EPIDEMIOLOGY, GENETIC DIVERSITY, AND COMPARATIVE GENOMICS OF 16SRV PHYTOPLASMAS INFECTING GRAPEVINE AND HAZELNUT

Zala Kogej Zwitter

Doctoral Dissertation
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia

Supervisor: Assoc. Prof. Dr. Nataša Mehle, National Institute of Biology, Ljubljana, Slovenia
Co-Supervisor: Assist. Prof. Dr. Denis Kutnjak, National Institute of Biology, Ljubljana, Slovenia

Evaluation Board:
Assist. Prof. Dr. Špela Baebler, Chair, National Institute of Biology and Jožef Stefan International Postgraduate School, Ljubljana, Slovenia
Prof. Dr. Jernej Jakše, Member, Biotechnical Faculty, University of Ljubljana, Slovenia
Prof. Dr. Martina Šeruga Musić, Member, Faculty of Science, University of Zagreb, Croatia

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL



Zala Kogej Zwitter

EPIDEMIOLOGY, GENETIC DIVERSITY, AND COMPARATIVE GENOMICS OF 16SRV PHYTOPLASMAS INFECTING GRAPEVINE AND HAZELNUT

Doctoral Dissertation

EPIDEMIOLOGIJA, GENETSKA RAZNOLIKOST IN PRIMERJALNA GENOMIKA FITOPLAZEM SKUPINE 16SRV, KI OKUŽUJEJO TRTO IN LESKO

Doktorska disertacija

Supervisor: Assoc. Prof. Dr. Nataša Mehle

Co-Supervisor: Assist. Prof. Dr. Denis Kutnjak

Ljubljana, Slovenia, June 2026

To all wonderers, big and small, who stop, look, and think.

Acknowledgments

This thesis would not have been possible without the support, knowledge, and dedication of many people who accompanied me during my doctoral studies. First, I would like to thank the researchers and technicians of the Department of Biotechnology and Systems Biology at the National Institute of Biology. Their contributions during meetings, laboratory work, fieldwork, and everyday discussions shaped this thesis and my development as a scientist. I am grateful to everyone in the FITO department for sharing their knowledge, experience, and advice, and for creating a supportive working environment.

I am especially grateful to my supervisor, Nataša Mehle, and my co-supervisor, Denis Kutnjak. Together, they provided an invaluable combination of expertise, guidance and scientific enthusiasm. Their doors were always open, and their comments on my manuscripts and thesis helped me improve and learn. I could always rely on their guidance when difficulties arose. I would also like to thank the members of my dissertation committee, Špela Baebler, Jernej Jakše, and Martina Šeruga Musić, for their careful reading, constructive comments, and timely support during the final stages of this dissertation.

During my PhD, I also spent part of my research training at INRAE in Bordeaux, in Xavier Foissac's team, supported by a FEMS Research and Training Grant (FEMS-GO-2021-070). I am grateful to Xavier, Sylvie, and the MOLLICUTES team for welcoming me, sharing their knowledge of grapevine diseases, and broadening my scientific perspective.

I want to express my sincere thanks to Gabrijel Seljak, who accompanied us during fieldwork and leafhopper collection, even after his retirement. His exceptional expertise and enthusiasm were essential for including the insect vector component in this thesis. I am also grateful to all hazelnut growers who allowed us to sample in their orchards and who were open to a scientific approach to understanding hazelnut decline, even though concrete solutions are not yet available. Throughout this work, I tried to keep in mind the people facing these problems in practice.

This research was funded by the Slovenian Research and Innovation Agency (ARIS) through a Young Researcher grant and the core financing program P4-0463. The NanoPhyto project (L7-2632) was funded as a Slovenian-Croatian research project by ARIS and the Croatian Science Foundation (HRZZ), with additional Slovenian support from the Ministry of Agriculture, Forestry and Food and BIA d.o.o. Additional support was provided by the Slovenian Ministry of Agriculture, Forestry and Food through the Administration of the Republic of Slovenia for Food Safety, Veterinary Sector and Plant Protection, including support for the Slovenian contribution to the Euphresco projects FLADO-VIGILANT (2020-A-344) and VIRNOTree (2023-A-446). Part of the work for the first reports was performed within the framework of the European Union Reference Laboratory for pests of plants on viruses, viroids and phytoplasmas, financed by the European Commission, Directorate-General for Health and Food Safety. Research infrastructure was supported by ARIS grant IO-0004, and computational resources were provided by the Slovenian node of ELIXIR ERIC.

I am also grateful to my lunch break team for the conversations, encouragement, and laughter that made even the most demanding periods easier, and to my office mates, Vid and Neža, for always making me feel competent and ready to tackle any task. Finally, I would like to thank my friends and family. Thank you for your interest in my work, for asking questions, and for encouraging me to explain science as clearly and simply as possible. Because of you, I became a better science communicator. Most of all, thank you for your patience, support, and love throughout this journey.

Abstract

Flavescence dorée (FD) phytoplasma and related 16SrV phytoplasmas are associated with grapevine disease, hazelnut decline, and non-cultivated plants. Their epidemiology is commonly investigated using gene-based typing, especially genotyping of the *map* gene, which enables rapid comparison of closely related strains. However, identical *map* genotypes in different hosts do not prove genome-level relatedness. In this thesis, we examined the distribution of 16SrV phytoplasmas in grapevine, hazelnut, and two leafhoppers (*Scaphoideus titanus* and *Orientus ishidae*). We evaluated whether selected cross-host relationships inferred from *map* genotyping are supported by genome-wide data.

The first objective was to use *map* genotyping and additional genetic markers to compare 16SrV phytoplasmas from grapevine, hazelnut, and leafhoppers in Slovenia. It revealed a partial overlap between grapevine and hazelnut. Some *map* genotypes previously identified in grapevine outbreaks were also found in hazelnut, and the same genotypes were present in *O. ishidae* collected from infected hazelnut orchards. First reports from the Czech Republic and the Netherlands generated in this work provided additional European context: FD-related phytoplasmas were detected in grapevine and *Clematis vitalba*, while a non-epidemic 16SrV genotype was found in asymptomatic alder. These findings suggest that hazelnut and non-cultivated plants in the studied landscapes contribute to the ecological background in which FD-related 16SrV phytoplasmas persist. However, only part of this diversity is directly relevant to grapevine epidemics.

To extend single-locus genotyping towards genome-wide comparison, the second objective addressed sequencing directly from naturally infected plant material. Because phytoplasmas cannot be cultured in cell-free media and occur at low titre in plant tissues, field-sample sequencing is constrained by host DNA background. We developed an enrichment protocol using naturally infected grapevine material. The optimised approach increased phytoplasma read recovery and enabled reconstruction of a high-quality draft genome without propagation in test plants. This enabled the transition from marker-based comparison to genome-wide analysis.

The third objective applied genome-wide analyses to FD-related 16SrV phytoplasmas from grapevine, hazelnut, *S. titanus*, and *O. ishidae*. Genome-wide similarity patterns were consistent with *map* genotype classification for M38 and M50, indicating that *map*-based classification captured major genomic relationships in this dataset. Host species did not correspond to genome-wide differentiation, while variation within *map* genotype-related groups indicated an influence of geographic origin. Genome-scale analysis, therefore, added resolution beyond *map* typing by showing that, for the studied *map* genotypes, shared genotypes across hosts were also closely related at the genome level, while finer variation was linked more strongly to sampling area than to host.

The thesis supports the interpretation that, for the analysed *map* genotypes and sampled regions, FD-related 16SrV phytoplasmas occur in a shared multi-host system in which hazelnut is part of the epidemiological landscape rather than an isolated host. It provides a method for directly sequencing phytoplasmas from field material, enabling future comparative genomic studies without prior propagation in test plants.

Povzetek

Fitoplazma zlate trsne rumenice in sorodne fitoplazme iz skupine 16SrV so povezane z boleznimi vinske trte, propadanjem leske ter okužbami samoniklih rastlin. Njihova epidemiologija se pogosto proučuje z gensko tipizacijo, zlasti genotipizacijo gena *map*, saj omogoča hitro primerjavo sorodnih izolatov. Vendar identični genotipi *map* v različnih gostiteljih sami po sebi ne pokažejo, ali so fitoplazme sorodne tudi na ravni genoma. V tej nalogi smo preverjali razširjenost fitoplazem iz skupine 16SrV v vinski trti, leski ter dveh vrstah škržatkov (*Scaphoideus titanus* in *Orientus ishidae*) ter ocenili, ali izbrani odnosi med gostitelji, sklepani na podlagi gena *map*, dobijo podporo tudi v genomskih podatkih.
V prvem cilju smo z genotipizacijo gena *map* in dodatnimi genskimi markerji primerjali fitoplazme iz skupine 16SrV iz vinske trte, leske in škržatkov v Sloveniji. Analize so pokazale le delno prekrivanje med vinsko trto in lesko. Nekateri genotipi, prej povezani z epifitocijami na vinski trti, so bili odkriti tudi v leski, isti genotipi pa v *O. ishidae*, ujetih v okuženih nasadih leske. Prvi poročili iz Češke republike in Nizozemske, pripravljeni v okviru te naloge, sta podali dodatne podatke iz drugih evropskih držav: fitoplazme, sorodne fitoplazmi zlate trsne rumenice, so bile odkrite v vinski trti in navadnem srobotu (*Clematis vitalba*), neepidemični genotip iz skupine 16SrV pa v asimptomatski jelši. Ti rezultati kažejo, da leska in samonikle rastline v preučevanih okoljih prispevajo k ekološkemu ozadju, v katerem se sorodne fitoplazme iz skupine 16SrV ohranjajo, čeprav je le del te raznolikosti neposredno pomemben za epifitocije na vinski trti.

Za nadgradnjo tipizacije posameznega gena v genomske primerjave smo v drugem cilju obravnavali metodološko oviro pri sekvenciranju neposredno iz naravno okuženega rastlinskega materiala. Ker fitoplazem ni mogoče gojiti in so v rastlinskih tkivih navadno prisotne v nizkem titru, sekvenciranje vzorcev omejuje gostiteljska DNA. Razvili in ovrednotili smo protokol obogatitve fitoplazemske DNA na naravno okuženi vinski trti. Optimizirani pristop je povečal delež odčitkov fitoplazme in omogočil rekonstrukcijo kakovostnega osnutka genoma brez predhodnega namnoževanja v testnih rastlinah. S tem je bil omogočen prehod od primerjav na podlagi markerjev k genomskim analizam.

V tretjem cilju smo genomske analize uporabili za primerjavo sorodnih fitoplazem iz skupine 16SrV iz vinske trte, leske, *S. titanus* in *O. ishidae*. Vzorci genomske podobnosti so bili pri genotipih M38 in M50 skladni s klasifikacijo na podlagi gena *map*, kar kaže, da je ta klasifikacija v tem naboru podatkov zajela glavne genomske odnose. Pri proučevanih genotipih so bili vzorci genomske podobnosti skladni s klasifikacijo na podlagi gena *map*, tudi kadar so bili isti genotipi odkriti v različnih gostiteljih; podrobnejše razlike pa so bile bolj povezane z območjem vzorčenja kot z gostiteljem.

Doktorska naloga podpira razlago, da se za analizirane genotipe *map* in vzorčena območja fitoplazme iz skupine 16SrV, sorodne fitoplazmi zlate trsne rumenice, pojavljajo v skupnem večgostiteljskem sistemu, v katerem je leska del epidemiološkega okolja in ne izoliran gostitelj. Poleg tega naloga prinaša metodo za sekvenciranje fitoplazem neposredno iz okoljskega materiala, kar omogoča prihodnje primerjalne genomske raziskave brez namnoževanja fitoplazem v testnih rastlinah.

Contents

Abbreviations **xv**

1 Introduction **1**
1.1 Phytoplasmas and Their Impact on Plant Health 1
1.2 Flavescence Dorée Phytoplasma 3
1.2.1 Disease overview and impact 3
1.2.2 Molecular epidemiology and genetic diversity 3
1.2.3 Vector biology and FDp circulation in the environment 5
1.3 Sequencing of Phytoplasma Genomes 7
1.4 Scientific Problems and Research Aims 8
1.4.1 Assessing gene-based similarity and epidemiological connections among 16SrV phytoplasmas in grapevine, hazelnut, and leafhoppers 8
1.4.2 Developing and optimising enrichment strategies for phytoplasma genome sequencing from naturally infected plant material 9
1.4.3 Comparing genome-wide relationships of 16SrV phytoplasmas across hosts using genome sequencing 10
1.5 Research Hypotheses 10
1.6 Publications Included and Candidate's Contributions 11

2 Scientific Publications **13**
2.1 Epidemiology of Flavescence Dorée and Hazelnut Decline in Slovenia: Geographical Distribution and Genetic Diversity of the Associated 16SrV Phytoplasmas 13
2.2 First Report of Grapevine Flavescence Dorée Phytoplasma in the Czech Republic 26
2.3 First Detection of a Phytoplasma From the 16SrV-C (Elm Yellows) Group in *Alnus Glutinosa* in the Netherlands 28
2.4 From Vineyard to Genome: Optimised Enrichment and Sequencing of Flavescence Dorée Phytoplasma from Grapevine Samples 32
2.5 Genome-Wide Comparisons Confirm Locally Shared Flavescence Dorée-Related Phytoplasma Populations Across Grapevine, Hazelnut, and Leafhoppers 49

3 Discussion **69**
3.1 Gene-Based Epidemiological Relationships of 16SrV Phytoplasmas in Grapevine, Hazelnut, and Leafhopper Vectors 69
3.2 Optimised Enrichment Strategy for Phytoplasma Genome Sequencing from Naturally Infected Plant Material 71
3.3 Genome-Wide Relationships of 16SrV Phytoplasmas Across Host Plants 72

4 Conclusions **75**

Appendix A Supplementary Material of Included Publications **77**

A.1 Supplementary Material for Publication 2.1 77
A.2 Supplementary Material for Publication 2.2 77
A.3 Supplementary Material for Publication 2.4 77
A.4 Supplementary Material for Publication 2.5 77

References **83**

Bibliography **91**

Biography **93**

Abbreviations

16S rRNA	16S ribosomal ribonucleic acid
ANI	average nucleotide identity
ATP	adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
'*Ca.* P.'	'*Candidatus* Phytoplasma'
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
Cq	quantification cycle
DNA	deoxyribonucleic acid
EU	European Union
FD	Flavescence dorée
FDp	Flavescence dorée phytoplasma
HTS	high-throughput sequencing
LAMP	loop-mediated isothermal amplification
map	methionine aminopeptidase
MLST	multilocus sequence typing
NGS	next-generation sequencing
ONT	Oxford Nanopore Technologies
PCoA	principal coordinates analysis
PCR	polymerase chain reaction
PMU	potential mobile unit
qPCR	Real-time PCR
RPLF	restriction fragment length polymorphism
SNP	single nucleotide polymorphisms
Vmp	variable membrane protein

Chapter 1

Introduction

1.1 Phytoplasmas and Their Impact on Plant Health

Phytoplasmas are cell wall-less bacteria that live as obligate parasites in plant phloem. Their transmission depends mainly on phloem-feeding insects, and replication occurs in both plant and insect hosts (Hogenhout et al., 2008; Namba, 2019). Within the class *Mollicutes*, phytoplasmas are assigned to the provisional genus '*Candidatus* Phytoplasma' ('*Ca.* P.'). The provisional status is linked to the inability to culture phytoplasmas in cell-free media, so their classification relies on molecular sequence data (IRPCM, 2004; Namba, 2019). They are pleomorphic, exhibiting size variation typically ranging from approximately 80 to 900 nm within phloem sieve elements (Marcone, 2014).

Phytoplasma classification has largely relied on analysis of the 16S ribosomal RNA (16S rRNA) gene, which led to the establishment of the 16Sr group and subgroup system (IRPCM, 2004; Namba, 2019). This system classifies strains according to restriction fragment length polymorphism (RFLP) patterns derived from 16S rRNA gene sequences (Lee et al., 1998), obtained either through laboratory-based assays or *in silico* simulation of restriction patterns (Zhao et al., 2013). This system remains widely used in diagnostics and in epidemiological studies. However, the resolution of the 16S rRNA gene is often insufficient to distinguish closely related strains. For this reason, multilocus analyses often include additional molecular markers such as *tuf*, *secA*, *secY*, *groEL*, ribosomal protein genes, and other (Wei & Zhao, 2022).

In host plants, phytoplasma infections may cause morphological and physiological symptoms relevant for disease detection. Symptoms vary depending on the phytoplasma strain, host species and cultivars, plant developmental stage, timing of infection, pathogen titre, and environmental conditions (Lee et al., 2000; Marcone, 2014). Typical symptoms include virescence, phyllody, big bud formation, and floral proliferation, often leading to sterility. Other commonly reported symptoms include witches' broom, rosetting, abnormal shoot development, and shortening of internode length. In woody hosts, infections often appear as a general decline, including foliar yellowing or reddening, reduced leaf size, leaf curling or rolling, vein abnormalities, premature senescence, reduced fruit size, dieback, and overall stunting (Marcone, 2014; Namba, 2019). In some cases, infections may remain asymptomatic, complicating disease detection and epidemiological assessments (Holz et al., 2016). The development and progression of symptoms are closely linked to the systemic colonisation of host plants and their subsequent role as sources of inoculum in insect vector-mediated transmission (Marcone, 2014).

Phloem-feeding insect vectors are the main route of phytoplasma spread between host plants (Hogenhout et al., 2008). By the mid-2000s, over 90 insect species had been identified as vectors of phytoplasmas, some of which can transmit multiple phytoplasmas (Weintraub

& Beanland, 2006). The main vectors belong to the order Hemiptera, particularly leafhoppers, planthoppers, and psyllids, which acquire phytoplasmas while feeding on infected plants and subsequently transmit them to new hosts (Weintraub & Beanland, 2006). Transmission occurs in a persistent-propagative manner. After ingestion of phloem sap, phytoplasmas cross the gut epithelium, circulate in the hemocoel, and multiply before reaching the salivary glands, from which they are inoculated into plants during feeding. Following an incubation period of approximately one month, infected vectors remain infectious for life (Hogenhout et al., 2008).

Phytoplasmas can also spread through several non-vector routes. One such route involves parasitic plants of the genus *Cuscuta* (dodder), which form phloem connections between host plants and facilitate the transfer of phytoplasmas from infected to healthy plants (Bertaccini & Duduk, 2009). Phytoplasma transmission can also occur through natural root grafts or root bridges between neighbouring plants, in which direct phloem connections enable the pathogen to move between hosts (Ciccotti et al., 2007). Phytoplasmas may also be spread through vegetative propagation, including grafting, cuttings, micropropagation, and other clonal propagation techniques, thereby enabling the persistence and dissemination of infections in cultivated plant material (Dickinson et al., 2013; Lee et al., 2000).

These transmission dynamics, together with their broad host range, contribute to the impact of phytoplasmas on plant health and agriculture, and they are responsible for severe yield losses in a wide variety of crops worldwide (Bertaccini, 2007; Lee et al., 2000). Phytoplasmas are distributed globally, although the prevalence of specific taxonomic groups and subgroups varies across geographic regions (Marcone, 2014). They have been reported to infect more than 1,000 plant species, including major food and oil crops, vegetables, fruit trees, ornamental plants, forest species, and weeds, resulting in substantial economic losses in agricultural and natural ecosystems (Bertaccini, 2007; Marcone, 2014; Namba, 2019). The broad host range of phytoplasmas is closely linked to the feeding behaviour and host preferences of their insect vectors (Marcone, 2014).

Disease outcome also depends on the combination of phytoplasma lineage, plant host, and insect vector. Phytoplasma lineages differ in host specificity and virulence, with closely related strains ranging from nearly avirulent to highly pathogenic (Marcone, 2010). Mixed infections can occur, particularly in perennial plants, where long lifespans increase the likelihood of repeated vector-mediated inoculation events (Marcone, 2014). Within plants, phytoplasma colonisation is typically described as phloem-associated and systemic; however, uneven pathogen titres and distribution in plant tissues can complicate symptom interpretation and detection (Hogenhout et al., 2008).

Reliable detection is central to disease management because phytoplasmas often occur at low titres and cause variable or nonspecific symptoms (Bertaccini & Duduk, 2009). Diagnostic methods have progressed from traditional microscopy and symptom-based identification to highly sensitive molecular techniques, including nested PCR, real-time PCR (qPCR), loop-mediated isothermal amplification (LAMP), and, more recently, CRISPR-based detection systems, which enable rapid and specific identification of phytoplasmas in both plant and insect hosts (Namba, 2019; Xu et al., 2026).

Disease management remains difficult because effective curative treatments are lacking, and epidemiology is often complex (Namba, 2019). The broad host ranges of both phytoplasmas and their insect vectors increase the unpredictability of disease outbreaks, while long incubation periods in plant and insect hosts often delay detection until significant pathogen spread has occurred (Hogenhout et al., 2008). These factors reduce the effectiveness of reactive control measures and complicate efforts to disrupt disease cycles (Firrao et al., 2007). Current management, therefore, relies mainly on prevention,

including vector control, removal of infected plants, and the use of certified planting material, supported by monitoring programs and regulatory measures (Bianco et al., 2019).

1.2 Flavescence Dorée Phytoplasma

1.2.1 Disease overview and impact

Flavescence dorée phytoplasma (FDp) belongs to the 16SrV phytoplasma group, mainly subgroups 16SrV-C and 16SrV-D. It causes a severe grapevine disease in Europe and is regulated as a quarantine pest (Casati et al., 2017; Chuche & Thiéry, 2014; EU Comission, 2019; Plavec et al., 2019). FDp belongs to the grapevine yellows complex, in which similar symptoms can result from different causal agents and epidemiological cycles (Constable, 2009). Flavescence dorée (FD) is one of the most damaging grapevine yellows diseases because infection can reduce yield and grape quality by affecting photosynthesis and nutrient transport (Chuche & Thiéry, 2014; Credi, 1989). In susceptible cultivars, infection often results in vine death within a few years (Filippin et al., 2009). Economic losses also arise from mandatory control measures, including insecticide applications and the removal of infected plants (Chuche & Thiéry, 2014). In traditional viticultural regions, FDp also threatens grapevine genetic resources and cultural heritage (Plavec et al., 2019). Its ecological complexity, involving interactions between cultivated and wild compartments, further distinguishes FDp as a notable open pathosystem, characterised by epidemiological connectivity between vineyards and surrounding non-cultivated or unmanaged host plants, as well as the involvement of alternative insect vectors (Angelini et al., 2004; Arnaud et al., 2007; Casati et al., 2017; Filippin et al., 2009; Filippin & De Pra, 2011; Malembic-Maher et al., 2020).

FDp spread involves two main processes: movement of infected planting material and local vector transmission. The disease was first reported in south-western France in the 1950s and has since expanded across multiple European countries (Bonfils & Schvester, 1960; Malembic-Maher et al., 2020). Infected planting material is considered the main route of long-distance spread. The occurrence of highly clonal FDp strains supports this role of the grapevine trade (Arnaud et al., 2007; Bertin et al., 2007; Papura et al., 2012). At local and regional scales, vector-borne transmission amplifies infections within and between vineyards (Chuche & Thiéry, 2014). Epidemics develop where vector activity, infected vines, and connected surrounding vegetation allow repeated pathogen movement (Oggier et al., 2024). Molecular tools are therefore used to characterise the FDp genetic structure and reconstruct the spread across affected viticultural regions.

1.2.2 Molecular epidemiology and genetic diversity

Molecular epidemiological approaches are used to resolve how vectors, hosts and human activities shape the FDp population structure and spread. Multilocus sequence typing (MLST) provides the resolution needed to distinguish closely related strains that cannot be differentiated using the highly conserved 16S rRNA gene (Arnaud et al., 2007; Casati et al., 2017). The methionine aminopeptidase (*map*) gene is one of the most epidemiologically informative loci and forms the basis for classifying FDp into three principal genetic clusters: Map-FD1, Map-FD2, and Map-FD3 (Arnaud et al., 2007). Several *map* genotypes within these clusters are closely associated with specific non-grapevine hosts, such as alder (*Alnus* spp.) and *Clematis vitalba*, which act as natural

reservoirs of FDp-related strains (Malembic-Maher et al., 2020) and are discussed in detail below.

The Map-FD1 cluster, closely related to the historical reference strain FD70 and associated with the 16SrV-C subgroup, exhibits relatively low genetic variability compared to other clusters (Malembic-Maher et al., 2020). However, this cluster also includes several epidemiologically relevant genotypes, such as M50, M27, M34, and M112, which can sustain epidemic spread in vineyards (Malembic-Maher et al., 2020). M50 is a rare example of a *map* genotype that is both epidemic and frequently detected in alder populations, suggesting a direct ecological link between natural reservoirs and the grapevine pathosystem (Malembic-Maher et al., 2020). In contrast, the Map-FD2 cluster, which includes reference strains FD92 and FD-D and spans both 16SrV-C and 16SrV-D subgroups, is one of the most widespread lineages across Europe (Malembic-Maher et al., 2020). Its high genetic homogeneity, particularly for *map* genotype M54, provides strong evidence for human-mediated dissemination via infected planting material (Malembic-Maher et al., 2020; Plavec et al., 2019). Within the same cluster, closely related genotypes such as M38 are predominantly associated with alder, indicating that epidemic vineyard strains may have originated from ancestral populations maintained in non-cultivated hosts (Malembic-Maher et al., 2020). Additional *map* genotypes identified in this cluster, including M89, M148, and M155, further support the hypothesis of repeated transitions from alder-based cycles into the grapevine–vector cycle (Krstić et al., 2022). The Map-FD3 cluster, represented by the Italian reference strain FD-C and restricted to the 16SrV-C subgroup, displays a higher level of genetic variability (Arnaud et al., 2007; Rossi et al., 2019). Unlike the other clusters, Map-FD3 is more closely associated with *C. vitalba* as a natural reservoir, with several genotypes, such as M51 and M144, detected in both non-cultivated hosts and cultivated grapevines, reflecting a more complex and locally structured epidemiology (Filippin et al., 2009; Krstić et al., 2022; Malembic-Maher et al., 2020).

Across all clusters, the FDp population is highly diverse, with at least 173 distinct *map* genotypes (M1–M173) identified to date (INRAE *map* sequence collection, Bordeaux; S. Malembic-Maher, personal communication, 12 June 2025). However, only a small fraction of these (approximately 11-20 genotypes) are associated with epidemic outbreaks in vineyards (Krstić et al., 2022; Malembic-Maher et al., 2020). This distinction is closely linked to their compatibility with the primary vector, *Scaphoideus titanus*, which is linked to variation in the variable membrane protein genes (*vmpA* and *vmpB*) (Malembic-Maher et al., 2020). These genes encode surface-exposed transmembrane proteins that function as adhesins, enabling phytoplasmas to bind to and colonise epithelial tissues of the insect midgut and salivary glands (Arricau-Bouvery et al., 2018). Structural variation within these genes, particularly in the number and arrangement of repeated domains, is considered a key mechanism of vector adaptation, allowing certain phytoplasma strains to establish transmission cycles in different insect hosts (Arricau-Bouvery et al., 2021).

Variation in *vmpA* and *vmpB* is used to classify FDp strains into three main vectotype clusters that reflect differences in vector compatibility and transmission potential (Malembic-Maher et al., 2020). Vectotype I (Vmp-I) comprises genotypes typically associated with alder populations and transmitted by alternative vectors such as *Oncopsis alni* (Malembic-Maher et al., 2020). These strains, which include numerous alder-specific *map* genotypes (e.g., M52, M53, M43, M45, M47, M48, and M49), are generally incompatible with *S. titanus* and therefore considered to pose low epidemiological risk in vineyard systems (Malembic-Maher et al., 2020). In contrast, Vectotypes II and III include genotypes fully compatible with *S. titanus* and responsible for epidemic spread in grapevine (Malembic-Maher et al., 2020). Vectotype II encompasses highly widespread epidemic strains such as the M54 genotype, as well as alder-associated genotypes like M38, highlighting potential evolutionary transitions from wild reservoirs (Malembic-Maher et

al., 2020). Vectotype III includes genotypes such as M50 and M51, which are also efficiently transmitted by *S. titanus* and are regularly detected in non-cultivated hosts, such as *C. vitalba* (Malembic-Maher et al., 2020).

Combining *map* genotyping with vectotype classification provides a functional framework for interpreting FDp epidemiology. While *map* clusters (Map-FD1, Map-FD2, and Map-FD3) describe genetic lineages and their geographic and ecological distribution, vectotypes define the biological capacity of those lineages to participate in FDp epidemics in vineyards (Krstić et al., 2022; Malembic-Maher et al., 2020). This distinction has direct relevance for disease management, as detection of Vectotype I strains suggests a low epidemiological risk for vine-to-vine transmission, whereas the presence of Vectotypes II or III indicates high epidemic potential and the need for vector control measures (Malembic-Maher et al., 2020). Additional loci further refine epidemiological analyses and trace transmission pathways at finer spatial scales. Commonly employed markers include *secY* and *uvrB-degV*, which improve resolution within ribosomal groups (Arnaud et al., 2007; Plavec et al., 2019), while more variable genes such as *malG* and *dnaK* enable discrimination of strains at the vineyard or agroecosystem level (Rossi et al., 2019). In particular, *dnaK* has proven useful for distinguishing major epidemic lineages, inferring movements of strains between cultivated and wild compartments, and detecting rare mixed infections, thereby providing further insight into population structure and strain interactions not captured by *map* genotyping alone (Rossi et al., 2019).

1.2.3 Vector biology and FDp circulation in the environment

The leafhopper *S. titanus* is a North American species that became established in Europe and is now the main vineyard vector of FDp (Malembic-Maher et al., 2020). In Europe, this species is closely associated with the genus *Vitis*, completing its entire life cycle on the grapevine (Chuche & Thiéry, 2014). Eggs are laid under the bark, and subsequent nymphal development and adult activity occur exclusively on the grapevine (Chuche & Thiéry, 2014). This transmission process makes FDp highly epidemic, as it can lead to rapid disease development with strong annual increases in infection levels reported in the absence of control measures (Chuche & Thiéry, 2014). *S. titanus* disperses over relatively short distances, with most individuals remaining within 25–30 m of their point of origin, although occasional flights of several hundred meters have been recorded (Chuche & Thiéry, 2014; Lessio & Alma, 2004).

Although *S. titanus* is the only vector currently known to sustain vineyard-scale epidemics, other species such as *O. alni*, *Dictyophara europaea,* and *Orientus ishidae* have been shown to transmit FDp under specific conditions, particularly in association with non-grapevine host plants (Filippin et al., 2009; Krstić et al., 2018; Lessio et al., 2016; Malembic-Maher et al., 2020). Among these, *O. ishidae*, an invasive polyphagous leafhopper of East Asian origin, is relevant because of its broad host range and expanding distribution in Europe (Cvrković et al., 2021). Unlike *S. titanus*, which is largely restricted to grapevine, *O. ishidae* feeds on a broad spectrum of woody and herbaceous plants, including several known or suspected non-cultivated hosts of FD-related phytoplasmas (Cvrković et al., 2021). Its broad host range allows contact with both non-cultivated vegetation and vineyard systems.

Field surveys have repeatedly detected FDp and FD-related phytoplasmas in *O. ishidae* populations, indicating acquisition under natural conditions (Gaffuri et al., 2011; Malembic-Maher et al., 2020; Mehle et al., 2010; Rizzoli et al., 2021). Experimental transmission trials have also shown that it can transmit FDp from non-grapevine hosts, such as broad bean, to grapevine. However, its transmission efficiency appears lower and more variable than that of *S. titanus* (Lessio et al., 2016). Earlier studies mainly reported

O. ishidae carrying *map* genotypes associated with wild non-cultivated plants; however, recent data have reported individuals harbouring the M54 genotype, one of the most widespread epidemic strains typically transmitted by *S. titanus*, suggesting that *O. ishidae* may also acquire epidemic strains directly from grapevines (Oggier et al., 2024). These data support a secondary role for *O. ishidae*: not as the main epidemic vector in vineyards, but as a possible contributor to circulation between vineyards and surrounding vegetation.

Outside cultivated vineyards, FDp and related phytoplasmas persist in several non-cultivated hosts. Grapevine itself, including American rootstocks and hybrids such as *Vitis labrusca* and *Vitis riparia*, can, in some cultivars and rootstocks, serve as asymptomatic carriers, maintaining the pathogen within agricultural systems (Rossi et al., 2019). "Gone-wild" grapevines and abandoned vineyards are also important because they can provide both inoculum sources and habitats for vectors (Oggier et al., 2024). Such unmanaged sites may therefore support persistence and repeated reintroduction into managed vineyards. In natural environments, several woody and herbaceous species help maintain phytoplasma populations. Alder species (*Alnus glutinosa* and *Alnus incana*) are considered key reservoirs for specific genetic clusters, particularly those associated with non-epidemic cycles (Malembic-Maher et al., 2020; Mehle, Rupar, et al., 2011), while *C. vitalba* represents an important host linked to Map-FD3-related strains across central and southern Europe (Angelini et al., 2004; Cvrković et al., 2021; Filippin et al., 2009; Rossi et al., 2019). Additional species, including *Ailanthus altissima* (Filippin & De Pra, 2011; Miladinović et al., 2019; Plavec et al., 2019), *Salix* spp. (Casati et al., 2017), *Spartium junceum* (Rizza et al., 2021) and *Artemisia vulgaris* (Jurga & Zwolińska, 2020) have also been identified as hosts of FD-related phytoplasmas, thereby increasing host plant diversity and contributing to the pathogen's persistence in the environment.

Hazelnut (*Corylus avellana*) is increasingly recognised as an important component of the FDp epidemiological system in some regions, linking forest and agricultural compartments (Casati et al., 2017). Initially considered a minor or incidental host, hazelnut was later identified as a natural reservoir of FD-related phytoplasmas, with infection rates reaching up to 35% in some surveyed populations (Casati et al., 2017). Molecular characterisation showed that hazelnut can harbour different FDp lineages (Casati et al., 2017). Studies in Slovenia confirmed the presence of 16SrV phytoplasmas in declining cultivated hazelnuts, with isolates genetically identical to those causing FDp based on molecular analysis of the FD9 marker, ribosomal protein operon sequences, and *map* gene (Mehle et al., 2019). Infected shrubs exhibited symptoms such as branch dieback and mortality, with notable susceptibility reported among local cultivars. Asymptomatic infections have also been documented, indicating that hazelnut can maintain infection without visible decline symptoms (Casati et al., 2017).

Hazelnut is also relevant through its association with polyphagous vectors. It has been identified as a suitable developmental host for *O. ishidae*, which can complete its life cycle on this plant and acquire phytoplasmas under natural conditions (Casati et al., 2017; Lessio et al., 2016). This suggests a specific epidemiological setting in which FDp is maintained in forest habitats and occasionally transferred to vineyards by mobile vectors. Importantly, the primary epidemic vector *S. titanus* has not been observed on hazelnut in the studied environments, indicating that hazelnut's role in FDp epidemiology is mediated primarily through alternative vectors rather than direct vineyard transmission cycles (Casati et al., 2017). The detection of 16SrV-C phytoplasmas in hazelnut also points to localised transmission cycles outside vineyards, although the extent to which these contribute to vineyard infections remains unresolved (Malembic-Maher et al., 2020). Determining whether these populations represent the same or distinct phytoplasma lineages would benefit from genome-level data.

1.3 Sequencing of Phytoplasma Genomes

Genome sequencing is particularly useful in phytoplasma research, primarily because they cannot be maintained in axenic culture and are inaccessible using conventional microbiological methods. Early studies relied on single-gene markers such as the 16S rRNA gene, but these are insufficient to distinguish closely related strains or capture biologically relevant diversity (Jardim et al., 2023). Whole-genome data allow closer examination of phytoplasma biology and evolution, revealing that their genomes are highly reduced and lack many essential metabolic pathways, including components of central carbon metabolism and ATP synthesis, reflecting a strong dependence on host-derived resources (Namba, 2019; Oshima et al., 2004, 2013). Such genome reduction is consistent with reductive evolution in obligate intracellular bacteria. Despite their small size, phytoplasma genomes remain highly plastic, mainly due to repetitive and mobile elements such as potential mobile units (PMUs), which drive genome rearrangements and adaptation (Bai et al., 2006; Ku et al., 2013; Wei et al., 2008). These regions often encode secreted effector proteins involved in host manipulation, playing a central role in pathogenicity and host–pathogen interactions (Huang et al., 2022; Musić et al., 2025; Tomkins et al., 2018). Comparative analyses of phytoplasma genomes have also identified strain-specific putative effector and secreted protein genes, many of them located in PMU-like regions (Musić et al., 2025). For this reason, genome data are useful for refining taxonomy and tracing epidemiological patterns.

Obtaining phytoplasma genome sequences nevertheless remains technically difficult. Because host-free cultured isolates are not available, DNA must be extracted directly from infected plants or insects. In such samples, phytoplasma DNA typically represents a very small fraction of the total nucleic acid content, often less than 1%, and is present in the background of the host genomic and organellar DNA (Rodrigues Jardim et al., 2022; Zhang et al., 2023). Low pathogen titre in phloem tissues and uneven distribution in woody hosts add to this problem (Prezelj et al., 2013; Siddique et al., 1998). Genome composition also affects sequencing and assembly, including extremely low GC content, usually between 20% and 30%, which can introduce bias in amplification and coverage during library preparation, potentially reducing sequencing efficiency (Wang et al., 2024). Although phytoplasma genomes are relatively small, reported genome sizes range from approximately 530 to 1350 kb (IRPCM, 2004). They are highly repetitive due to duplicated regions and PMUs, complicating assembly and frequently resulting in fragmented draft genomes (Zhang et al., 2023).

To overcome these limitations, several enrichment strategies have been developed to enrich phytoplasma DNA prior to sequencing. One common approach involves using alternative host plants, such as *Catharanthus roseus*, which can support higher phytoplasma titres and provide more suitable material for sequencing (Tran-Nguyen et al., 2008). Physical separation techniques, including pulsed-field gel electrophoresis (Bai et al., 2006; Kube et al., 2008; Oshima et al., 2004; Tran-Nguyen et al., 2008; J. Wang et al., 2018), have been used to separate large phytoplasma DNA molecules from host DNA based on size, while density-gradient centrifugation methods exploit the AT-rich nature of phytoplasma DNA to separate it from host DNA (Tran-Nguyen & Schneider, 2013). Non-toxic separation approaches, such as iodixanol gradient centrifugation, have also been used to extract phytoplasma DNA (Rodrigues Jardim et al., 2022). Other strategies include immunoprecipitation-based methods (Tan et al., 2021) and techniques based on methyl-CpG binding domains (Nijo et al., 2021). DNA amplification approaches, such as multiple displacement amplification and rolling circle amplification, can be employed to increase the amount of target DNA (Mitrović et al., 2013). Other methods, including subtractive

hybridisation and mirror orientation selection, have also been shown to improve the efficiency of phytoplasma DNA enrichment (Tran-Nguyen & Gibb, 2007).

Advances in sequencing technologies have facilitated the generation of phytoplasma genome data. Early genome projects relied on capillary sequencing, which required cloning DNA fragments into plasmid vectors and typically produced moderate coverage that had to be supplemented by techniques such as primer walking and long-range PCR to close assembly gaps (Andersen et al., 2013; Kube et al., 2008; Oshima et al., 2004; Tran-Nguyen et al., 2008). The introduction of high-throughput sequencing (HTS), also referred to as next-generation sequencing (NGS) platforms, such as Illumina, enabled the production of large volumes of short reads without cloning, reducing some sources of bias and facilitating the assembly of draft genomes (Cho et al., 2020; Orlovskis et al., 2017; J. Wang et al., 2018). However, short-read data alone are often insufficient to resolve repetitive regions, resulting in fragmented assemblies. More recently, long-read sequencing technologies, including Oxford Nanopore Technologies (ONT) and PacBio, have provided the ability to span repetitive elements and improve genome contiguity, albeit with higher raw error rates (De Maio et al., 2019). Hybrid approaches combining short- and long-read data are now widely adopted to balance accuracy and assembly completeness (De Maio et al., 2019).

Available phytoplasma genome sequences have increased in recent years, although high-quality complete genomes remain relatively limited compared to other bacterial groups (R. Wang et al., 2024). Early studies reported only a few complete genomes, whereas more recent datasets include a larger collection of both complete and draft assemblies (R. Wang et al., 2024). However, reference genomes are still lacking for many phytoplasma taxa, which constrains comparative analyses and complicates the identification of novel strains.

1.4 Scientific Problems and Research Aims

The purpose of this dissertation was to investigate the genetic diversity and epidemiological connections of 16SrV phytoplasmas in grapevine and hazelnut, and to determine whether hazelnut-associated isolates correspond to FDp populations in grapevine or represent distinct genetic lineages. This required gene-based epidemiological analyses and the development of enrichment and sequencing approaches suitable for genome-level comparison of phytoplasmas directly from naturally infected plant material. The dissertation is structured according to the three main aims.

1.4.1 Assessing gene-based similarity and epidemiological connections among 16SrV phytoplasmas in grapevine, hazelnut, and leafhoppers

Current knowledge shows that grapevine, hazelnut, and non-cultivated hosts can harbour closely related 16SrV phytoplasmas (Malembic-Maher et al., 2020; Mehle et al., 2019). It remains unresolved whether hazelnut-associated isolates belong to the same epidemiologically relevant population as FD-associated phytoplasmas in grapevine or represent genetically differentiated populations. The gap, therefore, concerns the limited resolution of epidemiological relationships among these host compartments, rather than the mere occurrence of infection. In Slovenia, this question is of direct practical relevance because 16SrV phytoplasma infections cause severe problems in vineyards as well as in commercial hazelnut orchards (Mehle et al., 2019; Mehle, Ravnikar, et al., 2011).

The research problem arising from this gap was to determine whether the same 16SrV *map* genotypes circulate among grapevine, hazelnut, and leafhoppers, and if hazelnut orchards harbour *map* genotypes associated with epidemic FD outbreaks in vineyards.

Addressing this problem required a measurable assessment of *map* genotype overlap across plant and insect hosts, assignment of detected *map* genotypes to established Map-FD clusters, and discrimination between epidemic FD-associated genotypes and non-epidemic ones. Only within such a framework could it be tested whether the analysed host and vector compartments indicate potential epidemiological linkage.

Accordingly, the first aim of the dissertation was to compare 16SrV phytoplasmas from grapevine, hazelnut, and leafhoppers using established gene markers, with primary emphasis on *map* genotyping as the standard approach in FDp molecular epidemiology. The analysis evaluated whether *O. ishidae* and *S. titanus* could carry identical or closely related phytoplasma *map* genotypes and determined the extent of genotype overlap between vectors and infected plant hosts. The resulting *map* genotype assignments were interpreted in light of their epidemiological relevance and assembled into a harmonised dataset directly comparable with FDp *map* genotyping data generated in other EU countries, thereby providing a standardised basis for strain typing and molecular surveillance across regions and host systems. *Map* genotype analyses were also performed on the first findings of 16SrV phytoplasma in other countries (the Czech Republic and the Netherlands), and their epidemiological importance was evaluated.

1.4.2 Developing and optimising enrichment strategies for phytoplasma genome sequencing from naturally infected plant material

The technical constraints concerning phytoplasma genome sequencing outlined in the Introduction have a direct consequence for this dissertation: genome-wide comparison of 16SrV phytoplasmas from naturally infected grapevine and hazelnut requires genome data that cannot be obtained reliably by direct extraction alone. In field-collected woody tissues, low phytoplasma titres and high host DNA background reduce read recovery and compromise assembly quality. At the same time, the limited availability of representative reference genomes restricts mapping-based approaches and reduces the resolution of comparative and population-level analyses.

Although several enrichment approaches have been proposed to reduce host DNA background and improve recovery of phytoplasma reads, their performance is difficult to compare across studies. Many protocols are time-consuming or technically demanding, and few have been evaluated with respect to both sequencing output and practical applicability to naturally infected woody plant material. For this dissertation, the unresolved methodological gap was the lack of a robust, reproducible enrichment workflow suitable for generating phytoplasma genome data from naturally infected grapevine material and for assessing its transferability to hazelnut tissues.

The research problem addressed in this part of the dissertation was therefore methodological: identifying an enrichment strategy that enables reliable recovery of phytoplasma DNA from naturally infected tissues. Accordingly, the second aim was to develop, optimise, and systematically compare enrichment strategies applied to DNA extracted from naturally infected grapevine, with additional evaluation in hazelnut samples. The evaluation was based on direct sequencing outputs, allowing objective assessment of protocol performance. This aim also included identifying a workflow that not only increases the recovery of phytoplasma reads and improves assembly quality but also meets practical requirements for routine use, including reduced processing time, technical simplicity, and reproducibility.

1.4.3 Comparing genome-wide relationships of 16SrV phytoplasmas across hosts using genome sequencing

The interpretation of phytoplasma population structure and transmission across host plants and insects is currently limited by the resolution of commonly used gene-based markers. Epidemiological studies report identical *map* genotypes across multiple host species, indicating possible exchange between cultivated and non-cultivated compartments (Casati et al., 2017; Malembic-Maher et al., 2020). Marker-based approaches, however, do not resolve whether this similarity extends across the genome or is limited to the loci used for typing.

Marker loci provide a practical framework for strain classification and tracking, but they reflect only a limited subset of genome-wide relationships. It remains unclear whether isolates sharing the same *map* genotype across host plants show genome-wide similarity or retain host-associated differences. Addressing this question requires genome-scale data generated directly from naturally infected material. Without reliable enrichment and sequencing approaches, such comparisons could not be performed with sufficient resolution. The methodological developments described in the preceding section provided the basis for addressing this biological problem.

The research problem addressed in this part of the dissertation was to determine whether patterns of genome-wide similarity among phytoplasma isolates correspond to *map* genotype classification or reflect host-associated differences when isolates from different hosts share the same *map* genotype. Accordingly, the third aim was to perform genome sequencing and comparative genomic analyses of FDp isolates from grapevine, hazelnut, and leafhoppers sharing the same *map* genotypes, to assess whether patterns of genome-wide similarity correspond to *map* genotype classification or whether host-associated differences can be detected among isolates sharing the same *map* genotype.

1.5 Research Hypotheses

In this doctoral thesis, we hypothesised that:

1. *Map* genotypes of 16SrV group phytoplasmas with epidemic potential in grapevine can also be detected in hazelnut orchards, indicating that the same genotypes occur in both host systems.
2. Hazelnut and non-cultivated plants (e.g., alder) are expected to harbour, in addition to 16SrV phytoplasma *map* genotypes associated with epidemic outbreaks in grapevine, further *map* genotypes that do not cluster with epidemic grapevine-associated strains.
3. *Map* genotypes found in hazelnut samples are expected to be detectable in *O. ishidae*, supporting its potential involvement in the circulation of 16SrV phytoplasmas in hazelnut orchards.
4. Direct enrichment of FDp DNA from field-collected grapevine samples is feasible and can generate sequencing data of sufficient quality to enable genome assembly without the need for phytoplasma propagation in test plants.
5. Genomes of 16SrV phytoplasma isolates representing the same *map* genotype and detected in both grapevine and hazelnut are expected to show high genome-wide similarity.
6. Genomic differences between isolates from different host plants are expected to be smaller than differences observed between isolates belonging to different *map* genotypes.

1.6 Publications Included and Candidate's Contributions

The first publication, "Epidemiology of Flavescence Dorée and Hazelnut Decline in Slovenia: Geographical Distribution and Genetic Diversity of the Associated 16SrV Phytoplasmas", was published in *Frontiers in Plant Science* (doi: 10.3389/fpls.2023.1217425). In this study, we analysed the distribution and genetic diversity of 16SrV phytoplasmas in grapevine, hazelnut, and insect vectors in Slovenia using *map* genotyping. The results showed the predominance of specific genotypes in grapevine and documented their occurrence in hazelnut and leafhoppers, indicating possible connections between host compartments. We also identified previously unreported *map* genotypes in grapevine and hazelnut, extending the known diversity of 16SrV phytoplasmas. These findings contribute to understanding phytoplasma circulation across different host plants and to the development of epidemiological frameworks for FDp. The PhD candidate is the first author of the publication. She contributed to study design, field sampling, laboratory analyses, *map* genotype analysis, data curation, visualisation, and manuscript writing.

The second and third publications, "First Report of Grapevine Flavescence Dorée Phytoplasma in the Czech Republic" (*Plant Disease*, doi: 10.1094/PDIS-02-23-0315-PDN) and "First Detection of a Phytoplasma From the 16SrV-C (Elm Yellows) Group in *Alnus glutinosa* in the Netherlands" (*New Disease Reports*, doi: 10.1002/ndr2.70056), document the detection and molecular characterisation of 16SrV phytoplasmas in new geographical contexts. Both studies used *map* genotyping and sequence-based analyses to confirm the presence and classification of phytoplasma isolates and to relate them to previously described *map* genotypes. These reports extend the geographic scope of molecular epidemiological data and support the use of genotype-based frameworks for comparative analyses across different regions. Together, they contribute to a broader European dataset of 16SrV phytoplasma diversity and distribution. The PhD candidate is a co-author on both publications. She contributed to sequence analysis, phylogenetic reconstruction, figure preparation, and critical review of the manuscript.

The fourth publication, "From Vineyard to Genome: Optimised Enrichment and Sequencing of Flavescence Dorée Phytoplasma from Grapevine Samples", was published in *Microbial Genomics* (doi: 10.1099/mgen.0.001514). This study addressed methodological limitations in phytoplasma genomics by developing and evaluating enrichment strategies for DNA extracted directly from naturally infected grapevine samples. Multiple protocols were compared, combining pre- and post-extraction steps to increase phytoplasma DNA recovery and improve genome assembly quality. The results demonstrate that an optimised workflow can increase the proportion of phytoplasma-derived reads and enable high-quality draft genome reconstruction without propagation in test plants. This approach provides a practical framework for genome sequencing of phytoplasmas from field material and supports further comparative genomic analyses. The PhD candidate is the first author of the publication. She contributed to study design, laboratory work, bioinformatic analyses, data curation, visualisation and writing the manuscript.

The fifth contribution included in the thesis, "Genome-Wide Comparisons Confirm Locally Shared Flavescence Dorée-Related Phytoplasma Populations Across Grapevine, Hazelnut, and Leafhoppers", is a manuscript that was submitted to *Frontiers in Microbiology* (May 2026). This study applies a comparative genomics framework to phytoplasma isolates from different host plants and insects sharing the same *map* genotypes. Genome-wide analyses based on core gene phylogeny and genome-wide similarity metrics were used to assess relationships among isolates and to evaluate how marker-based classification relates to genome-scale patterns. The results indicate that

genome-wide similarity patterns are consistent with *map*-genotype classification for the analysed genotypes, whereas no consistent grouping by host species was observed. These findings contribute to the interpretation of phytoplasma population structure and support the integration of genome-scale data into current epidemiological frameworks. The PhD candidate is the first author of the manuscript. She contributed to study design, laboratory work, bioinformatic analyses, data curation, visualisation, and manuscript preparation.

Chapter 2

Scientific Publications

2.1 Epidemiology of Flavescence Dorée and Hazelnut Decline in Slovenia: Geographical Distribution and Genetic Diversity of the Associated 16SrV Phytoplasmas

Authors: **Zala Kogej Zwitter**, Gabrijel Seljak, Tjaša Jakomin, Jakob Brodarič, Ana Vučurović, Sandra Pedemay, Pascal Salar, Sylvie Malembic-Maher, Xavier Foissac, Nataša Mehle

Published in: *Frontiers in Plant Science*, vol 14, 2023
doi: 10.3389/fpls.2023.1217425

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 04 July 2023
DOI 10.3389/fpls.2023.1217425

Check for updates

OPEN ACCESS

EDITED BY
Gianfranco Romanazzi,
Marche Polytechnic University, Italy

REVIEWED BY
Fabio Quaglino,
University of Milan, Italy
Michael Maixner,
Julius Kühn-Institut, Germany

*CORRESPONDENCE
Zala Kogej Zwitter
zala.kogej.zwitter@nib.si

RECEIVED 05 May 2023
ACCEPTED 13 June 2023
PUBLISHED 04 July 2023

CITATION
Kogej Zwitter Z, Seljak G, Jakomin T, Brodarič J, Vučurović A, Pedemay S, Salar P, Malembic-Maher S, Foissac X and Mehle N (2023) Epidemiology of flavescence dorée and hazelnut decline in Slovenia: geographical distribution and genetic diversity of the associated 16SrV phytoplasmas. *Front. Plant Sci.* 14:1217425. doi: 10.3389/fpls.2023.1217425

COPYRIGHT
© 2023 Kogej Zwitter, Seljak, Jakomin, Brodarič, Vučurović, Pedemay, Salar, Malembic-Maher, Foissac and Mehle. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Epidemiology of flavescence dorée and hazelnut decline in Slovenia: geographical distribution and genetic diversity of the associated 16SrV phytoplasmas

Zala Kogej Zwitter[1,2*], Gabrijel Seljak[1], Tjaša Jakomin[1], Jakob Brodarič[1], Ana Vučurović[1], Sandra Pedemay[3], Pascal Salar[3], Sylvie Malembic-Maher[3], Xavier Foissac[3] and Nataša Mehle[1,4]

[1]National Institute of Biology, Department of Biotechnology and Systems Biology, Ljubljana, Slovenia, [2]Sensor Technologies, Jožef Stefan International Postgraduate School, Ljubljana, Slovenia, [3]National Research Institute for Agriculture, Food and Environment, Fruit Biology and Pathology, Bordeaux, France, [4]School for Viticulture and Enology, University of Nova Gorica, Nova Gorica, Slovenia

Flavescence dorée (FD) phytoplasma from 16SrV-C and -D subgroups cause severe damage to grapevines throughout Europe. This phytoplasma is transmitted from grapevine to grapevine by the sap-sucking leafhopper *Scaphoideus titanus*. European black alder and clematis serve as perennial plant reservoirs for 16SrV-C phytoplasma strains, and their host range has recently been extended to hazelnuts. In Slovenia, hazelnut orchards are declining due to 16SrV phytoplasma infections, where large populations of the non-autochthonous leafhopper *Orientus ishidae* have been observed. To better characterise the phytoplasma-induced decline of hazelnut and possible transmission fluxes between these orchards and grapevine, genetic diversity of 16SrV phytoplasmas in grapevine, hazelnut and leafhoppers was monitored from 2017 to 2022. The nucleotide sequence analysis was based on the *map* gene. The most prevalent *map* genotype in grapevine in all wine-growing regions of Slovenia was M54, which accounted for 84% of the 176 grapevines tested. Besides M54, other epidemic genotypes with lower frequency were M38 (6%), M51 (3%), M50 (2%) and M122 (1%). M38, M50 and M122 were also detected in infected cultivated hazelnuts and in specimens of *O. ishidae* leafhopper caught in declining hazelnut orchards. It suggests that this polyphagous vector could be responsible for phytoplasma infection in hazelnut orchards and possibly for some phytoplasma exchanges between hazelnuts and grapevine. We hereby describe new genotypes: M158 in grapevine as well as four never reported genotypes M159 to M162 in hazelnut. Of these four genotypes in hazelnut, one (M160) was also detected in *O. ishidae*. Analysis of additional genes of the new genotypes allowed us to assign them to the VmpA-III cluster, which corresponds to the 16SrV-C strains previously shown to be compatible with *S. titanus* transmission.

KEYWORDS

epidemiology, plant disease, vectotype, *Vitis vinifera*, *Corylus avellana*

1 Introduction

Phytoplasmas, the wall-less plant pathogenic bacteria of the Mollicutes class, have a complex cycle (Lim and Sears, 1989; Marcone, 2014). They survive and multiply intracellularly in two different environments: the phloem sieve tube elements of the host plants and the body of the insect vector. Phytoplasmas are transmitted from plant to plant by hemipteran phloem feeders as leafhoppers, planthoppers, and psyllids (Weintraub and Beanland, 2006). They are associated with many plant diseases worldwide and can cause economic damage (Bertaccini et al., 2014). One of the quarantine diseases in the EU is caused by grapevine Flavescence dorée (FD) phytoplasma from taxonomic subgroups 16SrV-C and 16SrV-D (Davis and Dally, 2001). FD first occurred in southwestern France in the 1950s (Caudwell, 1957), and the disease continually spread to vineyards of southern European countries in the following decades (Tramontini et al., 2020). It was detected in Slovenia for the first time in 2005 (Mehle et al., 2011). The rapid epidemic spread of FD phytoplasma was caused by the vector *Scaphoideus titanus*, which probably arrived in Europe from North America in a single large introduction event (Papura et al., 2012). *S. titanus* is an efficient vector for propagating the FD phytoplasma in vineyards due to its specialisation for *Vitis* spp. and its ability to reach high population abundance (Chuche and Thiéry, 2014). Despite mandatory control measures, which consist of uprooting and destroying infected plants and spraying insecticides against *S. titanus*, FD phytoplasma continues to spread in Europe (Tramontini et al., 2020).

FD phytoplasma was not introduced in Europe simultaneously with *S. titanus*. Malembic-Maher et al. (2020) demonstrated that European alders (*Alnus* sp.) are the original reservoir for 16SrV-C phytoplasmas. The discoveries were made based on the sequence of methionine aminopeptidase (*map*) genetic locus, which shows higher variability in 16SrV phytoplasmas than the 16S rRNA gene. More than 150 different *map* genotypes were detected in wild alders and secondary host plants from the environment of vineyards all around Europe, but only 12 were associated with FD outbreaks on grapevine (*Vitis vinifera*) (Radonjić et al., 2013; Atanasova et al., 2014; Krstić et al., 2018; Plavec et al., 2019; Malembic-Maher et al., 2020; Krstić et al., 2022). The secondary hosts of 16SrV-C phytoplasmas include clematis (*Clematis vitalba)* (Angelini et al., 2004; Filippin et al., 2009; Casati et al., 2017; Plavec et al., 2019; Rossi et al., 2019), tree of heaven (*Ailanthus altissima)* (Filippin and De Pra, 2011; Miladinović et al., 2019; Plavec et al., 2019), Spanish broom (*Spartium junceum*) (Rizza et al., 2021) and willow (*Salix* spp.) (Casati et al., 2017). FD phytoplasmas are divided into three *map* genetic clusters: Map-FD1, Map-FD2, and Map-FD3 (Arnaud et al., 2007). Genotypes that do not belong to one of the three Map-FD clusters are not transmitted by *S. titanus* and do not provoke FD outbreaks (Malembic-Maher et al., 2020). Whether a phytoplasma variant is transmissible by *S. titanus* can be deduced from the *VmpA* sequence, a gene coding for a surface protein that functions as an adhesin and binds to insect vector cells (Arricau-Bouvery et al., 2021). Sequences of *VmpA* form three clusters, out of which clusters II and III are associated with *S. titanus* transmission and thus potential outbreaks of FD (Malembic-Maher et al., 2020). Up to now epidemic isolates from clusters Map-FD1, Map-FD2, and Map-FD3 were never found clustering in VmpA cluster I (Malembic-Maher et al., 2020).

In Slovenia, phytoplasmas of 16SrV-C subgroup have also been detected in decayed cultivated hazelnuts (*Corylus avellana*) (Mehle et al., 2019). Hazelnut is one of the ten most cultivated fruits in Slovenia and was intensively cultivated on an area of 190 ha in 2021 (Food and Agriculture Organisation of the United Nations, 2022). In some Slovenian hazelnut orchards, symptoms of phytoplasma infection were observed as yellowing and curling of leaves, dead branches leading to withering of entire bushes. These symptoms were linked to phytoplasmas of 16SrXII ('*Candidatus* P. fragariae'), 16SrIX and 16SrV groups (Mehle et al., 2019). Molecular analysis of the FD9 marker and ribosomal protein operon sequences on 16SrV phytoplasma infected samples revealed the presence of FD-D and FD70 isolates, and sequencing of *map* gene classified them as Map-FD1 and Map-FD2 (Mehle et al., 2019). In addition, 16SrV phytoplasmas were detected on asymptomatic wild hazelnuts in the vicinity of FD infected vineyards in Switzerland (Casati et al., 2017). In this study, the presence of all three Map-FD clusters was evidenced in hazelnut shrubs as well as in the Asian leafhopper, *Orientus ishidae*, captured in the vicinity of vineyard (Casati et al., 2017).

In Slovenia, the 16SrV group phytoplasmas were detected for the first time in the Asian leafhopper *O. ishidae* in 2009 (Mehle et al., 2010). It is a polyphagous alien species that feeds mainly on broadleaf trees and shrubs (Lessio et al., 2016). Spatial distribution studies have shown that it is more abundant in host plants outside vineyards than in vineyards themselves (Lessio et al., 2016; Casati et al., 2017). Nevertheless, it can complete its life cycle on the grapevine (Desqué et al., 2019; Lessio et al., 2019) and it can transmit 16SrV phytoplasmas to grapevine after experimentally forced acquisition (Lessio et al., 2016). FD *map* genotypes transmitted by *O. ishidae* were demonstrated to be compatible with *S. titanus* transmission and therefore can potentially cause epidemics in grapevine (Malembic-Maher et al., 2020).

The objective of this study was to look at 16SrV phytoplasmas infected hazelnut orchards in Slovenia in order to determine which *map* genotypes might be the cause of their decline and whether these genotypes can be found also in potential vector *O. ishidae*. We compared the results with the structure of *map* genotypes in Slovenian FD infected vineyards. Newly detected genotypes were further genetically characterized to determine their possible origin and their epidemic potential on hazelnut and grapevine.

2 Materials and methods

2.1 Plant and insect material

Sampling plan is summarized in Figure 1 and Table 1. Names and geographical origins of the genotyped samples are listed in Supplementary Table 3.

The grapevine samples from symptomatic grapevines were collected as part of the official monitoring of grapevine yellows disease under the direction of Administration of the Republic of Slovenia for Food Safety, Veterinary and Plant Protection. This

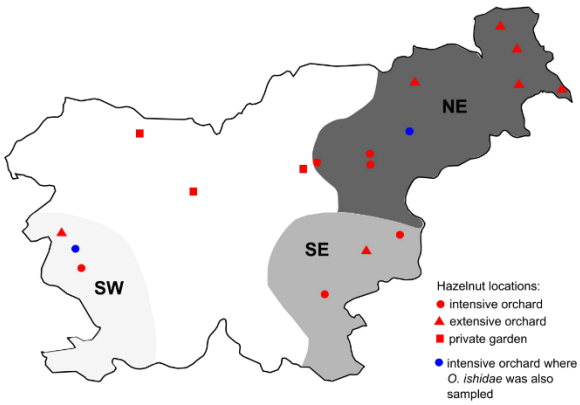


FIGURE 1
Map of grapevine, hazelnut, and leafhopper sampling regions/locations; Slovenian grapevine regions are shown in grey (NE in dark, SE in medium, and SW in light grey); locations of sampled hazelnuts are marked with dots (see legend); *S. titanus* samples were taken in NE and SE region.

TABLE 1 16SrV phytoplasma positive samples of grapevine, hazelnut, and leafhoppers from different regions of Slovenia between 2017 and 2022.

	Total number of samples*/number of 16SrV positive samples/number of samples with determined *map* sequence			
Region/host	Grapevine	Hazelnut	*O. ishidae*	*S. titanus*
NE	762 / 341 / 133	72** / 40** / 30**	48 / 8 / 6	15 / 7 / 5
NW	nt	2 / 1 / 0	nt	nt
SE	598 / 104 / 13	3 / 1 / 0	nt	6 / 0 / 0
SW	455 / 86 / 30	20 / 5 / 0	17 / 5 / 3	nt
total	1815 / 531 / 176	97 / 47 / 30	65 / 13 / 9	21 / 7 / 5

NE, northeast; NW, northwest; SE, southeast; SW, southwest; nt, not tested.
* Some samples of grapevine, *O. ishidae* and *S. titanus* were analysed as pools of up to five plants or up to seven leafhoppers.
** For some of these samples, roots and shoots were analysed separately.

study includes grapevine samples collected between 2017 and 2021 in 107 different sites in three different wine-growing regions of Slovenia: the northeast (NE), the southeast (SE) and the southwest (SW). Each sample consisted of leaves from at least three different parts of one or up to five plants from the same vineyard.

Symptomatic hazelnut samples were collected as part of a survey for phytoplasmas in Slovenian hazelnut orchards (seven extensive and eight intensive) and in three private gardens. They were collected from 2017 to 2021, from the end of June to the end of October. The sample consisted of leaves from three different parts of a decaying shrub. In most cases, samples were also taken from roots - excavated from three different parts of the shrub rooting system and treated as a separate subsample.

Leaf samples of grapevine and hazelnut were prepared in the same manner by excising 1 g of leaf mid-vein tissue. Hazelnut roots were washed extensively to remove all soil particles, and then 1 g of phloem tissue was excised. All samples were stored at -20°C until further analysis.

Leafhoppers were captured by sweep net within vineyards or hazelnut orchards during the period when plant symptoms were expressed: *S. titanus* between August and September in years 2021 and 2022 and *O. ishidae* in July 2019 and 2021. *S. titanus* were captured in the frame of the national monitoring in five different vineyards in NE region and stored in 70% ethanol until analysis. *O. ishidae* was captured in two hazelnut orchards infected with 16SrV phytoplasmas (Črešnjevec and Šempas) and immediately stored at -80°C or in 70% ethanol.

2.2 DNA extraction and phytoplasma detection

After homogenization with Plant DNA lysis buffer (QuickPick Plant DNA Kit; Bio-Nobile, Finland) or ELISA buffer (Kogovšek et al., 2015) of 1 g of leaf veins or roots using a FastPrep24™ instrument (MP Biomedicals, Illkirch-Graenstaden, France), total DNA was extracted using the QuickPick™ SML Plant DNA Kit (Bio-Nobile, Finland) and a KingFisher extraction device (ThermoScientific, USA) (Mehle et al., 2013a).

Leafhoppers were homogenised individually or in groups of two to seven specimens. In total 77 *S. titanus* specimen (analysed as 21 samples) and 105 *O. ishidae* (analysed as 65 samples) were tested. In case of storage in ethanol, they were rinsed in four consecutive water baths before homogenization. The leafhoppers were homogenised in a 1.5 ml tube with liquid nitrogen and a sterile plastic pestle. Total DNA was isolated following the same protocol as for plant material, with some adjustments: if <3 specimen was in sample, volume of lysis buffer used was 400 µl and DNA was eluted in 100 µl or if ≥3 specimen was in sample, volume of lysis buffer was 600 µl and DNA was eluted in 200 µl.

Tenfold diluted total plant or insect DNA (except undiluted insect DNA in case of one specimen per sample) were tested using a specific real-time PCR for 16SrV phytoplasma group (Hren et al., 2007). The setup of the real-time PCR was as described in (Mehle et al., 2013b). All samples were analysed also with universal phytoplasma real-time PCR assay (Christensen et al., 2004) and grapevine samples additionally with specific real-time PCR assay for Bois noir phytoplasma (Hren et al., 2007). In addition to the positive and negative amplification control in each real-time PCR run, the 18S rRNA assay (Applied Biosystems, Massachusetts, USA) was used as a control of DNA extraction.

2.3 Amplification and sequencing of genetic markers

Samples which tested positive with 16SrV phytoplasma-specific real-time PCR assay and whose Cq value was less than 32 were subjected to further amplification of the *secY-map* locus (*map* gene) by nested PCR (nPCR) with FD9f5/MAPr1 and FD9f6/MAPr2 primer pairs (Arnaud et al., 2007). First and second PCR reactions were performed in 50 µl with 1x High fidelity PCR Buffer (Invitrogen, Massachusetts, USA), 2mM $MgSO_4$, 0.2 mM dNTP, 0.2 µM of each primer and 1U of Platinum Taq DNA Polymerase High Fidelity (Invitrogen, Massachusetts, USA). 2 µl of 10-fold diluted plant or insect DNA (except in case of one specimen per sample where it was undiluted) was used as a template in the first PCR and 2 µl of 100-fold diluted PCR product was used for the second PCR. The conditions were the same for both PCR reactions: initial denaturation at 94°C for 2 min; 35 cycles consisting of denaturation at 94°C for 15 s, annealing at 52°C for 30 s and extension at 68°C for 1 min. Each run included negative template amplification control (water). Amplification of nPCR was checked by 1% agarose gel electrophoresis with ethidium bromide (EtBr) staining and visualisation under UV illumination. Nested PCR products were purified (outsourced to Macrogen Europe commercial service or with MinElute PCR Purification Kit (Qiagen, Hilden, Germany) and forward and reverse Sanger sequencing was performed by Macrogen Europe (Amsterdam, the Netherlands) or Eurofins GATC (Ebersberg, Germany).

Samples, in which new *map* genotypes were discovered, were further characterised with nPCR and Sanger sequencing of five different genes: *tuf*, *rplV*, *rplF*, *dnaK* and the first repeat of *vmpA* (*vmpA-R1*, information on tests in Supplementary Table 1). Primer sets for *tuf*, *rplF* and *dnaK* were changed compared to previous studies to improve sequence quality (Supplementary Table 1). PCR reactions were performed with *Taq* DNA Polymerase with Standard *Taq* Buffer (NEB, Massachusetts, USA), 3 mM final concentration of $MgCl_2$ and 5% DMSO in the first PCR reaction. Amplification of nPCR was checked by 1% agarose gel electrophoresis with EtBr staining and visualisation under UV illumination. Nested PCR products were purified and sequenced with Macrogen Europe (Amsterdam, the Netherlands).

2.4 Sequence analysis

Forward and reverse (or both forward for *VmpA-R1)* raw chromatograms of all genes were assembled, checked for quality and discrepancies, and trimmed in CLC Genomic Workbench (version 21.0.1, Qiagen) or with the Staden package (Bonfield et al., 1995). Contigs were compared to the local databases of various genes using BLAST algorithms (http://www.ncbi.nlm.nih.gov/blast). Location map and unrooted median-joining networks of *map* genotypes were generated using PopArt 1.7 software (Leigh and Bryant, 2015) with median-joining method and parameter e = 0 on alignment made with ClustalW in MEGA7 (Kumar et al., 2016). Sequence alignments of *VmpA-R1*, *dnaK* and multilocus sequence analysis (MLSA) of concatenated *tuf* (925 bp) – *rplV* (795 bp) – *rplF* (807 bp) – *map* (674 bp) were also performed using ClustalW in MEGA7. The phylogenetic trees were constructed using Maximum Parsimony algorithm in MEGA7. The statistical support of the branches was evaluated using the bootstrap method based on 500 replicates. All the accession numbers of the different genes from different isolates are listed in Supplementary Table 2 and in (Malembic-Maher et al., 2011) for the MLSA.

3 Results

3.1 Grapevine and hazelnut infection with 16SrV phytoplasmas

A total of 1815 symptomatic grapevine samples were sampled between 2017 and 2021 in three wine-growing regions of Slovenia (Table 1). The proportion of 16SrV phytoplasma infected grapevine samples increased over time, and was the highest in 2021 (49%, 169 out of 347 samples analysed). In 2021, there was a large outbreak in NE region and it was more extensively sampled. The five-year overall percentage of 16SrV positive samples was 29%. In each of the three wine-growing regions, 4-6% of 16SrV-positive samples were also infected with phytoplasma Bois noir (16SrXII, '*Ca.* Phytoplasma solani'). During this five-year period, 50% of all sampled grapevines were infected with Bois noir phytoplasma and 21% were phytoplasma negative.

From 2017 until 2021, a total of 97 hazelnut shrubs from 18 different locations were analysed. The 16SrV phytoplasmas were detected in 47 shrubs (Table 1). '*Ca.* P. fragariae' was also detected in some of the orchards and in one sample from a private garden in 2017 (Mehle et al., 2019), but no phytoplasma was detected in 41% of the sampled shrubs (tested with the universal phytoplasma assay, data not shown). 16SrV positive shrubs were found in intensive orchards (6 of 8 sampled), extensive orchards (3 of 7 sampled), and in one out of three private gardens. In one orchard monitored throughout the five-year period (Črešnjevec), the proportion of 16SrV-positive tests

of decaying shrubs increased from 30% in 2017 (3 out of 10 shrubs) to 90% in 2021 (9 out of 10 shrubs). All 16SrV phytoplasma-positive shrubs showed dieback symptoms - most commonly wilting of individual branches, yellowing of leaves, or withering of the entire shrub. The real-time PCR analyses of shoot and root subsamples, which were taken from the same shrub, showed that the phytoplasma concentration was higher in the roots (data not shown).

3.2 Leafhoppers infection with 16SrV phytoplasmas

FD phytoplasmas occurrence was analysed in 21 cumulative *S. titanus* samples (up to seven specimens per sample) collected between August and September 2021 and 2022. Using the real-time PCR assay of Hren et al. (2007) the presence of 16SrV phytoplasmas was detected in two samples of *S. titanus* in 2021 and five in 2022 - all caught in FD-infected vineyards in NE Slovenia.

In July 2019 and 2021, two hazelnut orchards infected with 16SrV phytoplasmas were subjected to insect vector searches - one in NE (Črešnjevec) and one in SW Slovenia (Šempas). *S. titanus* was not found in any of these hazelnut orchards. However, in 2021, the populations of *O. ishidae* in Črešnjevec and Šempas were estimated to be abundant, as 44 and 56 specimens were caught in one hour, respectively. In the NE orchard, 16SrV phytoplasmas were detected in two samples from 2019 (pools of four and two specimens) and in six samples composed of only one specimen in 2021. *O. ishidae* from the SW orchard were pooled in 17 pools of two to five specimens, and five pools tested positive for 16SrV phytoplasmas (Table 1).

3.3 *Map* sequencing of infected samples

The *map* gene of 16SrV phytoplasmas was sequenced in 176 grapevines and 30 hazelnuts, 11 of which were tested as separate subsamples of shoots (leaf veins) and roots (Table 1, details in Supplementary Table 3). Other samples had Cq value of 16SrV specific real-time PCR higher than 32, making the concentration of phytoplasma too low to submit for nPCR and Sanger sequencing. Out of 20 positive samples of leafhoppers, 14 samples were subjected to *map* sequencing: five were *S. titanus* (one sample of seven imagos and others of three imagos) and nine were *O. ishidae* (four single imago samples and five cumulative). Their Cq values ranged from 18 to 32 for the specific 16SrV phytoplasma assay. In total, 231 samples were submitted to *map* sequencing (Supplementary Table 3) with 2-fold coverage (except for nine samples, of which only one sequence was of sufficient quality). Single infections of a particular *map* genotype were identified in all samples.

Eleven different *map* genotypes were characterized, of which five, M158 to M162, are new (Table 2). In a phylogenetic analysis,

TABLE 2 Summary of discovered *map* genotypes from 2017 to 2022 in different hosts.

Map-FD cluster	Genotype	GenBank acc no.	Reference	Host	Number of samples
FD1	M50	AM384887	(Malembic-Maher et al., 2020)	grapevine	4
				hazelnut	8*
				O. ishidae	1
	M158	OX417123**	this study	grapevine	7
	M159	OX417124**	this study	hazelnut	1
	M160	OX417125**	this study	hazelnut	1
				O. ishidae	1
	M161	OX417126**	this study	hazelnut	1*
FD2	M54	AM384886	(Malembic-Maher et al., 2020)	grapevine	147
				S. titanus	5
	M38	LN850369	(Malembic-Maher et al., 2020)	grapevine	11
				hazelnut	12*
				O. ishidae	5
	M122	LT222007	(Plavec et al., 2019)	grapevine	2
				hazelnut	2*
				O. ishidae	2
	M162	OX417127**	this study	hazelnut	3*
FD3	M51	FN811141	(Malembic-Maher et al., 2020)	grapevine	5
grapevine non-epidemic isolate	M48	AM384893	(Malembic-Maher et al., 2020)	hazelnut	2*

* In 11 samples, roots and shoots were analysed separately (counted as one sample here; results of separate analysis are shown in Supplementary Table 3).
** sequences were deposited to European Nucleotide Archive (ENA) under https://www.ebi.ac.uk/ena/browser/view/OX417123-OX417147.

they were compared with reference sequences derived from epidemic FD strains, non-epidemic 16SrV-C isolates on grapevine (Palatinate Grapevine Yellows), and some 16SrV-C isolates from other plants described in previous studies (Malembic-Maher et al., 2020; Krstić et al., 2022) (Figure 2, Table 2). Four new genotypes (M158, M159, M160 and M161) group in the Map-FD1 cluster, which is the most variable in Slovenia. They are 99.85% identical to M50 with one nucleotide difference in the amplicon. A new *map* genotype found in three hazelnut shrubs (M162) belongs to Map-FD2 and differs from M38 by one nucleotide. The single nucleotide polymorphisms (SNP) in all newly discovered isolates represent transitions. The Map-FD3 cluster was the most homogeneous, as all *map* sequences were 100% identical to M51 (Figure 2; Table 2).

The *map* genotypes were determined on grapevine samples from all wine-growing regions of Slovenia: NE, SE, and SW. Six different genotypes were detected from all three Map-FD clusters (Table 2; Figure 2). The most prevalent genotype was M54 from Map-FD2, which was present in 84% of the sequenced samples. This genotype was also detected in all five samples of *S. titanus*. The other genotypes were much less abundant: M50 (2%) and M158 (4%) from Map-FD1, M38 (6%) and M122 (1%) from Map-FD2, and M51 (3%) from Map-FD3 (Table 2; Supplementary Figure 1). Genotypes M54, M51, and M158 were specific to grapevine. Interestingly, all six different genotypes were detected in NE, four in SE and only two in SW (Figure 3). The Map-FD2 genotypes M54 and M38 occurred in all wine-growing regions. They were the only ones present in SW with 28 samples of M54 in 16 different sites and two samples of M38 found at the same site in 2019 (Gradišče and Prvačino, Supplementary Table 3). Genotype M122 was detected only in NE, in two nearby sites (Stara Cesta and Žerovinci, Supplementary Table 3). The newly described M158 was also detected only in NE in seven samples from 2017 to 2019 at five nearby sites. Genotypes M50 and M51 were found in NE and SE. In NE, M50 was present in three different sites: at two it was the only genotype, but at one two other genotypes were detected in different years (M158 in 2017, M50 in 2019 and M54 in 2021). Interestingly, in every site harbouring more than one genotype, M54 was always detected except in Tolsti vrh, where it was M38 and M50 only. Up to three genotypes could be detected in a same site.

The hazelnuts infected with 16SrV phytoplasmas further characterized on *map*, all came from four orchards in NE Slovenia (Figure 3; Supplementary Table 3). Eight different genotypes were detected (Table 2; Figure 3). They grouped into two clusters - Map-FD1 and FD2, but none of the isolates belonged to Map-FD3 cluster. Six hazelnut shrubs (20%) were infected with *map* genotypes (M159-M162), which have not been described previously (Table 2; Supplementary Figure 2). These new genotypes were all found in an intensive orchard (Črešnjevec) in different years: M159 and M160 in root samples from 2020 (isolates D1448-20 and D1455-20), M161 in root and shoot samples of the same shrub from 2021 (isolate SB21-5-P), and M162 in root sample from 2020 and two shrubs (root and shoot samples) from 2021

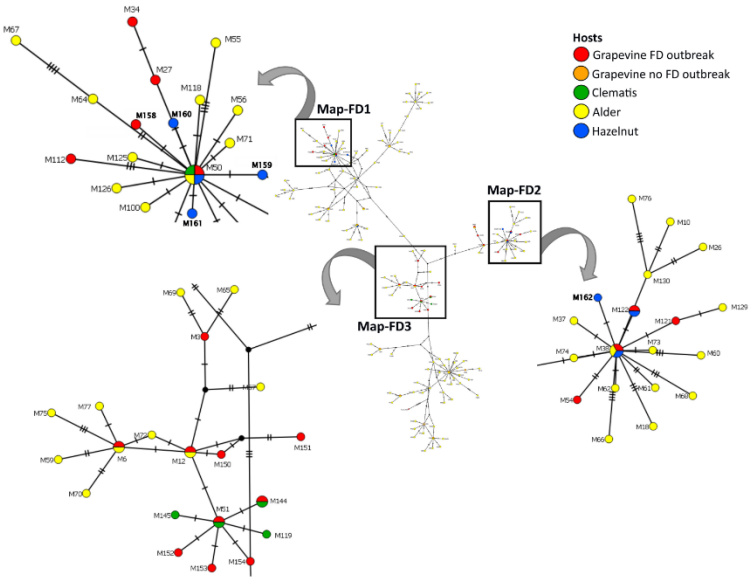


FIGURE 2
Median joining network of *map* genotypes (674 long amplicons); circles represent genotypes (paired genotypes are connected), black circles represent missing haplotypes and hatch mark represent one SNP; the legend shows color code of different hosts (in case of grapevine separated on genotypes causing FD and the ones not-causing FD outbreaks); squares mark Map-FD clusters that are enlarged; in the enlarged area new genotypes found in Slovenia are written in bold.

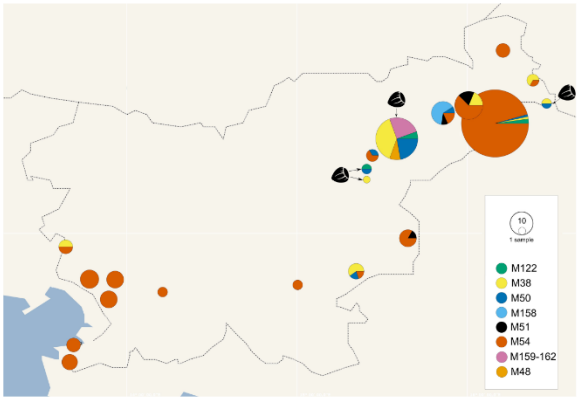


FIGURE 3
Geographical distribution of *map* genotypes in grapevines and hazelnuts (hazelnut samples marked with hazelnut image and arrow, other pie charts represent grapevine) collected in Slovenia from 2017 to 2021; the size of the circle corresponds to the number of samples and the colours to the *map* genotypes (see legend); where vineyards were close together, samples were combined in the same pie chart.

(isolate SB21-10-K). In this orchard, all the eight genotypes described in hazelnut were present in the years from 2017 to 2021. Of 30 hazelnut shrubs analysed, 22 were infected with the same genotype as grapevine: M38 (40%), M50 (27%), and M122 (7%). Interestingly, the most prevalent genotype in grapevine M54 was not detected in hazelnuts. The analysed hazelnut samples were located in the same NE geographic area as grapevines infected with the same genotypes. In addition, hazelnut was found to harbour genotype M48 (7%), which is not part of the Map-FD clusters and has been previously described in Palatinate Grapevine Yellows cases (non-epidemic in grapevine) and in alders (Malembic-Maher et al., 2020). Eleven hazelnut shrubs that were analysed in subsamples had the same genotype in shoots and roots (Supplementary Table 3).

O. ishidae collected on hazelnuts harboured three genotypes of 16SrV phytoplasmas that are present in both hazelnuts and grapevines: M38, M50, and M122 (Table 2). In the SW orchard (Šempas), only one genotype was detected in *O. ishidae* (M38), while in the NE orchard (Črešnjevec), four different genotypes were detected (Supplementary Table 3). In addition to the three genotypes listed, the new genotype M160 was also detected in one pooled sample. Interestingly, one year later, in 2020, it was also detected in hazelnuts in the same orchard.

3.4 Finer genetic characterisation of new *map* genotypes

The five new genotypes M158 to M162 were characterized more precisely by sequencing genetic markers *vmpA-R1, dnaK, tuf, rplF* and *rplV*. The results are shown in Figure 4. Based on the *dnaK* gene amplicon (475 bp), D1385-19 (Map-FD1, M158) detected in grapevine has the same sequence as the DnaK2 reference isolate collected in grapevine in northeast Italy (reference strain FD-C, MH547711) (Figure 4A). In the MLSA of concatenated *tuf* (925 bp), *rplV* (795 bp), *rplF* (807 bp), and *map* (674 bp) gene loci (Figure 4B), it is also clustered with Map-FD1 isolates collected in grapevine, clematis, and alder in northeast Italy and southwest France (with reference strain FD70 from grapevine). Analysis of the first *VmpA* repeat (VmpA-R1) showed that it belongs to the VmpA-III cluster that groups FD isolates transmitted by *S. titanus* (Figure 4C).

The four new *map* genotypes discovered in hazelnuts (M159-M162) differ from the reference sequences of *dnaK* (Figure 4A). M162 (isolate SB21-10-K) is most similar to *dnaK* from *Spartium junceum* (acc. no. MT629775, Sj1MS) with two nucleotide differences. M159 (isolate D1448-20) and M161 (isolate SB21-5-P) have one nucleotide difference from DnaK1 at the same nucleotide position towards the end of sequence and M160 (isolate D1455-20) one nucleotide difference from DnaK2 in the middle of the sequence. In the MLSA analysis, the isolates M159-161 group in a cluster with the greatest host diversity, which includes isolates from Map-FD1 along with the grapevine isolate D1385-19. Isolate SB21-10-K (genotype M162) is genetically related to isolates from Map-FD2 cluster (reference isolate FD92 from grapevine) and more closely related to an isolate collected in alder in southwest France. Interestingly, all hazelnut isolates also belong to the VmpA-III cluster gathering isolates compatible with *S. titanus* transmission (Figure 4C).

4 Discussion

FD disease has been present in Slovenia for more than a decade, and despite control measures, new outbreaks continue to occur. It causes damage and yield losses in all wine-growing regions of Slovenia. However, grapevine is not the only host of concern for

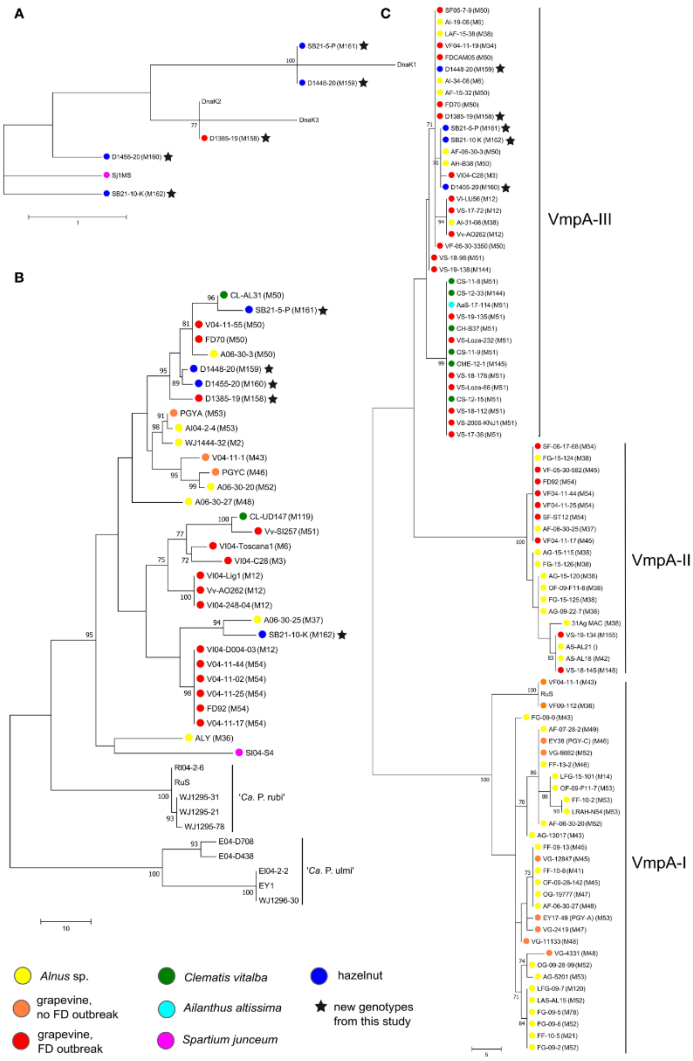


FIGURE 4
Unrooted phylogenetic trees constructed by maximum parsimony in MEGA7 software based on **(A)** *dnaK* (492 nt), **(B)** MLSA of concatenated *tuf* (925 bp) – *rplV* (795 bp) – *rplF* (807 bp) – *map* (674 bp) and **(C)** one of the trees constructed based on *VmpA-R1* (312 nt); sequences were aligned by ClustalW and number on branches are bootstrap values obtained from 500 replicates and indicated when ≥70%; the bar represents number of nucleotide changes; color circles represent different hosts (see legend); accession numbers are listed in Supplementary Table 2 and in (Malembic-Maher et al., 2011).

16SrV phytoplasmas as they were detected in several hazelnut orchards in Slovenia, where they caused severe damage to the production (Mehle et al., 2019). Our investigation of different *map* genotypes of 16SrV phytoplasmas present in grapevine, hazelnut, and the leafhopper vectors *O. ishidae* and *S. titanus* over a period of six years (2017-2022) in Slovenia shows that eleven *map* genotypes are present and of these, three are common to grapevine and hazelnut (M38, M50 and M122). Others were specifically detected only in grapevine (M51, M54, M158) or only in hazelnut (M48, M159-M162).

As we know from previous studies of FD phytoplasmas molecular diversity of Slovenian grapevine strains, they belong to

three FD clusters based on nucleotide sequence and restriction fragment length polymorphism analysis of FD9 marker (Mehle et al., 2011). In this study, we went beyond and monitored *map* genotypes. Most FD outbreaks in Slovenian vineyards are associated with genotype M54 from Map-FD2 cluster. This genotype is the most abundant in France and in Slovenian neighbouring countries Italy, and Croatia (Plavec et al., 2019; Rossi et al., 2019; Malembic-Maher et al., 2020; Rizzoli et al., 2021). In Croatia, in the region of Istra near the Slovenian SW wine-growing region, Map-FD2 was the only genetic cluster discovered, and the high number of infected *S. titanus* in Istra suggests a high transmission efficiency (Plavec et al., 2019). M54, the only *map* genotype of the 16SrV-D taxonomic subgroup, was the only genotype detected in our *S. titanus* samples. Upon a broad continental survey, M54 was never found in alder trees, clematis, and alder leafhoppers (Malembic-Maher et al., 2020). In our study, M54 was also absent from hazelnut and *O. ishidae*. So, like in other European vineyards, this genotype seems restricted and adapted to the grapevine-*S. titanus* pathosystem, with a high epidemic potential. However, a study from Switzerland has shown that M54 can also occur in hazelnuts in the vicinity of vineyards infected with FD phytoplasma M54 genotype, and we cannot exclude a possible transmission to the wild compartment (Casati et al., 2017).

The other *map* genotypes present in Slovenian grapevines were much less frequent. The newly described M158, belonging to Map-FD1 cluster, was found specific to grapevine but present at five sites in the NE region in three consecutive years. Our MLSA analyses showed it was clustering together with isolates detected in alder, clematis, and grapevine, which might indicate a transfer from wild plants to grapevine or an evolution from a strain present on grapevine. Based on *VmpA-R1* sequence analysis, it belongs to the VmpA-III cluster, which means that it is compatible with *S. titanus* transmission and has epidemic potential on grapevine. The only Map-FD3 genotype detected in grapevine in Slovenia was M51. M51 whose natural non-vitis host is clematis, is the dominant genotype causing epidemics on grapevine in Serbia (Krstić et al., 2022). In Croatia its finding was restricted to the borders of Slovenia and Serbia (Plavec et al., 2019). M51 was also found in other hosts, *C. vitalba* or *A. altissima*, in Italy, Croatia, Hungary and Serbia (Plavec et al., 2019; Rossi et al., 2019; Malembic-Maher et al., 2020; Krstić et al., 2022). In Slovenia, clematis samples were previously analysed on the DNA segment of the FD9 marker, and they also belonged to the FD3 strains (Mehle et al., 2011). Thus, the M51 cases in Slovenia could either originate from locally infected clematis or be due to a transport of infected *S. titanus* or grapevine material from the neighbouring infected areas in Croatia.

In this study we detected three *map* genotypes that were common to grapevine and hazelnut (M38, M50 and M122). All three were previously found in the *S. titanus*-grapevine pathosystem causing possible outbreaks (Plavec et al., 2019; Malembic-Maher et al., 2020). A two SNP distant genotype of M54 is M38, which has been previously detected in alder and grapevine, is thought to be an ancestral genotype of M54 (Malembic-Maher et al., 2020). We observed the presence of M38 in all Slovenian wine-growing regions and in samples of hazelnut. The other genotype M122 detected in both hosts, grapevine and hazelnut, also belongs to the Map-FD2 cluster. M122, which differs from M38 by one SNP, was previously found in France, Hungary, and Croatia (Plavec et al., 2019). In addition to M38, M50 from Map-FD1 cluster is a genotype that appears to have escaped from alders. It is successful in the grapevine-*S. titanus* pathosystem (Malembic-Maher et al., 2020) and was here detected in 2% of Slovenian grapevine samples and 27% among hazelnut samples. Interestingly, M38, M50 and M122 genotypes from grapevine were also detected in hazelnuts sampled in the same large NE region, suggesting that phytoplasma exchanges could have occurred between both species, probably mediated by insect vectors. However, the locations of hazelnut orchards, where we found genotypes detected also in grapevine, are relatively far away from FD infected vineyards. The closest vineyard infected with M50 was 3.5 km of air distance away from hazelnut orchard where we also found M50. Based on these data, we could not determine a direct link between the 16SrV phytoplasmas in grapevine and in hazelnut. Other woody hosts between orchards and vineyards could also serve as a source of these genotypes, such as alder, which is a common original host of M38 and M50 in Europe (Malembic-Maher et al., 2020) and hazelnut might be, like grapevine, a secondary host. *Map* genotypes had previously been studied in Slovenia in only two hazelnut shrubs – one belonged to Map-FD1 and one to Map-FD2 (Mehle et al., 2019). In this study, where we analysed 30 hazelnuts, we also detected only genotypes from Map-FD1 and FD2 clusters. However, phytoplasmas of the 16SrV group were identified in asymptomatic, non-cultivated hazelnut bushes near FD-infected vineyards in Switzerland (Casati et al., 2017). There they detected genotypes from all three *map* clusters. In future studies, wild hazelnuts in Slovenia will be subjected to a phytoplasma search to determine whether they are infected and, if so, which genotypes are present in hazelnuts in the environment.

In a broad search for present *map* genotypes in Slovenia, we found one new genotype in grapevine (M158) and four new ones in hazelnut (M159-162). To better understand about their origin and their epidemic potential, we performed some additional analyses. Whether the genotype detected is compatible with transmission with *S. titanus* can be determined based on the *VmpA-R1* sequence, which codes for a protein important for adhesion to insect vector cells (Arricau-Bouvery et al., 2021). The three distinct clusters divide isolates into those that are transmissible by *S. titanus* (vectotype II and III) and those that are not (vectotype I) (Malembic-Maher et al., 2020). All new *map* genotypes discovered in this study, regardless of host (grapevine/hazelnut), belonged to vectotype III cluster, which groups isolates from the Map-FD1 and Map-FD3 clusters found in clematis, alder, and grapevine outbreaks (Malembic-Maher et al., 2020). In the same study, it was also shown that some isolates transmitted by *O. ishidae* belong to vectotype III cluster and that they are compatible with *S. titanus* transmission. Based on the genetic marker *dnaK* (Rossi et al., 2019), the newly discovered genotypes belong to two clusters - DnaK1 and DnaK2. One of the hazelnut samples - isolate SB21-10-K - was separated from the DnaK2 cluster along with an isolate from *Spartium junceum*. It is also the only isolate of the newly discovered *map* genotypes that clustered in the MLSA together with the reference strain FD92 and not with FD70 as the other four. The MLSA was based on four genes (*tuf, rplV, rplF, map*), but it would be better to look for genes throughout the whole genome in the future.

Regarding the spread of the disease, it is of outmost importance to also look for insect vectors in diseased hazelnut orchards. *O. ishidae* is

an invasive species that has been efficiently expanding its territory since its first detection in the EU in 1998 in northern Italy, probably due to its wide range of feeding preferences, from woody to herbaceous plants (Guglielmino, 2005). Since invasive Hemiptera species are known to have an important impact on the spread of phytoplasmas, *O. ishidae* came into focus as a possible vector of FD (Mehle et al., 2010; Lessio et al., 2016; Casati et al., 2017; Lessio et al., 2019). Populations of *O. ishidae* in the two hazelnut orchards studied in 2021 were numerous and the insects harboured 16SrV phytoplasmas. High populations of *O. ishidae* were also observed in Switzerland, in a forest adjacent to a FD-infected vineyard and composed mainly of hazel and willow (Casati et al., 2017). The genotypes detected in this study in *O.ishidae* were the same as those found in hazelnuts: M38, M50, M122 and M160. Previous studies collecting *O. ishidae* from infected alders in the wild showed that insects harboured the potentially epidemic *map* genotypes M38 and M50 and have been shown to efficiently transmit them back to alders (Desqué et al., 2019).

The diversity of different genotypes seems to be greater in hazelnut than in grapevine, with more genotypes discovered in a single orchard at NE than in the entire vineyards of Slovenia. A possible reason for this could be that polyphagous vectors feeding on a variety of plants (e.g. *O. ishidae*) introduce different genotypes from outside the orchards (e.g. from alders, clematis). However, infection rates are higher for specialised vector species such as *S. titanus* in vineyards (Bressan et al., 2006) than for polyphagous vectors, which tend to feed on a variety of also non-infected plants (Lessio et al., 2016; Casati et al., 2017). The high diversity of detected *map* genotypes in the one hazelnut orchard does not support a possible transmission by planting of infected material, which could then also spread inside the orchard by root bridges (Lešnik et al., 2008). We can assume that polyphagous *O. ishidae* may be responsible for genotype diversity in hazelnut and possibly also for genotype fluxes between orchards and vineyard compartments. Other woody plant hosts present in the wild (alder or hazel) could serve as a source or relay between both compartments. Other autochthonous vectors might also be involved, as indicated by our finding of genotype M48 in hazelnuts. It was previously found in German grapevines and belongs to vectotype I, which is compatible with Macropsinae but not with Deltocephalinae (*S. titanus* and perhaps *O. ishidae*) (Malembic-Maher et al., 2020). M48 has been frequently detected in alders and in the alder Macropsinae *Oncopsis alni* in France and Germany (Malembic-Maher et al., 2020). This finding contributes to the hypothesis that alders are the original reservoir also for the genotypes found on hazelnut.

This study provides evidence that grapevine and hazelnut can harbour the same epidemic *map* genotypes of 16SrV phytoplasmas. Considering our results, the invasive *O. ishidae* is most likely the causal agent of transmission, as it was prevalent in hazelnut orchards harbouring the same genotypes as the infected plants. It could also potentially transmit these phytoplasmas to grapevines, as it has been shown to transmit 16SrV phytoplasmas from alders to grapevines at low frequency (Desqué et al., 2019). The goal of confirming *O. ishidae* as a vector between hazelnuts and testing its potential to transmit from hazelnuts to grapevines will be addressed in future research, but we should be cautious not to exclude other possible vectors. If transmission is proven, action would need to be taken against this vector, whose ecology needs further research. This will provide more clarity on how the overlapping *map* genotypes got onto the hazelnut: direct linkage from grapevine through the vector or a linkage through alder as the original reservoir. Future research will also dive into ecological studies to decipher host plants, vectors and the diversity of *map* genotypes in wild environment outside orchards and vineyards. Measuring fluxes of phytoplasma genotypes and their vectors that connect vineyards and orchards through wild environments with high throughput metabarcoding technologies will help to monitor new risks of future emergence and outbreak development.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Ethics statement

Ethical review and approval were not required for the study on animals in accordance with the local legislation and institutional requirements.

Author contributions

ZZ participated in design of study, collection of material, laboratory experiments, genotyping analyses, data curation, visualisation and writing the manuscript. GS collected insect material. TJ and JB performed part of experiments in Slovenia. SP and PS helped perform MLSA experiments in France. Supervision and review were done by AV, NM, SM-M, XF. NM was included in conceptualisation and study design, formal analyses, funding acquisition and coordinated the study. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the Slovenian Ministry of Agriculture, Forestry, and Food through the Administration of the Republic of Slovenia for Food Safety, Veterinary, and Plant Protection. Part of the work was done in the frame of the Euphresco project FLADO-VIGILANT (2020-A-344). It was also supported by the Slovenian Research Agency (grant number P4-0165, L7_2632) and FEMS Research and Training Grant (FEMS-GO-2021-070). ZZ is a recipient of a Ph.D. research grant from the Slovenian Research Agency.

Acknowledgments

We thank the Slovenian phytosanitary inspectors and expert surveyors for providing the field samples. We also thank to Maja

Ravnikar, Denis Kutnjak, Martina Šeruga Musić and Erika Orešek for productive discussions, Nejc Jakoš and Luka Zwitter for assistance in field work and Irena Bajde for help in the laboratory.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer MM declared a shared consortium/research group with the authors NM, XF to the handling editor.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1217425/full#supplementary-material

References

Angelini, E., Squizzato, F., Lucchetta, G., and Borgo, M. (2004). Detection of a phytoplasma associated with grapevine flavescence dorée in clematis vitalba. *Eur. J. Plant Pathol.* 110, 193–203. doi: 10.1023/B:EJPP.0000015361.95661.37

Arnaud, G., Malembic-Maher, S., Salar, P., Bonnet, P., Maixner, M., Marcone, C., et al. (2007). Multilocus sequence typing confirms the close genetic interrelatedness of three distinct flavescence dorée phytoplasma strain clusters and group 16SrV phytoplasmas infecting grapevine and alder in Europe. *Appl. Environ. Microbiol.* 73, 4001–4010. doi: 10.1128/AEM.02323-06

Arricau-Bouvery, N., Duret, S., Dubrana, M. P., Desqué, D., Eveillard, S., Brocard, L., et al. (2021). Interactions between the flavescence dorée phytoplasma and its insect vector indicate lectin-type adhesion mediated by the adhesin VmpA. *Sci. Rep.* 11, 11222. doi: 10.1038/s41598-021-90809-z

Atanasova, B., Spasov, D., Jakovljević, M., Jović, J., Krstić, O., Mitrović, M., et al. (2014). First report of alder yellows phytoplasma associated with common alder (Alnus glutinosa) in the republic of Macedonia. *Plant Dis.* 98, 1268. doi: 10.1094/PDIS-03-14-0315-PDN

Bertaccini, A., Duduk, B., Paltrinieri, S., and Contaldo, N. (2014). Phytoplasmas and phytoplasma diseases: a severe threat to agriculture. *Am. J. Plant Sci.* 05, 1763–1788. doi: 10.4236/ajps.2014.512191

Bonfield, J. K., Smith, K. F., and Staden, R. (1995). A new DNA sequence assembly program. *Nucleic Acids Res.* 23, 4992–4999. doi: 10.1093/nar/23.24.4992

Bressan, A., Larrue, J., and Boudon Padieu, E. (2006). Patterns of phytoplasma-infected and infective scaphoideus titanus leafhoppers in vineyards with high incidence of flavescence dorée. *Entomol Exp. Appl.* 119, 61–69. doi: 10.1016/j.neulet.2018.05.038

Casati, P., Jermini, M., Quaglino, F., Corbani, G., Schaerer, S., Passera, A., et al. (2017). New insights on flavescence dorée phytoplasma ecology in the vineyard agro-ecosystem in southern Switzerland. *Ann. Appl. Biol.* 171, 37–51. doi: 10.1111/aab.12359

Caudwell, A. (1957). Deux années d'étude sur la flavescence dorée, nouvelle maladie grave de la vigne. *Annales D'amélioration Des. Plantes* 4, 359–363.

Christensen, N. M., Nicolaisen, M., Hansen, M., and Schulz, A. (2004). Distribution of phytoplasmas in infected plants as revealed by real-time PCR and bioimaging. *Mol. Plant-Microbe Interact.* 17, 1175–1184. doi: 10.1094/MPMI.2004.17.11.1175

Chuche, J., and Thiéry, D. (2014). Biology and ecology of the flavescence dorée vector scaphoideus titanus: a review. *Agron. Sustain Dev.* 34, 381–403. doi: 10.1007/s13593-014-0208-7

Davis, R. E., and Dally, E. L. (2001). Revised subgroup classification of group 16SrV phytoplasmas and placement of flavescence dorée-associated phytoplasmas in two distinct subgroups. *Plant Dis.* 85, 790–797. doi: 10.1094/PDIS.2001.85.7.790

Desqué, D., Salar, P., Danet, J., Lusseau, T., Garcion, C., Moreau, E., et al. (2019). Impact of orientus ishidae on "flavescence dorée" emergence in the vineyards of riparian ecosystems. *Phytopathogenic Mollicutes* 9, 69–70. doi: 10.5958/2249-4677.2019.00035.5

Filippin, L., and De Pra, V. (2011). Nucleotide sequencing of imp gene in phytoplasmas associated to 'flavescence dorée' from ailanthus altissima. *Bull. Insectol.* 64, S49–S50.

Filippin, L., Jović, J., Cvrković, T., Forte, V., Clair, D., Toševski, I., et al. (2009). Molecular characteristics of phytoplasmas associated with flavescence dorée in clematis and grapevine and preliminary results on the role of dictyophara europaea as a vector. *Plant Pathol.* 58, 826–837. doi: 10.1111/j.1365-3059.2009.02092.x

Food and Agriculture Organisation of the United Nations (2022) *FAOSTAT: crop and livestock products*. Available at: https://www.fao.org/faostat/en/#data/QCL.

Guglielmino, A. (2005). Observations on the genus orientus (Rhynchota cicadomorpha cicadellidae) and description of a new species: o. amurensis n. sp. from Russia (Amur region and maritime territory) and China (Liaoning province). *Marburger Entomologische Publikationen* 3, 99–110.

Hren, M., Boben, J., Rotter, A., Kralj, P., Gruden, K., and Ravnikar, M. (2007). Real-time PCR detection systems for flavescence dorée and Bois noir phytoplasmas in grapevine: comparison with conventional PCR detection and application in diagnostics. *Plant Pathol.* 56, 785–796. doi: 10.1111/j.1365-3059.2007.01688.x

Kogovšek, P., Hodgetts, J., Hall, J., Prezelj, N., Nikolić, P., Mehle, N., et al. (2015). LAMP assay and rapid sample preparation method for on-site detection of flavescence dorée phytoplasma in grapevine. *Plant Pathol.* 64, 286–296. doi: 10.1111/ppa.12266

Krstić, O., Cvković, T., Marinković, S., Jakovljević, M., Mitrović, M., Toševski, I., et al. (2022). Genetic diversity of flavescence dorée phytoplasmas in vineyards of Serbia: from the widespread occurrence of autochthonous map-M51 to the emergence of endemic map-FD2 (Vectotype II) and new map-FD3 (Vectotype III) epidemic genotypes. *Agronomy* 12, 448. doi: 10.3390/agronomy12020448

Krstić, O., Cvrković, T., Mitrović, M., Radonjić, S., Hrnčić, S., Toševski, I., et al. (2018). Wolbachia infection in natural populations of dictyophara europaea, an alternative vector of grapevine flavescence dorée phytoplasma: effects and interactions. *Ann. Appl. Biol.* 172, 74–64. doi: https://doi.org/10.1111/aab.12400

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 2870–2874. doi: 10.1093/molbev/msw054

Leigh, J., and Bryant, D. (2015). PopART: full-feature software for haplotype network construction. *Methods Ecol. Evol.* 6, 1110–1116. doi: 10.1111/2041-210X.12410

Lešnik, M., Brzin, J., Mehle, N., and Ravnikar, M. (2008). Transmission of "Candidatus phytoplasma mali" by natural formation of root bridges in M9 apple rootstock. *Agricultura (Slovenia)* 5, 43–46.

Lessio, F., Bocca, F., and Alma, A. (2019). Development, spatial distribution, and presence on grapevine of nymphs of orientus ishidae (Hemiptera: cicadellidae), a new vector of flavescence dorée phytoplasmas. *J. Econ Entomol* 112, 2558–2564. doi: 10.1093/jee/toz177

Lessio, F., Picciau, L., Gonella, E., Mandrioli, M., Tota, F., and Alma, A. (2016). The mosaic leafhopper orientus ishidae: host plants, spatial distribution, infectivity, and transmission of 16SrV phytoplasmas to vines. *Bull. Insectol.* 69, 277–289.

Lim, P. O., and Sears, B. B. (1989). 16S rRNA sequence indicates that plant-pathogenic mycoplasmalike organisms are evolutionarily distinct from animal mycoplasmas. *J. Bacteriol* 171, 5901–5906. doi: 10.1128/jb.171.11.5901-5906.1989

Malembic-Maher, S., Desqué, D., Khalil, D., Salar, P., Bergey, B., Danet, J.-L., et al. (2020). When a palearctic bacterium meets a nearctic insect vector: genetic and ecological insights into the emergence of the grapevine flavescence dorée epidemics in Europe. *PloS Pathog.* 16, 1–28. doi: 10.1371/journal.ppat.1007967

Malembic-Maher, S., Salar, P., Filippin, L., Carle, P., Angelini, E., and Foissac, X. (2011). Genetic diversity of European phytoplasmas of the 16SrV taxonomic group and proposal of "Candidatus phytoplasma rubi. *Int. J. Syst. Evol. Microbiol.* 61, 2129–2134. doi: 10.1099/ijs.0.025411-0

Marcone, C. (2014). Molecular biology and pathogenicity of phytoplasmas. *Ann. Appl. Biol.* 165, 199–221. doi: 10.1111/aab.12151

Mehle, N., Jakoš, N., Mešl, M., Miklavc, J., Matko, B., Rot, M., et al. (2019). Phytoplasmas associated with declining of hazelnut (Corylus avellana) in Slovenia. *Eur. J. Plant Pathol.* 155, 1117–1132. doi: 10.1007/s10658-019-01839-3

Mehle, N., Nikolić, P., Rupar, M., Boben, J., Ravnikar, M., and Dermastia, M. (2013a). "Automated DNA extraction for Large numbers of plant samples," in *Phytoplasma. methods in molecular biology (Methods and protocols)*. Eds. M. Dickinson and J. Hodgetts (Totowa, NJ: Humana Press), 139–145.

Mehle, N., Prezelj, N., Hren, M., Boben, J., Gruden, K., Ravnikar, M., et al. (2013b). "A real-time PCR detection system for the Bois noir and flavescence dorée phytoplasmas and quantification of the target DNA," in *Phytoplasma: methods and protocols, methods in molecular biology*. Eds. M. Dickinson and J. Hodgetts (Totowa, NJ: New York: Humana Press).

Mehle, N., Rupar, M., Seljak, G., Ravnikar, M., and Dermastia, M. (2011). Molecular diversity of "flavescence dorée" phytoplasma strains in Slovenia. *Bull. Insectol.* 64, 29–30.

Mehle, N., Seljak, G., Rupar, M., Ravnikar, M., and Dermastia, M. (2010). The first detection of a phytoplasma from the 16SrV (Elm yellows) group in the mosaic leafhopper orientus ishidae. *New Dis. Rep.* 22, 11–11. doi: 10.5197/j.2044-0588.2010.022.011

Miladinović, Z., Radulović, M., Hrnčić, S., and Delić, D. (2019). Genetic variability of scaphoideus titanus populations and new host plant of 16SrV group phytoplasmas in Bosnia and Herzegovina. *Phytopathog. Mollicutes* 9, 111–112. doi: 10.5958/2249-4677.2019.00056.2

Papura, D., Burban, C., Helden, M., Giresse, X., Nusillard, B., Guillemaud, T., et al. (2012). Microsatellite and mitochondrial data provide evidence for a single major introduction for the neartic leafhopper scaphoideus titanus in Europe. *PloS One* 7 (5), e36882. doi: 10.1371/journal.pone.0036882

Plavec, J., Budinšćak,, Križanac, I., Škorić, D., Foissac, X., and Šeruga Musić, M. (2019). Multilocus sequence typing reveals the presence of three distinct flavescence dorée phytoplasma genetic clusters in Croatian vineyards. *Plant Pathol.* 68, 18–30. doi: 10.1111/ppa.12934

Radonjić, S., Hrnčić, S., Krstić, O., Cvrković, T., Mitrović, M., Jović, J., et al. (2013). First report of alder yellows phytoplasma infecting common and grey alder (Alnus glutinosa and a. incana) in Montenegro. *Plant Dis.* 97, 686. doi: 10.1094/PDIS-11-12-1087-PDN

Rizza, S., Rossi, M., Spallino, R. E., Tedeschi, R., Marzachì, C., Catara, V., et al. (2021). Ecology-based analysis of a recent association between spartium junceum and 16SrV phytoplasma. *Plant Pathol.* 70, 305–317. doi: 10.1111/ppa.13300

Rizzoli, A., Belgeri, E., Jermini, M., Conedera, M., Filippin, L., and Angelini, E. (2021). Alnus glutinosa and orientus ishidae (Matsumura 1902) share phytoplasma genotypes linked to the ' flavescence dorée ' epidemics. *J. Appl. Entomology*, 145, 1015–1028. doi: 10.1111/jen.12933

Rossi, M., Pegoraro, M., Ripamonti, M., Abbà, S., Bealb, D., Giraudoa, A., et al. (2019). Genetic diversity of flavescence dorée phytoplasmas at the vineyard scale. *Appl. Environ. Microbiol.* 85, 1–15. doi: 10.1128/AEM.03123-18

Tramontini, S., Delbianco, A., and Vos, S. (2020). Pest survey card on flavescence dorée phytoplasma and its vector scaphoideus titanus. *EFSA Supporting Publications*. 17 (8), EN-1909. doi: 10.2903/sp.efsa.2020.en-1909

Weintraub, P. G., and Beanland, L. (2006). Insect vectors of phytoplasmas. *Annu. Rev. Entomol* 51, 91–111. doi: 10.1146/annurev.ento.51.110104.151039

2.2 First Report of Grapevine Flavescence Dorée Phytoplasma in the Czech Republic

Authors: Šárka Linhartová, Kateřina Tománková, Hana Orságová, Miroslava Hejlová, Václav Čermák, Jakub Beránek, Nataša Mehle, **Zala Kogej Zwitter**, Ales Eichmeier

Published in: *Plant Disease*, vol 107 (10), 2023
doi: 10.1094/PDIS-02-23-0315-PDN

Disease Note

Diseases Caused by Bacteria and Phytoplasmas

First Report of Grapevine Flavescence Dorée Phytoplasma in the Czech Republic

Š. Linhartová,[1] K. Tománková,[1] H. Orságová,[1] M. Hejlová,[1] V. Čermák,[1,2] J. Beránek,[1,2] N. Mehle,[3,4] Z. Kogej Zwitter,[3,5] and A. Eichmeier[6,†]

[1] National Reference Laboratory, Division of Plant Pest Diagnostics, Department of Plant Pest Diagnostics Olomouc, Central Institute for Supervising and Testing in Agriculture, 779 00 Olomouc, Czech Republic
[2] Faculty of Forestry and Wood Technology, Department of Forest Protection and Game Management, Mendel University in Brno, 613 00 Brno, Czech Republic
[3] Department of Biotechnology and Systems Biology, National Institute of Biology, SI-1000 Ljubljana, Slovenia
[4] School for Viticulture and Enology, University of Nova Gorica, SI-5271 Vipava, Slovenia
[5] Jožef Stefan International Postgraduate School, SI-1000 Ljubljana, Slovenia
[6] Faculty of Horticulture, Mendeleum – Institute of Genetics, Mendel University in Brno, 691 44 Lednice, Czech Republic

Š. Linhartová and H. Orságová contributed equally and are co-first authors.

Funding: Funding was provided by Ministerstvo Zemědělství (grant no. QK22010031) and Ministerstvo Školství, Mládeže a Tělovýchovy (grant no. CZ.02.1.01/0.0/0.0/16_025/0007314). Plant Dis. 107:3273, 2023; published online as https://doi.org/10.1094/PDIS-02-23-0315-PDN. Accepted for publication 17 July 2023.

Viticulture is a traditional branch of agriculture in the Czech Republic. Grapevines (*Vitis vinifera* L.) are cultivated on more than 18,000 ha in the wine-growing regions of Bohemia and South Moravia. South Moravia alone accounts for more than 90% of the total wine-growing area in the country. Grapevine yellows are a complex of diseases associated with the presence of phytoplasmas. Phytoplasmas of at least five different groups can cause similar symptoms in grapevines, and they can be distinguished only on a molecular basis (EPPO Bulletin 2016). One of them, grapevine Flavescence dorée phytoplasma (GFDP), which belongs to the 16SrV group, is listed in Annex II, Part B, of the Commission Implementing Regulation (EU) 2019/2072 of 28 November 2019 as a Union quarantine pest known to occur in the Union territory. Official surveys for GFDP in the Czech Republic have been carried out since 2007. In 2016, the first occurrence of *Scaphoideus titanus* Ball, 1932, the main vector of GFDP, was reported in the South Moravian region (EPPO Reporting Service 2016). This is a matter of concern because it indicates that there is a risk of disease dissemination to other geographical locations. In September 2021, a total of 250 samples of *V. vinifera* (preferentially focused on symptomatic plants) and four samples of the wild plant host *Clematis vitalba* L. were collected from 50 vineyards in South Moravia. Total DNA was extracted using the High Pure PCR Template Preparation Kit (Roche, Basel, Switzerland). For phytoplasma screening, a real-time PCR test for generic detection of phytoplasmas was used (Christensen et al. 2004). Samples evaluated as positive were further tested by PCR using phytoplasma universal P1 and P7 primers (Deng and Hiruki 1991; Schneider et al. 1995), followed by nested PCR using the 16SrV group-specific primers fB1 and rULWS1 (Smart et al. 1996). For identification of 16SrV phytoplasma, sequence analysis of the *secY-map* genetic locus was performed. Two sets of primers were used: FD9f5/MAPr1 primers for the first PCR and FD9f6/MAPr2 for the nested PCR (Arnaud et al. 2007) with PCRBIO TaqMix (PCR Biosystems, London, U.K.). The nested PCR products were purified and sequenced (Eurofins Genomics, Ebersberg, Germany). The sequences were compared with sequences from the GenBank database. Phytoplasma of the 16SrV group was detected in three samples: *V. vinifera* cv. Gewürztraminer with symptoms of leaf reddening with no rolling and no other typical symptoms, *C. vitalba* with leaf curling, and symptomless *C. vitalba*. The obtained sequences of the *secY-map* locus of all three 16SrV-positive samples were identical to the sequence of GFDP, isolate Vv-SI257 (accession no. FN811141), detected in grapevine in Tuscany, Italy, which belongs to the 16SrV group. The sequence of the *V. vinifera* cv. Gewürztraminer isolate was submitted to GenBank under accession number OQ185203. This isolate belongs to the Map-FD3 cluster, and the genotype identified is M51, which has already been found in *C. vitalba* and outbreaks of Flavescence dorée in grapevines in some other European countries (Malembic-Maher et al. 2020). Based on the abovementioned results, this is the first report of GFDP in the Czech Republic.

References:

Arnaud, G., et al. 2007. Appl. Environ. Microbiol. 73:4001.
Christensen, N. M., et al. 2004. Mol. Plant-Microbe Interact. 17:1175.
Deng, S., and Hiruki, C. 1991. J. Microbiol. Methods 14:53.
EPPO Bulletin. 2016. 46:78.
EPPO Reporting Service. 2016. 9-2016, Num. article 2016/165.
Malembic-Maher, S., et al. 2020. PLoS Pathog. 16:e1007967.
Schneider, B., et al. 1995. Mol. Diagn. Proc. Mycoplasmol. 1:369.
Smart, C. D., et al. 1996. Appl. Environ. Microbiol. 62:2988.

The author(s) declare no conflict of interest.

***e*-Xtra**

Keywords: *Clematis*, GFDP, grapevine

†Indicates the corresponding author.
A. Eichmeier; ales.eichmeier@mendelu.cz

© 2023 The American Phytopathological Society

2.3 First Detection of a Phytoplasma From the 16SrV-C (Elm Yellows) Group in *Alnus Glutinosa* in the Netherlands

Authors: Pier Paul Max de Koning, Ruben Schoen, **Zala Kogej Zwitter**, Naomi te Braak, Anne K. J. Giesbers, Marcel Westenberg, Nataša Mehle, Marleen Botermans

Published in: *New Disease Reports*, vol 52 (1), 2025
doi: 10.1002/ndr2.70056

New Disease Reports

WILEY

NEW DISEASE REPORT OPEN ACCESS

First Detection of a Phytoplasma From the 16SrV-C (Elm yellows) Group in *Alnus glutinosa* in the Netherlands

P. P. M. de Koning[1] | R. Schoen[1] | Z. Kogej Zwitter[2,3] | N. te Braak[1] | A. K. J. Giesbers[1] | M. Westenberg[1] | N. Mehle[2,4] | M. Botermans[1]

[1]Netherlands Institute for Vectors, Invasive Plants and Plant Health (NIVIP), National Plant Protection Organization (NPPO), Netherlands Food and Consumer Product Safety Authority (NVWA), Wageningen, the Netherlands | [2]Department of Biotechnology and Systems Biology, National Institute of Biology, Ljubljana, Slovenia | [3]Jozef Stefan International Postgraduate School, Ljubljana, Slovenia | [4]School for Viticulture and Enology, University of Nova Gorica, Vipava, Slovenia

Correspondence: P. P. M. de Koning (p.p.m.dekoning@nvwa.nl)

Received: 21 May 2025 | **Accepted:** 14 August 2025

Keywords: black alder | M44 | *secY-map*

In September 2024, a phytoplasma was detected in six mature black alder (*Alnus glutinosa*) trees growing naturally in the municipality of Wageningen, the Netherlands. The trees displayed no phytoplasma- or virus-like symptoms. The samples were collected as part of a project, which aimed to enhance knowledge of viruses and other non-culturable organisms in forestry tree species. Black alder was included in this study because it has been reported as a host for the EU quarantine pest grapevine flavescence dorée phytoplasma (GFDP) (Malembic-Maher et al. 2020). GFDP is regulated due to its role in causing 'flavescence dorée' (FD), a grapevine disease characterised by severe symptoms and decline. GFDP was previously considered as '*Candidatus* Phytoplasma vitis'; however, as its 16S rRNA gene sequences share over 98.65% identity with those of other phytoplasmas (Bertaccini et al. 2022), it is not formally classified as a distinct '*Candidatus* Phytoplasma' species. This genetic similarity complicates its classification and regulatory status.

Leaves were collected from different branches of each tree. DNA was extracted (DNeasy Plant Mini Kit; Qiagen, the Netherlands) from the midribs and secondary veins. Phytoplasma was detected in all six trees by generic real-time PCR primers JH-F1/JH-F all/JH-R and probe JH-P (Hodgetts et al. 2009). One sample was further tested by nested PCR using the P1/P7 and R16F2n/R16R2 primers following EPPO Standard PM7/133 (European and Mediterranean Plant Protection Organization 2018). The PCR amplicon obtained through conventional PCR was purified (QIAquick PCR Purification Kit; Qiagen, the Netherlands) and sequenced using Illumina technology. According to the *i*PhyClassifier (Zhao et al. 2013), the resulting 16S rRNA sequence (GenBank Accession No. PV565616) belonged to the 16SrV-C subgroup. A BLASTn search showed 99.92% identity with several 16SrV phytoplasmas including GFDP (AF176319), Alder yellows phytoplasma (PP097681) and '*Candidatus* phytoplasma ulmi' (MH279539). To differentiate between these 16SrV phytoplasmas, the *secY-map* sequence was amplified from the same sample by nested PCR with primer pairs FD9f5/MAPr1 and FD9f6/MAPr2 (Arnaud et al. 2007). The PCR amplicon was purified and Sanger sequenced (Macrogen Europe, the Netherlands). The *secY-map* sequence (PV563858) was compared with known map genotypes in CLC Genomics Workbench (version 24, Qiagen, Slovenia) with BLAST tool and matched genotype M44 (Figure 1). Genotype M44 is not linked to epidemic clusters Map-FD1, Map-FD2 or Map-FD3, which include genotypes associated with *Scaphoideus titanus* transmission and FD outbreaks. Instead, the M44 genotype has previously been reported in black alder in France, Germany and Italy, and in the vector *Oncopsis alni* in France and Germany (Malembic-Maher et al. 2020).

To our knowledge, this is the first report of phytoplasma from the 16SrV-C (Elm yellows) group detected in the Netherlands. Following previous reports that 16SrV phytoplasmas do not always cause symptoms in black alder (Malembic-Maher et al. 2020),

P.P.M. de Koning and R. Schoen contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *New Disease Reports* published by British Society for Plant Pathology and John Wiley & Sons.

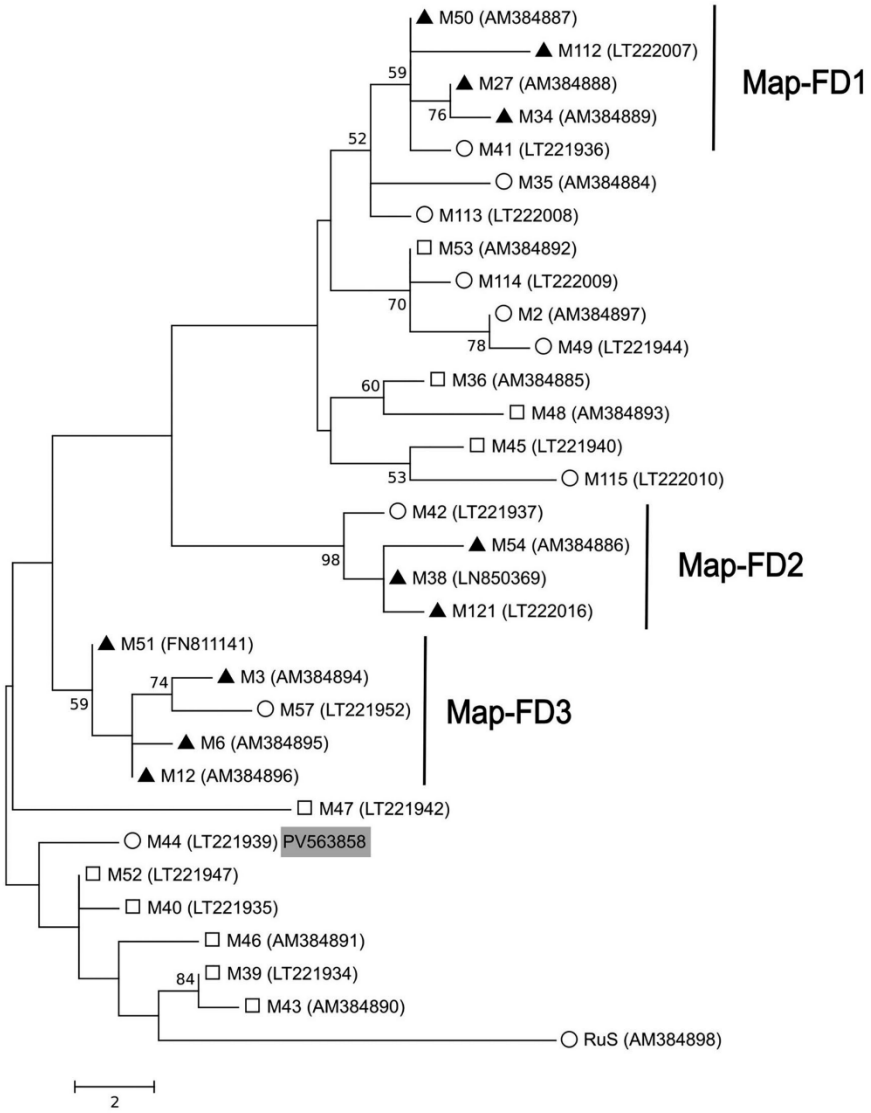


FIGURE 1 | Unrooted maximum parsimony phylogenetic tree of the *secY-map* sequence of the M44 isolate from *Alnus glutinosa* in the Netherlands (marked in grey) and reference sequences of 16SrV phytoplasmas. GFDP isolates are marked with filled triangles, grapevine non-epidemic isolates with squares and 16SrV isolates from other plants with circles; RuS—'*Candidatus* Phytoplasma rubi'; the tree was constructed in MEGA7 with 500 bootstrap replicates (bootstrap values indicated on branches when ≥50) based on a ClustalW alignment of 674 bp sequences; the scale bar represents the number of nucleotide changes per site.

the sampled trees in this study appeared healthy. The identified phytoplasma was not considered a quarantine pest, and therefore, no phytosanitary measures were taken. This finding highlights the value of baseline studies, including asymptomatic sampling, to identify reservoir hosts and gain insight in phytoplasma distribution, diversity and ecology. Such studies are essential for risk assessment, regulation refinement and improving early detection strategies. Moreover, this example underlines the complexity of phytoplasma taxonomy and the challenges in diagnosing regulated phytoplasma species.

Acknowledgements

The sampling and analysis at NIVIP was conducted within the Euphresco project 2023-A-446, Surveillance for viruses and other non-culturable pathogens of forestry tree species (VIRNOTree). Map genotyping has been performed within the framework of the activities of the EU reference laboratory for Pests of Plants on Viruses, Viroids and Phytoplasmas.

Data Availability Statement

The nucleotide sequences from this study can be found in GenBank (https://www.ncbi.nlm.nih.gov/genbank/) with GenBank accession no. PV565616 and PV563858.

References

Arnaud, G., S. Malembic-Maher, P. Salar, et al. 2007. "Multilocus Sequence Typing Confirms the Close Genetic Interrelatedness of Three Distinct *Flavescence dorée* Phytoplasma Strain Clusters and Group 16SrV Phytoplasmas Infecting Grapevine and Alder in Europe." *Applied and Environmental Microbiology* 73: 4000–4011. https://doi.org/10.1128/AEM.02323-06.

Bertaccini, A., Y. Arocha-Rosete, N. Contaldo, et al. 2022. "Revision of the '*Candidatus* Phytoplasma' Species Description Guidelines." *International Journal of Systematic and Evolutionary Microbiology* 72: 005353. https://doi.org/10.1099/ijsem.0.005353.

European and Mediterranean Plant Protection Organization. 2018. "PM 7/133 (1) Generic Detection of Phytoplasmas." *EPPO Bulletin* 48: 414–424. https://doi.org/10.1111/epp.12541.

Hodgetts, J., N. Boonham, R. Mumford, and M. Dickinson. 2009. "Panel of 23S rRNA Gene-Based Real-Time PCR Assays for Improved Universal and Group-Specific Detection of Phytoplasmas." *Applied and Environmental Microbiology* 75: 2945–2950. https://doi.org/10.1128/AEM.02610-08.

Malembic-Maher, S., D. Desqué, D. Khalil, et al. 2020. "When a Palearctic Bacterium Meets a Nearctic Insect Vector: Genetic and Ecological Insights Into the Emergence of the Grapevine Flavescence Dorée Epidemics in Europe." *PLoS Pathogens* 16: e1007967. https://doi.org/10.1371/journal.ppat.1007967.

Zhao, Y., W. Wei, I. M. Lee, J. Shao, X. Suo, and R. E. Davis. 2013. "The *i*PhyClassifier, an Interactive Online Tool for Phytoplasma Classification and Taxonomic assignment." In *Phytoplasma. Methods in Molecular Biology, vol 938*, edited by M. Dickinson and J. Hodgetts, 329–338. Humana Press. https://doi.org/10.1007/978-1-62703-089-2_28.

2.4 From Vineyard to Genome: Optimised Enrichment and Sequencing of Flavescence Dorée Phytoplasma from Grapevine Samples

Authors: **Zala Kogej Zwitter**, Denis Kutnjak, Nataša Mehle

Published in: *Microbial Genomics*, vol 11 (9), 2025
doi: 10.1099/mgen.0.001514

MICROBIAL GENOMICS

RESEARCH ARTICLE
Kogej Zwitter *et al.*, *Microbial Genomics* 2025;11:001514
DOI 10.1099/mgen.0.001514

MICROBIOLOGY SOCIETY
OPEN DATA OPEN ACCESS

From vineyard to genome: optimized enrichment and sequencing of Flavescence dorée phytoplasma from grapevine samples

Zala Kogej Zwitter[1,2,*], Denis Kutnjak[1] and Natasa Mehle[1,3,*]

Abstract

Phytoplasmas are non-culturable obligate intracellular bacteria that cause considerable economic losses in agriculture. Genome sequencing provides crucial insights into their biology and vector dependence. However, genome studies on phytoplasmas are often hampered by their low abundance in naturally infected plants. Propagation in test plants is usually necessary but time-consuming and resource-intensive, especially for quarantine phytoplasmas such as the phytoplasma causing Flavescence dorée (FD), a serious threat to European viticulture. To overcome these challenges, we aimed to develop a protocol for efficient enrichment of phytoplasma DNA directly from field-collected samples, enabling genome sequencing using both Illumina and Oxford Nanopore Technologies platforms. We evaluated six sample preparation protocols that included stepwise enrichment steps to improve phytoplasma genome coverage and assembly quality. The most effective approach combined differential centrifugation, CTAB extraction and removal of CpG-methylated host DNA and resulted in a notable increase in the relative abundance of phytoplasma reads compared to other protocols. Rarefaction analysis of the dataset generated using this protocol demonstrated that the entire phytoplasma genome was covered by reads in a dataset comprising 3 billion nucleotides. We also evaluated and compared *de novo* phytoplasma genome assemblies generated from short Illumina reads and long nanopore sequencing reads. While Illumina sequencing yielded more accurate assemblies with longer total lengths, the assemblies derived from nanopore sequencing data contained longer individual contigs. This advantage was reflected in hybrid assemblies that combined both technologies, yielding longer phytoplasma contigs than assemblies from Illumina datasets and lower mismatch rates compared to assemblies from nanopore sequencing datasets. A hybrid *de novo* assembled genome of the Slovenian FD phytoplasma isolate achieved 96% reference genome coverage, with high contiguity and low error rates. This streamlined and accessible protocol enables high-quality genome sequencing of phytoplasma-infected grapevines without the need for propagation in test plants. This facilitates broader phytoplasma research and can potentially be extended to other naturally infected phytoplasma hosts or organisms infected with other non-culturable microbes.

Impact Statement

Flavescence dorée (FD) disease is a major threat to European viticulture. It is associated with a quarantine phytoplasma that is difficult to analyse due to its low abundance in infected plants and the high proportion of host DNA in samples. Current methods for sequencing phytoplasma genomes often require propagation in test plants, a time-consuming and resource-intensive process. In this study, we developed an accessible, field sample-based protocol for phytoplasma DNA enrichment that eliminates the need for specialized facilities such as quarantine greenhouses. By optimizing a combination of pre- and post-extraction enrichment steps, we achieved high genome coverage and high-quality assembly of a Slovenian FD phytoplasma

[*Continued on next page*]

Received 23 April 2025; Accepted 05 September 2025; Published 29 September 2025
Author affiliations: [1]Department of Biotechnology and Systems Biology, National Institute of Biology, Ljubljana, Slovenia; [2]Jožef Stefan International Postgraduate School, Ljubljana, Slovenia; [3]School for Viticulture and Enology, University of Nova Gorica, Nova Gorica, Slovenia.
***Correspondence:** Natasa Mehle, natasa.mehle@nib.si; Zala Kogej Zwitter, zala.kogej.zwitter@nib.si
Keywords: enrichment; grapevine; HTS; nanopore; plant pathogen.
Abbreviations: FD, Flavescence dorée; FDp, Flavescence dorée phytoplasma; ONT, Oxford Nanopore Technologies; WGA, whole-genome amplification; BLAST, Basic Local Alignment Search Tool; QUAST, Quality Assessment Tool.
Eleven supplementary materials are available with the online version of this article. All supporting data, code and protocols have been provided within the article or through supplementary data files.
001514 © 2025 The Authors

This is an open-access article distributed under the terms of the Creative Commons Attribution License.

isolate directly from infected grapevines using high-throughput sequencing. This approach simplifies phytoplasma genomics research and enables more efficient monitoring, epidemiological studies and development of diagnostic tools. In addition, this protocol could be adapted for sequencing other naturally infected phytoplasma hosts and plants infected with other non-culturable bacteria, expanding its potential applications in plant pathology and microbial genomics.

DATA SUMMARY

The developed protocol is accessible at DOI: dx.doi.org/10.17504/protocols.io.8epv52w5jv1b/v1. All genome sequencing data, including sequencing reads (SRX28207723, SRX28087407, SRX28087406, SRX28087405, SRX28087404 and SRX28087403) and the final genome assembly, are available in the Sequence Read Archive and NCBI genome repositories under the BioProject number PRJNA1231790 and GenBank number GCA_050897995. The authors confirm that all supporting data, codes and protocols have been provided in this article or in supplementary data files.

INTRODUCTION

Phytoplasmas are obligate intracellular phytopathogenic bacteria that infect more than 1,000 plant species worldwide [1]. By disrupting the growth of stem, leaf or fruit tissue, phytoplasmas reduce the yield or survival of infected plants, including commercially valuable crops, ornamentals and fruit trees [2]. These non-culturable bacteria belong to the class *Mollicutes* and the taxon '*Candidatus* Phytoplasma', which is categorized into groups, subgroups and species based on the genetic information of 16S rRNA and other conserved genes [3, 4]. Phytoplasmas have no cell wall and reproduce in the phloem of plants and in the haemolymph of insect vectors [5]. Insects that feed on the phloem sap of infected plants, such as leafhoppers, planthoppers and psyllids, transmit phytoplasmas from one plant to another [6]. The host range of phytoplasmas is therefore determined by the feeding preferences of these vectors [7]. Phytoplasmas can also spread by vegetative propagation methods such as cuttings, grafts and micropropagation [1].

As they cannot grow in cell-free media, molecular methods targeting DNA sequences are essential for the detection and investigation of phytoplasmas. Genome sequencing is particularly valuable for understanding their biology, host dependence and interactions with insect vectors and plants [8]. However, sequencing of complete phytoplasma genomes remains a major challenge, as phytoplasma DNA typically comprises less than 1% of the total nucleic acids extracted from infected plant material [9]. This is not only due to their smaller genome size compared to their hosts [10], but also to their generally low concentrations in natural hosts, which can vary considerably depending on the plant organ [11, 12] and growth stage [13]. The high proportion of host DNA complicates sequencing and increases costs, while the high AT content (often more than 70%) and the numerous duplicated genes in phytoplasma genomes cause difficulties in assembly based on short-read sequencing [14].

To overcome these challenges, many studies propagate phytoplasmas in alternative hosts, such as *Catharanthus roseus* or *Vicia faba*, by transferring them to these hosts through grafting [15], dodder (*Cuscuta* sp.) [16] or insect vectors [17]. These test plants often harbour higher phytoplasma concentrations [18], making them suitable starting material for genome sequencing. Propagation in test plants is time-consuming and requires specialized, pest-proof greenhouses, especially for quarantine pests such as Flavescence dorée phytoplasma (FDp). Unfortunately, there is a lack of protocols that would allow direct sequencing of field samples, especially of woody plants. Enrichment methods can help solve these problems by increasing the phytoplasma DNA concentration after total DNA extraction. Common techniques include PFGE [15, 16, 19–21] and caesium chloride buoyant-density gradient centrifugation [16, 22]. Non-toxic alternatives such as iodixanol gradients have also been used for DNA separation [23]. Other enrichment approaches include immunoprecipitation-based methods [24], methyl-CpG binding domain-mediated techniques [25] and DNA amplification strategies such as multiple displacement amplification [26] and rolling circle amplification. Methods such as subtractive hybridization and mirror orientation selection have also been shown to be effective for the enrichment of phytoplasma DNA [9].

As of 10 July 2025, the NCBI Genome database lists 38 complete genomes of 16 different phytoplasma species, generated using different enrichment methods and sequencing platforms [15–17, 19–22, 24, 27–42]. While short-read platforms such as Illumina are commonly used, long-read platforms such as Oxford Nanopore Technologies (ONT) and PacBio are becoming increasingly popular. These platforms can be combined in hybrid assemblies to improve the quality of the genome [43]. Nanopore sequencing offers advantages for *de novo* assembly as it can generate long reads without the need for DNA amplification or synthesis during the sequencing process. However, the achievement of high-quality long reads is highly dependent on sample and library preparation, especially the extraction and handling of intact, high-molecular-weight DNA [44].

Nanopore sequencing in combination with Illumina sequencing data was used to complete the first and only available genome sequence of FDp propagated in test plants [17]. FDp from taxonomic subgroups 16SrV-C and 16SrV-D [45] is a quarantine pest in the EU (Regulation 2019/2072) and the causal agent of a serious grapevine disease that is widespread in European vineyards [46].

First documented in France in the 1950s [47], the spread of FDp has increased rapidly, fuelled by its vector, the invasive species *Scaphoideus titanus* Ball (Hemiptera: Cicadellidae), originally from North America [48]. Despite control measures, the disease persists, emphasizing the need for further research.

To overcome the challenges in obtaining the FDp genome sequence, we have developed an efficient and easily accessible protocol for the direct preparation of field-collected grapevine samples for whole-genome sequencing of phytoplasmas that eliminates the need for propagation in test plants. This protocol integrates and optimizes existing enrichment techniques both pre- and post-DNA extraction to maximize the yield of phytoplasma DNA. The protocol has also been adapted for use with the ONT MinION sequencer, a platform suitable for small laboratories. Different combinations of enrichment methods were evaluated based on their impact on reference genome coverage, sequencing depth and quality of the *de novo* assembled genome. By integrating Illumina and nanopore sequencing data, we successfully generated a high-quality draft genome of a Slovenian FDp isolate, demonstrating the effectiveness and applicability of the protocol for phytoplasma genomics.

METHODS

Plant material and homogenization

The plant material used in this study was a symptomatic grapevine sample from a vineyard in the northeast of Slovenia, variety 'Renski rizling', collected in the second half of August 2020 (sample ID D933/20). The sample consisted of leaves taken from different parts of the plant. The leaf veins were cut out and chopped into small pieces with a sterile scalpel. To ensure that the sampled grapevine was infected with FDp, DNA was extracted from 1 g of leaf veins using the QuickPick™ SML Plant DNA kit (Bio-Nobile) and the KingFisher mL Purification System (Thermo Fisher Scientific) [49]. The DNA was then subjected to nested PCR and Sanger sequencing of the secY-map locus with FD9f5/MAPr1 and FD9f6/MAPr2 primer pairs [50]. Determination of the *map* genotype and conformation of infection with FDp was performed according to the procedure described by Kogej Zwitter *et al.* [51]. Subsequently, a total of 10 g of small pieces of leaf veins was homogenized with a mortar and pestle using liquid nitrogen. Using a pre-cooled spatula, the powdered plant tissue sample was transferred into pre-cooled 15 ml tubes. Six aliquots containing 1 g of homogenized plant tissue were prepared from this sample. The tubes containing the plant material were then immediately stored at −80 °C.

To assess the consistency of the enrichment, two additional samples of grapevine and two samples of hazelnut were prepared and homogenized using the same procedure as described above. The grapevine samples were collected in northeast Slovenia – one at the end of August 2020 (sample ID: D1062/20, unknown cultivar) and the other at the beginning of June 2025 (sample ID: 572/25, cultivar Pinot Gris). Both hazelnut samples came from phytoplasma-infected orchards in the same region. One was collected in mid-July 2021 (sample ID: BS21-P-1) and the other in mid-October 2020 (sample ID: 1461/20). The latter differed from the others in that it originated from root phloem tissue and not from leaf veins.

Comparison of different protocols for phytoplasma sequencing

To find out how phytoplasmas can be reliably and rapidly enriched and their genomic sequences obtained from a field grapevine sample (D933/20), aliquots of the homogenized sample were analysed using six different protocols (named as protocols A–F). In these protocols, one step was added at a time, so that we could assess the impact of each step on the final result. The steps for each protocol are listed below and shown schematically in Fig. 1, with detailed procedures described in the following sections:

- –A: DNA extraction with KingFisher (ThermoScientific, USA) and Illumina sequencing
- –B: DNA extraction with CTAB and Illumina sequencing
- –C: differential centrifugation, DNA extraction with CTAB and Illumina sequencing
- –D: differential centrifugation, DNA extraction with CTAB, enrichment with NebNext Microbiome DNA Enrichment Kit (New England Biolabs, MA, USA) and Illumina sequencing
- –E: differential centrifugation, DNA extraction with CTAB, enrichment with NebNext Microbiome DNA Enrichment Kit (New England Biolabs, MA, USA), whole-genome amplification (WGA) and Illumina sequencing
- –F: differential centrifugation, DNA extraction with CTAB, enrichment with NebNext Microbiome DNA Enrichment Kit (New England Biolabs, MA, USA), WGA and ONT sequencing.

The four additional samples of grapevine and hazelnut (D1062/20, 572/25, BS21-P-1, 1461/20) were prepared according to protocols A, C and D up to the sequencing step (Fig. 1).

Pre-extraction enrichment: differential centrifugation

The first step in the enrichment of phytoplasmas was differential centrifugation. 1 g of the homogenised sample was processed with 10 ml of Phytoplasma Grinding Buffer (PGB: 96 mM K_2HPO_4, 30 mM KH_2PO_4, 10% sucrose, 0.15% bovine serum albumin, 2% polyvinylpyrrolidone-10, 30 mM ascorbic acid, pH 7.6). The mixture was kept on ice for 10 min to allow the cells to plasmolyse. The extract was then poured into fresh, chilled 15 ml tubes and centrifuged (Sigma, 3-18KS, rotor 12171) at 2,500 ***g*** for 5 min at 4 °C, allowing environmental and cell debris to settle to the bottom. The supernatant was poured into fresh, chilled 15 ml tubes

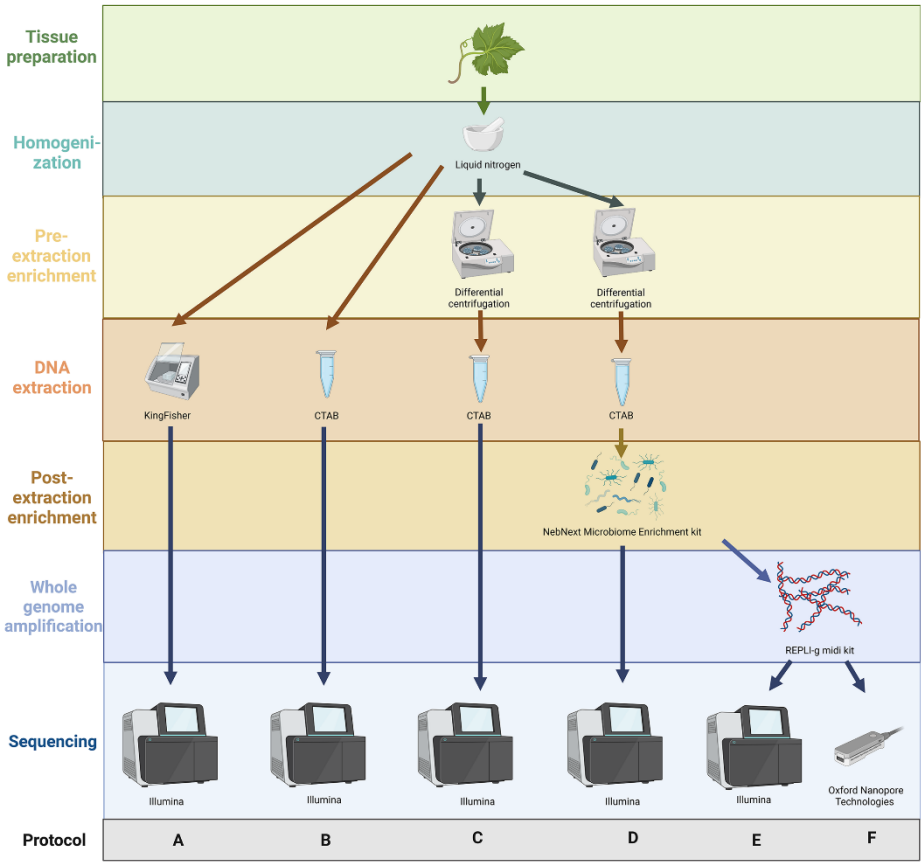


Fig. 1. Schematic representation of the workflow and testing of different protocols (A–F) used to obtain the genomic sequence of FDp. All protocols started with a homogenized sample of grapevine leaves (D933/20) crushed in liquid nitrogen (top). The diagram outlines the steps involved in each of the six protocols, labelled A–F at the bottom of the figure. Arrows illustrate the flow of the different steps, which are explained on the left-hand side of the figure, with additional information next to each step.

and equilibrated in pairs with PGB. These tubes were then centrifuged at 18,000 ***g*** for 25 min at 4 °C (Sigma, 3-18KS, rotor 12171). The supernatant was then poured off, and the resulting pellet with enriched phytoplasma material was immediately processed using the CTAB extraction method.

DNA extraction

KingFisher total DNA extraction

Semi-automated extraction of total DNA with magnetic particles using the KingFisher device was included in one of the protocols, as it is a rapid extraction option that can be used when processing a large number of samples. To the homogenized plant material, 2 ml of Plant DNA lysis buffer (QuickPick Plant DNA Kit; Bio-Nobile, Finland) was added and vortexed thoroughly for 1 min. The following steps of the extraction protocol using the QuickPick™ SML Plant DNA Kit (Bio-Nobile, Finland) and a KingFisher extraction device (ThermoScientific, USA) followed the procedure described in [49].

CTAB extraction of total DNA

The CTAB extraction protocol (adapted from [52, 53]) was included in five of the compared protocols. In 15 ml tubes, 3 ml of CTAB buffer (4% CTAB, 1% PVP40, 0.2 M Tris-HCl, 1.4 M NaCl, 20 mM EDTA, pH 8) heated to 65 °C was added to homogenized plant material (protocol B), and 2 ml of CTAB buffer with the same composition was added to the pellet after differential centrifugation (protocols C–F). The material was resuspended by vortexing. The tubes were incubated at 65 °C for 1 h, vortexed in between (every 20 min) and then centrifuged at 5,000 ***g*** for 5 min at room temperature. From this step onwards, the same sample was split and processed in two parallel extractions: twice, 1 ml of the sample was transferred to a 2 ml tube, and the same amount of chloroform:isoamyl alcohol (24:1) was added to each tube and vortexed. After centrifugation (10 min, 6,000 ***g***, room temperature), the upper aqueous phases were transferred to fresh 1.5 ml tubes. An equal volume of ice-cold isopropanol was added to each tube and mixed by inverting. The DNA was then precipitated for 20 min at room temperature. After centrifugation (15 min, 15,000 ***g***, 4 °C), the pellets were washed with ice-cold 70% ethanol and, after additional centrifugation (2 min, 10,000 ***g***, 4 °C) and drying, resuspended in 50 µl of nuclease-free water. The DNA extracted from two parallel extractions was combined in one tube.

Post-extraction enrichment: microbiome enrichment

For enrichment after DNA extraction, the NEBNext® Microbiome DNA Enrichment Kit was used (New England Biolabs, MA, USA), which removes methylated eukaryotic host DNA from the sample. The manufacturer's protocol was followed, using the maximum amount of DNA possible (up to 1 µg, in this case from 1,050 to 382 ng for different samples in the corresponding sample volume). The collected enriched microbial DNA was cleaned using Agencourt AMPure XP beads (option A in the NebNext Microbiome DNA enrichment protocol) (Beckman Coulter, Brea, CA, USA).

Whole-genome amplification

Since the selected nanopore sequencing library requires high amounts of input nucleic acids, that could not be obtained from our sample, an additional WGA step was performed in two of the protocols (E and F) as suggested in the ONT Ligation sequencing gDNA – WGA protocol (SQK-LSK110, version WAL_9115_v110_revH_10Nov2020). To obtain a sufficiently high concentration of total DNA suitable for nanopore sequencing (protocol F and, for comparison, protocol E, using Illumina sequencing), phytoplasma-enriched DNA was subjected to WGA using the REPLIg-midi Kit (Qiagen, Hilden, Germany) with an input of 10 ng of enriched DNA according to the Protocol for Amplification of Purified Genomic DNA using the REPLI-g Midi Kit.

Sequencing

For protocols A–E, Illumina sequencing was performed by Novogene (Cambridge, UK) using the Plant and Animal Whole Genome sequencing service on Illumina NovaSeq (PE150 mode) aiming for 5 G bases of raw data per sample.

Nanopore sequencing was used in this study as an option for rapid in-house sequencing and to test whether potentially longer sequencing reads could be beneficial for genome assembly. Nanopore sequencing was performed using the SQL-LSK110 Ligation sequencing kit (Oxford Nanopore Technologies Limited, Oxford, UK). Library preparation followed the Oxford Nanopore protocol Ligation sequencing gDNA – WGA (version: WAL_9115_v110_revH_10Nov2020, Oxford Nanopore Technologies Limited) with some modifications: Mag-Bind Total Pure NGS beads were used for the purification steps (Omega Bio-tek, Norcross, Georgia), and after the endonuclease reaction, cleaning was performed with 1.8 volumes of Mag-Bind Total Pure NGS beads (instead of Custom buffer). In the DNA repair and end-prep step, 1 µl of DNA CS from the SQL-LSK110 kit was added along with 47 µl of prepared genomic DNA (in this case, 1269 ng) as a control of the sequencing run. Long Fragment Buffer was used in the final clean-up step after adapter ligation. Thirty-eight femtomoles of the finally prepared library were loaded onto the flow cell. Sequencing on the MinION with flow cell (R9.4.1) was run for 72 h, with basecalling switched off and a minimum of 20 bp read length.

Measurements and quality control of DNA in each protocol

To determine DNA quantity across different protocols, DNA concentrations were measured using the dsDNA HS kit on a Qubit fluorometer (Thermo Fisher Scientific, MA, USA) and with the ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, USA). For the whole-genome-amplified sample prepared for nanopore sequencing (protocol F), DNA concentration and fragment lengths were determined using the TapeStation and Genomic DNA ScreenTape Kit (Agilent Technologies, Santa Clara, USA) to ensure the presence of high molecular-weight DNA suitable for library preparation and accurate loading of the library onto the flow cell.

Real-time PCR (qPCR) was used to verify the presence of phytoplasma DNA, as well as its host eukaryotic DNA in the samples prepared with each of the protocols prior to sequencing and to evaluate the consistency of the enrichment between samples. Tenfold dilutions of the isolated DNA (added amounts of sample D933/20 DNA ranged between 0.4 and 77.2 ng depending on the protocol, as indicated in Table 1) were tested using the 18S rRNA assay (Applied Biosystems, MA, USA) to control extraction and estimate the level of eukaryotic host DNA, and the specific assay for the 16SrV phytoplasma group [54] in two replicates. The qPCR reactions were performed as described in [55]. In addition to the samples, negative controls of extraction (with water instead of a biological sample) and negative controls of amplification were also tested in two replicates.

Table 1. Results of specific tests and bioinformatic analyses on grapevine sample D933/20 prepared with different protocols (A–F, see Fig. 1); DNA yield measured after isolation and enrichment (when performed); mean Cq values of two technical replicates for 18S rRNA and specific 16SrV real-time PCR assay; total HTS dataset size in billion nucleotides, which was subsampled in five replicates to the smallest dataset size for comparison; HTS read mapping results represent the average of five replicates, each subsampled to 4.8 billion nucleotides and mapped to the host (*Vitis vinifera*, GCF_000003745.3) and FDp reference genome (CP097583); consensus sequence length corresponds to the mapping-based consensus

Protocol	A	B	C	D	E	F
DNA concentration (ng µl^{-1})	3.9	28.0	2.2	2.4	386.0	
Total DNA yield (ng)	784	2,800	216	118	11,580	
18S rRNA assay (Cq)	19.1	17.3	21.7	26.0	19.4	
16SrV-specific assay (Cq)	28.9	25.0	28.3	27.3	18.9	
Total nt (billion)	6.1	4.8	4.9	6.0	7.0	7.4
Subsampled total nt (billion)	4.8	4.8	4.8	4.8	4.8	4.8
Nt mapped to host (%)	95.84	96.02	93.89	89.53	91.27	92.48
Nt mapped to FD ref (%)	0.03	0.04	0.07	0.26	0.17	0.15
Consensus length (FDp)	547,983	501,505	576,564	653,574	650,919	653,544
Fraction of FDp ref covered	0.84	0.77	0.88	1.00	1.00	1.00
Average depth (FDp)	3	3	5	19	12	10

Cq, quantification cycle; nt, nucleotides; ref, reference.

Bioinformatic analyses

All bioinformatic analyses were performed on the grapevine sample D933/20 using two high-performance servers (for characteristics, see Material S1, available in the online Supplementary Material, and Table 1). The software tools were run via command line, and CLC Genomic Workbench (Qiagen, Hilden, Germany) was running on Server 1.

Read quality control, subsampling and mapping to the reference

The short Illumina reads were analysed using CLC Genomic Workbench V23 (Qiagen, Hilden, Germany). The reads were first trimmed of adapters, and the quality limit was set to 0.05 without trimming homopolymers. The sequencing runs resulted in different output sizes. In order to compare the datasets, they were therefore normalized to the same size. The read datasets from the different protocols were randomly subsampled to 4.8 billion nucleotides using the CLC tool Subsample Sequence List, which corresponded to the size of the smallest dataset. Reads were mapped to the *V. vinifera* genome available at NCBI (RefSeq assembly GCF_000003745.3) to evaluate the proportion of reads mapped to the host and to the FDp reference genome sequence CP097583 from GenBank [17], using the parameters specified in Material S1.

The nanopore sequencing reads were first basecalled with Guppy (version 6.4.2). All passed files were merged into one fastq file, and based on the NanoPlot results, reads were trimmed by 50 nucleotides at the head and tail, and reads shorter than 50 nucleotides were discarded, all using NanoFilt. The statistics of the filtered reads were generated using nanoQC, NanoStat and NanoPlot. The merged filtered data were imported into CLC Genomics Workbench (version 24) and analysed using the Long Read Support plugin. The dataset was randomly subsampled to 4.8 billion nucleotides using the CLC Subsample Sequence List tool, estimating the number of reads based on the average read length after NanoFilt filtering. The reads were then mapped to the host (grapevine) and the FDp reference (as indicated above) using the Map Long Reads to Reference tool using the parameters stated in Material S1.

To investigate how the size of the datasets affects the genome coverage and to see how much data is needed to cover the majority of the FDp reference genome by read mapping, all 6 originally obtained datasets were downsized to 0.9, 1.8, 2.4, 3, 3.6 and 4.8 billion nucleotides in 5 replicates as stated above. Reads from each replicate were mapped to the FDp reference genome as described above. Data visualization of the fraction of the FDp reference covered and the average depth of all dataset points was performed in R (version 2022.12.0+353) using the *ggplot2* package.

Evaluation and comparison of *de novo* assemblies

De novo assemblies of reads from datasets obtained using different protocols were performed to assess the efficiency in *de novo* reconstruction of the phytoplasma genome. For this comparison, one replicate of each protocol (A–F) with the largest normalized read dataset size (4.8 billion nucleotides) was used. The Illumina sequencing datasets (protocols A–E) were assembled using

SPAdes version 3.14.1 [56] with default parameters. For the nanopore sequencing dataset (protocol F), assemblers designed for long reads were used. Initially, two *de novo* assemblers were compared on the complete dataset from protocol F: Flye [57], which was previously used for the *de novo* assembly of the FDp genome [17], and the *De Novo* Assemble Long Reads tool in CLC Genomics Workbench (Qiagen, Hilden, Germany), with parameters detailed in Material S1. As the CLC assembler produced superior results (Material S2), it was used for further analyses of dataset F. The assemblies were evaluated using the genome Quality Assessment Tool (QUAST, version 5.2) [58] with the FDp reference genome (CP097583). QUAST was used to generate comparative diagrams of *de novo* assemblies from different protocols.

We further investigated the performance of *de novo* assembly of phytoplasma genomes using datasets generated with the protocol that yielded the highest fraction of phytoplasma reads and included Illumina sequencing (D) and a protocol with nanopore sequencing (F). For these comparisons, genome assemblies were performed on subsampled datasets of different sizes, as well as on a mixed dataset to also test a hybrid assembly mode (assembling Illumina and nanopore sequencing data together). The datasets obtained with protocols D and F were downsized to 2.4, 4.8 and 6 billion nucleotides in 5 replicates. For each replicate, *de novo* assembly was performed as described above for Illumina or nanopore sequencing data. For the hybrid assembly, the SPAdes hybrid option (version 4.0.0., default settings) was used for datasets of the same sizes, with half of the data (1.2, 2.4 and 3 billion nucleotides) coming from the dataset generated with protocol D and the other half from the dataset generated with protocol F (1.2, 2.4 and 3 billion nucleotides). All assemblies performed were compared in QUAST, and the results were displayed in a diagram in R as described above. The quality metrics of the assemblies and contig lengths were plotted in a ridgeplot group distribution in R.

To perform a reference-independent comparison of the *de novo* assembled genomes, the assembled contigs were also analysed using the Basic Local Alignment Search Tool (BLAST) to identify all contigs belonging to FDp. Contigs generated from a randomly selected repeat of the *de novo* assembly using the largest comparable dataset (6 billion nucleotides) for each of the three approaches (D, F and hybrid) were analysed. First, a BLASTN search was performed using a custom phytoplasma database created by downloading all nucleotide sequences associated with the search term txid33926[Organism] from the NCBI database on 3 October 2024. All contigs with BLAST hits were then subjected to a second BLASTN search against the complete NCBI nt database (on 19 February 2023). The resulting output files were used for taxonomic classification of the contigs using MEGAN (version 6.25.10), and the contigs identified as hits for '*Candidatus* Phytoplasma vitis', 16SrV Elm yellows group and '*Ca*. P vitis' contigs incorrectly assigned to the unclassified NCBI group were further analysed.

The final assembly of the Slovenian FDp isolate was performed using the whole dataset of Illumina reads from protocol D and the whole dataset from nanopore sequencing (protocol F) in hybrid assembly with the SPAdes hybrid mode (version 4.0.0, default settings). Additionally, the Illumina reads obtained with protocol D were remapped to contigs (parameters in Material S1), and corrected consensus sequences of contigs were extracted. This corrected assembly was evaluated against the FDp genome reference (CP097583) using QUAST, as mentioned above, and the average nucleotide identity (ANI) value was calculated using FastANI version 1.34 [59] with the FDp reference genome. The final draft genome was submitted to GenBank (under BioProject PRJNA1231790).

RESULTS

Characterization of plant material

The grapevine and hazelnut samples used in this study were confirmed to be infected with FDp using qPCR. The grapevine sample D933/20 used for bioinformatic comparison of the protocols was determined to be genotype M54 [60] based on *map* genotyping with Sanger sequencing of the nested PCR product (GenBank accession number PV085772). The only available FDp reference genome (CP097583), which was also used as a reference in our analyses, corresponds to the same *map* genotype.

DNA concentrations and qPCR results

The total DNA yield varied considerably depending on the extraction and enrichment methods used in the preparation of sample D933/20 with different protocols (Table 1). The CTAB extraction method (protocol B) yielded significantly more DNA than the KingFisher method (protocol A), while the enrichment steps, including differential centrifugation (protocol C) and post-extraction treatment (protocol D), reduced total DNA yield. The qPCR analysis confirmed the successful extraction of FDp in all protocols. The relationship between the qPCR values of the 16SrV phytoplasma-specific assay and the 18S rRNA host DNA marker illustrated the effects of enrichment: a lower 16SrV Cq value indicated a higher phytoplasma DNA concentration, while a higher 18S rRNA Cq value reflected a depletion of host DNA. This inverse relationship indicates that the enrichment steps effectively increase the proportion of phytoplasma DNA while minimizing the presence of host DNA (Table 1). The additional step of WGA (protocol E and F) increased the overall DNA yield by amplifying both phytoplasma and host DNA (Table 1). Nanodrop measurements of DNA concentration and purity (A260/280 and A260/230) and Qubit DNA concentration values can be found in the Material S3 for the different steps of each protocol. Negative controls confirmed the absence of contamination.

Evaluation of enrichment consistency using qPCR

To evaluate the consistency of the enrichment protocols and their applicability to different hosts, we tested four additional 16SrV phytoplasma-infected samples – two from grapevine and two from hazelnut – using three protocols selected to represent the methodological range: a rapid extraction (protocol A), a pre-extraction enrichment (protocol C) and a combined pre- and post-extraction enrichment approach (protocol D).

qPCR analysis confirmed that the enrichment approaches shown to be effective in sample D933/20 were also effective in other grapevine samples, as well as in hazelnut samples from both leaf and root tissue (Material S4). The level of host DNA, as determined by the 18S rRNA assay, decreased consistently with each additional enrichment step, as indicated by increased Cq values: between protocols A and C, 18S Cq increased by 2.0–4.1 cycles, and between protocols C and D by an additional 3.5–8.6 cycles (Material S4). Meanwhile, the phytoplasma signal measured by an assay specific to the 16SrV phytoplasma group remained approximately the same, with Cq changes between protocols ranging from −1.6 to 1.3, indicating stable phytoplasma DNA content during enrichment (Material S4).

Evaluation of different protocols based on mapping to the FDp reference genome using subsampled datasets

The sequencing runs performed with different protocols (A–F) with sample D933/20 yielded outputs in the range of 4.8–7.4 billion nucleotides (Table 1). The quality control with TapeStation of the high-molecular-weight DNA and the final preparation of the library confirmed the suitability for nanopore sequencing (Material S5). This resulted in the highest output achieved with MinION nanopore sequencing, with DNA CS confirming the successful sequencing run (Material S5). Mapping these complete datasets from different protocols to the FDp reference revealed distinct distributions of mapped read lengths. As expected, for the Illumina sequencing runs, the mean read length was between 146.98 and 149.36 bp, according to the sequencing mode, while for the nanopore sequencing, it was 5,818.82 bp (Material S6).

To ensure comparability, all datasets of the different protocols were randomly subsampled to the smallest dataset size and compared based on the percentage of nucleotides mapped to the host genome (*V. vinifera*) and the FDp reference genome (CP097583). Protocols A and B, which did not include any enrichment step, yielded the highest percentage of reads mapping to the host genome (Table 1). Among these protocols, protocol A yielded the lowest percentage of phytoplasma reads. Changing the DNA extraction method from KingFisher (protocol A) to CTAB (protocol B) slightly improved the yield of phytoplasma reads by 1.3-fold (Table 1). The inclusion of an enrichment step prior to CTAB extraction (protocol C) led to a further increase in phytoplasma reads and yielded almost double the relative phytoplasma reads compared to protocol B (Table 1). Protocol D showed the most significant improvement and achieved a 4-fold increase in phytoplasma reads compared to protocol C (Table 1). This enrichment resulted in complete coverage of the FDp reference genome with an average depth of 19× and the longest consensus sequence among all protocols tested at the normalized subsample size. Incorporating WGA in protocol E maintained complete genome coverage and produced a high-quality consensus sequence but resulted in a slightly lower (1.5-fold) percentage of phytoplasma reads compared to protocol D (Table 1), while the percentage was similar (1.1-fold lower) to protocol E when using nanopore sequencing (protocol F). While WGA also increased the percentage of host reads, protocols E and F still achieved 12-fold and 10-fold average depth of the FDp reference genome at the normalized subsample size, respectively.

Performing a rarefaction analysis, in which randomly subsampled datasets (at six dataset sizes) were analysed in five replicates, allowed an unbiased comparison of the performance of the different protocols (A–F) when mapping to the FDp reference genome (Fig. 2a). Across all protocols, the proportion of the average depth of the FDp reference genome increased with the size of the sequencing data subsample (0.9–4.8 billion nucleotides), although the rate of increase varied considerably between protocols (Fig. 2b).

Protocols D, E and F performed best and achieved comparable genome coverage of 100% for D and 99% for E and F with a subsample size of 3 billion nucleotides (Fig. 2a). At this subsample size, the sequences of protocol A had an average genome coverage of 71%, protocol B 67% and protocol C 82%. Below this dataset size, protocol D provided better reference genome coverage than E and F (Fig. 2a). Protocols A and B performed the worst and did not achieve the genome coverage achieved by D, E and F, even at the largest subsample size (Fig. 2a). Protocol C, which included only one enrichment step of differential centrifugation, showed intermediate performance. The inclusion of enrichment steps (protocols C–F) contributed to higher genome coverage, with additional enrichment after extraction (D–F) resulting in better performance than enrichment before extraction alone (C) (Fig. 2a). For all protocols, a clear linear increase in average depth was observed with increasing dataset size (Fig. 2b). However, protocol D showed a steeper increase in average depth than the others (Fig. 2b). In contrast, protocols A and B showed a minimal improvement in average depth with larger datasets (Fig. 2b).

Evaluation of different protocols based on quality of genome assembly of subsampled datasets

To determine the most suitable *de novo* assembler for long-read sequencing data, two assemblers were tested with the full-size nanopore sequencing dataset (protocol F) and evaluated using QUAST (Material S2). The *De novo* Assemble Long Reads tool in

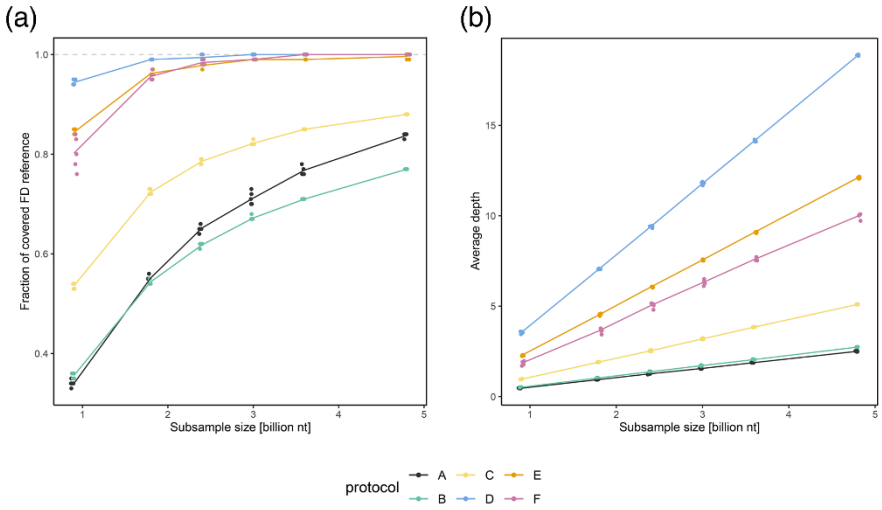


Fig. 2. Rarefaction analysis showing mapping results at 6 different subsampled dataset sizes (0.9, 1.8, 2.4, 3, 3.6 and 4.8 billion nucleotides) for 6 compared sample preparation protocols (A–F, colour-coded according to the legend). For each protocol/subsample dataset size combination, analysis was performed in five replicates (each dot represents the result for a single replicate). (a) Proportion of the FDp reference genome covered as a function of a subsample size. (b) Average depth of the FDp reference genome as a function of a subsample size.

CLC Genomics Workbench outperformed Flye and covered a higher fraction of the FDp reference genome with fewer contigs (Material S2). Based on these results, the CLC tool was selected for further comparisons.

Subsequently, randomly subsampled datasets of equal size (4.8 billion nucleotides) from protocols using Illumina sequencing (A–E) were assembled using SPAdes, while the nanopore subsample was assembled with CLC Genomics Workbench, and all assemblies were compared in QUAST (Material S7). The results showed significant differences in reference genome coverage and assembly quality between the protocols (Fig. 3a). Protocols A and B, in which no enrichment step was performed, resulted

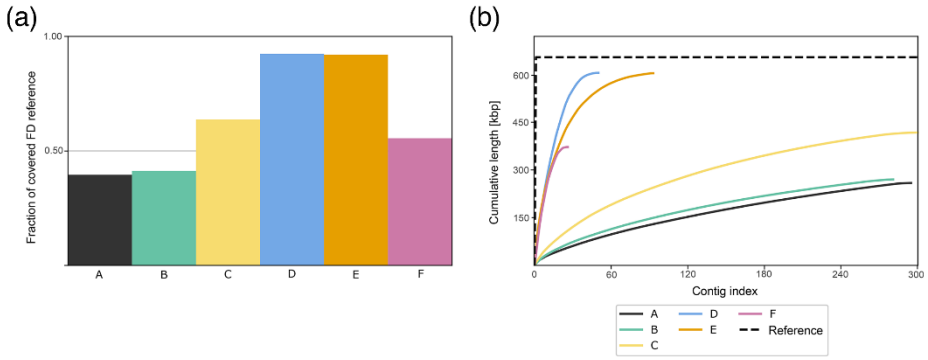


Fig. 3. QUAST-based quality metrics for the *de novo* assemblies of randomly subsampled datasets (size of 4.8 billion nucleotides) of different sample preparation protocols (A–F, colour-coded according to the legend on the right, on both panels). (a) The proportion of the FDp reference genome covered with assembled contigs. (b) Cumulative plot from QUAST showing the total assembled phytoplasma contig length as contigs are added sequentially from longest to shortest (the contig index represents the position of contigs sorted by length, with lower indices corresponding to longer contigs and higher indices to shorter contigs). kbp, 1,000 bp.

in assemblies in which less than 50% of the reference genome was covered by the assembled phytoplasma contigs. In contrast, the addition of a simple enrichment step, such as differential centrifugation (protocol C), resulted in assemblies with genome coverage improved to 64%. Protocols D and E with additional enrichment steps after DNA extraction resulted in assemblies that achieved over 90% of the reference genome coverage.

Protocols A–C resulted in more than 270 contigs mapped to the reference genome, whereas protocols D–F yielded fewer than 100 contigs, reflecting a lower degree of fragmentation (Material S7). The steep curves of the cumulative plot of aligned contigs indicate that protocols D, E and F contained long contigs that quickly accumulate a large portion of the reference genome (Fig. 3b). Looking at the NGA50 values (Material S7), i.e. the length of the shortest contig that, together with longer contigs, covers at least 50% of the reference genome, protocol D shows the best performance with longer and more continuous contigs (NGA50=18,116). Protocols E and F have lower NGA50 values (11,880 and 10,569, respectively). Protocol F, which used the same enrichment and WGA sample as protocol E but was sequenced using nanopore sequencing technology, achieved lower genome coverage and assembly quality for this subsampled dataset size. The largest alignment was 12,895 nucleotides shorter than the largest alignment achieved with protocol E (Material S7). In summary, the *de novo* assembly of the FDp genome was most successful using the dataset obtained with protocol D.

Comparative analysis of assembly performance across best-performing sample preparation protocols including Illumina and nanopore sequencing and their hybrid mode

To further evaluate the *de novo* assemblies, protocol D (the best-performing protocol including Illumina sequencing), protocol F (protocol including nanopore sequencing) and their hybrid mode assembly (D and F) were compared across 3 dataset sizes: 2.4, 4.8 and 6 billion nucleotides (Fig. 4). For the hybrid assemblies, half of the input data were sampled from Illumina and the other half from the nanopore sequencing dataset. Each *de novo* assembly was repeated five times with randomly subsampled datasets. The *de novo* assembly from nanopore sequencing datasets (protocol F) was strongly influenced by random subsampling and dataset size, which affected the coverage of the FDp reference genome by contigs (Material S8). However, this effect was much less pronounced in the hybrid assembly, as shown by the coefficient of variation of genome coverage between replicates for the largest dataset size (0.05 for protocol F compared to 0.004 for the hybrid assembly) (Material S8).

At the smallest dataset size (2.4 billion nucleotides), the phytoplasma contigs obtained from the nanopore sequencing dataset (F) covered only a small portion (7–17%) of the reference genome (Fig. 4). Coverage improved with increasing dataset size but

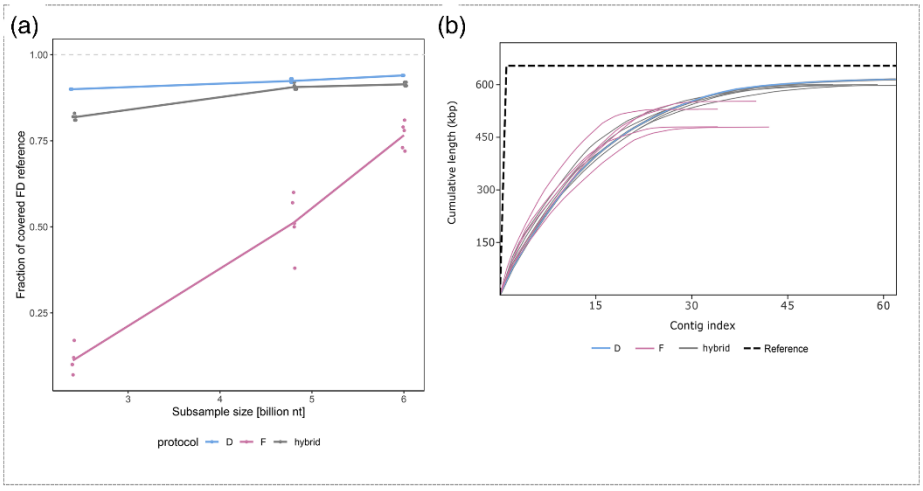


Fig. 4. (a) Rarefaction analysis showing FDp reference genome coverage by contigs from *de novo* assemblies generated using protocols D and F and their hybrid assembly (combining equal parts of data from protocols D and F). Assemblies are colour-coded according to the legend and analysed at 3 subsampled dataset sizes (2.4, 4.8 and 6 billion nucleotides). Each dot represents the result of a single replicate, with five replicates per assembly and dataset size combination. (b) Cumulative plot from QUAST for assemblies of 6-billion-nucleotide datasets using protocols D and F and their hybrid mode (colour-coded according to the legend); the contig index represents the position of phytoplasma contigs sorted by length, with lower indices corresponding to longer contigs and higher indices to shorter contigs.

remained lower than the coverage of protocol D and hybrid mode assemblies. In contrast, assemblies from datasets obtained with protocol D consistently achieved the highest genome coverage, exceeding 90% even for the smallest dataset (Fig. 4). Hybrid assemblies performed better than protocol F but did not outperform protocol D. They showed moderate genome coverage at smaller dataset sizes and approached the performance of protocol D at the largest dataset (Fig. 4).

Assemblies of datasets using protocol D produced fewer phytoplasma contigs than the hybrid protocol at a subsample size of 2.4 billion nucleotides (an average of 97 with protocol D and 227 with the hybrid mode), but the opposite happened at the largest subsample size (protocol D had an average of 60 contigs and the hybrid mode 46) (Material S8). Protocol F consistently produced the lowest number of phytoplasma contigs, averaging 24 with the 6-billion-nucleotide big dataset (Material S8). The cumulative plot of the QUAST analysis of the 6-billion-nucleotide datasets showed that the assemblies from nanopore and hybrid approaches produced longer phytoplasma contigs in some replicates than those obtained with protocol D (Fig. 4b). Nevertheless, at the 6-billion-nucleotide dataset point, the assemblies using protocol D produced an average of 60 contigs covering 94% of the reference genome, while the hybrid assemblies averaged 46 contigs covering 92% (Material S8).

To assess whether nanopore sequencing read assembly or hybrid assembly results in longer phytoplasma contigs, we compared the distribution of phytoplasma contig lengths in different datasets (D, F and hybrid) for a 6-billion-nucleotide subsample size (Fig. 5 and Material S9). The protocol D assemblies yielded relatively uniform phytoplasma contig lengths across replicates, with ~60% of the contigs in each replicate exceeding ~4,000 bp. In contrast, protocol F assemblies yielded fewer but generally longer contigs, all exceeding 5,000 bp, although these assemblies exhibited high variability between replicates, particularly in maximum contig length (Fig. 5). The hybrid assemblies contained features of both approaches and had a wide range of contig sizes, from ~500 bp to ~91,600 bp (Fig. 5 and Material S9).

Metrics for assembly quality varied between the three sequencing approaches when considering assemblies from datasets containing 6 billion nucleotides. Protocol D resulted in consistent NGA50 values between repeats and the shortest longest alignment, with low misassembly and mismatch rates (Fig. 5 and Material S8). Protocol F produced fewer phytoplasma contigs, had greater variability in NGA50 values and longest alignment lengths between 72,154 and 47,265 but had the highest mismatch and misassembly rates (Fig. 5 and Material S8). Hybrid assemblies had intermediate NGA50 values and longest alignment lengths between 56,110 and 41,063, with mismatch and misassembly rates higher than protocol D but lower than protocol F (Fig. 5 and Material S8).

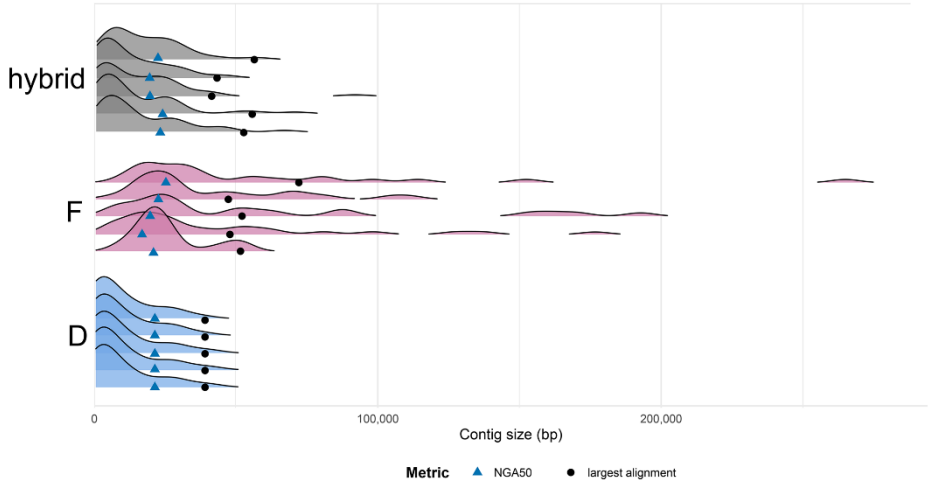


Fig. 5. Contig length distributions and assembly quality metrics (calculated by QUAST) for 6-billion-nucleotide datasets assembled using protocol D, protocol F and their hybrid mode in five replicates each. Contig length distributions are shown as ridgeline density plots, where the x-axis represents contig size and the y-axis indicates the estimated probability density within each assembly repetition. The height of each curve reflects the relative probability density, not absolute counts, and is scaled independently for each group to allow visual comparison of distribution shapes. Assembly quality metrics include NGA50 (blue triangles) and the largest alignment for each assembly (black dots).

To determine whether a reference-free analysis of the assembled phytoplasma contigs could reveal additional genomic regions that differ from the reference, an additional BLAST analysis was performed to search for possible additional FDp contigs within the assemblies. This analysis focussed on a random replicate of each assembly of the largest datasets obtained with the three protocols (D, F and hybrid). When assembling the datasets obtained with protocol D and the hybrid assembly, BLAST identified more contigs than reported by QUAST (140 contigs in protocol D compared to 61 found by QUAST, and 42 contigs in the hybrid assembly compared to 37 found by QUAST). However, all but one of the additional contigs identified by BLAST were short (<500 bp) and were therefore excluded by QUAST, which only considers contigs ≥500 bp. An additional contig identified by BLAST in the hybrid assembly was a hybrid of sequences from *V. vinifera* (26% query coverage, 95.72% identity) and FDp (40% query coverage, 95.42% identity). When assembling the dataset obtained with protocol F, BLAST did not identify any additional contigs beyond those reported in the QUAST analysis.

Slovene isolate of FDp genome

The draft genomic sequence of the FDp isolate (D933/20) analysed in this study was generated using a hybrid *de novo* assembly approach combining complete datasets from Illumina and nanopore sequencing and additionally corrected with Illumina reads mapped to contigs. The addition of the complete nanopore sequencing dataset (dataset size: 7.4 billion nucleotides) to the Illumina-derived data from protocol D (dataset size: 6 billion nucleotides) improved the quality of the genome assembly (Material S10). The 17 assembled contigs from the hybrid assembly covered 96% of the FDp reference genome. The largest aligned contig covered one quarter of the entire genome. The assembly showed good contiguity with a high NGA50, minimal misassembly rates, only three misassembled contigs and no local or scaffold gap misassemblies detected (Material S10). Low mismatch and indel rates are further evidence of the accuracy of the assembly (Material S10). The calculated ANI value of 99.97% shows that the Slovenian isolate is highly similar to the only FDp reference available in GenBank, which was obtained from a grapevine sampled in Switzerland.

DISCUSSION

Enrichment of phytoplasma DNA from naturally infected host plants would provide a significant improvement for whole-genome sequencing, as phytoplasmas cannot be cultured on media. This study focussed on a grapevine sample infected with the FDp M54 *map* genotype, which is efficiently transmitted by *S. titanus* and is the main culprit for epidemics in Slovenian vineyards [51]. The aim of this study was to establish an optimal, rapid and efficient protocol for sequencing the FDp genome directly from infected grapevine tissue. It is known that grapevine samples pose a challenge for DNA extraction due to their high content of polyphenols, polysaccharides and other secondary metabolites. These can co-purify with the DNA and interfere with spectrophotometric purity measurements [61, 62], which probably explains the suboptimal A260/280 and A260/230 ratios observed in our study.

Comparison of two protocols without phytoplasma enrichment (A and B) resulted in a similar percentage of mapped reads to the FDp reference genome and had similar performance in the *de novo* assembly of the FDp genome. From a practical point of view, the KingFisher extraction included in protocol A is faster and can be easily scaled up in a semi-automated way [49]. It has been shown to be efficient in processing of grapevine samples and comparable to the CTAB extraction method in downstream analyses such as PCR and real-time PCR [55]. However, due to the lower total DNA yield obtained with the KingFisher extraction, we used the CTAB method for downstream enrichment protocols and additional comparisons (C–F). We did not test how the enrichment steps would affect the DNA extracted with KingFisher. In future applications, incorporating an RNase treatment during the DNA extraction phase could help improve DNA purity and increase quality scores during library preparation. In our case, DNA quantification for comparisons and later downstream applications was based on the fluorometric Qubit measurements, which specifically measure dsDNA and are not affected by the presence of RNA.

Differential centrifugation is a simple and effective method for the enrichment of phytoplasma DNA, addressing the challenge of low phytoplasma concentrations in grapevine samples. As previously reported by Palmano [63], this method yields higher amounts of phytoplasma DNA compared to CTAB extraction alone – a finding initially based on PCR analyses and which we confirmed using HTS, allowing the analysis of larger genome segments. Differential centrifugation (protocol C) enhanced genome assembly by increasing the average coverage of reads mapping to the FDp reference genome and improving genome coverage compared to CTAB extraction alone (protocol B). This simple enrichment step, which does not require specialized equipment, also improved *de novo* assembly, enabling contigs to cover a substantial portion of the FD genome even at moderate dataset sizes.

By adding the post-extraction enrichment step, genome coverage and average depth were further improved. In this step, we employed the NebNext Microbiome Enrichment Kit to separate prokaryotic and eukaryotic DNA using the methyl-CpG binding domain protein [64]. This method selectively depletes plant host DNA, which naturally undergoes CpG methylation, while the non-methylated DNA of phytoplasmas is retained. The approach has been successfully applied to previous phytoplasma genome assemblies [17, 25, 65, 66] and has proven to be a valuable tool for obtaining genomes from natural hosts as well. We included it in protocol D, which demonstrated the highest efficiency in capturing the FDp genome with reads mapping, achieving over 90% coverage even at the smallest dataset size. *De novo* assembly using datasets derived from this protocol outperformed all other protocols, producing the most complete and contiguous reconstruction of the reference genome. Protocol D yielded the highest

assembly quality, with minimal errors and the greatest contiguity, highlighting the robustness of short-read Illumina sequencing for high-accuracy assemblies.

To determine whether the two enrichment steps perform consistently across samples, we included additional samples of grapevine and hazelnut infected with 16SrV phytoplasma. Due to resource constraints, these samples were not sequenced; however, we evaluated enrichment efficiency using qPCR assays targeting host DNA (18S rRNA) and phytoplasma DNA (16SrV-specific assay). Both enrichment steps proved to be effective compared to the simplest protocol (protocol A) and resulted in a consistent reduction of host DNA in all samples. Importantly, the enrichment strategy was successful not only in grapevine, but also in hazelnut and even in another tissue matrix (root phloem), suggesting that the method is broadly applicable across hosts and sample types. These results support the robustness and transferability of the enrichment protocol.

To complete the genomes of isolated phytoplasmas, long reads from nanopore sequencing can be used. However, the amount of DNA obtained after using the NebNext Microbiome Enrichment Kit was too low to successfully prepare a library for nanopore sequencing. To solve this problem, we used WGA, which employs multiple displacement amplification. The isothermal amplification reaction has been reported to produce DNA with a low bias in sequence representation and achieve greater genome coverage compared to PCR-based WGA methods [58]. The sample enriched with differential centrifugation, isolated with the CTAB method, further enriched with the NebNext Microbiome Enrichment Kit and then amplified with the REPLIg-midi WGA kit was sequenced using both platforms – Illumina and ONT. The WGA did not significantly alter the end result of the FDp reference genome coverage and enabled the acquisition of high-quality genomic data from limited amounts of DNA, while ensuring compatibility with long-read sequencing technologies that require higher input. The sequencing approach that followed WGA did not affect the percentage of reads that covered the FDp reference genome. Both methods achieved complete coverage of the FDp reference genome with an average depth of more than tenfold, demonstrating their effectiveness for phytoplasma genome analysis.

Nanopore sequencing used in protocol F has demonstrated its strength in generating much longer reads – the longest read that could be mapped to the reference genome was 95,584 bp long. Consequently, assemblers can output longer contigs, with the largest alignments surpassing those from protocol D at higher dataset sizes, highlighting the potential of long-read sequencing for resolving repetitive and structurally complex regions. However, when performing rarefaction analyses for *de novo* assemblies of smaller datasets where genome coverage was minimal and the assemblies contained more mismatches and misassemblies, its limitations became apparent. These results emphasize the need for larger nanopore sequencing datasets to achieve more reliable assemblies, as the lower base-level accuracy of nanopore reads compared to Illumina sequencing remains a challenge.

The hybrid *de novo* assembly approach combines the strengths of Illumina and nanopore sequencing, making it a compelling choice for phytoplasma genome reconstruction. This integration resulted in improved genome coverage and accuracy with fewer mismatches and misassemblies compared to assemblies performed with nanopore reads only (protocol F). In terms of assembly contiguity, the hybrid mode outperformed assemblies performed with Illumina reads only (protocol D) and assembled longer phytoplasma contigs with higher NGA50 at the largest dataset size. However, in hybrid mode, careful post-assembly validation is required to ensure the accuracy of the assembly. This was the case when some chimeric sequences of phytoplasmas and grapevine were detected by an additional BLAST analysis. To take advantage of both Illumina and nanopore sequencing, we used the hybrid mode to obtain the whole genome of the Slovenian isolate of FDp. Using the entire datasets obtained with protocols D and F in the hybrid assembly of short Illumina and long nanopore sequencing reads enabled the assembly of an almost complete FDp reference genome (96% coverage) with only 17 contigs. Compared to other phytoplasma genome drafts [67], the number of contigs is low, and the N50 value is high, indicating a high quality of assembly. In the future, different assemblers for long and short reads and their hybrid mode could also be benchmarked to further improve the assembly.

Although hybrid assembly offers advantages in terms of the completeness and integrity of the genome, its feasibility depends not only on technological considerations but also on the available financial resources. As sequencing strategies vary greatly in cost, it is important to evaluate the cost-effectiveness of sample preparation and sequencing to determine the most practical approach for a particular study (Material S11). Sample preparation cost can be divided into three main categories: (1) the cost of sample preparation chemicals, (2) the cost of library preparation and sequencing and (3) labour cost. The addition of differential centrifugation increases the costs, but mainly due to an extra hour of labour, not because of notable additional expenses for chemicals. Enrichment with the NebNext Microbiome Enrichment Kit and WGA incurs additional costs and additional time to the overall procedure. For the protocols tested, the total cost without sequencing ranged from just over €100 to nearly €900 per sample, and the hands-on time ranged from 2 to 9 h, depending on the complexity of the steps included (Material S9). However, in our case, the biggest cost differences arose at the sequencing phase. For Illumina sequencing, we outsourced library preparation and sequencing to a commercial provider using a high-throughput system such as NovaSeq. This approach offers a competitive cost of around €100 per sample, especially when sequencing is bundled into large runs, but turnaround times depend on the vendor (we typically obtained Illumina sequencing results within 2 weeks). In our case, in-house nanopore sequencing involved higher upfront costs of around €800 for library preparation and sequencing and required a full day of hands-on work but had the advantage of much faster results (after 3 days). Ultimately, the decision between Illumina and nanopore sequencing depends on factors such as the specific goals of the study, access to in-house equipment, availability of commercial vendors and the timeline

of the project. Although the hybrid assembly provides the highest quality, our results show that protocol D alone can provide very high-quality assemblies, making it a sufficient choice considering cost, time and bioinformatics resources.

New approaches to sequencing phytoplasma genomes provide valuable insights. While earlier genome projects using Sanger sequencing faced significant obstacles due to their labour-intensive and time-consuming nature [19], recent advances in HTS technologies have made it possible to generate draft sequences or even complete phytoplasma genomes with greater efficiency [17, 24, 29, 34, 35]. Despite the drawbacks of draft genomes, phytoplasma sequence drafts have a significant impact as they expand our understanding by revealing previously undiscovered genetic information in phytoplasma strains and providing new information for the design of new primer sets for use in population genetics and diagnostics. In our study, the best protocol for whole-genome sequencing of FDp included CTAB extraction, pre-extraction enrichment (differential centrifugation), post-extraction enrichment (NEbNext Microbiome Enrichment kit) and Illumina sequencing (protocol D). The data obtained with this protocol were also effectively utilized in hybrid assembly with nanopore sequencing reads and led to a draft genome assembly of the Slovenian FDp isolate. This study shows that the optimized enrichment protocols can facilitate genome sequencing of naturally infected phytoplasma hosts, not only in grapevine but also in other woody plants with similarly low phytoplasma concentrations, as shown in hazelnut samples infected with 16SrV phytoplasma. To date, research of FDp epidemiology has been based on Sanger sequencing of specific genes, especially the *map* gene [51, 60, 68–70]. However, our protocol has great potential to enable whole-genome sequencing of a larger number of different isolates of quarantine FDp, which will contribute to a better understanding of the epidemiology and consequently to a more effective implementation of measures to prevent the spread of this dangerous disease.

Funding information
The study was financially supported by the Slovenian Research Agency through the project NanoPhyto (L7-2632), core financing (P4-0165) and young researcher grant (Z.K.Z.).

Acknowledgements
We thank the Slovenian phytosanitary inspector, expert surveyor, Ana Vučurović and Irena Bajde for providing the field samples and Miha Kitek, Irena Bajde and Tjaša Jakomin for help in the laboratory. We also thank Maja Ravnikar, Martina Šeruga Musić, Xavier Foissac and Sylvie Malembic-Maher for productive discussions.

Author contributions
Z.K.Z. participated in the design of the study, laboratory experiments, bioinformatic analyses, data curation, visualization and writing of the manuscript. Supervision and review were done by N.M. and D.K. Both were included in conceptualization and study design, formal analyses and funding acquisition and coordinated the study. All authors contributed to the article and approved the submitted version.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. **Namba S**. Molecular and biological properties of phytoplasmas. *Proc Jpn Acad Ser B Phys Biol Sci* 2019;95:401–418.
2. **Bertaccini A, Duduk B, Paltrinieri S, Contaldo N**. Hytoplasmas and phytoplasma diseases: a severe threat to agriculture. *AJPS* 2014;05:1763–1788.
3. **The IRPCM Phytoplasma/Spiroplasma Working Team – Phytoplasma taxonomy group**. '*Candidatus Phytoplasma*', a taxon for the wall-less, non-helical prokaryotes that colonize plant phloem and insects. *Int J Syst Evol Microbiol* 2004;54:1243–1255.
4. **Bertaccini A, Arocha-Rosete Y, Contaldo N, Duduk B, Fiore N, *et al.*** Revision of the '*Candidatus Phytoplasma*' species description guidelines. *Int J* 2022;72.
5. **Bertaccini A, Lee IM**. Phytoplasmas: an update. In: **Rao GP, Bertaccini A, Fiore N** and **Liefting LW** (eds). *Phytoplasmas: Plant Pathogenic Bacteria - I: Characterisation and Epidemiology of Phytoplasma-Associated Diseases*. Springer Singapore; 2018. pp. 1–29.
6. **Weintraub PG, Beanland L**. Insect vectors of phytoplasmas. *Annu Rev Entomol* 2006;51:91–111.
7. **Rao GP, Bertaccini A, Fiore N, Liefting LW**. *Phytoplasmas: Plant Pathogenic Bacteria - I*, 1st Edn. Singapore: Singapore; 2018, p. 345.
8. **Kube M, Mitrovic J, Duduk B, Rabus R, Seemüller E**. Current view on phytoplasma genomes and encoded metabolism. *ScientificWorldJournal* 2012;2012:185942.
9. **Tran-Nguyen LTT, Gibb KS**. Optimizing *Phytoplasma* DNA purification for genome analysis. *J Biomol Tech* 2007;18:104–112.
10. **Bertaccini A, Duduk B**. *Phytoplasma* and phytoplasma diseases: a review of recent research. *Phytopathol Mediterr* 2009:355–378.
11. **Prezelj N, Nikolić P, Gruden K, Ravnikar M, Dermastia M**. Spatiotemporal distribution of *Flavescence dorée* phytoplasma in grapevine. *Plant Pathology* 2013;62:760–766.
12. **Siddique ABM, Guthrie JN, Walsh KB, White DT, Scott PT**. Histopathology and within-plant distribution of the phytoplasma associated with Australian papaya dieback. *Plant Dis* 1998;82:1112–1120.
13. **Saracco P, Bosco D, Veratti F, Marzachì C**. Quantification over time of *Chrysanthemum yellows* phytoplasma (16Sr-I) in leaves and roots of the host plant *Chrysanthemum carinatum* (Schousboe) following inoculation with its insect vector. *Physiol Mol Plant Pathol* 2005;67:212–219.
14. **Oyola SO, Otto TD, Gu Y, Maslen G, Manske M, *et al.*** Optimizing Illumina next-generation sequencing library preparation for extremely AT-biased genomes. *BMC Genomics* 2012;13:1.
15. **Tran-Nguyen LTT, Kube M, Schneider B, Reinhardt R, Gibb KS**. Comparative genome analysis of "*Candidatus Phytoplasma* australiense" (subgroup tuf-Australia I; rp-A) and "Ca. *Phytoplasma asteris*" strains OY-M and AY-WB. *J Bacteriol* 2008;190:3979–3991.
16. **Kube M, Schneider B, Kuhl H, Dandekar T, Heitmann K, *et al.*** The linear chromosome of the plant-pathogenic mycoplasma "*Candidatus Phytoplasma mali*". *BMC Genomics* 2008;9:1–14.
17. **Debonneville C, Mandelli L, Brodard J, Groux R, Roquis D, *et al.*** The complete genome of the "*Flavescence dorée*" phytoplasma reveals characteristics of low genome plasticity. *Biology* 2022;11.

18. **Marcone C**, **Hergenhahn F**, **Ragozzino A**, **Seemüller E**. Dodder transmission of pear decline, European stone fruit yellows, rubus stunt, *Picris echioides* yellows and cotton phyllody phytoplasmas to periwinkle. *J Phytopathol* 1999;147:187–192.
19. **Oshima K**, **Kakizawa S**, **Nishigawa H**, **Jung H-Y**, **Wei W**, ***et al.*** Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. *Nat Genet* 2004;36:27–29.
20. **Bai X**, **Zhang J**, **Ewing A**, **Miller SA**, **Jancso Radek A**, ***et al.*** Living with genome instability: the adaptation of phytoplasmas to diverse environments of their insect and plant hosts. *J Bacteriol* 2006;188:3682–3696.
21. **Wang J**, **Song L**, **Jiao Q**, **Yang S**, **Gao R**, ***et al.*** Comparative genome analysis of jujube witches'-broom phytoplasma, an obligate pathogen that causes jujube witches'-broom disease. *BMC Genomics* 2018;19:689.
22. **Andersen MT**, **Liefting LW**, **Havukkala I**, **Beever RE**. Comparison of the complete genome sequence of two closely related isolates of "*Candidatus Phytoplasma australiense*" reveals genome plasticity. *BMC Genomics* 2013;14:1–15.
23. **Rodrigues Jardim B**, **Tran-Nguyen LTT**, **Gambley C**, **Rodoni B**, **Constable FE**. Iodixanol density gradients as an effective phytoplasma enrichment approach to improve genome sequencing. *Front Microbiol* 2022;13:937648.
24. **Tan CM**, **Lin Y-C**, **Li J-R**, **Chien Y-Y**, **Wang C-J**, ***et al.*** Accelerating complete phytoplasma genome assembly by immunoprecipitation-based enrichment and MinION-based DNA sequencing for comparative analyses. *Front Microbiol* 2021;12.
25. **Nijo T**, **Iwabuchi N**, **Tokuda R**, **Suzuki T**, **Matsumoto O**, ***et al.*** Enrichment of phytoplasma genome DNA through a methyl-CpG binding domain-mediated method for efficient genome sequencing. *J Gen Plant Pathol* 2021;87:154–163.
26. **Mitrović J**, **Siewert C**, **Duduk B**, **Hecht J**, **Mölling K**, ***et al.*** Generation and analysis of draft sequences of "stolbur" phytoplasma from multiple displacement amplification templates. *J Mol Microbiol Biotechnol* 2014;24:1–11.
27. **Orlovskis Z**, **Canale MC**, **Haryono M**, **Lopes JRS**, **Kuo CH**, ***et al.*** A few sequence polymorphisms among isolates of Maize bushy stunt phytoplasma associate with organ proliferation symptoms of infected maize plants. *Ann Bot* 2017;119:869–884.
28. **Cho S-T**, **Zwolińska A**, **Huang W**, **Wouters RHM**, **Mugford ST**, ***et al.*** Complete genome sequence of "*Candidatus Phytoplasma asteris*" RP166, a plant pathogen associated with rapeseed phyllody disease in Poland . *Microbiol Resour Announc* 2020;9:1–2.
29. **Huang C-T**, **Cho S-T**, **Lin Y-C**, **Tan C-M**, **Chiu Y-C**, ***et al.*** Comparative genome analysis of "*Candidatus Phytoplasma luffae*" reveals the influential roles of potential mobile units in phytoplasma evolution. *Front Microbiol* 2022;13:773608.
30. **Duckeck D**, **Zübert C**, **Böhm JW**, **Carminati G**, **Schneider B**, ***et al.*** Complete genome of "*Candidatus Phytoplasma rubi*" RS, a phytopathogenic bacterium associated with Rubus stunt disease . *Microbiol Resour Announc* 2023;12.
31. **Cao Y**, **Sun G**, **Zhai X**, **Xu P**, **Ma L**, ***et al.*** Genomic insights into the fast growth of paulownias and the formation of Paulownia witches' broom. *Molecular Plant* 2021;14:1668–1682.
32. **Böhm JW**, **Huettel B**, **Schneider B**, **Kube M**. The complete genome of "*Candidatus Phytoplasma fraxini*" AshY1 from the ash yellows group. *Microbiol Resour Announc* 2024;13:e0031824.
33. **Toth R**, **Huettel B**, **Eini O**, **Varrelmann M**, **Kube M**. The complete genome sequence of the stolbur pathogen "*Candidatus Phytoplasma solani*" from *Pentastiridius leporinus* . *Microbiol Resour Announc* 2025;14.
34. **Zhang R-Y**, **Wang X-Y**, **Li J**, **Shan H-L**, **Li Y-H**, ***et al.*** Complete genome sequence of "*Candidatus Phytoplasma sacchari*" obtained using a filter-based DNA enrichment method and Nanopore sequencing. *Front Microbiol* 2023;14:1252709.
35. **Rodrigues Jardim B**, **Tran-Nguyen LTT**, **Gambley C**, **Al-Sadi AM**, **Al-Subhi AM**, ***et al.*** The observation of taxonomic boundaries for the 16SrII and 16SrXXV phytoplasmas using genome-based delimitation. *Int J Syst Evol Microbiol* 2023;73.
36. **Cho ST**, **Chen AP**, **Chou SJ**, **Yan XH**, **Lin CP**, ***et al.*** Complete genome sequence of "*Candidatus Phytoplasma cynodontis*" GY2015, a plant pathogen associated with Bermuda grass white leaf disease in Taiwan. *Microbiol Resour Announc* 2023;12:e0045723.
37. **Li Y**, **Yan X-H**, **Liu Y**, **Pei S-C**, **Kuo C-H**, ***et al.*** Complete genome sequence of "*Candidatus Phytoplasma aurantifolia*" TB2022, a plant pathogen associated with sweet potato little leaf disease in China. *Microbiol Resour Announc* 2023;12:e0030823.
38. **Yan X-H**, **Lin J**, **Liu Y**, **Huang P**, **Liu J**, ***et al.*** Complete genome sequence of "*Candidatus Phytoplasma asteris*" QS2022, a plant pathogen associated with lettuce chlorotic leaf rot disease in China. *Microbiol Resour Announc* 2023;12:e0030623.
39. **Pei SC**, **Chen AP**, **Chou SJ**, **Hung TH**, **Kuo CH**. Complete genome sequence of "*Candidatus Phytoplasma pruni*" PR2021, an uncultivated bacterium associated with poinsettia (Euphorbia pulcherrima). *Microbiol Resour Announc* 2023;12:e0044323.
40. **Toth R**, **Ilic AM**, **Huettel B**, **Duduk B**, **Kube M**. Divergence within the taxon "*Candidatus Phytoplasma asteris*" confirmed by comparative genome analysis of carrot strains. *Microorganisms* 2024;12:1016.
41. **Xue C**, **Zhang Y**, **Li H**, **Liu Z**, **Gao W**, ***et al.*** The genome of *Candidatus Phytoplasma ziziphi* provides insights into their biological characteristics. *BMC Plant Biol* 2023;23:251.
42. **Li N-P**, **Yan X-H**, **Pei S-C**, **Hung T-H**, **Chang T-C**, ***et al.*** Complete genome sequence of *Candidatus Phytoplasma australasiaticum* WF_GM2021, a plant pathogen associated with soybean witches' broom disease in Taiwan. *Microbiol Resour Announc* 2024;13.
43. **De Maio N**, **Shaw LP**, **Hubbard A**, **George S**, **Sanderson ND**, ***et al.*** Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microb Genom* 2019;5:e000294.
44. **Kono N**, **Arakawa K**. Nanopore sequencing: review of potential applications in functional genomics. *Dev Growth Differ* 2019;61:316–326.
45. **Davis RE**, **Dally EL**. Revised subgroup classification of group 16SrV phytoplasmas and placement of flavescence dorée-associated phytoplasmas in two distinct subgroups. *Plant Dis* 2001;85:790–797.
46. **Tramontini S**, **Delbianco A**, **Vos S**, **European Food Safety Authority (EFSA)**. Pest survey card on flavescence dorée phytoplasma and its vector *Scaphoideus titanus*. *EFS3* 2020;17.
47. **Caudwell A**. Deux années d'étude sur la flavescence dorée, nouvelle maladie grave de la vigne. *Annales D'amélioration Des Plantes* 1957;4:359–363.
48. **Chuche J**, **Thiéry D**. Biology and ecology of the Flavescence dorée vector Scaphoideus titanus: a review. *Agron Sustain Dev* 2014;34:381–403.
49. **Mehle N**, **Nikolić P**, **Rupar M**, **Boben J**, **Ravnikar M**, ***et al.*** Automated DNA extraction for large numbers of plant samples. In: **Dickinson M** and **Hodgetts J** (eds). *Phytoplasma Methods in Molecular Biology (Methods and Protocols)*. Totowa, NJ: Humana Press; 2013. pp. 139–145.
50. **Arnaud G**, **Malembic-Maher S**, **Salar P**, **Bonnet P**, **Maixner M**, ***et al.*** Multilocus sequence typing confirms the close genetic interrelatedness of three distinct flavescence dorée phytoplasma strain clusters and group 16SrV phytoplasmas infecting grapevine and alder in Europe. *Appl Environ Microbiol* 2007;73:4001–4010.
51. **Kogej Zwitter Z**, **Seljak G**, **Jakomin T**, **Brodarič J**, **Vučurović A**, ***et al.*** Epidemiology of flavescence dorée and hazelnut decline in Slovenia: geographical distribution and genetic diversity of the associated 16SrV phytoplasmas. *Front Plant Sci* 2023;14:1217425.
52. **Rogers SO**, **Bendich A**. Extraction of DNA from plant tissues. In: **Gelvin SB**, **Schilpoort RA** and **Verma DPS** (eds). *Plant Molecular Biology Manual*. Dordrecht: Kluwer Academic Publishers; 1998.
53. **Saponari M**, **Loconsole G**, **Potere O**, **Palmisano F**, **Boscia D**, ***et al.*** Current protocols for the detection of *Xylella fastidiosa* in host plants and vectors. 2014:12–13 .

54. **Hren M**, **Boben J**, **Rotter A**, **Kralj P**, **Gruden K**, *et al.* Real-time PCR detection systems for flavescence dorée and bois noir phytoplasmas in grapevine: comparison with conventional PCR detection and application in diagnostics. *Plant Pathology* 2007;56:785–796.
55. **Mehle N**, **Prezelj N**, **Hren M**, **Boben J**, **Gruden K**, *et al.* A real-time PCR detection system for the bois noir and flavescence dorée phytoplasmas and quantification of the target DNA. In: **Dickinson M** and **Hodgetts J** (eds). *Phytoplasma: Methods and Protocols, Methods in Molecular Biology*. New York: Humana Press; 2013. pp. 269–281.
56. **Prjibelski A**, **Antipov D**, **Meleshko D**, **Lapidus A**, **Korobeynikov A**. Using SPAdes de novo assembler. *Curr Protoc Bioinformatics* 2020;70:e102.
57. **Kolmogorov M**, **Yuan J**, **Lin Y**, **Pevzner PavelA**. Assembly of long error-prone reads using repeat graphs. *Bioinformatics* 2018;Available from. DOI: 10.1101/247148.
58. **Mikheenko A**, **Prjibelski A**, **Saveliev V**, **Antipov D**, **Gurevich A**. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics 2018;34:i142–i150.
59. **Jain C**, **Rodriguez-R LM**, **Phillippy AM**, **Konstantinidis KT**, **Aluru S**. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;9:5114.
60. **Malembic-Maher S**, **Desqué D**, **Khalil D**, **Salar P**, **Bergey B**, *et al.* When a Palearctic bacterium meets a Nearctic insect vector: genetic and ecological insights into the emergence of the grapevine flavescence dorée epidemics in Europe. *PLoS Pathog* 2020;16:e1007967.
61. **Öncü Öner T**, **Temel M**, **Pamay S**, **Abaci AK**, **Kaya Akkale HB**. An Improved Method for Efficient DNA Extraction from Grapevine. *Int J Life Sci Biotechnol* 2023;6:21–36.
62. **Rezaian MA**, **Krake LR**. Nucleic acid extraction and virus detection in grapevine. *J Virol Methods* 1987;17:277–285.
63. **Palmano S**. A comparison of different phytoplasma DNA extraction methods using competitive PCR. *Phytopathol Mediterr* 2001:99–107.
64. **Feehery GR**, **Yigit E**, **Oyola SO**, **Langhorst BW**, **Schmidt VT**, *et al.* A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. *PLoS One* 2013;8:e76096.
65. **Town JR**, **Wist T**, **Perez-Lopez E**, **Olivier CY**, **Dumonceaux TJ**. Genome sequence of a plant-pathogenic bacterium, "*Candidatus Phytoplasma asteris*" strain TW1. *Microbiol Resour Announc* 2018;7.
66. **Cai W**, **Shao J**, **Zhao Y**, **Davis RE**, **Costanzo S**. Draft genome sequence of '*Candidatus Phytoplasma pini*'-related strain MDPP: a resource for comparative genomics of gymnosperm-infecting phytoplasmas . *Plant Disease* 2020;104:1009–1010.
67. **Kube M**, **Duduk B**, **Oshima K**. Genome sequencing. In: **Bertaccini A**, **Kube M**, **Oshima K** and **Rao GP** (eds). *Phytoplasmas: Plant Pathogenic Bacteria. First*. Singapore: Springer; 2019. pp. 11–26.
68. **Oggier A**, **Conedera M**, **Debonneville C**, **Schumpp O**, **Rizzoli A**. Gone-wild grapevines in forests host phytoplasma genotypes linked to grapevine's flavescence dorée epidemics in cultivated vineyards and competent vectors. *J Plant Pathol* 2024;106:1537–1548.
69. **Rossi M**, **Pegoraro M**, **Ripamonti M**, **Abbà S**, **Beal D**, *et al.* Genetic diversity of flavescence dorée phytoplasmas at the vineyard scale. *Appl Environ Microbiol* 2019;85.
70. **Plavec J**, **Budinšćak Ž**, **Križanac I**, **Škorić D**, **Foissac X**, *et al.* Multilocus sequence typing reveals the presence of three distinct flavescence dorée phytoplasma genetic clusters in Croatian vineyards. *Plant Pathology* 2019;68:18–30.

The Microbiology Society is a membership charity and not-for-profit publisher.

Your submissions to our titles support the community – ensuring that we continue to provide events, grants and professional development for microbiologists at all career stages.

Find out more and submit your article at microbiologyresearch.org

2.5 Genome-Wide Comparisons Confirm Locally Shared Flavescence Dorée-Related Phytoplasma Populations Across Grapevine, Hazelnut, and Leafhoppers

Authors: **Zala Kogej Zwitter**, Nataša Mehle, Denis Kutnjak

Submitted to: *Frontiers in Microbiology* (May 2026)

frontiers

Genome-wide comparisons confirm locally shared Flavescence dorée-related phytoplasma populations across grapevine, hazelnut, and leafhoppers

Zala Kogej Zwitter[1,2*], Nataša Mehle[1,3*], Denis Kutnjak[1]

[1] Department of Biotechnology and Systems Biology, National Institute of Biology, Ljubljana, Slovenia

[2] Jožef Stefan International Postgraduate School, Ljubljana, Slovenia

[3] School for Viticulture and Enology, University of Nova Gorica, Nova Gorica, Slovenia

*** Correspondence:**
Zala Kogej Zwitter and Nataša Mehle
zala.kogej.zwitter@nib.si; natasa.mehle@nib.si

Keywords: phytoplasma, Flavescence dorée, hazelnut, genomics, network

Abstract

Flavescence dorée-related phytoplasmas from the 16SrV group are major plant pathogens in European agroecosystems, affecting grapevine and occurring in a range of alternative hosts and insect vectors. Although molecular epidemiology has largely relied on single-locus markers, such as the methionine aminopeptidase (*map*) gene, it remains unclear how well these reflect genome-wide relationships and host-associated genomic patterns. In this study, we applied a comparative genomics framework to Flavescence dorée-related phytoplasma isolates from grapevine, hazelnut, and two leafhoppers (*Scaphoideus titanus* and *Orientus ishidae*) in Slovenia, focusing on *map* genotypes M38 and M50 detected across hosts. Draft genomes were generated from Illumina sequencing of phytoplasma-enriched DNA and analysed using core gene phylogeny and genome-wide similarity metrics. Across analyses, genome-wide relationships corresponded to *map* genotype classification for M38 and M50. In contrast, no consistent clustering or genomic differentiation by host species was observed, whereas geographic origin contributed to finer-scale structuring. Network analyses revealed reticulate patterns, indicating conflicting phylogenetic signals, although their biological interpretation remains uncertain. Overall, these results show that *map* classification reflects genome-wide similarity, while host-associated differentiation is not supported. The findings support a multi-host system in which related phytoplasmas occur across different hosts and highlight genome-scale approaches for refining Flavescence dorée epidemiology.

1 Introduction

Phytoplasmas are plant-pathogenic, wall-less bacteria that colonise phloem sieve elements in plants and establish persistent infections in phloem-feeding insect vectors (IRPCM, 2004; Hogenhout et al., 2008). They depend on both plant and insect hosts. Their efficient spread is driven by phloem-feeding insect vectors, while infected vegetatively propagated planting material can additionally contribute to dissemination (Wang et al., 2024). In perennial woody hosts, infections

often cause progressive decline of affected plants and can lead to substantial economic losses (Wang et al., 2024).

Because phytoplasmas have not yet been successfully cultured in axenic media, our current understanding of their genetic diversity, phylogenetic relationships, and taxonomic classification relies largely on molecular markers and, more recently, genome sequencing (Kube et al., 2012; Wei and Zhao, 2022). The most widely used marker is the 16S rRNA gene, and the resulting 16Sr group/subgroup framework remains the standard for detection and broad classification, although it has limited resolution for distinguishing epidemiologically relevant lineages within major groups (Wei and Zhao, 2022). For outbreak tracing and inference of strain movement across hosts and landscapes, higher-resolution markers and multilocus or genome-scale approaches are therefore needed to complement 16Sr assignment (Arnaud et al., 2007; Wei and Zhao, 2022).

Within this context, Flavescence dorée (FD) represents one of the best-studied epidemic phytoplasma diseases in Europe (Malembic-Maher et al., 2020). The associated FD phytoplasma (FDp) is primarily linked to 16SrV-C and 16SrV-D lineages and causes a major yellows disease of the cultivated grapevine (*Vitis vinifera*) (Chuche and Thiéry, 2014; Malembic-Maher et al., 2020). FD epidemics in European vineyards are tightly connected to the presence and ecology of the invasive leafhopper *Scaphoideus titanus* Ball (Hemiptera: Cicadellidae), which enables efficient within-vineyard spread and contributes to regional epidemic dynamics (Chuche and Thiéry, 2014; Malembic-Maher et al., 2020; Krstić et al., 2022). A key insight into FDp emergence is that epidemic dynamics reflect an interaction between an introduced North American vector, *S. titanus*, and a pre-existing European reservoir of genetically diverse 16SrV phytoplasmas, thereby suggesting that surrounding vegetation and environmental reservoirs can strongly influence outbreak risk and strain composition (Malembic-Maher et al., 2020).

Studies have identified European alders (*Alnus* spp.) as important natural hosts harbouring an extensive diversity of FDp-related 16SrV-C genotypes, whereas only a limited subset of this diversity is repeatedly associated with grapevine epidemics (Arnaud et al., 2007; Malembic-Maher et al., 2020; Marcone et al., 2024). Beyond alders, FD-related 16SrV-C phytoplasmas have been reported in several plant species occurring near vineyards, including *Clematis vitalba* (Angelini et al., 2004; Filippin et al., 2009; Casati et al., 2017; Plavec et al., 2019; Rossi et al., 2019), *Spartium junceum* (Rizza et al., 2021), *Salix* spp. (Casati et al., 2017), and hazelnut (*Corylus avellana*) (Casati et al., 2017; Mehle et al., 2019); FDp-related lineages have also been detected in the invasive tree *Ailanthus altissima* (Plavec et al., 2019).

To characterise FDp diversity and distribution, molecular epidemiological studies have primarily relied on sequence variation in the methionine aminopeptidase (*map*) gene, which provides sufficient resolution to distinguish closely related strains and define *map* genotypes (Arnaud et al., 2007). More than 170 *map* genotypes have been identified across Europe in diverse plant hosts (Marcone et al., 2024). However, only a subset of these genotypes are consistently associated with symptomatic grapevine outbreaks and are considered epidemic strains in vineyards (Arnaud et al., 2007; Radonjić et al., 2013; Atanasova et al., 2014; Krstić et al., 2018, 2022; Plavec et al., 2019; Malembic-Maher et al., 2020). Epidemic FDp strains are grouped into three major *map*-based clusters (Map-FD1, FD2 and FD3) (Arnaud et al., 2007), within which genotypes differ in host associations and involvement in vineyard outbreaks. Strains outside these clusters are less likely to sustain *S. titanus*-driven epidemics (Malembic-Maher et al., 2020). At the vineyard scale, multiple *map* genotypes can co-circulate, and the co-occurrence of related genotypes in grapevine and surrounding vegetation suggests exchanges between cultivated and non-cultivated host compartments (Rossi et al., 2019).

While *map*-based classification provides a framework for describing FDp genetic diversity, additional markers and multilocus approaches have been used to further resolve relationships and trace transmission pathways (Angelini et al., 2004; Rossi et al., 2019; Malembic-Maher et al., 2020).

Slovenia provides a useful setting for studying FD-related 16SrV phytoplasmas, as infections have been documented in grapevine and in several other hosts (Mehle et al., 2011, 2019). FDp outbreaks in Slovenian vineyards have been characterised by *map* genotyping, with genotype M54 reported as dominant in grapevine, while several other epidemic genotypes, including M50, M51, and M38, have been detected at lower frequencies (Kogej Zwitter et al., 2023). Notably, 16SrV phytoplasmas in Slovenia have also been detected in the leafhopper *Orientus ishidae*, indicating circulation of FDp-related lineages beyond the specialist vineyard vector *S. titanus* (Mehle et al., 2010). In Slovenia, *O. ishidae* was abundant in hazelnut orchards with symptomatic shrubs infected with 16SrV phytoplasmas (Kogej Zwitter et al., 2023). In hazelnut, eight different *map* genotypes were detected, four of which were identified for the first time. Three genotypes (M38, M50, and M122) were detected in all the different analysed hosts – grapevine, hazelnut, and *O. ishidae* - which suggests possible movement between hosts (Kogej Zwitter et al., 2023).

Although the *map* gene and other single-locus markers provide a practical basis for surveillance and outbreak investigation, they offer only limited information on genomic similarity, and it has not yet been tested how well they reflect genome-wide evolutionary history. Addressing this gap would enable a more precise understanding of possible FDp host-associated adaptation and/or geographic structuring driven by local transmission and diversification. Genome-wide comparisons are limited by the technical challenges of phytoplasma genomics, including the inability to isolate clonal phytoplasma in culture, low pathogen titre, and overwhelming host DNA, which can hinder sequencing depth, assembly continuity, and downstream comparative analyses (Kogej Zwitter et al., 2025). Genome-scale approaches, particularly when combined with effective enrichment strategies and reproducible comparative frameworks, therefore offer an opportunity to link marker-based molecular epidemiology with higher-resolution genomic insights.

To address the above-mentioned questions, we applied a comparative genomics framework to FD-related 16SrV phytoplasmas from grapevine, hazelnut, and leafhoppers (*S. titanus* and *O. ishidae*) in Slovenia, focusing on *map* genotypes M38 and M50 detected across hosts (Kogej Zwitter et al., 2023). We tested whether genome-wide similarity patterns correspond to *map* genotype classification and whether isolates sharing the same *map* genotype exhibit host-associated genomic differentiation. In addition, we assessed whether finer-scale variation in genome-wide similarity is associated with geographic origin.

2 Material and methods

2.1 Sample selection

We selected samples for this study from a collection of plant and insect material previously confirmed to be infected with 16SrV phytoplasma and characterised for *map* genotype, as described in Kogej Zwitter et al. (2023). We specifically chose symptomatic grapevine and hazelnut plants and leafhopper samples (*S. titanus* and *O. ishidae*) infected with the two *map* genotypes M38 and M50, because these genotypes had been detected across all host types included in this study. Samples were collected between 2019 and 2024 from four sampling areas in Slovenia (Novo mesto, NM; Slovenska Bistrica, SB; Ptuj, PT; Ljutomer, LT; Figure 1). Detailed metadata, including sampling area, year, host type, sample type, and *map* genotype, are provided in Supplementary Table 1.

Hazelnut samples originated from a single infected orchard in the area of Slovenska Bistrica (SB) (Figure 1) and were collected in previous studies in 2020 and 2021 (Kogej Zwitter et al., 2023). Grapevine samples were obtained as part of the official grapevine yellows monitoring programme conducted under the authority of the Administration of the Republic of Slovenia for Food Safety, Veterinary and Plant Protection, with each sample originating from a different vineyard within the four defined areas (Figure 1). Hazelnut samples and two grapevine samples from 2019 were previously reported in Kogej Zwitter et al. (2023).

Leafhopper samples originated from vineyards and one hazelnut orchard and were collected in previous studies (Kogej Zwitter et al., 2023) and in national monitoring during periods of phytoplasma symptom expression. *S. titanus* specimens were collected in vineyards (NM and PT) as part of the national monitoring programme and stored in 70% ethanol until analysis. *O. ishidae* specimens were collected in the infected hazelnut orchard (SB) and stored at −80 °C. Each *S. titanus* sample consisted of pooled adult individuals, whereas *O. ishidae* samples consisted of single adults (Supplementary Table 1).

DNA used in this study was derived from the selected sample set described above and originated either from previously extracted material or from newly extracted DNA, depending on sample type. For hazelnut samples, DNA was newly extracted from archived plant material stored at −80 °C, following protocol D described in Kogej Zwitter et al. (2025). For grapevine samples, previously extracted DNA from earlier studies (Kogej Zwitter et al., 2023) using the KingFisher (ThermoScientific, USA) and stored at −20 °C was used. For all these samples, we performed post-extraction enrichment using the NEBNext® Microbiome DNA Enrichment Kit (New England Biolabs, MA, USA), following the protocol described in Kogej Zwitter et al. (2025). For insect samples, DNA was obtained from previously extracted DNA samples stored at −20 °C. DNA extraction had been performed using the KingFisher system (ThermoScientific, USA) as described in Kogej Zwitter et al. (2023) for *S. titanus* samples and one *O. ishidae* sample, while one *O. ishidae* sample (O-SB-21-2) was extracted using the OmniPrep™ for Plant kit (G-Biosciences, MO, USA) according to the manufacturer's instructions. We subsequently enriched all insect DNA samples using the NEBNext® Microbiome DNA Enrichment Kit as described above.

2.2 DNA quality control, enrichment efficiency measurements and sequencing

We measured DNA concentrations for each sample using the dsDNA HS kit on a Qubit fluorometer (Thermo Fisher Scientific, MA, USA) and used real-time PCR (qPCR) to verify the presence of 16SrV phytoplasma DNA and host eukaryotic DNA by analysing samples before and after post-extraction enrichment. We tested tenfold dilutions of the extracted or enriched DNA with an 18S rRNA assay (Applied Biosystems, MA, USA) to monitor host DNA and with a specific assay for the 16SrV phytoplasma group to estimate phytoplasma titre (Hren et al., 2007), running each reaction in duplicate. We performed qPCR reactions as described in (Mehle et al., 2013) and included negative extraction controls and negative amplification controls in duplicate in each run. Illumina sequencing was performed by Novogene (Cambridge, UK) using the Plant and Animal Whole Genome sequencing service on Illumina NovaSeq (PE150 mode).

2.3 Initial sequence data processing, *de novo* assembly, and genome validation

We processed raw sequencing reads through quality control (FastQC v0.12) and read trimming (fastp v0.23.4) to remove low-quality bases and sequencing artefacts that could affect downstream assembly and alignment accuracy. Next steps in the pipeline used the trimmed reads for *de novo* assembly with SPAdes v3.15 and for contig validation, sequence similarity searches with Basic Local

Alignment Search Tool v2.14 (BLAST). To sensitively identify phytoplasma-derived contigs, we first performed a blastn search against a custom phytoplasma database constructed by downloading all nucleotide sequences associated with the taxonomic identifier '*Candidatus* Phytoplasma' ('*Ca.* P.'; NCBI, txid33926) on 6 October 2025. This step allowed targeted detection of contigs with similarity to phytoplasma. Contigs with positive hits were then subjected to a second blastn search against the complete NCBI nt database (downloaded on 12 December 2024) to confirm their taxonomic assignment and to exclude non-specific matches. We implemented the bioinformatic analyses as a reproducible workflow in Snakemake (Köster and Rahmann, 2012). We executed all the tools within the conda environments defined in the workflow. The complete Snakemake pipeline, details of environment files, and needed configuration files have been published in Github (https://github.com/zalak-fito/FDp-genome-comparisson-workflows) and deposited in Zenodo, available through a restricted-access review link: [https://zenodo.org/records/20024946?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6IjZmMDRlZjljLTc3MGMtNGFlNi1iMGJhLWUxYWM4YzU4MmJjMCIsImRhdGEiOnt9LCJyYW5kb20iOiI1NmMxM2Q0MmNiZDcwM2I2ZmUzZmE4NDVmZTIzYjM4NyJ9.AXfa5FwEo4gjySQGymk6e1GzAjb5oq9d8KHWc9bzfSltfv51c9KWfB5KPS80EtxoVaGTAZSOX0uSpVesAU7OAg]. [The archive will be made publicly available upon publication.]

We analysed the BLAST output from the second (NCBI nt) blastn search using MEGAN v6.25.10 to assign contigs to taxonomic groups based on sequence similarity. Contigs classified as FDp were extracted for downstream analyses. We then filtered the selected contigs using SeqKit (v2.8.2) to retain only sequences longer than 300 bp. This threshold was applied to exclude very short contigs that provide limited information for reliable sequence alignment and annotation, are more likely to result in non-specific matches, and commonly arise from fragmented or low-support assemblies. We performed all downstream analyses using these filtered contigs, which we treated as the representative assemblies for each sample. We evaluated quality parameters of assembled genomes with Seqkit and the Genome Quality Assessment Tool v5.2 (QUAST).

For reference-based analyses, we mapped the trimmed reads to the FDp isolate CH as reference genome (CP097583) using CLC Genomics Workbench v26 (Qiagen, Hilden, Germany) with a length fraction of 0.9 and similarity fraction of 0.95, assigning non-specific matches randomly. We assessed genome coverage and evaluated the proportion of phytoplasma-derived reads relative to the reference genome based on these mappings.

2.4 Genome annotation and core genes-based phylogenetic analysis

We annotated the assemblies with Prokka v1.14.6 (Seemann, 2014), providing a custom phytoplasma protein database via the --proteins parameter to improve gene prediction in highly reduced and fragmented genomes. We generated this database by downloading annotated protein sequences from the NCBI genome entry for FDp isolate CH (CP097583) and supplying the resulting FASTA file to Prokka. We excluded assemblies with fewer than 200 annotated genes because they lacked sufficient gene content for reliable core gene analysis. We retained the remaining annotated assemblies for comparative analyses.

We used Panaroo v1.5.2 (Tonkin-Hill et al., 2020) to conduct gene-based comparative analysis and define the core genes across fragmented phytoplasma assemblies, enabling the construction of a reliable core gene alignment for downstream phylogenetic analyses. In Panaroo, we applied moderate cleaning mode, enabled removal of invalid genes, set the core gene threshold to 0.9, and used MAFFT for core gene alignment. We constructed the core gene alignment using annotated

assemblies from this study (n = 10) together with publicly available 16SrV phytoplasma genomes, including two M54 isolates (FDp isolate CH, CP097583; FDp isolate FD-Vv-SLO1, GCA_050897995), '*Ca.* P. rubi' (GCA_026821955), and '*Ca.* P. ziziphi' (GCA_003640545). We used '*Ca.* P. fraxini' (from 16SrVII phytoplasma group) as the outgroup. We reconstructed a maximum likelihood phylogeny from the core gene alignment using MEGA (v12). Model testing identified GTR+I as the best-fitting nucleotide substitution model, and we assessed branch support using 1,000 bootstrap replicates.

To further assess conflicting phylogenetic relationships within the FDp-related cluster, a Neighbor-Net network was constructed in SplitsTree v6 (Huson and Bryant, 2006) based on the Panaroo core gene alignment of a subset of genome assemblies, including 10 isolates from this study and two publicly available M54 genomes (FDp isolate CH, CP097583; FDp isolate FD-Vv-SLO1, GCA_050897995). Distances were calculated using the uncorrected P method. We used SplitsTree to calculate the overall delta score and Q-residual score, which describe the extent of deviation from a strictly tree-like structure and the fit of the observed distances to a tree-like metric. The robustness of splits was evaluated by bootstrap analysis, and phylogenetic incongruence in the core gene alignment was tested using the Phi test.

2.5 Genome-wide comparisons using MUMmer

To complement gene-based comparisons, we used MUMmer v4 (Marçais et al., 2018) to perform all-versus-all genome-wide sequence alignments, enabling us to assess nucleotide-level similarity and identify single nucleotide polymorphisms (SNPs) between fragmented phytoplasma assemblies. We implemented this analysis as a reproducible Snakemake workflow. The complete Snakemake pipeline, details of environment files, and Phyton scripts have been published in Github (https://github.com/zalak-fito/FDp-genome-comparisson-workflows) and deposited in Zenodo, available through a restricted-access review link: [https://zenodo.org/records/20024946?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6IjZmMDRlZjljLTc3MGMtNGFlNi1iMGJhLWUxYWM4YzU4MmJjMCIsImRhdGEiOnt9LCJyYW5kb20iOiI1NmMxM2Q0MmNiZDcwM2I2ZmUzZmE4NDVmZTIzYjM4NyJ9.AXfa5FwEo4gjySQGymk6e1GzAjb5oq9d8KHWc9bzfSltfv51c9KWfB5KPS80EtxoVaGTAZSOX0uSpVesAU7OAg].[The archive will be made publicly available upon publication.] We generated all unique pairwise genome combinations and aligned assemblies with nucmer using the --maxmatch option. We filtered alignments with the delta-filter (-r -q) to retain reciprocal best matches. Next, we extracted alignment coordinates with show-coords (-rclT).

To identify sequence variation between genomes, we analysed filtered alignments using show-snps, which reports SNPs and indels. For each genome pair, we calculated the total number of SNPs and indels and normalised SNP counts by the number of aligned base pairs. We performed downstream analyses in R v4.2.2. Using pairwise whole-genome alignments generated with MUMmer, we calculated two genomic distance measures: an ANI-like distance defined as $1 - ANI/100$ and a SNP distance defined as the number of SNPs divided by the total number of aligned base pairs. We further summarised pairwise genome similarity as ANI-like values (%) and visualised these as a heatmap.

To visualise relationships among genomes in a low-dimensional space, we performed principal coordinate analysis (PCoA) based on the matrix of pairwise ANI-like genomic distances using the cmdscale function in R. This approach projects genomes into a reduced number of dimensions while preserving pairwise distances as closely as possible. The first two principal coordinates were used for visualisation.

We assessed differences in genomic distances among groups using permutational multivariate analysis of variance (PERMANOVA) based on pairwise ANI-like distance matrices with 999 permutations in R. We used this analysis to test whether the centroids of genomic distance distributions differed according to assigned *map* genotype, host species, and geographic origin. We evaluated homogeneity of multivariate dispersion among groups using the betadisper function to ensure that observed differences were not due to unequal within-group variability. To ensure that differences in genomic distances were not caused by incomplete genome alignments, we calculated the proportion of each genome that aligned between pairs of sequences and averaged these values for each comparison.

We constructed a Neighbor-Net phylogenetic network from pairwise SNP distance matrices derived from MUMmer whole-genome comparisons to visualise relationships among genomes without imposing a strictly tree-like structure. We extracted SNP distance values and assembled a symmetric distance matrix that included all 16SrV phytoplasma isolates from grapevine and hazelnut, the two leafhoppers from this study, and two publicly available FDp M54 genomes. Diagonal values were set to zero, and we verified that no distances were missing, ensuring the matrix was complete. The resulting matrix was exported in PHYLIP distance-matrix format. We then imported the PHYLIP matrix into SplitsTree v6.7.4 and constructed a Neighbor-Net network using default settings. We calculated the overall delta score and Q-residual score in SplitsTree to describe the degree to which the SNP distance matrix deviated from a strictly tree-like structure.

3 Results

3.1 Sample quality and enrichment efficiency

All samples included in the comparative genomics analysis had confirmed 16SrV phytoplasma infection and were assigned to *map* genotypes M38 or M50 based on Sanger sequencing of the *secY–map* locus. We evaluated DNA yield, host DNA depletion, and phytoplasma DNA retention after NEBNext enrichment across grapevine, hazelnut, and leafhopper samples using Qubit measurements and qPCR assays targeting host 18S rRNA and 16SrV phytoplasma DNA.

The concentration of DNA after enrichment ranged from 0.7 to 7.5 ng/µl, with no clear association with extraction method or host type (Supplementary Table 2). Enrichment efficiency was assessed by comparing the quantification cycle (Cq) values for the host 18S rRNA qPCR and the 16SrV-specific qPCR assay before and after enrichment. For grapevine samples processed with the KingFisher DNA extraction protocol followed by NEBNext Microbiome enrichment, the Cq values of the host 18S rRNA assay increased by 5.8–9.4 cycles, indicating substantial depletion of host DNA (Supplementary Table 2). In contrast, Cq values of the 16SrV-specific assay remained largely stable, generally within 0.1–2.1 cycles of the unenriched DNA, suggesting that phytoplasma DNA was largely retained during the enrichment process (Supplementary Table 2).

Hazelnut samples exhibited variable enrichment efficiency. Changes in the 18S rRNA assay indicated both increases and decreases in host DNA after enrichment, with Cq differences (after – before enrichment) ranging from –12.2 to +3.6 cycles (Supplementary Table 2). In three samples, lower Cq values after enrichment indicated an increase in host DNA (–12.2, –6.9 and –1.2), whereas in two samples, higher Cq values indicated a decrease (+3.5 and +3.6). Changes in phytoplasma DNA amount, assessed by the 16SrV-specific assay, were generally moderate. Most samples showed lower Cq values after enrichment (–1.3 to –4.5), indicating an increase in phytoplasma signal. Notably, two hazelnut root samples (H-SB-20-1 and H-SB-20-3) showed pronounced decreases in Cq values for both assays after enrichment (18S: –6.9 and –12.2; 16SrV: –4.1 and –4.5).

In insect samples, enrichment effects were generally smaller than in grapevine samples. For the host 18S rRNA assay, Cq values after enrichment were higher in all insect samples, with Cq differences (after − before) ranging from 0.7 to 3.4 cycles (Supplementary Table 2), indicating a decrease in host DNA. Similarly, the 16SrV-specific assay showed higher Cq values after enrichment in all insect samples, with Cq differences ranging from 0.8 to 3.1 cycles, indicating a decrease in phytoplasma signal after enrichment. The strongest effect was observed in one *O. ishidae* sample (O-SB-21-2), which had been extracted using the OmniPrep kit and showed the largest increase in Cq for both the host assay (3.4) and the 16SrV-specific assay (3.1) (Supplementary Table 2).

3.2 Sequencing data quality

All libraries were initially sequenced to approximately 12 Gb of raw data, and selected samples were subsequently resequenced to increase the read depth for phytoplasma (Table 1). The proportion of reads mapping to the FDp isolate CH as reference genome (CP097583) was generally low, ranging from 0.0005% to 0.25% of the total reads per sample (Table 1). Despite the low proportion of phytoplasma reads, mapping to the FDp reference genome yielded between 14% to 85% genome coverage across samples (Table 1). Samples with higher reference genome coverage were found among grapevine, hazelnut, and insect samples, indicating that high-quality phytoplasma data could be obtained across multiple host types.

De novo assembly of phytoplasma-enriched reads produced fragmented draft genomes with variable total assembly lengths. We deposited all resulting draft genomes in a public database (accession numbers provided in Supplementary Table 3). The total length of assembled contigs longer than 300 bp ranged from 36,122 bp to 682,383 bp, and the number of contigs per assembly ranged from 42 to 601 (Table 1). Assemblies with shorter total contig lengths typically showed lower N50 values and a higher proportion of short contigs, while the longest assemblies combined greater total length with higher N50 (Table 1). These differences in assembly size were consistent with the proportion of reads mapping to the phytoplasma reference genome, with samples containing a higher fraction of phytoplasma reads producing longer assemblies (Supplementary Figure 1A). GC content varied substantially among assemblies with short total contig lengths, whereas assemblies approaching at least 60% of the expected genome size showed GC values close to the 22% reported for the FDp genome (Debonneville et al., 2022) (Table 1, Supplementary Figure 1B).

3.3 Genome annotation and core genes-based phylogeny

Highly fragmented assemblies yielded low numbers of annotated genes (Table 1), and only genomes with more than 200 annotated genes were retained for core gene analysis. In total, 10 of 18 newly assembled genomes met this criterion, with annotated gene counts ranging from 243 to 547 (Table 1). These genomes were analysed together with publicly available genomes of other 16SrV phytoplasmas (FDp isolate CH, FD-Vv-SLO1, '*Ca.* P. rubi', '*Ca.* P. ziziphi') and '*Ca.* P. fraxini' as a representative of the 16SrVII phytoplasma group.

Panaroo identified 17 core genes and the concatenated core gene alignment used for phylogenetic reconstruction had a total length of 18,035 bp (Supplementary Data 1). The phylogenetic tree reconstructed from the Panaroo core gene alignment showed that the analysed genomes separated into two clusters corresponding to *map* genotypes M38 and M50 (Figure 2A). Samples originating from grapevine, hazelnut, and insect vectors were interspersed within these genotype-defined clusters and did not group according to host species. The two M54 genomes included in the analysis clustered with the M38 group. Genomes of other 16SrV phytoplasmas, '*Ca.* P. rubi' and '*Ca.* P. ziziphi', branched outside the FDp cluster, whereas '*Ca.* P. fraxini' (16SrVII group) was set as the outgroup.

To further examine relationships within the FDp cluster, we conducted a phylogenetic network analysis on a reduced dataset comprising 10 genomes from this study and two publicly available M54 genomes, using Panaroo to identify 39 core genes (Figure 2B). The concatenated core gene alignment (52,447 bp, Supplementary Data 2) was used to construct a Neighbor-Net network. The resulting network revealed a non-tree-like structure with several reticulations among the analysed genomes (Figure 2B). The overall delta score, quantifying the deviation of the distance data from a strictly tree-like structure, was 0.2249, whereas the Q-residual score, reflecting the fit of the observed distances to a tree-like metric, was 0.01096. Both metrics range from 0 to 1, with values approaching 0 indicating a perfectly tree-like signal, while increasing values reflect progressively greater deviation from a tree-like structure. The Phi test detected statistically significant evidence of reticular events in the core genome alignment ($p < 0.001$), based on incompatibilities among phylogenetic signals in the alignment. The analysed genomes were also annotated by sampling area, and isolates from the same geographic area (NM and SB) formed distinct groups within the network (Figure 2B). Host origin did not correspond to clustering, as isolates from different hosts were intermixed within *map* genotype and location groups.

3.4 Genome-wide comparison of phytoplasma genomes using MUMmer

As not all assembled genomes met the minimum annotation threshold required for inclusion in the Panaroo core gene analysis, but still contained usable genomic sequence data, we performed an additional all-to-all genome-wide comparison using MUMmer. Pairwise genome similarity was estimated using an ANI-like identity metric derived from whole-genome alignments, and genomic distances were calculated as $1 - ANI\text{-}like/100$, representing the proportion of nucleotide differences across aligned regions.

The resulting ANI-like heatmap revealed a clear genomic structure among the analysed phytoplasma strains, characterised by blocks of high similarity that correspond to *map* genotype classification (Figure 3A). Genomes assigned to the same *map* genotype consistently exhibited high ANI-like values, forming visually coherent clusters along the diagonal of the matrix, whereas lower similarity values were observed between genotypes. Within the M38 cluster, one sample (S-PT-24) showed reduced similarity to other M38 genomes, appearing as a less tightly associated member of the cluster (Figure 3A). Notably, genomes assigned to the M54 genotype, including both isolates CH and FD-Vv-SLO1, grouped within the same high-similarity block, further supporting their close genomic relationship. A PCoA of ANI-like genomic distances showed a similar separation of genomes, with genome-wide similarity patterns corresponding to *map* genotype classification with samples from grapevine, hazelnut, and leafhoppers intermixed within *map* genotype clusters rather than grouping by host species (Figure 3B). One sample (again S-PT-24) deviated from the main M38 cluster and was clearly separated in the ordination space (Figure 3B).

PERMANOVA analysis of ANI-like genomic distances showed that, in this dataset, genome-wide variation among phytoplasma strains corresponds more closely to *map* genotype classification than to host species, while geographic origin also contributes significantly (Table 2).

The Neighbor-Net network based on genome-wide SNP distances (Figure 3C) included 18 draft genomes from this study and two publicly available FDp isolates (M54), and it showed that genomes from this study formed two main groups corresponding to *map* genotypes M38 and M50. This topology was consistent with the clustering patterns observed in the core genes phylogeny and heatmap. Also within these clusters, samples from grapevine, hazelnut, and leafhoppers were

intermixed and did not group by host. Several subgroups within the network corresponded to the sampling area of the isolates, indicated in Figure 3C (SB, NM, LT, and PT).

The Neighbor-Net network displayed several reticulations, indicating some conflicting phylogenetic signal within the distance data. The overall delta score of the network was 0.3453, calculated across 20 taxa; the delta score quantifies the deviation of the distance data from a strictly tree-like structure, with values close to 0 indicating a perfectly tree-like signal. The Q-residual score was 0.0503, which measures the fit of the observed distances to a tree-like metric; lower values indicate better fit, thereby complementing the delta score by assessing how well the distance data conform to a tree-like model. Delta scores calculated for individual taxa ranged from 0.2690 to 0.4468, with the highest values observed for H-SB-20-1 (0.4468), S-PT-24 (0.4198), G-LT-23 (0.4146), and O-SB-21-2 (0.4037). The highest Q-residual values were detected for H-SB-20-1 (0.1062), S-PT-24 (0.0856), and G-PT-19 (0.0820).

4 Discussion

Genome-wide analyses in this study show that the observed patterns of genomic similarity of FD-related 16SrV phytoplasma isolates from grapevine, hazelnut, and leafhoppers in Slovenia correspond closely to *map* genotype classification, while no consistent grouping by host species was observed. Within the analysed *map* genotypes (M38 and M50), isolates also formed subgroups associated with sampling areas. Within these location-associated groups, isolates from different hosts were consistently interspersed across all analyses, with no evidence of host-associated lineage separation or distinct "grapevine" and "hazelnut" lineages. These results indicate that, in this dataset, *map* genotype classification captures genome-wide similarity at the broader level, while finer-scale variation reflects geographic origin rather than host species. This pattern is consistent with previous observations that the same *map* genotypes are detected in multiple hosts in Slovenia (Kogej Zwitter et al., 2023); however, in this study, the genome-wide data directly support such cross-host connectivity for the *map* genotypes M38 and M50. Additional *map* genotypes occur in grapevine and hazelnut and were not necessarily detected across hosts, possibly indicating that the observed cross-host connectivity applies to specific lineages rather than the full diversity of 16SrV phytoplasmas.

The consistency of the observed patterns across analytical approaches indicates that the correspondence between genome-wide similarity patterns and *map* genotype classification is robust, extending previous *map*-based studies (Arnaud et al., 2007; Plavec et al., 2019; Rossi et al., 2019; Malembic-Maher et al., 2020) by supporting these relationships with genome-wide data. The placement of phytoplasma genomes typed as M54 within the M38-associated cluster in maximum-likelihood phylogeny further aligns with previous work describing close relationships between these genotypes (Malembic-Maher et al., 2020). In that study, M38 is reported as a two-SNP variant of M54 in the *map* gene, and M54 has been proposed to arise from an ancestral M38, although alternative scenarios remain possible. Compared to single-locus typing, the genome-wide approach applied here provides higher resolution, confirms *map* genotype relationships at the genome level, and reveals additional complexity not captured by *map*-based classification alone. Taken together, these results support the continued use of *map* genotypes as a practical proxy for interpreting relationships in epidemiological studies of FDp and related strains, at least for the genotypes analysed in this study.

Statistical analyses indicate that the *map*-based genotype classification is strongly reflected in whole-genome distance patterns, whereas host species does not significantly correspond to genome-wide similarity patterns. Geographic origin provided a secondary signal, but its

interpretation is limited by uneven sampling and the fact that the dataset represents a spatial and temporal snapshot rather than a densely sampled time series. These results, therefore, describe patterns of genomic similarity rather than underlying processes such as local adaptation or dispersal dynamics. From an epidemiological perspective, the detection of closely related genomes across grapevine, hazelnut, and analysed leafhoppers is consistent with a multi-host system, in which FD-related phytoplasmas circulate across host compartments, at least for the genotypes analysed in this study (Malembic-Maher et al., 2020; Kogej Zwitter et al., 2023). However, genomic similarity alone does not resolve transmission pathways, reservoir roles, or vector competence. In particular, the presence of the same *map* genotypes in *S. titanus* and *O. ishidae* should not be interpreted as evidence of equivalent vector roles, as the vector competence of *O. ishidae* for FDp transmission from hazelnut to grapevine, and vice versa, has not been demonstrated and would require experimental validation. These interpretations should also be considered in light of variable enrichment efficiency across host types, which influenced genome recovery and limited analyses primarily to genome-wide similarity rather than detailed gene content or structural genome comparisons.

Network analyses revealed reticulate patterns, suggesting the possible presence of more complex evolutionary signals. In other phytoplasma systems, similar patterns have been associated with genome rearrangements, mobile genetic elements, and potential recombination (Bai et al., 2006; Wei et al., 2008). However, these interpretations remain tentative in our case, as network-based approaches cannot distinguish between recombination, mixed infections, or technical artefacts arising from fragmented assemblies and incomplete genome recovery. Mixed infection is a plausible explanation for the outlier behaviour observed in individual samples, such as *S. titanus* (S-PT-24), particularly given the feeding ecology of insect vectors and the potential for exposure to multiple infected plants (Rossi et al., 2019; Malembic-Maher et al., 2020). However, this interpretation is not directly demonstrated by the data. Importantly, the analytical framework applied here, including distance-based approaches such as network reconstruction and PCoA, was selected to capture overall relationships between phytoplasma within-host populations and is therefore relatively robust to individual outliers. Thus, even if S-PT-24 were a mixed infection, it would most likely affect fine-scale relationships and reticulation-like signals rather than the primary *map* genotype-level clustering.

In conclusion, genome-scale data show that, for the analysed genotypes, *map* genotype classification reliably reflects genome-wide similarity, while host species does not structure genomic relationships in our dataset. These findings are consistent with the presence of shared FDp-related populations across grapevine, hazelnut, and leafhoppers within the studied system. These results provide a genome-wide framework for interpreting FDp epidemiology, but further studies are needed to resolve transmission dynamics and the roles of specific hosts and vectors. The observed connectivity across hosts underscores the importance of integrated monitoring of plants and insect vectors within FDp surveillance programmes.

5 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

6 Author Contributions

ZKZ participated in the design of the study, laboratory experiments, bioinformatic analyses, data curation, visualisation and writing the manuscript. Supervision and review were done by NM and DK. Both were included in conceptualisation and study design, formal analyses, funding acquisition and coordinated the study. All authors contributed to the article and approved the submitted version.

7 Funding

The study was financially supported by the Slovenian Research Agency through the project NanoPhyto (L7-2632), core financing (P4-0463), and a young researcher grant (ZKZ).

8 Acknowledgments

We thank the Slovenian phytosanitary inspectors and expert surveyors for providing the field samples, and Irena Bajde, Jakob Brodarič and Tjaša Jakomin for help in the laboratory. We thank Primož Pajk for help with the location map preparation. We also thank Maja Ravnikar for productive discussions. The authors acknowledge the Slovenian Research Agency (ARIS) for supporting research infrastructure used in this work through grant IO-0004 and the Slovenian node of Elixir ERIC for providing access to servers.

9 References

Angelini, E., Squizzato, F., Lucchetta, G., and Borgo, M. (2004). Detection of a phytoplasma associated with grapevine Flavescence dorée in Clematis vitalba.

Arnaud, G., Malembic-Maher, S., Salar, P., Bonnet, P., Maixner, M., Marcone, C., et al. (2007). Multilocus sequence typing confirms the close genetic interrelatedness of three distinct flavescence dorée phytoplasma strain clusters and group 16SrV phytoplasmas infecting grapevine and alder in Europe. *Appl. Environ. Microbiol.* 73, 4001–4010. doi: 10.1128/AEM.02323-06

Atanasova, B., Spasov, D., Jakovljević, M., Jović, J., Krstić, O., Mitrović, M., et al. (2014). First report of alder yellows phytoplasma associated with common alder (Alnus glutinosa) in the Republic of Macedonia. *Plant Dis.* 98, 1268.

Bai, X., Zhang, J., Ewing, A., Miller, S. A., Radek, A. J., Shevchenko, D. V., et al. (2006). Living with genome instability: The adaptation of phytoplasmas to diverse environments of their insect and plant hosts. *J. Bacteriol.* 188, 3682–3696. doi: 10.1128/JB.188.10.3682-3696.2006

Casati, P., Jermini, M., Quaglino, F., Corbani, G., Schaerer, S., Passera, A., et al. (2017). New insights on Flavescence dorée phytoplasma ecology in the vineyard agro-ecosystem in southern Switzerland. *Annals of Applied Biology* 171, 37–51. doi: 10.1111/aab.12359

Chuche, J., and Thiéry, D. (2014). Biology and ecology of the Flavescence dorée vector Scaphoideus titanus: A review. *Agron. Sustain. Dev.* 34, 381–403. doi: 10.1007/s13593-014-0208-7

Debonneville, C., Mandelli, L., Brodard, J., Groux, R., Roquis, D., and Schumpp, O. (2022). The Complete Genome of the "Flavescence Dorée" Phytoplasma Reveals Characteristics of Low Genome Plasticity. *Biology (Basel).* 11. doi: 10.3390/biology11070953

512 Filippin, L., Jović, J., Cvrković, T., Forte, V., Clair, D., Toševski, I., et al. (2009). Molecular
513 characteristics of phytoplasmas associated with Flavescence dorée in clematis and
514 grapevine and preliminary results on the role of Dictyophara europaea as a vector. *Plant*
515 *Pathol.* 58, 826–837. doi: https://doi.org/10.1111/j.1365-3059.2009.02092.x

516 Hogenhout, S. A., Oshima, K., Ammar, E. D., Kakizawa, S., Kingdom, H. N., and Namba, S.
517 (2008). Phytoplasmas: Bacteria that manipulate plants and insects. *Mol. Plant Pathol.* 9,
518 403–423. doi: 10.1111/j.1364-3703.2008.00472.x

519 Hren, M., Boben, J., Rotter, A., Kralj, P., Gruden, K., and Ravnikar, M. (2007). Real-time PCR
520 detection systems for Flavescence dorée and Bois noir phytoplasmas in grapevine:
521 Comparison with conventional PCR detection and application in diagnostics. *Plant Pathol.*
522 56, 785–796. doi: 10.1111/j.1365-3059.2007.01688.x

523 Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary
524 studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030

525 IRPCM (2004). "Candidatus Phytoplasma", a taxon for the wall-less, non-helical prokaryotes
526 that colonize plant phloem and insects. *Int. J. Syst. Evol. Microbiol.* 54, 1243–1255. doi:
527 10.1099/ijs.0.02854-0

528 Kogej Zwitter, Z., Kutnjak, D., and Mehle, N. (2025). From vineyard to genome: optimized
529 enrichment and sequencing of Flavescence dorée phytoplasma from grapevine samples.
530 *Microb. Genom.* 11. doi: 10.1099/mgen.0.001514

531 Kogej Zwitter, Z., Seljak, G., Jakomin, T., Brodarič, J., Vučurović, A., Pedemay, S., et al.
532 (2023). Epidemiology of flavescence dorée and hazelnut decline in Slovenia: geographical
533 distribution and genetic diversity of the associated 16SrV phytoplasmas. *Front. Plant Sci.*
534 14. doi: 10.3389/fpls.2023.1217425

535 Köster, J., and Rahmann, S. (2012). Snakemake-a scalable bioinformatics workflow engine.
536 *Bioinformatics* 28, 2520–2522. doi: 10.1093/bioinformatics/bts480

537 Krstić, O., Crvković, T., Marinkovič, S., Jakovljević, M., Mitrović, M., Toševski, I., et al.
538 (2022). Genetic Diversity of Flavescence Dorée Phytoplasmas in Vineyards of Serbia:
539 From the Widespread Occurrence of Autochthonous Map-M51 to the Emergence of
540 Endemic Map-FD2 (Vectotype II) and New Map-FD3 (Vectotype III) Epidemic
541 Genotypes. *Agronomy* 12.

542 Krstić, O., Cvrković, T., Mitrović, M., Radonjić, S., Hrnčić, S., Toševski, I., et al. (2018).
543 Wolbachia infection in natural populations of Dictyophara europaea, an alternative vector
544 of grapevine Flavescence dorée phytoplasma: effects and interactions. *Annals of Applied*
545 *Biology* 172, 74–64.

546 Kube, M., Mitrovic, J., Duduk, B., Rabus, R., and Seemüller, E. (2012). Current view on
547 phytoplasma genomes and encoded metabolism. *The Scientific World Journal* 2012. doi:
548 10.1100/2012/185942

549 Malembic-Maher, S., Desqué, D., Khalil, D., Salar, P., Bergey, B., Danet, J.-L., et al. (2020).
550 When a Palearctic bacterium meets a Nearctic insect vector: Genetic and ecological

551 insights into the emergence of the grapevine Flavescence dorée epidemics in Europe. *PLoS*
552 *Pathog.* 16, 1–28. doi: 10.1371/journal.ppat.1007967

553 Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., and Zimin, A. (2018).
554 MUMmer4: A fast and versatile genome alignment system. *PLoS Comput. Biol.* 14. doi:
555 10.1371/journal.pcbi.1005944

556 Marcone, C., Pierro, R., and Palmieri, C. (2024). Occurrence, Impact, and Multilocus Sequence
557 Analysis of Alder Yellows Phytoplasma Infecting Common Alder and Italian Alder in
558 Southern Italy. *Microorganisms* 12. doi: 10.3390/microorganisms12061140

559 Mehle, N., Jakoš, N., Mešl, M., Miklavc, J., Matko, B., Rot, M., et al. (2019). Phytoplasmas
560 associated with declining of hazelnut (Corylus avellana) in Slovenia. *Eur. J. Plant Pathol.*
561 155, 1117–1132. doi: 10.1007/s10658-019-01839-3

562 Mehle, N., Prezelj, N., Hren, M., Boben, J., Gruden, K., Ravnikar, M., et al. (2013). "A Real-
563 Time PCR Detection System for the Bois Noir and Flavescence Dorée Phytoplasmas and
564 Quantification of the Target DNA," in *Phytoplasma : Methods and protocols, methods in*
565 *molecular biology*, eds. M. Dickinson and J. Hodgetts (New York: Humana Press, Totowa,
566 NJ), 269–281.

567 Mehle, N., Rupar, M., Seljak, G., Ravnikar, M., and Dermastia, M. (2011). Molecular diversity
568 of "flavescence dorée" phytoplasma strains in Slovenia. *Bull. Insectology* 64, 29–30.

569 Mehle, N., Seljak, G., Rupar, M., Ravnikar, M., and Dermastia, M. (2010). The first detection of
570 a phytoplasma from the 16SrV (Elm yellows) group in the mosaic leafhopper Orientus
571 ishidae . *New Dis. Rep.* 22, 11–11. doi: 10.5197/j.2044-0588.2010.022.011

572 Plavec, J., Budinšćak, Križanac, I., Škorić, D., Foissac, X., and Šeruga Musić, M. (2019).
573 Multilocus sequence typing reveals the presence of three distinct flavescence dorée
574 phytoplasma genetic clusters in Croatian vineyards. *Plant Pathol.* 68, 18–30. doi:
575 10.1111/ppa.12934

576 Radonjić, S., Hrnčić, S., Krstić, O., Cvrković, T., Mitrović, M., Jović, J., et al. (2013). First
577 report of alder yellows phytoplasma infecting common and grey alder (Alnus glutinosa and
578 A. incana) in Montenegro. *Plant Dis.* 97, 686.

579 Rizza, S., Rossi, M., Spallino, R. E., Tedeschi, R., Marzachì, C., Catara, V., et al. (2021).
580 Ecology-based analysis of a recent association between Spartium junceum and 16SrV
581 phytoplasma. *Plant Pathol.* 70, 305–317. doi: 10.1111/ppa.13300

582 Rossi, M., Pegoraro, M., Ripamonti, M., Abbà, S., Bealb, D., Giraudoa, A., et al. (2019).
583 Genetic Diversity of Flavescence Dorée Phytoplasmas at the Vineyard Scale. *Appl.*
584 *Environ. Microbiol.* 85, 1–15.

585 Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–
586 2069. doi: 10.1093/bioinformatics/btu153

Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J. A., et al. (2020). Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* 21. doi: 10.1186/s13059-020-02090-4

Wang, R., Bai, B., Li, D., Wang, J., Huang, W., Wu, Y., et al. (2024). Phytoplasma: A plant pathogen that cannot be ignored in agricultural production—Research progress and outlook. *Mol. Plant Pathol.* 25. doi: 10.1111/mpp.13437

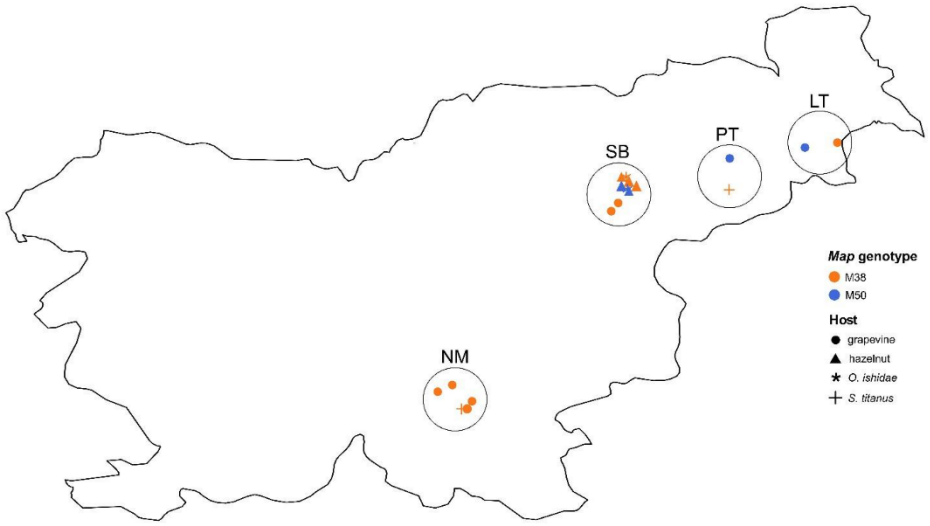
Wei, W., Davis, R. E., Jomantiene, R., and Zhao, Y. (2008). Ancient, recurrent phage attacks and recombination shaped dynamic sequence-variable mosaics at the root of phytoplasma genome evolution. *Proc. Natl. Acad. Sci. U. S. A.* 105, 11827–11832. doi: 10.1073/pnas.0805237105

Wei, W., and Zhao, Y. (2022). Phytoplasma Taxonomy: Nomenclature, Classification, and Identification. *Biology (Basel).* 11. doi: 10.3390/biology11081119

10 Data Availability Statement

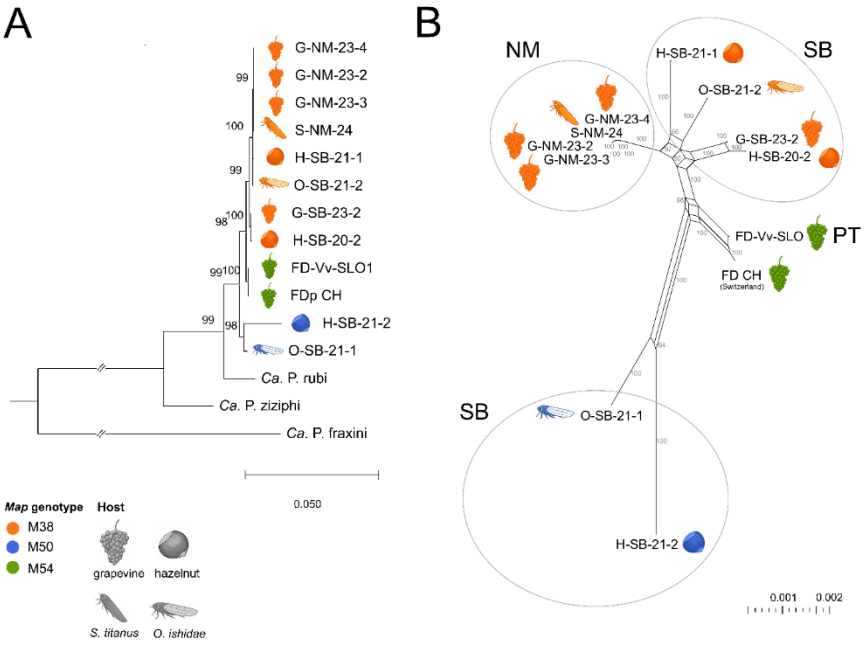
All genome sequencing data generated in this study are available from the European Nucleotide Archive (ENA) under BioProject accession PRJEB111790. Raw sequencing reads and draft genome assemblies are available under the ENA/INSDC accessions listed in Supplementary Table 3. The Snakemake workflows and associated scripts used for read processing, *de novo* assembly, contig validation, and genome-wide MUMmer comparisons will be made publicly available at https://github.com/zalak-fito/FDp-genome-comparisson-workflows upon publication. All other data supporting the findings of this study are included in the article and its Supplementary Material.

11 Figures



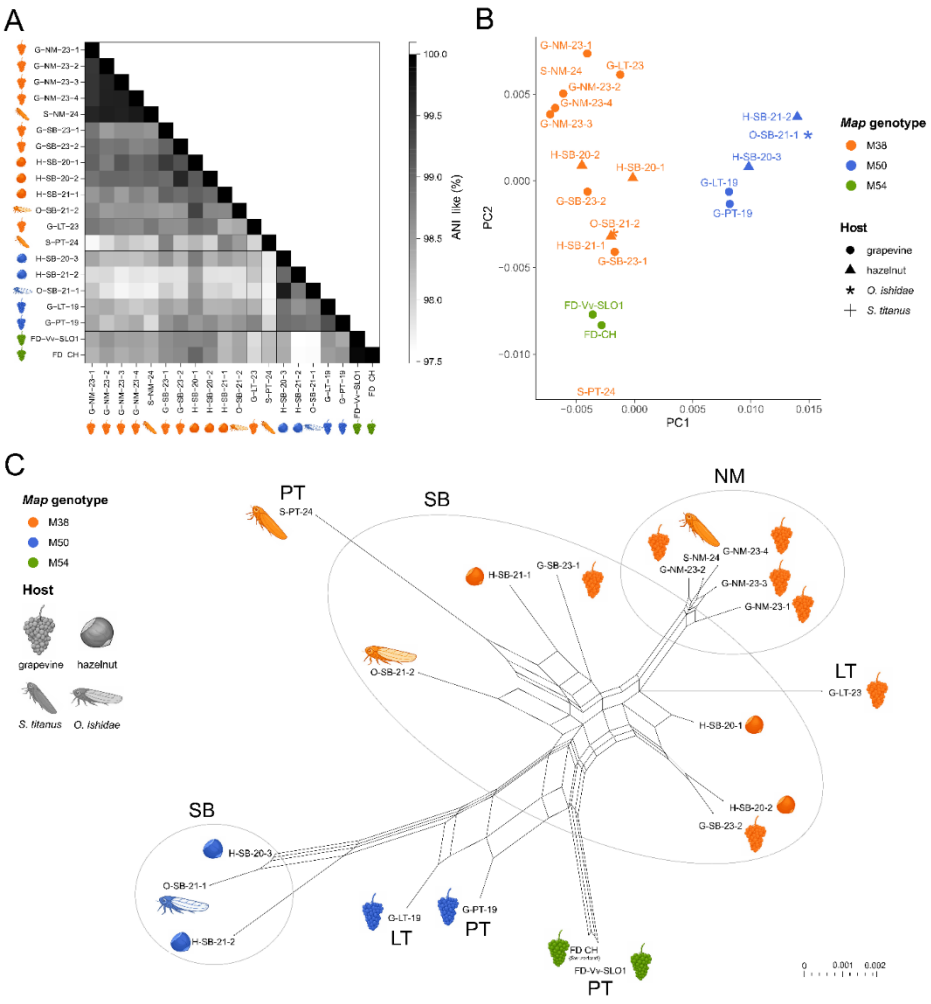
609

610 **Figure 1**. Sampling locations of FDp-related isolates in Slovenia. Colours indicate *map* genotype and
611 point shapes denote host species. All hazelnut and *O. ishidae* samples were collected from a single
612 hazelnut orchard, and point positions were jittered for visual clarity. Circles outline sampling areas,
613 and labels indicate sampling areas (NM = Novo Mesto, SB = Slovenska Bistrica, PT = Ptuj, LT =
614 Ljutomer).



615

616 **Figure 2**. Core gene phylogenetic analysis of 16SrV phytoplasmas. **(A)** Maximum likelihood
617 phylogenetic tree inferred from the Panaroo core gene alignment (18,035 bp, Supplementary Data 1)
618 of 16SrV phytoplasma genomes, including isolates from this study (genome assembly accession
619 numbers given in Supplementary Table 3), FDp isolate CH (CP097583), FDp isolate FD-Vv-SLO1
620 (GCA_050897995), '*Ca.* P. rubi' (GCA_026821955), and '*Ca.* P. ziziphi' (GCA_003640545), with
621 '*Ca.* P. fraxini' (16SrVII) as the outgroup. The tree was reconstructed in MEGA using the maximum
622 likelihood method with the GTR+I substitution model and 1,000 bootstrap replicates. Bootstrap
623 values ≥70% are shown at internal nodes, and the scale bar indicates nucleotide substitutions per site.
624 The branch leading to '*Ca.* P. fraxini' was truncated for visualization purposes. **(B)** Neighbour-Net
625 network inferred from a core gene alignment of FDp-related isolates only (52,447 bp, Supplementary
626 Data 2). The network was constructed in SplitsTree using the uncorrected P-distance. Edge lengths
627 are proportional to sequence divergence, and reticulations indicate conflicting phylogenetic signal.
628 Bootstrap support values ≥60% are shown on the network edges, and the scale bar represents
629 nucleotide substitutions per site. In panels A and B, colours indicate *map* genotype, icons represent
630 host species, and sampling locations are indicated by large labels, with grey circles grouping samples
631 from the same location where relevant.



632

633 **Figure 3**. Genome-wide comparison of FDp-related isolates based on MUMmer analysis, using 18
634 draft genomes generated in this study and two publicly available FDp genomes (M54). **(A)** Heatmap
635 of pairwise genome-wide similarity among phytoplasma genomes based on ANI-like values derived
636 from MUMmer alignments. Each cell shows the average nucleotide identity (ANI-like, %) between
637 two genomes, calculated across aligned genomic regions. Genomes are arranged along both axes
638 according to *map* genotype and sampling location. **(B)** Principal coordinate analysis of pairwise
639 genomic distances among phytoplasma genomes. Distances were calculated as *1 − ANI-like/100*,
640 where ANI-like identity was derived from all-versus-all genome-wide alignments generated with
641 MUMmer. Each point represents a genome; colours indicate *map* genotype and point shapes denote
642 host species. Axes correspond to the first two principal coordinates. **(C)** Neighbour-Net network

based on distances derived from whole-genome comparisons of FDp-related isolates. The distance matrix was calculated as the number of SNPs divided by the total number of aligned base pairs between each pair of genomes. The scale bar represents SNP distance. In panels A and C, colours indicate *map* genotype, icons represent host species, and sampling locations are indicated by large labels, with grey circles grouping samples from the same location where relevant.

12 Tables

Table 1. Sequencing and assembly statistics for 16SrV phytoplasma-infected samples from grapevine, hazelnut, and leafhopper hosts. Columns show host and sample identifiers, total Illumina sequencing data (Gb), the proportion of reads mapping to the FDp isolate CH as reference (ref) genome (CP097583), the percentage of the reference genome covered by mapped reads, the total length and number of assembled contigs longer than 300 bp, Quast-derived assembly metrics (largest contig, N50, and GC content of assembled contigs) and gene count obtained from genome annotation with Prokka.

Sample	Host	*Map* genotype by Sanger sequencing	Final data amount [Gb]	Mapped reads to FDp ref [%]	FDp ref coverage with reads [%]	Total contig length (≥300 bp)	N of contigs (≥300 bp)	Largest contig	N50	GC (%)	Gene count
G-SB-23-1	grapevine	M38	87.1	0.0030	52	196,750	315	6,270	906	28.3	142
G-SB-23-2	grapevine	M38	65.6	0.0066	67	509,707	601	8,437	1,270	25.8	341
G-NM-23-1	grapevine	M38	16	0.0084	34	106,979	174	9,130	940	31.3	80
G-NM-23-2	grapevine	M38	12.7	0.0370	66	377,621	340	13,392	1,595	25.4	280
G-NM-23-3	grapevine	M38	37.7	0.0170	83	549,648	309	29,146	2,860	23.6	475
G-NM-23-4	grapevine	M38	52.5	0.0194	68	390,102	364	12,471	1,702	25.9	294
G-LT-23	grapevine	M38	94.4	0.0031	27	75,779	109	6,391	1,002	32.5	79
G-LT-19	grapevine	M50	85.7	0.0047	37	126,846	176	6,727	923	30.7	108
G-PT-19	grapevine	M50	112.9	0.0019	18	36,122	48	6,197	836	35.7	49
H-SB-20-1	hazelnut	M38	95	0.0040	14	37,994	42	6,377	2,187	40.1	59
H-SB-20-2	hazelnut	M38	59.6	0.0091	55	311,229	322	21,949	1,542	31.6	243
H-SB-20-3	hazelnut	M50	90.8	0.0005	29	122,971	157	6,128	969	32.4	95
H-SB-21-1	hazelnut	M38	36	0.0417	83	592,805	245	16,580	4,201	22.7	497
H-SB-21-2	hazelnut	M50	18.5	0.0155	78	662,011	373	41,313	2,384	26.7	547
O-SB-21-1	*O. ishidae*	M50	16.6	0.0978	73	580,621	293	12,569	3,197	25.69	421
O-SB-21-2	*O. ishidae*	M38	13.2	0.2515	85	584,799	278	16,044	3,731	24.14	445
S-NM-24	*S. titanus*	M38	12.6	0.0911	78	682,383	321	32,643	3,269	26.72	494
S-PT-24	*S. titanus*	M38	13.3	0.0035	28	95,709	132	4,495	1,201	32.37	55

Table 2. Results of PERMANOVA testing the effects of *map* genotype, host species, and geographic origin on genomic distances among phytoplasma genomes. Genomic distances were calculated as *1−ANI-like/100* from all-versus-all whole-genome alignments generated with MUMmer. The proportion of explained variance (R^2), pseudo-F statistics, and permutation-based p-values (999 permutations) are shown for each factor. R^2 values represent marginal effects estimated independently for each factor and are therefore not additive.

	R^2	F	p-value
Map genotype	0.30	7.13	0.001
Location	0.21	3.38	0.001
Host	0.08	1.32	0.206
Residual	0.21		

Chapter 3

Discussion

3.1 Gene-Based Epidemiological Relationships of 16SrV Phytoplasmas in Grapevine, Hazelnut, and Leafhopper Vectors

The first publication in this thesis showed that 16SrV phytoplasma diversity in Slovenia is not uniformly distributed across grapevine, hazelnut, and insect vectors. Eleven *map* genotypes were identified, including representatives of all three epidemic Map-FD clusters described in grapevine and one grapevine non-epidemic *map* genotype. Only a limited subset was shared between grapevine and hazelnut, while most *map* genotypes were detected in a single host. This pattern indicates partial overlap between the grapevine and hazelnut phytoplasma populations but does not support a simple host-independent distribution of all *map* genotypes. Instead, it suggests that only some lineages occur across both grapevine and hazelnut, while others remain associated either with the grapevine–*S. titanus* cycle or with hazelnut and surrounding vegetation.

Genotype M54 (from the Map-FD2 cluster) is often associated with vineyard epidemics. In the Slovenian dataset, this *map* genotype was detected only in grapevine and *S. titanus*, but not in hazelnut or *O. ishidae*. This observation fits its frequent association with vineyard epidemics in Europe, where it is often reported among the main epidemic *map* genotypes (Malembic-Maher et al., 2020; Plavec et al., 2019; Rizzoli et al., 2021). Still, the Slovenian pattern should be interpreted as dataset-specific rather than as evidence of strict host restriction. In southern Switzerland, wild hazelnut growing near FDp-infected vineyards carried 16SrV phytoplasmas from all three Map-FD clusters (Casati et al., 2017). One Map-FD2-type phytoplasma from a Swiss wild hazelnut was later assigned to *map* genotype M54 (Oggier et al., 2024), indicating that M54 can also occur in hazelnut in some ecological settings.

Map genotypes M38, M50, and M122 were detected in both grapevine and hazelnut. Previous studies have also reported these genotypes from non-Vitis hosts, including alder, indicating that they are not confined to vineyard systems (Malembic-Maher et al., 2020; Plavec et al., 2019). Their presence in hazelnut is therefore compatible with a broader host range, but it does not resolve the transmission route. Because the infected hazelnut orchards analysed in this study were not located near vineyards with the same *map* genotypes, the Slovenian data do not permit a direct inference of a grapevine–hazelnut connection. One possible explanation is the involvement of other host plants. Alder is frequently reported as a reservoir of 16SrV phytoplasmas and harbours substantial genetic diversity of these phytoplasmas (Malembic-Maher et al., 2020; Marcone et al., 2024).

The role of alder in maintaining 16SrV phytoplasma diversity was again shown by the third publication of this thesis, where 16SrV phytoplasma was detected in asymptomatic *Alnus glutinosa* in the Netherlands. The detected *map* genotype, M44, does not belong to the epidemic Map-FD1, Map-FD2, or Map-FD3 clusters and has previously been linked to alder and leafhopper *O. alni*, rather than to the grapevine–*S. titanus* cycle. Its detection in asymptomatic alder supports the idea that 16SrV phytoplasmas may persist in non-cultivated hosts without visible symptoms and outside the *map* genotypes repeatedly associated with grapevine epidemics. For the Slovenian hazelnut decline, the combined evidence of shared *map* genotypes and *map* genotypes found in non-cultivated plants is more compatible with hazelnut phytoplasma infection being associated with surrounding vegetation than with an independent host-specific cycle.

The occurrence of *map* genotype M51 (from the cluster Map-FD3) in grapevine further supports the involvement of non-cultivated host-associated cycles in the epidemiology of 16SrV phytoplasmas. M51 has been linked to *C. vitalba* and reported from grapevine outbreaks in several European regions (Krstić et al., 2022; Malembic-Maher et al., 2020). Its detection in Slovenia, together with the previous identification of Map-FD3-related strains in *C. vitalba* (Mehle et al., 2011), is consistent with either local transmission from *C. vitalba* or introduction from neighbouring regions where this genotype has already been reported (Malembic-Maher et al., 2020; Plavec et al., 2019; Rossi et al., 2019). The second publication of this thesis provides a comparable example from the Czech Republic, where 16SrV phytoplasmas were detected in both grapevine and *C. vitalba*, and the grapevine isolate was assigned to *map* genotype M51. Together, the Slovenian and Czech observations indicate that Map-FD3-related infections in grapevine may be linked to non-cultivated host-associated epidemiological cycles.

A similar mixture of vineyard-associated and non-cultivated host-associated 16SrV diversity has been reported from neighbouring countries. The closest parallel is Croatia, where FD-related phytoplasmas were first reported in northwestern vineyards and later surveys described a strongly affected area west and southwest of Zagreb, near the Slovenian border. These early reports already suggested variability within 16SrV-C phytoplasmas rather than a single uniform genotype (Šeruga Musić et al., 2011; Škorić et al., 2011). Later multilocus analyses in Croatia showed the presence of all three Map-FD clusters. Map-FD2, including genotype M54, was linked mainly with vineyard epidemics in western Croatia and Istria, whereas Map-FD3 was found in a broader host and geographic context (Plavec et al., 2019). A comparable contrast between vineyard and extra-vineyard diversity has been reported in northern Italy, where symptomatic grapevines were dominantly infected by M54, while surrounding plants and insects carried additional FD-related *map* genotypes, including M38, M50, M51, M43, and M48 (Rigamonti et al., 2023). In Hungary, the available data point more towards non-cultivated hosts, such as alder and *C. vitalba* near vineyards, which were infected with 16SrV phytoplasmas. *C. vitalba* isolates were assigned to M51, and alder also carried genotypes found in grapevine outbreaks, including M50 (Ember et al., 2011; Malembic-Maher et al., 2020). Austrian data are less detailed at the *map*-genotype level, but outbreaks in South and Southeast Styria were associated with 16SrV-C phytoplasmas and nearby infected *C. vitalba*, again suggesting a possible link with non-cultivated hosts (Reisenzein & Steffek, 2011). Overall, data from neighbouring regions show that FD-related 16SrV phytoplasmas occur in both vineyard epidemics and non-cultivated hosts, with the relative importance of these compartments varying across local epidemiological settings.

Additional phytoplasma diversity detected exclusively in hazelnut indicates that this host harboured 16SrV phytoplasma *map* genotypes not observed in grapevine in the analysed dataset. These included the previously described genotype M48 and the newly

described genotypes M159–M162. The presence of several *map* genotypes within individual orchards suggests repeated introductions from surrounding vegetation rather than the spread of a single lineage. Their epidemiological interpretation varies by VmpA type, as VmpA encodes a membrane protein involved in adhesion to insect vector cells and is used to infer compatibility with vector transmission (Malembic-Maher et al., 2020). M48 had previously been assigned to the VmpA-I vectotype, which is associated with the alder–Macropsinae cycle rather than the grapevine–*S. titanus* cycle (Malembic-Maher et al., 2020). Its detection in hazelnut is therefore consistent with circulation associated with non-cultivated plants. For the newly described hazelnut genotypes M159–M162, VmpA analysis in this study placed them in the VmpA-III cluster, which includes isolates compatible with transmission by *S. titanus* (Malembic-Maher et al., 2020). Thus, hazelnut harboured both a previously described VmpA-I genotype linked to the alder–Macropsinae cycle and newly described VmpA-III genotypes with genetic features compatible with transmission by *S. titanus*.

The vector data further support the distinction between the grapevine–*S. titanus* cycle and the more diverse *map* genotype pattern observed in hazelnut orchards. The two leafhopper species carried distinct *map* genotype sets: *S. titanus* harboured only M54, whereas *O. ishidae* carried M38, M50, M122, and M160. Field surveys have detected 16SrV phytoplasmas in *O. ishidae* collected from different host plants, showing natural acquisition but not, by itself, proving transmission (Casati et al., 2017; Lessio et al., 2016). *O. ishidae* can complete its life cycle on the grapevine (Desqué et al., 2019; Lessio et al., 2019). Experimental transmission to grapevine has been shown after forced acquisition of 16SrV phytoplasmas from infected broad bean (Lessio et al., 2016). This supports a possible contributory role of *O. ishidae* in 16SrV phytoplasma spread, although transmission from hazelnut to grapevine has not been demonstrated.

These results support hypotheses 1–3: *map* genotypes with epidemic potential in grapevine were also detected in hazelnut, hazelnut and alder harboured additional *map* genotypes associated with non-cultivated hosts, and the overlap between hazelnut and *O. ishidae* is consistent with a possible contributory role of this species in 16SrV phytoplasma circulation in hazelnut orchards.

3.2 Optimised Enrichment Strategy for Phytoplasma Genome Sequencing from Naturally Infected Plant Material

The fourth publication of this thesis addressed the methodological step required to transition from marker-based epidemiology to genome-wide comparison of 16SrV phytoplasmas. As phytoplasmas cannot be cultured in cell-free media and are present in naturally infected tissues together with a large excess of host DNA, direct genome sequencing from field material remains technically demanding (Rodrigues Jardim et al., 2022; Saracco et al., 2006; Zhang et al., 2023). Although several enrichment strategies have been described, their use in direct genome sequencing from naturally infected woody hosts has remained limited, particularly in terms of both sequencing output and practical applicability.

The protocol comparison showed that the limiting factor was not primarily the extraction chemistry, but the biological composition of the sample. When host DNA remained dominant, changes in DNA extraction alone had little effect on phytoplasma recovery or assembly quality. The improvement became evident only after enrichment steps were introduced, especially when differential centrifugation was combined with depletion

of methylated host DNA. Performance improved stepwise, with the largest gain achieved when pre-extraction enrichment was combined with post-extraction host DNA depletion. These findings are consistent with earlier reports demonstrating enrichment of phytoplasma DNA by physical separation (Palmano, 2001) and selective host DNA removal (Debonneville et al., 2022; Feehery et al., 2013; Nijo et al., 2021) but extend these observations to genome-scale sequencing of naturally infected material.

At comparable dataset sizes, assemblies from protocols without enrichment covered substantially less of the FDp reference genome, whereas the optimised protocol achieved complete FDp reference genome coverage at the normalised dataset size and reached 100% coverage at approximately 3 billion nucleotides in the rarefaction analysis. This result demonstrates the practical value of enrichment: after host DNA reduction, sequencing depth was used more effectively to recover the phytoplasma genome. Short-read Illumina sequencing data yielded the most accurate assemblies at comparable dataset sizes, whereas long-read ONT sequencing produced longer reads and longer individual contigs. Hybrid assembly combined some of these advantages, yielding improved contiguity, but did not consistently outperform Illumina-only assemblies from the optimised protocol in reference genome coverage.

Whole-genome amplification was used in one protocol because host DNA depletion left too little DNA for the needed ONT library preparation. In the tested sample, this step did not substantially change the FDp reference coverage or the proportion of reads assigned to the phytoplasma genome. It therefore provided enough input DNA for long-read sequencing without a detectable loss of genome representation. Combining Illumina and ONT sequencing data yielded a high-quality draft that spanned most of the FDp reference genome, demonstrating that genome-scale data can be recovered directly from field-collected grapevine material. However, the detection of chimeric phytoplasma–host contigs by BLAST analysis shows that hybrid assemblies require careful post-assembly validation.

To assess whether the enrichment effect was reproducible beyond the sequenced grapevine sample, additional grapevine and hazelnut samples were evaluated by qPCR. Both enrichment steps reduced host DNA levels, while the 16SrV signal in qPCR remained approximately stable. This indicates that host DNA was reduced without a comparable loss of phytoplasma DNA. Comparable effects were observed in grapevine and hazelnut samples, suggesting that the approach is not restricted to a single host type. However, this conclusion is based on qPCR data, while sequencing-based evaluation was performed only for grapevine. From a practical perspective, the optimised protocol reduces the need to propagate phytoplasma in test plants, a time-consuming step that requires specialised facilities, particularly for quarantine organisms such as FDp. Within this thesis, this was essential because it made naturally infected field material suitable for genome comparison and enabled testing of marker-based relationships between grapevine and hazelnut isolates at a broader genomic level. These results support hypothesis 4: direct enrichment of FDp DNA from field-collected grapevine material generated sequencing data of sufficient quality for draft genome assembly without prior propagation in test plants. Of note, the hypothesis is supported for the tested grapevine material, while its broader application to hazelnut is supported by qPCR-based evidence of host DNA reduction and will require future sequencing-based validation.

3.3 Genome-Wide Relationships of 16SrV Phytoplasmas Across Host Plants

The third objective of this thesis addressed a limitation of marker-based epidemiology: single-locus typing can identify shared *map* genotypes across hosts but cannot determine

whether this similarity extends to the genome level. Molecular epidemiological studies of FDp have primarily used variation in the *map* gene to distinguish closely related strains (Arnaud et al., 2007; Malembic-Maher et al., 2020). Identical *map* genotypes have been reported from grapevine, alternative hosts such as alder and *C. vitalba*, and insect vectors (Plavec et al., 2019; Rossi et al., 2019). For the host comparisons addressed in this thesis, the unresolved question was whether such shared *map* genotypes represent closely related phytoplasma populations across hosts or whether they mask host-associated genomic differentiation.

Genome comparisons in the fifth manuscript showed that, for the analysed *map* genotypes M38 and M50, patterns of genome-wide similarity were consistent with *map* genotype classification. This correspondence was observed in both core gene phylogeny and genome-wide similarity analyses, indicating that the result was not method-specific. Samples from grapevine, hazelnut, and leafhoppers were not grouped by host in any of the analyses. This observation extends previous marker-based studies by supporting the *map*-based relationships with genome-wide data (Arnaud et al., 2007; Malembic-Maher et al., 2020; Rossi et al., 2019).

Within the context of the thesis, this finding directly informs the interpretation of the epidemiological patterns described in the first objective. The detection of identical *map* genotypes in grapevine, hazelnut, and *O. ishidae* suggested possible movement between host compartments, as shown in the first publication. The genome-wide data indicate that, at least for M38 and M50, this overlap is not restricted to the marker locus but is consistent with broader genomic similarity. This supports the interpretation of shared phytoplasma populations across multiple hosts, in line with ecological frameworks that place FDp within both cultivated and non-cultivated host plants (Malembic-Maher et al., 2020). Because *map* genotyping is widely used in FD epidemiology (Wei & Zhao, 2022), this genome-wide correspondence strengthens the interpretation of surveillance data for these lineages.

The genome-wide evidence of cross-host relatedness is also relevant to disease management. In Slovenia, control measures for FD focus on the grapevine and its primary vector, *S. titanus*. However, both M38 and M50 have been associated with grapevine outbreaks and were also detected in symptomatic cultivated hazelnut. Although the present data do not resolve transmission pathways, their occurrence in hazelnut suggests that these hosts may contribute to the persistence of epidemiologically relevant *map* genotypes outside vineyards. This supports the need to consider systematic monitoring of 16SrV phytoplasmas in hazelnut as part of a broader surveillance framework.

Genome-wide data also provided additional resolution to the analysed population structure. The similarity patterns did not show consistent host-associated separation, indicating that host identity was not the main source of genome-wide differentiation. Finer-scale variation was more consistent with geographic origin, suggesting an influence of local transmission. This is consistent with previous observations that FDp diversity reflects local circulation and interactions with surrounding vegetation (Rossi et al., 2019). However, the extent to which host-associated differences could be assessed was constrained by the material available for genome sequencing. The genome comparisons in this section of the thesis were conducted using material available from the epidemiological study. Specifically, for grapevine samples, fresh or frozen plant material of these selected *map* genotypes was no longer accessible, so the analyses relied on stored DNA previously extracted using KingFisher protocol, which is not optimal for subsequent genome sequencing. Because this material had not undergone pre-extraction enrichment, such as differential centrifugation, and its DNA quality was lower than that of CTAB extraction, it was less suitable for subsequent genome sequencing. As a result, the available assemblies were fragmented draft genomes rather than complete closed genomes, which limited reliable assessment of host-specific gene content or accessory regions. Therefore, the analysis was most suitable for

comparing overall genome-wide similarity and broad population structure, rather than for detecting individual host-associated genes.

Network-based phylogenetic analyses added another layer to the genome-wide comparison by showing that relationships among some isolates were not strictly tree-like. Similar reticulate patterns have been reported in phytoplasma genomes and have been linked to genome rearrangements, mobile elements, or recombination (Bai et al., 2006; Wei et al., 2008). In this dataset, such patterns were interpreted cautiously, because incomplete assemblies and possible mixed infections, particularly in insect-derived samples exposed to multiple infected hosts, may also contribute to these signals (Malembic-Maher et al., 2020; Rossi et al., 2019). Importantly, they did not alter the main finding that genome-wide similarity patterns corresponded to *map* genotype classification and did not separate according to host species.

These results support hypothesis 5, as isolates representing the same *map* genotype in grapevine and hazelnut showed high genome-wide similarity. This confirms that, for the analysed genotypes M38 and M50, shared *map* genotype classification across hosts was not limited to the marker locus but was consistent with broader genome-wide relatedness. Hypothesis 6 was partially supported: isolates did not separate by host plant, and host-associated differences were smaller than those between the analysed *map* genotypes. However, finer-scale variation within *map* genotypes was more consistent with geographic origin than with host plant. In addition, the available assemblies were fragmented draft genomes rather than complete, closed genomes, which meant that the analysis could assess overall genome-wide similarity but could not reliably compare host-specific genes or accessory genome content. Thus, the data do not support a simple host-associated separation of grapevine and hazelnut isolates. However, they indicate that local population structure contributes to genomic differentiation within shared *map* genotypes.

Chapter 4

Conclusions

In this thesis, we examined the epidemiology and genomic diversity of 16SrV phytoplasmas in grapevine, hazelnut, and associated leafhoppers by combining gene-based sequence analyses, development of an enrichment approach for phytoplasma genome sequencing, and genome-wide comparisons. This combination of approaches clarifies how 16SrV phytoplasmas were distributed across the analysed hosts and provides a basis for interpreting selected cross-host relationships at the genome level.

The evidence from this thesis supports the interpretation of FD-related 16SrV phytoplasmas as part of a shared multi-host system in the studied Slovenian and European contexts, in which only a subset of *map* genotypes is associated with grapevine epidemics. Gene-based comparisons across grapevine, hazelnut, and additional hosts showed that, in the analysed material, identical *map* genotypes can occur in different host compartments. At the same time, other lineages were detected only in non-cultivated hosts and were not necessarily associated with disease expression. The detection of shared *map* genotypes in hazelnut and *O. ishidae* is consistent with a possible contribution of this vector to the circulation of 16SrV phytoplasmas within hazelnut orchards. Observations from Slovenia, together with findings from the Czech Republic and the Netherlands, further support that the analysed 16SrV phytoplasma genotypes are not confined to a single host species but can occur across multiple hosts. In this context, hazelnut should be regarded as part of this system rather than as an isolated host. It may contribute to the maintenance of *map* genotypes also associated with grapevine epidemics.

However, the use of a single genetic marker does not allow assessment of whether FDp-related isolates detected in different hosts are similar across the genome. Addressing this limitation required a methodological advance. The optimised enrichment approach for HTS developed in this thesis enabled the direct recovery of FDp genome data from field-collected grapevine material without prior propagation in test plants. This provided a practical basis for extending phytoplasma research beyond marker-based typing and for making genome-wide comparisons of isolates from different hosts feasible.

Genome-scale analysis then tested this relationship directly for *map* genotypes M38 and M50. For M38 and M50, genome-wide similarity patterns were consistent with *map* genotype classification. At the same time, variation within these groups was not associated with host species and instead indicated an influence of geographic origin. In this context, genome-wide analysis supported the relevance of *map* genotype classification for the analysed lineages. It resolved additional variation within genotypes that single-locus data did not capture. This provides a more precise basis for interpreting phytoplasma relationships across the sampled hosts and regions.

Overall, the results support hypotheses 1–5, while hypothesis 6 is partially supported. Genome-wide differences between isolates from different host plants were smaller than

differences between the analysed *map* genotypes, and host-associated differentiation was not the main source of genome-wide variation. However, fragmented draft genomes prevented a reliable assessment of host-specific gene content. The scope of these conclusions is limited by the number of analysed *map* genotypes, the use of draft genome assemblies, and the geographical range of sampling; broader generalisation will require additional genome data from more hosts, vectors, and regions.

The results of this thesis show that, in the studied systems, FD-related 16SrV phytoplasmas occur across multiple hosts, with only a subset of *map* genotypes repeatedly associated with grapevine outbreaks. At the same time, additional diversity is present in non-cultivated hosts and the cultivated hazelnut. Gene-based analyses captured these broad relationships, whereas genome-wide comparisons provided higher resolution for the analysed genotypes M38 and M50. The enrichment approach developed in this work enabled such comparisons directly from naturally infected material and enables the inclusion of a wider range of isolates in future genomic analyses.

Appendix A

Supplementary Material of Included Publications

A.1 Supplementary Material for Publication 2.1

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1217425/full#supplementary-material

A.2 Supplementary Material for Publication 2.2

The Supplementary Material for this article can be found online at: https://apsjournals.apsnet.org/doi/suppl/10.1094/PDIS-02-23-0315-PDN/suppl_file/PDIS-02-23-0315-PDN.sf1.pdf

A.3 Supplementary Material for Publication 2.4

The Supplementary Material for this article can be found online at: https://www.microbiologyresearch.org/docserver/fulltext/mgen/11/9/001514_1.pdf?expires=1778245186&id=id&accname=guest&checksum=D2AFFC9C639C7326AD28281AB805BE4F

A.4 Supplementary Material for Publication 2.5

The manuscript "Genome-wide comparisons confirm locally shared Flavescence dorée-related phytoplasma populations across grapevine, hazelnut, and leafhoppers" was under review at the time of thesis submission; therefore, the final publisher-hosted supplementary material was not yet available. Supplementary Tables S1–S3 and Supplementary Figure S1 are included below because they support the interpretation of the manuscript. Large sequence files and workflow files are not reproduced in the thesis and are available through the repositories listed in the manuscript.

Supplementary Table 1: Metadata for all samples included in this study. Each sample is defined by laboratory ID, host species, sampling area, sampling year, sampled tissue, and map genotype determined by Sanger sequencing prior to this study. Sample names are composed of host (G = grapevine, H = hazelnut, S = Scaphoideus titanus, O = Orientus

ishidae), sampling area (NM = Novo Mesto, S = Slovenska Bistrica, PT = Ptuj, LT = Ljutomer), and sampling year, followed by a sequential number where multiple samples were collected at the same location in the same year.

Sample	Lab ID	Host	Sampling area	Sampling year	Sample type	*Map* genotype
G-SB-23-1	D1082/23	grapevine	SB	2023	leaf veins	M38
G-SB-23-2	D1083/23	grapevine	SB	2023	leaf veins	M38
G-NM-23-1	D1754/23	grapevine	NM	2023	leaf veins	M38
G-NM-23-2	D1755/23	grapevine	NM	2023	leaf veins	M38
G-NM-23-3	D1758/23	grapevine	NM	2023	leaf veins	M38
G-NM-23-4	D1761/23	grapevine	NM	2023	leaf veins	M38
G-LT-23	D915/23	grapevine	LT	2023	leaf veins	M38
G-LT-19	D887/19	grapevine	LT	2019	leaf veins	M50
G-PT-19	D1208/19	grapevine	PT	2019	leaf veins	M50
H-SB-20-1	D1446/20	hazelnut	SB	2020	root	M38
H-SB-20-2	D1461/20	hazelnut	SB	2020	root	M38
H-SB-20-3	D1459/20	hazelnut	SB	2020	root	M50
H-SB-21-1	SB21-1-P	hazelnut	SB	2021	leaf veins	M38
H-SB-21-2	SB21-8-P	hazelnut	SB	2021	leaf veins	M50
O-SB-21-1	SB21-OI-7	*O. ishidae*	SB	2021	1 imago	M50
O-SB-21-2	SB21-OI-36	*O. ishidae*	SB	2021	1 imago	M38
S-NM-24	D917/24	*S. titanus*	NM	2024	6 imago	M38
S-PT-24	D1287/24	*S. titanus*	PT	2024	7 imago	M38

Supplementary Table 2: qPCR-based assessment of enrichment efficiency and DNA concentration for all samples included in this study. Each sample is defined by a sample identifier, host species, map genotype, sample tissue type, and DNA isolation method. DNA concentration after enrichment is reported in ng/µl. Results are shown for the 18S rRNA assay (host DNA) and the 16SrV-specific assay (phytoplasma DNA), including Cq values before and after enrichment and the corresponding Cq differences for each assay.

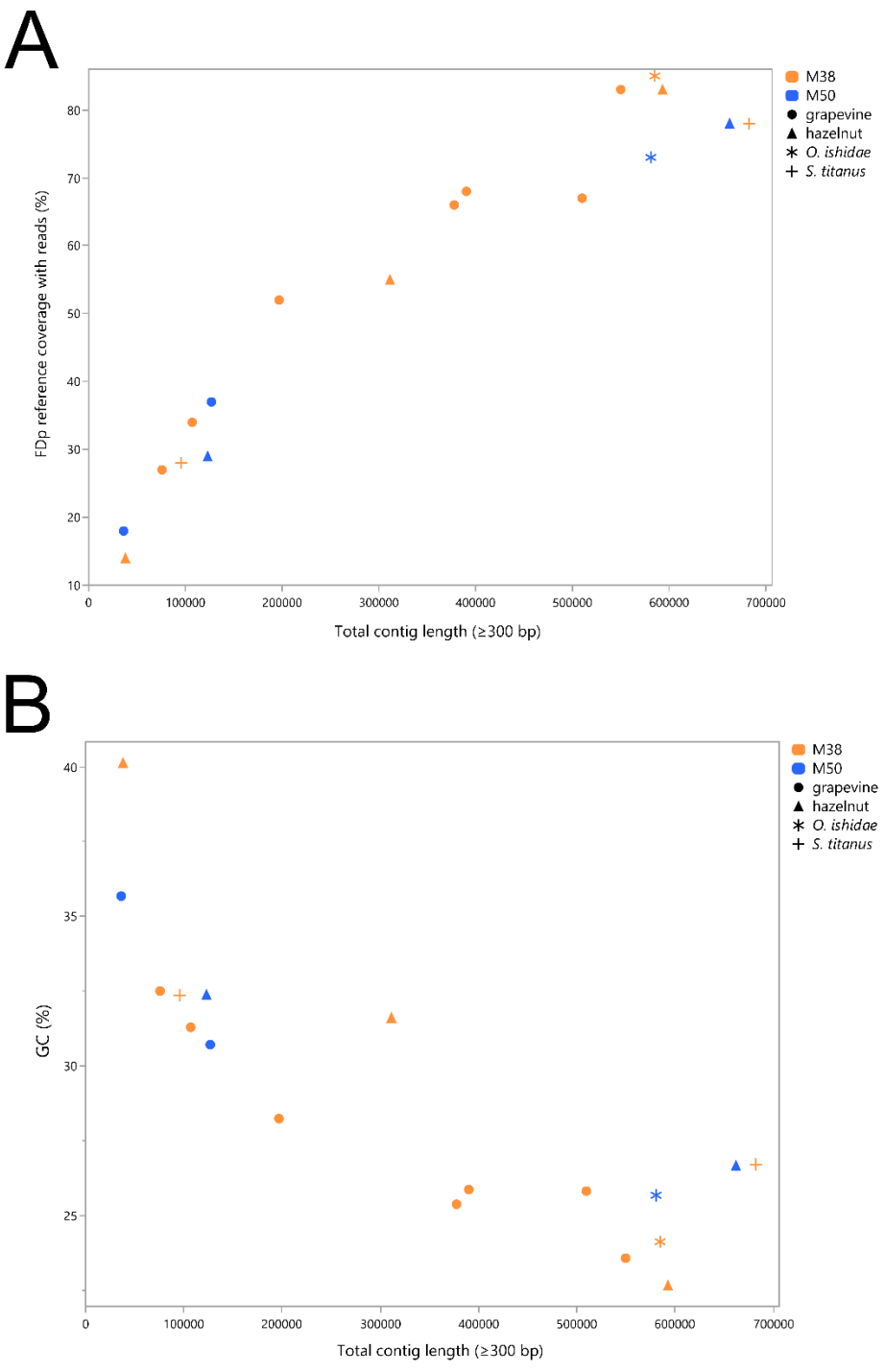
Sample	Host	*Map* genotype	Sample type	DNA isolation	DNA concentration after enrichment [ng/µl]	18S rRNA assay before enrichment [Cq]	18S rRNA assay after enrichment [Cq]	Cq difference for 18S rRNA assay	16SrV specific assay before enrichment [Cq]	16SrV specific assay after enrichment [Cq]	Cq difference for 16SrV specific assay
G-SB-23-1	grapevine	M38	leaf veins	KingFisher	0.9	18.7	26.9	8.2	31.5	31.6	0.1
G-SB-23-2	grapevine	M38	leaf veins	KingFisher	1.2	18.2	25.9	7.7	30.5	31.5	1.0
G-NM-23-1	grapevine	M38	leaf veins	KingFisher	1.4	21.3	27.1	5.8	29.6	29.7	0.2
G-NM-23-2	grapevine	M38	leaf veins	KingFisher	1.9	16.5	25.8	9.3	24.8	27.0	2.1
G-NM-23-3	grapevine	M38	leaf veins	KingFisher	6.5	17.7	24.3	6.6	28.3	28.8	0.4
G-NM-23-4	grapevine	M38	leaf veins	KingFisher	7.5	17.2	24.3	7.1	26.2	26.8	0.6
G-LT-23	grapevine	M38	leaf veins	KingFisher	1.9	19.3	27.4	8.1	30.0	30.2	0.1
G-LT-19	grapevine	M50	leaf veins	KingFisher	1.4	17.1	26.5	9.4	27.4	28.9	1.5
G-PT-19	grapevine	M50	leaf veins	KingFisher	1.5	17.5	24.9	7.4	30.1	31.6	1.5

H-SB-20-1	hazelnut	M38	root	differential centrifugation and CTAB	0.7	35.1	28.2	-6.9	35.6	31.5	-4.1
H-SB-20-2	hazelnut	M38	root	differential centrifugation and CTAB	6.5	23.2	26.7	3.5	31.5	30.2	-1.3
H-SB-20-3	hazelnut	M50	root	differential centrifugation and CTAB	1.6	38.1	25.9	-12.2	36.7	32.3	-4.5
H-SB-21-1	hazelnut	M38	leaf veins	differential centrifugation and CTAB	4.7	19.4	23.1	3.6	28.5	28.8	0.3
H-SB-21-2	hazelnut	M50	leaf veins	differential centrifugation and CTAB	1.6	20.5	19.2	-1.2	30.2	26.8	-3.4
O-SB-21-1	*O. ishidae*	M50	1 imago	KingFisher	2.8	23.5	24.5	1.0	25.3	26.7	1.3
O-SB-21-2	*O. ishidae*	M38	1 imago	OmniPrep for plant	4.6	17.8	21.2	3.4	19.0	22.2	3.1
S-NM-24	*S. titanus*	M38	6 imago	KingFisher	3.0	24.5	25.3	0.7	25.5	26.3	0.8
S-PT-24	*S. titanus*	M38	7 imago	KingFisher	3.6	22.9	24.8	1.9	29.3	30.6	1.3

Supplementary Table 3: Sample identifiers and corresponding ENA accession numbers for sequencing data generated in this study under BioProject accession PRJEB111790. For each sample, the table includes the BioSample accession, the raw sequencing read accession numbers, and the associated draft genome assembly accession numbers. For samples represented by multiple sequencing runs, all corresponding ENA run accessions are listed.

Sample	BioSample accession	Raw sequencing read accession(s)	Assembly accession
G-SB-23-1	ERS29724434	ERR17336863, ERR17336864	GCA_983199565
G-SB-23-2	ERS29724435	ERR17336857, ERR17336858, ERR17336859	GCA_983197445
G-NM-23-1	ERS29724436	ERR17037494	GCA_983216095
G-NM-23-2	ERS29724437	ERR17037495	GCA_983204435
G-NM-23-3	ERS29724438	ERR17336849, ERR17336850	GCA_983204275
G-NM-23-4	ERS29724439	ERR17336851, ERR17336852, ERR17336853	GCA_983216105
G-LT-23	ERS29724440	ERR17336868, ERR17336869, ERR17336870	GCA_983198065
G-LT-19	ERS29724441	ERR17336860, ERR17336861, ERR17336862	GCA_983202855
G-PT-19	ERS29724442	ERR17336874, ERR17336875, ERR17336876	GCA_983204195
H-SB-20-1	ERS29724443	ERR17336871, ERR17336872, ERR17336873	GCA_983207345
H-SB-20-2	ERS29724444	ERR17336854, ERR17336855, ERR17336856	GCA_983203905
H-SB-20-3	ERS29724445	ERR17336865, ERR17336866, ERR17336867	GCA_983198385
H-SB-21-1	ERS29724446	ERR17216624, ERR17216654, ERR17216696	GCA_983198345
H-SB-21-2	ERS29724447	ERR17214383, ERR17215134	GCA_983203915
O-SB-21-1	ERS29724448	ERR17038918	GCA_983207365

O-SB-21-2	ERS29724449	ERR17038927	GCA_983198405
S-NM-24	ERS29724450	ERR17038931	GCA_983197805
S-PT-24	ERS29724451	ERR17038946	GCA_983198355



Supplementary Figure 1: Sequencing parameters plots. (A) Percentage of the FDp isolate CH as reference genome (CP097583) covered by mapped sequencing reads plotted against the total length of assembled contigs longer than 300 bp for each sample. Points are coloured by host species. (B) GC content (%) of assembled contigs plotted against the total length of assembled contigs longer than 300 bp for each sample. Points are coloured by host species.

References

Andersen, M. T., Liefting, L. W., Havukkala, I., & Beever, R. E. (2013). Comparison of the complete genome sequence of two closely related isolates of "Candidatus Phytoplasma australiense" reveals genome plasticity. *BMC Genomics*, *14*(1), 1–15. https://doi.org/10.1186/1471-2164-14-529

Angelini, E., Squizzato, F., Lucchetta, G., & Borgo, M. (2004). Detection of a phytoplasma associated with grapevine Flavescence dorée in Clematis vitalba. In *European Journal of Plant Pathology* (Vol. 110).

Arnaud, G., Malembic-Maher, S., Salar, P., Bonnet, P., Maixner, M., Marcone, C., Boudon-Padieu, E., & Foissac, X. (2007). Multilocus sequence typing confirms the close genetic interrelatedness of three distinct flavescence dorée phytoplasma strain clusters and group 16SrV phytoplasmas infecting grapevine and alder in Europe. *Applied and Environmental Microbiology*, *73*(12), 4001–4010. https://doi.org/10.1128/AEM.02323-06

Arricau-Bouvery, N., Duret, S., Dubrana, M. P., Batailler, B., Desqué, D., Béven, L., Danet, J. L., Monticone, M., Bosco, D., Malembic-Maher, S., & Foissac, X. (2018). Variable membrane protein A of flavescence dorée phytoplasma binds the midgut perimicrovillar membrane of Euscelidius variegatus and promotes adhesion to its epithelial cells. *Applied and Environmental Microbiology*, *84*(8). https://doi.org/10.1128/AEM.02487-17

Arricau-Bouvery, N., Duret, S., Dubrana, M. P., Desqué, D., Eveillard, S., Brocard, L., Malembic-Maher, S., & Foissac, X. (2021). Interactions between the flavescence dorée phytoplasma and its insect vector indicate lectin-type adhesion mediated by the adhesin VmpA. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-90809-z

Bai, X., Zhang, J., Ewing, A., Miller, S. A., Radek, A. J., Shevchenko, D. V., Tsukerman, K., Walunas, T., Lapidus, A., Campbell, J. W., & Hogenhout, S. A. (2006). Living with genome instability: The adaptation of phytoplasmas to diverse environments of their insect and plant hosts. *Journal of Bacteriology*, *188*(10), 3682–3696. https://doi.org/10.1128/JB.188.10.3682-3696.2006

Bertaccini, A. (2007). Phytoplasmas: diversity, taxonomy, and epidemiology. *Frontiers in Bioscience*, *12*, 673–689.

Bertaccini, A., & Duduk, B. (2009). Phytoplasma and phytoplasma diseases: a review of recent research. *Phytopathol. Mediterr.*, (48), 355–378.

Bertin, S., Guglielmino, C. R., Karam, N., Gomulski, L. M., Malacrida, A. R., & Gasperi, G. (2007). Diffusion of the Nearctic leafhopper Scaphoideus titanus Ball in Europe: A consequence of human trading activity. *Genetica*, *131*(3), 275–285. https://doi.org/10.1007/s10709-006-9137-y

Bianco, P. A., Romanazzi, G., Mori, N., Myrie, W., & Bertaccini, A. (2019). Integrated Management of Phytoplasma Diseases. In A. Bertaccini, P. G. Weintraub, G. P. Rao, & N. Mori (Eds.), *Phytoplasmas: Plant Pathogenic Bacteria - II: Transmission and Management of Phytoplasma - Associated Diseases* (pp. 237–258). Springer Singapore. https://doi.org/10.1007/978-981-13-2832-9_11

Bonfils, J., & Schvester, D. (1960). Les cicadelles (Homoptera Auchenorhyncha) dans leurs rapports avec la vigne dans le Sud-Ouest de la France. *Ann Epiphyt*, (3), 325–336.

Casati, P., Jermini, M., Quaglino, F., Corbani, G., Schaerer, S., Passera, A., & Bianco, P. A. (2017). New insights on Flavescence dorée phytoplasma ecology in the vineyard agro-ecosystem in southern Switzerland. *Annals of Applied Biology*, *171*, 37–51. https://doi.org/10.1111/aab.12359

Cho, S., Zwolin, A., Hogenhout, S. A., & Kuo, C. (2020). Complete Genome Sequence of " Candidatus Phytoplasma asteris " RP166 , a Plant Pathogen Associated with Rapeseed Phyllody Disease in Poland. *Microbiology Resource Announcements*, *0*(35), 1–2.

Chuche, J., & Thiéry, D. (2014). Biology and ecology of the Flavescence dorée vector Scaphoideus titanus: A review. *Agronomy for Sustainable Development*, *34*(2), 381–403. https://doi.org/10.1007/s13593-014-0208-7

Ciccotti, A. M., Bianchedi, P. L., Bragagna, P., Deromedi, M., Filippi, M., Forno, F., & Mattedi, L. (2007). Transmission of "Candidatus Phytoplasma mali" by root bridges under natural and experimental conditions. *Bulletin of Insectology*, *60*(2), 387–388.

Constable, F. E. (2009). Phytoplasma epidemiology: grapevines as a model. In P. G. Weintraub & P. Jones (Eds.), *Phytoplasmas: Genomes, Plant Hosts and Vectors* (pp. 188–212). Cabi.

Credi, R. (1989). Flavescenza dorata della Vite in Emilia-Romagna: evoluzione della malattia nelle piante e suoi effetti sulla produzione e sullo sviluppo vegetativo. *Phytopathologia Mediterranea*, *28*(2), 113–121. http://www.jstor.org/stable/42685688

Cvrković, T., Jović, J., Jakovljević, M., Krstić, O., Marinković, S., Mitrović, M., & Toševski, I. (2021). The "code red" for balkan vineyards: Occurrence of orientus ishidae (matsumura, 1902) (hemiptera: Cicadellidae) in Serbia. *BioInvasions Records*, *10*(3), 544–554. https://doi.org/10.3391/BIR.2021.10.3.04

De Maio, N., Shaw, L. P., Hubbard, A., George, S., Sanderson, N. D., Swann, J., Wick, R., Oun, M. A., Stubberfield, E., Hoosdally, S. J., Crook, D. W., Peto, T. E. A., Sheppard, A. E., Bailey, M. J., Read, D. S., Anjum, M. F., Sarah Walker, A., & Stoesser, N. (2019). Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microbial Genomics*, *5*(9). https://doi.org/10.1099/mgen.0.000294

Debonneville, C., Mandelli, L., Brodard, J., Groux, R., Roquis, D., & Schumpp, O. (2022). The Complete Genome of the "Flavescence Dorée" Phytoplasma Reveals

Characteristics of Low Genome Plasticity. *Biology*, *11*(7). https://doi.org/10.3390/biology11070953

Desqué, D., Salar, P., Danet, J., Lusseau, T., Garcion, C., Moreau, E., Dubus, C., Dureuil, J., Delbac, L., Binet, D., Auriol, A., Thiéry, D., Foissac, X., & Malembic-Maher, S. (2019). Impact of Orientus ishidae on "flavescence dorée" emergence in the vineyards of riparian ecosystems. *Phytopathogenic Mollicutes*, *9*(1), 69–70.

Dickinson, M., Tuffen, M., & Hodgetts, J. (2013). The Phytoplasmas: an Introduction. In M. Dickinson & J. Hodgetts (Eds.), *Phytoplasma: Methods and Protocols* (Vol. 938, pp. 1–14). Springer Science+Business Media.

Ember, lbolya, Acs, Z., Salar, P., Danet, J.-L., Foissac, X., & Kolber, M. (2011). Survey and genetic diversity of phytoplasmas from the 16SrV-C and-D subgroups in Hungary. *Bulletin of Insectology*, *64*, 33–34. https://cabidigitallibrary.org

EU Comission. (2019, November 28). *Commission Implementing Regulation (EU) 2019/2072*. EU Regulations.

Feehery, G. R., Yigit, E., Oyola, S. O., Langhorst, B. W., Schmidt, V. T., Stewart, F. J., Dimalanta, E. T., Amaral-Zettler, L. A., Davis, T., Quail, M. A., & Pradhan, S. (2013). A Method for Selectively Enriching Microbial DNA from Contaminating Vertebrate Host DNA. *PLoS ONE*, *8*(10). https://doi.org/10.1371/journal.pone.0076096

Filippin, L., & De Pra, V. (2011). Nucleotide sequencing of imp gene in phytoplasmas associated to 'flavescence dorée' from Ailanthus altissima. *Bulletin of Insectology*, *64*, S49–S50. www.bulletinofinsectology.org

Filippin, L., Jović, J., Cvrković, T., Forte, V., Clair, D., Toševski, I., Boudon-Padieu, E., Borgo, M., & Angelini, E. (2009). Molecular characteristics of phytoplasmas associated with Flavescence dorée in clematis and grapevine and preliminary results on the role of Dictyophara europaea as a vector. *Plant Pathology*, *58*(5), 826–837. https://doi.org/https://doi.org/10.1111/j.1365-3059.2009.02092.x

Firrao, G., Garcia-Chapa, M., & Marzachì, C. (2007). Phytoplasmas: Genetics, Diagnosis and Relationships with the Plant and Insect Host. In *Frontiers in Bioscience* (Vol. 12).

Gaffuri, F., Sacchi, S., & Cavagna, B. (2011). First detection of the mosaic leafhopper, Orientus ishidae , in northern Italian vineyards infected by the flavescence dorée phytoplasma . *New Disease Reports*, *24*(1), 22–22. https://doi.org/10.5197/j.2044-0588.2011.024.022

Hogenhout, S. A., Oshima, K., Ammar, E. D., Kakizawa, S., Kingdom, H. N., & Namba, S. (2008). Phytoplasmas: Bacteria that manipulate plants and insects. In *Molecular Plant Pathology* (Vol. 9, Number 4, pp. 403–423). https://doi.org/10.1111/j.1364-3703.2008.00472.x

Holz, S., Duduk, B., Büttner, C., & Kube, M. (2016). Genetic variability of Alder yellows phytoplasma in Alnus glutinosa in its natural Spreewald habitat. *Forest Pathology*, *46*(1), 11–21. https://doi.org/10.1111/efp.12206

Huang, C. T., Cho, S. T., Lin, Y. C., Tan, C. M., Chiu, Y. C., Yang, J. Y., & Kuo, C. H. (2022). Comparative Genome Analysis of 'Candidatus Phytoplasma luffae' Reveals

the Influential Roles of Potential Mobile Units in Phytoplasma Evolution. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.773608

IRPCM. (2004). "Candidatus Phytoplasma", a taxon for the wall-less, non-helical prokaryotes that colonize plant phloem and insects. *International Journal of Systematic and Evolutionary Microbiology*, *54*(4), 1243–1255. https://doi.org/10.1099/ijs.0.02854-0

Jardim, B. R., Tran-Nguyen, L. T. T., Gambley, C., Al-Sadi, A. M., Al-Subhi, A. M., Foissac, X., Salar, P., Cai, H., Yang, J. Y., Davis, R., Jones, L., Rodoni, B., & Constable, F. E. (2023). The observation of taxonomic boundaries for the 16SrII and 16SrXXV phytoplasmas using genome-based delimitation. *International Journal of Systematic and Evolutionary Microbiology*, *73*(7). https://doi.org/10.1099/ijsem.0.005977

Jurga, M., & Zwolińska, A. (2020). Artemisia vulgaris, a new host of 16SrV-C phytoplasma related strains infecting black alder in Poland. *Journal of Phytopathology*, *168*(11–12), 659–667. https://doi.org/10.1111/jph.12946

Krstić, O., Crvković, T., Marinkovič, S., Jakovljević, M., Mitrović, M., Toševski, I., & Jović, J. (2022). Genetic Diversity of Flavescence Dorée Phytoplasmas in Vineyards of Serbia: From the Widespread Occurrence of Autochthonous Map-M51 to the Emergence of Endemic Map-FD2 (Vectotype II) and New Map-FD3 (Vectotype III) Epidemic Genotypes. *Agronomy*, *12*(448).

Krstić, O., Cvrković, T., Mitrović, M., Radonjić, S., Hrnčić, S., Toševski, I., & Jović, J. (2018). Wolbachia infection in natural populations of Dictyophara europaea, an alternative vector of grapevine Flavescence dorée phytoplasma: effects and interactions. *Annals of Applied Biology*, *172*(1), 74–64.

Ku, C., Lo, W.-S., & Kuo, C.-H. (2013). Horizontal transfer of potential mobile units in phytoplasmas. *Mobile Genetic Elements*, *3*(5), e26145. https://doi.org/10.4161/mge.26145

Kube, M., Schneider, B., Kuhl, H., Dandekar, T., Heitmann, K., Migdoll, A. M., Reinhardt, R., & Seemüller, E. (2008). The linear chromosome of the plant-pathogenic mycoplasma "Candidatus Phytoplasma mali." *BMC Genomics*, *9*, 1–14. https://doi.org/10.1186/1471-2164-9-306

Lee, I., Davis, R. E., & Gundersen-Rindal, D. E. (2000). Phytoplasma: Phytopathogenic Mollicutes. *Annual Review of Microbiology*, *54*, 221–255.

Lee, I., Gundersen-rindal, D. E., Davis, R. E., & Bartoszyk, M. (1998). Revised classification scheme of phytoplasmas based on RFLP analyses of 16s rRNA and ribosomal protein gene sequences. *International Journal of Systematic Bacteriology*, (48), 1153–1169.

Lessio, F., & Alma, A. (2004). Dispersal patterns and chromatic response of Scaphoideus titanus Ball (Homoptera Cicadellidae), vector of the phytoplasma agent of grapevine flavescence dorée. *Agricultural and Forest Entomology*, *6*(2), 121–128. https://doi.org/10.1111/j.1461-9563.2004.00212.x

Lessio, F., Bocca, F., & Alma, A. (2019). Development, spatial distribution, and presence on grapevine of nymphs of Orientus ishidae (Hemiptera: Cicadellidae), a new vector

of flavescence dorée phytoplasmas. *Journal of Economic Entomology*, *112*(6), 2558–2564. https://doi.org/10.1093/jee/toz177

Lessio, F., Picciau, L., Gonella, E., Mandrioli, M., Tota, F., & Alma, A. (2016). The mosaic leafhopper Orientus ishidae: Host plants, spatial distribution, infectivity, and transmission of 16SrV phytoplasmas to vines. *Bulletin of Insectology*, *69*(2), 277–289.

Malembic-Maher, S., Desqué, D., Khalil, D., Salar, P., Bergey, B., Danet, J.-L., Duret, S., Dubrana-Ourabah, M.-P., Beven, L., Ember, I., Acs, Z., Della Bartola, M., Materazzi, A., Filippin, L., Krnjajic, S., Krstić, O., Toševski, I., Lang, F., Jarausch, B., … Foissac, X. (2020). When a Palearctic bacterium meets a Nearctic insect vector: Genetic and ecological insights into the emergence of the grapevine Flavescence dorée epidemics in Europe. *PLOS Pathogens*, *16*(3), 1–28. https://doi.org/10.1371/journal.ppat.1007967

Marcone, C. (2010). Movement of phytoplasmas and the development of disease in the plant. In P. G. Weintraub & P. Jones (Eds.), *Phytoplasmas: Genomes, Plant Hosts and Vectors* (pp. 114–131). CAB International.

Marcone, C. (2014). Molecular biology and pathogenicity of phytoplasmas. *Annals of Applied Biology*, *165*(2), 199–221. https://doi.org/10.1111/aab.12151

Marcone, C., Pierro, R., & Palmieri, C. (2024). Occurrence, Impact, and Multilocus Sequence Analysis of Alder Yellows Phytoplasma Infecting Common Alder and Italian Alder in Southern Italy. *Microorganisms*, *12*(6). https://doi.org/10.3390/microorganisms12061140

Mehle, N., Jakoš, N., Mešl, M., Miklavc, J., Matko, B., Rot, M., Ferlež Rus, A., Brus, R., & Dermastia, M. (2019). Phytoplasmas associated with declining of hazelnut (Corylus avellana) in Slovenia. *European Journal of Plant Pathology*, *155*(4), 1117–1132. https://doi.org/10.1007/s10658-019-01839-3

Mehle, N., Ravnikar, M., Seljak, G., Knapic, V., & Dermastia, M. (2011). The most widespread phytoplasmas, vectors and measures for disease control in Slovenia. *Phytopathogenic Mollicutes*, *1*(2), 65–76.

Mehle, N., Rupar, M., Seljak, G., Ravnikar, M., & Dermastia, M. (2011). Molecular diversity of "flavescence dorée" phytoplasma strains in Slovenia. *Bulletin of Insectology*, *64*, 29–30.

Mehle, N., Seljak, G., Rupar, M., Ravnikar, M., & Dermastia, M. (2010). The first detection of a phytoplasma from the 16SrV (Elm yellows) group in the mosaic leafhopper Orientus ishidae . *New Disease Reports*, *22*(1), 11–11. https://doi.org/10.5197/j.2044-0588.2010.022.011

Miladinović, Z., Radulović, M., Hrnčić, S., & Delić, D. (2019). Genetic variability of Scaphoideus titanus populations and new host plant of 16SrV group phytoplasmas in Bosnia and Herzegovina. *Phytopathog. Mollicutes*, *9*, 111–112.

Mitrović, J., Siewert, C., Duduk, B., Hecht, J., Mölling, K., Broecker, F., Beyerlein, P., Büttner, C., Bertaccini, A., & Kube, M. (2013). Generation and Analysis of Draft Sequences of 'Stolbur'' Phytoplasma from Multiple Displacement Amplification Templates.' *Journal of Molecular Microbiology and Biotechnology*, 1–11. https://doi.org/10.1159/000353904

Musić, M. Š., Polak, B., Drčelić, M., Pei, S. C., & Kuo, C. H. (2025). Sequencing and comparative analyses of 'Candidatus Phytoplasma solani' genomes reveal diversity of effectors and potential mobile units. *Microbial Genomics*, *11*(4). https://doi.org/10.1099/mgen.0.001401

Namba, S. (2019). Molecular and biological properties of phytoplasmas. In *Proceedings of the Japan Academy Series B: Physical and Biological Sciences* (Vol. 95, Number 7, pp. 401–418). Japan Academy. https://doi.org/10.2183/pjab.95.028

Nijo, T., Iwabuchi, N., Tokuda, R., Suzuki, T., Matsumoto, O., Miyazaki, A., Maejima, K., Oshima, K., Namba, S., & Yamaji, Y. (2021). Enrichment of phytoplasma genome DNA through a methyl-CpG binding domain-mediated method for efficient genome sequencing. *Journal of General Plant Pathology*, *87*(3), 154–163. https://doi.org/10.1007/s10327-021-00993-z

Oggier, A., Conedera, M., Debonneville, C., Schumpp, O., & Rizzoli, A. (2024). Gone-wild grapevines in forests host phytoplasma genotypes linked to grapevine's flavescence dorée epidemics in cultivated vineyards and competent vectors. *Journal of Plant Pathology*. https://doi.org/10.1007/s42161-024-01775-0

Orlovskis, Z., Canale, M. C., Haryono, M., Lopes, J. R. S., Kuo, C. H., & Hogenhout, S. A. (2017). A few sequence polymorphisms among isolates of Maize bushy stunt phytoplasma associate with organ proliferation symptoms of infected maize plants. *Annals of Botany*, *119*(5), 869–884. https://doi.org/10.1093/aob/mcw213

Oshima, K., Kakizawa, S., Nishigawa, H., Jung, H. Y., Wei, W., Suzuki, S., Arashida, R., Nakata, D., Miyata, S. I., Ugaki, M., & Namba, S. (2004). Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. *Nature Genetics*, *36*(1), 27–29. https://doi.org/10.1038/ng1277

Oshima, K., Maejima, K., & Namba, S. (2013). Genomic and evolutionary aspects of phytoplasmas. *Frontiers in Microbiology*, *4*(AUG), 1–8. https://doi.org/10.3389/fmicb.2013.00230

Palmano, S. (2001). A comparison of different phytoplasma DNA extraction methods using competitive PCR. *Phytopathologia Mediterranea*, *40*(2), 99–107. https://doi.org/10.2307/26456586

Papura, D., Burban, C., Helden, M. Van, Giresse, X., Nusillard, B., Guillemaud, T., & Kerdelhué, C. (2012). Microsatellite and Mitochondrial Data Provide Evidence for a Single Major Introduction for the Neartic Leafhopper Scaphoideus titanus in Europe. *PLoS ONE*, 1–14. https://doi.org/10.1371/journal.pone.0036882

Plavec, J., Budinšćak, Križanac, I., Škorić, D., Foissac, X., & Šeruga Musić, M. (2019). Multilocus sequence typing reveals the presence of three distinct flavescence dorée phytoplasma genetic clusters in Croatian vineyards. *Plant Pathology*, *68*(1), 18–30. https://doi.org/10.1111/ppa.12934

Prezelj, N., Nikolić, P., Gruden, K., Ravnikar, M., & Dermastia, M. (2013). Spatiotemporal distribution of flavescence dorée phytoplasma in grapevine. *Plant Pathology*, *62*(4), 760–766. https://doi.org/10.1111/j.1365-3059.2012.02693.x

Reisenzein, H., & Steffek, R. (2011). First outbreaks of grapevinèflavescence doree' in Austrian viticulture. *Bulletin of Insectology*, *64*, 223–224. www.bulletinofinsectology.org

Rigamonti, I. E., Salvetti, M., Girgenti, P., Bianco, P. A., & Quaglino, F. (2023). Investigation on Flavescence Dorée in North-Western Italy Identifies Map-M54 (16SrV-D/Map-FD2) as the Only Phytoplasma Genotype in Vitis vinifera L. and Reveals the Presence of New Putative Reservoir Plants. *Biology*, *12*(9). https://doi.org/10.3390/biology12091216

Rizza, S., Rossi, M., Spallino, R. E., Tedeschi, R., Marzachì, C., Catara, V., & Tessitori, M. (2021). Ecology-based analysis of a recent association between Spartium junceum and 16SrV phytoplasma. *Plant Pathology*, *70*(2), 305–317. https://doi.org/10.1111/ppa.13300

Rizzoli, A., Belgeri, E., Jermini, M., Conedera, M., Filippin, L., & Angelini, E. (2021). Alnus glutinosa and Orientus ishidae (Matsumura , 1902) share phytoplasma genotypes linked to the ‘ Flavescence dorée ’ epidemics. *Journal of Applied Entomology*, 1–14. https://doi.org/10.1111/jen.12933

Rodrigues Jardim, B., Tran-Nguyen, L. T. T., Gambley, C., Rodoni, B., & Constable, F. E. (2022). Iodixanol density gradients as an effective phytoplasma enrichment approach to improve genome sequencing. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.937648

Rossi, M., Pegoraro, M., Ripamonti, M., Abbà, S., Bealb, D., Giraudoa, A., Veratti, F., Malembic –Maher, S., Salar, P., Boscoa, D., & Marzachìa, C. (2019). Genetic Diversity of Flavescence Dorée Phytoplasmas at the Vineyard Scale. *Applied and Environmental Microbiology*, *85*(10), 1–15.

Saracco, P., Bosco, D., Veratti, F., & Marzachì, C. (2006). Quantification over time of chrysanthemum yellows phytoplasma (16Sr-I) in leaves and roots of the host plant Chrysanthemum carinatum (Schousboe) following inoculation with its insect vector. *Physiological and Molecular Plant Pathology*, *67*(3–5), 212–219. https://doi.org/10.1016/j.pmpp.2006.02.001

Šeruga Musić, M., Škorić, D., Haluška, I., Križanac, I., Plavec, J., & Mikec, I. (2011). First Report of Flavescence Dorée-Related Phytoplasma Affecting Grapevines in Croatia. *Plant Disease*, *95*(3), 353. https://doi.org/10.1094/PDIS-09-10-0664

Siddique, A. B. M., Guthrie, J. N., Walsh, K. B., White, D. T., & Scott, P. T. (1998). *Histopathology and Within-Plant Distribution of the Phytoplasma Associated with Australian Papaya Dieback*.

Škorić, D., Šeruga Musić, M., Plavec, J., & Križanac, I. (2011). Geographical distribution of “flavescence dorée” phytoplasmas in Croatian grapevines. *Bulletin of Insectology*, *64*, 243–244.

Tan, C. M., Lin, Y. C., Li, J. R., Chien, Y. Y., Wang, C. J., Chou, L., Wang, C. W., Chiu, Y. C., Kuo, C. H., & Yang, J. Y. (2021). Accelerating Complete Phytoplasma Genome Assembly by Immunoprecipitation-Based Enrichment and MinION-Based DNA Sequencing for Comparative Analyses. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.766221

Tomkins, M., Kliot, A., Marée, A. F., & Hogenhout, S. A. (2018). A multi-layered mechanistic modelling approach to understand how effector genes extend beyond phytoplasma to modulate plant hosts, insect vectors and the environment. In *Current*

Opinion in Plant Biology (Vol. 44, pp. 39–48). Elsevier Ltd. https://doi.org/10.1016/j.pbi.2018.02.002

Tran-Nguyen, L. T. T., & Gibb, K. S. (2007). Optimizing Phytoplasma DNA Purification for Genome Analysis. *Journal of Biomolecular Techniques*, *18*, 104–112.

Tran-Nguyen, L. T. T., Kube, M., Schneider, B., Reinhardt, R., & Gibb, K. S. (2008). Comparative genome analysis of "Candidatus Phytoplasma australiense" (subgroup tuf-Australia I; rp-A) and "Ca. phytoplasma asteris" strains OY-M and AY-WB. *Journal of Bacteriology*, *190*(11), 3979–3991. https://doi.org/10.1128/JB.01301-07

Tran-Nguyen, L. T. T., & Schneider, B. (2013). *Cesium Chloride-Bisbenzimide Gradients for Separation of Phytoplasma and Plant DNA BT - Phytoplasma: Methods and Protocols* (M. Dickinson & J. Hodgetts, Eds.; pp. 381–393). Humana Press. https://doi.org/10.1007/978-1-62703-089-2_32

Wang, J., Song, L., Jiao, Q., Yang, S., Gao, R., Lu, X., & Zhou, G. (2018). Comparative genome analysis of jujube witches'-broom Phytoplasma, an obligate pathogen that causes jujube witches'-broom disease 06 Biological Sciences 0604 Genetics. *BMC Genomics*, *19*(1), 1–12. https://doi.org/10.1186/s12864-018-5075-1

Wang, R., Bai, B., Li, D., Wang, J., Huang, W., Wu, Y., & Zhao, L. (2024). Phytoplasma: A plant pathogen that cannot be ignored in agricultural production—Research progress and outlook. In *Molecular Plant Pathology* (Vol. 25, Number 2). John Wiley and Sons Inc. https://doi.org/10.1111/mpp.13437

Wei, W., Davis, R. E., Jomantiene, R., & Zhao, Y. (2008). Ancient, recurrent phage attacks and recombination shaped dynamic sequence-variable mosaics at the root of phytoplasma genome evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(33), 11827–11832. https://doi.org/10.1073/pnas.0805237105

Wei, W., & Zhao, Y. (2022). Phytoplasma Taxonomy: Nomenclature, Classification, and Identification. In *Biology* (Vol. 11, Number 8). MDPI. https://doi.org/10.3390/biology11081119

Weintraub, P. G., & Beanland, L. (2006). Insect Vectors of Phytoplasmas. *Annual Review of Entomology*, *51*(1), 91–111. https://doi.org/10.1146/annurev.ento.51.110104.151039

Xu, Z., Peng, L., Xing, P., Gao, Y., Yu, Y., Wang, T., Song, Z., Zhao, W., Cheng, Y., & Hu, Q. (2026). Recent Advances in Diagnosing and Managing Phytoplasma Diseases. In *Agronomy* (Vol. 16, Number 5). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/agronomy16050504

Zhang, R. Y., Wang, X. Y., Li, J., Shan, H. L., Li, Y. H., Huang, Y. K., & He, X. H. (2023). Complete genome sequence of "Candidatus Phytoplasma sacchari" obtained using a filter-based DNA enrichment method and Nanopore sequencing. *Frontiers in Microbiology*, *14*. https://doi.org/10.3389/fmicb.2023.1252709

Zhao, Y., Wei, W., Lee, I.-M., Shao, J., Suo, X., & Davis, R. E. (2013). The iPhyClassifier, an Interactive Online Tool for Phytoplasma Classification and Taxonomic Assignment. In M. Dickinson & J. Hodgetts (Eds.), *Phytoplasma: Methods and Protocols* (Vol. 938, pp. 329–338). Springer Science+Business Media.

Bibliography

Publications Related to the Thesis

de Koning, P. P. M., Schoen, R., Kogej Zwitter, Z., te Braak, N., Giesbers, A. K. J., Westenberg, M., Mehle, N., & Botermans, M. (2025). First Detection of a Phytoplasma From the 16SrV-C (Elm yellows) Group in Alnus glutinosa in the Netherlands. *New Disease Reports*, 52(1). https://doi.org/10.1002/ndr2.70056

Kogej Zwitter, Z., Seljak, G., Jakomin, T., Brodarič, J., Vučurović, A., Pedemay, S., Salar, P., Malembic-Maher, S., Foissac, X., & Mehle, N. (2023). Epidemiology of flavescence dorée and hazelnut decline in Slovenia: geographical distribution and genetic diversity of the associated 16SrV phytoplasmas. *Frontiers in Plant Science*, 14. https://doi.org/10.3389/fpls.2023.1217425

Kogej Zwitter, Z., Kutnjak, D., & Mehle, N. (2025). From vineyard to genome: optimized enrichment and sequencing of Flavescence dorée phytoplasma from grapevine samples. *Microbial Genomics*, 11(9). https://doi.org/10.1099/mgen.0.001514

Kogej Zwitter, Z., Mehle, N., & Kutnjak, D (2026). Genome-wide comparisons confirm locally shared Flavescence dorée-related phytoplasma populations across grapevine, hazelnut, and leafhoppers. – submitted to *Frontiers in Microbiology* (May 2026)

Linhartová, Š., Tománková, K., Országová, H., Hejlová, M., Čermák, V., Beránek, J., Mehle, N., Kogej Zwitter, Z., & Eichmeier, A. (2023). First Report of Grapevine Flavescence Dorée Phytoplasma in the Czech Republic. *In Plant Disease* (Vol. 107, Number 10, p. 3273). American Phytopathological Society. https://doi.org/10.1094/PDIS-02-23-0315-PDN

Biography

Zala Kogej Zwitter was born in Ljubljana on 17 August 1993. She completed her primary and secondary education in Ljubljana, Slovenia. She obtained her BSc in Biology from the Biotechnical Faculty, University of Ljubljana, in 2015, and her MSc in Molecular and Functional Biology from the same institution in 2018. During her Master's studies, she conducted her thesis research under the supervision of Prof. Dr Maja Ravnikar and Dr Denis Kutnjak at the Department of Biotechnology and Systems Biology, National Institute of Biology (NIB), Ljubljana, focusing on the effects of mixed infections of potato virus Y and pepino mosaic virus on virus accumulation and symptom development in different host plants. During her studies, she also completed an Erasmus+ research internship at the Instituto de Biología Molecular y Celular de Plantas in Valencia, Spain, in the Evolutionary Systems Virology Group led by Prof. Dr Santiago F. Elena, which resulted in her co-authorship of a peer-reviewed publication in Molecular Biology and Evolution. After graduation, she worked as a junior expert at NIB, where she was involved in the diagnosis of plant viruses and phytoplasmas, the development of novel molecular detection methods, and activities within the framework of the European Union Reference Laboratory for plant viruses, viroids, and phytoplasmas. She contributed to the development of molecular and sequencing-based diagnostic tools for emerging and quarantine plant pathogens, including first and shared first authorship on peer-reviewed scientific articles. In October 2020, she enrolled in the doctoral study programme Sensor Technologies at the Jožef Stefan International Postgraduate School under the mentorship of Dr Nataša Mehle and co-mentorship of Dr Denis Kutnjak at the Department of Biotechnology and Systems Biology, NIB. Her doctoral research focused on the epidemiology, genetic diversity, and genome-level characterisation of 16SrV group phytoplasmas associated with grapevine and hazelnut in Slovenia. As part of her doctoral training, she conducted a three-month research visit at the French National Research Institute for Agriculture, Food and Environment (INRAE) in Bordeaux in the group of Dr Xavier Foissac. She has presented her research at several national and international scientific conferences and professional symposia, including receiving the Third Place Poster Presentation Award at the 16th International Symposium on Plant Virus Epidemiology (São Paulo, Brazil, 30 June–3 July 2025), and is actively engaged in science communication and outreach activities related to plant health and plant pathology.