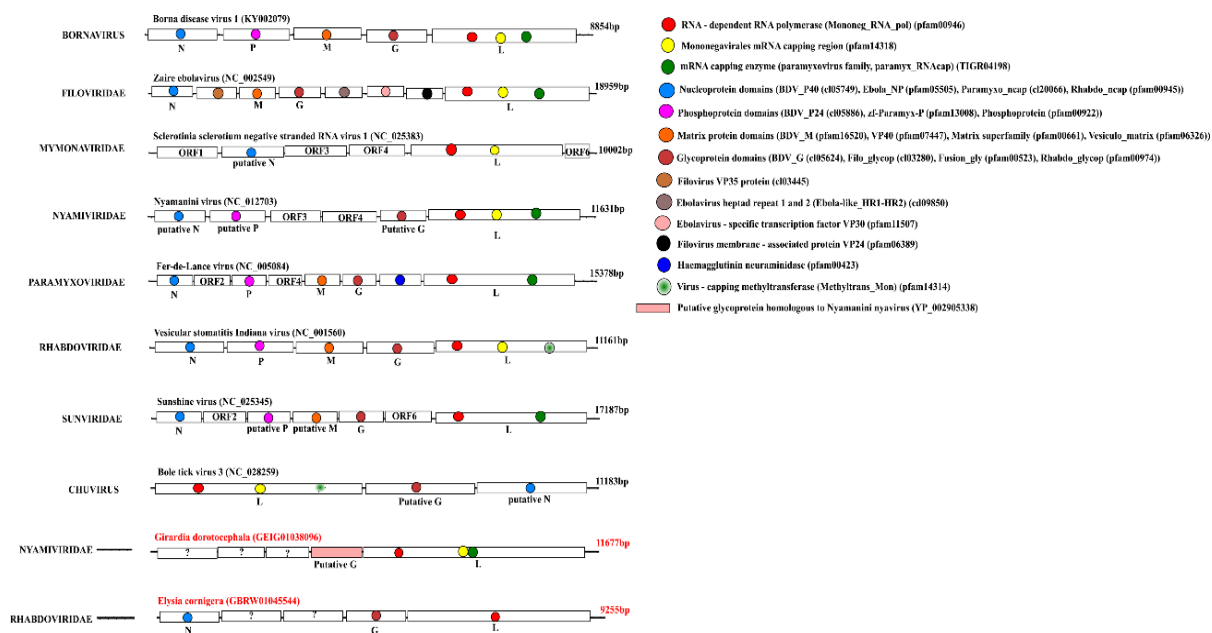
Slika C.35: Genomska organizacija virusov klade **Bunya-Arena**, ki smo jih našli pri skupini *Spiralia*.



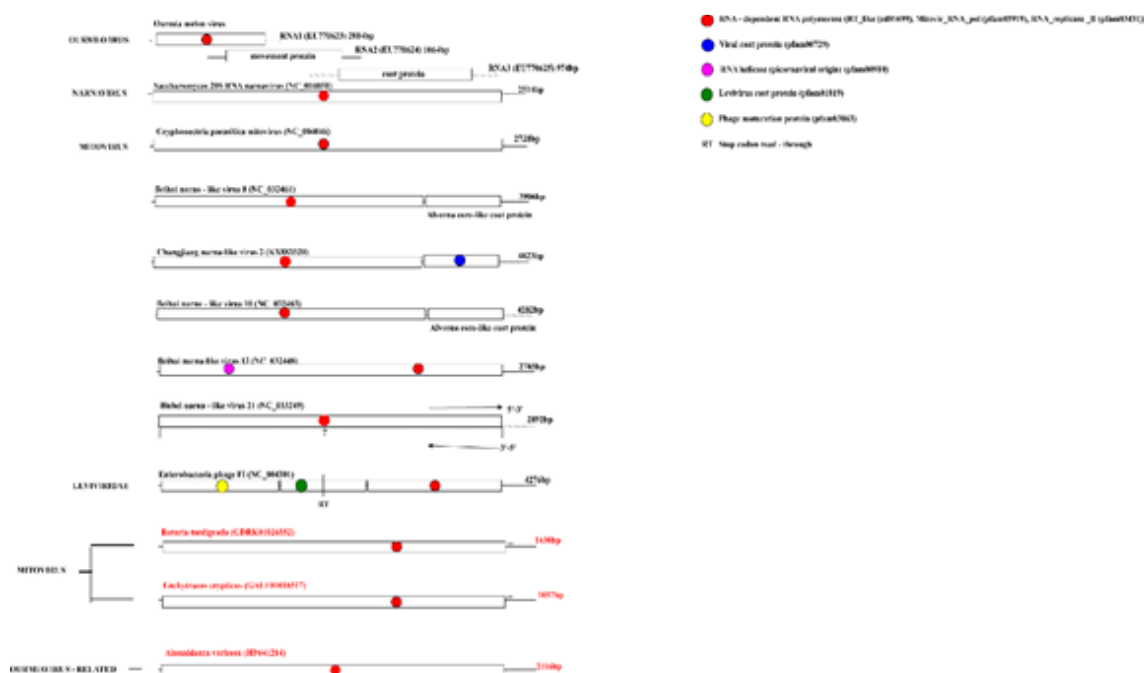




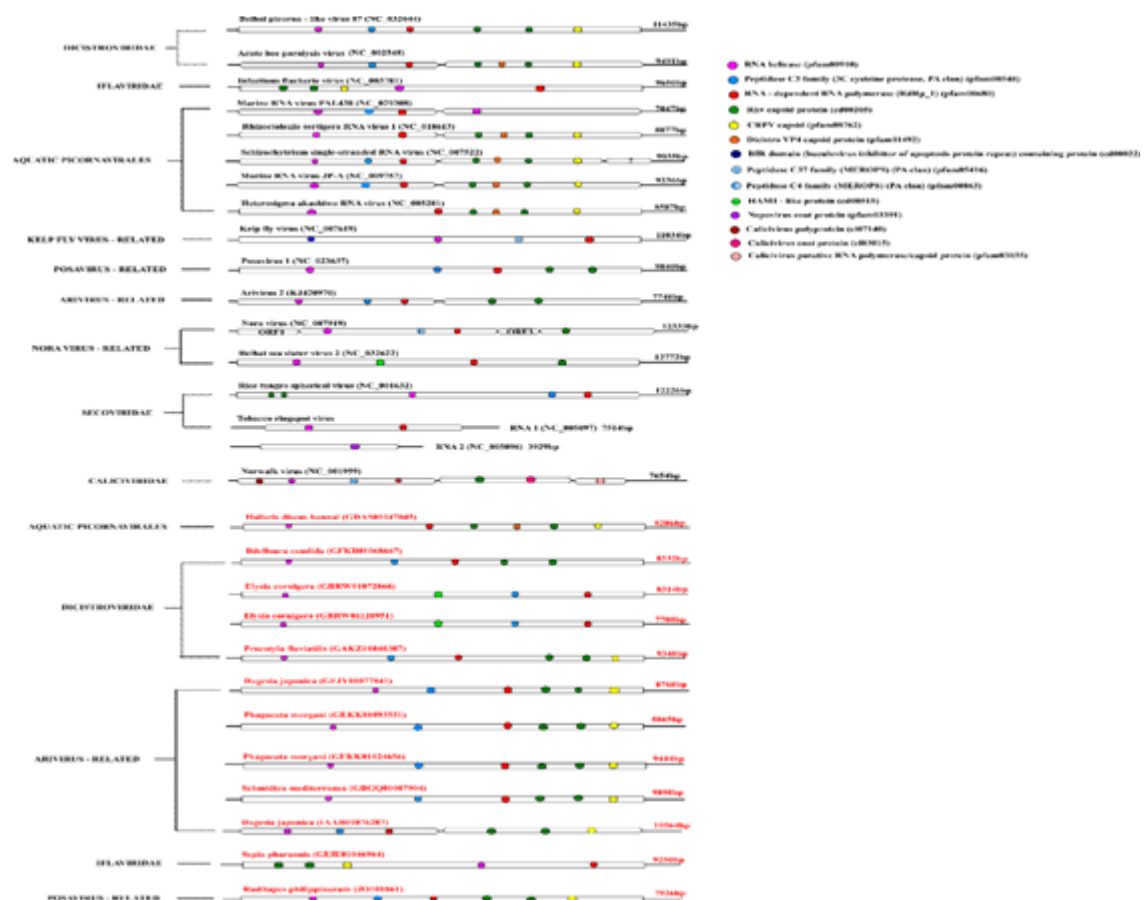
Slika C.38: Genomska organizacija virusov klade **Luteo-Sobemo**, ki smo jih našli pri skupini Spiralia.



Slika C.39: Genomska organizacija virusov klade **Mono-Chu**, ki smo jih našli pri skupini Spiralia.



Slika C.40: Genomska organizacija virusov klade **Narna-Levi**, ki smo jih našli pri skupini Spiralia.

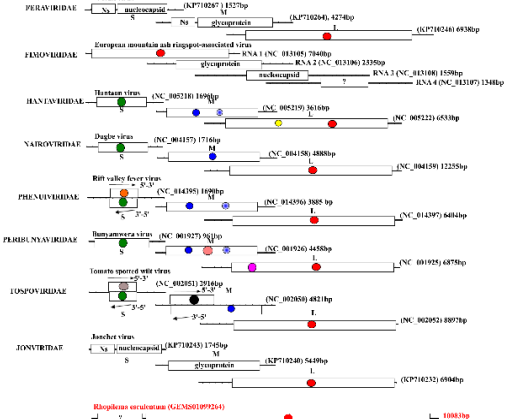


Slika C.41: Genomska organizacija virusov klade **Picorna-Calici**, ki smo jih našli pri skupini Spiralia.



skupini Spiralia.

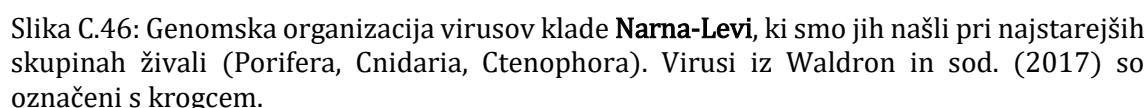
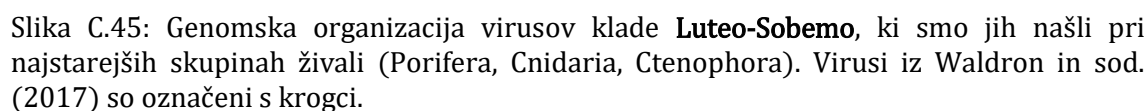
C.6 Najstarejše skupine živali: ožigalkarji (Cnidaria), spužve (Porifera) in rebrače (Ctenophora)

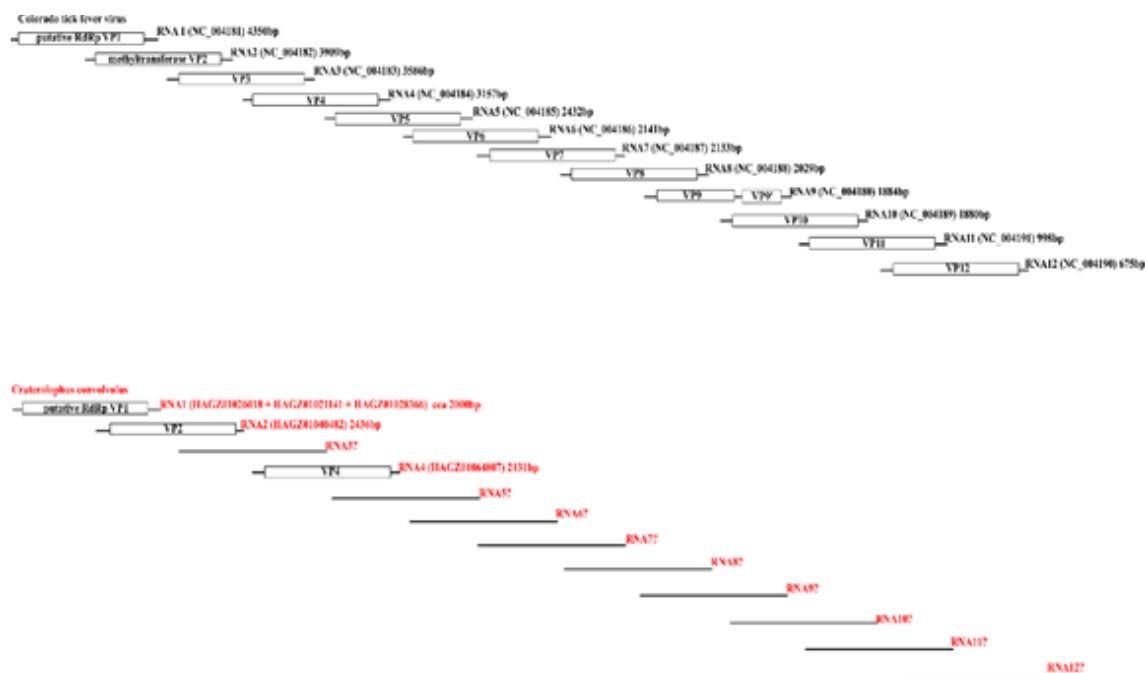


Slika C.43: Genomska organizacija virusov klade **Bunya-Arena**, ki smo jih našli pri najstarejših skupinah živali (Porifera, Cnidaria, Ctenophora). Novi RNA virusi so označeni z rdečo pisavo.

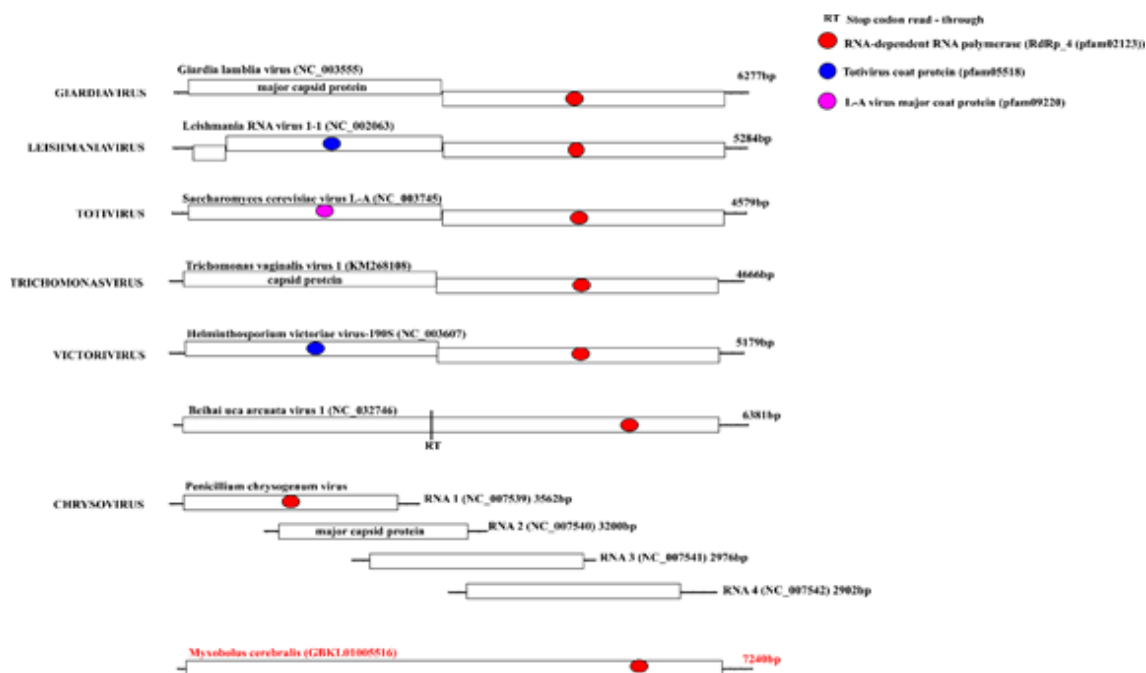


Slika C.44: Genomska organizacija virusov klade **Hepe-Virga**, ki smo jih našli pri najstarejših skupinah živali (Porifera, Cnidaria, Ctenophora). Virusi iz Waldron in sod. (2017) so označeni s krogci.





Slika C.49: Genomska organizacija virusov klade **Toti-Chryso**, ki smo jih našli pri najstarejših skupinah živali (Porifera, Cnidaria, Ctenophora).



Slika C.50: Genomska organizacija virusov klade **Reo**, ki smo jih našli pri najstarejših skupinah živali (Porifera, Cnidaria, Ctenophora).

Literatura

- Abrescia, N. G. A., Grimes, M. J., Fry E.E., Ravantti, J.J., Bamford, D.H., & Stuart, D.I. (2010). What does it take to make a virus: the concept of the viral 'self'. V P. G Stockley, & R. Twarock (Ured.), *Emerging Topics in Physical Virology* (str. 35-58). London: Imperial College Press.
- Abrescia, N.G., Bamford, D.H., Grimes, J.M., & Stuart, D.I. (2012). Structure unifies the viral universe. *Annual Reviews in Biochemistry*, 81, 795-822. doi: 10.1146/annurev-biochem-060910-095130
- Achong, B.G. (1972). EB virus in relation to tumours. *Journal of Clinical Pathology Supplement*, 6, 51-57.
- Aguinaldo, A.M., Turbeville, J.M., Linford, L.S., Rivera, M.C., Garey, J.R., Raff, R.A., & Lake, J.A. (1997). Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature*, 387(6632), 489-493.
- Ahlquist, P. (2006). Parallels among positive-strand RNA viruses, reverse-transcribing viruses and double-stranded RNA viruses. *Nature Reviews Microbiology*, 4 (5), 371-382.
- Aiewsakun, P., & Katzourakis, A. (2015). Endogenous viruses: connecting recent and ancient viral evolution. *Virology*, 479-480, 26-37. doi: 10.1016/j.virol.2015.02.011
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215 (3), 403-410.
- Ang, B.S.P., Lim, T.C.C., & Wang, L. (2018). Nipah virus infection. *Journal of Clinical Microbiology* 56 (6), 1-10. doi: 10.1128/JCM.01875-17
- Angly, F.E., Felts B., Breitbart, M., Salamon, P., Edwards, R.A., Carlson, C., Chan, A.M., Haynes, M., Kelley, S., Liu, H., Mahaffy, J.M., Mueller, J.E., Nulton, J., Olson, R., Parsons, R., Rayhawk, S., Suttle, C.A., & Rohwer, F. (2006). The marine viromes of four oceanic regions. *PLoS Biology*, 4 (11), 2121-2131.
- Apte-Sengupta, S., Sirohi, D., & Kuhn, R.J. (2014). Coupling of replication and assembly in flaviviruses. *Current Opinion in Virology*, 9, 134-142. doi: 10.1016/j.coviro.2014.09.020
- Ariumi, Y., Kuroki, M., Mak,i M., Ikeda, M., Dansako, H., Wakita, T., & Kato, N. (2011). The ESCRT system is required for hepatitis C virus production. *PLoS One*, 6 (1), 1-10. doi: 10.1371/journal.pone.0014517
- Assiri, A., McGeer, A., Perl, T.M., Price, C.S., Al Rabeeah, A.A., Cummings, D.A., Alabdullatif, Z.N., Assad, M., Almulhim, A., Makhdoom, H., Madani, H., Alhakeem, R., Al-Tawfiq, J.A., Cotton, M., Watson, S.J., Kellam, P., Zumla, A.I., & Memish, Z.A.; KSA MERS-CoV Investigation Team. (2013). Hospital outbreak of Middle East respiratory syndrome coronavirus. *The New English Journal of Medicine*, 369 (5), 407-416. doi: 10.1056/NEJMoa1306742
- Bahir, I., Fromer, M., Prat, Y., & Linial, M. (2009). Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Molecular Systems Biology*, 5 (311), 1-14. doi: 10.1038/msb.2009.71
- Baker, M., & Prasad, B.V. (2010). Rotavirus cell entry. *Current Topics in Microbiology and Immunology*, 343, 121-148. doi: 10.1007/82_2010_34
- Balique, F., Lecoq, H., Raoult, D., & Colson, P. (2015). Can plant viruses cross the kingdom border and be pathogenic to humans? *Viruses*, 7 (4), 2074-2098. doi: 10.3390/v7042074
- Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriological Reviews*, 35 (3), 235-241.
- Balvay, L., Soto Rifo, R., Ricci, E.P., Decimo, D., & Ohlmann, T. (2009). Structural and functional diversity of viral IRESes. *Biochimica et Biophysica Acta*, 1789 (9/10), 542-557. doi: 10.1016/j.bbagr.2009.07.005

- Banerjee, I., Miyake, Y., Nobs, S.P., Schneider, C., Horvath, P., Kopf, M., Matthias, P., Helenius, A., & Yamauchi, Y. (2014). Influenza A virus uses the aggresome processing machinery for host cell entry. *Science*, *346* (6208), 473-477. doi: 10.1126/science.1257037
- Bauer, S.P., & Murphy, F.A. (1975). Relationship of two arthropod-borne rhabdoviruses (kotonkan and Obodhiang) to the rabies serogroup. *Infection and Immunity*, *12* (5), 1157-1172.
- Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., & Sayers, E.W. (2013). GenBank. *Nucleic Acids Research*, *41*, 36-42. doi: 10.1093/nar/gks1195.
- Bergelson, J.M. (2008). New (fluorescent) light on poliovirus entry. *Trends in Microbiology*, *16*, (2), 44-47. doi: 10.1016/j.tim.2007.12.004
- Bergh, O., Børshiem, K.Y., Bratbak, G., & Haldal, M. (1989). High abundance of viruses found in aquatic environments. *Nature*, *340* (6233), 467-468.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., & Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28* (1), 235-242.
- Brandenburg, B., Lee, L.Y., Lakadamyali, M., Rust, M.J., Zhuang, X., & Hogle, J.M. (2007). Imaging poliovirus entry in live cells. *PLoS Biology*, *5* (7), 1543-1555.
- Branstetter, M.G., Danforth, B.N., Pitts, J.P., Faircloth, B.C., Ward, P.S., Buffington, M.L., Gates, M.W., Kula, R.R., & Brady, S.G. (2017). Phylogenomic insights into the evolution of stinging wasps and the origins of ants and bees. *Current Biology*, *27* (7), 1019-1025. doi: 10.1016/j.cub.2017.03.027
- Breitbart, M., Thompson, L.R., Suttle, C.A., & Sullivan, M. (2007). Exploring the vast diversity of marine viruses. *Oceanography*, *20* (2), 135-139.
- Bron, J.E., Frisch, D., Goetze, E., Johnson, S.C., Lee, C.E., & Wyngaard, G.A. (2011). Observing copepods through a genomic lens. *Frontiers in Zoology*, *8* (1): 22, 1-15. doi: 10.1186/1742-9994-8-22
- Brum, J.R., Ignacio-Espinoza, J.C., Roux, S., Doulier, G., Acinas, S.G., Alberti, A., Chaffron, S., Cruaud, C., de Vargas, C., Gasol, J.M., Gorsky, G., Gregory, A.C., Guidi, L., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Poulos, B.T., Schwenck, S.M., Speich, S., Dimier, C., Kandels-Lewis, S., Picheral, M., Searson, S., Tara Oceans Coordinators, Bork, P., Bowler, C., Sunagawa, S., Wincker, P., Karsenti, E., & Sullivan, M.B. (2015). Ocean plankton. Patterns and ecological drivers of ocean viral communities. *Science*, *348* (6237), 1-10. doi: 10.1126/science.1261498
- Budd, G.E., & Telford, M.J. (2009). The origin and evolution of arthropods. *Nature*, *457* (7231), 812-817. doi: 10.1038/nature07890
- Bui, C.M., Chughtai, A.A., Adam, D.C., & MacIntyre, C.R. (2017). An overview of the epidemiology and emergence of influenza A infection in humans over time. *Archives of Public Health*, *75* (15), 1-7. doi: 10.1186/s13690-017-0182-z.
- Burkard, C., Verheije, M.H., Wicht, O., van Kasteren, S.I., van Kuppeveld, F.J., Haagmans, B.L., Pelkmans, L., Rottier, P.J., Bosch, B.J., & de Haan, C.A. (2014). Coronavirus cell entry occurs through the endo-/lysosomal pathway in a proteolysis-dependent manner. *PLoS Pathogens*, *10* (11), 1-17. doi: 10.1371/journal.ppat.1004502
- Burki, F. (2014). The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harbor Perspectives in Biology*, *6* (5), 1-17. doi: 10.1101/cshperspect.a016147
- Burki, F., Kaplan, M., Tikhonenkov, D.V., Zlatogursky, V., Minh, B.Q., Radaykina, L.V., Smirnov, A., Mylnikov, A.P., & Keeling, P.J. (2016). Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proceedings of the Royal Society B: Biological Sciences*, *283* (1823), 1-9. doi: 10.1098/rspb.2015.2802
- Cann, A.J. (2012). Principles of molecular virology. 5. izdaja. Elsevier.
- Cantalupo, P.G., Calgua, B., Zhao, G., Hundesa, A., Wier, A.D., Katz, J.P., Grabe, M., Hendrix, R.W., Girones, R., Wang, D., & Pipas, J.M. (2011). Raw sewage harbors diverse viral populations. *MBio*, *2* (5), 1-11. doi: 10.1128/mBio.00180-11
- Carapeta, S., do Bem, B., McGuinness, J., Esteves, A., Abecasis, A., Lopes, Â., de Matos, A.P., Piedade, J., de Almeida, A.P., & Parreira, R. (2015). Negevirus found in multiple species of mosquitoes from southern Portugal: isolation, genetic diversity, and replication in insect cell culture. *Virology*, *483*, 318-328. doi: 10.1016/j.virol.2015.04.021
- Carneiro, F.A., Ferradosa, A.S., & Da Poian, A.T. (2001). Low pH-induced conformational changes in vesicular stomatitis virus glycoprotein involve dramatic structure reorganization. *Journal of Biological Chemistry*, *276* (1), 62-67.

- Carvalho, C.A.M., Silva, J.L.1, Oliveira, A.C., & Gomes, A.M.O. (2017). On the entry of an emerging arbovirus into host cells: Mayaro virus takes the highway to the cytoplasm through fusion with early endosomes and caveolae-derived vesicles. *PeerJ*, 5, 1-18. doi: 10.7717/peerj.3245
- Chiba, S., Kondo, H., Tani, A., Saisho, D., Sakamoto, W., Kanematsu, S., & Suzuki, N. (2011). Widespread endogenization of genome sequences of non-retroviral RNA viruses into plant genomes. *PLoS Pathology*, 7(7), 1-16. doi: 10.1371/journal.ppat.1002146
- Chuang, C.K., Yang, T.H., Chen, T.H., Yang, C.F., & Chen, W.J. (2015). Heat shock cognate protein 70 isoform D is required for clathrin-dependent endocytosis of Japanese encephalitis virus in C6/36 cells. *Journal of General Virology*, 96 (Pt 4), 793-803. doi: 10.1099/jgv.0.000015
- Cohen, S., Au, S., & Panté, N. (2011). How viruses access the nucleus. *Biochimica et Biophysica Acta*, 1813(9), 1634-1645. doi: 10.1016/j.bbamcr.2010.12.009
- Compeau, P.E., Pevzner, P.A., & Tesler, G. (2011). How to apply de Bruijn graphs to genome assembly. *Nature Biotechnology*, 29(11), 987-991. doi: 10.1038/nbt.2023
- Cong, H., Jiang, Y., & Tien, P. (2011). Identification of the myelin oligodendrocyte glycoprotein as a cellular receptor for rubella virus. *Journal of Virology*, 85 (21), 11038-11047. doi: 10.1128/JVI.05398-11
- Cotton, J.A., Steinbiss, S., Yokoi, T., Tsai, I.J., & Kikuchi, T. (2016). An expressed, endogenous Nodavirus-like element captured by a retrotransposon in the genome of the plant parasitic nematode *Bursaphelenchus xylophilus*. *Scientific Reports*, 6 (39749), 1-7. doi: 10.1038/srep39749
- Coyne, C.B., Kim, K.S., & Bergelson, J.M. (2007). Poliovirus entry into human brain microvascular cells requires receptor-induced activation of SHP-2. *EMBO Journal*, 26 (17), 4016-4028.
- Crochu, S., Cook, S., Attoui, H., Charrel, R.N., De Chesse, R., Belhouchet, M., Lemasson, J.J., de Micco, P., & de Lamballerie, X. (2004). Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of *Aedes* spp. mosquitoes. *Journal of General Virology*, 85 (Pt 7), 1971-1980.
- Cui, J., & Holmes, E.C. (2012). Endogenous RNA viruses of plants in insect genomes. *Virology*, 427(2), 77-79. doi: 10.1016/j.virol.2012.02.014
- Culley, A.I., Lang, A.S., & Suttle, C.A. (2006). Metagenomic analysis of coastal RNA virus communities. *Science*, 312(5781), 1795-1798.
- Czarnetzki, A.B., & Tebbe, C.C. (2004). Detection and phylogenetic analysis of *Wolbachia* in *Collembola*. *Environmental Microbiology*, 6(1), 35-44.
- Černý, J., Černá, B. B., Valdés, J.J., Grubhoffer, L., & Růžek, D. (2014). Evolution of tertiary structure of viral RNA dependent polymerases. *PLoS One*, 9 (5), 1-14. doi: 10.1371/journal.pone.0096070
- Černý, J., Černá, B. B., de A Zanotto, P.M., Grubhoffer, L., & Růžek, D. (2015). A deep phylogeny of viral and cellular right-hand polymerases. (2015). *Infection, Genetics and Evolution*, 36, 275-286. doi: 10.1016/j.meegid.2015.09.026
- Dacks, J.B., Field, M.C., Buick, R., Eme, L., Gribaldo, S., Roger, A.J., Brochier-Armanet, C., & Devos, D.P. (2016). The changing view of eukaryogenesis - fossils, cells, lineages and how they all come together. *Journal of Cell Science*, 129(20), 3695-3703
- Daehnel, K., Harris, R., Maddera, L., & Silverman, P. (2005). Fluorescence assays for F-pili and their application. *Microbiology*, 151(11), 3541-3548.
- Daugelavičius, R., Cvirkaite, V., Gaidelyte, A., Bakiene, E., Gabrenaite-Verkhovskaya, R., & Bamford D.H. (2005). Penetration of enveloped double-stranded RNA bacteriophages ϕ 13 and ϕ 6 into *Pseudomonas syringae* cells. *Journal of Virology*, 79(8), 5017-5026.
- Decroly, E., Ferron, F., Lescar, J., & Canard, B. (2011). Conventional and unconventional mechanisms for capping viral mRNA. *Nature Reviews Microbiology*, 10(1), 51-65. doi: 10.1038/nrmicro2675
- Deniz, N., Lenarcic, E.M., Landry, D.M., & Thompson, S.R. (2009). Translation initiation factors are not required for Dicistroviridae IRES function in vivo. *RNA*, 15(5), 932-946. doi: 10.1261/rna.1315109
- Desnues, C., Rodriguez-Brito, B., Rayhawk, S., Kelley, S., Tran, T., Haynes, M., Liu, H., Furlan, M., Wegley, L., Chau, B., Ruan, Y., Hall, D., Angly, F.E., Edwards, R.A., Li, L., Thurber, R.V., Reid, R.P., Siefert, J., Souza, V., Valentine, D.L., Swan, B.K., Breitbart, M., & Rohwer, F. (2008). Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature* 452, (7185), 340-343. doi: 10.1038/nature06735

- Dheilly, N.M., Duval, D., Mouahid, G., Emans, R., Allienne, J.F., Galinier, R., Genthon, C., Dubois, E., Du Pasquier, L., Adema, C.M., Grunau, C., Mitta G., & Gourbal, B. (2015). A family of variable immunoglobulin and lectin domain containing molecules in the snail *Biomphalaria glabrata*. *Developmental and Comparative Immunology*, 48 (1), 234-243. doi: 10.1016/j.dci.2014.10.009
- Dohrmann, M., & Wörheide, G. (2017). Dating early animal evolution using phylogenomic data. *Scientific Reports*, 7(1), 1-6. doi: 10.1038/s41598-017-03791-w
- Dolja, V.V., & Koonin, E.V. (2011). Common origins and host-dependent diversity of plant and animal viromes. *Current Opinion in Virology*, 1 (5), 322-331. doi: 10.1016/j.coviro.2011.09.007
- Dolja, V.V., & Koonin, E.V. (2018). Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. *Virus Research*, 244, 36-52. doi: 10.1016/j.virusres.2017.10.020.
- Dudas, G., & Obbard, D.J. (2015). Are arthropods at the heart of virus evolution? *eLife*, 4, 1-3. doi: 10.7554/eLife.06837.
- Duffy, S., Shackelton, L.A., & Holmes, E.C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nature Reviews Genetics*, 9 (4), 267-276. doi: 10.1038/nrg2323
- Dunn, C.W., Giribet, G., Edgecombe, G.D., & Hejnol, A. (2014). Animal phylogeny and its evolutionary implications. *The Annual Review of Ecology, Evolution and Systematics*, 45, 371-395. doi: 10.1146/annurev-ecolsys-120213-091627
- Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32 (5), 1792-1797.
- Engel, M.S. (2015). Insect evolution. *Current Biology* 25, (19), 868-872. doi: 10.1016/j.cub.2015.07.059.
- ExPASy - <https://www.expasy.org/>
- Ferrer-Orta, C., Arias, A., Escarmís, C., & Verdaguer, N. (2006). A comparison of viral RNA-dependent RNA polymerases. *Current Opinion in Structural Biology*, 16 (1), 27-34.
- Feschotte, C., & Gilbert, C. (2012). Endogenous viruses: insights into viral evolution and impact on host biology. *Nature Reviews Genetics*, 13 (4), 283-296. doi: 10.1038/nrg3199
- Finn, R.D., Coghill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., & Bateman, A. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* 44, (D1), 279-285.
- Fort, P., Albertini, A., Van-Hua, A., Berthomieu, A., Roche, S., Delsuc, F., Pasteur, N., Cappy, P., Gaudin, Y., & Weill, M. (2012). Fossil rhabdoviral sequences integrated into arthropod genomes: ontogeny, evolution, and potential functionality. *Molecular Biology and Evolution*, 29 (1), 381-390. doi: 10.1093/molbev/msr226
- Forzan, M., Marsh, M., & Roy, P. (2007). Bluetongue virus entry into cells. *Journal of Virology*, 81 (9), 4819-4827.
- Frank, J.A., & Feschotte, C. (2017). Co-option of endogenous viral sequences for host cell function. *Current Opinion in Virology*, 25, 81-89. doi: 10.1016/j.coviro.2017.07.021
- Freitas, N., & Cunha, C. (2009). Mechanisms and signals for the nuclear import of proteins. *Current Genomics*, 10 (8), 550-557. doi: 10.2174/138920209789503941
- Galinier, R., Tetreau, G., Portet, A., Pinaud, S., Duval, D., & Gourbal, B. (2017). First characterization of viruses from freshwater snails of the genus *Biomphalaria*, the intermediate host of the parasite *Schistosoma mansoni*. *Acta Tropica*, 167, 196-203. doi: 10.1016/j.actatropica.2016.12.021
- Gascuel, O., & Steel, M. (1987). Neighbor-joining revealed. *Molecular Biology and Evolution*, 23 (11), 1997-2000.
- Gebhard, L.G., Filomatori, C.V., & Gamarnik, A.V. (2011). Functional RNA elements in the dengue virus genome. *Viruses*, 3 (9), 1739-1756. doi: 10.3390/v3091739
- Giribet, G. (2016). New animal phylogeny: future challenges for animal phylogeny in the age of phylogenomics. *Organisms Diversity & Evolution*, 16 (2), 419-426. doi: 10.1007/s13127-015-0236-4
- Greber, U.F. (2016). Virus and host mechanics support membrane penetration and cell entry. *Journal of Virology*, 90 (8), 3802-3805. doi: 10.1128/JVI.02568-15
- Griffin, S.D., Beales, L.P., Clarke, D.S., Worsfold, O., Evans, S.D., Jaeger, J., Harris, M.P., & Rowlands, D.J. (2003). The p7 protein of hepatitis C virus forms an ion channel that is blocked by the antiviral drug, Amantadine. *FEBS Letters*, 535 (1/3), 34-38.

- Griffith, M., Walker, J.R., Spies, N.C., Ainscough, B.J., & Griffith, O.L. (2015). Informatics for RNA sequencing: A web resource for analysis on the cloud. *PLoS Computational Biology*, 11 (8), 1-21. doi: 10.1371/journal.pcbi.1004393
- Guindon, S., Delsuc, F., Dufayard, J.F., & Gascuel, O. (2009). Estimating maximum likelihood phylogenies with PhyML. *Methods in Molecular Biology*, 537, 113-137. doi: 10.1007/978-1-59745-251-9_6
- Gutiérrez, M., Isa, P., Sánchez-San, M. C., Pérez-Vargas, J., Espinosa, R., Arias, C.F., & López, S. (2010). Different rotavirus strains enter MA104 cells through different endocytic pathways: the role of clathrin-mediated endocytosis. *Journal of Virology*, 84 (18), 9161-9169. doi: 10.1128/JVI.00731-10
- Hackenberg, M., & Kotsyfakis, M. (2018). Exosome-Mediated Pathogen Transmission by Arthropod Vectors. *Trends in Parasitology*, pii: S1471-4922(18)30066-7, 1-3. doi: 10.1016/j.pt.2018.04.001.
- Halanych, K.M., Bacheller, J.D., Aguinaldo, A.M., Liva, S.M., Hillis, D.M., & Lake, J.A. (1995). Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. *Science*, 267(5204), 1641-1643.
- Halstead, S.B. (1966). Mosquito-borne haemorrhagic fevers of South and South-East Asia. *Bulletin of the World Health Organization*, 35 (1), 3-15.
- Han, B.A., Schmidt, J.P., Bowden, S.E., & Drake, J.M. (2015). Rodent reservoirs of future zoonotic diseases. *Proceedings of the National Academy of Sciences*, 112 (22), 7039-7044. doi: 10.1073/pnas.1501598112
- Harrison, B.D., Finch, J.T., Gibbs, A.J., Hollings, M., Shepherd, M.J., Valenta, V., & Wetter, C. (1971). Sixteen groups of plant viruses. *Virology*, 45 (2), 356-363.
- Harwig, A., Landick, R., & Berkhout, B. (2017). The battle of RNA synthesis: virus versus host. *Viruses*, 9 (10), 1-18. doi: 10.3390/v9100309.
- Havlak, P., Chen, R., Durbin, K.J., Egan, A., Ren, Y., Song, X.Z., Weinstock, G.M., & Gibbs R.A. (2004). The Atlas genome assembly system. *Genome Research*, 14 (4), 721-732.
- He, D., Fiz-Palacios, O., Fu, C.J., Fehling, J., Tsai, C.C., & Baldauf, S.L. (2014). An alternative root for the eukaryote tree of life. *Current Biology*, 24 (4), 465-470. doi: 10.1016/j.cub.2014.01.036
- Hogenhout, S.A., Ammar, el-D., Whitfield, A.E., & Redinbaugh, M.G. (2008). Insect vector interactions with persistently transmitted viruses. *Annual Review of Phytopathology*, 46, 327-359. doi: 10.1146/annurev.phyto.022508.092135
- Holmes, E.C. (2011). The evolution of endogenous viral elements. *Cell Host and Microbe*, 10 (4), 368-377. doi: 10.1016/j.chom.2011.09.002
- Horie, M., & Tomonaga, K. (2011). Non-retroviral fossils in vertebrate genomes. *Viruses*, 3 (10), 1836-1848. doi: 10.3390/v3101836
- Hull, R., & Shepherd, R.J. (1976). Cauliflower Mosaic Virus: an Improved Purification Procedure and Some Properties of the Virus Particles. *Journal of General Virology*, 31, 93-100.
- Hulo, C., de Castro, E., Masson, P., Bougueleret, L., Bairoch, A., Xenarios, I., & Le Mercier, P. (2011). ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Research*, 39 (Izdaja baze podatkov – Database issue), 576-582.
- Irie, T., & Harty, R.N. (2005). L-domain flanking sequences are important for host interactions and efficient budding of vesicular stomatitis virus recombinants. *Journal of Virology*, 79 (20), 12617-12622.
- Jadav, S.S., Kumar, A., Ahsan, M.J., & Jayaprakash, V. (2015). Ebola virus: current and future perspectives. *Infectious Disorders - Drug Targets*, 15 (1), 20-31.
- Junglen, S., & Drosten, C. (2013). Virus discovery and recent insights into virus diversity in arthropods. *Current Opinion in Microbiology*, 16 (4), 507-513. doi: 10.1016/j.mib.2013.06.005
- Katoh, K., & Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30 (4), 772-780. doi: 10.1093/molbev/mst010.
- Katzourakis, A., & Gifford, R.J. (2010). Endogenous viral elements in animal genomes. *PLoS Genetics*, 6 (11), 1-13. doi: 10.1371/journal.pgen.1001191
- Kawahara, A.Y., & Breinholt, J.W. (2014). Phylogenomics provides strong evidence for relationships of butterflies and moths. *Proceedings of the Royal Society B: Biological Sciences*, 281 (1788), 1-8. doi: 10.1098/rspb.2014.0970
- Kawakami, K., Kurnia, Y.W., Fujita, R., Ito, T., Isawa, H., Asano, S., Binh, N.D., & Bando, H. (2016). Characterization of a novel negevirus isolated from Aedes larvae collected in

- a subarctic region of Japan. *Archives of Virology*, 161 (4), 801-809. doi: 10.1007/s00705-015-2711-9
- Kiss, A.L., & Botos, E. (2009). Endocytosis via caveolae: alternative pathway with distinct cellular compartments to avoid lysosomal degradation? *Journal of Cellular and Molecular Medicine*, 13 (7), 1228-1237. doi: 10.1111/j.1582-4934.2009.00754.x
- Kolatkhar, P.R., Bella, J., Olson, N.H., Bator, C.M., Baker, T.S., & Rossmann, M.G. (1999). Structural studies of two rhinovirus serotypes complexed with fragments of their cellular receptor. *EMBO Journal*, 18 (22), 6249-6259.
- Koonin, E.V., Gorbalenya, A.E., & Chumakov, K.M. (1989). Tentative identification of RNA-dependent RNA polymerases of dsRNA viruses and their relationship to positive strand RNA viral polymerases. *FEBS Letters*, 252 (1/2), 42-46.
- Koonin, E.V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *Journal of General Virology*, 72 (9), 2197-2206. doi: 10.1099/0022-1317-72-9-2197
- Koonin, E.V., Senkevich, T.G., & Dolja, V.V. (2006). The ancient Virus World and evolution of cells. *Biology Direct*, 1 (29). doi: 10.1186/1745-6150-1-29
- Koonin, E.V., & Dolja, V.V. (2013). A virocentric perspective on the evolution of life. *Current Opinion in Virology*, 3 (5), 546-557. doi: 10.1016/j.coviro.2013.06.008
- Koonin, E.V., & Dolja, V.V. (2014). Virus world as an evolutionary network of viruses and capsidless selfish elements. *Microbiology and Molecular Biology Reviews*, 78 (2), 278-303. doi: 10.1128/MMBR.00049-13
- Koonin, E.V., Dolja, V.V., & Krupovic, M. (2015). Origins and evolution of viruses of eukaryotes: the ultimate modularity. *Virology*, 479-480, 2-25. doi: 10.1016/j.virol.2015.02.039
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33 (7), 1870-1874. doi: 10.1093/molbev/msw054
- Ladner, J.T., Wiley, M.R., Beitzel, B., Auguste, A.J., Dupuis, A.P., Lindquist, M.E., Sibley, S.D., Kota, K.P., Fetterer, D., Eastwood, G., Kimmel, D., Prieto, K., Guzman, H., Aliota, M.T., Reyes, D., Brueggemann, E.E., St John, L., Hyeroba, D., Lauck, M., Friedrich, T.C., O'Connor, D.H., Gestole, M.C., Cazares, L.H., Popov, V.L., Castro-Llanos, F., Kochel, T.J., Kenny, T., White, B., Ward, M.D., Loaiza, J.R., Goldberg, T.L., Weaver, S.C., Kramer, L.D., Tesh, R.B., & Palacios, G. (2016). A multicomponent animal virus isolated from mosquitoes. *Cell Host and Microbe*, 20 (3), 357-367. doi: 10.1016/j.chom.2016.07.011
- Lange, A., Mills, R.E., Lange, C.J., Stewart, M., Devine, S.E., & Corbett, A.H. (2007). Classical nuclear localization signals: definition, function, and interaction with importin α . *Journal of Biological Chemistry*, 282 (8), 5101-5105.
- Lartillot, N., & Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution*, 21 (6), 1095-1109.
- Laumer, C.E., Bekkouche, N., Kerbl, A., Goetz, F., Neves, R.C., Sørensen, M.V., Kristensen, R.M., Hejnol, A., Dunn, C.W., Giribet, G., & Worsaae, K. (2015). Spiralian phylogeny informs the evolution of microscopic lineages. *Current Biology*, 25 (15), 2000-2006. doi: 10.1016/j.cub.2015.06.068
- Lauterbach, S.E., Wright, C.M., Zentkovich, M.M., Nelson, S.W., Lorbach, J.N., Bliss, N.T., Nolting, J.M., Pierson, R.M., King, M.D., & Bowman, A.S. (2018). Detection of influenza A virus from agricultural fair environment: air and surfaces. *Preventive Veterinary Medicine*, 1 (153), 24-29. doi: 10.1016/j.prevetmed.2018.02.019
- Lefkowitz, E.J., Dempsey, D.M., Hendrickson, R.C., Orton, R.J., Siddell, S.G., & Smith, D.B. (2018). Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Research*, 46 (D1), 1-10. doi: 10.1093/nar/gkx932.
- Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, 44 (W1), 242-245. doi: 10.1093/nar/gkw290
- Li, C.X., Shi, M., Tian, J.H., Lin, X.D., Kang, Y.J., Chen, L.J., Qin, X.C., Xu, J., Holmes, E.C., & Zhang, Y.Z. (2015). Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *ELife*, 4, 1-26. doi: 10.7554/eLife.05378
- Liszewski, M.K., Post, T.W., & Atkinson, J.P. (1991). Membrane cofactor protein (MCP or CD46): newest member of the regulators of complement activation gene cluster. *Annual Review of Immunology*, 9, 431-455.

- Littlewood, D.T.J. (2006). The evolution of parasitism in flatworms. V A.G.Maule, N.J. Marks (Ured.), *Parasitic flatworms: Molecular Biology, Biochemistry, Immunology and Physiology*. CABI Publishing, 1-31.
- Liu, H., Fu, Y., Xie, J., Cheng, J., Ghabrial, S.A., Li, G., Yi, X., & Jiang, D. (2012). Discovery of novel dsRNA viral sequences by *in silico* cloning and implications for viral diversity, host range and evolution. *PLoS One*, 7(7), 1-10. doi: 10.1371/journal.pone.0042147
- Liu, Z., Chen, T., Ma, L., Zhao, Z., Zhao, P.X., Nan, Z., & Wang, Y. (2013). Global transcriptome sequencing using the Illumina platform and the development of EST-SSR markers in autotetraploid alfalfa. *PLoS One*, 8(12), 1-13. doi: 10.1371/journal.pone.0083549
- Lozach, P.Y., Mancini, R., Bitto, D., Meier, R., Oestereich, L., Overby, A.K., Pettersson, R.F., & Helenius, A. (2010). Entry of bunyaviruses into mammalian cells. *Cell Host and Microbe*, 7(6), 488-499. doi: 10.1016/j.chom.2010.05.007
- Lu, T.M., Kanda, M., Satoh, N., & Furuya, H. (2017). The phylogenetic position of dicyemid mesozoans offers insights into spiralian evolution. *Zoological Letters*, 3 (6), 1-9. doi: 10.1186/s40851-017-0068-5
- Maib, H., Smythe, E., & Ayscough, K. (2017). Forty years on: clathrin-coated pits continue to fascinate. *Molecular Biology of the Cell*, 28 (7), 843-847. doi: 10.1091/mbc.E16-04-0213
- Mandl, J.N., Ahmed, R., Barreiro, L.B., Daszak, P., Epstein, J.H., Virgin, H.W., & Feinberg, M.B. (2015). Reservoir host immune responses to emerging zoonotic viruses. *Cell*, 160 (1/2), 20-35. doi: 10.1016/j.cell.2014.12.003
- Marchler-Bauer, A., Bo, Y., Han, L., He J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Geer, L.Y., & Bryant, S.H. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*, 45 (D1), 200-203. doi: 10.1093/nar/gkw1129
- Martin, S.J., Highfield, A.C., Brettell, L., Villalobos, E.M., Budge, G.E., Powell, M., Nikaido, S., & Schroeder D.C. (2012). Global honey bee viral landscape altered by a parasitic mite. *Science*, 336 (6086), 1304-1306. doi: 10.1126/science.1220941.
- Maurer, K., Krey, T., Moennig, V., Thiel, H.J., & Rümenapf, T. (2004). CD46 is a cellular receptor for bovine viral diarrhea virus. *Journal of Virology*, 78 (4), 1792-1799.
- McDonald, B., & Martin-Serrano, J. (2009). No strings attached: the ESCRT machinery in viral budding and cytokinesis. *Journal of Cell Science*, 122 (Pt 13), 2167-2177. doi: 10.1242/jcs.028308
- McGarigal, K., Cushman, S., & Stafford, S. (2000). Multivariate statistics for wildlife and ecology research. 1. izdaja. Springer.
- Mizuno, C.M., Rodriguez-Valera, F., Kimes, N.E., & Ghai, R. (2013). Expanding the marine virosphere using metagenomics. *PLoS Genetics*, 9 (12), 1-13. doi: 10.1371/journal.pgen.1003987
- Mohl, B.P., & Roy, P. (2014). Bluetongue virus capsid assembly and maturation. *Viruses*, 6 (8), 3250-3270. doi: 10.3390/v6083250
- Mokili, J. L., Rohwer, F., & Dutilh, B.E. (2012). Metagenomics and future perspectives in virus discovery. *Current Opinion in Virology*, 2 (1), 63-77. doi: 10.1016/j.coviro.2011.12.004
- Moroz, L.L., Kocot, K.M., Citarella, M.R., Dosung, S., Norekian, T.P., Povolotskaya, I.S., Grigorenko, A.P., Dailey, C., Berezikov, E., Buckley, K.M., Ptitsyn, A., Reshetov, D., Mukherjee, K., Moroz, T.P., Bobkova, Y., Yu, F., Kapitonov, V.V., Jurka, J., Bobkov, Y.V., Swore, J.J., Girardo, D.O., Fodor, A., Gusev, F., Sanford, R., Bruders, R., Kittler, E., Mills, C.E., Rast, J.P., Derelle, R., Solovyev, V.V., Kondrashov, F.A., Swalla, B.J., Sweedler, J.V., Rogaev, E.I., Halanych, K.M., & Kohn, A.B. (2014). The ctenophore genome and the evolutionary origins of neural systems. *Nature*, 510 (7503), 109-114. doi: 10.1038/nature13400
- Mullis, K.B., & Faloona, F.A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in Enzymology*, 155, 335-350.
- Mutanen, M., Wahlberg, N., & Kaila, L. (2010). Comprehensive gene and taxon coverage elucidates radiation patterns in moths and butterflies. *Proceedings of the Royal Society B: Biological Sciences*, 277 (1695), 2839-2848. doi: 10.1098/rspb.2010.0392
- Nabi, I.R., & Le, P.U. (2003). Caveolae/raft-dependent endocytosis. *Journal of Cell Biology*, 161 (4), 673-677.
- Nasir, A., & Caetano-Anollés, G. (2015). A phylogenomic data-driven exploration of viral origins and evolution. *Science Advances*, 1 (8), 1-24. doi: 10.1126/sciadv.1500527

- NCBI (National Center for Biotechnology Information) - <https://www.ncbi.nlm.nih.gov/>
- NCBI Resource Coordinators. (2016). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 44 (D1), D7-19. doi: 10.1093/nar/gkv1290.
- NCBI Resource Coordinators. (2017). Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 45 (D1), D12-D17. doi: 10.1093/nar/gkw1071.
- Negi, S., Pandey, S., Srinivasan, S.M., Mohammed, A., & Guda, C. (2015). LocSigDB: a database of protein localization signals. *Database (Oxford)*, pii: bav003, 1-7. doi: 10.1093/database/bav003
- Nesnidal, M.P., Helmkamp, M., Meyer, A., Witek, A., Bruchhaus, I., Ebersberger, I., Hankeln, T., Lieb, B., Struck, T.H., & Hausdorf, B. (2013). New phylogenomic data support the monophyly of Lophophorata and an Ectoproct-Phoronid clade and indicate that Polyzoa and Kryptozoa are caused by systematic bias. *BMC Evolutionary Biology*, 13(253), 1-13. doi: 10.1186/1471-2148-13-253.
- Ng, W.C., Soto-Acosta, R., Bradrick, S.S., Garcia-Blanco, M.A., & Ooi, E.E. (2017). The 5' and 3' untranslated regions of the flaviviral genome. *Viruses*, 9 (6), 1-14. doi: 10.3390/v9060137
- Nguyen, B. A.N., Pogoutse, A., Provart, N., & Moses, A.M. (2009). NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics*, 10(202), 1-11. doi: 10.1186/1471-2105-10-202
- Nguyen, L.T., Schmidt, H.A., von Haeseler, A., & Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32 (1), 268-274. doi: 10.1093/molbev/msu300
- Niepmann, M. (2009). Internal translation initiation of picornaviruses and hepatitis C virus. *Biochimica et Biophysica Acta*, 1789 (9/10), 529-541. doi: 10.1016/j.bbarm.2009.05.002
- Nouri, S., Salem N., Nigg, J.C., & Falk, B.W. (2015). Diverse array of new viral sequences identified in worldwide populations of the asian citrus psyllid (*Diaphorina citri*) using viral metagenomics. *Journal of Virology*, 90 (5), 2434-2445. doi: 10.1128/JVI.02793-15
- Olival, K.J., Hosseini, P.R., Zambrana-Torrel, C., Ross, N., Bogich, T.L., & Daszak, P. (2017). Host and viral traits predict zoonotic spillover from mammals. *Nature*, 546 (7660), 646-650. doi: 10.1038/nature22975
- Otsuki, N., Sakata, M., Saito, K., Okamoto, K., Mori, Y., Hanada, K., & Takeda, M. (2018). Both sphingomyelin and cholesterol in the host cell membrane are essential for Rubella virus entry. *Journal of Virology*, 92 (1), 1-16. doi: 10.1128/JVI.01130-17
- Ozsolak, F., Platt, A.R., Jones, D.R., Reifengerger, J.G., Sass, L.E., McInerney, P., Thompson, J.F., Bowers, J., Jarosz, M., & Milos, P.M. (2009). Direct RNA sequencing. *Nature*, 461 (7265), 814-818. doi: 10.1038/nature08390
- Paez-Espino, D., Eloie-Fadrosch, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann, M., Mikhailova, N., Rubin, E., Ivanova, N.N., & Kyrpides, N.C. (2016). Uncovering Earth's virome. *Nature*, 536 (7617), 425-430.
- PDB (Protein Data Bank) - <https://www.rcsb.org/>
- Peters, R.S., Krogmann, L., Mayer, C., Donath, A., Gunkel, S., Meusemann, K., Kozlov, A., Podsiadlowski, L., Petersen, M., Lanfear, R., Diez, P.A., Heraty, J., Kjer, K.M., Klopstein, S., Meier, R., Polidori, C., Schmitt, T., Liu, S., Zhou, X., Wappler, T., Rust, J., Misof, B., & Niehuis, O. (2017). Evolutionary History of the Hymenoptera. *Current Biology*, 27 (7), 1013-1018. doi: 10.1016/j.cub.2017.01.027
- Pevsner, J. (2009). Bioinformatics and functional genomics. 2. izdaja. Wiley Blackwell.
- Pham-Dinh, D., Mattei, M.G., Nussbaum, J.L., Roussel, G., Pontarotti, P., Roedel, N., Mather, I.H., Artzt, K., Lindahl, K.F., & Dautigny, A. (1993). Myelin/oligodendrocyte glycoprotein is a member of a subset of the immunoglobulin superfamily encoded within the major histocompatibility complex. *Proceedings of the National Academy of Sciences*, 90 (17), 7990-7994.
- Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchiellini, C., Boury-Esnault, N., Vacelet, J., Renard, E., Houlston, E., Quéinnec, E., Da Silva, C., Wincker, P., Le Guyader, H., Leys, S., Jackson, D.J., Schreiber, F., Erpenbeck, D., Morgenstern, B., Wörheide, G., & Manuel, M. (2009). Phylogenomics revives traditional views on deep animal relationships. *Current Biology*, 19 (8), 706-712. doi: 10.1016/j.cub.2009.02.052

- Philosof, A., Yutin, N., Flores-Urbe, J., Sharon, I., Koonin, E.V., & B  j  , O. (2017). Novel abundant oceanic viruses of uncultured marine group II euryarchaeota. *Current Biology*, 27(9), 1362-1368. doi: 10.1016/j.cub.2017.03.052
- PIR (Protein Information Resource) - <https://pir.georgetown.edu/>
- Pisani, D., Pett, W., Dohrmann, M., Feuda, R., Rota-Stabelli, O., Philippe, H., Lartillot, N., & W  rheide, G. (2015). Genomic data do not support comb jellies as the sister group to all other animals. *Proceedings of the National Academy of Sciences*, 112(50), 15402-15407. doi: 10.1073/pnas.1518127112
- Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., & Claverie, J.M. (2004). The 1.2-megabase genome sequence of Mimivirus. *Science*, 306(5700), 1344-1350.
- Raoult, D., & Forterre, P. (2008). Redefining viruses: lessons from Mimivirus. *Nature Reviews Microbiology*, 6(4), 315-319. doi: 10.1038/nrmicro1858
- Rawlings, N.D., & Barrett, A.J. (1993). Evolutionary families of peptidases. *Biochemical Journal*, 290(1), 205-218.
- Rawlings, N.D., Barrett, A.J. & Finn, R.D. (2016). Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, 44, 343-350. doi: 10.1093/nar/gkv1118
- Rohwer, F., & Thurber, R.V. (2009). Viruses manipulate the marine environment. *Nature*, 459(7244), 207-212. doi: 10.1038/nature08060
- Rosani, U., & Gerdol, M. (2017). A bioinformatics approach reveals seven nearly-complete RNA-virus genomes in bivalve RNA-seq data. *Virus Research*, 239, 33-42. doi: 10.1016/j.virusres.2016.10.009
- Roossinck, M.J. (2015). Plants, viruses and the environment: Ecology and mutualism. *Virology*, 479-480, 271-277. doi: 10.1016/j.virol.2015.03.041
- Rossman, J.S., Jing, X., Leser, G.P., & Lamb, R.A. (2010). Influenza virus M2 protein mediates ESCRT-independent membrane scission. *Cell*, 142(6), 902-913. doi: 10.1016/j.cell.2010.08.029
- Rossmann, M.G., He, Y., & Kuhn, R.J. (2002). Picornavirus-receptor interactions. *Trends in Microbiology*, 10(7), 324-331.
- Ryan, J.F., Pang, K., Schnitzler, C.E., Nguyen, A.D., Moreland, R.T., Simmons, D.K., Koch, B.J., Francis, W.R., Havlak, P., NISC Comparative Sequencing Program, Smith S.A., Putnam, N.H., Haddock, S.H., Dunn, C.W., Wolfsberg, T.G., Mullikin, J.C., Martindale, M.Q., & Baxevanis, A.D. (2013). The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science*, 342(6164), 1-19. doi: 10.1126/science.1242592
- Saitou, N., & Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4), 406-425.
- Sanger, F., Nicklen, S., & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463-5467.
- Sanju  n, R., & Domingo-Calap, P. (2016). Mechanisms of viral mutation. *Cellular and Molecular Life Sciences*, 73(23), 4433-4448.
- Sasson, D.A., & Ryan, J.F. (2017). A reconstruction of sexual modes throughout animal evolution. *BMC Evolutionary Biology*, 17(1): 242, 1-10. doi: 10.1186/s12862-017-1071-3
- Schulz, M.H., Zerbino, D.R., Vingron, M., & Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*, 28(8), 1086-1092 doi: 10.1093/bioinformatics/bts094
- Seb  -Pedr  s, A., Degnan, B.M., & Ruiz-Trillo, I. (2017). The origin of Metazoa: a unicellular perspective. *Nature Reviews Genetics*, 18(8), 498-512. doi: 10.1038/nrg.2017.21
- Sherman, M.P., & Greene, W.C. (2002). Slipping through the door: HIV entry into the nucleus. *Microbes and Infection*, 4(1), 67-73.
- Shi, M., Lin, X.D., Tian, J.H., Chen, L.J., Chen, X., Li, C.X., Qin, X.C., Li, J., Cao, J.P., Eden, J.S., Buchmann, J., Wang, W., Xu, J., Holmes, E.C., & Zhang, Y.Z. (2016a). Redefining the invertebrate RNA virosphere. *Nature*, 540, 539-543. doi: 10.1038/nature20167
- Shi, M., Lin, X.D., Vasilakis, N., Tian, J.H., Li, C.X., Chen, L.J., Eastwood, G., Diao, X.N., Chen, M.H., Chen, X., Qin, X.C., Widen, S.G., Wood, T.G., Tesh, R.B., Xu, J., Holmes, E.C., & Zhang, Y.Z. (2016b). Divergent viruses discovered in arthropods and vertebrates revise the evolutionary history of the flaviviridae and related viruses. *Journal of Virology*, 90(2), 659-669. doi: 10.1128/JVI.02036-15

- Shi, M., Zhang, Y.Z., & Holmes, E.C. (2018a). Meta-transcriptomics and the evolutionary biology of RNA viruses. *Virus Research*, 243, 83-90. doi: 10.1016/j.virusres.2017.10.016
- Shi, J., Hu, Z., Deng, F., & Shen, S. Tick-Borne Viruses. (2018b). *Virologica Sinica*, 33 (1), 21-43. doi: 10.1007/s12250-018-0019-0
- SIB (Swiss Institute of Bioinformatics) - <https://www.sib.swiss/>
- Sicard, A., Michalakakis, Y., Gutiérrez, S., & Blanc, S. (2016). The strange lifestyle of multipartite viruses. *PLoS Pathology*, 12 (11), 1-19. doi: 10.1371/journal.ppat.1005819
- Sievers, F., & Higgins, D.G. (2014). Clustal omega. *Current Protocols in Bioinformatics*, 48 (3.13.), 1-16. doi: 10.1002/0471250953.bi0313s48
- Simion, P., Philippe, H., Baurain, D., Jager, M., Richter, D.J., Di Franco, A., Roure, B., Satoh, N., Quéinnec, E., Ereskovsky, A., Lapébie, P., Corre, E., Delsuc, F., King, N., Wörheide, G., & Manuel, M. (2017). A Large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. *Current Biology*, 27 (7), 958-967. doi: 10.1016/j.cub.2017.02.031
- Simmonds, P., Adams, M.J., Benkő, M., Breitbart, M., Brister, J.R., Carstens, E.B., Davison, A.J., Delwart, E., Gorbalenya, A.E., Harrach, B., Hull, R., King, A.M., Koonin, E.V., Krupovic, M., Kuhn, J.H., Lefkowitz, E.J., Nibert, M.L., Orton, R., Roossinck, M.J., Sabanadzovic, S., Sullivan, M.B., Suttle, C.A., Tesh, R.B., van der Vlugt, R.A., Varsani, A., & Zerbini, F.M. (2017). Consensus statement: Virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology*, 15 (3), 161-168. doi: 10.1038/nrmicro.2016.177
- Smith, E.C., Popa, A., Chang, A., Masante, C., & Dutch, R.E. (2009). Viral entry mechanisms: the increasing diversity of paramyxovirus entry. *FEBS Journal*, 276 (24), 7217-7227. doi: 10.1111/j.1742-4658.2009.07401.x
- Sobhy, H. (2016). A Review of Functional Motifs Utilized by Viruses. *Proteomes*, 4 (1), 1-21. doi: 10.3390/proteomes4010003
- Sousa, I.P. (Jr.), Carvalho, C.A.M., Mendes, Y.S., Weissmuller, G., Oliveira, A.C., & Gomes, A.M.O. (2017). Fusion of a new world Alphavirus with membrane microdomains involving partially reversible conformational changes in the viral spike proteins. *Biochemistry*, 56 (43), 5823-5830. doi: 10.1021/acs.biochem.7b00650
- Sterk, P., Kulikova, T., Kersey, P., & Apweiler, R. (2007). The EMBL Nucleotide Sequence and Genome Reviews Databases. *Methods in Molecular Biology*, 406, 1-21.
- Suttle, C.A. (2007). Marine viruses--major players in the global ecosystem. *Nature Reviews Microbiology*, 5 (10), 801-812.
- Suzuki, Y., Frangeul, L., Dickson, L.B., Blanc, H., Verdier, Y., Vinh, J., Lambrechts, L., & Saleh, M.C. (2017). Uncovering the repertoire of endogenous flaviviral elements in Aedes mosquito genomes. *Journal of Virology*, 91 (15), 1-20. doi: 10.1128/JVI.00571-17
- Stauffer, S., Feng, Y., Nebioglu, F., Heilig, R., Picotti, P., & Helenius, A. (2014). Stepwise priming by acidic pH and a high K⁺ concentration is required for efficient uncoating of influenza A virus cores after penetration. *Journal of Virology*, 88 (22), 13029-13046. doi: 10.1128/JVI.01430-14
- Stockman, L.J., Bellamy, R., & Garner, P. (2006). SARS: systematic review of treatment effects. *PLoS Medicine*, 3 (9), 1525-1531.
- te Velthuis, A.J., van den Worm, S.H., & Snijder, E.J. (2012). The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. *Nucleic Acids Research*, 40 (4), 1737-1747. doi: 10.1093/nar/gkr893
- te Velthuis, A.J. (2014). Common and unique features of viral RNA-dependent polymerases. *Cellular and Molecular Life Sciences*, 71 (22), 4403-4420. doi: 10.1007/s00018-014-1695-z
- The UniProt Consortium. (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, 45 (D1), 158-169. doi: 10.1093/nar/gkw1099
- Thézé, J., Leclercq, S., Moumen, B., Cordaux, R., & Gilbert, C. (2014). Remarkable diversity of endogenous viruses in a crustacean genome. *Genome Biology and Evolution*, 6 (8), 2129-2140. doi: 10.1093/gbe/evu163
- Thierry, E., Guilligay, D., Kosinski, J., Bock, T., Gaudon, S., Round, A., Pflug, A., Hengrung, N., El Omari, K., Baudin, F., Hart, D.J., Beck, M., & Cusack, S. (2016). Influenza Polymerase Can Adopt an Alternative Configuration Involving a Radical Repacking of PB2 Domains. *Molecular Cell*, 61 (1), 125-137. doi: 10.1016/j.molcel.2015.11.016

- Tokuriki, N., Oldfield, C.J., Uversky, V.N., Berezhovsky, I.N., & Tawfik, D.S. (2009). Do viral proteins possess unique biophysical features? *Trends in Biochemical Science*, *34* (2), 53-59. doi: 10.1016/j.tibs.2008.10.009
- Toribio, R., & Ventoso, I. (2010). Inhibition of host translation by virus infection in vivo. *Proceedings of the National Academy of Sciences*, *107* (21), 9837-9842. doi: 10.1073/pnas.1004110107
- Traub, L.M., & Bonifacino, J.S. (2013). Cargo recognition in clathrin-mediated endocytosis. *Cold Spring Harbor Perspectives in Biology*, *5* (11), 1-23. doi: 10.1101/cshperspect.a016790
- Trinh, Q.D., Pham, N.T.K., Takada, K., Komine-Aizawa, S., & Hayakawa, S. (2018). Myelin oligodendrocyte glycoprotein-independent rubella infection of keratinocytes and resistance of first-trimester trophoblast cells to rubella virus in vitro. *Viruses*, *10* (1), 1-6. doi: 10.3390/v10010023
- Ullmer, W., & Semler, B.L. (2016). Diverse strategies used by picornaviruses to escape host RNA decay pathways. *Viruses*, *8* (12), 1-19. doi: 10.3390/v8120335
- Virgin, H. W. (2014) The virome in mammalian physiology and disease. *Cell*, *157* (1), 142-150. doi: 10.1016/j.cell.2014.02.032
- von Reumont, B.M., Jenner, R.A., Wills, M.A., Dell'ampio, E., Pass, G., Ebersberger, I., Meyer, B., Koenemann, S., Iliffe, T.M., Stamatakis, A., Niehuis, O., Meusemann, K., & Misof, B. (2012). Pancrustacean phylogeny in the light of new phylogenomic data: support for Remipedia as the possible sister group of Hexapoda. *Molecular Biology and Evolution*, *29* (3), 1031-1045. doi: 10.1093/molbev/msr270
- Waldron, F.M., Stone, G.N., & Obbard, D.J. (2017). Metagenomic sequencing suggests arthropod antiviral RNA interference is highly derived, and identifies novel viral small RNAs in a mollusk and a brown alga. *BioRxiv*, 1-36. doi: <https://doi.org/10.1101/166488>
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10* (1), 57-63. doi: 10.1038/nrg2484
- Wasik, B.R., & Turner, P.E. (2013). On the biological success of viruses. *Annual Reviews in Microbiology*, *67*, 519-541. doi: 10.1146/annurev-micro-090110-102833
- Webster, C.L., Waldron, F.M., Robertson, S., Crowson, D., Ferrari, G., Quintana, J.F., Brouqui, J.M., Bayne, E.H., Longdon, B., Buck, A.H., Lazzaro, B.P., Akorli, J., Haddrill, P.R., & Obbard, D.J. (2015). The discovery, distribution, and evolution of viruses associated with *Drosophila melanogaster*. *PLoS Biology*, *13* (7), 1-33. doi: 10.1371/journal.pbio.1002210
- Webster, C.L., Longdon, B., Lewis, S.H., & Obbard, D.J. (2016). Twenty-five new viruses associated with the Drosophilidae (Diptera). *Evolutionary Bioinformatics Online*, *12* (Suppl 2), 13-25. doi: 10.4137/EBO.S39454
- Whitfield, A.E., Falk, B.W., & Rotenberg, D. (2015). Insect vector-mediated transmission of plant viruses. *Virology*, *479-480*, 278-289. doi: 10.1016/j.virol.2015.03.026
- Wilfert, L., Long, G., Leggett, H.C., Schmid-Hempel, P., Butlin, R., Martin, S.J., & Boots, M. (2016). Deformed wing virus is a recent global epidemic in honeybees driven by Varroa mites. *Science*, *351* (6273), 594-597. doi: 10.1126/science.aac9976
- Wit, E., van den Heuvel, E., & Romeijn, J.-W. (2012). All models are wrong...: an introduction to model uncertainty. *Statistica Neerlandica*, *66* (3), 217-236. doi: 10.1111/j.1467-9574.2012.00530.x
- Wudiri, G.A., Schneider, S.M., & Nicola, A.V. (2017). Herpes simplex virus 1 envelope cholesterol facilitates membrane fusion. *Frontiers in Microbiology*, *8* (2383), 1-10. doi: 10.3389/fmicb.2017.02383
- Xia, H., Wang, Y., Atoni, E., Zhang, B., & Yuan, Z. (2018). Mosquito-Associated Viruses in China. *Virologica Sinica*, *33* (1), 5-20. doi: 10.1007/s12250-018-0002-9
- Xiong, J.P., Stehle, T., Diefenbach, B., Zhang, R., Dunker, R., Scott, D.L., Joachimiak, A., Goodman, S.L., & Arnaout, M.A. (2001). Crystal structure of the extracellular segment of integrin $\alpha V\beta 3$. *Science*, *294* (5541), 339-345.
- Yang, Z., & Rannala, B. (2012). Molecular phylogenetics: principles and practice. *Nature Reviews Genetics*, *13* (5), 303-314. doi: 10.1038/nrg3186
- Yang, S.T., Kreutzberger, A.J.B., Kiessling, V., Ganser-Pornillos, B.K., White, J.M., & Tamm, L.K. (2017). HIV virions sense plasma membrane heterogeneity for cell entry. *Science Advances*, *3* (6), 1-12. doi: 10.1126/sciadv.1700338

- Zanotto, P.M., Gibbs, M.J., Gould, E.A., & Holmes, E.C. (1996). A reevaluation of the higher taxonomy of viruses based on RNA polymerases. *Journal of Virology*, 70 (9), 6083-6096

Bibliografija

Objave v povezavi z delom

Članki v revijah

- Ott Rutar, S., & Kordiš, D. (2020). Analysis of the RNA virome of basal hexapods. *Peer J* 8: e8336. doi: 10.7717/peerj.8336 (published).
- Ott Rutar, S., & Kordiš, D. (2020). Discovery and molecular characterisation of the first ambidensovirus in honey bees. *Acta agriculturae Slovenica* (accepted).

Prispevki na konferencah

- OTT, S., KORDIŠ, D. (2017). Phylogenomic analysis of RNA viruses in invertebrates. V: KLJUN, Jakob (ur.), GABER, Aljaž (ur.). Book of abstracts, Scientific Conference for Young Researchers, 19. 9. 2017. Ljubljana: Fakulteta za kemijo in kemijsko tehnologijo. 2017, str. 31. [COBISS.SI-ID 31285799]
- OTT, S., KORDIŠ, D. Phylogenomic analysis of RNA viruses in invertebrates. V: PAVLIN, Majda (ur.), et al. Zbornik = Proceedings, 9. študentska konferenca Mednarodne podiplomske šole Jožefa Stefana in 11. dan mladih raziskovalcev (Konferenca KMBO) = 9th Jožef Stefan International Postgraduate School Students' Conference and 11th Young researchers' Day, [19.-20. 4. 2017, Ljubljana, Slovenija]. Ljubljana: Mednarodna podiplomska šola Jožefa Stefana: = Jožef Stefan International Postgraduate School: Inštitut Jožef Stefan: = Jožef Stefan Institute. 2017, str. 58. El.virnanaslovu http://ipssc.mps.si/proceedings/Proceedings_IPSSC+CMBE_2017.pdfnivečdostopen (15.3.2018). [COBISS.SI-ID 30474023]

Biografija

Sabina Ott Rutar je kot zlata maturantka maturo zaključila leta 2008 z vsemi možnimi točkami (34/34) in najvišjim odstotkom doseženih točk pri biologiji v Sloveniji (99 %). Dodiplomski študij 1. bolonjske stopnje (BSc) – smer biotehnologija – je zaključila na Biotehniški fakulteti Univerze v Ljubljani leta 2011 in s tem pridobila naziv univerzitetni diplomirani biotehnolog.

Oktobra 2011 se je vpisala na podiplomski študij 2. bolonjske stopnje (Msc) – splošna biotehnologija na Biotehniški fakulteti. Študij je zaključila septembra 2013 in pridobila naziv magistra biotehnologije.

Tekom študija je pridobila priznanje Biotehniške fakultete za izvrsten študijski uspeh na dodiplomskem študiju prve stopnje s povprečno oceno 9,48, potrdilo o uvrstitvi med 5 % najboljših študentov v generaciji in fakultetno Prešernovo nagrado za delo: »*Razvoj novega raziskovalnega področja Fraktalomike s sintezo podatkov fraktalne geometrije v medicini in molekularni biologiji*«.

Leta 2014 se je vpisala na podiplomski študij 3. bolonjske stopnje (PhD) na Mednarodni podiplomski šoli Jožefa Stefana – smer Nanoznanosti in Nanotehnologije (NANO3). Novembra 2014 se je zaposlila kot mlada raziskovalka (MR) na »Institutu Jožef Stefan«.