MULTI-TARGET APPROACHES FOR DETECTION AND QUANTIFICATION OF GENETICALLY MODIFIED ORGANISMS

Alexandra Bogožalec Košir

Doctoral Dissertation
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia

Supervisor: Prof. Dr. Jana Žel, National Institute of Biology, Ljubljana, Slovenia and Jožef Stefan International Postgraduate School, Ljubljana, Slovenia
Co-Supervisor: Dr. David Dobnik, National Institute of Biology, Ljubljana, Slovenia

Evaluation Board:
Prof. Dr. Kristina Gruden, Chair, National Institute of Biology, Ljubljana, Slovenia and Jožef Stefan International Postgraduate School, Ljubljana, Slovenia
Prof. Dr. Jelka Šuštar Vozlič, Member, Agricultural Institute, Ljubljana, Slovenia
Dr. Esther J. Kok, Member, RIKILT, Wageningen University and Research, Wageningen, the Netherlands

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL



Alexandra Bogožalec Košir

MULTI-TARGET APPROACHES FOR DETECTION AND QUANTIFICATION OF GENETICALLY MODIFIED ORGANISMS

Doctoral Dissertation

VEČTARČNI PRISTOPI ZA DOLOČANJE IN KVANTIFIKACIJO GENSKO SPREMENJENIH ORGANIZMOV

Doktorska disertacija

Supervisor: Prof. Dr. Jana Žel

Co-Supervisor: Dr. David Dobnik

Ljubljana, Slovenia, January 2018

Acknowledgments

This thesis could not exist without the cooperation of many colleagues from different European institutions, the support of my friends and family, and foremost without the support of funding bodies who funded the research. The research presented in this thesis was supported by the funding from the European Union under Grant Agreement No. 613908, (Project DECATHLON) and Slovenian Research Agency (ARRS) through the PhD grant P4-0165 and project grant 1000-15-0105. My visit to RIKILT UR in Wageningen was supported by the Slovene Human Resources Development and Scholarship Fund (Contract Number 11012-51/2015). Complementary financial support was given by the Norwegian Research Council and the Dutch Ministry of Economic Affairs. The droplet digital PCR system (BioRad's QX100) used in the research was funded by the Metrology Institute of the Republic of Slovenia, with financial support from the European Regional Development Fund.

I would like to thank my supervisor, Jana Žel, and co-supervisor David Dobnik who supported and guided me throughout the past three and a half years of my PhD studies. A big thank you also goes to the rest of the GMO group, Mojca Milavec, Dejan Štebih, Tina Demšar and Jernej Pavšič for constructive debate and help with understanding the mosaic which is GMO testing, including the legislation. Thanks to Esther Kok from RIKILT UR for hosting me in her lab and giving me the opportunity to learn from her and her team, especially Jeroen van Dijk, Alfred Arulandhu, Marleen Voorhuijzen, Martijn Staats and Rico Hagelaar. Thank you! I would also like to thank Bjørn Spilsberg and Arne Holst-Jensen from the Norwegian Veterinary Institute and Carla Divieto from the Italian National Metrology Institute for a fruitful cooperation. Thanks also to the members of the dissertation committee, Kristina Gruden, Jelka Šuštar Vozlič, and Esther Kok.

I would also like to thank all the members of the FITO department. You are the ones creating a friendly and stimulating environment with all the good energy and encouragement, which makes great work possible. Thank you all!

A very special thanks goes to my family and friends for all the encouragement and support, especially, my husband. Thank you Uroš for all the patience and support. And last but not least I would also like to thank the extended (ex)DŠIWSO, Tjaša, Špela, Živa, Dejan and Andrej for every after hour coffee and game night laugh.

Abstract

Many countries regulate the presence of genetically modified organisms (GMOs) in food and feed products. A labelling threshold of 0.9% per ingredient has been implemented in the European Union (EU) for authorised GMOs and a zero tolerance is in use for unauthorised GMOs (UGMOs). Standard curve based quantitative real-time polymerase chain reaction (qPCR) is used as a golden standard for GMO detection and quantification. However, the rise in number of authorised GMOs renders qPCR laborious and expensive. Additionally, with the increased use of authorised GMOs in the food and feed industries the incidence of UGMOs is also rising. Within this doctoral thesis we aim to develop and characterise multi-target approaches for quantification of authorised GMOs and identification of UGMOs.

The most abundant GMO ingredients in terms of authorised GM lines in the EU are maize and soybean. Thus, we designed multi-target droplet digital PCR (ddPCR) assays for quantification of 12 EU-authorised maize and 11 EU-authorised soybean lines. Two assays were designed for each ingredient; a 4- and 10-plex for maize and 6-and 7-plex for soybean. In all cases the transgenes were labelled with one fluorescent reporter and the endogene with another, thus enabling the quantification per ingredient. There are four additional GM soybean lines in the process of authorisation, therefore we have extended the 7-plex assay into an 11-plex, hence eliminating the need for further method characterisation upon an authorisation of one or more of these four GM lines. All multi-target ddPCR assays produce specific results, and perform within the international recommendations set for GMO testing. Performance of soybean 6- and 7-plex assays was tested in an inter-laboratory assessment.

As accurate and precise nucleic acid quantification is important in GMO testing, an assessment of the critical factors for accurate copy number determination is needed. In ddPCR this factor is droplet volume variability. The results of an inter-laboratory comparison of different droplet generation platforms and different mastermixes demonstrated a clear discrepancy between the measured droplet volume and that assigned by the manufacturer, showing an underestimation of copy number when relying on calculations made by the corresponding software.

An adapted genome walking approach called amplification of linearly-enriched fragments (ALF) was developed for identification of UGMOs. By coupling ALF to next-generation sequencing (NGS) and designing a (semi)automated data analysis pipeline we aimed at simultaneous identification of all GMOs, including UGMOs. A DNA mixture of four reference materials in an uneven distribution was prepared, with one GM line for which only element order is known, serving as a model for a UGMO. By submitting the sequences gained after NGS to the (semi)automated data analysis pipeline, we successfully identified sequences matching element order of this UGMO serving as proof of principle for ALF as new UGMO detection strategy.

The presented methods are not relevant only for GMO community, they may also be introduced in fields such as medicine, where several targets can be quantified as a group or where only partial DNA sequence is known. Additionally, determination of factors influencing ddPCR droplet volume allows for a more accurate copy number determination.

Povzetek

V večini držav je prisotnost gensko spremenjenih organizmov (GSO) v hrani in krmi strogo regulirana. V Evropski uniji (EU) je prag za označevanje izdelkov, ki vsebujejo odobrene GSO, postavljen pri 0,9 % na sestavino. Hrana in krma, ki vsebuje višji odstotek odobrenih GSO, mora biti primerno označena, medtem ko je prisotnost neodobrenih GSO prepovedana. Verižna reakcija s polimerazo v realnem času (qPCR) je zlati standard za določanje in kvantifikacijo GSO, vendar z naraščajočim številom GSO linij postaja časovno in finančno prezahtevna. S povečanim številom odobrenih GSO se povečuje tudi prisotnost neodobrenih GSO. Cilj pričujoče doktorske disertacije je razvoj in karakterizacija metod za kvantifikacijo odobrenih in identifikacijo neodobrenih GSO.

Sestavini, ki jima pripada največ odobrenih GSO linij, sta v EU koruza in soja. Za kvantifikacijo 12 linij GSO koruze in 11 linij GSO soje smo razvili večtarčne teste kapljične digitalne verižne reakcije s polimerazo (ddPCR). Za koruzo tako imenovani 4- in 10-plex ter za sojo 6- in 7-plex. V vseh primerih so bili transgeni označeni z enim fluorescenčnim barvilom, medtem ko je bil endogen označen z drugim barvilom. Na ta način je omogočena kvantifikacija vseh GSO linij znotraj ene sestavine naenkrat. V primeru soje so v procesu avtorizacije še štiri GSO linije. Da bi se izognili potrebi po ponovni karakterizaciji večtarčnega testa vsakokrat, ko bi bila sprejeta še ena linija, smo 7-plex razširili v 11-plex in ga okarakterizirali. Vsi večtarčni testi ddPCR so specifični in v skladu z mednarodno sprejetimi priporočili za testiranje GSO. Dodatno sta bila 6- in 7-plex testirana v medlaboratorijskem testiranju.

Določitev kritičnih faktorjev, ki vplivajo na točnost in natančnost kvantifikacije nukleinskih kislin je izredno pomembna pri testiranju GSO. V primeru ddPCR je kritični faktor volumen kapljice oziroma variabilnost volumna. V medlaboratorijskem testiranju smo primerjali različne generatorje kapljic in različne osnovne zmesi (ang. *mastermix*). Primerjava je pokazala relevantno razliko med volumnom, ki je predpisan s strani proizvajalca, in izmerjenim volumnom, kar povzroči podcenjevanje števila kopij nukleinskih kislin v reakciji, kadar se zanesemo na izračun, ki nam ga poda programska oprema.

Za namen identifikacije neodobrenih GSO smo razvili prilagojen pristop hoje po genomu, imenovan pomnoževanje linearno obogatenih fragmentov ali ALF (ang. *amplification of linearly-enriched fragments*). Naš cilj je bil pripraviti delotok, ki omogoča (semi)avtomatizirano analizo podatkov, pridobljenih po sklopitvi ALF, z naslednjimi generacijami sekvenciranja, kar bi omogočalo hkratno identifikacijo tako odobrenih kot neodobrenih GSO. Pristop smo testirali na vzorcu, ki je bil pripravljen kar se da podobno vzorcem, ki jih srečamo v praksi. Vseboval je štiri GSO linije v različnih količinah. Za eno linijo popolno DNK zaporedje ni bilo poznano, ta linija je služila kot neodobren GSO v tej analizi. S pomočjo delotoka smo identificirali zaporedja, ki se ujemajo z zaporedjem neodobrenega GSO,in s tem potrdili princip delovanja tako za ALF kot za delotok.

Razvite metode pa niso relevantne le na področju GSO. Področja, kjer je možno zaznavati več tarč naenkrat kot skupino ali kjer je znan le delček tarčnega DNK zaporedja, na primer v medicini, lahko s prilagoditvami uvedejo novo razvite metode. Poleg tega smo z določitvijo faktorjev, ki odločilno vplivajo na velikost kapljic pri ddPCR, ključno pripomogli k bolj natančnemu določanju števila kopij.

Contents

Abbreviations **xiii**

1 Introduction **1**
1.1 Genetically Modified Organisms 1
1.2 GMO Detection and Quantification – Approaches and Issues 2
1.3 Novel Molecular Approaches in DNA-Based Detection and Quantification 3
1.3.1 Digital Polymerase Chain Reaction 4
1.3.2 Next-Generation Sequencing 4
1.4 Aims of the Research 5
1.4.1 Multi-target Digital Polymerase Chain Reaction 5
1.4.2 Study of Critical Factors for Accuracy of Absolute Quantification in Digital Polymerase Chain Reaction 5
1.4.3 Detection of Unknown/Unauthorised Genetically Modified Organism 5
1.5 Research Hypotheses 6
1.6 Publications Included and Candidate's Contributions 6

2 Scientific Publications **9**
2.1 Multiplex Quantification of Twelve European Union Authorized GM Maize Lines with Droplet Digital PCR 9
2.2 Development and Inter-Laboratory Assessment of Droplet Digital PCR Assays for Multiplex Quantification of 15 Genetically Modified Soybean Lines 21
2.3 Droplet Volume Variability as a Critical Factor for Accuracy of Absolute Quantification Using Droplet Digital PCR 33
2.4 ALF: A Strategy for Identification of Unauthorized GMO in Complex Mixtures by a GW-NGS Method and Dedicated Bioinformatics Analysis 45

3 Discussion **59**
3.1 ddPCR-Based Multi-Target Approach for GMO Quantification 59
3.2 Determination of Droplet Volume in ddPCR 60
3.3 ALF as a Strategy for UGMO Identification 61

4 Conclusions **65**

Appendix A **67**
A.1 Supplementary Material for Publication 2.1 68
A.2 Supplementary Material for Publication 2.2 82
A.3 Supplementary Material for Publication 2.3 105
A.4 Supplementary Material for Publication 2.4 132

Appendix B **141**
B.1 Permission for Reproduction of Publication 2.1 141
B.2 Permission for Reproduction of Publication 2.2 144

B.3 Permission for Reproduction of Publication 2.3 149
B.4 Permission for Reproduction of Publication 2.4 150

References **155**

Bibliography **163**

Biography **165**

Abbreviations

ALF ... Amplification of linearly enriched fractions
CCS ... Circular consensus sequence
CRM ... Certified reference material
ddPCR ... Droplet digital polymerase chain reaction
DNA ... Deoxyribonucleic acid
dPCR ... Digital polymerase chain reaction
EFSA ... European food safety authority
EU ... European Union
FDA ... United States Food and Drug Administration
GM ... Genetically modified
GMO ... Genetically modified organism
GW ... Gene walking
LAMP ... Loop-mediated isothermal amplification
MPP ... Minimal performance parameter
NGS ... Next-generation sequencing
NIB ... National Institute of Biology, Slovenia
PCR ... Polymerase chain reaction
qPCR ... Quantitative real-time polymerase chain reaction
rDNA ... Recombinant DNA
SMRT ... Single molecule real-time
UGMO ... Unauthorised genetically modified organism
WGS ... Whole genome sequencing

Chapter 1

Introduction

1.1 Genetically Modified Organisms

Genetically modified organisms (GMOs) are by a definition organisms, with the exception of humans, in which the genetic material has been altered in a way that does not occur naturally by mating and/or natural recombination [1]. The first commercial genetically modified organism, the FlavrSavr^{TM} tomato, gained approval from the United States Food and Drug Administration (FDA) in 1994 [2], triggering more or less successful placements of a variety of crops, ornamental plants, trees, and even fish on the worlds market [3]–[6]. The most important traits present in 80% of all genetically modified plants are herbicide tolerance and insect resistance with 334 and 291 lines approved worldwide, respectively [3]. Less abundant are modified production quality, pollinator control system, disease resistance, abiotic stress tolerance, and altered growth or yield [3].

The increased use of GMOs also influenced the legislation and today the majority of countries regulate the cultivation and trade of GM organisms [7], [8]. An authorisation system and mandatory labelling is most commonly set for food and feed that include GMOs above a certain threshold [8]–[10]. In the European Union (EU), the labelling threshold for food and feed products that contain, consist of, or are produced from authorised GMOs is set at 0.9% per ingredient by Regulation (EC) 1829/2003 [11]. A zero tolerance level is set for unauthorised GMOs (UGMOs). According to the so-called Low Level Presence (LLP) regulation, Commission Regulation EC 619/2011 [12], an exception to the zero tolerance are feed products where GMOs pending authorisation are allowed up to a level of 0.1%. However, strict requirements still need to be met in order for a certain GMO to be given such a status; an authorisation needs to be given in another country, a reference material and a method for identification of this GMO need to be available, and no adverse effects in health or the environment are identified by the European Food Safety Authority (EFSA) [12]. Cultivation or release of GMO is set in the Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EC [1]. Food and feed products, as well as seeds and whole organisms, therefore have to be rigorously tested to ensure that labelling complies with the legislation. Thus, to control the presence and quantity of GMOs in the food and feed supply chains, sensitive and specific detection and quantification approaches are needed. In order to sustain the increasing amount of samples tested, these methods also need to be time- and cost-effective.

With the increased number and use of authorised GMOs, we can also observe a gradual increase of UGMO incidents [13]. There are two categories of UGMOs, known and unknown UGMOs. Known UGMOs have been granted approval in countries other than the EU, and therefore the structure of the GM insert is known. Alternatively, unknown UGMOs have

not undergone any authorisation process and the insertion sequence and the point of insertion are unknown. Unknown UGMOs therefore cannot be identified using traditional approaches. According to the RASFF Portal [14], there were nine incidents of UGMOs reported in 2015, 15 in 2016 and 15 in the first 10 months of 2017. However, the majority of cases were related to known UGMOs.

So far no GM crops have been identified to have adverse effects on humans, animals or the environment [15], [16]. Nevertheless, with the rapid growth of GMO diversity in experimental settings, it is imperative to have the methodologies in place to detect and identify both known and unknown UGMOs. Thus, apart from the sensitive and specific detection and quantification of authorised GMOs, detection and identification of UGMOs and a distinction between authorised GMOs and UGMOs is required.

1.2 GMO Detection and Quantification – Approaches and Issues

GMOs express at least one additional trait encoded by one or more genes, introduced into the organism by means of recombinant DNA technology [10]. This trait differentiates them from their wild-type counterparts. GMO testing is based either on detection of recombinant DNA (rDNA) or on detection of the recombinant protein encoded by the inserted gene. Both approaches have their pros and cons. Protein-based methods are generally quicker and easier to execute, especially when immunochromatographic (lateral flow) strip tests (LFT) are used, they are also more cost-effective when looking at costs per analysis. However, detection relies on the expression of the protein, which can vary greatly, as it may happen that the protein is not even produced due to gene silencing mechanisms. Additionally, the rDNA containing the gene of interest can be inserted in different plant species or in different parts of the genome, which results in the presence of the same protein in different GMOs. DNA-based methods on the other hand are, compared to immunological methods, more time-consuming and less cost-effective, mostly due to the costs of the equipment, but provide us with much more information and are thus the most widely used methods.

The majority of DNA-based methods employ polymerase chain reaction (PCR) technology [10], [17]. PCR approaches can be categorised according to the information that we gain; screening, identification, and quantification. The rDNA is comprised of different GM elements; e.g. promoter, regulator, gene of interest, terminator. Screening approach is based on the detection of a pattern given by the presence or absence of a specific GM element. This pattern can be compared to a reference table of the corresponding elements in known GMOs and a spectrum of possible GMOs can be identified [18], [19]. Identification of a single GMO is, on the other hand, based on the determination of the point of insertion of the rDNA (GM-event) into the plant genome, and quantification is based on determination of the ratio between a species-specific reference gene (endogene) and the GM-event [10].

For more than a decade GM-event–specific real-time quantitative polymerase chain reaction (qPCR) has been the gold standard for GMO analysis. In recent years, more and more GMOs have been authorised and event–specific qPCR has become less and less time- and cost-effective. Therefore, an alternative approach was needed; and screening approaches became used among many laboratories [20]. However, with the increasing number of GMOs, the number of GM elements they consist of is also increasing. To further optimise the screening approach, qualitative multi-target or multiplex PCR [21]–[24] and

qPCR [25]–[29] screening assays [30] have been developed. As multi-target approaches enable the characterisation of more than one target in a reaction, a further time and cost reduction was obtained. Although, multiplexing has seemed to overcome several issues in GMO analysis, it has been, nonetheless, unable to solve the problems in GMO quantification. As the EU legislation dictates labelling of products based on the quantity of a certain GM ingredient, not only detection, but also quantification of authorised GMOs is critical. Despite the efforts towards development of quantitative multi-target qPCR approaches, success has been limited.

PCR-based methods are only suitable if the GMO sequence or at least a part of it is known. For UGMOs such is seldom the case. Whereas at least a part of the DNA sequence or the GM element order is available for known UGMOs, we have no sequence information for unknown UGMOs. For now, we rely on the fact that GM elements that comprise known GMOs are also present in unknown UGMOs and thus the presence and identification of a UGMO can be detected in the screening process, when the elements screened for do not correspond to any of the authorised GMOs. Many efforts are therefore directed into the development of screening approaches that could aid in the detection of UGMOs [31]–[35]. However, screening alone does not confirm the presence of UGMOs in a sample, to confirm the suspicion, further analyses need to be performed. Limited sequence information and absence of reference materials renders specific detection approaches for UGMOs practically impossible. Thus, in the last decades several genome walking (GW) approaches, aiming at identification of the GM insertion site or the construct sequence, have been developed and modified for application in the GMO field [36]–[46]. However, none of these GW approaches, with one exception [40], have been tested on a complex sample representing the real-life situation.

1.3 Novel Molecular Approaches in DNA-Based Detection and Quantification

Apart from qPCR screening and multiplexing, several novel approaches have been developed that enable DNA-based detection of GMOs. The most notable are digital PCR (dPCR) [47]–[52], loop-mediated isothermal amplification (LAMP) [53]–[61], PCR capillary gel electrophoresis (PGE) [62]–[65], and microarray [66]–[70] and Luminex [71]–[74] technologies. Although, PGE, microarrays and Luminex technology enable high levels of multiplexing, only dPCR enables quantification. Moreover, in many cases dPCR enables more accurate quantification than qPCR, as quantification does not rely on a calibration curve. Therefore, dPCR is the most promising of the novel approaches for detection and quantification of GMOs.

None of these approaches have the ability to identify UGMOs directly. An indication of a presence of a UGMO could be given, when GM elements are detected using these methods; however as for qPCR screening, the suspicion needs to be confirmed with sequencing. GW approaches can be used for confirmation of a UGMO in a sample. With the advancement of technology, and consistent drop of prices, next-generation sequencing is a good candidate for UGMO identification.

1.3.1 Digital Polymerase Chain Reaction

The concept of digital PCR was first introduced in 1992 [75]. Sykes et al. used a limiting dilution principle to quantify a rearranged immunoglobulin heavy chain from a leukemic clone. A reaction mixture is divided into many individual reactions called partitions. Each partition can contain none, one, or more target copies. At the end-point of PCR these partitions are either scored as negative or positive, and the input DNA concentration can thus be calculated using Poisson distribution. This principle was later named digital PCR by Vogelstein and Kinzler [76].

As dPCR is an end-point all-or-nothing detection system, it is less vulnerable to partial inhibition as qPCR. This principle also enables absolute quantification of nucleic acids without the need of a standard curve. Today two dPCR technologies are available; microfluidic/chip-based dPCR (cdPCR) and emulsion/droplet-based dPCR (ddPCR). In cdPCR the sample is distributed in wells on a chip, while in ddPCR the reaction mixture is partitioned in droplets. Both Taq-Man and intercalating dye chemistries can be applied in dPCR. Partitioning the reaction mixture in several thousand or in some cases even millions of partitions also increases the multiplex potential of dPCR. This technique thus represents a good candidate for reference methods for nucleic-acid copy-number determination not only in the GMO field, but also in human diagnostics and research [77], [78], provided, of course, that a thorough characterisation in terms of specificity, accuracy/trueness and source of uncertainty is performed.

1.3.2 Next-Generation Sequencing

With the advance of Next-Generation Sequencing (NGS) and lately third-generation sequencing technology, such as PacBio and Oxford Nanopore, sequencing of complex GMO samples has become feasible. As a method (e.g. qPCR assay) for quantification needs to be submitted by the applicant for authorised GMOs, NGS is predominantly used for identification of UGMOs. Here two approaches can be taken, depending on the kind of a UGMO.

At least a part of a sequence or an element order is known, for the majority of known UGMOs, thus targeted sequencing can be used to identify the UGMOs/GMOs in a sample. For simple samples, GW coupled to Sanger sequencing [36], [39], [41]–[45] yields good enough results. However, in complex samples, a common element can have different adjacent sequences from different materials present. Thus, GW coupled to NGS is a perfect approach for sequencing enriched complex samples. NGS in general outperforms Sanger sequencing, in terms of speed, throughput, and potential for multiplexing. For the purpose of UGMO detection, two GW approaches [37], [38], [40] have been coupled to NGS so far. The drawbacks of these two approaches are a potential sensitivity to contamination and a lack of automated or at least (semi)automated data analysis.

For unknown UGMOs whole genome sequencing (WGS) is needed. In WGS prior knowledge is not needed, as the entire DNA library is sequenced and reads assembled. Assembled reads can either be aligned to the reference genome or a de novo assembly can be performed when a reference genome is not available [38], [79]–[84].

As the majority of the laboratories use at least one form of screening, the GW-NGS approaches are better suited for routine use, as well as more cost-effective, compared to WGS. However, NGS provides us with an enormous amount of sequence information. In order to identify a UGMO and differentiate it from authorised GMOs, a bioinformatics

data analysis pipeline needs to be developed in general, and especially if such an analysis is to be used on a daily bases.

1.4 Aims of the Research

1.4.1 Multi-target Digital Polymerase Chain Reaction

In any analytical service, not just GMO testing, one must consider not only precision but also cost-effectiveness. New approaches therefore need to be at least as precise as previously used ones and, if possible, more cost-effective. A fitness for purpose study, practicability and applicability therefore need to be assessed. One of the most promising approaches for achieving higher precision and cost-effectiveness in GMO testing is multi-target dPCR.

Our first aim was to develop multi-target dPCR assays in order to achieve more precise time- and cost-effective quantification of the most abundant GM species, maize and soybean, authorised in the EU.

Our second aim was to characterise the novel multi-target assays in terms of compliance with minimum performance parameters in use in GMO testing, by testing accuracy, repeatability, reproducibility, robustness and trueness of this new approach.

1.4.2 Study of Critical Factors for Accuracy of Absolute Quantification in Digital Polymerase Chain Reaction

In dPCR there are two critical factors of absolute quantification accuracy, the partition volume and the number of partitions. In ddPCR the number of partitions varies and is impossible to predetermine, while the volume of droplets can be measured and used in the calculations, to accurately determine the copy number. The opposite is true for the second dPCR approach, chamber/microfluidics-based dPCR, where the number of partitions, i.e. wells in a chip is known but the volume of each well is nearly impossible to determine [85]. Thus to determine the accuracy of the ddPCR assays, the assessment of droplet volume is critical.

Our third aim was to measure and assess the partition volume of two BioRad droplet digital PCR droplet generator platforms, manual and automated, and different reaction conditions. And moreover, to use these data to assess the impact of droplet volume on the accuracy of copy number determination.

1.4.3 Detection of Unknown/Unauthorised Genetically Modified Organism

The only strategy of UGMO identification today is sequencing; either WGS or targeted sequencing (e.g. GW). Targeted sequencing approaches focus more and more on NGS, since NGS enables multiplexing and with that identification of UGMOs in complex samples. However, NGS creates a large amount of data, which makes identification of UGMOs,

without at least some automation of the data analysis, a laborious and lengthy process. Thus, in order to successfully introduce a sequencing approach into routine analysis, a (semi)automated data analysis pipeline is as important as the sequencing approach.

Our fourth aim was to design a (semi)automated bioinformatics data analysis pipeline for sequence analysis gained after a newly designed GW-NGS approach, with the aim of UGMO identification.

1.5 Research Hypotheses

1. dPCR is suitable for the development of a multi-target quantification assay that is more precise, time- and cost-effective compared to standard qPCR approaches used for the quantification of GMOs authorised in the EU.

2. UGMOs can be detected in a sample mimicking a real-life situation when applying a specific post GW-NGS (semi)automated bioinformatics data analysis pipeline.

1.6 Publications Included and Candidate's Contributions

In the first paper (*Multiplex quantification of twelve European Union authorised GM maize lines with droplet digital PCR*) a new strategy for multi-target quantification of GMOs has been shown for the first time. Two multiplex ddPCR assays have been designed and characterised, enabling quantification of 12 EU-authorised maize lines in just two reactions. The characterisation of the assays proved that this strategy is sensitive and accurate, and at least as time- and cost-effective as the most widely used qPCR-based methods. The PhD candidate is the co-author of this publication. She has performed ddPCR, analysed the data and characterised the assays. She contributed to the preparation of the manuscript by preparing the experimental section and drafting corresponding tables and figures. She also contributed to the drafting of the supplementary information.

Where the first paper dealt only with GMOs authorised in the EU, the second paper (*Development and inter-laboratory assessment of droplet digital PCR assays for multiplex quantification of 15 genetically modified soybean lines*) additionally tackles the issue of GMOs that are in the pipeline for authorisation. Here two multiplex assays have also been developed to enable quantification of 11 EU-authorised soybean lines, however one of the assays has been extended to incorporate additional four lines that are in the authorisation pipeline. With a comparison of the extended assay to the unextended one, we have proven that subtracting a target from the multiplex does not change its accuracy and precision for the remaining targets. Additionally, a successful inter-laboratory assessment has been performed. The PhD candidate drafted the experimental design, performed the ddPCR experiments, and data analysis at the National Institute of Biology (NIB). She also analysed the data for the inter-laboratory assessment and drafted the manuscript.

One of the critical factors for absolute quantification of ddPCR, the partition volume, is described in the third paper (*Droplet volume variability as a critical factor for accuracy of absolute quantification using droplet digital PCR*). Here we describe and compare two droplet measurement protocols. The most efficient protocol was used for the measurement of droplets generated by two types of BioRad droplet generators, manual and automated. Two manual and one automated droplet generators and two different mastermixes,

representing two different chemistries were tested. The droplet volumes for all different scenarios were determined and compared. The newly measured droplet volumes were used in a set of ddPCR experiments, where results were compared to those gained by using the assigned volume. The candidate drafted the experimental design, conducted the measurements at NIB and performed the ddPCR including the data analysis. She contributed to the statistical analysis and drafted the manuscript including the supplement.

In our fourth paper (*ALF: a strategy for identification of unauthorized GMOs in complex mixtures by a GW-NGS method and dedicated bioinformatics analysis*) we address the issue of UGMOs and their identification. Since GMOs, authorised, unauthorised and presumably unknown, contain just a handful of GM elements, a GW approach coupled to NGS has been developed to amplify and sequence GMOs including UGMOs that contain a known GM element, in complex samples. The bioinformatics analysis was conducted within the Galaxy platform, where a Galaxy workflow, a (semi)automated data analysis pipeline, was created. The PhD candidate designed the (semi)automated data analysis pipeline and drafted the bioinformatics part of the manuscript and the supplement.

Chapter 2

Scientific Publications

2.1 Multiplex Quantification of Twelve European Union Authorized GM Maize Lines with Droplet Digital PCR

David Dobnik, Bjørn Spilsberg, Alexandra Bogožalec Košir, Arne Holst-Jensen, Jana Žel

Analytical Chemistry, 2015, 87: 8218-8226. DOI: 10.1021/acs.analchem.5b01208

In this publication, we describe the design and characterisation of two multiplex ddPCR assays for quantification of EU-authorized GM maize lines. We have shown for the first time that assays with such a high level of multiplexing can be used for quantification of complex samples. Assessment of the multiplex assays showed compliance with the minimal performance parameters (MPPs) placed for detection and quantification of GMOs, and that these assays can compete with the most widely used qPCR methods. Moreover, in case of positive samples the new multiplex methods were more cost-effective. Additionally, a proof of principle on extension of one of the multiplex assays was shown, proving that in case of new authorization, an extension of the present multiplex system is possible.

The PhD candidate performed the comparison of simplex and multiplex assays and characterisation of both multiplex assays. She performed the ddPCR, analysed the data and assessed the assays in terms of compliance with MPPs. She also helped with the preparation of the paper by writing the experimental section and contributed to the preparation of tables and figures, as well as the supplementary information.

This is an open access article published under an ACS AuthorChoice License, which permits copying and redistribution of the article or any adaptations for non-commercial purposes.

analytical chemistry

Article

pubs.acs.org/ac

Multiplex Quantification of 12 European Union Authorized Genetically Modified Maize Lines with Droplet Digital Polymerase Chain Reaction

David Dobnik,*[,†,∥] Bjørn Spilsberg,[‡,∥] Alexandra Bogožalec Košir,[†,§] Arne Holst-Jensen,[‡] and Jana Žel[†]

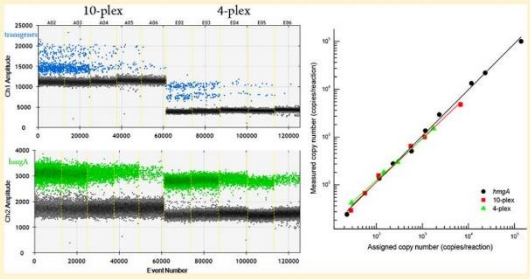
[†]Department of Biotechnology and Systems Biology, National Institute of Biology, Večna pot 111, 1000 Ljubljana, Slovenia

[‡]Section of Virology, Norwegian Veterinary Institute, P.O. Box 750, Sentrum, 0106 Oslo, Norway

[§]Josef Stefan International Postgraduate School, Jamova 39, 1000 Ljubljana, Slovenia

Supporting Information

ABSTRACT: Presence of genetically modified organisms (GMO) in food and feed products is regulated in many countries. The European Union (EU) has implemented a threshold for labeling of products containing more than 0.9% of authorized GMOs per ingredient. As the number of GMOs has increased over time, standard-curve based simplex quantitative polymerase chain reaction (qPCR) analyses are no longer sufficiently cost-effective, despite widespread use of initial PCR based screenings. Newly developed GMO detection methods, also multiplex methods, are mostly focused on screening and detection but not quantification. On the basis of droplet digital PCR (ddPCR) technology, multiplex assays for quantification of all 12 EU authorized GM maize lines (per April first 2015) were developed. Because of high sequence similarity of some of the 12 GM targets, two separate multiplex assays were needed. In both assays (4-plex and 10-plex), the transgenes were labeled with one fluorescence reporter and the endogene with another (GMO concentration = transgene/endogene ratio). It was shown that both multiplex assays produce specific results and that performance parameters such as limit of quantification, repeatability, and trueness comply with international recommendations for GMO quantification methods. Moreover, for samples containing GMOs, the throughput and cost-effectiveness is significantly improved compared to qPCR. Thus, it was concluded that the multiplex ddPCR assays could be applied for routine quantification of 12 EU authorized GM maize lines. In case of new authorizations, the events can easily be added to the existing multiplex assays. The presented principle of quantitative multiplexing can be applied to any other domain.



Many countries in the world regulate the cultivation and trade with genetically modified organisms (GMO).[1] Most commonly there is an approval system and mandatory labeling above a certain threshold. The enforcement of such regulations depends on the ability to detect and quantify the presence of GMOs in food, feed, and seed products. In the European Union (EU) this is enforced for all authorized GMOs through Regulation (EC) 1829/2003.[2] This regulation specifies that the developer of a GMO has to provide a method for quantifying the specific GMO event and funding for a collaborative trial to validate the method. Quantification of the GMO is done relative to a species specific reference gene. Thus, there exists a set of validated event specific GMO detection methods (available at http://gmo-crl.jrc.ec.europa.eu/gmomethods/). These are all standard curve based real-time PCR (qPCR) methods. In Europe, the labeling threshold for authorized GMO events is specified in Regulation 1829/2003.[2] This regulation states that labeling: *"shall not apply to foods containing material which contains, consists of or is produced from GMOs in a proportion no higher than 0.9% of the food ingredients considered individually or food consisting of a single ingredient..."* Thus, the labeling threshold is on a *per ingredient* basis, not on a *per GMO* basis. In analytical terms, ingredient in this context is interpreted as species.[3] So, it is the cumulative concentration, e.g., of all authorized maize GMOs relative to the total quantity of maize that determines if the threshold is exceeded or not. The technical guidance for sampling and detection of GMOs[4] states that the GMO content should be expressed as ratio of GM part in relation to the taxon specific part, meaning that GMO content can either be measured as additive concentration of individual authorized events or it can be measured as the total concentration of this group directly. The latter appears as a more cost-efficient approach, but hitherto it has not been possible to implement due to lack of suitable analytical methods.

Received: March 17, 2015
Accepted: July 13, 2015
Published: July 13, 2015

 © 2015 American Chemical Society

DOI: 10.1021/acs.analchem.5b01208
Anal. Chem. 2015, 87, 8218−8226

Standard curve based qPCR is sensitive and robust and is regarded as the gold standard for GMO analysis. As the number of authorized GMOs has increased over time, the required number of qPCR analyses has increased correspondingly. An analysis regime based only on separate event specific qPCRs is no longer cost efficient. As a more cost efficient approach, many laboratories now employ one of several alternative screening approaches, testing for presence of genetic elements commonly found in GMOs.[5] The observed presence/absence pattern for each sample can be compared with a reference table of presence/absence of the corresponding elements in known GMOs, e.g., using the GMOseek software.[6,7] At and above the limit of detection (LOD), the presence of GMOs containing elements that are not observed in the sample can effectively be excluded. The list of remaining candidate GMOs after screening serves as a starting point for quantification and confirmation of events present in the sample. Recently a multiplex version of one of these screening approaches[8] was developed as a single qualitative pentaplex PCR.[9]

Multiplex PCR analysis obviously has the potential to improve the cost efficiency, but a multiplex PCR is a complex experimental system with potential interference between oligonucleotides and amplification products (target and non-target). The development of quantitative methods to a high level of multiplexing has proven difficult. Although, several multiplex qPCRs exist, only two (duplex) are interlaboratory validated: one in Europe[10] and one in Japan.[11]

Digital PCR (dPCR) as a concept was invented in 1992.[12] In dPCR the reaction mixture is distributed into a large number of partitions, where each partition contains zero, one, or more copies of the target nucleic acid. Positive and negative signal in partitions is scored at end point of the PCR and the initial DNA concentration can be calculated using the Poisson distribution.[13] Digital PCR does not depend on PCR efficiency for DNA quantification to the same degree as standard curve based qPCR. Partial inhibition of a PCR will not affect quantification with dPCR as only discrete data on signal and not amplitude of the signal will be used in the calculations. Digital PCR can be performed in normal 96-well PCR plates using fluorescence probes,[14] and different approaches have been explored for the use in GMO analysis.[15,16] Although one of these was formally collaborative trial validated as a simplex dPCR approach for quantitation of GTS 40-3-2 (RoundupReady) soya (Marco Mazzara et al., personal communication), this experimental setup did not improve cost efficiency significantly due to large chemical consumption and low statistical power.

The dPCR is measuring the absolute number of copies in the sample; therefore, the final result for GM content is given as copy number ratio of GM event and taxon specific gene. Most widely used reporting unit for qPCR is nevertheless the mass ratio, as the majority of reference materials are certified for the GM% in mass. Thus, although the PCR reaction is amplifying the actual DNA copies, the final result is converted in mass ratio based on the standard curve of reference material. This fact made the topic unclear already from the beginning. The official unit of measurement has however never been exactly defined for the whole GMO field but rather for individual scopes (prepackaged food,[17] EU unauthorized GMOs[18] and recently for reporting of proficiency test results [EURL-GMFF Invitation letter, 2015]). As the EU Commission is apparently headed toward accepting only mass fraction as the accepted unit of measurement, an ENGL working group on the Unit of measurement has been recently established, who will try to explain the conversion of copy number ratio to mass fraction. For the purpose of this manuscript, the conversion factor set in EURL Technical Guidance 619/2011[19] was used.

Presently, two designated dPCR technologies are available: microfluidic/chip based dPCR and emulsion (droplet) based dPCR.[20] Several studies have tested the applicability of dPCR for GMO detection.[21−25] The practicability and applicability of using the droplet dPCR (ddPCR) in routine GMO diagnostics was recently shown for duplex reactions.[24] In the present study it was explored whether ddPCR could be multiplexed to a high degree and utilized to quantify a group of GMOs for compliance with the EU labeling regulation.[2] The analytical procedure of GMO testing include sample preparation, DNA extraction and purification, (q)PCR amplification, and data evaluation. These steps are nowadays often referred to as modules.[5] There are several advantages of using the modular approach in GMO testing, especially it is more favorable when the methods are being validated.[26] This study was focused only on the module of PCR amplification and does not include the evaluation of other modules. To be able to analyze GMO content per ingredient in a cost efficient manner and in line with the EU regulatory requirements,[2] a set of validated qPCR modules was used and combined for multiplex ddPCR. In this manuscript, the data on the performance of the developed assays are presented and their practicability, applicability, and potential impact is discussed.

EXPERIMENTAL SECTION

Test Material. A list of all CRM test materials is provided in Table S-15 the Supporting Information. They all have certified mass/mass (m/m) GM maize/wild-type maize material ratios. Routine diagnostic samples containing maize events (seeds or flour) and samples from proficiency programs (USDA and EURL) were also used in this study (Table S-12 in the Supporting Information). A specificity study was conducted on samples containing either GM maize events not authorized in the EU, DNA from non-GM maize/sweet corn, and GM and non-GM soybean, rape seed, rice, and linseed samples without maize (Table 2).

DNA Extraction. DNA was extracted and purified from 200 mg of starting material for all samples using a cetyltrimethylammonium bromide (CTAB) protocol with RNase-A solution and proteinase-K solution for removal of RNA and proteins from the sample (as described in annex A.3 of ISO 21570:2005[27]), with small adaptations. Dilutions of the extracted stock DNA solutions were made in nuclease- and protease-free water (Sigma-Aldrich Chemie GmbH, Munich, Germany). All samples were stored at −20 °C.

Primers and Probes Mix Preparation. The *hmgA* gene was used as the endogenous reference gene for maize, to which relative quantity of the GM events was estimated. For quantification of GM maize lines, the construct (for Bt11) or event (all others) specific qPCR modules were used. Nucleotide sequences of probes and primers used are presented in Tables S-4 and S-5 in the Supporting Information. The GM content in each sample was determined by simplex ddPCR. A total of 13 primer and probe sets were prepared for simplex reactions targeting individual events, constructs or the endogene, as well as 12 sets for duplex reactions where primers and probes for endogene (*hmgA*) and individual events or construct were mixed. The preparation of 10-plex and 4-plex primer and probe (PPP) mixes was done as follows. For 10-plex, primers and probes for the endogene (*hmgA*) and nine

DOI: 10.1021/acs.analchem.5b01208
Anal. Chem. 2015, 87, 8218−8226

events, GA21, MON810, MON863, DAS1507, MIR604, MON88017, MON89034, MIR162, and T25, were mixed (for final concentrations see Table S-4 in the Supporting Information). For 4-plex, primers, and probes for the endogene (*hmgA*), two events, DAS59122 and NK603, and one construct, Bt11, were mixed (for final concentrations see Table S-5 in the Supporting Information). Primers and probes were purchased from Eurofins MWG Operon (Ebersberg, Germany) or from Integrated DNA Technologies (Leuven, Belgium). Primers and probes were shipped lyophilized and diluted in nuclease- and protease-free water (Sigma-Aldrich Chemie GmbH, Munich, Germany) upon reception.

Droplet Digital PCR Reactions and Data Analysis. The reaction mix for all reactions, simplex as well as multiplex, was composed of 10 μL of 2xddPCR Supermix for probe (No dUTP) (Bio-Rad, Pleasanton, CA), 6 μL of primer and probe mix, 4 μL of DNA template (or 4 μL of nuclease- and protease-free water for nontemplate control), mixed into a 20 μL reaction volume. Droplets were generated in 8-well cartridges, using the QX100 droplet generator (Bio-Rad, Pleasanton, CA). Water-in-oil emulsions were transferred to a 96-well plate and amplified in a T100 PCR cycler (Bio-Rad, Pleasanton, CA). Thermal cycling conditions were 2 min at 50 °C, 10 min at 95 °C followed by 40 cycles of a two-step thermal profile comprising of 15 s at 95 °C and 60 s at 60 °C at ramp rate 2.5 °C/s. After cycling, each sample was incubated at 98 °C for 10 min and then cooled to 4 °C. Plates were then transferred to the QX100 droplet reader (Bio-Rad, Pleasanton, CA). Data acquisition and analysis was performed using QuantaSoft (Bio-Rad, Pleasanton, CA) software. Positive droplets, containing amplification products, were discriminated from negative droplets without amplification products by applying a fluorescence amplitude threshold (Figure S-1 in the Supporting Information). The threshold was set manually, using both the fluorescence amplitude vs event number (1D amplitude) and the histogram of events vs amplitude data streams, on each of the FAM and VIC channels (2D amplitude). Data generated by the QX100 droplet reader were rejected from subsequent analysis if a clog was detected by the Quantasoft software or if a low number of droplets (<8000) was measured per 20 μL PCR. After being exported, the data were further analyzed in Microsoft Excel spreadsheets. The number of template copies per μL was calculated using the volume of 0.85 nL per droplet. A digital MIQE checklist is made available in Table S-16 in the Supporting Information.

Comparison of Simplex and Multiplex Reactions. The performance of ddPCR 4-plex and 10-plex assays was compared to simplex assays. Both *hmgA* simplex and event/construct simplex reactions were performed on individual DNA samples in two dilutions (containing from 3 000 to 40 000 target DNA copies per reaction) each in a duplicate. The 4- and 10-plex assays were also tested on individual DNAs but only in duplicate of one dilution.

Dynamic Range, Repeatability, Limits of Detection, and Quantification. A dilution series in terms of target copies was prepared with mixture of DNA from all 12 transgenic maize events extracted from the CRMs (each line was present approximately in equal number of target copies). The quantity of each GM maize event in the DNA mix solution was determined by duplex ddPCR (individual event and *hmgA* in the same reaction on 6× dilution). These values were used to calculate the assigned copy number of targets in the dilution series of DNA mix. The assigned values were from 6724 to 0.01 copies of transgenes covered by 10-plex and from 1734 to 0.002 copies of transgenes covered by 4-plex. The dilution series for both multiplexes contained from 139 302 to 0.2 copies of the endogene *hmgA*. Twelve replicates of the dilution series were measured by ddPCR (three separate runs in 3 days each containing four replicates). Another dilution series was prepared in terms of transgene content, where the dilutions of DNA mix 1 for 10-plex contained 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 0.9, and 2% of all transgenes found in 10-plex, and 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, and 0.9% of all transgenes found in 4-plex for DNA mix 2. Six replicates of these dilution series were measured by ddPCR. The absolute/relative (copy number/%) limit of quantification (a/rLOQ) and absolute/relative limit of detection (a/rLOD) for ddPCR were determined based on these experimental results. The LOQ was determined as the aLOQ or rLOQ in the sample, where the relative standard deviation (RSD) of all replicates was below 25%. The LOD was determined as the aLOD or rLOD in the sample, where all replicates still produced a positive signal.

Specificity. To test specificity the DNA extracts from 12 different samples (3 GM maize events not authorized in the EU, 2 GM soybean events, one GM rice event, one GM oilseed rape event and five different nontransgenic samples; Table 2) were tested with the 4- and 10-plex ddPCR assay. Reactions were performed in duplicates per sample per assay.

Fitness for Purpose. To assess whether the new assays are fit for purpose, 13 different samples from proficiency programs (USDA and EURL) and four routine diagnostic samples (Table 3) were tested with both multiplex assays. To obtain a robust mean value of GM content, the proficiency test reports were used to make an average of all reported values per event, then to calculate the sum of all present events, and finally to calculate standard deviation and *Z*-score values. For the results to be directly comparable, the conversion of ddPCR results (measured as GM% in copy number ratio) into the mass ratio (GM% in mass fraction) was done according to EURL Technical Guidance 619/2011.[19] For sample 189/09, containing GM maize in which the transgene originated from the male parent, another conversion using the factor 0.36 (based on information from Holst-Jensen et al.[3] and our previous ddPCR analyses of those GM lines [data not shown]) was also performed. For routine diagnostic samples, the additive concentration of GM maize events was calculated based on results of quantification with qPCR performed in the NIB laboratory for official control under ISO17025 accreditation.

Robustness. The robustness of the ddPCR multiplex assay was evaluated by making small modifications to the protocol. These included (1) use of a PCR cycler with slower ramp rate (GeneAmp 9700 PCR cycler [Applied BioSystems, Foster City, CA]), (2) different annealing temperatures (61 and 59 °C), (3) use of another Mastermix (2xddPCR Supermix for probe containing dUTP [Bio-Rad, Pleasanton, CA]), (4) pipetting all eight reaction mixes to the cartridge in one step with a multichannel pipet, and (5) transfer of generated droplets to PCR plate with an automatic electric multichannel pipette. Two samples from the repeatability experiments (DNA mix, 60× and 120× dilution) were used in the robustness experiments, and the results for normal conditions were taken from the repeatability experiments. The 60× sample was tested in duplicates with both multiplex assays, whereas only one reaction was tested for the 120× sample per multiplex assay.

DOI: 10.1021/acs.analchem.5b01208
Anal. Chem. 2015, 87, 8218−8226

RESULTS AND DISCUSSION

Selection of PCR Modules. In the EU there is no consistent official definition of an event (see discussion in ref 3). Consequently, single transformations, multiple transformations, and crosses of transformations appear as events in the official EU registry (http://ec.europa.eu/food/dyna/gm_register/index_en.cfm). There are currently (per first April 2015) 37 such events of maize authorized in the EU (Table S-1 in the Supporting Information). Twelve of the authorized events (1−12 in Table S-1 in the Supporting Information) are what can be denoted single events, while the remaining 25 (13−37 in Table S-1 in the Supporting Information) are stacked events (hybrid crosses between the 12 single events). Validated event specific qPCR modules are available for each of these 12 single events (Table S-2 in the Supporting Information and http://gmo-crl.jrc.ec.europa.eu/gmomethods/). As far as technically possible, the multiplexes were designed using methods validated by the European Network of GMO Laboratories (ENGL). For GA21 there are two alternative validated modules from different developers (QT-EVE-ZM-007, 112 base pair (bp) and QT-EVE-ZM-014, 101 bp). Both modules target the 5′ junction between the maize genome and the inset. The QT-EVE-ZM-007 fulfilled the validation criteria in a relative concentration range of 0.98−4.26% GMO while the QT-EVE-ZM-014 fulfilled the criteria in the concentration range 0.09−8.0% GMO. The latter module thus appeared to be more robust. In addition, there is an 8 base pair (bp) perfect match between the probe in the QT-EVE-ZM-007 module and the reverse primer of the ENGL validated NK603 specific module (QT-EVE-ZM-008) that can be avoided by choosing QT-EVE-ZM-014. In the ENGL validated Bt11 module (QT-EVE-ZM-006), the forward primer and probe targets the pUC19 derived sequence. Consequently, commonly found pUC19 contamination in molecular biological reagents could possibly interfere with primer and probe. On the basis of previous experience, the module works with some mastermixes but not with others. Therefore, it was decided to substitute it with the construct specific module designed and in-house validated by Brodman et al.[28] The latter module is in-house verified for use in routine analyses for detection and quantification of Bt11 in one of our laboratories (NIB) and has a record of successful use by at least one other laboratory in GeMMa proficiency testing schemes over nearly a decade (see reports from the proficiency testing rounds; http://fapas.com/proficiency-testing-schemes/gemma/). A qPCR targeting the maize *hmgA* reference gene was chosen as the taxon specific module (QT-TAX-ZM-002).[27] By focusing on PCR modules, different extraction procedures of different samples were not tested. On the basis of the experiences from qPCR, there might be some influences of extraction procedures of different complex or processed real-life samples on the performance of PCR. However, as dPCR is less sensitive to different inhibitors coming from such samples,[25,29] the effect should not be as pronounced as with qPCR.

Combining Modules into Multiplex Assays. As the selected qPCR modules were designed and validated for use in simplex qPCRs, it needed to be checked carefully that the modules can be multiplexed. As a first step, obvious interactions between primers and probes were analyzed *in silico* using the software Autodimer.[30] All primers and probes were aligned with a local alignment algorithm with a score of +1 for match and a penalty of −1 for mismatch. Three putative interactions with a score ≥7 were identified (Table S-3 in the Supporting Information). There was an 18 bp perfect match between the probes for DAS59122 and MON88017. This interaction will most likely not result in the generation of false positive signals, but it can reduce the effective free probe concentration in the PCR and potentially lead to generation of false negatives. A 9 bp match between the probe of MIR604 and the reverse primer of DAS59122 and an 8 bp match between the probes of MIR604 and MON88017 were carefully evaluated and it was concluded that the latter was unlikely to cause significant interference. Pooling of all the 12 event/construct-specific modules and the *hmgA* maize reference gene module into one multiplex was considered. However, on the basis of the putative interactions (Table S-3 in the Supporting Information) it was decided to divide the 12 event/construct specific PCRs onto two multiplex assays. Initially, multiplexing was tested under qPCR conditions with two hexaplex combinations; one with GA21, MON810, MON863, MON88017, MON89034, and NK603, and another hexaplex with Bt11 (QT-EVE-ZM-006), DAS1507, DAS59122, MIR162, MIR604, and T25. Both hexaplexes were found to be acceptable (data not shown). Then the hexaplexes were transferred to ddPCR and observed that Bt11 did not yield a signal, while the GMO concentration was underestimated for NK603 and DAS59122 (data not shown). The event specific Bt11 module was substituted with the module developed by Brodmann et al.[28] as discussed above and decided to combine the modules for GA21, MON88017, MON89034, MIR162, MIR604, T25, MON810, MON863, DAS1507, and *hmgA* in a decaplex (10-plex) and the modules for NK603, Bt11, DAS59122 and *hmgA* in a tetraplex (4-plex) assay, respectively. These combinations were successively adopted as the final assay designs and circumvented the putative interactions described above.

Characterization of Multiplex Assays on Individual GMOs. The ddPCR system used in the experiments enables the use of two different fluorophores (FAM and VIC/HEX channel, respectively). The idea was to label all probes for GMOs with FAM and the probe for the maize endogene with HEX (Tables S-4 and S-5 in the Supporting Information). Quantification of individual GMOs using the 4-plex and 10-plex assays was compared to quantification under simplex conditions. Both multiplex assays were comparable to the corresponding simplex conditions (Table 1), as the bias was consistently below 25%. Thus, it was concluded that both multiplex assays were suitable for use in subsequent performance assessment experiments.

Limit of Quantification and Dynamic Range. The absolute limit of quantification (aLOQ) is the lowest target copy number in a sample that can be reliably quantified with an acceptable level of precision and accuracy.[31] The aLOQ of each target group in the ddPCR assays was estimated as the lowest copy number within the dynamic range with a relative standard deviation (RSD) of the measured copy number ≤25%.[31] Four replicates of each reaction were performed on 3 separate days on serial dilutions of DNA mixture. The RSD of all 12 replicates was used to determine the aLOQ, which was 42 and 29 copies per reaction for transgenic lines in the 4-plex (Table S-6 in the Supporting Information) and the 10-plex (Table S-7 in the Supporting Information), respectively. For the endogene, *hmgA*, the aLOQ was 24 copies (Table S-8 in the Supporting Information). Acceptable values of aLOQ for qPCR, published[15,32] as well as implemented in our routine laboratories, range from 30 to 100 copies. Both developed

DOI: 10.1021/acs.analchem.5b01208
Anal. Chem. 2015, 87, 8218−8226

Table 1. Performance Comparison of Multiplex Assay against Simplex Conditions on Individual DNA Material of GM Lines

	average GM %		
GM event	simplex	multiplex	bias multiplex to simplex (%)
DAS1507	5.71	6.09	6.65
DAS59122	3.48	3.59	3.16
GA21	34.09	27.11	−20.47
MIR162	58.99	59.12	0.22
MIR604	38.47	38.21	−0.68
MON810	3.83	3.63	−5.22
MON863	6.07	6.16	1.48
MON89034	60.27	55.77	−7.47
NK603	2.48	2.02	−18.47
T25	96.44	87.33	−9.45
Bt11	1.90	2.06	8.42
MON88017	48.90	54.78	12.02

multiplex assays fit within this range. As both multiplex assays are intended for quantification of total amount of 12 EU authorized GM maize lines, the sum of all transgene copies against the average of *hmgA* copies from both assays was also calculated. With this approach, the estimated aLOQ is 35 copies of transgenes (Table S-9 in the Supporting Information). The data is summarized in Figure 1. For a GMO detection method to be fit for purpose it must also be able to quantify low amounts of transgenic material in a high amount of nontransgenic material (e.g., ≥0.1%). The observed relative LOQ (rLOQ) for the 4-plex and 10-plex were 0.068% and 0.058%, respectively (Table S-10 in the Supporting Information). Quantitative performance at this level is compliant with the EU legal requirements (Regulation EC 1829/2003[2]) and the performance requirements for GMO detection methods as specified by the European Network of GMO Laboratories.[31]

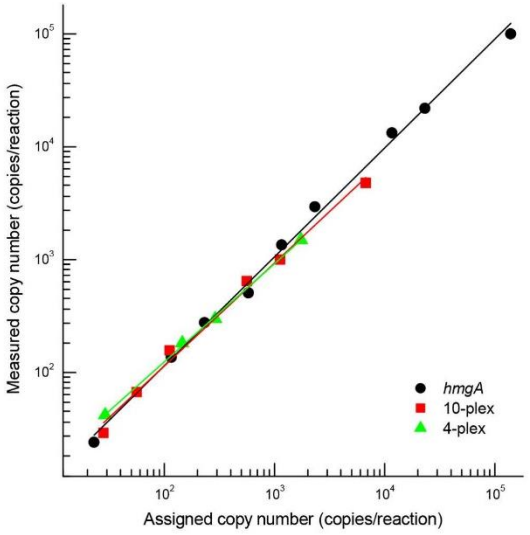


Figure 1. Dynamic range and correlation between measured copy number with the two multiplexes and assigned copy number on a mixture of equal amounts of each of the 12 maize events (as described in the Experimental Section). Each data point represents the average of three independent experiments with four replicates. Linear response for GM copy number mesured with the 4-plex (green triangles) is illustrated by the green line, linear response for GM copy number measured with the 10-plex (red squares) is illustrated by red line and linear response for maize reference gene copy numbers measured with *hmgA* in the 4-plex and in the 10-plex (black circles) is illustrated by black line. R^2 values originate from linear regression over the linear values and were 0.9898, 0.9943 for transgenes in 4-plex and 10-plex, respectively, and 0.9965 for *hmgA*.

Repeatability. To test whether both multiplex assays produce repeatable results, the performance was assessed on the level of target copies per reaction. Data from the experiments performed for aLOQ and dynamic range determination were used. All along the quantitative dynamic range, the RSD of the determined transgene copies in both multiplex sets remained below the threshold for acceptance for quantitative methods (RSD < 25%[31]) (Tables S-6−S-8 and S-10 in the Supporting Information).

Sensitivity. To assess the sensitivity, the absolute limit of detection (aLOD) was determined. The aLOD is the lowest target copy number in a sample that can be reliably detected but not necessarily quantified.[31] The aLOD was determined as the lowest copy number value for which all replicates were scored as positive. There does not yet exist a consensus on the number of positive droplets that are required to score a ddPCR replicate as positive, but 2, 3, and 5 droplets are commonly used thresholds. In this study, a threshold of 3 positive droplets was applied for scoring a replicate as positive. The aLOD was estimated to be 10 and 24 copies for the *hmgA* in the 4-plex and 10-plex, respectively, and 7 and 17 copies for the transgenes in the 4-plex and 10-plex, respectively. These results are compliant with the performance requirements for GMO testing methods in the EU[31] (i.e., aLOD < 25 copies).

Specificity. The specificity of the multiplex ddPCR assays was verified on the samples of unauthorized maize lines, nontransgenic maize, and other plant species (transgenic and nontransgenic) listed in Table 2. The false positive rate in dPCR could be presented in two ways: (a) as the whole

Table 2. False-Positive Rates Observed with ddPCR

sample	% of positive droplets for transgenes in 4-plex[a]	% of positive droplets for transgenes in 10-plex[a]
Bt176 nonauthorized GM maize	0	0.0038
DP98140 nonauthorized GM maize	0	0
MON87460 nonauthorized GM maize	0.0042	0
nontransgenic sweet corn	0	0.0034
nontransgenic maize seeds	0.0048	0
MON89788 soybean	0.0079	0.0042
MON40-3-2 soybean	0	0
GT73 rapeseed	0	0
nontransgenic rape seeds	0.0076	0
LL62 rice	0	0
nontransgenic rice grains	0	0
nontransgenic linseed	0	0
water	0.002	0

[a]Droplets from two replicates were considered for calculation.

DOI: 10.1021/acs.analchem.5b01208
Anal. Chem. 2015, 87, 8218−8226

replicate or (b) as per partition (droplet). Acceptance values of false negative rates ≤5% are widely adopted[33] and a false positive rate below 5% of the replicates (wells) is a translation of this threshold. These are also the thresholds proposed in a recently published guidance document.[34] In present experiments, a false-positive rate of 0% was observed over all replicates (Table S-11 in the Supporting Information), showing a good specificity of the assays. When the false positive rate was assessed over droplets in each replicate, the observed false positive rate was below 0.015% (data not shown). The latter satisfies the acceptance criterion for a false positive rate per partition set to <0.2% for dPCR.[34] The percentage of positive droplets for transgenes in both replicates of a sample is presented in Table 2. Detailed results of false-positive rates are presented in Table S-11 in the Supporting Information. In addition, for all experiments performed in this study, the no-template-controls (NTC) were negative.

Fitness for Purpose. Both proficiency test samples and certified reference materials were used to assess the closeness of agreement between the average value obtained with the multiplex ddPCR method and the accepted reference values of these materials.[31] Thirteen proficiency test samples containing a variety of currently EU authorized maize events were analyzed and for all of these ddPCR results deviated less that 25% from the robust mean value (Table S-12 in the Supporting Information). In proficiency tests, each result gets assigned a *Z*-score value, indicating the statistical distance from the mean value of all results. All values where $|Z| < 2$ pass the proficiency test. A theoretical *Z*-score were calculated to assess whether the results from multiplex ddPCR would have passed the proficiency test. In all 13 cases $|Z| < 2$, therefore the results would have passed the test. Four samples from our routine diagnostic activity were included to test different matrixes. These samples have a high concentration of GM maize. The measured values of ddPCR were compared to results obtained earlier with standard curve based qPCR. In all four cases were the ddPCR values within ±25% of the qPCR values. To compare measured ddPCR values to accepted reference value or certified values of CRMs, the mass/mass concentration were converted to copy number ratio. The conversion to GM% in mass fraction was done according to EURL Technical Guidance on implementation of Regulation 619/2011.[19] A conversion factor "GM % in DNA copy number ratio = 50 % [GM% in mass fraction]" is used for crops hemizygous for an event-specific insert (e.g., hemizygous GM maize), while the conversion factor "GM % in DNA copy number ratio = 100 % [GM % in mass fraction]" is used for crops homozygous for an event-specific insert (e.g., homozygous GM soybean). This factor is not a precise value applicable to each sample, but rather a pragmatic approach for harmonization among laboratories in the EU. Because of specific seed composition in maize, the actual conversion factors are higher or lower, based on the parental origin of GM plant and the tissue in question.[3] Nevertheless, over the entire dynamic range the values measured with ddPCR deviated <25% from the certified values (Table S-13 in the Supporting Information).

Robustness. Robustness is defined as a measure for the capacity of a method to remain unaffected by small, but deliberate deviations from the experimental conditions described in the procedure.[31] Within the EU, a GMO detection method is deemed robust, if despite varied parameters the determined analyte copy number does not change more than 30%.[31] Six different modifications to the original procedure were tested experimentally, including different annealing temperatures, different PCR cyclers, different Mastermix, and different pipetting techniques. When comparing the overall GM % determined by both multiplex assays together, none of the modifications caused a change in result >30% for both samples, indicating that the difference may be the result of increased variability between replicates. A significant effect of PCR cycler and multichannel pipetting was observed in only one of two samples (Table 3). A detailed look at the results, comparing the measured copy numbers for individual assays can confirm that the changes of the procedure affected the results randomly (Table S-14 in the Supporting Information). When the results of the experiments obtained with the original protocol were compared to results obtained with modified protocols as one group, the RSD of the determined copy number did not exceed the threshold of 25% at concentrations ≥ LOQ (data not shown). Thus, the overall robustness of both assays was shown to be acceptable.

Table 3. Robustness of the Multiplex Assays Tested by Introducing Small Changes to Original Protocol Presented As Bias % of Determined GM Content Compared to Original Protocol

	bias % of GM content		
changes introduced	DNA mix 60×	DNA mix 120×	average
annealing at 61 °C	−5.5	6.4	1.4
annealing at 59 °C	−20.0	−21.5	−19.9
different Mastermix	−8.7	22.4	7.8
different PCR cycler	38.6	−1.4	20
multichannel pipetting to cartridge	−5.1	39.5	18.2
automatic pipetting to plate	−9.6	10.9	1.6

Practicability, Applicability, and Impact. Prior to introduction of new technology in a laboratory, practicability for its daily use has to be verified.[35] To benchmark the multiplex ddPCR for quantification of 12 EU authorized GM maize lines against existing qPCR methods, calculations were made, using information on overall costs of performed analysis (personnel, consumables, indirect costs, etc.; Table 4). Routine sample analysis with qPCR consists of several different steps, depending on the customer's order. Usually the testing includes an initial screening analysis, identification of GM lines on the basis of positive screening results, and final quantification of specific GM lines identified in the sample. Genetically modified botanical impurities, such as presence of some GM soybean (the impurity) in a maize gluten (the ingredient) are commonly detected.[5] These will typically interfere with screening approaches and complicate the interpretation, e.g., giving the impression that the sample is likely to contain GM maize. Direct application of the presented multiplex ddPCR assays will emend the interpretation problem with respect to EU authorized GM maize events. However, screening could aid to uncover that a sample contains a botanical impurity or a GM maize event not authorized in the EU.[5] Thus, the choice to apply screening or not is also a matter of the scope of analysis. For the reasons given above, different combinations of methods were used in calculations for qPCR in combination with different numbers of samples analyzed simultaneously (1, 5 and 11). Multiplex ddPCR assays for quantification of 12 EU authorized GM maize lines were taken as a reference, to which

DOI: 10.1021/acs.analchem.5b01208
Anal. Chem. 2015, 87, 8218−8226

Table 4. Practicability and Cost-Effectiveness of the Developed Multiplex Assays in Comparison to Existing qPCR Approaches[a]

Testing pipeline	Steps prior quantification	Number of tested samples	Estimated hands-on time (hours)	Working days until final result (days)	Relative final price per sample (%)
Direct quantification of twelve approved GM maize lines with 4-plex and 10-plex multiplex assays with ddPCR	/	1	3	0.8	100
		5	4	0.9	100
		11	5	1.0	100
Direct quantification of twelve approved GM maize lines in simplex reactions with qPCR	/	1	8	2.0	272[b]
		5	16	4.3	292[b]
		11	25	7.5	300[b]
Initial screening with qPCR, all samples negative	5x Splex SE	1	2	0.5	83
		5	6	1.3	105[b]
		11	8	1.5	105[b]
Initial screening with qPCR, all samples negative	Pplex SE	1	1	0.4	31
		5	3	0.6	39
		11	5	1.0	39[b]
Initial screening, identification of six GM maize lines, quantification of one GM maize line with qPCR	Pplex SE, 6 specific lines	1	8	1.5	175[b]
		5	14	2.8	167[b]
		11	20	4.4	159[b]
Direct identification of twelve approved GM maize lines, quantification of two GM maize lines with qPCR	12 specific lines	1	10	1.8	226[b]
		5	14	3.1	195[b]
		11	18	4.8	185[b]
Initial screening with qPCR and quantification of twelve approved GM maize lines with 4-plex and 10-plex multiplex assays with ddPCR	Pplex SE	1	4	1.1	131[b]
		5	7	1.5	139[b]
		11	10	2.0	139[b]

[a]Testing pipeline, different scenarios of possible procedures for testing a diagnostic sample are taken into account; ddPCR, droplet digital PCR; qPCR, quantitative real-time PCR with standard curve; 5x Splex SE, screening phase using five simplex qPCRs to determine the presence of P-35S, T-nos, ctp2-cp4-epsps, bar, and pat elements; Pplex SE, screening phase using pentaplex assay to determine the presence of P-35S, T-nos, ctp2-cp4-epsps, bar, and pat elements.[9] Working days until final result, hands-on time and time needed for PCR cycling was considered. An 8-h working day was used for calculation. Relative final price per sample, ddPCR testing for 1, 5, or 11 samples with 4- and 10-plex assays was taken as a reference. For purpose of comparison the relative value of 100 was set to these setups. Values shaded with the same color should be compared, as they correspond to the equal number of tested samples. [b]More than one 96-well reaction plate is needed to perform all the reactions.

all other approaches were compared (Table 4). As a rule of thumb, the cost per sample is reduced by approximately 50% in the case of 5 samples and by approximately 60% in the case of 11 samples, independently of analysis strategy (data not shown). Only relative values are presented, as the costs for specific consumables, personnel, and other costs vary between laboratories and countries. Additionally, the working days until final result were calculated to visualize the influence of number of runs needed in some testing pipelines. The calculations demonstrate that qPCR testing is more cost efficient than multiplex ddPCR only when all samples are negative and analyzed with pentaplex qualitative qPCR screening[9] or when a single sample analyzed with simplex qualitative qPCR screening[8] yielded a negative result. Our experience is that GMOs are detected in approximately 50% of the analyzed food/feed samples, but the ratio of positive samples can of course vary from laboratory to laboratory. A common scenario is that an initial screening is followed by qualitative identification of 6 events and finally quantification of one of these. This scenario is 60−70% more costly than direct quantification with the multiplex ddPCR approach. As pentaplex screening of negative samples is highly cost efficient, an interesting approach could be to combine an initial pentaplex qPCR screening and use multiplex ddPCR to quantify the positive samples. As expected, the use of standard curve based singleplex qPCR directly is the most costly approach in the comparison. Unlike for qPCR, the cost of Mastermix is a large contributor to the cost of ddPCR, as can be read indirectly out of Table 4. Digital droplet PCR is a relatively new technology. The costs of Mastermix can be expected to decrease when the technology will gain volume and if new suppliers enter the market. This could further improve the cost efficiency of ddPCR. It has previously been demonstrated that ddPCR performs comparably to qPCR in duplex reactions for quantification of single events.[24] Although not reflected in the costs, the total analysis time is worth considering. The multiplex ddPCR analysis of as much as 11 samples (one full 96 well plate) can be finished in 1 day, whereas in cases other than pentaplex screening of negative samples, from 2 to almost 8 working days are needed (Table 4). The present study demonstrates that increasing the multiplexing using ddPCR is highly cost-effective. Therefore, it can be concluded that quantification of currently EU authorized GM maize lines in routine samples using multiplex ddPCR is

DOI: 10.1021/acs.analchem.5b01208
Anal. Chem. 2015, 87, 8218−8226

practical for GMO laboratories, with better throughput and cost-effectiveness than qPCR.

Extension of Multiplex Assays. There is a high possibility of authorization of new events in the future; therefore, it is important that it will be possible to add additional events to the multiplexes. The 10-plex system is already close to the limit, as the overall concentration of the probes in the reaction is much higher than recommended for the simplex reaction. This results in a high fluorescence value of negative droplets and makes the difference between negative and positive droplets smaller. With potential rain, the threshold setting could be rather challenging. Thus, the 10-plex could probably not be extended a lot, but the 4-plex assay still has some room left for extensions and is more appropriate for addition of future authorized events. In the process of manuscript revision, another GM maize event (MON87460) was authorized in the EU. Preliminary tests were performed, adding the event to the 4-plex resulting in a 5-plex reaction. All of the testing combinations were comparable to the equivalent assays (data not shown) and were within the limits of acceptance criteria.[31] Shortly, the 5-plex assay produced a comparable result on MON87460 DNA as simplex assay (bias of 7.8%). 5-plex and 4-plex assays on DNA which were used in the rLOQ 4-plex also produced comparable results (bias of 7.8%). The quantification of the DNA mixture of all four events targeted by the 5-plex assay produced a result comparable to the sum of quantifications of individual lines (bias of 6.1%). Although the full in-house validation was not performed with this 5-plex assay, it was shown that addition of new events to the multiplex system is possible in the case of new authorizations.

Summary. The purpose of this study was to take the ddPCR to a higher level by introducing a multiplex approach and evaluating its suitability and conformity to the strict minimum performance parameters in the field of GMO diagnostics. The applicability of ddPCR for GMO diagnostics has previously been investigated and demonstrated to be comparable to qPCR (reviewed in[36]). However, the multiplexing capabilities of ddPCR for GMO detection were previously only demonstrated as duplex amplification of transgene and endogene target.[24] The presently reported, newly developed multiplex assays, combining interlaboratory validated simplex qPCR assays, were implemented with only slight optimization of primer and probe concentrations. The multiplex assays achieved a wide dynamic range, with lower limit of quantification below 50 target copies per reaction and upper limit around 100 000 target copies. Both multiplex assays showed good sensitivity, suitable for GMO testing. All of the tested parameters comply with international recommendations[31,35] and are comparable to those reported for interlaboratory validated qPCR simplex assays. Quantitation using multiplex ddPCR appears to be highly robust. At least three factors contribute to this. First, each technical replicate consists of more than 10 000 individual PCRs. Second, the calculations are based on the distribution of discrete signals, which is a qualitative measure, and not signal intensity and Cq-values, which are quantitative measures. Third, the reference gene and the transgene target are analyzed in the same well and thus the pipetting errors associated with sample volumes will cancel in the calculation of GMO content. From the perspective of a diagnostic lab, one of the most important parameters of a method is its practicability. The characteristics of the ddPCR system enable quantification of GMO content in samples performed in much lower number of reactions, thus reducing the cost of analysis, and decreasing the uncertainty linked to dilution pipetting errors.[22] Comparison of cost-effectiveness in the case of routine GMO diagnostics of both the multiplex ddPCR assays and qPCR demonstrate an overall better cost efficiency of the first, whenever samples are positive for GM events and better cost efficiency of the latter, when the screening phase gives negative results. To be employed in routine testing, methods need to be properly validated through interlaboratory validations and verified during their introduction in laboratories to demonstrate their fitness for the purpose. The ddPCR performance demonstrated in the present study already verified the suitability of the method for testing routine samples. Collaborative trial validation, planned within the DECATHLON project (www.decathlon-project.eu), is expected to reinforce this conclusion. Nonetheless, it was not the purpose of this study to achieve better performance than qPCR, which is still a golden standard in GMO detection, but rather to show that new techniques based on PCR amplification are comparable in terms of performance parameters, but can outperform qPCR in terms of cost-effectiveness. The most cost efficient approach can be qPCR screening of all samples prior to multiplex ddPCR quantification of the screening positive samples or ddPCR alone of all samples. The best basis to decide is probably the expected GMO negative rate as experienced by the laboratory. To conclude, the presented ddPCR multiplex assays could enable more cost-effective GMO quantification, especially when the system is extended to include additional relevant plant species. Moreover, the principle of multiplexing presented here could be applied to any other domain where several targets can be quantified as a group.

■ ASSOCIATED CONTENT

Supporting Information

Additional information as noted in the text. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.5b01208.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: david.dobnik@nib.si.

Author Contributions

∥D.D. and B.S. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Carlos Goncalo das Neves for reviewing the manuscript. We would also like to thank Dejan Štebih for help with calculations of cost-effectiveness and for useful discussions and comments. Atiya R. Ali, Camilla Skjæret, and Coraline Basset are gratefully acknowledged for superb technical assistance. The research leading to these results has received funding from the European Union under Grant Agreement No. 613908, (Project DECATHLON). Complementary financial support was given by the Norwegian Research Council and the Slovenian Research Agency (Contract Numbers P4-0165 and 1000-15-0105).

■ REFERENCES

(1) Gruère, G. P.; Rao, S. R. *AgBioForum* **2007**, *10*, 51−64.
(2) European Commission. *Off. J. Eur. Union L* **2003**, *268*, 1−23.

DOI: 10.1021/acs.analchem.5b01208
Anal. Chem. 2015, 87, 8218−8226

(3) Holst-Jensen, A.; De Loose, M.; Van den Eede, G. *J. Agric. Food Chem.* **2006**, *54*, 2799−2809.
(4) European Commission. *Off. J. Eur. Union L* **2004**, *348*, 18−26.
(5) Holst-Jensen, A.; Bertheau, Y.; de Loose, M.; Grohmann, L.; Hamels, S.; Hougs, L.; Morisset, D.; Pecoraro, S.; Pla, M.; Van den Bulcke, M.; Wulff, D. *Biotechnol. Adv.* **2012**, *30*, 1318−1335.
(6) Morisset, D.; Novak, P. K.; Zupanič, D.; Gruden, K.; Lavrač, N.; Žel, J. *BMC Bioinf.* **2014**, *15*, 258.
(7) Block, A.; Debode, F.; Grohmann, L.; Hulin, J.; Taverniers, I.; Kluga, L.; Barbau-Piednoir, E.; Broeders, S.; Huber, I.; Van den Bulcke, M.; Heinze, P.; Berben, G.; Busch, U.; Roosens, N.; Janssen, E.; Žel, J.; Gruden, K.; Morisset, D. *BMC Bioinf.* **2013**, *14*, 256.
(8) Waiblinger, H.-U.; Grohmann, L.; Mankertz, J.; Engelbert, D.; Pietsch, K. *Anal. Bioanal. Chem.* **2010**, *396*, 2065−2072.
(9) Huber, I.; Block, A.; Sebah, D.; Debode, F.; Morisset, D.; Grohmann, L.; Berben, G.; Stebih, D.; Milavec, M.; Zel, J.; Busch, U. *J. Agric. Food Chem.* **2013**, *61*, 10293−10301.
(10) Waiblinger, H.-U.; Ernst, B.; Anderson, A.; Pietsch, K. *Eur. Food Res. Technol.* **2008**, *226*, 1221−1228.
(11) Takabatake, R.; Koiwa, T.; Kasahara, M.; Takashima, K.; Futo, S.; Minegishi, Y.; Akiyama, H.; Teshima, R.; Oguchi, T.; Mano, J.; Furui, S.; Kitta, K. *Shokuhin Eiseigaku Zasshi* **2011**, *52*, 265−269.
(12) Sykes, P. J.; Neoh, S. H.; Brisco, M. J.; Hughes, E.; Condon, J.; Morley, A. A. *Biotechniques* **1992**, *13*, 444−449.
(13) Pinheiro, L. B.; Coleman, V. A.; Hindson, C. M.; Herrmann, J.; Hindson, B. J.; Bhat, S.; Emslie, K. R. *Anal. Chem.* **2012**, *84*, 1003−1011.
(14) Vogelstein, B.; Kinzler, K. W. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9236−9241.
(15) Berdal, K. G.; Bøydler, C.; Tengs, T.; Holst-Jensen, A. *Eur. Food Res. Technol.* **2008**, *227*, 1149−1157.
(16) Lee, D.; Mura, M. La; Greenland, A.; Mackay, I. *Food Chem.* **2010**, *118*, 974−978.
(17) European Commission. *Off. J. Eur. Union* **2000**, *109*, 29−42.
(18) European Commission. *Off. J. Eur. Union* **2011**, *166*, 9−15.
(19) EURL-GMFF. Technical guidance document from the European Union Reference Laboratory for Genetically Modified Food and Feed on the implementation of Commission Regulation (EU) No. 619/2011, http://gmo-crl.jrc.ec.europa.eu/doc/Technical%20Guidance%20from%20EURL%20on%20LLP.pdf (accessed January 30, 2015).
(20) Baker, M. *Nat. Methods* **2012**, *9*, 541−544.
(21) Demeke, T.; Gräfenhan, T.; Holigroski, M.; Fernando, U.; Bamforth, J.; Lee, S.-J. *Food Control* **2014**, *46*, 470−474.
(22) Bhat, S.; Herrmann, J.; Armishaw, P.; Corbisier, P.; Emslie, K. R. *Anal. Bioanal. Chem.* **2009**, *394*, 457−467.
(23) Corbisier, P.; Bhat, S.; Partis, L.; Xie, V. R. D.; Emslie, K. R. *Anal. Bioanal. Chem.* **2010**, *396*, 2143−2150.
(24) Morisset, D.; Štebih, D.; Milavec, M.; Gruden, K.; Žel, J. *PLoS One* **2013**, *8*, e62583.
(25) Burns, M. J.; Burrell, a. M.; Foy, C. a. *Eur. Food Res. Technol.* **2010**, *231*, 353−362.
(26) Holst-Jensen, A.; Berdal, K. G. *J. AOAC Int.* **2004**, *87*, 927−936.
(27) International Organization for Standardization. *ISO 21570:2005 Foodstuffs − Methods of analysis for the detection of genetically modified organisms and derived products − Quantitative nucleic acid based methods*; ISO 21570; International Organization for Standardization: Geneva, Switzerland, 2005.
(28) Brodmann, P. D.; Ilg, E. C.; Berthoud, H.; Herrmann, A. *J. AOAC Int.* **2002**, *85*, 646−653.
(29) Rački, N.; Dreo, T.; Gutierrez-Aguirre, I.; Blejec, A.; Ravnikar, M. *Plant Methods* **2014**, *10*, 42.
(30) Vallone, P. M.; Butler, J. M. *BioTechniques* **2004**, *37*, 226−231.
(31) EURL-GMFF. Definition of Minimum Performance Requirements for Analytical Methods of GMO Testing European Network of GMO Laboratories (ENGL), http://gmo-crl.jrc.ec.europa.eu/doc/Min_Perf_Requirements_Analytical_methods.pdf (accessed January 30, 2015).
(32) Peccoud, J.; Jacob, C. *Biophys. J.* **1996**, *71*, 101−108.
(33) Standardization, I. O. for. *ISO 24276:2006 Foodstuffs — Methods of analysis for the detection of genetically modified organisms and derived products — General requirements and definitions*; ISO 24276; International Organization for Standardization: Geneva, Switzerland, 2006.
(34) Holst-Jensen, A.; Crespo, T.; Simplício, A.; Richl, P.; Welsche, M.; Dobnik, D.; Dreo, T. *Deliverable 6.1 MPP Decathlon*, http://www.decathlon-project.eu/sites/default/files/Deliverable6.1_MPP_Decathlon.pdf.
(35) Codex Committee on Methods of Analysis and Sampling. 2010, 1−22.
(36) Milavec, M.; Dobnik, D.; Yang, L.; Zhang, D.; Gruden, K.; Zel, J. *Anal. Bioanal. Chem.* **2014**, *406*, 6485−6497.

DOI: 10.1021/acs.analchem.5b01208
Anal. Chem. 2015, 87, 8218−8226

2.2 Development and Inter-Laboratory Assessment of Droplet Digital PCR Assays for Multiplex Quantification of 15 Genetically Modified Soybean Lines

Alexandra Bogožalec Košir, Bjørn Spilsberg, Arne Holst-Jensen, Jana Žel, David Dobnik

Scientific Reports, 2017, article number: 8601. DOI: 10.1038/s41598-017-09377-w

In this publication, we describe the design and inter-laboratory assessment of two multiplex ddPCR assays for quantification of EU-authorized GM soybean lines. We also describe and assess the expansion of one of the assays to include additional lines that are in the pipeline for authorisation, which contributes to the practicability of such assays for future use. An inter-laboratory assessment has been performed for the first time, on an assay for GMO quantification of such complexity.

The PhD candidate outlined the experiment design, and performed the ddPCR experiments, complete with data analysis at NIB. The candidate also performed the comparison of results gained between the two laboratories. Additionally she performed the practicability calculations and drafted the manuscript including the supplement.

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Development and inter-laboratory assessment of droplet digital PCR assays for multiplex quantification of 15 genetically modified soybean lines

Received: 22 November 2016
Accepted: 25 July 2017
Published online: 17 August 2017

Alexandra Bogožalec Košir[1,2], Bjørn Spilsberg[3], Arne Holst-Jensen[3], Jana Žel[1] & David Dobnik[1]

Quantification of genetically modified organisms (GMOs) in food and feed products is often required for their labelling or for tolerance thresholds. Standard-curve-based simplex quantitative polymerase chain reaction (qPCR) is the prevailing technology, which is often combined with screening analysis. With the rapidly growing number of GMOs on the world market, qPCR analysis becomes laborious and expensive. Innovative cost-effective approaches are therefore urgently needed. Here, we report the development and inter-laboratory assessment of multiplex assays to quantify GMO soybean using droplet digital PCR (ddPCR). The assays were developed to facilitate testing of foods and feed for compliance with current GMO regulations in the European Union (EU). Within the EU, the threshold for labelling is 0.9% for authorised GMOs per ingredient. Furthermore, the EU has set a technical zero tolerance limit of 0.1% for certain unauthorised GMOs. The novel multiplex ddPCR assays developed target 11 GMO soybean lines that are currently authorised, and four that are tolerated, pending authorisation in the EU. Potential significant improvements in cost efficiency are demonstrated. Performance was assessed for the critical parameters, including limits of detection and quantification, and trueness, repeatability, and robustness. Inter-laboratory performance was also determined on a number of proficiency programme and real-life samples.

Since the early 1990s when the first genetically modified (GM) food plant was introduced, the FlavrSavr tomato, many GM plants have been more or less successfully placed on the world market[1–3]. Most countries regulate the cultivation and trade of GM organisms (GMOs)[4,5], and many have an authorisation system and mandatory labelling of foods that include GMOs above a certain threshold[5,6]. In the European Union (EU), Regulation (EC) 1829/2003[7] sets the labelling threshold for food products that contain, consist of, or are produced from authorised GMOs, at 0.9%. A consequence of this (relatively low) threshold is that distinction between authorised and unauthorised GMOs is required, and thus sensitive and specific detection and quantification methods are needed.

Most accepted methods for the detection and quantification of GMOs are DNA-based. Quantification is based on determination of the ratio between a species-specific reference gene (endogene) and a specific sequence unique to the genetic modification (the GM-event)[8]. EU legislation requires that methods for detection and quantification of both the reference gene and the specific GM-event sequence are provided by the applicant for GMO authorisation[7]. This includes the funding for validation of the method by the EU Reference Laboratory for GM Food and Feed (EURL-GMFF), assisted by the European Network of GMO Laboratories (ENGL). The validated methods are published in a database of reference methods for GMO analysis (available at: http://gmo-crl.jrc.ec.europa.eu/gmomethods/). This database also includes other formally validated GMO detection methods. The

[1]Department of Biotechnology and Systems Biology, National Institute of Biology, Večna pot 111, SI-1000, Ljubljana, Slovenia. [2]Josef Stefan International Postgraduate School, Jamova 39, SI-1000, Ljubljana, Slovenia. [3]Norwegian Veterinary Institute, P.O. box 750 Sentrum, 0106, Oslo, Norway. Alexandra Bogožalec Košir and Bjørn Spilsberg contributed equally to this work. Correspondence and requests for materials should be addressed to A.B.K. (email: alexandra.bogozalec@nib.si)

Event/species[a]	Unique identifier[b]	Legal status	Detected by	EURL-GMFF method reference[c]
MON87701	MON-877Ø1-2	Authorised	6-plex	QT-EVE-GM-010
MON87708	MON-877Ø8-9	Authorised	6-plex	QT-EVE-GM-012
MON87769	MON-87769-7	Authorised	6-plex	QT-EVE-GM-002
305423	DP-3Ø5423-1	Authorised	6-plex	QT-EVE-GM-008
BPS-CV127-9	BPS-CV127-9	Authorised	6-plex	QT-EVE-GM-011
A2704-12	ACS-GMØØ5-3	Authorised	7-plex/11-plex	QT-EVE-GM-004
MON89788	MON-89788-1	Authorised	7-plex/11-plex	QT-EVE-GM-006
MON40-3-2	MON-Ø4Ø32-6	Authorised	7-plex/11-plex	QT-EVE-GM-005
356043	DP-356Ø43-5	Authorised	7-plex/11-plex	QT-EVE-GM-009
A5547-127	ACS-GMØØ6-4	Authorised	7-plex/11-plex	QT-EVE-GM-007
MON87705	MON-877Ø5-6	Authorised	7-plex/11-plex	QT-EVE-GM-003
DAS-68416-4	DAS-68416-4	Valid EFSA application	11-plex	QT-EVE-GM-013
FG72	MST-FGØ72-2	Valid EFSA application	11-plex	QT-EVE-GM-001
DAS-44406-6	DAS-444Ø6-6	Valid EFSA application	11-plex	QT-EVE-GM-015
DAS-81419-2	DAS-81419-2	Valid EFSA application	11-plex	QT-EVE-GM-014
Soy (*Le*1)	—	—	6-plex/7-plex/11-plex	QT-TAX-GM-002

Table 1. EU authorised and tolerated soybean GM-event, together with validated qPCR methods for their detection and quantification, and specified ddPCR multiplex. [a]Name of GMO used in the EU register of authorised GMOs[24], except soy (*Le*1). [b]Unique identifier according to OECD guidelines[34]. [c]http://gmo-crl.jrc.ec.europa.eu/gmomethods/.

majority of the methods included in the database at present are standard-curve-based real-time PCR (qPCR) methods.

For more than a decade, GM-event–specific qPCR has been the gold standard for GMO analysis. The number of authorised GMOs has grown rapidly during this time, and separate GM-event–specific detection methods are now only exceptionally cost effective. Development and application of multiplex and screening methods can lower the costs of GMO analysis considerably, although they can also face several limitations. Several PCR[9–12] and qPCR[13–17] multiplex methods have been developed. One of these, a qualitative screening qPCR multiplex (5-plex) approach that detects p35S, tNOS, ctp2-cp4-*epsps*, *bar* and *pat* screening elements has also been validated[13]. These multiplex methods are predominantly qualitative. It has proven difficult to develop quantitative qPCR methods with high levels of multiplexing. To the best of our knowledge, only two duplex methods have been inter-laboratory validated for quantitative analysis; one in Europe[15] and one in Japan[18]. The labelling threshold in the EU is not based on the individual concentration of GM-events. Instead, the labelling required is based on the total concentration of GM-events per ingredient, with an ingredient in this context interpreted as a species[8]. This opens the possibility to take a different approach, in which all GMOs from a species can be quantified simultaneously using one fluorophore[8].

Digital PCR (dPCR) as a technique was described by Vogelstein and Kindler for the first time in the early 1990s[19]. The concept of dPCR is relatively simple. The amplification of targets in a sample is performed in hundreds, thousands, or even millions of partitions. Each partition acts as a separate qualitative test with a digital (positive or negative) outcome (hence dPCR), regardless of the amplitude of the signal. By counting the negative and positive partitions, the number of DNA copies in a sample can be calculated using the Poisson distribution[20]. This gives dPCR an advantage over qPCR, as partial PCR inhibition does not affect the quantification. Today, two dPCR technologies are available: microfluidic or chip-based digital PCR (cdPCR), and emulsion-based or droplet digital PCR (ddPCR)[20]. The applicability of a ddPCR multiplex in GMO analysis has recently been shown for duplex[21–23], 4-plex, and 10-plex assays for GM maize[8].

At the time of writing, there were 12 soybean GM-events authorised in the EU under regulation (EC) 1829/2003[7, 24]: 11 single GM-events and one hybrid stack (cross of two authorised single GM-events: MON87701 × MON89788) (Supplementary Table S1). In the EU, a hybrid stack is legally treated as a new GM-event, and thus requires separate authorisation. These 12 GM-events can be detected with 11 GM-event–specific PCR assays (Table 1). By 22nd of July, 2016, valid European Food Safety Authority (EFSA) applications for authorisation have been submitted and are pending for an additional eight soybean GM-events. At present, these are regulated under the so-called Low Level Presence (LLP) Regulation EC 619/2011[25]. This means that the presence of each of these eight soybean GM-events is tolerated in feeds, but not in foods, at concentrations below 0.1%. Four of these eight GM-events are hybrid stacks, where the individual GM-events are authorised (305423 × 40-3-2; MON87769 × MON89788; MON87705 × MON89788; MON87708 × MON89788). These four stacks can be detected with five GM-event–specific assays that are already among the assays that detect EU-authorised single GM-events (Table 1). The remaining four GM-events that fall under LLP Regulation EC 619/2011[25] can be detected with four GM-event–specific assays with legal status as "valid EFSA applications" (Table 1).

The objective of the present study was to carry out an inter-laboratory comparative assessment to develop and test the performance of quantitative ddPCR multiplex assays that cover all of the present EU authorised soybean GM-events. Two quantitative ddPCR multiplex assays, a 6-plex and 7-plex, were designed, covering the 11 soybean GM-events authorised in the EU. With the addition of four tolerated GM soybean lines falling under LLP Regulation EC 619/2011[25], the 7-plex was extended to an 11-plex assay. This introduced sufficient flexibility in the assays developed to cover all pending authorisations.

The 6-plex, 7-plex and 11-plex assays quantify the species-specific gene *Le*1 and either five (MON88701, MON87708, MON87769, 305423, BPS-CV127-9), six (A2704-12, MON89788, MON40-3-2, 356043, A5547-127, MON87705) or ten (A2704-12, MON89788, MON40-3-2, 356043, A5547-127, MON87705, DAS-68416-4, FG72, DAS-44406-6, DAS-81419-2) GM-events, respectively (Table 1). Of note, the 6-plex assay targets all five soybean GM-events that do not contain any of the five most common screening elements (p35S, tNOS, ctp2-cp4-*epsps*, *bar* and *pat*)[15]. This assay is therefore suitable for application alongside qPCR screening with, e.g., the commonly used 5-plex screening assay published by Huber *et al.*[13].

The three multiplex ddPCR assays presented here (Table 1) are based on the combination of validated qPCRs and were designed to facilitate the testing for legal compliance in the EU. The performance with respect to sensitivity, repeatability, trueness and robustness of all three multiplexes was tested, including an inter-laboratory reproducibility study. The results and applications of the assays are discussed.

Results and Discussion

Selection of multiplexing approach. It is important to define the scope, performance criteria, and target values up-front for method development, validation, and application. This facilitates benchmarking and can lead to more efficient use of resources[26,27]. The multiplex ddPCR assays reported here were developed with a clearly defined scope: fitness for detection and reliable quantification of all EU-authorised and LLP-regulated (and tolerated) soybean GM-events, in line with current EU legal requirements for GMFF[7,25]. Quantification can be carried out per GM-event (line) or per species (ingredient). The EU labelling regulation[7] requires labelling on the basis of the GMO concentration per ingredient. This means that it is the sum of the GM-event concentrations per ingredient that must be determined, and thus the identification or quantification of each single GM-event is not necessary. On the other hand, the LLP regulation[25] applies only to feeds, not to foods, and LLP-regulated GMOs must be quantified individually in the EU[25]. However, these GMOs are rarely observed and their presence is tolerated up to a maximum of 0.1% (mass/mass) of an individual GMO per ingredient (species). If the presence of all EU-authorised or tolerated soybean GM-events does not exceed these thresholds in a feed product, then the product is legally compliant in the EU, except, if the product would contain other un-authorised GMO. Multiplex quantification of groups of soybean GM-events (authorised and tolerated) therefore appears to be an attractive, convenient, and potentially cost-efficient, solution.

The EURL for GMFF and the ENGL launched the first guidelines for performance assessment and acceptance of methods for legal compliance with EU GMO legislation in 2005. These guidelines have since been revised twice (2008, 2015). The latest version[27] contains specific performance criteria and target values, and these were adopted without modifications for the present study.

Specificity assessments. The specificity of the assays was determined *in silico*. The interactions between primers and probes were determined using the Autodimer[28] software and the multiplex specificity was predicted using the ePCR[29] software. No significant risk of dimers between pairs of oligonucleotides (primers/probes) was observed. In multiplex PCR, all of the possible combinations of primers can generate an unintended PCR product, if a matching template is present by chance. All of the possible primer combinations were analysed for each multiplex using 20 publically available GMO sequences (both EU authorised and unauthorised), as listed in Supplementary Table S2. In addition, the soybean and maize genomes were included. The ePCR[29] software was used as described (see Materials and Methods). For all three multiplexes, the expected products were found, except for A2701-12, where the sequence that covers the PCR is not publically available and thus not present in the template database. ePCR identified one hypothetical unintended 207-bp amplicon that was generated with the reverse primer for MON40-3-2 alone and the SYHT0H2 EU unauthorised GM soybean line as template (unique identifier, SYN-ØØØH2-5). SYHT02 contains a 254-bp inverted repeat spanning positions 3384 to 3637; in GeneBank accession JC050419. However, this hypothetical amplicon does not match any of the probes in the multiplex soybean assays. To test this, we analysed SYHT0H2 in duplex MON40-3-2 and *Le*1 ddPCR assay. SYHT0H2 alone yielded no signal, which demonstrated that this hypothetical PCR product does not yield false-positive results. In addition, we mixed MON40-3-2 and SYHT0H2 at concentrations of 7.3% and 1.1%, respectively. This mixture was quantified to 7.9%, which corresponded to a bias of 8.2%, which demonstrated that the quantification was not affected by the presence of SYHT0H2. Classical PCR was used to explore whether it was possible to visualise the 207-bp PCR product. First a gradient PCR with annealing temperature from 50 °C to 72 °C was performed. There were no bands of the expected size, although there was a shadow at approximately 200 bp at 50 °C. The PCR was repeated with the 50 °C annealing temperature and a fast cycling programme (and the normal programme), but the predicted PCR product could not be visualised. It is possible that the template can fold back on itself and outcompete primer binding. It was concluded that this hypothetical PCR product does not interfere with the ddPCR multiplexes.

Material characterisation and performance comparison. For pre-validation and to compare the performance of simplex, duplex (GM-event and reference gene) and multiplex (GM-events and reference gene) ddPCR assays, the absolute and relative GM-event concentrations were determined in reference materials (Table 2). The trueness of the methods were calculated as the bias compared to the certified values, and the associated method performance parameter was set to bias $\leq \pm 25\%$, as is commonly used for qPCR[26,27] minimum

GM soybean line	CRM (% m/m)	Quantitative performance (%)									
		Simplex		Duplex		6-plex		7-plex		11-plex	
		GM-event	Bias (%)	GM-event	Bias (%)	GM-event	Bias (%)	GM-event	Bias (%)	GM-event	Bias (%)
MON87701	99.94	104.1	4.2	97.6	−2.3	94.3	−5.6	—	—	—	—
MON87708	99.05	105.3	6.3	101.8	2.8	101.9	2.9	—	—	—	—
MON87769	99.94	101.6	1.7	95.8	−4.1	95.3	−4.6	—	—	—	—
305423	10	8.6	−14.0	8.7	−13.0	8.2	−18.0	—	—	—	—
BPS-CV127-9	96.32*	58.9	22.2	54.7	13.5	52.8	9.5	—	—	—	—
A2704-12	99.99	102.9	2.9	93.8	−6.2	—	—	96.7	−3.3	94.8	−5.2
MON89788	100	99.7	−0.3	95.4	−4.6	—	—	89.7	−10.3	90.4	−9.6
MON40-3-2	10	10.4	4.0	9.3	−7.0	—	—	10.7	7.0	9.2	−8.0
356043	10	10.1	1.0	11.0	10.0	—	—	10.9	9.0	9.9	−1.0
A5547-127	99.99	88.4	−11.6	94.4	−5.6	—	—	93.7	−6.3	93.0	−7.0
MON87705	99.4	95.3	−4.1	99.8	0.4	—	—	98.4	−1.0	98.0	−1.4
DAS-68416-4	10	12.2	22.0	12.0	20.2	—	—	—	—	9.8	−2.0
FG72	99.99	106.3	6.3	102.8	2.8	—	—	—	—	111.1	11.1
DAS-44406-6	98.60	90.9	−7.8	117.6	19.3	—	—	—	—	110.1	10.7
DAS-81419-2	98.60	97.7	−0.9	88.1	−10.6	—	—	—	—	85.1	−13.7

Table 2. Comparison of quantitative performance for simplex, duplex and multiplex ddPCR assays. CRM, Certified reference material. *CRM diluted by a factor of 2, and measured value thus compared to 48.2% (m/m), and not certified value of >96.3%.

performance. The bias was <25% in all three cases. Multiplex assays were thus deemed suitable for further performance assessment. Based on this dataset (Supplementary Table S3), GM-event and endogene concentrations (copy numbers) were assigned to each individual reference material. A DNA mix of the 15 GM-events was then prepared with the copy numbers as described in Supplementary Table S4.

Estimation of the limits of detection and quantification. The lower limit of detection (LOD), the lower limit of quantification (LOQ), and the dynamic range (quantification range) were estimated using a dilution series of the mixed DNA that targeted all of the GM-events in the three multiplex assays. The observed copy numbers for the *Le*1 reference gene in all three assays and at all dilutions were compared and used as an additional control for data comparability across assays.

A total of 10 test samples (DNA mixes 1–10) were prepared from the reference materials, with the purpose of determining the absolute LOD (LOD_{abs}) and LOQ (LOQ_{abs}). DNA mix 1 was a mix of all 15 reference materials, and DNA mixes 2–10 were prepared as serial dilutions of DNA mix 1 in sterile de-ionised water (see Methods). The assigned copy numbers for these DNA mixes were calculated from the material characterisation data (Supplementary Table S4). For DNA mix 1, the copy number was verified by duplex ddPCR assays for each of the 15 GM soybean lines (Supplementary Table S5), and the sum of the soybean GM-event copy numbers present in the individual multiplexes was calculated and compared to the assigned copy number (Supplementary Table S6). The bias between the assigned and experimentally determined copy numbers in DNA mix 1 was well below the 25% acceptance threshold for all three assays (Supplementary Tables S5 and S6). Therefore, the assigned values for further comparisons were 26282 copies for *Le*1, and 2557, 2612 and 4658 copies of GM-events covered by 6-plex, 7-plex and 11-plex, respectively (Supplementary Table S6).

Twelve replicates were tested for the 6-plex and 11-plex ddPCR assays, and 14 replicates were tested for the 7-plex ddPCR assay, on three separate days. The LOD_{abs} was estimated to 8 copies per reaction for 6-plex, 10 copies per reaction for 7-plex, 18 copies per reaction for 11-plex, and 17 copies per reaction for *Le*1. The LOQ_{abs} was estimated as 38 copies per reaction for 6-plex, 39 copies per reaction for 7-plex and 11-plex, and 36 copies per reaction for endogene *Le*1 (Table 3 and Supplementary Tables S7–S10). The observed repeatability standard deviation RSD_r between replicates was consistently <25% (acceptance limit[26,27]), with 39 or more target copies/reaction (Table 3). It is therefore reasonable to estimate that the LOQ_{abs} is 40 copies/reaction. As there were no negative replicates observed with 18 or more target copies/reaction (Table 3), it is therefore reasonable to estimate that the LOD_{abs} is 20 copies/reaction. Bias to assigned copy numbers was also determined for all three multiplexes and for *Le*1 (Supplementary Tables S7–S10). Bias was below 25% for all of the assays and DNA mixes, with the exception of the 11-plex assay for DNA mix 2, where the bias was slightly over 25% (26.11%).

To estimate the relative LOD (LOD_{rel}) and relative LOQ (LOQ_{rel}), a second set of DNA samples (DNA mix I–VII) was prepared with an assigned constant *Le*1 copy number per reaction, as 50000, and decreasing GM-event copy numbers (Supplementary Table S11). The observed bias in the calculated GM-event percentages (GM-event copy number relative to *Le*1 copy number) was ≤25% for all of the tests that contained 100 or more GM-event copies (assigned) per reaction, except for DNA mix III with the 11-plex ddPCR assay, where the bias was 31.64% (Supplementary Table S12). The LOQ_{rel} for the 6-plex, 7-plex and 11-plex assays were 0.20%, 0.21% and 0.44%, respectively, whereas the LOD_{rel} was 0.015%, 0.017%, and 0.028%, respectively (Supplementary Table S12). Comparison of the copy numbers in LOQ_{rel} to LOQ_{abs} (39 copies) suggested that the LOQ_{rel} might be lower than

DNA mix	6-plex		7-plex		11-plex		*Le*1	
	Mean copy number	RSD%	Mean copy number	RSD%	Mean copy number	RSD%	Mean copy number	RSD%
1	2229	5.4	2429	7.5	3633	7.3	22903	5.9
2	331	10.7	391	10.1	574	7.1	3382	6.0
3	189	6.8	186	9.1	316	10.1	1831	6.1
4	38	12.6	39	15.0	68	7.1	378	7.6
5	18	31.6	22	25.9	39	11.7	194	11.0
6	**8**	**26.6**	**10**	**34.1**	**18**	**36.7**	87	13.1
7	5[a]	ND[b]	4[a]	ND[b]	12	34.2	36	17.6
8	Neg.	Neg.	Neg.	Neg.	7[a]	ND	**17**	**28.5**
9	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	5[a]	ND
10	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.

Table 3. LOD_{abs} (bold) and LOQ_{abs} (underlined) according to total copy numbers per reaction. [a]At least one replicate was negative. ND, not determined due to negative replicate(s). Neg., all replicates were negative.

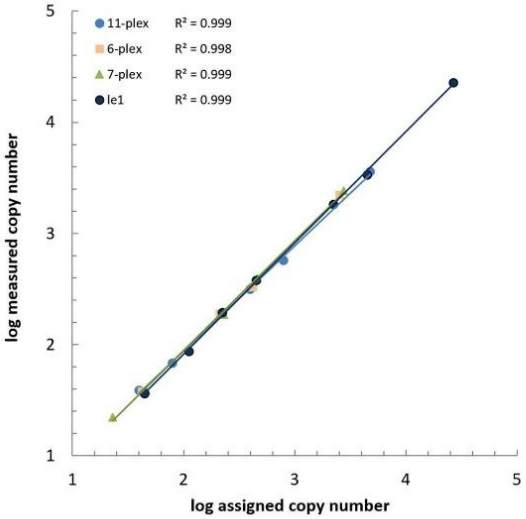


Figure 1. Dynamic range of the multiplex ddPCR assays. All of the assays demonstrate a very high degree of linearity ($R^2 > 0.998$) for all concentrations above the estimated limit of quantification (R^2, 0.9985 for GM-event targets in the 6-plex, 0.9995 for GM-event targets in the 7-plex, 0.9987 for GM-event targets in the 11-plex, and 0.9989 for the *Le*1 reference gene). Each data point represents the mean of three independent experiments with four replicates, except for the 7-plex, where two additional replicates were included for each data point in the third experiment.

that seen in Supplementary Table S12, as closer to 0.08% for all three multiplex assays. Also the LOD_{abs} determined on DNA mix VII (12 copies) suggested that the overall LOD_{abs} might be six copies lower than estimated (18 copies) for the 11-plex assay (Supplementary Table S12 and Table 3).

Dynamic range. The dynamic range is the range of concentrations that is compliant with quantifiability and includes the LOQ. The quality criteria and acceptance values for the dynamic range include linearity ($R^2 \geq 0.98$), precision ($RSD_r \leq 25\%$) and trueness (bias $< 25\%$)[26,27]. Furthermore, the current EU guidelines for acceptance of qPCR methods for legal compliance for GMO testing propose that the upper end of the dynamic range includes at least 2520 GM-event target copies/reaction and 56000 reference gene copies/reaction. The dynamic range should also include the relative concentration range from 0.09% to 4.5%[27].

The experimentally assigned numbers of copies of DNA mix 1 and its dilutions were plotted against the measured numbers of copies (Fig. 1) for the *Le*1 reference gene and the total GM-event content for each DNA mix. For all of the concentrations that exceeded the estimated LOQ_{abs}, the coefficient of variation, R^2, was >0.998, which indicated that the dynamic range included at least 40 to 2200 copies of the GM-event and 40 to 22000 copies of the endogene for all of the multiplex ddPCR assays. The mean observed GM-event copy numbers for the 6-plex

Sample	Assigned GM-event (%)[a]	Sum GM-event (6-plex + 7-plex) (%)		Bias (%)	
		NIB	NVI	NIB	NVI
G023/11	0.91	1.03	1.06	13.2	16.5
G071/15	94.20	91.39	91.92	−3.0	−2.4
226/14	2.70	2.42	2.47	−10.4	−8.5
272/14[b]	1.29	1.40	1.39	8.5	7.8
274/14[b]	2.26	2.34	2.63	3.5	16.4
1/15	0.94	0.95	1.12	1.1	19.1
75/15	91.38	90.59	96.29	−0.9	5.4
116/15	1.68	1.73	1.91	3.0	13.7

Table 4. Inter-laboratory comparison of results for 6-plex and 7-plex ddPCR assays for soybean GM-event quantification. [a]Assigned GM-event, value determined by NIB using GM-event–specific qPCR methods under ISO 17025 accreditation. [b]USDA proficiency test sample.

and 7-plex assays for the least diluted samples were slightly lower than the 2520 required for qPCR, whereas the GM-event copy number for the 11-plex was well above this (Table 3). The mean observed *Le*1 reference gene copy numbers for all three assays for the least diluted sample were approximately 40% of the 56000 required for qPCR (Table 3). With further experiments where the constant *Le*1 copy number was 50000, the RSD was <25% (Supplementary Table S13), and the dynamic range was thus extended, as much closer to 56000 copies. The relative concentration range of 0.09% to 4.5% was partially satisfied by all of the assays. The lower level (LOQ_{rel}) was 0.20%, 0.21% and 0.30% for 6-plex, 7-plex and 11-plex assays, respectively (Supplementary Table S12), and the upper level was well above 4.5%, at 9.61%, 10.50% and 18.07%, respectively (calculated from the data in Supplementary Tables S7–S10). Given the jump in RSD (Supplementary Table S12) additional dilutions between mix III and IV, could additionally lower the lower level of LOQ_{rel}, and shift it more toward the 0.09% limit. The observed linearity (Fig. 1) and accuracy (Table 3) were acceptable for all of the tested concentrations ≥LOQ.

Interlaboratory performance assessment. To assess the fitness for purpose and to test the inter-laboratory performance of the multiplex ddPCR assays, six real-life samples from routine testing for the presence of GMOs, and two proficiency-test samples from the US Department of Agriculture (USDA)/Grain Inspection, Packers and Stockyards Administration (GIPSA) proficiency-testing programme were selected (Supplementary Table S14). The robust means of the proficiency tests and qPCR results obtained by the National Institute of Biology (NIB, Slovenia) using GM-event–specific methods under ISO 17025 accreditation were used as accepted reference values. The results of the inter-laboratory comparison are presented in Table 4 and Supplementary Tables S15–S17.

The final protocols for the 6-plex and 7-plex assays were tested independently at the NIB and the Norwegian Veterinary Institute (NVI) to determine the method performance on real-life samples. This provided information on the accuracy (i.e., trueness [bias] and precision [repeatability *RSD*$_r$ and reproducibility *RSD*$_R$ relative standard deviation]). The overall performances were completely comparable between these two laboratories, with bias <21% (Table 4 and Supplementary Table S15). When compared in detail, all of the results fulfilled the criteria, with the exception of the repeatability between days for sample 274/14 tested at NVI (Supplementary Table S17), where RSD_r was slightly higher than acceptable (41.78% > 35%) for the 7-plex assay (mean GM-event target copies/reaction, 135).

The acceptance criteria were very strict and were originally developed for single GM-event–specific qPCR methods. At very low GM-event concentrations (<100 copies/reaction, or <0.09%) the EURL-GMFF/ENGL requirements are less strict (the EU guidelines for acceptance of qPCR methods accepts $RSD_r \leq 50\%$)[27]. The observed exceptions from compliance were almost consistently associated with concentrations near or below 100 copies/reaction. For sample 274/14, the RSD_r observed for 6-plex (46.66%) at the NVI was therefore acceptable (Supplementary Table S16). At the NIB, the repeatability was just above the acceptable value (52.48%) for 6-plex for sample 116/15; on the other hand, the RSD_r for the GM-event percentage was well below 25% (15.32%), and the result was considered acceptable (Supplementary Table S16). The observed reproducibility standard deviation (RSD_R) and bias were consistently acceptable for all samples (<25%; Table 4). It is therefore reasonable to conclude that the ddPCR assays are reliable and fit for purpose.

Accuracy (trueness and precision). The accuracy of the 11-plex assay was independently tested at NIB, and the results of these experiments were combined with the results for the 6-plex assay from the inter-laboratory trial (Table 5 and Supplementary Table S18), using the same routine and proficiency-test samples (Supplementary Table S13). Two additional samples (49/15, G158/15) were also included (Supplementary Table S13). The bias of the measured GM-event concentration was <25% for all of the samples, except for sample 274/14, for which the observed bias was 27.9%, which was just above the threshold (Table 5). However, the RSD_r was acceptable for both the 6-plex and 11-plex assays (<25%) (Supplementary Table S17), although the combined data led to a bias that was slightly higher than acceptable. The observed RSD_r for sample 49/15, which contained only targets for the 11-plex, was just above an acceptable level (35.31% > 35%), but the bias to the assigned GM-event was <25% (Supplementary Table S18), and with LOQ_{rel} as 0.44% for the 11-plex (Supplementary Table S12), these results were acceptable.

Sample	Assigned GM-event (%)[a]	Sum GM-event (6-plex + 11-plex) (%)	Bias (%)
G023/11	0.91	0.97	6.6
G071/15	94.20	90.23	−4.2
G158/15	<0.10	0.07	ND[b]
226/14	2.70	2.49	−7.8
272/14[c]	1.29	1.31	1.6
274/14[c]	2.26	2.89	27.9[d]
1/15	0.94	1.03	9.6
49/15	0.48	0.53	10.4
75/15	91.38	91.88	0.5
116/15	1.68	1.92	14.3

Table 5. Accuracy (trueness and precision) determined by analysis of routine and proficiency-test samples at NIB with 6-plex and 11-plex ddPCR assays for quantification of GMO soybean. [a]Assigned GM-event, value determined by NIB using GM-event–specific qPCR methods under ISO 17025 accreditation. [b]Bias cannot be calculated because assigned value is below level of quantification for qPCR, and it can only be determined that there is <0.1% of GM soy present. [c]USDA proficiency test sample. [d]Value not compliant with acceptance values for dynamic range (quantifiability); i.e., $RSD_r \leq 25\%$, and bias ≤25%[26,27].

		Protocol									
		original		1[a]			2[b]			3[c]		
System	dilution	GM-event (%)	RSD %	GM-event (%)	RSD %	Bias %	GM-event (%)	RSD %	Bias %	GM-event (%)	RSD %	Bias %
6-plex	60×	10.54	27.52	11.26	19.55	6.88	8.15	11.29	−22.63	10.73	14.86	1.77
	240×	13.76	68.12	7.49	47.18	45.56	13.53	30.35	−1.64	12.44	23.22	−9.6
7-plex	60×	9.88	8.95	10.04	10.81	1.58	9.11	7.71	−7.79	10.62	12.48	7.47
	240×	12.25	32.64	11.41	66.83	−6.82	16.03	30.76	30.9	9.9	29.92	−19.2
11-plex	60×	20.31	7.2	22.66	37.36[d]	11.54	18.04	9.6	−11.2	17.65	16.36	−13.12
	240×	20.75	26.16	20.47	35.4	−1.37	18.51	27.15	−10.8	22.64	34.22	9.12

Table 6. Robustness testing. [a]Protocol 1: use of automated droplet generator (DG32). [b]Protocol 2: reduced reaction volume (18 µL instead of 20 µL). [c]Protocol 3: use of thermal cycler with slower ramp rate. [d]RSD higher than acceptance value (35%) for dynamic range (quantifiability).

Robustness. Robustness is assessed through deliberate introduction of minor modifications to the protocol, accompanied by a comparison of the results with those obtained with the original protocol. The acceptance criteria adopted here were that the modified protocol should meet the same target values as the original protocol, with accepted deviation of <25%. Robustness was assessed using three deliberate modifications: (1) use of an automated droplet generator (DG32); (2) reduced reaction volume (18 µL instead of 20 µL); and (3) use of a thermal cycler with a slower ramp rate (Table 6). The experiments were carried out with two sample concentrations (dilutions), which were intended to challenge the LOQ and LOD. The three multiplex assays demonstrated compliance with all of the performance parameters and acceptance values[27], with a few minor exceptions observed.

For modification (1), but not for (2) and (3), the 11-plex assay yielded unacceptably high RSD_r for the 60× dilution (Table 6). The observed copy numbers were lower than the estimated LOQ_{abs} for the 60× dilution for the 6-plex and 7-plex assays with the original protocol and with modification (2) (i.e., reduced reaction volume; Supplementary Table S19). For the 240× dilution, the observed copy numbers were lower than the estimated LOD_{abs} for the 6-plex and 7-plex assays in combination with the DG32 protocol (i.e., modification (1)), and also for the 7-plex assay with the alternative thermal cycler protocol (i.e., modification (3); Supplementary Table S19). However, no negative replicates were observed (Table 6 and Supplementary Table S19). Furthermore, all of the reactions were positive, which again indicated that the LOD_{abs} is considerably lower than 20 copies/reaction (Table 6). Taken together, the three multiplex ddPCR assays demonstrated satisfactory robustness.

Practicabillity. The cost-efficiency of applying standard-curve-based single-event specific qPCRs for routine GMO detection has been challenged almost since the first methods were described[30]. Numerous multiplexing methods have been described, but very few of these have become established in routine analyses for various reasons. A common problem with qPCR multiplex methods is that they can be applied for qualitative analysis only, and not for quantification. Other problems include the need to handle amplified DNA for post-amplification detection and identification, the limited number of detection channels, and the potential interference between fluorescence labels in probe-based assays. The first really successful multiplex qPCR assays for GMO detection were qualitative element screening assays that were developed in 2009[30].

System	Tested samples (n)	Relative final price/sample	Hands on time (including analysis) (h)	96-well plates required (n)
Direct quantification of 11 EU authorised GM soybean lines with multiplex ddPCR assays (6-plex and 7-plex)	1	100	4	1
	5	100	7	1
	10	100	8	1
Initial screening with qPCR + identification of five transgenic soybean lines, all negative	1	100	5	2
	5	110	7	3
	10	92	11	5
Initial screening with qPCR + identification of five transgenic soybean lines PLUS quantification of one positive line from screening (5-plex)	1	153	8	3
	5	173	13	5
	10	145	19	9
Initial screening with qPCR + identification of five transgenic soybean lines PLUS quantification of three positive lines from screening (5-plex) + one positive line from identification	1	232	10	5
	5	279	18	9
	10	258	30	17
Direct quantification of 11 GM soybean lines	1	368	14	7
	5	445	31	18
	10	483	63	36

Table 7. Cost-effectiveness of the developed multiplex ddPCR assays.

To determine the cost-effectiveness of the multiplex ddPCRs, a comparison was made between: (i) applications of the multiplex ddPCR; (ii) representative screening with qPCR; and (iii) GM-event–specific qPCRs, for detection of GM soybean lines (Table 7). In these comparisons, a premise was that positive results in screening, were followed by identification and quantification of individual GM-events. The qualitative 5-plex qPCR[13] commonly used for screening covers only six of the 11 authorised soybean lines. For the remaining five, individual identification was needed alongside the 5-plex screening. For this reason, calculations were carried out to compare multiplex ddPCR assays to this most common approach. To evaluate the cost-effectiveness of the multiplex ddPCR assays against the screening approach, comparisons were carried out for four different scenarios: (1) 5-plex screening with qPCR[13] that yielded consistently negative results; (2) presence of only the most commonly detected GM soybean line (MON40-3-2); (3) presence of the four most common soybean lines (MON40-3-2, MON89788, A2704-12, MON87701); and (4) direct quantification of 11 authorised GM soybeans (Table 7). Comparisons were carried out for simultaneous processing of one, five or 10 samples. All of the comparisons were relative, which means that the price of the direct quantification with multiplex ddPCR was set as 100, and the other prices were calculated relative to the ddPCR.

As expected, the multiplex ddPCR approach became considerably more cost-effective (from 45% to 383%) in comparison to qPCR where a sample was positive for at least one GM-events (Table 7). The costs were comparable when the sample was negative, although here the time (in hours) needed to achieve the final analytical result was longer for the qPCR analysis (Table 7). This arose because up to 10 samples can be analysed in a single 96-well plate in multiplex ddPCR, while when qPCR analysis is considered (with the exception of the initial screening and identification of five transgenic lines for one sample), it requires from two (scenario 1, one sample) up to 36 (scenario 4, 10 samples) 96-well plates (Table 7). Therefore, while up to 10 samples (one 96-well plate) can be analysed in 1 day (8 h) with the multiplex ddPCR assay, up to 8 days are needed for the qPCR assays (Table 7). Multiplexing here is thus more cost-effective and has higher throughput.

Recently, a multiplex ddPCR assay for detection and quantification of EU-authorised GM-events in maize was reported[8]. The multiplex ddPCR assays first reported in the present study complement this maize assay. The soybean multiplex ddPCR assays were approved by the Slovene accreditation body, and are included among the methods under ISO 17025 accreditation at NIB. This type of multiplexing is very useful for quantitative detection of GMOs, and potentially also for qualitative and quantitative screening for the presence of specific groups of GMOs.

Methods

Test materials. Certified reference materials (CRMs) were purchased from the Institute for Reference Materials and Measurements, Geel, Belgium (IRMM) and the American Oil Chemists' Society (AOCS), and used for all 15 targeted soybean GM-events. These CRMs are consistently certified for mass/mass (m/m) GM soybean/wild-type soybean ratios. A list of all of the CRMs used in the present study is provided in Supplementary Table S20. Routine diagnostic samples that contained soybean events and samples from the USDA/GIPSA proficiency testing[31] programme were also used in this study (Supplementary Table S13).

DNA extraction and purification. DNA was extracted and purified from 200 mg starting material for all of the samples, using a cetyltrimethyl-ammonium bromide (CTAB) protocol, with RNase-A and proteinase-K

for removal of RNA and protein from the samples, respectively (as described in Annex A.3 of ISO21570:2005[32]), with small modifications as described in Dobnik *et al.* (2015)[8]. Dilutions of the extracted stock DNA solutions were made in nuclease-free and protease-free water (Sigma-Aldrich Chemie GmbH, Munich, Germany). All of the DNA extracts and samples were stored at −20 °C until further use.

Interactions between primers and probes. All primers and probes were investigated for potential interactions through alignment with Autodimer[28]. Autodimer makes ungapped local alignments of all primers and probes in a multiplex PCR assay, giving +1 as a reward for a match, and −1 as a penalty for a mismatch. As recommended, a score of 7 was considered as a significant interaction.

***In-silico* multiplex specificity prediction.** The specificity of the multiplex assays was assessed using electronic PCR (ePCR)[29], the programme that runs primer-BLAST on the web[33]. The *in-silico* specificity of PCRs is usually monitored by alignment of primers to a sequence database; e.g., using Primer-BLAST[33]. In a multiplex assay, any combination of two primers can generate a PCR product. For 6-plex, 7-plex and 11-plex assays, there are 72, 98 and 242 possible combinations of primer pairs, respectively. A custom python script called ePCR multiplex was written to generate all of the combinations of these primers in the format specified for ePCR. The script is deposited at github (https://github.com/karinlag/ePCR_multiplex). The ePCRs were run for each multiplex assay with the complete insert and flanking sequences of a total of 19 GM-soybean events (Supplementary Table S2) combined with the genomes of soybean (RefSeq: GCF_000004515.3) and maize (RefSeq: GCF_000005005.1), while allowing for one mismatch for each primer, but no gaps.

Preparation of primer and probe mixes. To facilitate the direct comparison of quantities of GM-events and the endogene, all of the probes for the GM-events were labelled with 6-carboxyfluorescein (FAM), and the probe for the lectin *Le*1 endogene was labelled with 5′-hexachloro-fluorescein phosphoramidite (HEX; Supplementary Table S21). For the estimation of the relative quantity of GM-soybean events, the *Le*1 gene was used as the endogenous reference gene. For quantification of GM soybean lines, GM-event–specific qPCR modules were used. Primer and probe sets were prepared for the simplex, duplex and multiplex reactions. A total of 15 primer and probe sets were prepared for the simplex reactions that targeted the individual soybean GM-event. A primer and probe set was also prepared for *Le*1. In total, 15 sets for duplex reactions were prepared, each of which contained the primers and probe for *Le*1 and the primers and probe of one of the 15 individual GM-events. The preparation of 6-plex, 7-plex and 11-plex primer and probe mixes was carried out as follows. For 6-plex, the primers and probes for *Le*1 and the five GM-events of MON87701, 305423, MON87708, BPS-CV127-9 and MON87769 were mixed (for final concentrations, see Supplementary Table S21). For 7-plex, the primers, and probes for *Le*1 and the six GM-events of MON40-3-2, A5574-127, A2704-12, 356043, MON89788 and MON87705 were mixed (for final concentrations, see Supplementary Table S21). For 11-plex, the same primers and probes were used as for 7-plex, with the addition of primers and probes for the four soybean GM-events of FG72, DAS-44406-6, DAS-81419-2 and DAS-68416-4. All primers and probes were purchased from Eurofins MWG Operon (Ebersberg, Germany) or from Integrated DNA Technologies (Leuven, Belgium). Primers and probes were shipped lyophilised and diluted in nuclease-free and protease-free water (Sigma-Aldrich Chemie GmbH, Munich, Germany) upon receipt.

Droplet digital PCR and data analysis. All ddPCRs (i.e., simplex, duplex, multiplex) were performed by mixing 10 µL 2 × ddPCR Supermix for probe (no dUTP) (Bio-Rad, Pleasanton, CA, USA), 6 µL of the corresponding 6-plex, 7-plex or 11-plex primer and probe mix, and 4 µL DNA template (or 4 µL nuclease-free and protease-free water for the non-template control), in a 20 µL reaction volume. Droplets were generated in DG8 cartridges, using a QX100 Droplet Digital System droplet generator (Bio-Rad). Water-in-oil emulsions were transferred to a 96-well plate and amplified in a T100 PCR cycler (Bio-Rad). The thermal cycling conditions were: 10 min at 95 °C, followed by 40 cycles of a two-step thermal profile (15 s at 95 °C and 60 s at 60 °C, at a ramp rate of 2.5 °C/s), and final incubation at 98 °C for 10 min, followed by cooling to 4 °C. After the thermal cycling, the plates were transferred to a QX100 Droplet Digital System droplet reader (Bio-Rad). Data acquisition and analysis was performed using the QuantaSoft software, version 1.6.6 (Bio-Rad). Positive droplets (Supplementary Fig. S1) that contained the amplification products were discriminated from negative droplets (Supplementary Fig. S2) without the amplification products by applying a fluorescence amplitude threshold. The threshold was set manually (for channel 1 [FAM] at amplitudes of approximately 13200, 9300 and 18500 for 6-plex, 7-plex and 11-plex, respectively, and for channel 2 [HEX] at amplitudes of approximately 1600, 1700 and 2000 for 6-plex, 7-plex and 11-plex, respectively), using both fluorescence amplitude *versus* event number (i.e., one-dimensional amplitude), and histogram of events *versus* amplitude data streams on both the FAM and 2′-chloro-7′-phenyl-1, 4-dichloro-6-carboxyfluorescein (VIC) channels (i.e., two-dimensional amplitude). The data generated by the QX100 droplet reader were rejected from subsequent analysis if a clog was detected by the Quantasoft software or if a low number of droplets (<8000) was measured per 20 µL PCR. After exporting the data, they were further analysed using Microsoft Excel spreadsheets. A digital 'Minimum Information for Publication of Quantitative Real-Time PCR Experiments' (MIQE) checklist is available in Supplementary Table S22, and the list for the compliance of the data with the minimum performance parameters for method acceptance is available in Supplementary Table S23.

Comparison of simplex, duplex and multiplex reactions. The performances of the 6-plex, 7-plex and 11-plex ddPCRs were compared to both simplex and duplex assays. All of the comparisons were carried out using data obtained with duplicate analyses of DNA samples obtained from single CRMs in two dilutions (i.e., four data points). The mean copy numbers determined by simplex, duplex, 6-plex and 11-plex analyses of single

CRM-derived DNA samples were used to establish the assigned concentrations (copies/µL) for the preparation of dilutions and DNA mixes containing multiple events.

Dynamic range, repeatability, and limits of detection and quantification. A dilution series in terms of target copies was prepared by mixing DNA from all 15 targeted transgenic soybean GM-events. Each GM-event was assigned with 128 target copies/µL, except for A2704-12 and MON89788, which had 49 and 92 copies/µL, respectively. The quantity of each soybean GM-event in the DNA mix was verified by duplex ddPCR (individual GM-event and *Le*1 in the same reaction; Supplementary Table S5). The verified values for the DNA mix were used to calculate the assigned copy numbers of the targets in a dilution series of the DNA mix (referred to as DNA mixes 1–10, with DNA mixes 2–10 representing 6×, 12×, 60×, 120×, 240×, 600×, 1200×, 6000× and 12000× dilutions of DNA mix 1, respectively). The final assigned values of the mixed DNA solutions were from 2527 to 0.02 copies of GM-event targets covered by the 6-plex, from 2761 to 0.05 copies of GM-event targets covered by the 7-plex, and from 4752 to 0.04 copies of GM-event targets covered by the 11-plex. The dilution series for all three multiplexes contained from 26789 to 0.07 copies of the endogene *Le*1. Twelve replicates of the dilution series were measured by ddPCR (three separate runs, in 3 days, each containing four replicates) for the 6-plex and 11-plex, and 14 replicates of the dilution series (three separate runs, in 3 days, with the first two days containing four replicates, and the last day containing six replicates) for the 7-plex. The absolute (copy number) limit of quantification (LOQ_{abs}) and detection (LOD_{abs}) for ddPCR were determined based on these experimental results. Another dilution series was prepared in terms of the GM-event target content, where DNA mix 1 was diluted in non-GM soybean DNA to obtain final concentrations from 0.005% to 2% of the GM-event target relative to *Le*1 (Supplementary Table S11). Three replicates of this dilution series were measured by ddPCR for each multiplex. The relative (%) limit of quantification (LOQ_{rel}) and detection (LOD_{rel}) for ddPCR were determined based on these experimental results. The LOQ (both absolute and relative) was determined as the lowest concentration in a series of concentrations that consistently yielded a relative standard deviation (RSD) <25%[26,27]. The LOD (both absolute and relative) was determined as the lowest concentration in a series of concentrations that consistently yielded positive signals with all replicates[26,27].

Fitness for purpose and inter-laboratory comparison. To determine whether the new assays were fit for purpose, two different samples from the USDA/GIPSA proficiency-testing programme[31] and samples obtained from routine diagnostics at NIB (Supplementary Table S13) were tested with all three multiplex assays. To obtain robust means of the GM-event content, the values per event reported in the proficiency-test reports were used to calculate the sum of all of the present GM-events, and finally to calculate the standard deviation. For routine diagnostic samples, the additive concentrations of the soybean GM-events were calculated based on the results of the quantification with GM-event-specific qPCRs performed at NIB, under ISO 17025 accreditation. However, the quantitative qPCR results were not always available for all of the detected GM-events in a sample. Complementary simplex ddPCRs were performed in these cases to obtain the quantitative data for all of the GM-events detected as present in a sample. All of the proficiency-test samples and routine samples were tested on two different days, and for each day at two concentrations and with two replicates (altogether eight data points per sample/assay). The inter-laboratory comparisons were carried out by transfer of protocols and samples to the NVI. Only the 6-plex and 7-plex assays were tested at the NVI, using the proficiency-test samples and routine samples (all except G158/15 and 49/15).

Robustness. The robustness of the ddPCR multiplex assays was evaluated by comparison of the original protocol to three modified protocols. These included the use of the following: (1) automated droplet generator (AutoDG; Bio-Rad); (2) lower reaction volume (18µL instead of 20µL), and (3) PCR cycler with slower ramp rate (GeneAmp 9700 PCR cycler; Applied BioSystems, Foster City, CA, USA). Two samples from the repeatability experiments, one close to the LOQ (DNA mix 60× dilution) and one close the LOD (DNA mix 240× dilution) were tested in triplicate in the robustness experiments.

References

1. ISAAA. ISAAA GM Approval Database. (2016). Available at: http://www.isaaa.org/gmapprovaldatabase/. (Accessed: 8 June 2016).
2. Biosafety Clearing House. LMO quick links. (2013). Available at: https://bch.cbd.int/resources/quicklinks.shtml. (Accessed: 8 June 2016).
3. United States Department of Agriculture. Petitions for Determination of Nonregulated Status. (2016). Available at: https://www.aphis.usda.gov/biotechnology/petitions_table_pending.shtml. (Accessed: 8 June 2016).
4. Strategic Plan for the Cartagena Protocol on Biosafety for Period 2011–2020 (BS-V/16, Annex I). (2014). Available at: https://bch.cbd.int/protocol/issues/cpb_stplan_txt.shtml#oo1_1. (Accessed: 8 June 2016).
5. Vigani, M., Raimondi, V. & Olper, A. International trade and endogenous standards: the case of GMO regulations. *World Trade Rev.* **11**, 415–437 (2012).
6. Gruère, G. P. A Review of international labeling policies of genetically modified food to evaluate India's proposed rule. *AgBioforum* **10**, 51–64 (2007).
7. European Commission. Regulation (EC) No. 1829/2003 of the European Parliament and of the Council of 22 September 2003 on genetically modified food and feed. *Off. J. Eur. Union L* 1–23 (2003).
8. Dobnik, D., Spilsberg, B., Bogožalec Košir, A., Holst-Jensen, A. & Žel, J. Supporting Information - Multiplex quantification of 12 European union authorized genetically modified maize lines with droplet digital polymerase chain reaction. *Anal. Chem.* **87**, 8218–8226 (2015).
9. Randhawa, G. J., Chhabra, R. & Singh, M. Multiplex PCR-based simultaneous amplification of selectable marker and reporter genes for the screening of genetically modified crops. *J. Agric. Food Chem.* **57**, 5167–5172 (2009).
10. Randhawa, G. J., Chhabra, R. & Singh, M. Decaplex and real-time PCR based detection of MON531 and MON15985 Bt cotton events. *J. Agric. Food Chem.* **58**, 9875–9881 (2010).
11. Luan, F. X., Tao, R., Xu, Y. G., Wu, J. & Guan, X. J. High-throughput detection of genetically modified rice ingredients in foods using multiplex polymerase chain reaction coupled with high-performance liquid chromatography method. *Eur. Food Res. Technol.* **234**, 649–654 (2012).

2.3 Droplet Volume Variability as a Critical Factor for Accuracy of Absolute Quantification Using Droplet Digital PCR

Alexandra Bogožalec Košir, Carla Divieto, Jernej Pavšič, Stefano Pavarelli, David Dobnik, Tanja Dreo, Roberto Belloti, Maria Paola Sassi, Jana Žel

Analytical Bioanalytical Chemistry, 2017, 409(28): 6689-6697. DOI: 10.1007/s00216-017-0625-y

In this publication, we used an optical profilometer and optical microscopy to determine one of the critical factors of absolute quantification accuracy in ddPCR, the partition volume. Partition volume, in our case droplet volume, was measured for two types of BioRad droplet generators, manual and automated, at NIB. Additionally, manual droplet generator was tested at the Italian Metrological Institute (INRiM). As different chemistries can have an effect on droplet volume, two different mastermixes were tested for manual droplet generator. A thorough statistical analysis was conducted to compare and study all of the different scenarios and to determine the droplet volume for each one. Two droplet volume measurement protocols, traceable to the SI, were compared and evaluated. Lastly, a ddPCR experiment was performed and the data analysis conducted using the newly gained knowledge of droplet volume. A comparison of results with the new and assigned droplet volume was conducted.

The PhD candidate designed the experiments. She conducted the droplet volume measurements, according to the automatic measurement protocol, at NIB and performed the ddPCR including the data analysis. The candidate also contributed to the statistical analysis of droplet volume and to the preparation of tables and figures. The candidate also performed the comparisons between the measurement protocols and operators in the two cooperating laboratories. She drafted the manuscript including the supplement.

Anal Bioanal Chem
DOI 10.1007/s00216-017-0625-y

RESEARCH PAPER

Droplet volume variability as a critical factor for accuracy of absolute quantification using droplet digital PCR

Alexandra Bogožalec Košir[1,2] · Carla Divieto[3] · Jernej Pavšič[1,2] · Stefano Pavarelli[3] · David Dobnik[1] · Tanja Dreo[1] · Roberto Bellotti[3] · Maria Paola Sassi[3] · Jana Žel[1]

Received: 19 July 2017 / Revised: 30 August 2017 / Accepted: 5 September 2017
© The Author(s) 2017. This article is an open access publication

Abstract Accurate and precise nucleic-acid quantification is crucial for clinical and diagnostic decisions, as overestimation or underestimation can lead to misguided treatment of a disease or incorrect labelling of the products. Digital PCR is one of the best tools for absolute nucleic-acid copy-number determination. However, digital PCR needs to be well characterised in terms of accuracy and sources of uncertainty. With droplet digital PCR, discrepancies between the droplet volume assigned by the manufacturer and measured by independent laboratories have already been shown in previous studies. In the present study, we report on the results of an inter-laboratory comparison of different methods for droplet volume determination that is based on optical microscopy imaging and is traceable to the International System of Units. This comparison was conducted on the same DNA material, with the examination of the influence of parameters such as droplet generators, supermixes, operators, inter-cartridge and intra-cartridge variability, and droplet measuring protocol. The mean droplet volume was measured using a QX200™ AutoDG™ Droplet Digital™ PCR system and two QX100™ Droplet Digital™ PCR systems. The data show significant volume differences between these two systems, as well as significant differences in volume when different supermixes are used. We also show that both of these droplet generator systems produce droplets with significantly lower droplet volumes (13.1%, 15.9%, respectively) than stated by the manufacturer and previously measured by other laboratories. This indicates that to ensure precise quantification, the droplet volumes should be assessed for each system.

Keywords Droplet digital PCR · Droplet volume · DNA quantification · Optical microscopy imaging

Alexandra Bogožalec Košir and Carla Divieto contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00216-017-0625-y) contains supplementary material, which is available to authorized users.

✉ Alexandra Bogožalec Košir
alexandra.bogozalec@nib.si

[1] Department of Biotechnology and Systems Biology, National Institute of Biology, Večna pot 111, 1000 Ljubljana, Slovenia

[2] Jožef Stefan International Postgraduate School, Jamova 39, 1000 Ljubljana, Slovenia

[3] Instituto Nazionale di Ricerca Metrologica, Strada delle Cacce, 91-10135 Turin, Italy

Introduction

Accurate and precise nucleic-acid quantification is fundamental in many fields, from basic research to molecular diagnostics, and in preclinical and clinical research and industrial processes. Absolute DNA quantification allows the reduction of measurement bias between laboratories, which is essential for target DNA measurements in medicine, viral load analysis in diagnostics, and microbial quantification in microbiology. Currently, the most used method for DNA quantification is real-time polymerase chain reaction, which depends on the standard curve approach for quantification. However, the choice of the standard can in some fields be arbitrary and is often not uniform among laboratories [1]. Absolute quantification of nucleic acids without the need of a standard curve can be achieved with digital PCR (dPCR). Taking into account the high potential of absolute quantification by dPCR, this technique represents a good candidate for reference methods for nucleic-acid copy-number determination [1, 2].

Published online: 18 September 2017

Although the concept of dPCR was defined in the 1990s [3, 4], dPCR has gained further popularity in more recent years [1, 5–14], with 442 papers listed in the Scopus database for 2016, compared to 125 in 2012 (Fig. S1 in the Electronic Supplementary Material, ESM). The main characteristic of dPCR is the dispersal of the reaction volume into partitions, either chambers in chamber/microfluidics-based digital PCR, or droplets in droplet digital PCR (ddPCR). The underlying assumptions of dPCR which need to be met include a random distribution of DNA molecules into the partitions and amplification of a single target in a partition [15, 16]. To achieve this, the sample and reaction preparation need to be carefully controlled, and the dPCR needs to be well characterised in terms of accuracy and sources of uncertainty.

In dPCR, the copy number concentration in the sample is calculated according to Eq. (1),

$$Tc = -ln\left(1-\frac{P}{R}\right) \times \left(\frac{1}{Vd}\right) \times D \quad (1)$$

where Tc is the mean target concentration (copies/μL), P is the number of positive partitions, and R is the number of analysed partitions. The copy number concentration depends also on the partition volume (Vd) and the dilution factor of the original solution in the PCR reaction (D). Consequently, the accuracy of absolute quantification is strongly dependent on correct determination of the partition volume, which is the droplet volume for ddPCR, and on the number of partitions. The confidence of measurements can thus be enhanced by accurate determination of the partition volume and by raising the number of reactions (i.e. the number of partitions) for any one sample.. If the droplet volume varies among reactions or droplet generators, the copy number can be either underestimated or overestimated, depending on whether the droplet volume is overestimated or underestimated, respectively. Additionally, as was clearly illustrated by Huggett et al. [5], if there is partition volume variability within one reaction, this can lead to a smaller number of partitions with a positive signal [5]. Hence, if partition volume variability is not considered in the calculation of DNA copy numbers, bias in the estimated number of copies per reaction can be attributed to incorrect droplet volume assessment and to non-homogeneous distribution of molecules among partitions.

Discrepancies between droplet volumes assigned by the manufacturer and measured by independent laboratories have been shown in previous studies [17–19], where the droplet volumes were measured for QX100™ Droplet Digital™ PCR systems (hereafter indicated as a DG8 manual droplet generator; Bio-Rad). The QuantaSoft™ software version 1.3.2.0 used the pre-set volume of 0.91 nL [18]. Further on, droplet volumes were measured independently at the Australian National Measurement Institute (NMIA) and by Bio-Rad. NMIA measured droplets in 16 wells across five separate cartridges and determined a volume of 0.868 nL with a relative standard deviation of the mean inter-well droplet volume of 2.8%. The Bio-Rad measure for droplets generated from the eight-channel droplet generator cartridges was 0.89 nL [17], which showed that the droplet volume was smaller than previously believed. Additionally, even lower droplet volume, measured as 0.834 nL, was determined by Corbisier et al. [18]. It is prudent here to point out that measurements at the NMIA were conducted on an early access beta-prototype ddPCR system; thus, it is not necessary that this system generated droplets of same volume as current versions of QX100™ Droplet Digital™ PCR and QX200™ AutoDG™ Droplet Digital™ PCR system. And although the same droplet volume was pre-set regardless of the droplet generator used, a direct comparison cannot be made.

Abovementioned measures by Corbisier et al. [18] were performed on a discontinued series of cartridges (DG8-186-3008). More recently, a new series of cartridges (DG8-186-4008) became available on the market. At the same time the QuantaSoft™ software was updated and the new version had a new droplet volume of 0.85 nL incorporated in the calculations. Although there have been assessments of supermix effects on droplet volume (ddPCR™ Supermix™ for Probes [17–19]; ddPCR™ Supermix™ for Probes (no dUTP) [19]; QX200™ ddPCR™ Eva Green™ Supermix™ [18]), all of the studies to date have been focused only on the DG8 manual droplet generator and to the best of our knowledge no study was conducted on the new series of cartridges.

In the present study, the partition volume was considered in an inter-laboratory comparison with different methods for volume determination. A comparison of droplet volume measurements was conducted on the same DNA material, while taking into account further parameters, including droplet generator, supermix type, laboratory operator, inter-cartridge and intra-cartridge variability, and droplet measuring protocol. Mean droplet volumes were measured for a QX200™ AutoDG™ Droplet Digital™ PCR system (hereafter indicated as DG32 automated droplet generator) at the National Institute of Biology (NIB; Ljubljana, Slovenia) and for two DG8 manual droplet generators, one at NIB and the other at the Italian National Metrology Institute - Instituto Nazionale di Ricerca Metrologica (INRiM; Turin, Italy).

Materials and methods

Test material, and methods

The American Oil Chemists' Society 0707-B4 certified reference material that contains 99.99 mass/mass% of A2704-12 (ACS-GMØØ5-3) genetically modified (GM) soybean was used as the DNA sample. A duplex assay that targets the

GM event (A2704-12) and the endogene (*Le1*) was used. The primer and probe sequences were taken from GMOMETHODS, a European Union database of reference methods for analysis of genetically modified organisms (GMOs) (http://gmo-crl.jrc.ec.europa.eu/gmomethods/), with the entry names of QT-EVE-GM-004 and QT-TAX-GM-002 for A2704-12 and lectin, respectively. To exclude the sample as a cause of droplet volume variation, the same sample was used for the measurement of all droplets at NIB and INRiM.

Droplet volume determination

Experimental design

The volume of the droplets was calculated by measuring the individual droplet diameters. Droplet volume was determined for three different experimental set-ups: (1) DG8 manual droplet generator (Bio-Rad, Pleasanton, CA, USA) and ddPCR Supermix™ for probes (no dUTP) (Bio-Rad, Pleasanton, CA, USA); (2) DG32 automated droplet generator (Bio-Rad, Pleasanton, CA, USA) and ddPCR Supermix™ for probes (no dUTP); and (3) DG8 manual droplet generator and QX200 ddPCR EvaGreen™ Supermix™ (Bio-Rad, Pleasanton, CA, USA) (Fig. S2, see ESM). For experimental set-ups (1) and (2), three cartridges were analysed over three consecutive days, and for experimental set-up (3), three cartridges were analysed on the same day. Experimental set-up (1) was carried out using two different machines (one at NIB, one at INRiM), to determine the variability of the DG8 droplet generators. The same reaction mix containing supermix (either ddPCR Supermix™ for probes (no dUTP) or EvaGreen™ Supermix™), DNA, primers and probe (200 nM final concentration) was used for all set-ups. The same Lot N° of DG8 cartridges (000036132) was used in experimental set-ups (1) and (3). For experimental set-ups (1) and (2), six wells in each cartridge were randomly selected, and for experimental set-up (3), four wells in each cartridge were randomly selected.

Optical microscopy

Droplet preparation and optical microscopy imaging were performed on the same day. After droplet generation, 10 μL of droplets was transferred into 24-well polystyrene cell culture plates (Corning Inc., Corning, NY, USA). The droplets were left to form a uniform monolayer by positioning the plates at an angle of approximately 45° to the horizontal axis for a few seconds. The droplets for each of the experimental set-ups were imaged using an optical microscope (NIB: Nikon Eclipse™ T*i*-E; INRiM: Zeiss Axio Observer™ Z1), with monochromatic, high-resolution digital CCD cameras (NIB: Clara™, Andor Technology; INRiM: digital Olympus XM10™). All of the images were recorded in bright field under uniform illumination and × 100 apparent magnification. Digital images were captured using Nikon NIS-Elements™ Advanced Research software at NIB, and the Olympus cellSens™ microscope imaging software at INRiM. The droplets were imaged immediately after their formation, to avoid evaporation and shrinking (Fig. S3, see ESM). An average of six images per well was acquired, and 86 to 284 droplets were measured for each well. For each experimental set-up, the following total number of droplets were measured: set-up (1) at NIB, 3091 droplets; set-up (1) at INRiM, 3567 droplets; set-up (2), 3533 droplets; and set-up (3), 2140 droplets (Fig. S2, see ESM).

Optical profilometer

To ensure traceability to the International System of Units (SI), an optical profilometer (Plμ 2300™, Sensofar) and software for image analysis were used as the reference method for droplet diameter measurements. The software was validated within the EURAMET iMERA-Plus "Traceable Characterisation of Nanoparticles" project [20]. The x–y measurement system was traceable to the SI, so the diameter measurement was traceable to the SI. The optical profilometer and the entire reference method are described in Methods S1 in the ESM.

Image analysis and validation

The Fiji™ software [21], a distribution of ImageJ™ [22, 23], was used to analyse the digital images of the droplets. Two droplet measurement protocols were compared, both of which used the Fiji™ software. The droplet measurement protocol referred to as 'manual' was developed at INRiM and is described in Methods S2 in the ESM. The traceability of the diameter measurements to the SI unit of 'metre' was assured by comparison of the manual and automated droplet measurements with the optical profilometer-based method, which was used as a reference method, as described in Methods S1 (see ESM). The 'automatic' droplet measurement protocol was adapted from Pinheiro et al. [17], and is described in Methods S3 in the ESM. The automatic droplet measurements were used to compare the data from different operators, one at NIB and one at INRiM, and to eliminate the possibility of any *between-operators effect*. For this, images obtained from experimental set-up (1) for day 1 at INRiM were analysed independently by two different operators. All of the sequential image analysis was performed with the automatic droplet measurement protocol, at both NIB and INRiM.

Droplet digital PCR

PCR conditions

A 20-μL reaction mixture was prepared comprising of 10 μL ddPCR Supermix™ for probes (no dUTP) (Bio-Rad,

Pleasanton, CA, USA), 6 μL primers and probe mix, and 4 μL DNA. The final concentration of both primers and probe was 200 nM. For each droplet generator, 24 ddPCR reactions were performed on three consecutive days, which included six negative template controls. The amplification conditions were 10 min DNA polymerase activation at 95 °C, followed by 40 cycles of a two-step thermal profile of 30 s at 94 °C for denaturation, and 60 s at 60 °C for annealing and extension, followed by a final hold of 10 min at 98 °C for droplet stabilisation, and cooling to 4 °C. A thermal cycler (T100™; Bio-Rad, Pleasanton, CA, USA) was used, and the temperature ramp rate was set to 2.5 °C/s, with the lid heated to 105 °C, according to the Bio-Rad recommendations.

Data analysis

After the thermal cycling, the plates were transferred to a droplet reader (QX100™; Bio-Rad, Pleasanton, CA, USA). The software package provided with the ddPCR system was used for data acquisition (QuantaSoft™ 1.6.6.0320; Bio-Rad). The rejection criterium for the exclusion of a reaction from subsequent analysis was a low number of droplets measured (< 10,000 per 20 μL PCR). The data from the ddPCR are given in target copies/μL reaction.

Analysis of results

Statistical analysis of droplet volume

To evaluate the differences between the droplet volumes calculated for each individual droplet generator, single-factor analysis of variance (ANOVA) and Tukey's tests were used in R studio™, version 1.0.136 [24]. Outliers were determined for each experimental set-up independently, using Grubbs tests (R studio™, 'outliers' package), and were excluded from further analysis.

The bottom-up approach was used to determine the expanded measurement uncertainty [25]. A combined measurement uncertainty u_c was calculated according to Eq. (2) [26],

$$u_c = \sqrt{u_r{}^2 + u_{ip}{}^2} \tag{2}$$

where u_r is the uncertainty associated with the repeatability, and u_{ip} is the uncertainty associated with the intermediate precision. u_r and u_{ip} (u_{ip1} or u_{ip2}) were calculated according to Eqs. (3), (4) and (5):

$$u_r = \frac{\sqrt{MS_{within}}}{\sqrt{n}} \tag{3}$$

$$u_{ip1} = \sqrt{\frac{MS_{between} - MS_{within}}{n \times N}} \tag{4}$$

$$u_{ip2} = \frac{\sqrt{\frac{MS_{within}}{n}} \times \sqrt[4]{\frac{2}{N \times (n-1)}}}{\sqrt{N}} \tag{5}$$

where n is the number of independent replicates per experiment, N is the number of experiments performed on one platform, MS_{within} is the mean square value within groups, and MS_{between} is the mean square value between groups. Both of these mean squares were calculated using ANOVA in Microsoft Excel™ 2016, with all of the measurements taken into account for each droplet generator. If $MS_{\text{between}} > MS_{\text{within}}$, Eq. (4) was used to calculate the intermediate precision, otherwise Eq. (5) was used. A coverage factor (k) of 2.1 was chosen at the 95% level of confidence, based on the degrees of freedom, and was applied to obtain the expanded measurement uncertainty.

Results and discussion

Digital PCR is becoming more and more accepted as a tool for absolute quantification, and thus there is the need to assure the accuracy of its copy-number determination. For this, it is critical to determine the droplet volume, as this is one of the key factors for accuracy of absolute quantification in ddPCR. Earlier studies have shown that differences in droplet volumes can arise due to different supermix types [18, 19], with potential differences also seen for different droplet generator [18]. For the purpose of an inter-laboratory comparison of how different factors affect droplet volume, the mean droplet volume was measured independently both at NIB and INRiM using the same droplet measurement protocol. For this purpose, three experimental set-ups were prepared: (1) DG8 manual droplet generator and ddPCR Supermix™ for probes (no dUTP); (2) DG32 automated droplet generator and ddPCR Supermix™ for probes (no dUTP); and (3) DG8 manual droplet generator and QX200 ddPCR™ EvaGreen™ Supermix™ (Fig. S2 in the ESM).

Droplet diameter and droplet volume measurements are traceable and accurate

To calculate the droplet volume, measurements of droplet diameters were obtained using microscopic imaging. The volume measurement must be metrologically traceable [27]. To determine the SI traceability and the accuracy of the droplet imaging protocols for droplet diameter measurement, the diameters obtained by the manual and automatic droplet measurement protocols were compared with results obtained using an optical profilometer-based reference method (Tables S1

and S2 in the ESM). The reference method and the traceability process are fully described in Methods S1 in the ESM. The accuracy of the reference method was higher than 0.8%, while the variability of the droplet diameter measurement was 2% (k = 1) (Table S1 in the ESM). Data from the comparison between the reference method and each of the manual and automatic droplet measurement protocols showed that the results were comparable with the reference method within a bias of 1.3 and 0.9%, respectively (Table S2 in the ESM). The optical-microscopy-based manual and automatic droplet measurement protocols can thus be considered traceable to the SI through comparison with the reference method.

A comparison between the manual and automatic droplet measurement protocols was used on three sub-sets of measurements (sub-sets of wells 1 and 2 in experimental set-up (1) for day 1 at NIB). Data from these three sub-sets showed that the droplet measurement protocols give comparable results, with the maximum of 1.9% bias, and comparable relative standard deviations (RSDs) (Table S3 in the ESM). As the automatic droplet measurement protocol is quicker, further analysis was conducted using this automatic droplet measurement only. Additionally, a validation of the image analysis used for the automatic droplet measurement protocol was performed, to eliminate the possibility of differences in volume due to different operators. No *between-operators effect* was observed: in both cases, the mean determined droplet volume was 0.708 nL, with standard deviations of 0.025 and 0.022 for NIB and INRiM, respectively (Table S4 in the ESM). The bias between the droplet volumes was − 0.02%, and the bias between the mean area equivalent diameters was − 0.01%. As different operators did not have any effects on the droplet volume measurements, comparisons were then possible for the droplet volumes measured in the different laboratories following the same droplet measurement protocol.

Droplet generators have an effect on droplet volume

The mean droplet volume for experimental set-ups (1) and (2) was measured for each of the analysis days (Table 1). The mean volume of the droplets generated by the DG8 manual droplet generator was 0.715 nL at NIB and 0.720 nL at INRiM (Table 1). No significant differences were observed between the droplet volumes for each of the DG8 droplet generators (p value, 0.939). The RSDs for the inter-well mean droplet volumes were 5.46 and 4.37% for the DG8 manual droplet generator at NIB and the DG8 manual droplet generator at INRiM, respectively. The mean droplet volume between the two laboratories was calculated as 0.718 nL, which is considerably lower than the 0.850 nL that is used in the QuantaSoft software™ from version 1.6.6.0320 onwards. The measured volume is also lower than the 0.834 nL determined by Corbisier et al. [18] and the volumes measured at the NMIA, from where volumes of 0.868 [17] and 0.833 [18] nL were reported. However, a direct comparison cannot be made, as Corbisier et al. used a different supermix (ddPCR Supermix™ for probes) and different cartridges, and as mentioned above, droplets were generated using a beta version of the droplet generator at NMIA. Nevertheless, the difference in droplet volume shows that a fixed volume used in QuantaSoft software™ may cause bias in copy-number determination. The droplet volumes determined in the present study are closer to the values measured at the National Institute of Standards and Technology (USA) in 2016 [19], where the mean droplet volume for ddPCR Supermix™ for probes (no dUTP) was 0.780 and 0.767 nL using diluted and concentrated methods, respectively. The principle of the diluted method is the same as that used in the present study.

The mean droplet volume measured for the DG32 droplet generator was 0.739 nL, with RSD of 5.84% (Table 1). This was significantly different to the DG8 results for both NIB and INRiM (p values, 0.011, 0.029, respectively) (Fig. 1). Although the DG32 droplet generator volume was larger than the volume measured for both DG8 manual droplet generators, the volume was smaller than those reported in previous studies [17–19].

The same Lot N° of cartridges was tested with each of the DG8 manual droplet generators, which indicated that the DG8 manual droplet generator on its own does not have any effect on droplet volume. However, the droplet volume for the DG32 automated droplet generator was significantly higher than that of DG8. This can lead to the conclusion that droplet volumes can vary depending to the type of droplet generators (i.e. DG8 and DG32 droplet generators).

Intra-cartridge and inter-cartridge variability is not significant

The intra-cartridge and inter-cartridge droplet volume variabilities were determined for all of the experimental set-ups. Grubbs tests showed that none of the wells can be considered as an outlier for the DG8 droplet generator at NIB, while there was one outlier (day 1) for the DG8 droplet generator at INRiM and two outliers for the DG32 droplet generator (days 1, 3) (Table S5 in the ESM). The use of ANOVA followed by Tukey's tests showed no significant intra-cartridge (p values, 0.206 to 0.761) and inter-cartridge (p values, 0.151 to 0.252) effects. Thus, potential effects that a specific position in a cartridge (specific well) would have on copy number determination for a sample can be excluded as a significant uncertainty source that affects the accuracy of ddPCR in the present study.

Supermix™ has a significant influence on droplet volume

The mean droplet volume measured for the three cartridges using EvaGreen™ Supermix™ was 0.778 nL, with RSD of

Table 1 Comparisons of the mean area equivalent diameters and mean droplet volumes (±standard deviation) produced by different droplet generators for the three individual analysis days

Droplet generator	Measurement day	Area equivalent diameter (μm)	Droplet volume (nL)	Expanded measurement uncertainty (nL)
DG8 NIB	1	110.97 ± 2.00	0.716 ± 0.04	
	2	110.23 ± 1.89	0.702 ± 0.03	
	3	111.33 ± 2.08	0.723 ± 0.04	
	Mean*	110.89 ± 2.05	0.715 ± 0.04[a]	0.014
DG8 INRiM	1	110.58 ± 1.17	0.708 ± 0.02	
	2	111.47 ± 1.46	0.725 ± 0.03	
	3	111.50 ± 2.18	0.726 ± 0.04	
	Mean*	111.28 ± 1.73	0.720 ± 0.03[a]	0.013
DG32	1	112.73 ± 1.24	0.750 ± 0.02	
	2	112.25 ± 2.32	0.741 ± 0.04	
	3	111.52 ± 2.65	0.727 ± 0.05	
	Mean*	112.12 ± 2.28	0.739 ± 0.04[b]	0.016

Data with different superscript letters are significantly different ($p < 0.05$; ANOVA)

*Means are calculated from all of the accepted droplets during all 3 days of measurement

3.08% (Table S6 in the ESM). With the same DNA sample and the same Lot N° of the cartridges used in the comparisons of different droplet generators, the droplet volume was significantly larger when EvaGreen™ Supermix™ was used, compared to the droplet volume measured when using Supermix™ for Probes (p value, < 0.01) (Fig. 2). The droplet volume is thus not only dependent upon the type of droplet generator, but also on the type of supermix.

Differences in copy-number determination are correlated with droplet volume

Copy numbers were determined for one dilution (100×) of the certified reference material that contained 99.9% A2704-12 GM soybean. The copy numbers were determined over three consecutive days to ensure sufficient numbers of test replicates ($n = 18$), repeatability and reproducibility, to be consistent with the 'Definition of Minimum Performance Requirements for Analytical Methods of GMO Testing' and the 'Minimum Performance Parameters and Validation Aspects on PCR Amplification' defined within the Decathlon Project [28, 29]. Each day one cartridge was prepared, with six replicates of reference DNA and two replicates of negative template control. Negative control wells were positioned randomly. Droplets were prepared with both the DG8 and DG32 droplet generators. The data were analysed using both the droplet volume of 0.85 nL, as specified by the manufacturer, and the volumes measured within this study,

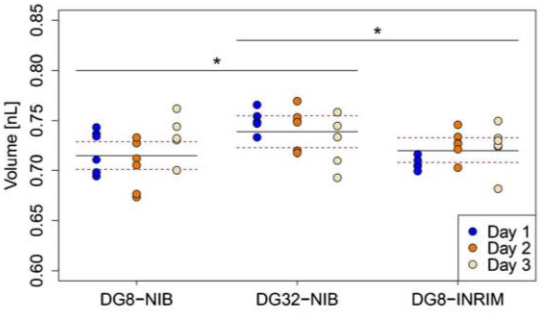


Fig. 1 Inter-cartridge variability and differences between the droplet generators. The inter-cartridge variability is shown by the dispersal of the dots that represent mean droplet volumes of one well in a cartridge on that day. The full black line indicates mean droplet volume and the dotted red lines indicate expanded measurement uncertainty. The droplet volumes between the DG8 (manual) and DG32 (automatic) droplet generators are significantly different (*$p < 0.05$; ANOVA), while there are no significant differences between the two DG8 droplet generators

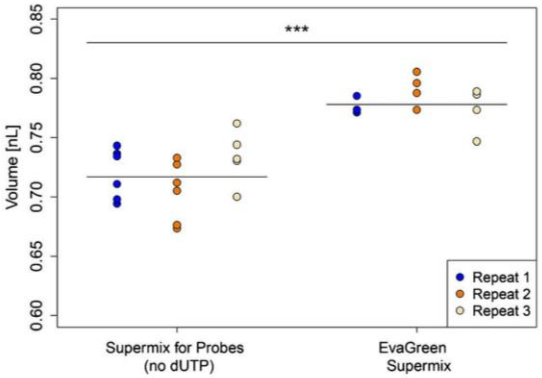


Fig. 2 Comparison of droplet volumes between the two different supermixes, Supermix™ for probes (no dUTP) and EvaGreen™ Supermix™. The inter-cartridge variability is shown by the dispersal of the dots that represent mean droplet volumes of one well in a cartridge for each of the three repeats. The full black line indicates mean droplet volume. The droplet volumes between the different supermixes are significantly different (***$p < 0.001$; ANOVA)

0.715 nL for the DG8 manual droplet generator, and 0.739 nL for the DG32 automated droplet generator.

The relative repeatability standard deviation (RSDr) for the copy number, which provides the precision of the method, was 5.73% for A2704-12 and 5.37% for *Le1* using the DG8 droplet generator, and 5.69% for A2704-12 and 5.39% for *Le1* using the DG32 droplet generator. These are well below the minimum performance criteria of ≤ 25% (Tables S7 and S8 in the ESM) [28, 29]. The mean copy number for all of the 18 measurements with the DG8 manual droplet generator was compared to the measurements obtained with the DG32 automated droplet generator, and the bias between these copy-numbers was calculated (Table 2).

The relative reproducibility standard deviation (RSD_R) was calculated, which is defined as the relative standard deviation of the results obtained with the same method on the same test items using different equipment and with different operators. For a method to be compliant with the minimum performance parameters, the RSD_R should be ≤ 35% [28, 29]. The RSD_R between the DG8 manual droplet generator and the DG32 automated droplet generator using the fixed droplet volume was 5.65% for A2704-12 and 5.37% for *Le1*, and using the measured droplet volume was 6.05% for A2704-12 and 5.86% for *Le1* (Table S9 in the ESM). Thus, although the ANOVA shows that there are statistically significant differences between the droplet volumes for the DG8 manual droplet generator and the DG32 automated droplet generator, the difference has less of an effect on the copy number than the overall difference in droplet volume between 0.85 nL and the measured volume. If the measured droplet volumes are used, the bias in the DNA copy number calculated for the DG8 and DG32 systems are 15.9 and 13.1%, respectively (Table S10 in the ESM). This bias is within the limits set in the currently accepted minimum performance parameters [28, 29]. The bias due to the droplet volume measurements can be taken into account as components of the combined uncertainty of the DNA copy number measured by ddPCR.

In the field of GMO diagnostics, apart from copy number, the GM% (percentage of genetically modified ingredient) is also critical, and is calculated as the ratio between GM and endogene copy number. As the droplet volume affects both of the PCR amplicons, the GMO (A2704-12) and endogene (*Le1*), there is no difference in GM% when comparing fixed and measured droplet volumes. In both cases, the mean GM% with expanded measurement uncertainty was 96.76 ± 2.29 and 97.24 ± 1.76 for the DG8 manual droplet generator and the DG32 automated droplet generator, respectively. Unlike the case of copy number (Table 2), although the expanded measurement uncertainty for the DG8 manual droplet generator was still very low, it was higher than that of the DG32 automated droplet generator. However, the bias between the GM% of the DG8 manual droplet generator and the DG32 automated droplet generator was − 0.5%, which is again well within the limits set in the minimum performance parameters [28, 29].

Conclusions

This study shows the importance of correct droplet volume determination for absolute quantification of nucleic acids by ddPCR, and highlights the factors that can influence the droplet volume. The droplet volume was based on the measurement of the droplet diameter, from which the droplet volume was calculated, assuming the droplets are (nearly) perfect spheres. To make sure the droplet measurements were reliable, and thus provided accurate results, a reference method traceable to the SI was used for diameter measurements and compared with the two methods used in the present study. This is a calibrated profilometer that is directly traceable to the SI, and was validated in an inter-laboratory comparison study among national metrology institutes. This allowed metrological validation of the methods used, which assures traceable results, and provides a verifiable link to the relevant SI unit—'metre'. Additionally, the reference method (using the optical profilometer and the unit cell software method) was applied to measure the minor and major axes of the

Table 2 Mean estimated copy numbers of the A2704-12 and *Le1* targets in the certified reference material, and the expanded measurement uncertainty, and bias between copy numbers obtained with the DG8 and DG32 droplet generators, for the fixed (0.85 nL) and measured (0.715 nL, 0.739 nL, for DG8, DG32, respectively) droplet volumes

Droplet volume (nL)	Target	DG8		DG32		Bias copies/μL DG8/DG32 (%)
		Copies/μL	Measurement uncertainty	Copies/μL	Measurement uncertainty	
0.85	A2704-12	115,351	6034	114,131	5934	1.07
	Le1	119,230	5855	117,369	5789	1.59
0.715/0.739	A2704-12	137,131	7174	131,274	6826	4.46
	Le1	141,742	6961	134,998	6659	5.00

droplets and an average difference between the axes length of 1.8% was obtained. Thus, the assumption that droplets are (nearly) perfect spheres was taken. The microscopy imaging procedure was well characterised when the effects of the dimension calibration of the microscopes were taken into account. The results of the inter-laboratory comparisons between NIB and INRiM were highly comparable, and revealed a droplet volume that was significantly smaller than the previous values measured and used by the manufacturer. The results from this study show that, if the droplet volume is not accurately measured, this can be a source of bias that will influence the DNA copy number measured using this ddPCR platform. Indeed, several factors can affect the droplet volume, which confirms some previous data and gives some new insights into the DG32 system. The mean droplet volumes measured in this inter-laboratory study for experimental set-ups (1), (2) and (3) were 0.715, 0.739 and 0.778 nL, respectively. These are significantly lower than the 0.85 nL that is used in version 1.7.4 of the QuantaSoft™ software from Bio-Rad. The application of an incorrect volume will lead to an underestimation or overestimation of the copy number, which can in turn have effects on clinical and diagnostic decisions, and potentially lead to incorrect labelling of a product or misguided treatment or diagnosis. It should also be taken into account that when transferring the method from EvaGreen™ chemistry to Taq-Man™ chemistry, the supermix used needs to be changed, and that the droplet volume needs to be considered when assessing the copy number. Thus, to accurately determine copy number, the droplet volume should be measured for all possible droplet generators and supermixes used in a laboratory. However, this is time consuming and impractical for the majority of laboratories. Therefore, laboratories where the droplet volume cannot be assessed, and particularly for laboratories with applications in clinical, diagnostic and food-related fields, and in research fields, the measurement uncertainty of the data should be expanded to encompass any influences of droplet volume variability.

Acknowledgements The authors would like to thank Phillip Corbisier and the Biomolecular Detection and Quantification group of the European Commission Joint Research Centre, for help with the automatic image analysis, and Mojca Milavec from the Department of Biotechnology and Systems Biology at NIB, for help with data interpretation and measurement uncertainty assignment. We also thank Christopher Berrie for language revision. This study was conducted under Bio-SITrace, Traceability for biologically relevant molecules and entities (Number JRP SIB54), a European Metrology Research Programme (EMRP) project. The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union. Financial support was also provided by the Slovenian Research Agency (Contract No. P4-0165, 1000-15-0105). The QX100 system used in this study was financed by the Metrology Institute of the Republic of Slovenia, with financial support from the European Regional Development Fund.

Compliance with ethical standards This study did not involve research on human participants or animals.

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Pavšič J, Devonshire AS, Parkes H, Schimmel H, Foy CA, Karczmarczyk M, et al. Standardization of nucleic acid tests for clinical measurements of bacteria and viruses. J Clin Microbiol. 2015;53:2008–14. https://doi.org/10.1128/JCM.02136-14.
2. BioSitrace: Traceability for biologically relavant molecules and entities. http://biositrace.lgcgroup.com/. Accessed 24 April, 2017.
3. Vogelstein B, Kinzler KW. Digital PCR. Proc Natl Acad Sci U S A. 1999;96:9236–41. https://doi.org/10.1073/pnas.96.16.9236.
4. Sykes PJ, Neoh SH, Brisco MJ, Hughes E, Condon J, Morley AA. Quantitation of targets for PCR by use of limiting dilution. BioTechniques. 1992;13:444–9.
5. Huggett JF, Cowen S, Foy CA. Considerations for digital PCR as an accurate molecular diagnostic tool. Clin Chem. 2015;61:79–88. https://doi.org/10.1373/clinchem.2014.221366.
6. Pavšič J, Žel J, Milavec M. Digital PCR for direct quantification of viruses without DNA extraction. Anal Bioanal Chem. 2016;408:67–75. https://doi.org/10.1007/s00216-015-9109-0.
7. Dreo T, Pirc M, Ramšak Ž, Pavšič J, Milavec M, Žel J, et al. Optimising droplet digital PCR analysis approaches for detection and quantification of bacteria: a case study of fire blight and potato brown rot. Anal Bioanal Chem. 2014;406:6513–28. https://doi.org/10.1007/s00216-014-8084-1.
8. Morisset D, Štebih D, Milavec M, Gruden K, Žel J. Quantitative analysis of food and feed samples with droplet digital PCR. PLoS One. 2013;8:e62583. https://doi.org/10.1371/journal.pone.0062583.
9. Block A, Debode F, Grohmann L, Hulin J, Taverniers I, Kluga L, et al. The GMOseek matrix: a decision support tool for optimizing the detection of genetically modified plants. BMC Bioinformatics. 2013;14:256. https://doi.org/10.1186/1471-2105-14-256.
10. Dobnik D, Spilsberg B, Bogožalec Košir A, Holst-Jensen A, Žel J. Multiplex quantification of 12 European Union authorized genetically modified maize lines with droplet digital polymerase chain reaction. Anal Chem. 2015;87:8218–26. https://doi.org/10.1021/acs.analchem.5b01208.
11. Dobnik D, Štebih D, Blejec A, Morisset D, Žel J. Multiplex quantification of four DNA targets in one reaction with bio-rad droplet digital PCR system for GMO detection. Sci Rep. 2016;6:35451. https://doi.org/10.1038/srep35451.
12. Hudecova I. Digital PCR analysis of circulating nucleic acids. Clin Biochem. 2015;48:948–56. https://doi.org/10.1016/j.clinbiochem.2015.03.015.
13. Kline MC, Romsos EL, Duewer DL. Evaluating digital PCR for the quantification of human genomic DNA: accessible amplifiable

targets. Anal Chem. 2016;88:2132–9. https://doi.org/10.1021/acs.analchem.5b03692.

14. Sedlak RH, Liang S, Niyonzima N, De Silva Feelixge HS, Roychoudhury P, Greninger AL, et al. Digital detection of endonuclease mediated gene disruption in the HIV provirus. Sci Rep. 2016;6:20064. https://doi.org/10.1038/srep20064.
15. Whale AS, Huggett JF, Cowen S, Speirs V, Shaw J, Ellison S, et al. Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation. Nucleic Acids Res. 2012;40:1–9. https://doi.org/10.1093/nar/gks203.
16. Dube S, Qin J, Ramakrishnan R. Mathematical analysis of copy number variation in a DNA sample using digital PCR on a nanofluidic device. PLoS One. 2008;3:e2876. https://doi.org/10.1371/journal.pone.0002876.
17. Pinheiro LB, Coleman VA, Hindson CM, Herrmann J, Hindson BJ, Bhat S, et al. Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. Anal Chem. 2012;84:1003–11. https://doi.org/10.1021/ac202578x.
18. Corbisier P, Pinheiro L, Mazoua S, Kortekaas A, Chung PYJ, Gerganova T, et al. DNA copy number concentration measured by digital and droplet digital quantitative PCR using certified reference materials. Anal Bioanal Chem. 2015;407:1831–40. https://doi.org/10.1007/s00216-015-8458-z.
19. Dagata JA, Farkas N, Kramer JA. Method for measuring the volume of nominally 100 μm diameter spherical water-in-oil emulsion droplets. NIST Spec Publ. 2016; https://doi.org/10.6028/NIST.SP.260-184.
20. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open source platform for biological image analysis. Nat Methods. 2012;9:676–82. https://doi.org/10.1038/nmeth.2019.Fiji.
21. Schindelin J, Rueden CT, Hiner MC, Eliceiri KW. The ImageJ ecosystem: an open platform for biomedical image analysis. Mol Reprod Dev. 2015;82:518–29. https://doi.org/10.1002/mrd.22489.
22. Schneider CA, Rasband WS, Eliceiri KW. NIH image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9:671–5.
23. Meli F, Klein T, Buhr E, Frase CG, Gleber G, Krumrey M, et al. Traceable size determination of nanoparticles, a comparison among European metrology institutes. Meas Sci Technol. 2012;23:125005. https://doi.org/10.1088/0957-0233/23/12/125005.
24. Rstudio Team. RStudio: integrated development for R. Boston: RStudio, Inc.; 2016.
25. Joint Committee for Guides in Metrology. Evaluation of measurement data - guide to the expression of uncertainty in measurement. 2008. http://www.bipm.org/en/publications/guides/gum.html. Accessed 26 Feb, 2017
26. Corbisier P, Zobell O, Trapmann S, Auclair G, Emons H. Training manual on GMO quantification: proper calibration and estimation of measurement uncertainty. Luxembourg: Publications Office of the European Union. 2014. https://doi.org/10.2787/10085.
27. Kline MC, Duewer DL. Evaluating droplet digital polymerase chain reaction for the quantification of human genomic DNA: lifting the traceability fog. Anal Chem. 2017;89:4648–54. https://doi.org/10.1021/acs.analchem.7b00240.
28. Marchezi U, Mazzara M, Broll H, De Giacomo M, Grohmann L, Herau V, Holst-Jensen A, Hougs L, Hübert P, Laurensse E, Morisset D, Papazova N, Pietsch K, Pla M, Sandberg M, Savini C, Waiblinger H-U, Woll K. Definition of minimum performance requirements for analytical methods of GMO testing. European Union Reference Laboratory for GM Food and Feed and European Network of GMO Laboratories 2015.
29. Decathlon Project. Minimum performance parameters. 2015;1–27. http://www.decathlon-project.eu/content/d-61-minimum-performance-parameters. Accessed 24 April, 2017.

2.4 ALF: A Strategy for Identification of Unauthorized GMO in Complex Mixtures by a GW-NGS Method and Dedicated Bioinformatics Analysis

Alexandra Bogožalec Košir, Alfred J. Arulandhu, Marleen M. Voorhuijzen, Hongmei Xiao, Rico Hagelaar, Martijn Staats, Adalberto Costessi, Jana Žel, Esther J. Kok, Jeroen P. van Dijk

Scientific Reports, 2017, article number: 14155. DOI: 10.1038/s41598-017-14469-8

In this paper, we describe a novel approach for identification of UGMOs called ALF. ALF, amplification of linearly enriched fragments, is a strategy where a targeted GW coupled to NGS provides us with a plethora of sequences containing clues of a presence of UGMOs in a complex sample. In order to decipher these clues a (semi)automated data analysis pipeline was created using the Galaxy platform. In order to prove the principle of the method we prepared a sample containing DNA of four GMOs. As the full insertion sequence is not known for one of the GMOs, this GMO acted as a mock UGMO. Submitting the sequences gained from NGS to the (semi)automated pipeline, we found sequences related to this mock UGMO, thus proving the principle of the method.

The PhD candidate designed the (semi)automated data analysis pipeline and helped with the preparation of figures and tables. She drafted the bioinformatics part of the manuscript and the supplement.

www.nature.com/scientificreports



OPEN

ALF: a strategy for identification of unauthorized GMOs in complex mixtures by a GW-NGS method and dedicated bioinformatics analysis

Received: 24 March 2017
Accepted: 9 October 2017
Published online: 26 October 2017

Alexandra Bogožalec Košir[1,2], Alfred J. Arulandhu[3,4], Marleen M. Voorhuijzen[3], Hongmei Xiao[5], Rico Hagelaar[3], Martijn Staats[3], Adalberto Costessi[6], Jana Žel[1], Esther J. Kok[3] & Jeroen P. van Dijk[3]

The majority of feed products in industrialised countries contains materials derived from genetically modified organisms (GMOs). In parallel, the number of reports of unauthorised GMOs (UGMOs) is gradually increasing. There is a lack of specific detection methods for UGMOs, due to the absence of detailed sequence information and reference materials. In this research, an adapted genome walking approach was developed, called ALF: Amplification of Linearly-enriched Fragments. Coupling of ALF to NGS aims for simultaneous detection and identification of all GMOs, including UGMOs, in one sample, in a single analysis. The ALF approach was assessed on a mixture made of DNA extracts from four reference materials, in an uneven distribution, mimicking a real life situation. The complete insert and genomic flanking regions were known for three of the included GMO events, while for MON15985 only partial sequence information was available. Combined with a known organisation of elements, this GMO served as a model for a UGMO. We successfully identified sequences matching with this organisation of elements serving as proof of principle for ALF as new UGMO detection strategy. Additionally, this study provides a first outline of an automated, web-based analysis pipeline for identification of UGMOs containing known GM elements.

Nowadays, the vast majority of feed products in industrialised countries contain materials that are derived from genetically modified organisms (GMOs). For food products the situation is still very different, primarily due to the lack of public acceptance in some countries and amongst specific groups of consumers, but here also a slow trend can be observed of increased use of products derived from GMOs. From the Food and Agriculture Organization of the United Nations (FAO) report of 2014[1], it can furthermore be seen that the number of incidents with unauthorised GMOs (UGMOs) is gradually increasing. The reported incidents related primarily to identified UGMOs, that had received market approval in other countries. However, in a growing number of cases the incidents related to unknown UGMOs where the mere combination of the crop at hand and the detected GMO elements were deemed sufficient to take action. It should be stressed that so far no genetically modified (GM) crops have been identified to have adverse effects on humans, animals or the environment[2,3]. Nevertheless, in the light of the rapidly expanding diversity of GMOs in experimental settings, it seems prudent to have the methodologies in place to detect and identify UGMOs, including the as yet unknown ones. There is a general lack of specific detection methods for UGMOs, usually due to the absence of detailed sequence information and reference materials.

In recent years, a number of strategies have been developed that focus on the detection of UGMOs. The strategy generally used is based on the screening of samples for a range of GMO elements. Subsequently, the

[1]Department of Biotechnology and Systems Biology, National Institute of Biology, Večna pot 111, SI-1000, Ljubljana, Slovenia. [2]Jožef Stefan International Postgraduate School, Jamova 39, SI-1000, Ljubljana, Slovenia. [3]RIKILT Wageningen UR, P.O. Box 230, 6700 AE, Wageningen, The Netherlands. [4]Food Quality and Design Group, Wageningen University and Research, P.O. Box 8129, 6700 EV, Wageningen, The Netherlands. [5]College of Food Science and Technology, Nanjing Agricultural University, Jiangsu, 210095, P. R. China. [6]BaseClear, Einsteinweg 5, 2333 CC, Leiden, The Netherlands. Alexandra Bogožalec Košir and Alfred J. Arulandhu contributed equally to this work. Correspondence and requests for materials should be addressed to J.P.v.D. (email: jeroen.vandijk@wur.nl)

presence of the observed elements is compared with the confirmed presence of authorised GMOs or known UGMOs in the same sample. A mismatch between observed elements and observed GMOs indicates the potential presence of a UGMO. Clearly, only GMOs for which adequate methods for identification are available[4–15] can be taken into account here. Additional analyses are necessary to determine whether the identified GMO element is indeed linked to an unknown, unauthorised GMO, or rather to, for instance, the native organism of the element. Therefore, in these cases, additional experiments will be required, that are currently usually based on variants of gene walking (GW) strategies where the unexplained elements can be used as a starting point for GW to obtain adjacent sequence information. This sequence information could lead to the identification of a specific UGMO, especially when the sequence information stretches into the flanking genomic region of the GMO insert. This information can ultimately, if deemed necessary, be used to develop a specific method for the identified UGMO.

In the last two decades several GW approaches have been developed and modified for application in the GMO field. Examples are Long template-Rapid Amplification of Genomic DNA Ends (LT-RADE)[16,17], SiteFinding-PCR[18,19], APAgene GOLD Genome Walking Kit[20–22], A-T linker adapter PCR[23], Randomly broken fragment PCR (RBF-PCR)[24], Locus-finding PCR (LF PCR)[25] and Loop-linker PCR[26] recently reviewed in Arulandhu *et al.*[27]. Most of the approaches have been initially evaluated in a pure GMO sample. The APAgene GOLD Genome Walking Kit[20–22] was also tested in low GM percentages in combination with fragment cloning and Sanger sequencing[20,22]. Only very recently this method was also succesfully applied in combination with NGS and more complex GMO mixtures[21].

Any enrichment approach applied to a mixed sample may yield a mixture of sequences, because even a single common element may have several adjacent sequences, from different GMOs in the mixture. Next Generation Sequencing (NGS) is ideally suited for sequencing all amplified fragments in a mixture. In the development of a new GW approach we chose the PacBio RSII as NGS platform, as long reads of single molecules minimize the chance of artefacts, frequently referred to as chimeric sequences, due to incorrect assembly of short sequence reads[19]. Target molecules can be sequenced multiple times or 'passes', as hairpin adaptors are ligated to both ends, creating a circular DNA molecule, serving as a polymerase template without an end. The initial output is a polymerase read with multiple copies of the target sequence interspaced by adaptor sequences. When such a subread is present at least four times, a consensus sequence is made after removing the adaptor sequences. This is the circular consensus sequence (CCS) output of the PacBio RS II[28] (Supplementary Fig. S1). With 80–85% accuracy[29], subreads have a relatively high error rate[29–32]. As the errors are random, with the error pattern of 10% insertions and 5% deletions[31], the accuracy of the CCS read is higher than that of the subread, and it increases with the number of subreads. SiteFinding-PCR[18,19], and APAgene GOLD Genome Walking Kit[21], are the only enrichment approaches that have been published in combination with NGS analysis to detect and identify GMOs/UGMOs. However, given the random start and two nested PCRs of SiteFinding-PCR, and two semi nested PCRs and the use of DRT primers in APAgene GOLD Genome Walking Kit, these approaches might be sensitive to contamination. Therefore, we adapted the LT-RADE method with a biotinylation primer mediated clean-up prior to PCR and only one round of PCR. Coupling this approach to NGS should allow for simultaneous detection and identification of all GMOs in a sample, including UGMOs, in a single analysis.

In this research, an adapted GW approach was developed, hence referred to as Amplification of Linearly-enriched Fragments (ALF), encompassing the advantages of the available methods, particularly the rapid amplification of cDNA ends (RACE)[33] approach and the use of biotinylated primers. The efficacy of the ALF approach was assessed on a complex mixture consisting of four GMOs: maize MON810, MON89034, MON88017 and cotton MON15985, using the GMO elements p35S promoter and tNOS terminator as starting-points for the elongation. After amplification of the elongated fragments, all obtained fragments were sequenced using the PacBio RSII platform. The complete insert and genomic flanking regions were known for the first three GM maize events, while for the MON15985 only partial sequence information was available; combined with a known organisation of elements, this GMO served as a model for a UGMO in this set-up. In the present study, the ALF approach is presented and applied on the mixture of all four GMOs. The first outline of an automated, web-based analysis pipeline for identification of UGMOs containing known GM elements was set, and the results of the new strategy are evaluated in the light of the necessity to have adequate methodology in place to identify GMOs for which limited information is available.

Results

Amplification of Linearly-enriched Fragments (ALF). A protocol was developed for the identification of unknown GMO-related sequences starting from known GMO elements, called ALF: Amplification of Linearly-enriched Fragments (Fig. 1). The success of the ALF protocol was determined both in terms of quantity and length of the molecules of interest. In the reference materials used, upstream and downstream GMO elements were known. By performing qPCR analysis prior to, and at different steps of the procedure, relative increase in target fragments was estimated. Targets for the qPCR analysis were chosen at various distances from the start of the LE (Fig. 2). To increase the reliability of the quantification outcomes, Joint Research Centre (JRC)-validated qPCR methods were used in combination with certified reference material for >99.05% MON88107. qPCRs were performed before the procedure in the starting material (SM), after LE, and after snPCR. Apart from the qPCRs targeting the GM elements of interest, also the abundance of the maize endogenous high mobility group (*hmg*) gene was evaluated with qPCR, to monitor the removal of genomic DNA (Fig. 3). After the LE step, a slight increase was observed for targets of interest, more prominently for targets closer to the starting point of LE. As expected, the genomic DNA amount was very similar before and after LE. After snPCR, a large increase in targets of interest was observed, again more prominent for target sequences close to the LE start, indicating a size-dependent enrichment. Loss of gDNA beyond detection was observed for all targets after the snPCR step (Fig. 3).

The relative amount of the different targets was expressed as the relative Cq value and calculated by subtracting the Cq after LE or snPCR from the Cq of the starting material. The largest increase in relative Cq was observed

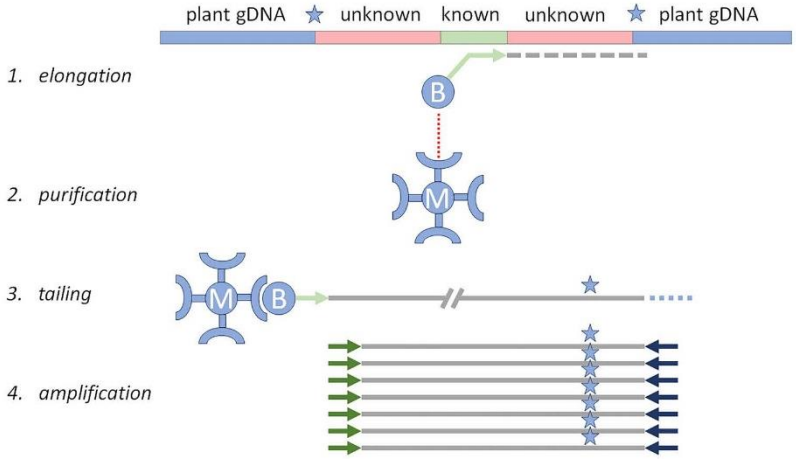


Figure 1. Schematic overview of the ALF procedure. The procedure yields dsDNA molecules after initial primer elongation, of various length, the longest of which will pass a construct-genome boundary (star), leading to GMO identification.

for GW starting from tNOS in the upstream direction, for the element closest to the LE start, Cp4-epsps (Fig. 3, panel B). The fold change for the different targets upstream of tNOS was estimated using a ΔCq approach, assuming a qPCR efficiency of 2. The Cq for the Cp4-epsps element, 303 nt upstream of the tNOS LE primer, was 30.3 in the starting material. During development of the method, we found it to be necessary to dilute the snPCR material 100-fold prior to qPCR evaluation, in order to obtain Cq values higher than five. Even in the diluted snPCR material the Cq value for the Cp4-epsps element was as low as 9.6. This was converted to the theoretical value of 2.96, by subtracting the 6.64 cycles from this Cq value, corresponding to a 100-fold dilution. The ALF related ΔCq for the Cp4-epsps element was therefore calculated to be 27.3 cycles, indicating an estimated increase of 165 million-fold of fragments at least 303 nt long. A fraction of these molecules was at least 1609 nt long, evidenced by the increase in relative Cq value for the Ctp2-cp4epsps element, located 1609 nt upstream of the tNOS LE primer. This fraction was estimated to be enriched 30 thousand fold. Likewise, a subfraction of these fragments was at least 1922 nt long, based on relative rAct1 Cq increase. This fraction was estimated to be enriched 3.5 thousand fold.

PacBio sequencing of a GMO mixture after ALF. A DNA-based mixture was made using DNA extracts from four reference materials, in an uneven distribution, mimicking a real life situation, where one or several GMOs may be present in a much lower concentration than others in the same mixed sample. DNA of MON810 from a co-existence field trial[34] was used as the most abundant GMO and was present in 97%. Of the other three GMO DNA isolations, 1% relative weight was added for each (Table 1). Two maize lines were used, MON89034 and MON88017, and one cotton line, MON15985. This mixture was subjected to the ALF protocol in four separate reactions for the different elements and directions: the p35S promoter and the tNOS terminator, both in upstream and downstream direction. The reactions were pooled and sequenced on the PacBio RSII platform. The results of sequencing were 411 CCS reads (length distribution is shown in Supplementary Fig. S2).

Sequence analysis: database construction. To analyse the CCS reads gained after PacBio RSII sequencing three databases were constructed: an event, a Constructs-And-Flanks (CAF) and an element database. The event database (Supplementary Data S1) consisted of the most likely amplicon sequences of the event specific qPCR methods used for quantification of events (MON810, MON88017, MON89034 and MON15985) in the sample. Sequences of primers and probes, together with the number of unknown nucleotides between them, were taken from the EU Database of Reference Methods for GMO Analysis (GMOMETHOD) (http://gmo-crl.jrc.ec.europa.eu/gmomethods/) and from the method validation reports[35–38]. These sequences were queried against the NCBI patent database (Table 2). After inspection, the top hit for each of them was added to the event database.

The CAF database consisted of the available construct sequences and flanking plant genomic regions of individual events in the sample (Supplementary Data S2). In case of MON15985 only partial 3′ and 5′ insert sequences with corresponding flanking genomic sequences were available. For MON810 two reference sequences were available: sequence JQ406879, covering the 5′ flank and the p35S promoter, and AY326434, covering the insert from the p35S promoter onwards to the 3′ flank. These sequences showed a 75-nucleotide overlap. In the initial workflow both sequences were in the CAF database, but the merged MON810 sequence of JQ406879 and AY326434, named RIKILT20151130, was used in the final workflow (Supplementary Data S2).

The third, element database, consisted of element sequences present in the experimental mixture. All element sequences from individual lines were gathered. For MON810, MON89034 and MON88017 complete reference sequences were already known and annotated. These reference sequences were divided in elements based on

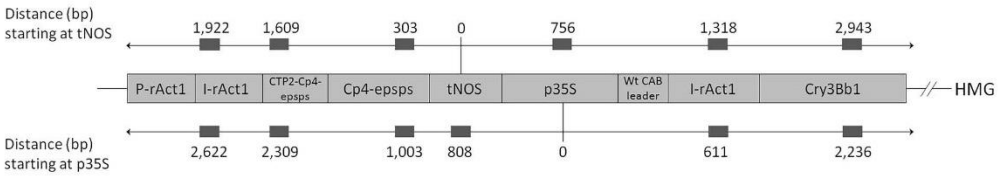


Figure 2. Distances of several qPCR targets from the enrichment starting points in the MON88107 GMO, used for evaluation of the ALF protocol. The light grey boxes indicate the different elements in the GMO. The upper and lower line indicate the enrichment for tNOS (upper) and p35S (lower), 0 indicates the starting points of enrichment and arrows indicate the direction of enrichment (upstream or downstream). The dark grey boxes indicate qPCR targets used for evaluation of the ALF protocol and distance of the qPCR targets from the starting points.

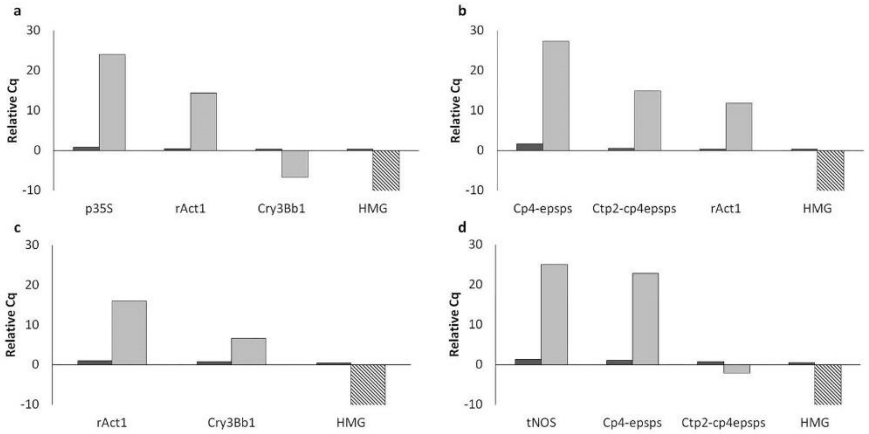


Figure 3. Length-dependent increase of specific targets and loss of genomic background shown with qPCR. In all cases the amplicon closest to the element targeted for linear enrichment showed the highest increase in signal, expressed as the relative Cq value and calculated by subtracting the Cq after LE (dark bars) or snPCR (light bars) from the Cq of the starting material. This means the relative Cq value for each starting point is zero. Panel a: Enrichment for tNOS downstream. (**b**) Enrichment for tNOS upstream. (**c**) Enrichment for p35S downstream. (**d**) Enrichment for p35S upstream. In all panels, the dashed bars for HMG indicate a reduction beyond detection.

their annotation, for which a General Feature Format (GFF) file was constructed (Supplementary Data S3). For MON15985, for which only the order of elements was known, generic sequences for these elements were taken from the NCBI patent database. Initially, this database contained all element sequences (Supplementary Data S4). To reduce redundancy, the final version contained only the longest sequence of elements with several entries in the first version (Supplementary Data S5). All designed databases were transformed into BLAST+ databases[39] and imported into the open, web-based analysis platform Galaxy[40].

Building the workflow for GMO sequence identification. The workflow for GMO sequence identification in the samples was constructed in Galaxy to answer the following questions: (1) is there any potential evidence for UGMOs, (2) can known GMOs be identified and (3) can a list of potential GMOs be prepared. Considering these questions, a workflow with six main steps was constructed (Fig. 4 and Supplementary Fig. S3).

Step 1 consisted of a trimming and length selection; removing artificial sequences from the CCS reads and subsequently discarding short reads. The poly-dC tail and AAP adaptor sequences were removed and a length filtering was applied at >50 nt. All 411 CCS reads were used as the workflow input, a total of 401 reads proceeded to step 2.

Step 2 was selection of reads containing an enrichment primer, i.e. the nested tNOS and p35S LE primers, to make sure only specific sequences were kept. A total of 383 sequences passed the LE primer filtering; 368 were enriched for the p35S and 15 for the tNOS.

Step 3 was a clustering, using UPARSE[41], of CCS reads to further reduce the number of sequences. A three-step clustering approach was used. In this approach, sequences were dereplicated, sorted by abundance and clustered.

GMO	Content in mixture (%)	Estimated copy number	Supplier	Code	Description	Element order*	Donor organism
MON810	97	~40.000	Field trial[34]	In-house	50% MON810, ground corn	P-e35S	Cauliflower mosaic virus
						I-hsp70	*Zea mays*
						CS-cry1Ab	*Bacillus thuringiensis*
MON89034	1	~400	AOCS	0906-E	>99.42% MON89034, ground corn	V-LB	*Agrobacterium tumefaciens*
						P-e35S	Cauliflower mosaic virus
						L-cab	*Triticum aestivum*
						I-1_act1	*Oryza sativa*
						CS-cry1A_105	Synthetic
						T-hsp17_3	*Triticum aestivum*
						P-FMV	Figwort mosaic virus
						I-hsp70	*Zea mays*
						I-1_rbcS	*Zea mays*
						CS-cry2Ab2	*Bacillus thuringiensis ssp. Kurstaki*
						T-nos	*Agrobacterium tumefaciens*
						V-RB	*Agrobacterium tumefaciens*
MON88017	1	~400	AOCS	0406-D	>99.05% MON88017, ground corn	P-act1	*Oryza sativa*
						I-1_act1	*Oryza sativa*
						TP-ctp	*Arabidopsis thaliana*
						CS-CP4epsps	*Agrobacterium tumefaciens ssp. CP4*
						T-nos	*Agrobacterium tumefaciens*
						P-e35S	Cauliflower mosaic virus
						L-cab	*Triticum aestivum*
						I-1_act1	*Oryza sativa*
						CS-cry3Bb1	*Bacillus thuringiensis ssp. Kumamotoensis*
						T-hsp17_3	*Triticum aestivum*
MON15985	1	~400	AOCS	0804-D	>98.45% Bollgard II cotton, ground cotton seed	P-e35S	Cauliflower mosaic virus
						L-hsp70	Petunia hybrida
						TP-ctp	*Arabidopsis thaliana*
						CS-cry2Ab2	*Bacillus thuringiensis ssp. Kurstaki*
						T-nos	*Agrobacterium tumefaciens*
						P-e35S	Cauliflower mosaic virus
						CS-cry1Ac	*Bacillus thuringiensis ssp. Kurstaki*
						T-7Salpha	*Glycine max*
						P-35S	*Cauliflower mosaic virus*
						CS-nptII	*Escherichia coli*
						T-nos	*Agrobacterium tumefaciens*
						P-e35S	Cauliflower mosaic virus
						CS-uidA	*Escherichia coli*
						T-nos	*Agrobacterium tumefaciens*
						CS-aadA	*Escherichia coli*

Table 1. Details of reference materials[49]. *Initial capital characters: P = promoter, I = intron, CS = coding sequence, V = vector, LB = left border, L = leader, T = terminator, RB = right border, TP = transit peptide.

Due to the nature of PacBio sequencing a CCS read either starts or ends with the enrichment target element, in this case the tNOS or p35S. An additional step of reverse complementing was applied to all the reads containing the enrichment target element on the 3′ end before dereplication, since the clustering algorithm aligned sequences only in plus/plus manner. With a minimum cluster identity, also termed cluster radius, of 0.97, 383 CCS reads were clustered in 16 clusters. For each cluster, the longest CCS read was used as the representative sequence, and termed cluster representative CCS read (crCCS read).

Step 4 was a Megablast of the crCCS read against the event database. Seven of the 16 crCCS reads showed a match to a sequence in the database with an identity of 95% or higher and a minimum coverage of 97%. The main characteristic of these reads was that they included a *confirmed event*, and were not analysed further. The nine remaining crCCS reads were further processed in step 5.

Step 5 was a Megablast of the remaining crCCS reads against the CAF database. Two out of nine crCCS reads aligned completely to a CAF database sequence, i.e. full query coverage, and were annotated using the intersect interval tool (part of BEDtools[42]). The main characteristic of these annotated crCCS reads was the establishment that they contained *only known information*.

Step 6 was a Megablast of the remaining seven crCCS reads against the element database, aiming at identification of reads that contain both known and unknown sequence, as is expected for UGMOs. The crCCS reads were

GM event	Event sequences	Best hit accession number
MON810	TCGAAGGACGAAGGACTCTAACGT***TT***AACATCCTTTGCCATTGCCCAGC ***TATCTGTCACTTTATTGTG***AAGATAGTGGAAAAGGAAGGTGGC*	AR490568
MON89034	TTCTCCATATTGACCATCATACTCATT***GC***ATCCCCGGAAATTATGTT***TT***TTTAAAAACCACGGTATTATAGATACCG	FV532179
MON88017	GAGCAGGACCTGCAGAAGCT***AGCTTGATGGGGATCAGATTGTCGTT*** TCCCGCCTTCAGTTTAAACAGAGTCGGGT***T***TGGATGGTCAACTCCGGCA	DJ058152/DJ058151
MON15985	GTTACTAGATCGGGGATATCC***CCGGGGCGG***CCGCTCTAGAACTAGTGGATCTGCACTGAAA***T***CCCATCCATTTAGCAACCTT	EA135634

Table 2. Event database sequences and corresponding NCBI accession numbers. *Nucleotides in italics denote the string of unknown nucleotides between primer and probe sequences in the search template.

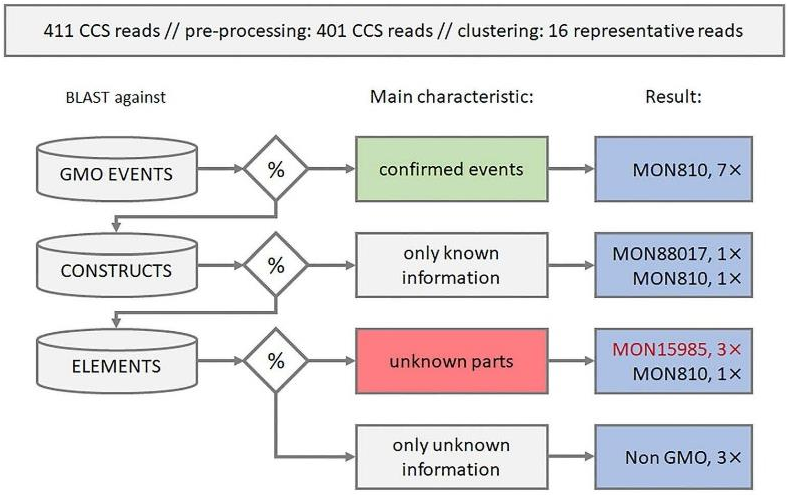


Figure 4. Schematic overview of the data analysis workflow. CCS reads are first processed to reduce noise and redundancy. Representative reads are then grouped in relevant bins based on homology with sequences in different databases, using blast. In the results, reads likely to be related to MON15985 are marked in red as this authorised GMO with incomplete sequence information served as a model UGMO in this study.

divided into two bins, also based on their main characteristic, either containing *unknown parts*, or containing *only unknown information*. The reads containing unknown parts also contained homologies to known elements, by the definition of this workflow. The Blast output of these reads against the element database was sorted in a way that the top hit was the one with the longest sequence alignment. The alignments with different elements were ordered according to their position in the CCS read, from 5′ to 3′. All four output bins, *confirmed events*, *only known information*, *unknown parts*, and *only unknown information* were imported in an excel template, constructed to give a user-friendly output (Supplementary spreadsheet 1).

Connecting sequences with the experimental set-up. Relating to the three main question (1) is there any potential evidence for UGMOs, (2) can known GMOs be identified and (3) can a list of potential GMOs be prepared, the crCCS reads in the *unknown parts* bin were the most informative for the first question (Fig. 4 and Supplementary Fig. S3). The four crCCS reads in that bin showed three different element orders. One was *tNOS – nnn - NPT II – nnn - p35S*, present in two crCCS reads in both orientations, where *nnn* denotes an unknown sequence. This element order was consistent with that of the MON15985 parental line MON531. The second order was *p35S – nnn – tNOS – nnn – UidA*, consistent with the order in the retransformation construct of MON15985 (Fig. 5 and Supplementary Data S6). Sequences with these two element orders were previously not linked to any MON15985 designated sequence in a public database. The third element order was *p35S – nnn - hsp70 - cry1ab*, corresponding to MON810. The gap between the p35S and HSP70 was 16 nucleotides, and was the result of an incomplete alignment with the p35S element reference sequence.

In addition to the four crCCS reads in the *unknown parts* bin, three crCCS reads (*120091*, *43434*, and 106520) ended up in the *only unknown information* bin (Fig. 4). To find the origin of these reads an NCBI Blastn, against nucleotide database, and a Megablast and a Blastn against NCBI patent database were run. For read 120091 a 99% identity and a 98% CCS read coverage against *maize genotype CMS-S mitochondrion* (DQ490951) was observed. For read 106520 top hits were *nucleotide sequences and polypeptides encoded thereby useful for modifying plant*

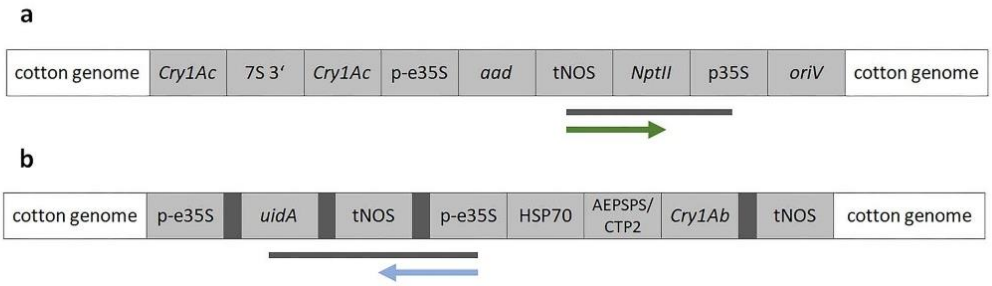


Figure 5. Two partial sequences of the MON531 and MON15985 inserts. Panel a shows the alignment of crCCS reads 133951 and 45207 to the insert and flanking region of MON531, with the position of enrichment primer (downstream NOS terminator). Panel b shows the alignment of and crCCS read 156962 to the insert and flanking region of MON15985, and the enrichment primer (upstream 35S promotor). Both sequences cover a previously unknown insert sequence.

crCCS read	Identified GMO	% of identical match	Alignment length	Length of db sequence	Alignment length in %
33879	MON810	100.00	92	92	100
40876	MON810	95.65	92	92	100
78816	MON810	98.92	93	92	101
119658	MON810	96.81	94	92	102
128024	MON810	100.00	92	92	100
132414	MON810	96.74	92	92	100
158734*	MON810	100.00	92	92	100
158734*	MON810	100.00	92	92	100

Table 3. Results of NCBI BLAST– against the event database. *CCS read 158734 showed two perfect alignments with the database sequence.

characteristics (GP689587) and *long intergenic noncoding RNAs in maize* (JC761506). A total of 76% of the read was covered. For read 43434 the top hits were described as a *composition and method for therapy and diagnosis of ovarian cancer* (DL059073), with 90% identity and query coverage of 60% (Supplementary Table S4).

The *confirmed events* bin, as the result of step 4, provided the answer to question (2) on identification of known GMOs. In order for a crCCS read to be acknowledged as containing a confirmed event, a 97% minimal coverage of the event sequence in the database was required, with at least 95% identity (Table 3). All seven reads that fulfilled the criteria of step 4 originated from MON810 (Fig. 4). In one of these, the read aligned to the same database event sequence twice, both with a 100% identity match. After closer investigation, the CCS read turned out to contain a complete repetition of one sequence; an AAP primer with a poly G tail prior to a 5′ MON810 flanking region followed by a p35S promoter sequence. The sequence was kept in the *identified GMOs*.

In step 5, a bin was made for the fully annotated CCS reads, containing *only known information*. In this step the answers to the question (3) on the list of potential GMOs were found. An indication of MON88017 and MON810 was observed (Fig. 4). Two crCCS reads completely aligned with a database sequence, the first CCS read had the element order of *p35S - hsp70 - cry1ab*, present only in MON88017 and the second *cp4epsps – tNOS – 35S* present only in MON810, considering the GMOs used in this experiment.

The merged MON810 reference sequence. The merged MON810 read RIKILT20151130 (deposited at The European GMO reference database - Euginious http://www.euginius.eu/euginius/pages/sequence_string.jsf?sequence=4165286415439512999) that comprised of sequences JQ406879 and AY326434 was compared to the CCS reads. Bowtie2[43] mapping with default settings of the CCS reads was performed using very sensitive local alignment (Supplementary Fig. S4). The CCS reads were first filtered by quality, only those that had a quality score of 33 or more over at least 90% of the sequence were accepted. Out of the initial 411 CCS reads, 165 filtered sequences aligned to this reference. A total of 133 CCS reads covered the overlap between sequences JQ406879 and AY326434, confirming the trueness of the new, merged MON810 reference sequence.

Discussion

The main goal of this study was to develop a protocol for identification of UGMOs in a complex mixture. The new approach was successful in its aim, and a proof of principle was found by identifying previously unknown sequences corresponding to the element order in a GMO.

The adapted GW approach, ALF, aimed to enrich the adjacent regions of the target prior to sequencing. Enrichment decreased the required sequencing depth and cost, as well as the data analysis time and cost. A new

combination of GW steps was conceived and tested. Two common approaches were combined: gDNA reduction and the RACE/LT-RADE principle of tailing and semi-universal PCR. Biotinylation of the enrichment primer was used for background reduction. Since genomic DNA background was already reduced beyond detection, we decided against the use of a nested PCR approach. The semi-nested PCR in this paper is actually a first round of PCR. It is called semi-nested because one of the primers is nested, i.e. slightly downstream of the primer used in the LE step. PCR in general, and nested PCR in particular, is prone to contamination. For most diagnostic laboratories, the use of nested PCR is something to be avoided, especially in case of many repetitive tests. A contamination of a GMO testing laboratory with especially common GM elements would severely compromise an efficient GMO screening using the matrix approach. Target-specific LE primes as well as the nested primers of the snPCR were designed on the basis of the available qPCR primer-probe sequences for the targeted GMO elements, in our case p35S promoter and tNOS terminator. Primer and probe sequences of an unexplained GMO element test might be the only basis for finding related sequences, in case of looking for unknown GMOs as the result of a GMO matrix approach outcome. Therefore, the primer and probe sequences are a logical starting point for enrichment.

The high quality of the CCS reads warranted the omission of an extra quality check at the beginning of the analysis pipeline. This is important, as the library preparation requires a high amount of DNA, with low CCS read counts resulting from a lower DNA input. In this experiment, 411 CCS reads were generated despite a high amplification of relevant molecules in the ALF protocol. A clustering step was added to the workflow to further reduce the number of redundant CCS reads. To make sure no information is lost during clustering, a workflow omitting this step was tested, and it showed no extra information. The largest decrease in redundancy was observed in the *confirmed event* bin, where the decrease of sequences was more than 97% (295 *vs* 7 sequences). However, the main difference influencing the time of further analysis was observed in the *only known information* and *unknown parts* bins, where the number of sequences was decreased more than 92%; 28 *vs* 2 sequences after clustering in the *only known information* bin and 53 *vs* 3 sequences in the *unknown parts* bin. To reduce the time for clustering, a length filtering set at 50 nucleotides was applied in the very first step of the workflow. Reads of this length are not informative, as the shortest transgenic element in our element database is 60 base pairs.

The experimental sample was composed of known GMOs in different quantities: MON810 maize– 97%, MON89034 maize– 1%, MON88017 maize– 1%, and MON15985 cotton– 1%, or expressed in estimated copy numbers: the MON810: ~40.000 estimated copies, and the other three: ~400 copy numbers. This approach is somewhat different from the approach taken by Fraiture *et al.*[21]. They used lower copy numbers (20 in their lowest mixtures), but in equal amounts, for three GMOs. Both situations may actually occur in real samples that may be contaminated by UGMOs. For three GMOs the complete insertion sequence including the corresponding flanking genomic sequence was known. For one GMO, MON15985 cotton, only the element order and partial 5′ and 3′ insert sequences with corresponding flanking genomic sequence were available. Furthermore, MON15985 is a retransformation of another GMO cotton, MON531. The MON531 sequence was also unknown, with the exception of the element order. Therefore, MON15895 mimicked a UGMO in this experimental set-up.

The results can be looked at from two different perspectives. The first is whether or not the workflow enabled identification of UGMOs. The second is whether or not all the input sequences were detected. The transgenic cotton mimicked a UGMO, as explained above. Element orders corresponding to MON15985, to both constructs, were found in the *unknown parts* bin. By finding these two sequences, it was shown that a UGMO containing a known element can be identified. The sequences as such were not before specifically annotated to be part of the MON19585 and MON531 constructs. Event-specific sequences for either event were not found. The MON15985 event-specific test is at 3′ end of the insert. We therefore expected MON15985 in the *confirmed events* bin, through reads starting from tNOS in downstream direction. However, the tNOS enrichment was less successful than p35S enrichment. Out of 383 sequences, 368 were enriched starting from p35S and only 15 from tNOS, and upstream enrichment was in both cases seven times more successful than downstream. This is indeed a striking difference, and seemingly in conflict with the observation in the qPCR evaluation. Potential explanations would be variability in the procedure and or the samples, as the NGS sample (mixture of GMOs) was different from the qPCR evaluated sample (single GMO), or, some kind of bias in the library preparation/sequencing run. Noteworthy, Fraiture *et al.* also found a large variation in numbers of reads related to the different starting points and directions they used for GW[21]. Of the 16 CCS reads, two were tNOS enriched, one upstream and one downstream. The one downstream was indicative of the MON15985 GMO, via the tNOS-nnn-NTPII-p35S sequence present in the MON531 construct of MON15985 (Fig. 5).

From the other perspective, finding back the input transgenes was partially successful. Out of the 16 crCCS reads, 13 could be unequivocally linked to one of the four input sequences, MON810, through reads starting from p35S in upstream direction (Fig. 4). We could not find back event-specific sequences for MON89034 and MON19585 since the event-specific tests are at 3′ end of insert, and tNOS reads were much less abundant than p35S ones. Still, there was sequence evidence linked to the presence of MON88017 and MON15985, the latter one is discussed above. For MON88017, a partial insert sequence covering cp4epsps, the tNOS and the p35S was found. This clearly indicated the presence of MON88017, as no other transgenic line contained this sequence in this experiment. The MON88017 event-specific sequence was actually not expected to be found in the dataset, as both p35S and tNOS are present in the middle of the inserted sequence and in order to reach the point of insertion the upstream or downstream enrichments should cover ~3.5 kb. The longest CCS read was approximately 2.5 kb. Without detection of the point of insertion, the presence of MON88017 and MON15985 cannot be asserted with the same certainty as the presence of MON810. Some other transgenic lines containing the same element order could have been present in the reference material used for sample preparation. As a matter of fact, the *nptII* cassette with the CDS flanked by a p35S and tNOS is found in several GMOs, e.g. MON87460 and MON863 maize, and MON757 and MON1698 cotton.

Finding back specific sequences depends on the databases used. Elements are the building blocks of a transgenic sequences. Elements with the same function, the p35S for instance, do not always have the same sequence length in all transgenic lines. This makes building of an element database more difficult and the database larger than desired. For this reason, a local database was built containing only the longest sequence for each element. The drawback of such a database is that although element sequences are very similar they are not identical. This needs to be considered when interpreting the results. A 75% coverage of a p35S does not necessarily mean that there is an incompletely explained read with a gap between elements: it could be that this certain promoter sequence is not completely identical to the one in the transgenic line. Such was the case for the crCCS read with the event sequence *p35S – nnn - hsp70 - cry1ab*, where a 16 nucleotide gap was found between 35S promoter and HSP70. This sequence was placed in the *unknown parts* bin, where potential UGMOs are expected, although further investigation of this read showed that it belonged to MON810. Such a bias might be resolved by addition of all versions of a particular sequence to the database and considering the bit score of the hits after BLAST analysis. Implementing the 'filter and sort' tool in Galaxy would allow this, and this addition is recommended for future use of this pipeline. Another database challenge lies in element sequences not always being annotated in the same manner in all database entries. For instance, there is a possibility that a part of a plasmid used in the process of transformation can be included into the element annotation in some database entries and not in others.

The identification of three crCCS reads without any known GMO homology was not expected. All three reads contained at least partial enrichment primer sequences at 5′ or 3′ end, and therefore passed through the enrichment primer filter, pointing to an unspecific binding of enrichment primers to a sequence other than p35S or tNOS. In case of crCCS reads 120091 and 43434 the sequences can be explained as maize mitochondrion and a non-plant sequence, respectively. For crCCS read 106520, however, NCBI patent database Blast identified this sequence as part of patents US756989 and WO2014036048A1, sequences accession number GP69587 and JC61506, respectively, with the field of invention for both patents described as *related to methods altering gene expression* (Supplementary Table S4). While the crCCS read contained only the p35S enrichment primer and not the p35S promoter sequence, there is still a possibility that this sequences could be left over from the genomic transformation. In the current version of the element database we only included the known elements of the GMOs in the study, and actually left out the elements known to be present in the mimic-UGMO, such as nptII. Obviously, *nptII* should be added to the database in a real life situation. Likewise, any homology leading to identification of previously unknown elements should be added to this database, when verified. Within the current three hits, besides the titles, there were no real solid clues for a GMO related sequence. In Table S4, the query and subject starts and ends are given, plus the description of this region in the subject. None of the three reads showed any known GMO related sequence besides the presence of the primer. Therefore, the identification of a previously unknown GMO element, harbouring the 35S primer sequence, is a possible explanation. Another, perhaps more likely explanation would be off-target hybridisation of the 35S primer. In either case, if proven repeatable, this sequence might perhaps be added to one of the dedicated databases in the pipeline, probably the element database, with an annotation such as: previously found potential GMO-related sequence.

A new MON810 reference sequence was combined from previously known sequences JQ406879 and AY326434. There is overlap between sequences JQ406879 and AY326434 in the p35S promoter region. With 133 CCS reads mapping to this overlap, sufficient evidence was given for merging these two reference sequences in one reference sequence named RIKILT20151130. RIKILT20151130 replaced the JQ406879 and AY326434 sequence in the final workflow (Supplementary Data S2).

The main objective of the analysis pipeline was to guide the end-user as quickly as possible to those sequences that require further investigation. For identification of UGMOs, known authorized GMO events are merely nice to know, and reads completely covered by a known construct sequence are of even lower priority. Artefacts should be recognized and filtered, independent of where in the procedure they might occur. For this reason, the reads were set apart that contained only known primer sequences and not any further GM element homology. The placement of reads in separate bins instead of removing reads from the output, enables the end-user to further detail the analysis at any point in the procedure. The 18 CCS reads without a primer were not investigated further. The reasoning behind this was that all sequences should be related to the experiment that was performed. If none of the primers used in the experiment could be found, the link of such a sequence to the experiment itself is weaker. On top of that, if some of these did actually contain a primer, but with too many errors, the quality of the rest of the sequence would also be less reliable, increasing the risk of having to manually explore false positives. The contents of all other bins were analysed and a likely artefact was found in the *confirmed events* bin. One read showed homology to two event sequences instead of one. The raw CCS read, before trimming and primer filtering, turned out to consist of an AAP primer with a poly G tail, 5′ MON810 flanking region followed by p35S sequence, with this motive being repeated. The repetition in itself is already rather strange, and less likely to be a true molecule because of the simple reason that if so, chances are that it was published already, given the well sequenced nature of the MON810 GMO. The presence of primers used in the experiment at the very borders of the repeated sequence makes it near impossible that this sequence is not an experimental artefact. A very likely explanation is a ligation artefact, i.e. the same molecule was ligated twice in a row between smartbell adapters. In the potential UGMO output, an element order consistent with MON810 was identified, besides the relevant sequences for the partially unknown GMO. This read should have been placed in the *only known information* bin, but was not, due to threshold settings. We did choose to keep the threshold as it was. A lowering might easily solve the problem in this dataset, yet it might cause false negative discoveries in others. In this setting, it serves to illustrate the point that UGMO discovery may never be free of false positives. Therefore, all these potential discoveries should be confirmed independently, preferably by design of a novel PCR, based on the newly found sequence. Within the current dataset 2372 polymerase reads containing 11.538 subreads were present. Of those, 411 polymerase reads contained more than four subreads of high enough quality to be merged into a CCS read. Those 411 could be further clustered, based on sequence homology, into 16 distinct sequences. Only four of

those needed hands-on evaluation, of which three proved 'unknown'. All this was automated, meaning that out of a potential 11.538 reads, only 4 had to be manually checked, and three of those rightly so. In case this is a true new, potential UGMO sequence, this sequence should then be confirmed through Sanger sequencing of the PCR amplicon, from an independent DNA isolation of the suspect sample.

In summary, we conceived and tested an integral approach for a lab-based target enrichment based on a suspect GM element, followed by NGS and data analysis by design. We successfully showed the identification of partial sequences of a model UGMO. Future experiments will be aimed at further testing the approach in more complex mixtures, and the current settings of the pipeline design. This study provides a first outline of an automated, web-based analysis pipeline for identification of UGMOs containing known GM elements.

Methods

Description of certified reference materials. Certified reference materials (CRMs) of MON89034, MON88017, and MON15985 and the well-characterized reference material MON810 from a field trial[34] were used for the preparation of complex mixtures. For detailed information on the reference materials see Table 1.

DNA isolation and preparation of the mixture. Per CRM DNA was isolated using a CTAB extraction followed by the Qiagen DNeasy plant mini kit (Qiagen) according to Scholtens *et al.*[5]. 100 ± 10 mg of dry material was weighed and extraction was performed by adding 700 µl of CTAB buffer (20 g/L CTAB, 1.4 M NaCl, 0.1 M Tris, 20 mM Na2EDTA), 200 µl of nuclease-free water (Life Technologies) and 5 µl of Rnase A (Qiagen,100 mg/µl) and incubated for 15 min at 65 °C in a thermo shaker at 250 rpm. Subsequently, 20 µl of proteinase K solution (Fermentas; 20 ng/µl) was added and the mixture was incubated in the thermo shaker for another 30 min at 65 °C, 250 rpm. To precipitate detergent, proteins, and polysaccharides 200 µl of Buffer P3 (Qiagen, DNeasy plant mini kit) was added to the lysate, the mixture was mixed and cooled on ice for 5 min. After cooling on ice, the manufacturer's protocol (Qiagen, DNeasy Plant Handbook 10/2012) was followed starting from step 10. The quantity and purity of the isolated DNA was assessed from Nanodrop absorbance measurements (Nanodrop 1000 instrument, Thermo Fisher Scientific). The mixture was prepared by combining 2.68 µl of MON15985 (18.68 ng/µl), 0.91 µl of MON88017 (54.92 ng/µl), 1.09 µl of MON89034 (45.79 ng/µl), 38.4 µl of MON810 (126.4 ng/µl) and 57 µl of nuclease-free water (Life Technologies).

Linear enrichNment. Linear enrichment (LE) was performed in four separate reactions: p35S up, p35S down, tNOS up and tNOS down. Each 20 µl reaction contained: 1× Buffer 1 (17.5 mM MgCl2, Expand Long Template PCR System, Roche), 200 µM dNTPs (10 mM dNTP mix each, Invitrogen), 3.75 U polymerase blend Taq + Tgo (Expand Long Template PCR System, Roche), 125 nM biotinylated enrichment primer and 200 ng genomic DNA. Copy numbers were estimated to be the ~40.000 for MON810 and ~400 for the other three GMOs. A 1 C value was used of 2.725 for maize and 2.33 for cotton[44] and a convension factor for 50% for hetero/hemizygous GMOs as recommended by the JRC[45] to calculate in the 200 ng mixture: 35596 MON810 copies, 367 MON88017 and 367 MON89034 copies, and 429 MON15985 copies, which we rounded to 1 significant number. The following program was performed in a thermal cycler (iCycler, Bio-Rad): 2 min at 95 °C, 20 cycles of 1 min at 95 °C, 5 sec at 60 °C, ramp to 72 °C, over 1 min and 5 min at 72 °C.

Column and bead purification. Column purification was performed (Qiaquick PCR Purification Kit, Qiagen) according to the manufacturer's instructions, for removal of surplus primers and primer dimers. The linearly enriched fragments were eluted using 30 µl elution buffer provided with the kit. Streptavidin coated magnetic beads were used to select for the biotinylated enriched fragments from the genomic background. 15 µl Dynabeads MyOne Streptavidin C1 (Invitrogen) per sample were washed according to the manufacturer's instructions and resuspended in 30 µl 2× B&W buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 2 M NaCl). To immobilize the fragments, 30 µl column purified biotinylated enriched fragments were added to 30 µl washed beads and incubated for 30 min at 20 °C using a thermoshaker at 600 rpm. After immobilization, the DNA-bead complexes were washed for three times with 1× B&W buffer. Finally, samples were resuspended in 10 µl 10 mM Tris.

Tailing. The tailing reaction consisted of 5× tailing buffer (5′ RACE System for Rapid Amplification of cDNA Ends, Invitrogen), 200 µM dCTPs and 10 µl purified sample. Water was added to a final volume of 24 µl. After mixing well, the mixture was incubated for 3 min at 94 °C and chilled for 1 min on ice. 1 µl of TdT (15 U/µl, Life Technologies) was added to each sample and incubated at 37 °C for 10 min. To heat inactivate the enzyme a last step of 10 min at 65 °C was performed then the samples were put on ice.

Semi nested PCR. Expand Long Template PCR Buffer 1 (1×), dNTPs (200 µM; each, Invitrogen), semi-nested primer (0.4 µM) and Abridged Anchor Primer (0.4 µM, 5′ RACE System for Rapid Amplification of cDNA Ends, Invitrogen) (Supplementary Table S1) and *Taq* + *Tgo* polymerase blend (2.5 U, Expand Long Template PCR System, Roche) were combined with 10 µl of tailed product in a total volume of 50 µl. Amplification was achieved using the following cycling program: 94 °C for 2 min, 45 cycles of 95 °C for 10 sec and 59 °C for 6 min and a final extension of 72 °C for 5 min.

Real-time polymerase chain reaction (qPCR). Prior to q-PCR, reactions were diluted based on the amount of DNA put in the LE reaction. The following concentrations were used: 0.8 ng/µl for starting material (SM) and, 0.008 ng/µl for LE. qPCR reactions consisted of Diagenode (GMO-MM2X-A300), primers and probe (Supplementary Table S2) and 5 µl template in a total volume of 50 µl. Amplification was performed as follows: 2

min at 50 °C for Uracil N-Glycosylase (UNG) decontamination, denaturation for 10 min at 95 °C, 45 cycles of 15 s at 95 °C and 1 min at 60 °C using a MyiQ or CFX real-time PCR machine (Bio-Rad). Data-analysis was performed using CFX Manager Software Version 3.1 (Bio-Rad).

PacBio analysis. A sequencing library for the PacBio platform was generated and sequenced at BaseClear BV (Leiden, The Netherlands). The ALF reactions were pooled, concentrated using 1.8 volume AMPureXP beads (Beckman Coulter Life Sciences) and subject to size-selection on agarose gel. The DNA fraction larger than 1 kb was isolated from gel, with a yield of 9.3 ng according to a measurement with the dsDNA high sensitivity Qubit assay (Life Technologies). The standard library preparation kits and protocols of the manufacturer (Pacific Biosciences) were used for library preparation, without DNA fragmentation, and despite the low input. The final library had a very low concentration of 0.1 ng/µl and was sequenced on one PacBio RSII SMRT cell. The run generated 2372 polymerase reads with an average subread length of 1031 bp. PacBio SMRT Portal's CCS workflow was used to generate 411 CCS reads (deposited at NCBI's Biosample database; accession number SAMN06461360).

NGS data processing. CCS reads gained by PacBio RS II NGS sequencing were processed with Galaxy platform[46–48]. A pipeline, i.e. workflow (Supplementary pipeline online, https://github.com/RIKILT/ALF), was built using tools available in Galaxy (Supplementary Table S3). Three databases were built: event, CAF (Construct And Flanks) and element database (Supplementary Data S1, S2, S4 and S5). CAF database comprises of annotated reference sequences of MON810 (RIKILT20151130 joined JQ406879 and AY326434), MON89034 (DI362404), MON88017 (HV702026) and two partial MON15985 (EA135634 3′ flank, EA135633 5′ flank). Annotations were written in a General Feature Format (GFF) file (Supplementary Data S3).

References

1. Neufeld, J. L. *The State of Food and Agriculture. Canadian Journal of Comparative Medicine* **34** (2014).
2. Prusak, A., Strojny, J. & Rowe, G. Is GMO 'sustainable'? A review of the environmental risk of GM plants in comparison with conventional andorganic crops. *Mod. Manag. Rev.* https://doi.org/10.7862/rz.2014.mmr.55 (2014).
3. *Genetically Engineered Crops: Experiences and prospectives*. https://doi.org/10.17226/23395 (National Academies Press, 2016).
4. Dinon, A. Z. *et al.* Development and validation of real-time PCR screening methods for detection of cry1A.105 and cry2Ab2 genes in genetically modified organisms. *Anal. Bioanal. Chem.* **400**, 1433–1442 (2011).
5. Scholtens, I. *et al.* Practical experiences with an extended screening strategy for genetically modified organisms (GMOs) in real-life samples. *J. Agric. Food Chem.* **61**, 9097–9109 (2013).
6. Holst-Jensen, A., Rønning, S. B., Løvseth, A. & Berdal, K. G. PCR technology for screening and quantification of genetically modified organisms (GMOs). *Anal. Bioanal. Chem.* **375**, 985–993 (2003).
7. Block, A. *et al.* The GMOseek matrix: a decision support tool for optimizing the detection of genetically modified plants. *BMC Bioinformatics* **14**, 256 (2013).
8. European Committee for standardization. NPR-CEN/TS 16707:2014 en. 2016-06 Foodstuffs — Methods of analysis for the detection of genetically modified organisms and derived products — Nucleic acid extraction (2014).
9. Rosa, S. F. *et al.* Development and applicability of a ready-to-use PCR system for GMO screening. *Food Chem.* **201**, 110–119 (2016).
10. Broeders, S. *et al.* New qualitative trait-specific SYBR® Green qPCR methods to expand the panel of GMO screening methods used in the CoSYPS. *Eur. Food Res. Technol.* **241**, 275–287 (2015).
11. Park, S.-B., Kim, H.-Y. & Kim, J.-H. Multiplex PCR system to track authorized and unauthorized genetically modified soybean events in food and feed. *Food Control* **54**, 47–52 (2015).
12. Fu, W. *et al.* A temperature-tolerant multiplex elements and genes screening system for genetically modified organisms based on dual priming oligonucleotide primers and capillary electrophoresis. *Food Chem.* **229**, 396–402 (2017).
13. Wang, C. *et al.* GMO detection in food and feed through screening by visual loop-mediated isothermal amplification assays. *Anal. Bioanal. Chem.* **407**, 4829–4834 (2015).
14. Scholtens, I. M. J. *et al.* Semiautomated TaqMan PCR screening of GMO labelled samples for (unauthorised) GMOs. *Anal. Bioanal. Chem.* **409**, 3877–3889 (2017).
15. Prins, T. W., van Hoof, R. A., Scholtens, I. M. J. & Kok, E. J. Novel TaqMan PCR screening methods for element cry3A and construct gat/T-pinII to support detection of both known and unknown GMOs. *Eur. Food Res. Technol.* **243**, 481–488 (2017).
16. Spalinskas, R., V den Bulcke, M., Van den Eede, G. & Milcamps, A. LT-RADE: An Efficient User-Friendly Genome Walking Method Applied to the Molecular Characterization of the Insertion Site of Genetically Modified Maize MON810 and Rice LLRICE62. *Food Anal. Methods* 705–713, https://doi.org/10.1007/s12161-012-9438-y (2013).
17. Spalinskas, R., Van den Bulcke, M. & Milcamps, A. Efficient retrieval of recombinant sequences of GM plants by Cauliflower Mosaic Virus 35S promoter-based bidirectional LT-RADE. *Eur. Food Res. Technol.* **237**, 1025–1031 (2013).
18. Tan, G. *et al.* SiteFinding-PCR: A simple and efficient PCR method for chromosome walking. *Nucleic Acids Res.* **33**, 1–7 (2005).
19. Liang, C. *et al.* Detecting authorized and unauthorized genetically modified organisms containing vip3A by real-time PCR and next-generation sequencing. *Anal. Bioanal. Chem.* **406**, 2603–2611 (2014).
20. Fraiture, M. A. *et al.* An innovative and integrated approach based on DNA walking to identify unauthorised GMOs. *Food Chem.* **147**, 60–69 (2014).
21. Fraiture, M.-A. *et al.* An integrated strategy combining DNA walking and NGS to detect GMOs. *Food Chem.* **232**, 351–358 (2017).
22. Fraiture, M.-A. *et al.* Validation of a sensitive DNA walking strategy to characterise unauthorised GMOs using model food matrices mimicking common rice products. *Food Chem.* **173**, 1259–1265 (2015).
23. Trinh, Q., Xu, W., Shi, H., Luo, Y. & Huang, K. An A-T linker adapter polymerase chain reaction method for chromosome walking without restriction site cloning bias. *Anal. Biochmeistry* **425**, 62–67 (2012).
24. Xu, W. *et al.* Randomly broken fragment PCR with 5′ end-directed adaptor for genome walking. *Sci. Rep.* **3** (2013).
25. Thirulogachandar, V., Pandey, P., Vaishnavi, C. S. & Reddy, M. K. An affinity-based genome walking method to find transgene integration loci in transgenic genome. *Anal. Biochem.* **416**, 196–201 (2011).
26. Trinh, Q. *et al.* Loop-linker PCR: An advanced PCR technique for genome walking. *IUBMB Life* **64**, 841–845 (2012).
27. Arulandhu, A. J. *et al.* DNA enrichment approaches to identify unauthorized genetically modified organisms (GMOs). *Anal. Bioanal. Chem.* **408**, 4575–4593 (2016).
28. Rhoads, A. & Au, K. F. PacBio Sequencing and Its Applications. *Genomics. Proteomics Bioinformatics* **13**, 278–289 (2015).
29. Hackl, T., Hedrich, R., Schultz, J. & Forster, F. proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics* **30**, 3004–3011 (2014).

30. Ono, Y., Asai, K. & Hamada, M. PBSIM: PacBio reads simulator—toward accurate genome assembly. *Bioinformatics* **29**, 119–121 (2013).
31. Ross, M. G. *et al.* Characterizing and measuring bias in sequence data. *Genome Biol.* **14**, R51 (2013).
32. Hestand, M. S., Houdt, J. V., Cristofoli, F. & Vermeesch, J. R. Polymerase specific error rates and profiles identified by single molecule sequencing. *Mutat. Res. Mol. Mech. Mutagen.* **784–785**, 39–45 (2016).
33. Frohman, M. A., Dush, M. K. & Martin, G. R. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci.* **85**, 8998–9002 (1988).
34. Van De Wiel, C. C. M. *et al.* Pollen-mediated gene flow in maize tested for coexistence of GM and non-GM crops in the Netherlands: effect of isolation distances between fields. *NJAS - Wageningen J. Life Sci.* **56**, 405–423 (2009).
35. Savini, C., Mazzara, M., Munaro, B. & Van den Eede, G. *Event-specific Method for the Quantification of Cotton Line MON 15985 Using Real-time PCR Protocol. JRC Scientific and Technical Reports.* https://doi.org/10.2788/4378 (2008).
36. Savini, C. *et al. Event-specific Method for the Quantification of Maize Line MON 89034 Using Real-time PCR Protocol. JRC Scientific and Technical Reports.* https://doi.org/10.2788/68713 (2008).
37. Mazzara, M., Cordeil, S., Savini, C. & Eede, G. Van den. *Report on the Verification of the Performance of a MON 810 Event-specific Method on Maize Line MON 810 Using Real-time PCR. JRC Scientific and Technical Reports.* https://doi.org/10.2788/59036 (2009).
38. Delobel, C., Foti, N., Grazioli, E., Mazzara, M. & Van den Eede, G. *Event-specific Method for the Quantification of Maize Line MON 88017 Using Real-time PCR v. 1.01.* https://doi.org/10.2788/22579 (2013).
39. Cock, P. J. A., Chilton, J. M., Grüning, B., Johnson, J. E. & Soranzo, N. *NCBI BLAST+ integrated into Galaxy. bioRxiv.* https://doi.org/10.1101/014043 (GigaScience, 2015).
40. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* **44**, W3–W10 (2016).
41. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Meth* **10**, 996–998 (2013).
42. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
43. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
44. Nuclear DNA content of some important plant species. **9**, 415–415 (1991).
45. European Commission's Joint research Centre. *Technical guidance document from the European Union Reference Laboratory for Genetically Modified Food and Feed on the implementation of Commission Regulation (EU) No 619/2011* (2011).
46. Goecks, J., Nekrutenko, A. & Taylor, J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome biology* **11**, R86 (2010).
47. Giardine, B., Riemer, C. & Hardison, R. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* **15**, 1451–1455 (2005).
48. Blankenberg, D. *et al.* In *Current Protocols in Molecular Biology.* Chapter 19, Unit 19.10.1-21 (John Wiley & Sons, Inc., 2010).
49. EUginius - the European GMO reference database. Available at: http://www.euginius.eu/euginius/pages/home.jsf. (Accessed: 30th December 2016).

Acknowledgements

This study received funding from the European Union under Grant Agreement No. 613908 (Project DECATHLON). Complementary financial support was provided by the Slovenian Research Agency (Contract Numbers P4-0165 and 1000-15-0105) and Slovene Human Resources Development and Scholarship Fund (Contract Number 11012-51/2015), and the Dutch Ministry of Economic Affairs.

Author Contributions

A.B.K. contributed to the building of the Galaxy workflow, data analysis and preparation and writing of the manuscript, A.J.A. contributed to the design of experiments in this study and preparation and writing of the manuscript, M.M.V. contributed to the design of experiments in this study and writing of the manuscript, H. X. contributed to the experimental part of the study, R.H. and M.S. contributed to the building of the Galaxy workflow, A.C. performed the primary analysis of the PacBio data and contributed to the writing of the manuscript, J.Ž., E.K. and J.v.D. contributed to the writing of the manuscript and jointly supervised the work. All authors revised the manuscript.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017

Chapter 3

Discussion

3.1 ddPCR-Based Multi-Target Approach for GMO Quantification

A new strategy for GMO quantification is more and more needed, as the current qPCR-based system is becoming cumbersome. Thus, our first focus was to employ novel technologies for DNA-based GMO quantification. From all of the new and evolving technologies dPCR has shown the greatest potential in this field. In our first two publications, we show the design and characterisation of five multi-target droplet digital polymerase chain reaction assays that enable quantification of 12 EU-authorised GM maize lines and 11 EU-authorised soybean lines, with the addition of four GM soybean lines awaiting authorisation.

The most represented of the GM species authorised in the EU are maize and soybean. The four assays, maize 4- and 11-plex and soybean 6- and 7-plex, designed and characterised in the first two of our publications thus complement each other. We have shown that all the designed assays are in compliance with the MPPs [86], [87] set out for GMO testing. Meaning that the newly designed methods show at least as high a level of repeatability, sensitivity, specificity, and robustness as do individual qPCR assays in use as the golden standard, while reducing the time and costs of the analysis.

Without the need of standard curve and due to a high resilience to partial inhibition, which enables testing of fewer dilutions per sample, a sample confirmed to contain maize and soybean can be tested for the presence of 11 and 12 EU-authorised maize and soybean lines, respectively, in just 32 reactions (taking in to account 2 DNA extraction parallels, 2 dilutions and 2 technical repeats of each dilutions). In order to do the same with qPCR at least 14 96-well plates would be needed. This is indeed an enormous difference, which consequently makes ddPCR much more time- and cost-effective. We must consider the fact that a direct quantification of a sample using qPCR is uncommon, thus the difference seen above is somewhat overestimated. Although we can decrease the number of needed reactions in qPCR, by first screening approaches followed by an individual line identification and quantification, the number of reactions needed in ddPCR is still significantly lower.

In the case of maize assays we have focused only on authorised GMOs. A new GM maize line was authorised while our first paper was in the process of revision. This meant that in routine analysis, either one additional line would need to be analysed separately, either by ddPCR or by qPCR, or we could extend one of the assays, in which case the newly formed assays would need to be characterised from the start. We decided for the former. However we did extend the maize 4-plex in to a 5-plex, but only as a proof of principle. We did not characterise the performance of this 5-plex assay. To avoid the same

thing from happening with the soybean multi-target assays, four GM soybean lines that were in the pipeline for authorisation were added to the 7-plex assay, extending it to the 11-plex. As the assays give comparable results, it was concluded that subtraction of the lines from the assay does not change the results and that additional characterisation is not needed in this case.

We have also shown that direct transfer of already validated qPCR methods from EURL database of reference methods for GMO analysis (http://gmo-crl.jrc.ec.europa.eu/gmomethods/) to ddPCR is not only possible, but also relatively straightforward. Individual qPCR assays were directly transferred to ddPCR, and tested as simplex as well as duplex assays, on individual DNA and on a DNA mix comprising several maize or soybean GM lines. Although simplex assays were simple to transfer between platforms, combining PCR modules into multi-target assays needs to be approached with caution. *In silico* analysis can show an interaction between multiple primers and probes in a multi-target assay; however, this interaction needs to be confirmed experimentally. As was shown in the case of maize assays, even where *in silico* analysis did not show significant interactions (first division of maize targets into two 6-plex assays) experiments showed a false negative signal, and a rearrangement of targets was needed.

The inter-laboratory assessment of soybean 6- and 7-plex assays proved the transferability to other laboratories is feasible, giving base for submission of the assays for accreditation. Slovene accreditation body accredited both 6- and 7-plex assays, and both were placed on the list of ISO 17025 accredited methods at the National Institute of Biology. These two studies thus demonstrated that ddPCR is a good approach for time and cost reduction in the GMO field.

3.2 Determination of Droplet Volume in ddPCR

In the first two studies, we demonstrated that multi-target ddPCR approaches are a good candidate for further method development in the GMO field. We characterised the assays in terms of MPPs. However, if on one side we have the accuracy that is a derivative of the assay at hand, we have on the other hand limitations set by the platform.

In ddPCR the critical factor of accuracy is the droplet volume. The most used ddPCR platform today is BioRad's QX100 or QX200 platform. The analysis software has a predetermined droplet volume of 0.85 nL (from version 1.6.6 on [88]). This volume is used to calculate the copy number that the software gives. The correctness of the volume assigned by the company has been challenged before [88]–[90]. Based on previous studies conducted in our laboratory, where different dPCR platforms were tested, QX100/200 gave consistently a lower copy number [91], [92]. Since this bias is likely to come from incorrect droplet volume used in the copy number calculations, we have measured the droplets produced by two BioRad's ddPCR droplet generation platforms, in order to calculate the droplet volume.

Different mastermixes have been shown to have an effect on droplet volume [89], therefore we have decided to test two mastermixes that are used in two different PCR chemistries; probe-based or TaqMan chemistry and intercalating dye chemistry. As QX100/200 platform uses a proprietary mastermix, we tested BioRad's supermix for probes (No dUTP) and EvaGreen supermix. As we have both droplet generator systems, QX100™ Droplet Digital™ PCR systems (hereafter indicated as a DG8 manual droplet generator) and QX200™ AutoDG™ Droplet Digital™ PCR system (hereafter indicated as DG32 automated droplet generator) in our laboratory we measured droplets produced by both

systems and additionally another DG8 manual droplet generator was tested at the Italian National Metrology Institute - Instituto Nazionale di Ricerca Metrologica (INRiM; Turin, Italy).

There was a statistically significant difference between the volume assessed for the two mastermixes, and the two droplet generation systems, and what is even more crucial there has been a statistically significant difference between the droplet volume assigned by the manufacturer and the droplet volume measured in this study. For intercalating dye chemistry (EvaGreen supermix) we showed that the droplet volume is 8.5% lower than the assigned value and for probe-based chemistry the droplet volume is 15.9% lower. When we used the DG32 automated droplet generator and probe-based chemistry this difference was slightly lower than with the DG8 manual droplet generator but at 13.1% still very high. Our findings tell us that when relying on the droplet volume assigned by the manufacturer we underestimate the copy number in all systems be it probe-based or intercalating dye chemistry, manual or automated droplet generation.

Partition volume variability also has an effect on the accuracy of the copy number determination [93]. If droplets are of different sizes then the distribution of the targets is uneven, bigger droplets can contain more targets than smaller ones, and since we count positive partitions, we get an underestimation of the copy number. We have shown that the relative standard deviation of the volume of the measured droplets is around 3% for all measurements. Confirming that the source of the underestimation of the copy number is the droplet volume.

Additionally the inter-laboratory comparison showed there is no significant difference between the droplet volume of the two DG8 manual droplet generators and that there is no significant difference in the two approaches of droplet measurement. There is also no significant inter- and intra- cartridge variability.

Since in GMO testing we use relative quantification in terms of a part of the whole species e.g. soybean in the sample, the droplet volume does not have a direct effect on the GMO%. However, droplet volume has an effect on the limit of quantification (LOQ) and limit of detection (LOD) of the method. Since MPPs set the threshold of LOQ to < 50 copies per reaction and LOD to < 25 copies per reaction, assays that show LOQ and LOD close to the threshold values may exceed it when considering the corrected volume. When applying the droplet volume correction to the calculations of the LOQ and LOD for the multi-target assays designed and characterised in our first two papers we see that although both limits are now higher, they are still below the 50 and 25 copies per reaction threshold.

3.3 ALF as a Strategy for UGMO Identification

The continuation of our studies shifted focus from quantification of authorised GMOs to identification of UGMOs. In our fourth paper we have developed a UGMO identification pipeline, and demonstrated for the first time that GW coupled to NGS, can be used to identify a UGMO in a complex sample.

Looking at the drawbacks of the already designed GW approaches for UGMO identification, a new approach called ALF: amplification of linearly enriched fragments was devised. ALF incorporates the reduction of gDNA by biotinylation and the principle of tailing and semi-universal PCR found in the RACE/LT-RADE approach. As the enrichment in ALF is targeted, this approach also reduces the possibility of false positive UGMO identification.

In this study we have coupled ALF with NGS, more specifically with PacBio RSII single molecule real-time or SMRT sequencing. As SMRT sequencing is one of the two sequencing options that produces long reads; the other being nanopore sequencing by Oxford NanoPore. The drawback of both of these sequencing approaches is the error rate that is much higher than is the error rate of the most used NGS platform, Illumina. Thus the circular consensus sequence (CCS) option of PacBio was chosen to ensure a high accuracy of the reads. The compromise with this kind of sequencing is the reduction of the number of reads. Still, NGS produces an enormous amount of data that cannot be analysed manually. Thus a bioinformatics pipeline needed to be designed. The data analysis pipeline needs to be as automated and as user friendly as possible, as the final user of this kind of an approach would be a GMO testing laboratory. As the personnel of such a laboratory is usually not very familiar with the command line environment, the data analysis pipeline was constructed in the web-based analysis platform Galaxy [94].

The analysis pipeline was constructed in such a way that the end user would get to the desired information, the potential presence of UGMOs, as quick as possible, but without losing any additional potentially useful information. The idea behind the construction was step by step exclusion of the sequences with known information. First we excluded sequences that can be with certainty attributed to authorised GMOs, i.e. sequences that include the point of insertion. Then sequences that include only known information, followed by sequences with gaps of unknown information and finally sequences that contain only unknown information. However none of these sequences were discarded, rather they were placed into designated bins, enabling the end user to first go through the sequences in the bin containing partially known information as the probability that these sequences belong to a UGMO is the highest. This bin should, and we showed that it does, contain sequences with the element we enriched for followed by an unknown sequence – a potential UGMO. Since the initial steps of the pipeline filter for the presence of the primers we used for enrichment, and since these primers belong to one of the GM elements we enriched for, the last bin, the bin with only unknown information should be empty. However this was not the case. We found two sequences that are potentially linked to the sample and one that is an artefact of the library preparation, a contamination. These sequences need to be checked. When and if these sequences are found not to be a part of a UGMO, they can be added to the database so that with the next sample they would be recognised.

The sample was prepared as a non-even mixture of four GM lines, enabling the assessment of the sensitivity of the method. Three GM lines were added at 1% and one at 97%. We showed that finding back a GM line present in a high percentage is not problematic, as the majority of the sequences belonged to MON810, which was present in 97%. However out of three lines with 1% we could only find the evidence of two. This shows that the sensitivity of the method is close to 1%. Here we must point out that sequencing gave us only 411 CCS reads, and that the sensitivity could be higher, with higher sequencing yield. To truly characterise the method in terms of sensitivity, more mixtures with GM lines in different ratios should be tested.

Not only did we prove that we can identify a UGMO with ALF, we have also determined two partial sequences belonging to the GM cotton line MON15985 we used as a UGMO mimetic. These two sequences cover a previously unknown part of the insert. And additionally we have been able to combine a new reference sequence for MON810 from previously known sequences JQ406879 and AY326434. A day to day use of such an approach would give us new information, e.g. a new step and a new database could be added to the pipeline once the point of insertion would be known for UGMOs and a new bin containing confirmed UGMOs would appear, or a new qPCR or dPCR method could be developed based on the new sequence information. While the databases would become

more and more comprehensive, so would the information of GMO content in a food or feed sample.

Chapter 4

Conclusions

The present thesis offers the insight into the development and study of novel multi-target techniques for identification and quantification of GMOs, authorised and unauthorised. Acknowledging the need for new approaches in GMO quantification, multi-target ddPCR assays have been designed and characterised and even inter-laboratory assessed in our first two papers. The approach of GMO quantification per ingredient has been shown for the first time. This approach was by no means designed to discredit qPCR, which still remains the golden standard in GMO testing, however it does demonstrate that ddPCR gives specific, precise and accurate results, and is more time- and cost-effective.

The critical factor determining the accuracy of droplet digital PCR as a method is droplet volume. Our third study showed the importance of correct droplet volume determination and highlighted the factors that can influence the volume, which include mastermixes and droplet generators. An inter-laboratory study confirmed that the droplet volume is smaller than the one stated by the manufacturer, leading to the underestimation of the DNA copy number in the reaction. We have also shown that our method for droplet volume determination is metrologically sound and traceable to the SI, as it was compared to a reference method, a calibrated optical profilometer.

A genome walking approach, ALF, coupled to NGS was designed and used to enrich for sequences containing at least one of the two most common GM elements, 35S promotor and NOS terminator, in order to identify a UGMO in a complex sample. By designing a (semi)automated bioinformatics data analysis pipeline a proof of principle for finding a UGMO was shown, making ALF a potential new UGMO identification strategy. New insights into the sequence of the GM lines used in this study will also contribute to the GMO testing, either by giving options for designing new qPCR assays targeting these sequences or by adding these sequences into the databases behind the bioinformatics pipeline, making the analysis more comprehensive and the differentiation between authorised GMOs and UGMOs easier.

The developed approaches can be either directly applicable in the GMO testing field, in the case of ddPCR multiplex assays, or present a proof of principle that could become a standard testing approach in the case of ALF. Additionally, these approaches may be introduced in fields such as medicine or other fields where several targets can be quantified as a group or where only partial DNA sequence is known.

Appendix A

Supplementary Material of Included Publications

A.1 Supplementary Material for Publication 2.1

Table S-1. Maize GM events approved for food and feed use in EU per 04/01/2014

#	Name of transgenic maize event	OECD unique identifier of transgenic maize event
1	Bt11	SYN-BTØ11-1
2	DAS59122	DAS-59122-7
3	DAS1507	DAS-Ø15Ø7-1
4	GA21	MON-ØØØ21-9
5	MIR162	SYN-IR162-4
6	MIR604	SYN-IR6Ø4-5
7	MON810	MON-ØØ81Ø-6
8	MON863	MON-ØØ863-5
9	MON88017	MON-88Ø17-3
10	MON89034	MON-89Ø34-3
11	NK603	MON-ØØ6Ø3-6
12	T25	ACS-ZMØØ3-2
13	59122xNK603	DAS-59122-7 x MON-ØØ6Ø3-6
14	1507xMON88017x59122	DAS-Ø15Ø7-1 x MON-88Ø17-3 x DAS-59122-7
15	1507x59122	DAS-Ø15Ø7-1 x DAS-59122-7
16	1507xMON88017	DAS-Ø15Ø7-1 x MON-88Ø17-3
17	59122x1507xNK603	DAS-59122-7 x DAS-Ø15Ø7-1 x MON-ØØ6Ø3-6
18	Bt11xGA21	SYN-BTØ11-1 x MON-ØØØ21-9
19	Bt11xMIR604	SYN-BTØ11-1 x SYN-IR6Ø4-5
20	Bt11xMIR604xGA21	SYN-BTØ11-1 x SYN-IR6Ø4-5 x MON-ØØØ21-9
21	DAS1507xNK603	DAS-Ø15Ø7-1 x MON-ØØ6Ø3-6
22	MIR604xGA21	SYN-IR6Ø4-5 x MON-ØØØ21-9
23	MON 89034x1507x MON88017x59122	MON-89Ø34-3 x DAS-Ø15Ø7-1 x MON-88Ø17-3 x DAS-59122-7
24	MON863xMON810	MON-ØØ863-5 x MON-ØØ81Ø-6
25	MON863xMON810xNK603	MON-ØØ863-5 x MON-ØØ81Ø-6 x MON-ØØ6Ø3-6
26	MON863xNK603	MON-ØØ863-5 x MON-ØØØ6Ø3-6
27	MON88017x59122	MON-88Ø17-3 x DAS-59122-7
28	MON88017xMON810	MON-88Ø17-3 x MON-ØØ81Ø-6
29	MON89034x1507x59122	MON-89Ø34-3 x DAS-Ø15Ø7-1 x DAS-59122-7
30	MON89034x1507xMON88017	MON-89Ø34-3 x DAS-Ø15Ø7-1 x MON-88Ø17-3
31	MON89034x1507xNK603	MON-89Ø34-3 x DAS-Ø15Ø7-1 x MON-ØØ6Ø3-6
32	MON89034xMON88017x59122	MON-89Ø34-3 x MON-88Ø17-3 x DAS-59122-7
33	MON89034x1507	MON-89Ø34-3 x DAS-Ø15Ø7-1
34	MON89034x59122	MON-89Ø34-3 x DAS-59122-7
35	MON89034xMON88017	MON-89Ø34-3 x MON-88Ø17-3
36	MON89034xNK603	MON-89Ø34-3 x MON-ØØØ6Ø3-6
37	NK603xMON810	MON-ØØ6Ø3-6 x MON-ØØ81Ø-6

Table S-2. Methods and test material used in this study

Event/construct/gene	UI	Multiplex group	Method reference	Amplicon size [bp]
GA21	MON-ØØØ21-9	10-plex	QT-EVE-ZM-014	101
MON88017	MON-88Ø17-3	10-plex	QT-EVE-ZM-016	95
MON89034	MON-89Ø34-3	10-plex	QT-EVE-ZM-018	77
MIR162	SYN-IR162-4	10-plex	QT-EVE-ZM-022	92
MIR604	SYN-IR6Ø4-5	10-plex	QT-EVE-ZM-013	76
T25	ACS-ZMØØ3-2	10-plex	QT-EVE-ZM-011	102
MON810	MON-ØØ810-6	10-plex	QT-EVE-ZM-020	92
MON863	MON-ØØ863-5	10-plex	QT-EVE-ZM-009	84
DAS1507	DAS-Ø15Ø7-1	10-plex	QT-EVE-ZM-010	58
NK603	MON-ØØ6Ø3-6	4-plex	QT-EVE-ZM-008	108
Bt11	SYN-BTØ11-1	4-plex	Brodmann et al[1]	75
DAS59122	DAS-59122-7	4-plex	QT-EVE-ZM-012	86
hmgA	-	10-plex and 4-plex	QT-TAX-ZM-002	79

UI: OECD unique identifier of the transgenic maize event.

Multiplex group: multiplex assay in which the event/construct/gene is present.

Method reference: reference for the method for quantification of event/construct/gene. Methods starting with QT- were validated by the EURL-GMFF.

Table S-3. Putative interferences between the primers and probes for selected qPCR modules.

#	Sequence One	Sequence Two	Match Sequence	Score	Est. Tm (°C)	ΔG (kcal/mol)	Solution
1	DAS59122P	MON88017P	TCCCGCCTTCAGTTTAAA	18	54.4	-16.5	Divide[1]
2	MIR604P	DAS59122R	CATGATCAG	9	19.6	-5.6	Divide[1]
3	MIR604P	MON88017P	TCCCGCCT	8	30.5	-8.2	Accept[2]

[1] Divide = the two PCR modules were not combined in the same multiplex assay.

[2] Accept = the putative interference was evaluated and concluded to be insignificant.

Table S-4. Primers and probes used in the 4-plex assay

Name	DNA sequence of the oligonucleotide (5'-sequence-3')	Final concentration in the PCR [nmol/L]	Concentration in PPP mix [nmol/L]
	hmgA as the target sequence:		
Fw-hmgA	TTGGACTAGAAATCTCGTGCTGA	900	3000
R-hmgA	GCTACATAGGGAGCCTTGTCCT	900	3000
P-hmgA	HEX-CAATCCACACAAACGCACGCGTA-BHQ-1	180	600
	Bt11[a] as the target sequence:		
Fw-Bt11	GCGGCTTATCTGTCTCAGGG	600	2000
R-Bt11	CAACTGGTCTCCTCTCCGGA	600	2000
P-Bt11	6-FAM-CGTGTTCCCTCGGATCTCGACATGT-BHQ-1	180	600
	NK603 as the target sequence:		
Fw-NK603	ATGAATGACCTCGAGTAAGCTTGTTAA	900	2000
R-NK603	AAGAGATAACAGGATCCACTCAAACACT	900	2000
P-NK603	6-FAM-TGGTACCACGCGACACACTTCCACTC-BHQ-1	180	600
	DAS59122 as the target sequence:		
Fw-DAS59122	GGGATAAGCAAGTAAAAGCGCTC	600	2000
R-DAS59122	CCTTAATTCTCCGCTCATGATCAG	600	2000
P-DAS59122	6-FAM- TTTAAACTGAAGGCGGGAAACGACAA -BHQ-1	200	667

PPP mix: a mix of forward primer, reverse primer and probe; 6-FAM: 6-carboxyfluorescein; BHQ1: Black Hole Quencher 1; HEX: hexachloro-6-carboxyfluorescein.

[a] construct specific method.

Table S-5. Primers and probes used in the 10-plex assay

Name	DNA sequence of the oligonucleotide (5'-sequence-3')	Final concentration in the PCR [nmol/L]	Concentration in PPP mix [nmol/L]
	hmgA as the target sequence:		
Fw-hmgA	TTGGACTAGAAATCTCGTGCTGA	900	3000
R-hmgA	GCTACATAGGGAGCCTTGTCCT	900	3000
P-hmgA	HEX-CAATCCACACAAACGCACGCGTA-BHQ-1	180	600
	DAS1507 as the target sequence:		
Fw-DAS1507	TAGTCTTCGGCCAGAATGG	450	1500
R-DAS1507	CTTTGCCAAGATCAAGCG	450	1500
P-DAS1507	6-FAM-TAACTCAAGGCCCTCACTCCG-BHQ-1	90	300
	GA21 as the target sequence:		
Fw-GA21	CGTTATGCTATTTGCAACTTTAGAACA	450	1500
R-GA21	GCGATCCTCCTCGCGTT	450	1500
P-GA21	6-FAM-TTTCTCAACAGCAGGTGGGTCCGGGT-BHQ-1	90	300
	MIR604 as the target sequence:		
Fw-MIR604	GCGCACGCAATTCAACAG	150	500
R-MIR604	GGTCATAACGTGACTCCCTTAATTCT	150	500
P-MIR604	6-FAM-AGGCGGGAAACGACAATCTGATCATG-BHQ-1	90	300
	MIR162 as the target sequence:		
Fw-MIR162	GCGCGGTGTCATCTATGTTACTAG	150	500
R-MIR162	TGCCTTATCTGTTGCCTTCAGA	150	500
P-MIR162	6-FAM-TCTAGACAATTCAGTACATTAAAAACGTCCGCCA-BHQ-1	90	300
	MON810 as the target sequence:		
Fw-MON810	TCGAAGGACGAAGGACTCTAACGT	150	500
R-MON810	GCCACCTTCCTTTTCCACTATCTT	150	500
P-MON810	6-FAM-AACATCCTTTGCCATTGCCCAGC-BHQ-1	90	300
	MON863 as the target sequence:		
Fw-MON863	TGTTACGGCCTAAATGCTGAACT	450	1500
R-MON863	GTAGGATCGGAAAGCTTGGTAC	450	1500
P-MON863	6-FAM-TGAACACCCATCCGAACAAGTAGGGTCA-BHQ-1	90	300
	MON88017 as the target sequence:		
Fw-MON88017	GAGCAGGACCTGCAGAAGCT	150	500
R-MON88017	GCCGGAGTTGACCATCCA	150	500
P-MON88017	6-FAM-TCCCGCCTTCAGTTTAAACAGAGTCGGGT-BHQ-1	90	300
	MON89034 as the target sequence:		
Fw-MON89034	TTCTCCATATTGACCATCATACTCATT	450	1500
R-MON89034	CGGTATCTATAATACCGTGGTTTTTAAA	450	1500
P-MON89034	6-FAM-ATCCCCGGAAATTATGTT-MGBNFQ	180	300
	T25 as the target sequence:		
Fw-T25	ACAAGCGTGTCGTGCTCCAC	450	1500
R-T25	GACATGATACTCCTTCCACCG	450	1500
P-T25	6-FAM-TCATTGAGTCGTTCCGCCATTGTCG-BHQ-1	90	300

PPP mix: a mix of forward primer, reverse primer and probe; 6-FAM: 6-carboxyfluorescein; BHQ1: Black Hole Quencher 1; HEX: hexachloro-6-carboxyfluorescein; MGBNFQ: Molecular-Groove Binding Non-fluorescence Quencher.

Table S-6. Results of experiments for determining the transgene copies with 4-plex assay on a dilution series (evaluating absolute limits of quantification and detection).

Sample - dilution	avg cp repeat 1	cv% repeat 1	avg cp repeat 2	cv% repeat 2	avg cp repeat 3	cv% repeat 3	avg cp all	cv% all
DNA mix - stock	nd	nd	1458	3	1531	2	1495	4
DNA mix - 6x	309	3	295	8	288	10	298	7
DNA mix - 12x	182	11	206	3	163	8	182	12
DNA mix - 60x	41	20	44	18	43	17	42	17
DNA mix - 120x	22	18	24	35	20	43	22	31
DNA mix - 240x	6	21	8	20	7	46	7	31
DNA mix - 600x	6[a]	[b]	8[a]	[b]	8[a]	[b]	7[a]	[b]
DNA mix - 1200x	neg	neg	neg	neg	neg	neg	neg	neg
DNA mix - 6000x	neg	neg	neg	neg	neg	neg	neg	neg
DNA mix - 12000x	neg	neg	neg	neg	neg	neg	neg	neg
DNA mix - 60000x	neg	neg	neg	neg	neg	neg	neg	neg
DNA mix - 120000x	neg	neg	neg	neg	neg	neg	neg	neg
DNA mix - 600000x	neg	neg	neg	neg	neg	neg	neg	neg

[a] three out of four replicates were negative (negative reactions were excluded from calculation of average copies).

[b] cv% was not calculated as only one replicate was positive.

nd: no data available due to technical error.

neg: all replicates negative.

avg cp repeat: average copy number of transgenes determined for each sample (from four replicates) per repeat.

avg cp all: overall average copy number of transgenes determined for each sample (from twelve replicates).

cv% repeat: percentage value of coefficient of variation per sample determined from the results of four replicates in each repeat (not calculated for the sets containing at least one negative replicate).

cv% all: percentage value of coefficient of variation per sample determined from all results of twelve replicates (not calculated for the sets containing at least one negative replicate).

Cells highlighted in green represent the values at aLOQ.

Cells highlighted in blue represent the values at aLOD.

Table S-7. Results of experiments for transgene copies with 10-plex assay on a dilution series (evaluating absolute limits of quantification and detection).

Sample - dilution	avg cp repeat 1	cv% repeat 1	avg cp repeat 2	cv% repeat 2	avg cp repeat 3	cv% repeat 3	avg cp all	cv% all
DNA mix - stock	nd	nd	4832	5	4693	4	4763	5
DNA mix - 6x	973	5	1026	5	1001	6	1000	5
DNA mix - 12x	646	8	651	9	630	1	642	6
DNA mix - 60x	167	16	143	15	160	5	157	14
DNA mix - 120x	75	19	62	14	63	13	67	17
DNA mix - 240x	24	17	32	16	30	26	29	23
DNA mix - 600x	20	32	19	28	13	60	17	39
DNA mix - 1200x	9[a]	30	8[a]	[b]	12[a]	64	10[a]	42
DNA mix - 6000x	neg	neg	neg	neg	neg	neg	neg	neg
DNA mix - 12000x	neg	neg	neg	neg	neg	neg	neg	neg
DNA mix - 60000x	neg	neg	neg	neg	neg	neg	neg	neg
DNA mix - 120000x	neg	neg	neg	neg	neg	neg	neg	neg
DNA mix - 600000x	neg	neg	neg	neg	neg	neg	neg	neg

[a] at least one replicate was negative (negative reactions were excluded from calculation of average copies).

[b] cv% was not calculated as only one replicate was positive.

nd: no data available due to technical error.

neg: all replicates negative.

avg cp repeat: average copy number of transgene determined for each sample (from four replicates) per repeat.

avg cp all: overall average copy number of transgene determined for each sample (from twelve replicates).

cv% repeat: percentage value of coefficient of variation per sample determined from the results of four replicates in each repeat (not calculated for the sets containing at least one negative replicate).

cv% all: percentage value of coefficient of variation per sample determined from all results of twelve replicates (not calculated for the sets containing at least one negative replicate).

Cells highlighted in green represent the values at aLOQ.

Cells highlighted in blue represent the values at aLOD.

Table S-8. Results of experiments for *hmgA* copies with 4-plex and 10-plex assay on a dilution series (evaluating absolute limits of quantification and detection).

Sample - dilution	avg cp repeat 1	cv% repeat 1	avg cp repeat 2	cv% repeat 2	avg cp repeat 3	cv% repeat 3	avg cp all	cv% all
DNA mix - stock	nd	nd	99211	8	100463	7	99837	7
DNA mix - 6x	22025	2	22158	1	21582	4	21922	3
DNA mix - 12x	13415	3	13308	2	12983	2	13224	3
DNA mix - 60x	3041	4	2830	9	2951	3	2945	6
DNA mix - 120x	1476	5	1265	7	1309	4	1350	9
DNA mix - 240x	495	11	524	11	501	12	507	11
DNA mix - 600x	322	9	279	10	229	9	277	17
DNA mix - 1200x	149	13	128	12	130	13	136	14
DNA mix - 6000x	25	28	25	26	22	13	24	24
DNA mix - 12000x	11	33	10[a]	26	10[b]	19[b]	10	32
DNA mix - 60000x	9[c]	[d]	neg	neg	neg	neg	9[e]	[e]
DNA mix - 120000x	6[f]	17	neg	neg	neg	neg	6[e]	17[e]
DNA mix - 600000x	neg	neg	neg	neg	neg	neg	neg	neg

[a] one replicate was negative (negative reactions were excluded from calculation of average copies).

[b] value represents both aLOQ and aLOD

[c] one out of eight replicates was positive (negative reactions were excluded from calculation of average copies).

[d] cv% was not calculated as only one replicate was positive.

[e] only the data from repeat 1 are considered

[f] two out of eight replicates were positive (negative reactions were excluded from calculation of average copies).

nd: no data available due to technical error.

neg: all replicates negative.

avg cp repeat: average copy number of transgene determined for each sample (from four replicates) per repeat.

avg cp all: overall average copy number of transgene determined for each sample (from twelve replicates).

cv% repeat: percentage value of coefficient of variation per sample determined from the results of four replicates in each repeat (not calculated for the sets containing at least one negative replicate).

cv% all: percentage value of coefficient of variation per sample determined from all results of twelve replicates (not calculated for the sets containing at least one negative replicate).

Cells highlighted in green represent the values at aLOQ.

Cells highlighted in blue represent the values at aLOD.

Table S-9. Sum of results for copies of all maize transgenes in 4-plex and 10-plex assay on a dilution series (evaluating absolute limits of quantification).

Sample - dilution	avg cp repeat 1	cv% repeat 1	avg cp repeat 2	cv% repeat 2	avg cp repeat 3	cv% repeat 3	avg cp all	cv% all
DNA mix - stock	nd	nd	6290	4	6225	3	6257	3
DNA mix - 6x	1282	4	1321	4	1289	5	1297	4
DNA mix - 12x	830	9	838	8	793	2	818	6
DNA mix - 60x	208	16	176	8	203	2	198	12
DNA mix - 120x	98	17	86	19	83	14	89	17
DNA mix - 240x	29	16	40	10	37	23	35	20
DNA mix - 600x	21[a]	40	21[a]	22	15[a]	45	19[a]	36
DNA mix - 1200x	9[b]	30	8[b]	[c]	12[b]	64	10[b]	42
DNA mix - 6000x	neg	neg	neg	neg	neg	neg	neg	neg
DNA mix - 12000x	neg	neg	neg	neg	neg	neg	neg	neg
DNA mix - 60000x	neg	neg	neg	neg	neg	neg	neg	neg
DNA mix - 120000x	neg	neg	neg	neg	neg	neg	neg	neg
DNA mix - 600000x	neg	neg	neg	neg	neg	neg	neg	neg

[a] three out of four 4-plex reactions were negative (negative reactions were excluded from calculation of GM%).

[b] all 4-plex reactions and at least one replicate of 10-plex assay were negative (negative reactions were excluded from calculation of GM%).

[c] cv% was not calculated due to only one positive reaction.

nd: no data available due to technical error.

neg: all replicates negative.

avg cp repeat: average copy number of transgene determined for each sample (from four replicates) per repeat.

avg cp all: overall average copy number of transgene determined for each sample (from twelve replicates).

cv% repeat: percentage value of coefficient of variation per sample determined from the results of four replicates in each repeat (not calculated for the sets containing at least one negative replicate).

cv% all: percentage value of coefficient of variation per sample determined from all results of twelve replicates (not calculated for the sets containing at least one negative replicate).

Cells highlighted in green represent the values at aLOQ.

Table S-10. Results of experiments performed on GM DNA sample diluted in non-GM maize DNA for 4-plex and 10-plex assay (evaluating relative limit of quantification).

Sample – theoretical GM content	Assay	avg transgene cp/reaction	cv% cp/reaction	avg GM%	cv% for GM%
DNA mix 1 - 2.00%	10-plex	582	2	1.35	2
DNA mix 1 - 0.9%		255	11	0.55	11
DNA mix 1 - 0.50%		127	13	0.28	12
DNA mix 1 - 0.10%		28	17	0.058	16
DNA mix 1 - 0.05%		12	68	0.027	70
DNA mix 1 - 0.01%		7[a]	16	0.015[a]	16
DNA mix 1 - 0.05%		neg	neg	neg	neg
DNA mix 1 - 0.001%		neg	neg	neg	neg
DNA mix 2 - 0.9%	4-plex	339	6	0.85	5
DNA mix 2 - 0.50%		168	8	0.38	7
DNA mix 2 - 0.10%		33	18	0.068	19
DNA mix 2 - 0.05%		12	34	0.025	34
DNA mix 2 - 0.01%		4[b]	57	0.013[b]	24
DNA mix 2 - 0.05%		neg	neg	neg	neg
DNA mix 2 - 0.001%		neg	neg	neg	neg

[a] two out of six reactions were negative for presence of transgenes (negative reactions were excluded from calculations).

[b] two out of six reactions were negative for presence of transgenes (negative reactions were excluded from calculations).

Theoretical GM content: calculated based on copy number values from characterization of individual DNAs, which were later on mixed to get the DNA mix 1 and 2. This was expected GM content, however the resulting average GM% should not necessary be the same due to measurement uncertainty in DNA characterization and errors in pipetting.

avg transgene cp/reaction: average copy number of transgene determined for the sample based on six replicates.

cv% cp/reaction: percentage value of coefficient of variation for copy number of transgenes per sample determined from the results of six replicates.

avg GM%: percentage GM content in the sample calculated from the six replicates (transgene copy number/endogene copy number).

cv% for GM: percentage value of coefficient of variation for GM% per sample determined from the results of six replicates.

neg: all replicates negative for transgenes (*hmgA* was positive).

Cells highlighted in green represent the values at rLOQ.

Cells highlighted in blue represent the values at rLOD.

Table S-11. False-positive rates observed with ddPCR in both multiplex assays

Sample	False positive wells	positive droplets (% of total) for *hmgA* in 4-plex	positive droplets (% of total) for transgenes in 4-plex	total droplets (4-plex)	positive droplets (% of total) for *hmgA* in 10-plex	positive droplets (% of total) for transgenes in 10-plex	total droplets (10-plex)
Bt176 non-authorized GM maize	0	22257 (88.4)	0 (0)	25179	23051 (88.4)	1 (3.8e-3)	26071
DP98140 non-authorized GM maize	0	18443 (73.1)	0 (0)	25235	19239 (72.7)	0 (0)	26469
MON87460 non-authorized GM maize	0	21989 (92.5)	1 (4.2e-3)	23763	21780 (92.6)	0 (0)	23532
non-transgenic sweet corn	0	7784 (31.1)	3.4e-3	25060	9104 (31.0)	1 (3.4e-3)0	29339
non-transgenic maize seeds	0	16193 (77.4)	1 (4.8e-3)	20925	19558 (75.7)	0 (0)	25823
MON89788 soybean	0	1 (4.0 e-3)	2 (7.9e-3)	25276	0 (0)	1 (4.2e-3)	23604
MON40-3-2 soybean	0	1 (4.1 e-3)	0 (0)	24477	1 (4.9 e-3)	0 (0)	20248
GT73 rapeseed	0	0 (0)	0 (0)	23165	1 (4.0 e-3)	0 (0)	24844
non-transgenic rape seeds	0	0 (0)	2 (7.6e-3)	26201	0 (0)	0 (0)	27247
LL62 rice	0	0 (0)	0 (0)	25980	0 (0)	0 (0)	25431
non-transgenic rice grains	0	0 (0)	0 (0)	24756	0 (0)	0 (0)	24919
non-transgenic linseed	0	0 (0)	0 (0)	23018	1 (4.1 e-3)	0 (0)	24475
water	0	2 (4.0 e-3)	1 (2.0e-3)	49672	0 (0)	0 (0)	53314

Table S-12. Trueness of ddPCR tested on proficiency samples and routine diagnostics samples

Organizer	Sample	Robust mean (sum of events; m/m)[a] [%]	Interval of ±25% of the robust mean value	ddPCR quantification (sum of 4-plex and 10-plex converted to m/m)[b] [%]	Z-score for ddPCR	sample description
EURL	G226/13	2.30	1.73 – 2.88	2.57	0.5	biscuit flour
	G227/13	0.89	0.67 – 1.12	0.86	-0.2	
USDA	179/08	3.49	2.62 – 4.36	3.02	-0.5	maize flour
	180/08	4.34	3.25 – 5.42	3.52	-1.1	
	181/08	3.54	2.66 – 4.43	3.34	-0.3	
	186/09	2.98	2.24 – 3.73	3.19	0.1	
	188/09	5.85	4.39 – 7.31	5.77	-0.1	
	189/09	1.77	1.32 – 2.21	1.34 (1.85)[c]	-1.4 (0.3)[c]	
	190/09	3.35	2.51 – 4.18	2.93	-0.5	
	254/06	2.44	1.83 – 3.05	1.83	-0.6	
	255/06	7.67	5.75 – 9.59	8.19	0.2	
	257/06	5.49	4.12 – 6.86	5.63	0.1	
	258/06	0.58	0.43 – 0.72	0.51	0.4	
routine sample	G154/13	5.69	4.27 – 7.11	5.65	n.a.	maize flour
	G147/08	54.6	41.0 – 68.3	41.5	n.a.	wheat and maize flour
	G187/14	175	131 - 219	193.2	n.a.	white hominy corn
	G189/14	157	117 - 196	159.3	n.a.	white corn flour

[a] quantification results based on qPCR (see experimental section for details on calculations); for routine samples the value represents the quantification with qPCR performed in our lab (not a robust mean)

[b] ddPCR result was converted from cp/cp to m/m (see experimental section) to be comparable to qPCR results

[c] value in brackets represents the adjusted value, taking into account that all GM seeds (used for preparation of this sample) were of male GM parental origin.

n.a. not applicable as the samples were not part of proficiency tests

Table S-13. Comparison of measured GM content with certified and assigned values (4-plex and 10-plex assay summed together)

Sample - dilution	Measured GM% (cp/cp %)	Certified GM% (converted to cp/cp %)	% bias to certified	Assigned GM% (cp/cp %)	% bias to assigned
DNA mix - stock	6.27	6.75	-7.18	6.10	2.75
DNA mix - 6x	5.92		-12.36		-2.98
DNA mix - 12x	6.18		-9.56		0.12
DNA mix - 60x	6.71		-1.59		8.94
DNA mix - 120x	6.58		-2.52		7.91
DNA mix - 240x	7.00		3.71		14.81

Measured GM%: GM content measured for individual sample with multiplex ddPCR

Certified GM%: GM content calculated based on the certified values of individual CRMs in the mix. The conversion to copy/copy % was done according to the EURL Technical Guidance on implementation of Regulation 619/2011[2] (taking into consideration the zygosity, but not gender).

% bias to certified: bias of measured GM% value against certified GM%

Assigned GM%: summed GM content in DNA mix calculated from copy numbers resulting from simplex ddPCR assays.

% bias to characterized: bias of measured GM% value against characterized GM%.

Table S-14. Detailed presentation of robustness results by introducing different changes to the normal protocol.

Sample - dilution	Assay	original protocol		annealing at 61°C				annealing at 59°C				different mastermix			
		average cp/reaction	GM%	average cp/reaction	GM%	% bias to original (copies)	% bias to original (GM%)	average cp/reaction	GM%	% bias to original (copies)	% bias to original (GM%)	average cp/reaction	GM%	% bias to original (copies)	% bias to original (GM%)
DNA mix - 60x	4-plex (3 GM lines)	42	1.46	26	0.98	-37.6	-33.1	36	1.29	-15.6	-11.7	42	1.60	-1.7	9.4
	4-plex (*hmgA*)	2897		2702		-6.7		2767		-4.5		2604		-10.1	
DNA mix - 120x	4-plex (3 GM lines)	22	1.62	12	1.15	-47.1	-29.4	15	1.47	-30.3	-9.6	26	2.32	19.0	43.1
	4-plex (*hmgA*)	1348		1009		-25.2		1038		-23.0		1121		-16.8	
DNA mix - 60x	10-plex (9 GM lines)	157	5.25	142	5.36	-9.5	2.2	109	4.08	-30.3	-22.3	128	4.53	-18.7	-13.7
	10-plex (*hmgA*)	2990		2650		-11.4		2684		-10.2		2815		-5.8	
DNA mix - 120x	10-plex (9 GM lines)	67	4.96	59	5.86	-12.6	18.2	41	3.70	-38.9	-25.3	68	5.73	0.9	15.6
	10-plex (*hmgA*)	1352		1001		-26.0		1106		-18.2		1181		-12.7	
		original protocol		different PCR cycler				multichannel pipetting to cartridge				automatic multichannel pipetting to plate			
		average cp/reaction	GM%	average cp/reaction	GM%	% bias to original (copies)	% bias to original (GM%)	average cp/reaction	GM%	% bias to original (copies)	% bias to original (GM%)	average cp/reaction	GM%	% bias to original (copies)	% bias to original (GM%)
DNA mix - 60x	4-plex (9 GM lines)	42	1.46	42	1.55	-1.6	6.0	42	1.53	-0.8	4.9	31	1.19	-26.0	-18.8
	4-plex (*hmgA*)	2897		2688		-7.2		2740		-5.4		2640		-8.9	
DNA mix - 120x	4-plex (9 GM lines)	22	1.62	17	1.51	-23.6	-6.8	16	1.40	-28.1	-13.4	17	1.63	-21.4	0.6
	4-plex (*hmgA*)	1348		1105		-18.0		1119		-16.9		1053		-21.8	
DNA mix - 60x	10-plex (9 GM lines)	157	5.25	204	7.75	29.8	47.7	128	4.84	-18.6	-7.9	130	4.88	-17.1	7.0
	10-plex (*hmgA*)	2990		2627		-12.1		2641		-11.7		2665		-10.9	
DNA mix - 120x	10-plex (9 GM lines)	67	4.96	56	4.97	-15.8	0.4	89	7.77	32.9	56.8	61	5.66	-9.4	14.3
	10-plex (*hmgA*)	1352		1134		-16.1		1146		-15.3		1072		-20.7	

Results for the protocols with changes to the original protocol are presented below each noted change (highlighted in blue).

avg cp/reaction: average copy number of transgenes or *hmgA* determined for the sample .

GM%: percentage GM content in the sample (transgene copy number/endogene copy number).

% bias to original (copies): percentage bias of average copy number result determined using changed protocol in comparison to original protocol.

% bias to original (GM%): percentage bias of GM% determined from results gathered with changed protocol in comparison to original protocol.

Cells highlighted in grey represent values above the 30% threshold (acceptance criteria for robustness).

For DNA mix – 120x the target concentration in 4-plex assay was below the LOQ (also with the original protocol) and reliable quantitative results could therefore not be expected.

Table S-15. List of all CRM test materials used in this study

Event/construct/gene	UI	Producer
Bt 176	SYN-EV176-9	IRMM
Bt11	SYN-BTØ11-1	IRMM
DAS1507	DAS-Ø15Ø7-1	IRMM
DAS59122	DAS-59122-7	IRMM
DP98140	DP-Ø9814Ø-6	IRMM
GA21	MON-ØØØ21-9	IRMM
GT73	MON-ØØØ73-7	AOCS
LL62	ACS-OSØØ2-5	AOCS
MIR162	SYN-IR162-4	AOCS
MIR604	SYN-IR6Ø4-5	IRMM
MON40-3-2	MON-04032-6	AOCS
MON810	MON-ØØ810-6	IRMM
MON863	MON-ØØ863-5	IRMM
MON87460	MON-8746Ø-4	AOCS
MON88017	MON-88Ø17-3	AOCS
MON89034	MON-89Ø34-3	AOCS
MON98788	MON-89788-1	AOCS
NK603	MON-ØØ6Ø3-6	IRMM
T25	ACS-ZMØØ3-2	AOCS

UI: OECD unique identifier of the transgenic maize event.

IRMM: Institute for Reference Materials and Measurements, Geel, Belgium.

AOCS: American oil chemists' society.

Additional data is presented in the supplementary table S-15. This table consists of a MIQE checklist and is omitted here, since it represents an extensive additional text, not needed for the understanding of the study.

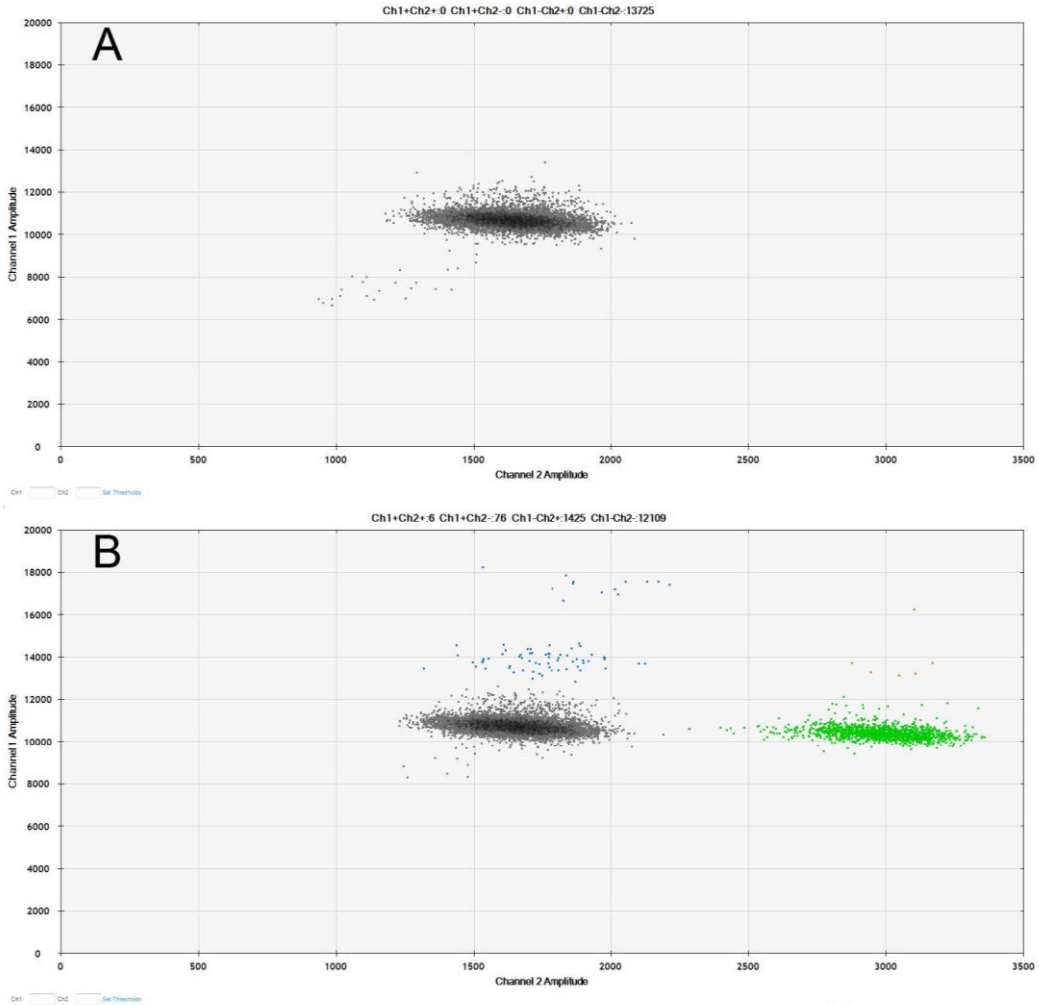


Figure S-1. Example of positive and negative experimental result. Fluorescence of negative droplets (A) as a result of NTC sample amplification. Fluorescence of droplets after 10-plex amplification of a positive sample (B). Droplets positive for GM events are in blue, droplets positive for *hmgA* are presented in green and droplets positive for both GM events and *hmgA*) are presented in orange.

A.2 Supplementary Material for Publication 2.2

Table S1. Soybean GM-events authorised and in the process of authorisation for food and feed use in the EU (to September 2016).

Line/species	Unique identifier	Legal status
A2704-12	ACS-GMØØ5-3	Authorised
MON89788	MON-89788-1	Authorised
MON40-3-2	MON-Ø4Ø32-6	Authorised
356043	DP-356Ø43-5	Authorised
A5547-127	ACS-GMØØ6-4	Authorised
MON87705	MON-877Ø5-6	Authorised
MON87701	MON-877Ø1-2	Authorised
MON87708	MON-877Ø8-9	Authorised
MON87769	MON-87769-7	Authorised
305423	DP-3Ø5423-1	Authorised
BPS-CV127-9	BPS-CV127-9	Authorised
MON87701 x MON89788	MON-877Ø1-2 x MON-89788-1	Authorised
FG72*	MST-FGØ72-2	Authorised
MON 87705 x MON 89788	MON-877Ø5-6 × MON-89788-1	Authorised
MON 87708 x MON 89788	MON-877Ø8-9 x MON-89788-1	Authorised
DAS-68416-4	DAS-68416-4	Valid EFSA application
FG72*	MST-FGØ72-2	Valid EFSA application
DAS-44406-6	DAS-444Ø6-6	Valid EFSA application
DAS-81419-2	DAS-81419-2	Valid EFSA application
305423 x 40-3-2	DP-305423-1 x MON-04032-6	Valid EFSA application
MON 87769 x MON 89788	MON-87769-7 × MON-89788-1	Valid EFSA application
MON 87705 x MON 89788	MON-877Ø5-6 × MON-89788-1	Valid EFSA application
MON 87708 x MON 89788	MON-877Ø8-9 x MON-89788-1	Valid EFSA application
MON 87705 x MON 87708 x MON89788*	MON-877Ø5-6 × MON-877Ø8-9 x MON-89788-1	Valid EFSA application

* On July 22, after the presented experimental work was completed, FG72 was authorised, together with two hybrid stacks, MON 87705 × MON 89788 and MON 87708 × MON 89788, additionally for the stack MON 87705 x MON 87708 x MON89788 a valid EFSA application was submitted.

Table S2. Soybean GM-events used in the ePCR.

Soybean GM-events	GenBank accession
A2704-12*	CS447634
A2704-12**	CS447635
A5547-127	CS444252
BPS-CV127-9	HD112081
DAS-44406-6	JC006926
DAS-68416-4	JB090933
DAS-81419-2	JC006928
305423	GM037889
356043	FB727428
Event1606	HK037693
FG72	HK747743
MON40-3-2	HW479040
MON87712	HW503815
MON87701	HB386793
MON87705	JA486740
MON87708	JB048350
MON87754	HK762611
MON87769	HB976128
MON89788	FB572992
SYHT0H2	JC050419

*3′ junction between the host and insert, not including the target for QT-EVE-GM-004

**5′ junction between the host and insert, not including the target for QT-EVE-GM-004

Table S3. Certified reference material characterisation in copies per reaction.

Soybean GM-events	Mean copy number			
	Simplex	Duplex	Multiplex	Mean
MON87701	329548	312255	246741	296181
*Le*1	333848	323910	261405	306387
MON87708	170792	186888	170838	176172
*Le*1	165436	183505	167695	172212
MON87769	2574882	2409375	2441373	2475210
*Le*1	2613941	2544248	2562539	2573576
305423	23491	31063	25673	26742
*Le*1	298139	358422	314061	323541
BPS-CV127-9	235109	233961	213966	227679
*Le*1	397970	425314	404040	409108
A2704-12	422058	390567	361810	391478
*Le*1	411862	414867	381314	402681
MON89788	130258	122812	103137	118736
*Le*1	138150	131760	114036	127982
MON40-3-2	46314	36459	27697	36823
*Le*1	426855	392449	279546	366283
356043	54104	48933	35155	46064
*Le*1	534809	444045	331342	436732
MON87705	320407	345418	302935	322920
*Le*1	336037	346189	309110	330445
A5547-127	316956	335020	289262	313746
*Le*1	312398	354759	311134	326097
DAS-68416-4	46390	45262	28432	40028
*Le*1	410511	381669	272252	354811
FG72	345804	331541	315753	331032
*Le*1	325130	322551	334878	327519
DAS-44406-6	230476	261977	156168	216207
*Le*1	227793	226744	145001	199846
DAS-81419-2	359756	348622	287884	332087
*Le*1	368132	285900	338532	330855

Table S4. DNA mix dilutions and their assigned copy numbers per reaction for each multiplex.

DNA mix		Assigned copy number		
		6-plex	7-plex	11-plex
1	GM-events	2527	2761	4752
	*Le*1	26294	26294	26294
2	GM-events	421	460	792
	*Le*1	4382	4382	4382
3	GM-events	211	230	396
	*Le*1	2191	2191	2191
4	GM-events	42	46	79
	*Le*1	438	438	438
5	GM-events	21	23	40
	*Le*1	219	219	219
6	GM-events	11	12	20
	*Le*1	110	110	110
7	GM-events	4	5	8
	*Le*1	44	44	44
8	GM-events	2	2	4
	*Le*1	22	22	22
9	GM-events	0	0	1
	*Le*1	4	4	4
10	GM-events	0	0	0
	*Le*1	2	2	2

Table S5. Verification of assigned copy numbers per µL for each individual transgene in DNA mix 1, with duplex ddPCR.

Soybean GM-events	Duplex		Assigned		Bias	
	Soybean GM- event copy number	*Le*1 copy number	Soybean GM- event copy number	*Le*1 copy number	Soybean GM- event (%)	*Le*1
MON87701	128	6914	128	6570	0.44	5.23
MON87708	124	6875	128	6570	-3.15	4.63
MON87769	161	6888	128	6570	26.24	4.84
305423	112	6677	128	6570	-12.19	1.62
BPS-CV127-9	106	6449	128	6570	-17.28	-1.85
A2704-12	43	6342	49	6570	-13.84	-3.48
MON89788	81	6410	92	6570	-12.14	-2.45
MON40-3-2	156	6626	128	6570	21.63	0.85
356043	150	6868	128	6570	17.37	4.53
A5547-127	126	7406	128	6570	-1.72	12.71
MON87705	135	6781	128	6570	5.86	3.20
DAS68416	109	6590	128	6570	-14.88	0.30
FG72	137	6415	128	6570	6.82	-2.37
DAS44406	135	6731	128	6570	5.73	2.44
DAS81419	117	6487	128	6570	-8.44	-1.27

Table S6. Bias between summed up assigned and experimental (duplex ddPCR verified) transgene and endogene copy numbers per reaction in DNA mix 1.

	6-plex			7-plex			11-plex		
	Experimental copy number	**Assigned copy number**	**Bias (%)**	**Experimental copy number**	**Assigned copy number**	**Bias (%)**	**Experimental copy number**	**Assigned copy number**	**Bias (%)**
GM-event	2527	2557	-1.19	2761	2612	5.69	4752	4658	2.01
*Le*1	27042	26282	2.89	26955	26282	2.56	26662	26282	1.45
GM-events (%)	9.34	9.73	-3.97	10.24	9.94	3.05	17.82	17.72	0.55

Table S7. Absolute limit of detection (LOD_{abs} **bold**), absolute limit of quantification (LOQ_{abs} underlined), and relative standard deviation (RSD) for the 6-plex assay.

DNA mix	Assigned copy number	Repeat 1		Repeat 2		Repeat 3		All		Bias to assigned copy number (%)
		Mean copy number	RSD (%)	Mean copy number	RSD (%)	Mean copy number	RSD (%)	Mean copy number	RSD (%)	
1	2557	2176	5.59	2252	4.49	2260	6.16	2229	5.41	-12.81
2	426	312	13.82	353	9.53	328	8.64	331	10.66	-22.34
3	213	185	7.86	194	3.23	189	9.20	189	6.76	-11.15
4	43	32	11.36	44	18.61	39	7.99	38	12.65	-10.23
5	21	18	32.91	20	31.26	16	30.59	18	31.58	-13.38
6	11	**6**	**23.00**	**10**	**28.40**	7	28.37	**8**	**26.59**	**-26.60**
7	4	5 [a]	ND[b]	6 [a]	ND[b]	**5**	**16.73**	5 [a]	ND[b]	ND[b]
8	2.1	Neg.	Neg.	Neg.	Neg.	6 [a]	ND[b]	5 [a]	ND[b]	ND[b]
9	0.4	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.
10	0.2	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.

[a] At least one replicate was negative.

[b] ND, not determined because of negative replicate(s).

Neg., all replicates were negative

Table S8. Absolute limit of detection (LOD_{abs} **bold**), absolute limit of quantification (LOQ_{abs} <u>underlined</u>), and relative standard deviation (RSD) for the 7-plex assay.

DNA mix	Assigned copy number	Repeat 1		Repeat 2		Repeat 3		All		Bias to assigned copy number (%)
		Mean copy number	RSD (%)	Mean copy number	RSD (%)	Mean copy number	RSD (%)	Mean copy number	RSD (%)	
1	2612	2525	1.79	2441	5.80	2418	9.10	2429	6.46	-6.99
2	435	384	2.19	392	15.21	390	4.97	391	9.39	-10.22
3	218	185	10.87	194	10.36	178	7.78	186	9.72	-14.56
4	44	<u>33</u>	<u>5.99</u>	<u>42</u>	<u>18.62</u>	41	9.16	<u>39</u>	<u>15.28</u>	<u>-10.64</u>
5	22	27	27.02	23	28.20	<u>22</u>	<u>20.91</u>	22	25.90	1.59
6	11	**10**	**11.96**	**11**	**27.79**	**9**	**40.48**	**10**	**30.20**	**-8.48**
7	4	5[a]	ND[b]	3[a]	ND[b]	3[a]	ND[b]	3[a]	ND[b]	ND[b]
8	2.2	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.
9	0.4	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.
10	0.2	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.

[a] At least one replicate was negative.

[b] ND, not determined because of negative replicate(s).

Neg., all replicates were negative

Table S9. Absolute limit of detection (LOD_{abs} **bold**), absolute limit of quantification (LOQ_{abs} underlined), and relative standard deviation (RSD)) for the 11-plex assay.

DNA mix	Assigned copy number	Repeat 1		Repeat 2		Repeat 3		All		Bias to assigned copy number (%)
		Mean copy number	RSD (%)	Mean copy number	RSD (%)	Mean copy number	RSD (%)	Mean copy number	RSD (%)	
1	4658	3943	3.06	3535	6.74	3422	12.00	3633	7.27	-22.00
2	776	583	8.66	574	5.08	564	7.49	574	7.08	-26.11
3	388	352	9.27	278	12.70	319	8.40	316	10.12	-18.54
4	78	66	11.62	78	6.15	61	3.50	68	7.09	-11.98
5	39	41	7.33	41	11.54	35	16.37	39	11.76	0.09
6	19	19	51.41	22	12.81	**14**	**45.85**	**18**	**36.69**	**-5.89**
7	8	9	27.89	**15**	**40.47**	9 [a]	ND[b]	12	34.18	59.57
8	3.9	**6**	**26.86**	8 [a]	ND[b]	Neg.	Neg.	7 [a]	ND[b]	ND[b]
9	0.8	9 [a]	ND[b]	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.
10	0.4	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.

[a] At least one replicate was negative.

[b] ND, not determined because of negative replicate(s).

Neg., all replicates were negative

Table S10. Absolute limit of detection (LOD_{abs} **bold**), absolute limit of quantification (LOQ_{abs} underlined), and relative standard deviation (RSD) for *Le*1 endogene.

DNA mix	Assigned copy number	Repeat 1		Repeat 2		Repeat 3		All		Bias to assigned copy number (%)
		Mean copy number	RSD (%)	Mean copy number	RSD (%)	Mean copy number	RSD (%)	Mean copy number	RSD (%)	
1	26788	23313	4.17	23207	6.26	22189	7.26	22903	2.71	-12.79
2	4465	3282	6.97	3510	6.26	3355	4.70	3382	3.44	-22.73
3	2232	1859	7.51	1874	6.24	1760	4.48	1831	3.37	-16.33
4	446	380	8.98	401	9.25	352	4.68	378	6.42	-13.69
5	223	202	13.16	201	9.84	180	9.90	194	6.33	-11.23
6	112	92	14.21	85	13.70	85	11.33	87	4.58	-20.25
7	45	40	21.56	34	19.00	35	12.33	36	8.65	-17.17
8	22.3	**16**	**29.68**	**21**	**23.59**	**14**	**32.32**	**17**	20.19	**-21.87**
9	4.5	6 [a]	ND [b]	3 [a]	ND [b]	5 [a]	ND [b]	5 [a]	ND [b]	ND [b]
10	2.23	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.	Neg.

[a] At least one replicate was negative.

[b] ND, not determined because of negative replicate(s).

Neg., all replicates were negative

Table S11. DNA mixes I to VII, assigned overall copy number and copy number for each multiplex.

DNA mix	Assigned overall[a]		Assigned 6-plex		Assigned 7-plex		Assigned 11-plex	
	GM-event (%)	Copy number	GM-event (%)	Copy number	GM-event (%)	Copy number	GM-event (%)	Copy number
I	2	1000	0.80	333	1.33	400	1.33	667
II	0.9	450	0.36	150	0.60	180	0.60	300
III	0.5	250	0.20	83	0.33	100	0.33	167
IV	0.1	50	0.04	17	0.07	20	0.07	33
V	0.05	25	0.02	8	0.03	10	0.03	17
VI	0.02	10	0.008	3	0.013	4	0.013	7
VII	0.005	2.5	0.002	1	0.003	1	0.003	2

[a] assigned number of copies for *Le*1 is 50000, for all dilutions

Table S12. Relative limit of detection (LOD_{rel} **bold**) and quantification (LOQ_{rel} underlined), for all three multiplex assays.

Assay	DNA mix	Assigned[a]		Experimental				Bias (%)	
		GM-event (%)	Copy number	Mean GM-event (%)	RSD (%) GM-event	Mean copy number	RSD (%) copy number	GM-event (%)	Copy number
6-plex	I	0.67	333	0.73	6.12	355	5.49	9.50	6.50
	II	0.30	150	0.36	21.07	172	5.85	19.67	14.67
	III	0.17	83	0.20	3.14	96	13.08	18.20	15.20
	IV	0.03	17	0.04	46.79	20	44.80	26.00	20.00
	V	**0.017**	**8**	**0.015**	**59.66**	**7**	**64.21**	**-10.00**	**-16.00**
	VI	0.007	3	0.001	173.21	1[b]	ND[c]	ND[c]	ND[c]
	VII	0.002	1	0.002	173.21	1[b]	ND[c]	ND[c]	ND[c]
7-plex	I	0.80	400	0.82	5.72	418	8.57	2.00	4.50
	II	0.36	180	0.42	5.47	212	6.42	17.50	17.78
	III	0.20	100	0.21	11.36	107	12.94	3.50	7.00
	IV	0.04	20	0.05	35.49	24	34.72	22.50	20.00
	V	**0.02**	**10**	**0.017**	**18.92**	**9**	**13.26**	**-15.00**	**-10.00**
	VI	0.008	4	0.007	108.75	3[b]	ND[c]	ND[c]	ND[c]
	VII	0.002	1	0.008	144.35	4[b]	ND[c]	ND[c]	ND[c]
11-plex	I	1.33	667	1.49	14.22	704	11.71	11.49	5.62
	II	0.60	300	0.68	12.97	333	16.83	13.02	11.12
	III	0.33	167	0.44	19.97	210	20.49	31.64	25.95
	IV	0.07	33	0.05	73.58	24	73.97	-20.53	-27.02
	V	0.03	17	0.02	38.62	10	41.11	-40.44	-42.60
	VI	**0.013**	**7**	**0.028**	**48.91**	**12**	**49.41**	**107.49**	**77.54**
	VII	0.003	2	0.015	102.44	7[b]	ND[c]	ND[c]	ND[c]

[a] assigned number of copies for *Le*1 is 50000, for all dilutions.

[b] At least one replicate was negative.

[c] ND, not determined due to negative replicate(s).

Table S13. Bias between assigned and experimental endogene copy numbers per reaction.

DNA mix	Assigned copy number	6-plex		7-plex		11-plex		Mean copy number	Bias (%)
		Experimental copy number	Bias (%)	Experimental copy number	Bias (%)	Experimental copy number	Bias (%)		
I	50000	48696	-2.61	51283	2.57	47392	-5.22	49124	-1.75
II	50000	47995	-4.01	50205	0.41	49159	-1.68	49120	-1.76
III	50000	48743	-2.51	51760	3.52	47936	-4.13	49480	-1.04
IV	50000	47685	-4.63	49743	-0.51	45457	-9.09	47628	-4.74
V	50000	48083	-3.83	53565	7.13	49146	-1.71	50265	0.53
VI	50000	47012	-5.98	49793	-0.41	44921	-10.16	47242	-5.52
VII	50000	48445	-3.11	51577	3.15	46036	-7.93	48686	-2.63

Table S14. Routine and USDA proficiency programme samples.

Sample	Code	Source/ use	Total determined (%)[a]	Targeted by 6-plex (%)	Targeted by 7-plex (%)
Real-life	G023/11	Flour	0.91	0.00	0.91
	G071/15	Soybean meal	94.20	2.82	91.34
	G158/15	Strawberry milkshake	<0.10	0.00	<0.10
	1/15	Chicken feed	0.94	0.00	0.94
	49/15	Soybean flour	0.48[b]	0.00	0.00
	75/15	Feed mix	91.40	15.87	75.52
	116/15	Feed mix	1.68	<0.10	1.68
	226/14	PU-starter premium	2.70	0.00	2.70
USDA	272/14	Flour	1.29	0.07	1.22
proficiency	274/14	Flour	2.26	0.74	1.52

[a] Determined by qPCR under ISO 17025 accreditation

[b] 0.48 (LLP), only DAS68416, so will only be detected by 11-plex

Table S15. Inter-laboratory comparison of combined results for 6-plex and 7-plex ddPCR assay for genetically modified soybean quantification.

Sample	Assigned GM-events (%) [a]	Sum GM-events (6-plex + 7-plex) (%)		Bias (%)	
		NIB	NVI	NIB	NVI
G023/11	0.91	1.03	1.06	14.13	17.52
G071/15	94.20	91.39	91.92	-2.98	-2.42
226/14	2.70	2.42	2.47	-10.37	-8.52
272/14	1.29	1.40	1.39	8.60	7.45
274/14	2.26	2.34	2.63	3.58	16.53
1/15	0.86	0.95	1.12	2.67	20.62
75/15	91.38	90.59	96.29	-0.86	5.37
116/15	1.68	1.73	1.91	2.70	13.53

[a] Assigned GM-event (%), value determined by NIB using event-specific qPCR or simplex ddPCR methods

Table S16. Inter-laboratory comparison of results for 6-plex ddPCR assay for genetically modified soybean quantification.

Sample	GM-event (%) [a]				RSD_r% for GM-event		Mean GM-event copy number		RSD_r (%)GM-event copy number		Mean *Le*1 copy number		RSD_r (%) *Le*1 copy number	
	Mean	RSD_R (%)	NIB	NVI	NIB	NVI	NIB	NVI	NIB	NVI	NIB	NVI	NIB	NVI
G023/11	NS[b]	NS[b]	NS[b]	NS[b]	NS[b]	NS[b]	NS[b]	NS[b]	NS[b]	NS[b]	41716	46392	7.23	5.96
G071/15	2.64	3.08	2.60	2.68	13.50	22.02	1633	1654	14.13	24.31	62870	60887	4.60	5.98
226/14	ND[c]	ND[c]	NS[b]	0.05 [d]	NS[b]	185.67 [d]	NS[b]	8 [e,d]	NS[b]	185.50 [d]	15927	15929	5.32	2.80
272/14	0.06		0.07	0.05	1.51	17.28	82	72	13.01	20.72	131216	141689	4.63	12.22
274/14	1.00	9.28	0.95	1.04	22.35	23.55	73	73	24.70	46.66	7693	8459	5.74	23.58
1/15	NS[b]	NS[b]	NS[b]	NS[b]	NS[b]	NS[b]	NS[b]	NS[b]	NS[b]	NS[b]	151429	135436	1.94	4.78
75/15	16.40	6.50	15.88	16.91	5.38	4.82	4373	4857	13.66	6.62	28930	28714	4.31	3.32
116/15	0.05	30.42	0.05	0.06	15.32	35.77	29 [e]	37 [e]	52.48 [f]	29.38	54445	54047	5.22	12.66

[a] Means observed at NIB and NVI (RSD_R, reproducibility relative standard deviation between NIB and NVI), (value observed at NIB/value observed at NVI)

[b] NS, negative sample

[c] ND, cannot be determined

[d] Six out of eight replicates negative

[e] Copy number lower than estimated limit of quantification

[f] Value not compliant with acceptance value for dynamic range (quantifiability)

Table S17. Inter-laboratory comparison of results for 7-plex ddPCR assay for genetically modified soybean quantification.

Sample	GM-event (%) [a]				RSD_r (%) for GM-event		Mean GM-event copy number		RSD_r (%) GM-event copy number		Mean *Le*1 copy number		RSD_r (%) *Le*1 cpopy number	
	Mean	RSD_R	NIB	NVI	NIB	NVI	NIB	NVI	NIB	NVI	NIB	NVI	NIB	NVI
G023/11	1.04	2.97	1.03	1.06	8.45	10.34	465	455	9.29	8.84	45050	35124	5.45	43.12 [b]
G071/15	89.02	0.51	88.79	89.24	5.22	2.38	58143	52292	4.96	5.61	65510	58593	2.57	33.66
226/14	2.39	-2.24	2.42	2.36	20.20	19.51	405	376	20.67	13.77	16825	15857	4.90	3.65
272/14	1.34	-0.25	1.34	1.33	7.57	7.86	1742	1700	4.86	13.63	131322	124938	10.83	14.37
274/14	1.49	14.72	1.39	1.59	13.95	32.84	104	135	15.33	41.78 [b]	7515	7126	11.62	49.28 [b]
1/15	1.04	17.49	0.95	1.12	5.33	6.71	1373	1410	10.36	10.59	144551	125880	12.39	10.73
75/15	77.05	6.25	74.71	79.38	17.68	4.28	22766	21228	6.84	11.92	30464	26923	6.10	12.17
116/15	1.76	9.97	1.68	1.84	9.13	14.56	916	935	10.58	20.27	54661	51585	5.78	14.55

[a] Means observed at NIB and NVI, (RSD_R = reproducibility relative standard deviation between NIB and NVI), (value observed at NIB/value observed at NVI)

[b] Value not compliant with acceptance value for dynamic range (quantifiability)

Table S18. Accuracy (trueness and precision) of the ddPCR assays.

Sample	Combined 6-plex and 11-plex			6-plex						11-plex					
	Assigned GM-event (%) [a]	Sum GM-event (%)	Bias (%)	GM-event (%)	RSD_r (%) for GM-line	Mean GM-event copy number	RSD_r (%) GM-event copy number	Mean *Le1* copy number	RSD_r (%) for *Le1* copy number	GM-event (%)	RSD_r (%) for GM-line	Mean GM-event copy number	RSD_r (%) GM-event copy number	Mean *Le1* copy number	RSD_r (%) for *Le1* copy number
G023/11	0.91	0.97	6.6	NS[b]	NS[b]	NS[b]	NS[b]	41716	7.2	0.97	6.2	410	8.66	42335	6.5
G071/15	94.2	90.23	-4.2	2.6	13.5	1633	14.13	62870	4.6	87.63	4.62	54517	3.83	62317	5.7
G158/15	<0.1	0.07	NA[c]	0.0	0.0	0.0	0.0	583852	28.5	0.07	18.94	286	15.47	448603	27.8
226/14	2.7	2.49	-7.8	0.0	0.0	0.0	0.0	15927	5.3	2.49	22.69	397	22.33	15994	4.4
272/14	1.29	1.31	1.6	0.07	1.51	82	13.01	131216	4.6	1.24	8.42	70	16.41	5706	16.2
274/14	2.26	2.89	27.9	0.95	22.35	73	24.7	7693	5.7	1.94	0.48	160	23.74	8226	6.6
1/15	0.94	1.03	9.6	0.0	0.0	0.0	0.0	151429	1.9	1.03	12.62	1472	12.92	143445	5.4
49/15	0.48	0.53	10.4	0.0	0.0	0.0	0.0	66306	13.2	0.53	0.21	342	35.31[d]	63939	14.0
75/15	91.38	91.88	0.5	15.88	5.38	4373	13.66	28930	4.3	76	4.43	17137	5.3	22548	2.4
116/15	1.68	1.92	14.3	0.05	15.32	29[e]	52.48[d]	54445	5.2	1.87	15.32	937	15.18	50202	3.5

[a] Assigned GM-event (%),determined by NIB using event-specific qPCR methods under ISO 17025 accreditation.

[b] NS, negative sample

[c] NA, not applicable

[d] Value not compliant with acceptance value for dynamic range (quantifiability)

[e] Copy number lower than estimated limit of quantification

Table S19. Robustness test: comparison of copy numbers.

		6-plex		7-plex		11-plex	
Protocol	**Dilution** / **Factor**	**60×** [a]	**240×** [b]	**60×**	**240×**	**60×**	**240×**
Original	Mean GM-event copy number/reaction	36[c]	10	34[c]	9[d]	67	12
	RSD (%) GM-event copy number/reaction	13.28	57.86[e]	9.44	31.06	13.53	24.73
	Mean *Le*1 copy number/reaction	352	71	346	75	329	76
	RSD (%) *Le*1 copy number/reaction	16.85	16.73	3.77	8.00	8.36	6.13
DG32	Mean GM-event copy number/reaction	39	6[d]	39	9[d]	80	18
	RSD (%) GM-event copy number/reaction	17.72	24.73	4.95	58.05[e]	42.39[e]	20.00
	Bias (%)to original GM-event copy number	7.69	-66.67	12.82	0.00	16.25	33.33[f]
	Mean *Le*1 copy number/reaction	351	76	391	83	350	93
	RSD (%) *Le*1 copy number/reaction	4.66	18.45	10.77	8.46	4.87	22.58
	Bias (%) to original *Le*1 copy number	-0.28	6.58	11.51	9.64	6.00	18.28
18 µL	Mean GM-event copy number/reaction	30[c]	10	34[c]	11	70	16
sample	RSD (%) GM-event copy number/reaction	8.64	50.50[e]	8.85	13.77	11.13	32.48
	Bias (%)to original GM-event copy number	-20.00	0.00	0.00	18.18	4.29	25.00
	Mean *Le*1 copy number/reaction	367	71	373	68	388	86
	RSD (%) *Le*1 copy number/reaction	10.78	13.34	4.09	20.78	8.22	14.49
	Bias (%) to original *Le*1 copy number	4.09	0.00	7.24	-10.29	15.21	11.63
Thermal	Mean GM-event copy number/reaction	42	10	40	8[d]	64	18
cycler	RSD (%) GM-event copy number/reaction	10.49	15.94	17.30	22.51	21.16	39.86[e]
	Bias (%)to original GM-event copy number	14.29	0.00	15.00	-12.50	-4.69	33.33[f]
	Mean *Le*1 copy number/reaction	391	81	372	84	362	78
	RSD (%) *Le*1 copy number/reaction	4.45	13.46	4.84	17.13	10.84	7.99
	Bias (%) to original *Le*1 copy number	9.97	12.35	6.99	10.71	9.12	2.56

[a] 60× dilution was chosen to challenge the LOQ.

[b] 240× dilution was chosen to challenge the LOD.

[c] Copy number lower than estimated limit of quantification.

[d] Copy number lower than estimated limit of detection.

[e] RSD higher than acceptance value (35%) for dynamic range (quantifiability).

[f] Bias higher than acceptance value (25%) for dynamic range (quantifiability).

Table S20. **Certified reference material used in this study.**

Event	Available at	Reference material ID	Certified GM content (mass/mass; %)
MON87705	AOCS[a]	AOCS 0210-A	> 99.4
MON89788	AOCS	AOCS 0906-B	100
MON40-3-2	JRC-IRMM[b]	ERM-BF410gk	10
356043	JRC-IRMM	ERM-BF425d	10
A5547-127	AOCS	AOCS 0707-C2	> 99.99
A2704-12	AOCS	AOCS 0707-B4	> 99.99
MON87708	AOCS	AOCS 0311-A	> 99.05
305423	JRC-IRMM	ERM-BF526d	10
BPS-CV127-9	AOCS	AOCS 0911-C	> 96.32
MON87769	AOCS	AOCS 0809-B	> 99.94
MON87701	AOCS	AOCS 0512-A	> 99.94
DAS-68416-4	JRC-IRMM	ERM-BF432d	10
FG72	AOCS	AOCS 0610-A2	> 99.99
DAS-44406-6	JRC-IRMM	ERM-BF436b	98.60
DAS-81419-2	JRC-IRMM	ERM-BF437b	98.60

[a]AOCS (American Oil Chemist Society) and [b] JRC-IRMM (Joint Research Centre, Institute for Reference Materials and Measurements) are providers of certified reference material.

Table S21. Primers and probes used in this study.

Multiplex	Event	Forward / reverse / probe (F/ R/P)	Sequence	Amplicon length (bp)	Final concentration in PCR (nM)
Endogene	*Le*1	F	5'-CCAGCTTCGCCGCTTCCTTC-3'	74	650
	*Le*1	R	5'-GAAGGCAAGCCCATCTGCAAGCC-3'		650
	*Le*1	P	5'-HEX-CTTCACCTTCTATGCCCCTGACAC-BHQ-3'		180
6-plex	305423	F	5'-CGTGTTCTCTTTTTGGCTAGC-3'	93	800
	305423	R	5'-GTGACCAATGAATACATAACACAAACTA-3'		500
	30542	P	5'-FAM-TGACACAAATGATTTTCATACAAAAGTCGAGA-BHQ-3'		220
	MON87708	F	5'-TCATACTCATTGCTGATCCATGTAG-3'	91	300
	MON87708	R	5'-AGAACAAATTAACGAAAAGACAGAACG-3'		300
	MON87708	P	5'-FAM-TCCCGGACTTTAGCTCAAAATGCATGTA-BHQ-3'		150
	BPS-CV127-9	F	5'-AACAGAAGTTTCCGTTGAGCTTTAAGAC-3'	88	400
	BPS-CV127-9	R	5'-CATTCGTAGCTCGGATCGTGTAC-3'		400
	BPS-CV127-9	P	5'-FAM-TTTGGGGAAGCTGTCCCATGCCC-BHQ-3'		100
	MON87769	F	5'-CATACTCATTGCTGATCCATGTAGATT-3'	87	600
	MON87769	R	5'-GCAAGTTGCTCGTGAAGTTTTG-3'		600
	MON87769	P	5'-FAM-CCCGGACATGAAGCCATTTACAATTGAC-BHQ-3′		600
	MON87701	F	5'-TGGTGATATGAAGATACATGCTTAGCAT-3'	89	600
	MON87701	R	5'-CGTTTCCCGCCTTCAGTTTAAA-3'		600
	MON87701	P	5'-FAM-TCAGTGTTTGACACACACACTAAGCGTGCC-BHQ-3′		250
7-plex	MON40-3-2	F	5'-TTCATTCAAAATAAGATCATACATACAGGTT-3'	84	600
	MON40-3-2	R	5'-GGCATTTGTAGGAGCCACCTT-3'		600
	MON40-3-2	P	5'-FAM-CCTTTTCCATTTGGG-BHQ-3'		200
	A2704-12	F	5'-GCAAAAAAGCGGTTAGCTCCT-3'	64	200
	A2704-12	R	5'-ATTCAGGCTGCGCAACTGTT-3'		200
	A2704-12	P	5'-FAM-CGGTCCTCCGATCGCCCTTCC-BHQ-3'		200
	A5547-127	F	5'-GCTATTTGGTGGCATTTTTCCA-3'	75	200
	A5547-127	R	5'-CACTGCGGCCAACTTACTTCT-3'		200
	A5547-127	P	5'-FAM-CCGCAATGTCATACCGTCATCGTTGT-BHQ-3'		200
	MON89788	F	5'-TCCCGCTCTAGCGCTTCAAT-3'	139	600
	MON89788	R	5'-TCGAGCAGGACCTGCAGAA-3'		600
	MON8978	P	5'-FAM-CTGAAGGCGGGAAACGACAATCTG-BHQ-3'		200

	356043	F	5'-GTCGAATAGGCTAGGTTTACGAAAAA-3'	99	650
	356043	R	5'-TTTGATATTCTTGGAGTAGACGAGAGTGT-3'		650
	35604	P	5'-FAM-CTCTAGAGATCCGTCAACATGGTGGAGCAC-BHQ-3'		180
	MON87705	F	5'-TTCCCGGACATGAAGCCATTTAC-3'	86	450
	MON87705	R	5'-ACAACGGTGCCTTGGCCCAAAG-3'		450
	MON87705	P	5'-FAM-AAGAGACTCAGGGTGTTGTTATCACTGCGG-BHQ-3'		250
11-plex	DAS68416-4	F	5′GTACATTAAAAACGTCCGCAATGTGT-3′	130	550
(addition	DAS68416-4	R	5′GTTTAAGAATTAGTTCTTACAGTTTATTGTTAG-3′		550
to 7-plex)	DAS68416-4	P	5′-FAM-TTAAGTTGTCTAAGCGTCAATA–MGBNFQ-3′		150
	FG72	F	5'-AGATTTGATCGGGCTGCAGG-3'	70	400
	FG72	R	5'-GCACGTATTGATGACCGCATTA-3'		400
	FG72	P	5'-FAM-AATGTGGTTCATCCGTCTT-MGBNFQ-3'		200
	DAS44406-6	F	5'-TTATTGTTCTTGTTGTTTCCTCTTTAGG–3′	99	300
	DAS44406-6	R	5'-CCTCAATTGCGAGCTTTCTAATTT-3′		300
	DAS44406-6	P	5′-FAM-ATTCGGACCTCCATGATGACCTTACCGTT-BHQ-3′		180
	DAS81419-2	F	5'-TCTAGCTATATTTAGCACTTGATATTCAT-3′	105	400
	DAS81419-2	R	5'-GCTTCAAGATCCCAACTTGCG-3′		400
	DAS81419-2	P	5′-FAM-ATCAACAGGCACCGATGCGCACCG-BHQ-3′		120

Table S23. Compliance with minimum performance parameters for methods acceptance according to ref. 26[a]

Additional data is presented in supplementary tables S22 and S23. Table S22 is a MIQE checklist, which is, as in Chapter 2.1, omitted here, since it represents an extensive additional text, not needed for the understanding of the study. Table S23 (title written above) is omitted for the same reason.

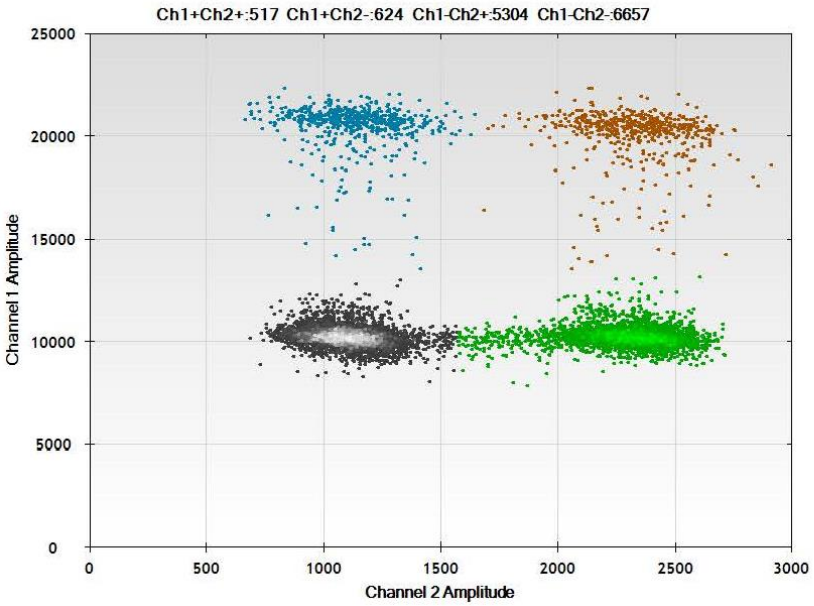


Figure S1. Example of positive experimental results. Fluorescence of droplets in a positive sample (6-plex assay). Droplets in black are negative, droplets in blue are positive for GM-events, droplets in green are positive for endogene (Le1), and those in brown are double positive (positive both for GM-events and Le1).

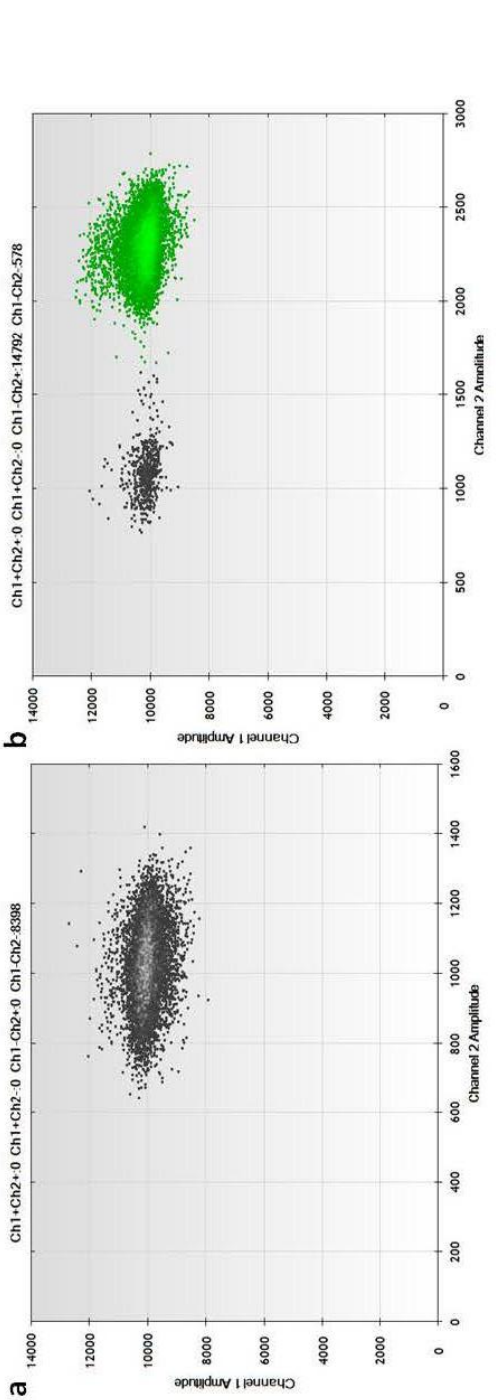


Figure S2. Example of negative experimental results. (a) fluorescence of droplets (black) in no-template control (NTC), (b) fluorescence of negative droplets (black) and only le1 positive droplets (GM negative sample).

A.3 Supplementary Material for Publication 2.3

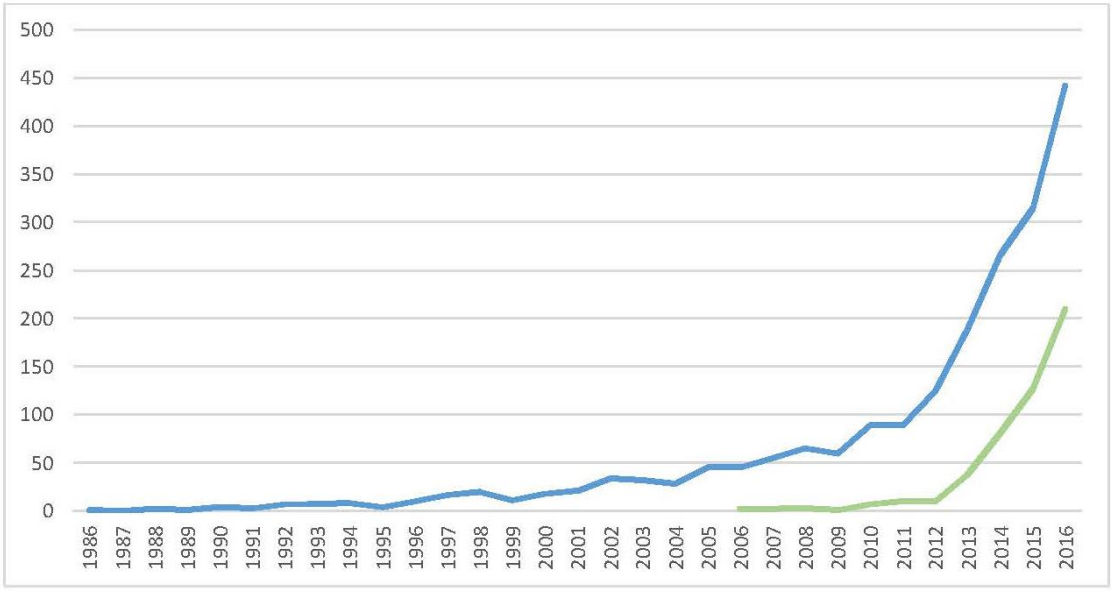


Fig S1 Papers reporting on digital PCR (blue) and droplet digital PCR (green) from 1986 to 2016. Data from Scopus database (accessed 1 February, 2017). In 2017 there were 44 papers listed for digital PCR in general, and of that 27 for droplet digital PCR, in particular. The number of publications in 2017 for both is likely to exceed the number in 2016. Search key word were digital PCR and digital PCR droplet, for dPCR and ddPCR respectively

	Aparatus	Supermix	Number cartridges/ wells per cartridge/ Droplets measured	
Experiment setup (1)	QX100™ Droplet Digital™ PCR System	ddPCR Supermix for probes (No dUTP)	3/6/3091	3/6/3567
Experiment setup (2)	QX200™ AutoDG™ Droplet Digital™ PCR System	ddPCR Supermix for probes (No dUTP)	3/6/3533	
Experiment setup (3)	QX100™ Droplet Digital™ PCR System	QX200 ddPCR EvaGreen Supermix	3/4/2140	

Fig S2 Description of the experimental set-ups. Operators from NIB and INRiM worked on set-ups (1), (2) and (3) at NIB and operators from INRiM worked on set-up (1) at INRiM. All of the droplets were imaged under an optical microscope and all of the images were recorded in bright field under uniform illumination and 100× apparent magnification

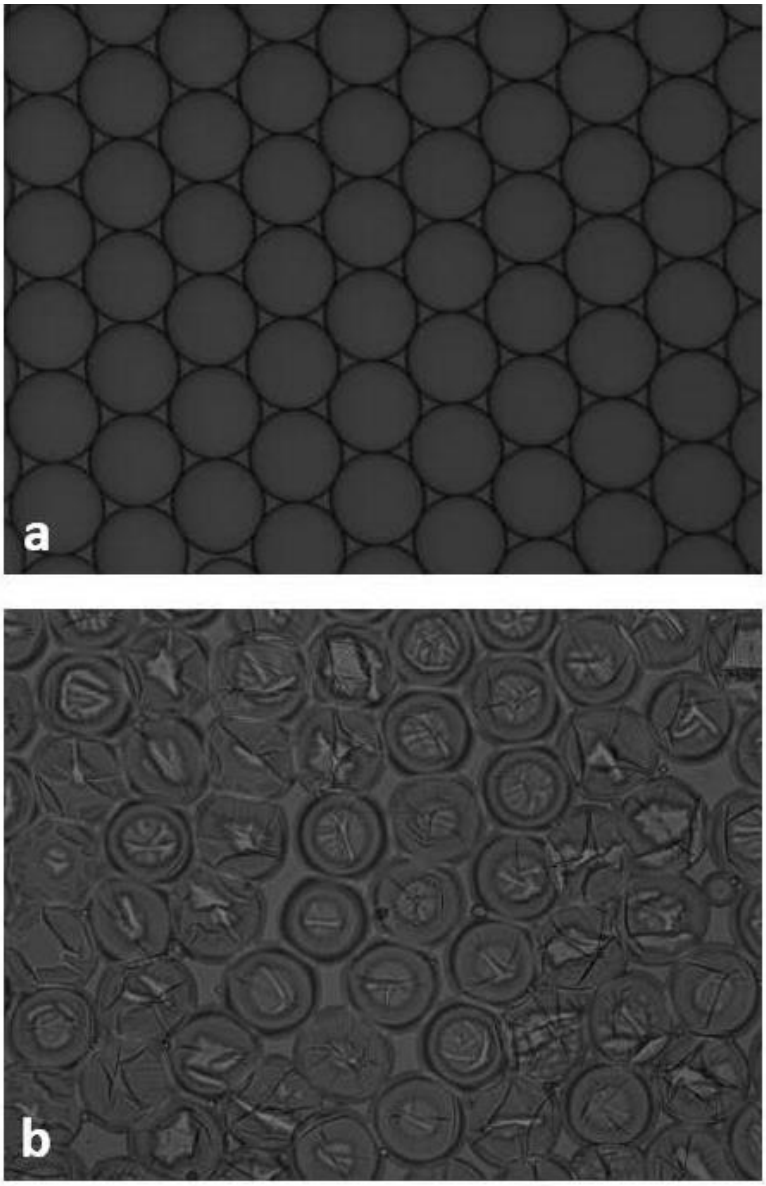


Fig S3 Comparison of droplets imaged immediately after generation (a) and after one hour (b). Droplets left on the plate for one hour were subjected to evaporation. Evaporation causes shrinking and wrinkling, thus these droplets are no longer suitable for imaging

Table S1 Assigned value of droplet diameter measure by optical profilometer. The assigned value was calculated as mean of results from the 3 different methods: based on autocorrelation, unit cell detection and length measurements of lines of packed droplets

Methods	Mean diameter [µm]	SD [µm]	N° of droplets	RSD %
Autocorrelation	109.68	1.83	303	1.7
Unit cell	109.55	2.36	303	2.2
Line	109.23	2.40	48	2.2
Reference Value	**109.48**	**2.20**		**2.0**

Table S2 Comparison of manual and automatic droplet measurement protocols with the assigned value gained by optical profilometer. The comparison of manual and automatic droplet measurement protocols with the reference method was done on 3 different sets of images

Method	Mean diameter [µm]	SD [µm]	N° of droplets	RSD %	Bias to reference %
Reference	109.48	2.20	303	2.0	
Manual	110.91	2.64	136	2.4	1.3
Automatic	110.52	1.77	6175	1.6	0.9

Table S3 Droplet volumes statistics measured with the manual and automatic protocol for the three datasets

Dataset	Manual measurements			Automatic measurements			Mean volume (nL)	Bias (%)
	Volume (nL)	RSD (%)	N° of droplets	Volume (nL)	RSD (%)	N° of droplets		
(1)	0.668	1.06	58	0.669	2.04	44	0.669	-0.15
(2)	0.709	1.54	46	0.723	1.38	41	0.716	-1.94
(3)	0.770	2.19	32	0.777	1.16	37	0.774	-0.90

Table S4 Validation of image analysis used for the automatic droplet measurements

Droplet parameter	Statistic	NIB	INRiM	Bias NIB/ INRiM (%)
Number	(n)	612	1028	
Volume	Mean volume (nL)	0.708	0.708	-0.02
	Standard deviation (nL)	0.025	0.022	
	RSD (%)	3.56	3.05	
Diameter	Mean area equivalent diameter (µm)	110.56	110.58	-0.01
	Standard deviation (µm)	1.39	1.17	
	RSD (%)	1.25	1.06	

Table S5 Mean droplet volumes of each individual well, for set-ups (1) and (2), including outliers (grey), for the ddPCR Supermix™ for probes (no dUTP)

Droplet generator	Set-up	Day 1		Day 2		Day 3	
		Mean droplet volume (nL)	No of analysed droplets	Mean droplet volume (nL)	No of analysed droplets	Mean droplet volume (nL)	No of analysed droplets
DG8	(1) NIB	0.694	229	0.699	133	0.744	155
		0.743	177	0.728	160	0.700	206
		0.711	232	0.733	121	0.762	151
		0.737	165	0.673	130	0.700	284
		0.698	215	0.676	169	0.731	182
		0.734	111	0.705	165	0.732	106
	(1) INRiM	0.654	231	0.734	231	0.682	141
		0.704	149	0.727	200	0.724	186
		0.716	269	0.746	194	0.725	141
		0.705	175	0.727	137	0.750	207
		0.710	244	0.721	140	0.732	220
		0.699	191	0.703	301	0.730	21
DG32	(2) NIB	0.733	117	0.720	236	0.710	241
		0.749	154	0.749	201	0.693	202
		0.747	213	0.753	185	0.745	157
		0.654	189	0.718	228	0.595	265
		0.766	162	0.770	208	0.758	231
		0.754	144	0.748	174	0.733	226

Table S6 Mean area equivalent diameters and droplet volumes ±standard deviation from the EvaGreen™ Supermix™ for the three repeats for the DG8 droplet generator at NIB. Mean droplet volume are calculated from all of the accepted droplets from all four measured wells

Measurement cartridge repeat	Mean area equivalent diameter ±SD (µm)	Mean droplet volume ±SD (nL)
1	114.01 ±1.15	0.777 ±0.02
2	114.51 ±1.12	0.786 ±0.02
3	113.86 ±1.22	0.773 ±0.02
Overall mean	**114.11 ±1.20**	**0.778 ±0.02**

Table S7 Repeatability and intermediate precision for copy numbers obtained with the DG8 droplet generator for both fixed and measured droplet volumes

Measurement cartridge repeat	Target	Fixed droplet volume as 0.85 nL		Measured droplet volume as 0.715 nL	
		Copy number ±SD	RSD_r (%)	Copy number ±SD	RSD_r (%)
1	A2704-12	113457 ±6597	5.81	134879 ±7843	5.81
	Le1	116242 ±6648	5.72	138190 ±7903	5.72
2	A2704-12	116479 ±4712	4.05	138471 ±5602	4.05
	Le1	121084 ±3938	3.25	143947 ±4682	3.25
3	A2704-12	116117 ±8731	7.52	138041 ±10379	7.52
	Le1	120364 ±7970	6.62	143090 ±9475	6.62
Overall mean	**A2704-12**	**115351 ±6609**	**5.73**	**137131 ±7856**	**5.73**
	Le1	**119230 ±6408**	**5.37**	**141742 ±7618**	**5.37**

Table S8 Repeatability and intermediate precision for copy number obtained with the DG32 droplet generator for both fixed and measured droplet volumes

Measurement cartridge repeat	Target	Fixed droplet volume as 0.85 nL		Measured droplet volume as 0.739 nL	
		Copy number ±SD	RSD$_r$ (%)	Copy number ±SD	RSD$_r$ (%)
1	A2704-12	110714 ±4676	4.22	127343 ±5379	4.22
	Le1	113916 ±3302	2.90	131026 ±3798	2.90
2	A2704-12	116811 ±3542	3.03	134357 ±4074	3.03
	Le1	120047 ±3405	2.84	138078 ±3917	2.84
3	A2704-12	114868 ±9250	8.05	132122 ±10639	8.05
	Le1	118144 ±9496	8.04	135889 ±10923	8.04
Overall mean	**A2704-12**	**114131 ±6491**	**5.69**	**131274 ±7466**	**5.69**
	Le1	**117369 ±6332**	**5.39**	**134998 ±7283**	**5.39**

Table S9 Relative reproducibility standard deviation (RSD_R) for fixed (0.85 nL) and measured (0.715 nL, 0.739 nL, for DG8, DG32, respectively) droplet volumes

Droplet generator	Measurement cartridge repeat	Target (copies/µL)			
		Fixes droplet volume at 0.85 nL		Measured droplet volumes at 0.715 (DG8) nL / 0.739 (DG32) nL	
		A2704	*Le1*	A2704	*Le1*
DG8	1	101011	104266	120083	123953
		111394	117838	132426	140087
		118441	120009	140804	142668
		118114	123980	140415	147389
		115682	115313	137524	137085
		116100	116046	138021	137957
	2	122316	121267	145410	144164
		110660	119063	131554	141544
		118403	119025	140758	141498
		111417	116338	132453	138304
		120149	127564	142835	151650
		115928	123249	137816	146519
	3	114696	119363	136352	141899
		125549	126380	149254	150242
		121861	125867	144869	149632
		100711	106608	119726	126737
		113907	116706	135414	138742
		119980	127259	142634	151288
DG32	1	110570	115338	127178	132662
		108081	114649	124315	131870
		104276	109147	119939	125541
		117429	117478	135067	135124
		109440	110607	125879	127221
		114487	116275	131683	133740
	2	112042	115756	128871	133143
		117026	123916	134603	142528
		122745	123826	141182	142426
		117832	120712	135530	138843
		116253	118979	133714	136850
		114971	117092	132239	134680
	3	96755	101671	111287	116943
		118469	119057	136263	136939
		119933	123335	137947	141861
		122033	129767	140363	149259
		117900	120333	135609	138407
		114120	114699	131261	131927
SD (copies)		**6486**	**6349**	**8116**	**8102**
RSD_R (%)		**5.65**	**5.37**	**6.05**	**5.86**

Table S10 Bias between copy numbers when using fixed (0.85 nL) and measured (0.715 nL, 0.739 nL, for DG8, DG32, respectively) droplet volumes

System	Target (copies per µL)			
	DG8		DG32	
	A2704-12	*Le1*	A2704-12	*Le1*
Fixes droplet volume at 0.85 nL	115351	119230	114131	117369
Measured droplet volumes at 0.715 (DG8)/ 0.739 (DG32) nL	137131	141742	131274	134998
Bias (%)	-15.88	-15.88	-13.06	-13.06

Methods S1: Reference method for droplet diameter measurement

A reference method for droplet diameter measurement was used to ensure traceability to the International System of Units (SI). The reference method used a combination of a calibrated measurement instrument for image acquisition and two validated image analysis protocols. The calibrated instruments include an optical profilometer (Plμ 2300™, Sensofar) equipped with a motorised stage for xy movements, a microscope with a 10× objective, and a CCD camera for image acquisition. The profilometer was calibrated with an INRiM two-dimensional standard grating referred to the SI units system. The profilometer is located in a temperature controlled laboratory at INRiM.

The image analysis protocols were validated within the EURAMET iMERA-Plus project "Traceable Characterisation of Nanoparticles" [1] for traceable calibration of nanoparticle sizes at National Metrology Institutes. Within the project, comparison measurements aimed at the validation of methods and uncertainty estimations across six National Metrology Institutes was carried out. All of the measurements were directly traceable to the SI unit 'metre'.

The reference method 'procedure' of the images acquisition by the profilometer and the image analysis was used for the droplet diameter measurements. As a reference method, this is traceable to the SI.

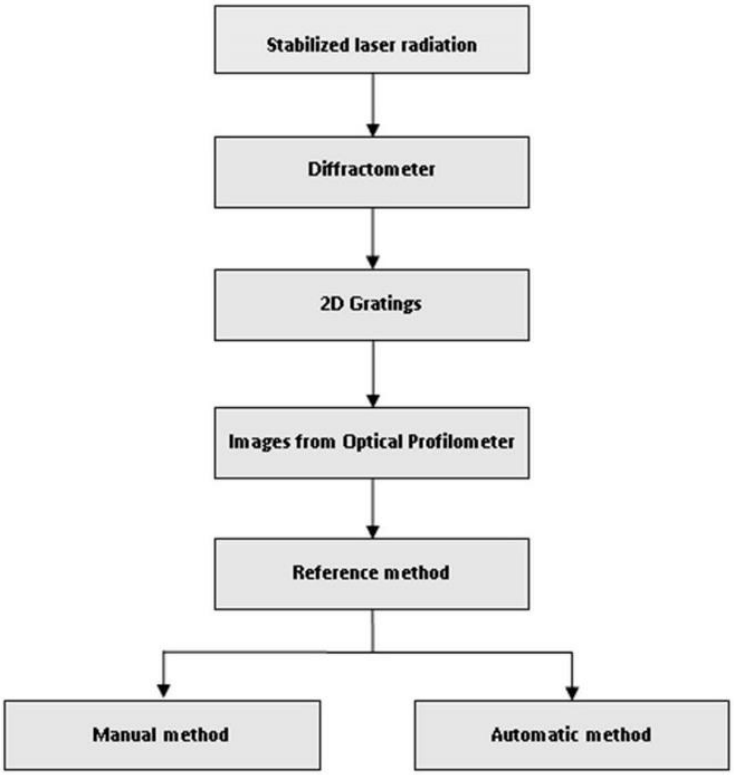
Reference method procedure description

Several images of packed droplets were acquired with an area of 1.27 × 0.95 mm (768 × 576 pixels) using the optical profilometer. The scanning probe image processor (Image Metrology) software was used for image elaboration, to determine the mean diameters of the droplets by lateral measurements using two different methods: (i) as based on autocorrelation and unit cell detection in the image; and (ii) as based on length measurements of lines of packed droplets. The mean droplet diameters were obtained from the scanning probe image processor elaboration for each acquired image. The mean droplet diameter measurement uncertainty was calculated according to [2].

The autocorrelation and unit cell detection functions were used to enhance periodic structures and to extract the correlation length parameters. The advantage of measuring the mean diameters of lines of packed droplets is that the uncertainty due to border determination is only in the first and last droplets of the line.

Validation of manual and automatic methods through the reference method

The use of the profilometer and scanning probe image processor software image elaboration is a primary reference method for length measurements with known uncertainties. To validate the droplet diameter obtained with the manual and automatic droplet measurement protocols, comparisons with the reference value of the mean droplet diameter obtained by the reference method were carried out. The results of these comparisons were the evaluation of the measurement accuracies of the manual and automatic droplet measurement protocols. The comparisons showed that the both protocols are traceable to the SI through the reference method. The following scheme (scheme M1) shows the traceability chain of the measurement of the diameter of droplets. The first element of the chain is a stabilized laser radiation which is a reference for length. The laser radiation in combination with a diffractometer (instrument able to analyse the structure of a material from the scattering pattern produced when a light radiation interacts with it) is used to measure distances within a 2D grating (a grid with nominal distance between regularly spaced elements). The 2D calibrated grating is used as reference for calibrating distances in images produced with the optical profilometer. The optical profilometer is finally used to measure droplet diameter as reference method.



Scheme M1 Traceability chain of droplet diameter measurement

Methods S2: Manual image analysis

Briefly, the Image J™ [3, 4] segmented line tool was used to measure the distance between the first and the last droplets of a line, which were then divided by the number of droplets to calculate the mean diameter. From the mean diameter, the volume was calculated after calibrating the pixel size using a certified micrometer stage. A detailed description is provided below.

The manual analysis of droplet images required: (i) measurement of the length scale of the microscope, by measurement of the pixel dimension with an X-Y ruler (or calibration grid) (Fig. M1); and (ii) measurement of the droplet diameter on lines of droplets that were touching each other (Fig. M2). These images were acquired at 10× magnification, and the analysis was carried out only on lines of droplets that were self-packed within each image. The X-Y ruler used by INRiM had nominal X and Y pitch scales of 25 µm, and additionally, it had been calibrated by the INRiM profilometer. The ruler calibration uncertainty was 0.15%.



Fig M1 XY ruler with a cross on the metric scale (left, circled in red) and a cross on the inch scale (right)

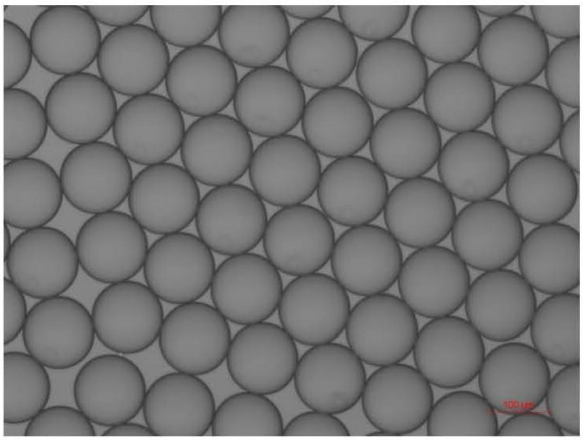


Fig M2 Lines of self-packed droplets

Pixel dimensions measurement

A calibrated XY micrometer (or ruler) was used for the pixel dimensions as follows:

1. Position the XY calibrated ruler on the microscope stage.
2. Set the microscope for phase contrast imaging and the same magnification that is to be used for acquiring the droplets.
3. Place the metric scale of the ruler in the centre of the camera view.
4. Adjust the z-axis to focus the XY ruler scale and set the exposure time to achieve the optimum optical contrast (Fig. M3, red square shows the minimum image portion to acquire).

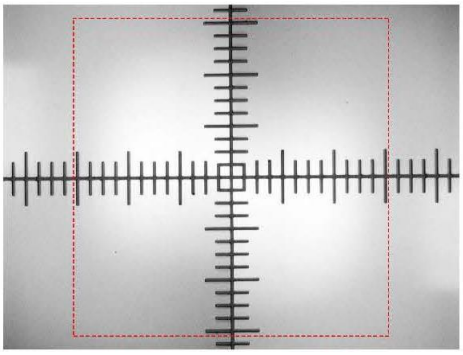


Fig M3 Image of the XY ruler portion for the acquisition

5. Acquire the image of the cross in the metric scale so that at least 600 µm × 600 µm (Fig. M3, red square) is visualised, and save this; the dimension of this square is required to have a statistically sufficient number of pixels (the two certified segments are less than the side of the red square).
6. With Image J™ (version 1.50a used here), open the cross image and select the tool 'Rectangular' on the software main window (Fig. M4).

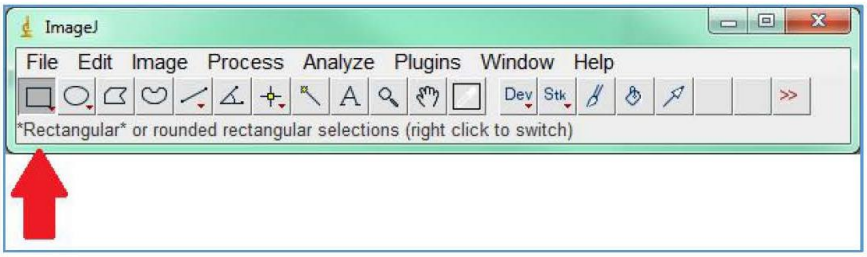


Fig M4 Image J 'Rectangular' tool

7. Use the 'Rectangular' tool to select the portion of the image, and measure the number of pixels on the x and y axes of that portion (pixel numbers of X axis and Y axis correspond to the two certified segments).
8. Draw a square as close as possible to the area that includes the certified segments (Fig. M5).

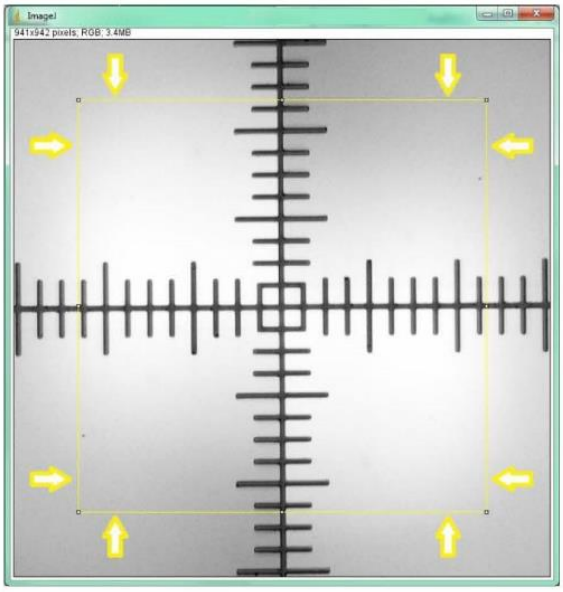


Fig M5 Square drawn around the portion of the XY ruler for the acquisition

9. After this coarse selection, refine the selection of the square using the '+' key of the keyboard; zoom-in on the image while keeping the mouse indicator on the centre of the view to zoom (Fig. M6).

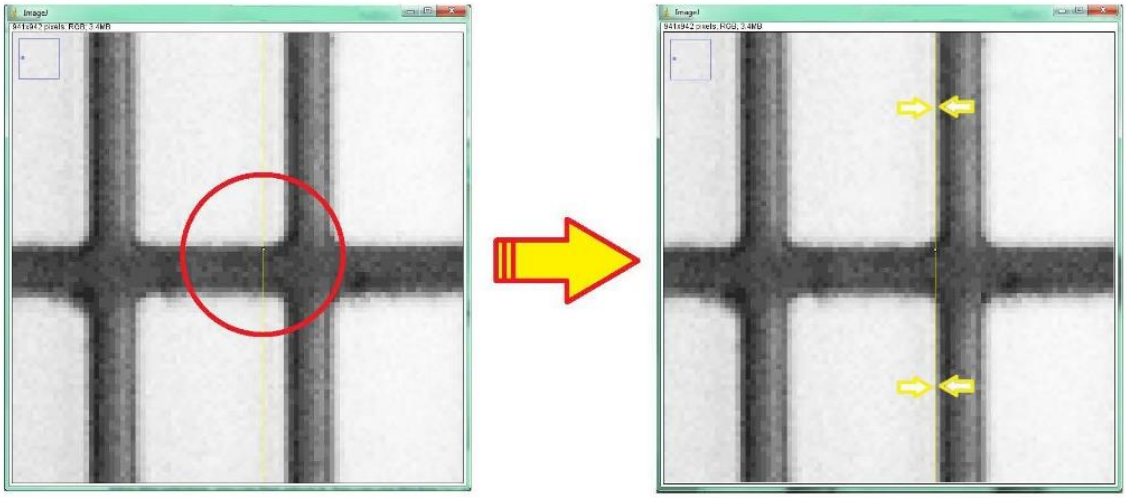


Fig M6 Left side of the cross metric scale (-225 μm)

10. To resize the yellow square, go on the little white square on the yellow line (Fig. M6, red circle; the mouse indicator will change from an arrow to a hand when it is possible to drag the line) and drag the line to the end of the black line (Fig. M6, on the left side of the nominal -225 μm).

11. Repeat point 9 for all of the other square sides. When the yellow square is resized, crop the portion of the image (Fig. M7).

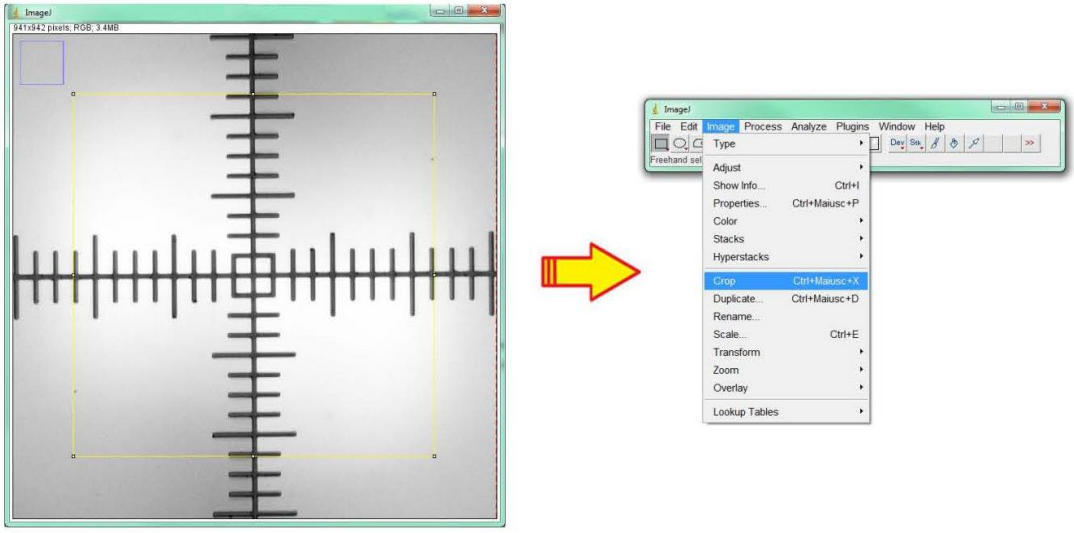


Fig M7 Crop the image

11. The number of pixels will appear at the top left of the image windows (Fig M8); the first value is the number of pixel for the horizontal axis (x), and the second for the vertical axis (y).

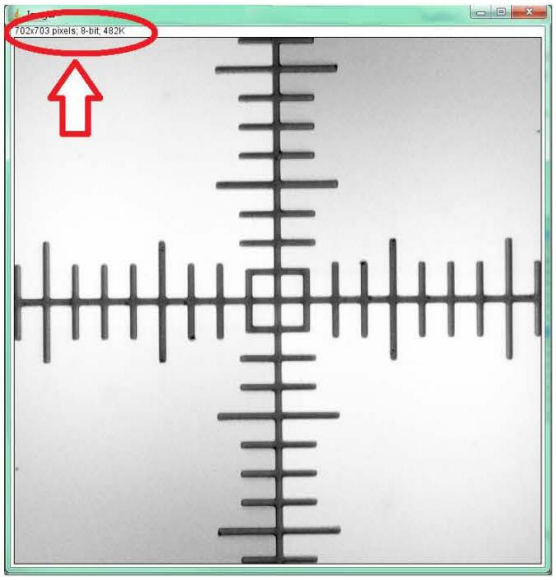


Fig M8 Number of pixels shown at the top left corner

12. Calculate the pixel dimension (*D*) in micrometres, for both the horizontal (x axis) and the vertical (y axis) scales using Equation (M1):

$$D = \frac{L}{N} \text{ (µm)} \qquad \text{(M1)},$$

where *N* is the number of pixels in the segment considered, and *L* is its length in micrometres (the nominal value of this length is -225 µm to +225 µm; considering the external edge of the scale lines, the total length (*L*) measured is x = 458 ±0.7 µm and y = 458 ±0.7 µm; Fig. M8).

13. The pixel dimensions (D) in micrometres for both the horizontal and vertical scales given were calculated as:

D = 458 µm/ 702 pixel = 0.652 µm/pixel → pixel width (x)

D = 458 µm/ 703 pixel = 0.651 µm/pixel → pixel height (y)

It is possible to assume that each pixel is a square with 0.65 µm per side, and to express the droplet diameters in micrometres (and not only in pixels).

Droplet diameter measurement

The manual droplet measurement protocol is described here.

1. Open ImageJ™ and an image showing lines of self-packed droplets (Fig. M9).

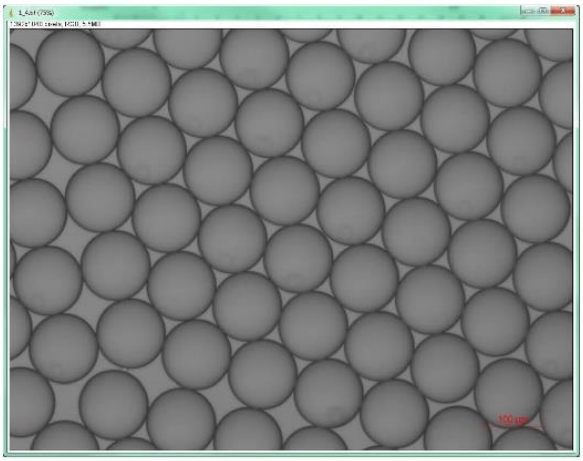


Fig M9 Droplets forming lines imaged at 10× magnification

2. Right click on the 'Straight line' tool (Fig. M10).

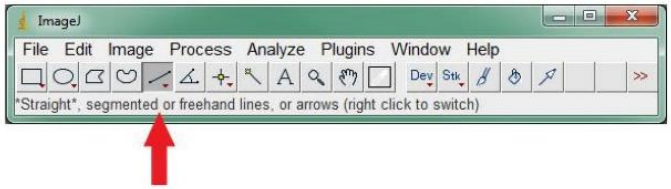


Fig M10 The 'Straight line' tool in Image J™

3. Left click to select the 'Segmented line' tool (Fig. M11). The tool icon will change. With this tool it is possible to draw a line on the droplets. Left click to set the 'beginning', and right click to set the 'end' of the line selection.

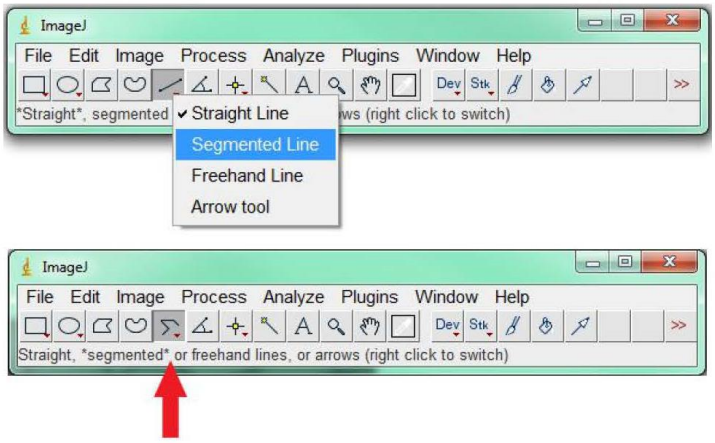


Fig M11 The 'Segmented line' tool in Image J™

4. It is recommended to use the zoom function ('+' and '–' of the keyboard) for accurate drawing of the segmented line. Zoom in at the edge of the first droplet (move the mouse cursor onto the portion of interest and press the '+' key). Left click on the edge to set the line 'beginning' (Fig. M12).

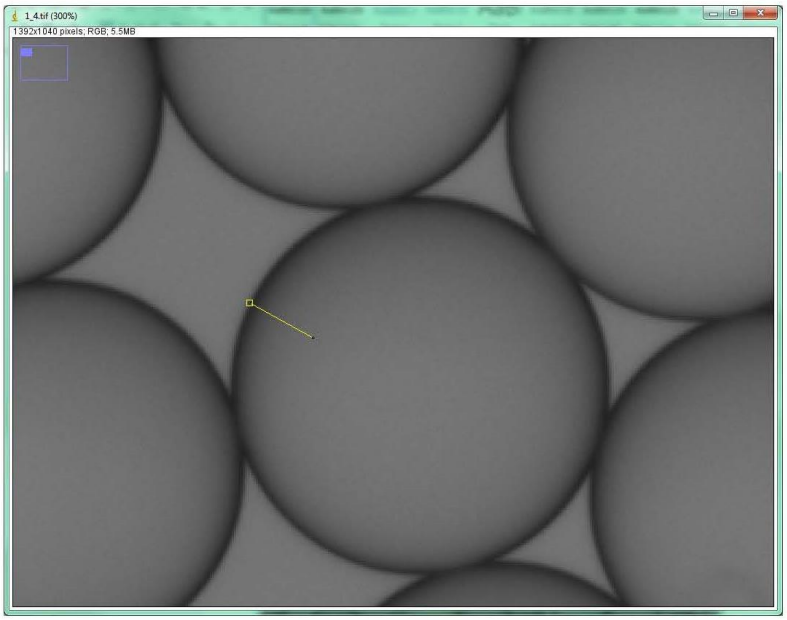


Fig M12 Setting the line beginning

5. Zoom out (using the '–' key) to see the entire zone again, and move the cursor to the external edge of the last droplet to be included in the selection (Fig. M13).

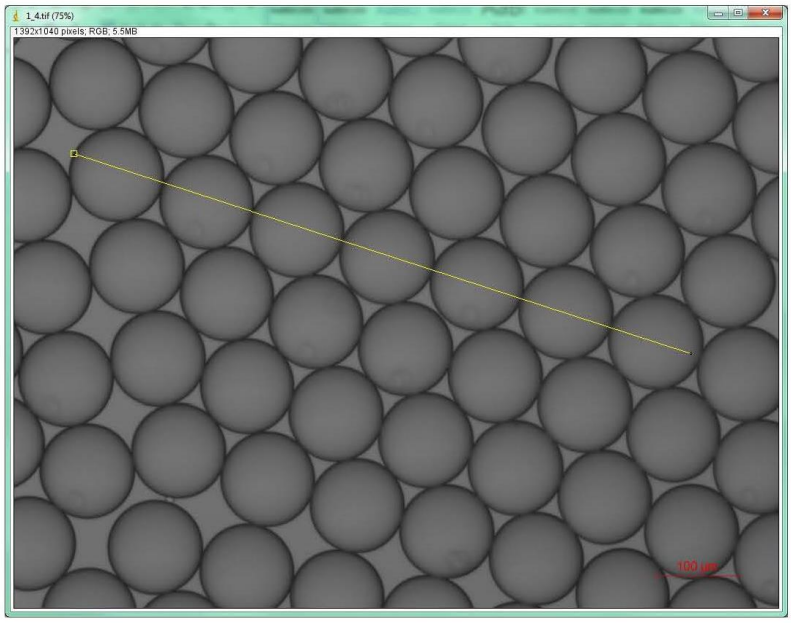


Fig M13 Selecting the last droplet

6. Zoom in around the external edge of the last droplet, and right click on the external edge to select the end of the line of the self-packed droplets (Fig. M14).

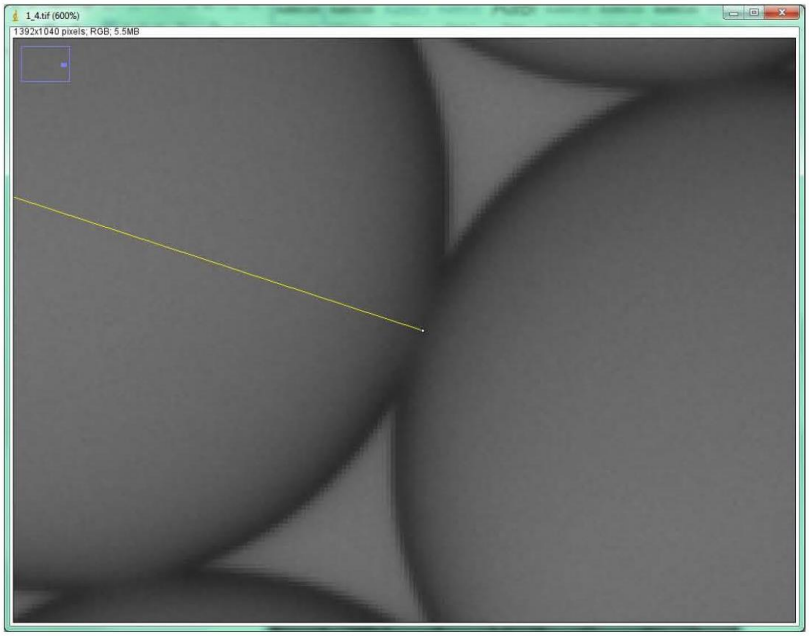


Fig M14 Setting the end of the line

7. Zoom out until the entire image can be seen, to check the line of the selected droplets (Fig. M15).

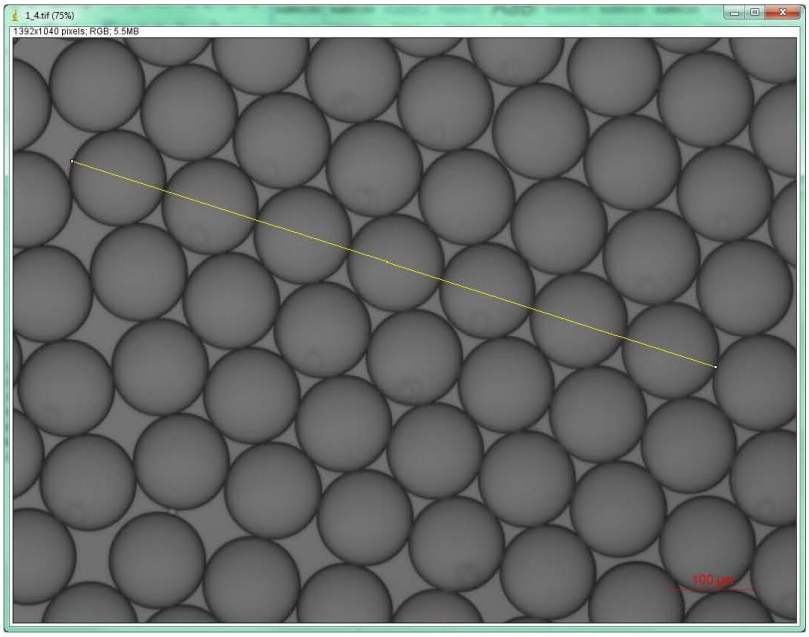


Fig M15 The selected line of droplets

8. On the main ImageJ™ window, click on 'Analyze → Plot Profile' (alternatively, use the keyboard shortcut of 'CTRL + K') (Fig. M16).

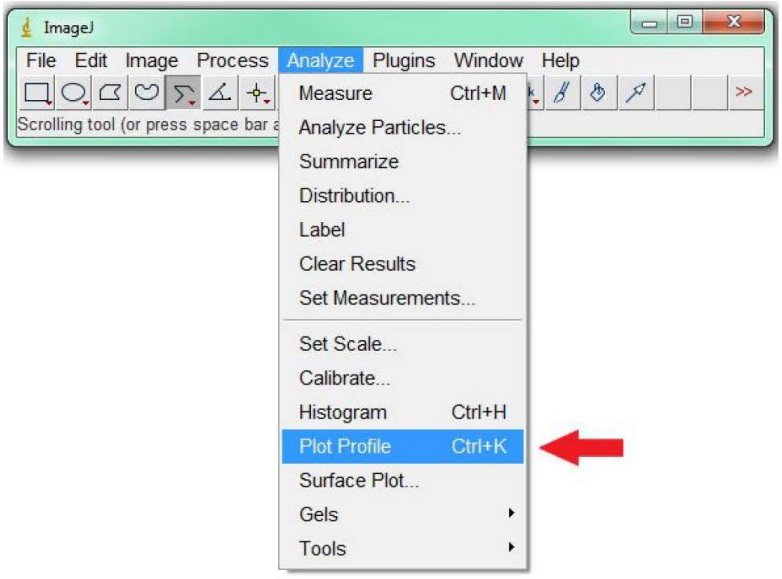


Fig M16 Selecting 'Plot Profile'

9. A new window will open with the results of the line selection: the number of peaks in the 'Plot' window represents the number of droplets in the selection (Fig. M17).

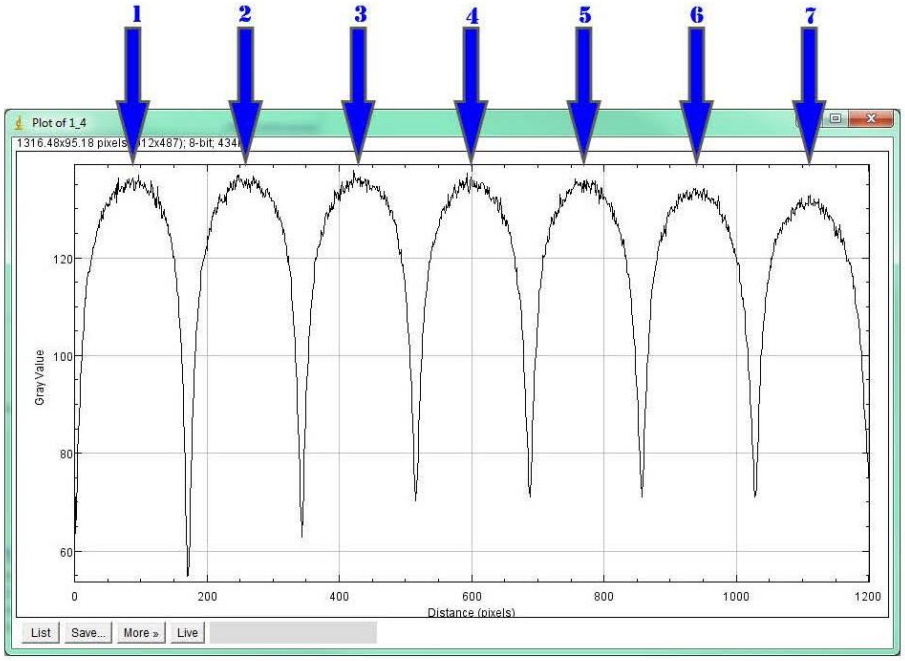


Fig M17 The 'Plot' window

10. Click on 'List' (bottom left) to see the list of Plot values (Fig. M18).

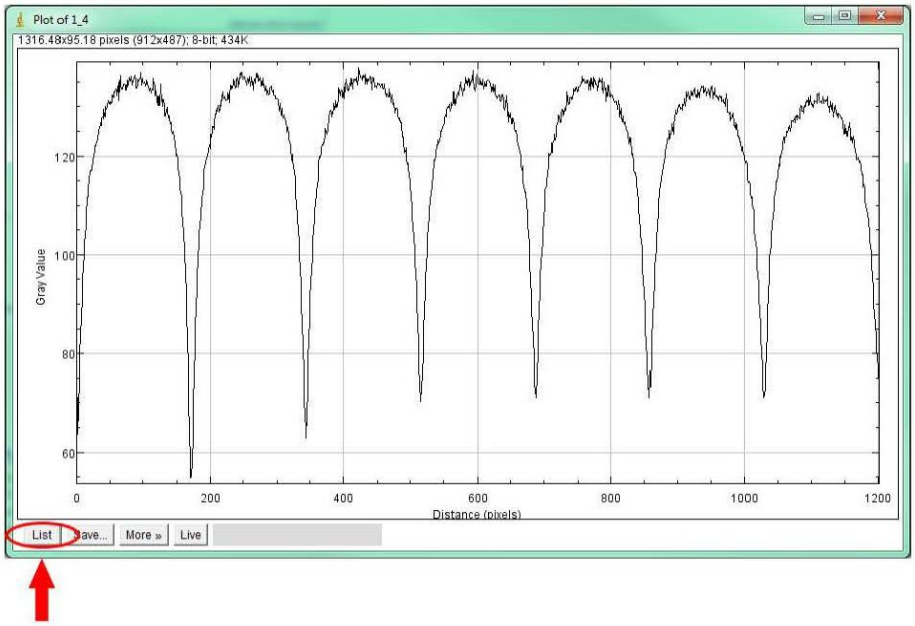


Fig M18 Click on 'List'

11. The last plot value of the list is the total length (in pixels) of the line of droplets selected (Fig. M19).

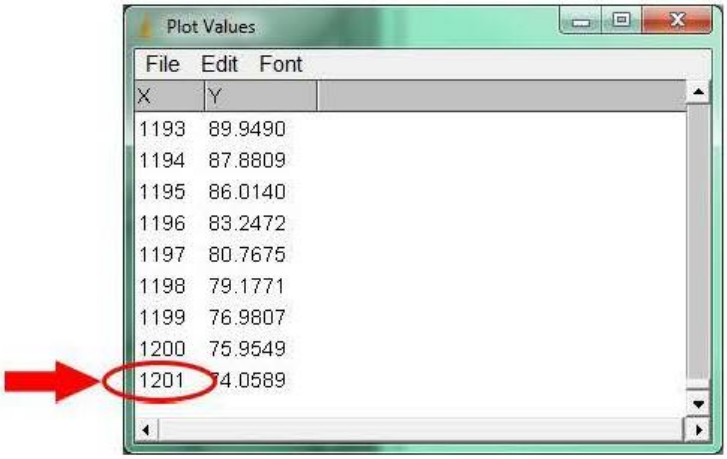


Fig M19 Total length of line of droplets selected

12. Calculate the mean droplet diameter using Equation M2:

$$Mean\ droplet\ diameter = \frac{Length\ of\ droplets\ line}{Number\ of\ droplets}\ [\text{pixels}] \qquad (M2).$$

Then convert the mean diameter in pixels to the mean diameter in µm, using the known pixel dimension.

Example:

Length of droplet line: 1201 pixels

Number of droplets: 7

Pixel dimension: 0.65 µm

$$Mean\ droplet\ diameter\ [pixels] = \frac{1201\ pixels}{7} = 171.6\ pixels$$

$$Mean\ droplet\ diameter\ [\mu m] = 171.6\ pixels \times 0.65\ {}^{\mu m}/_{pixel} = 111.54\ \mu m$$

13. Calculate the mean droplet volume using Equation M3

$$Mean\ droplet\ volume = \frac{4}{3}\pi r^3 \quad [\mu m^3] \qquad (M3).$$

Then convert the mean droplet volume in μm^3 to the mean droplet volume in nL.

Example:

Droplet radius: 111.54 μm/ 2 = 55.77 μm

$$Mean\ droplet\ volume\ [\mu m^3] = \frac{4}{3}\pi \times 55.77^3 = 726592$$

$Mean\ droplet\ volume\ [nL] = \frac{726592}{10^6}$ **= 0.727Methods S3: Automatic image analysis using the Fiji™ software**

For the automatic analysis, for all of the droplets in one image to be analysed at once, the next 14 steps were followed.

Analysis in ImageJ™

1. Open ImageJ™.
2. Open image (File → Open).
3. 'Set scale' (make sure you tick off 'Global'; Fig. M20).
4. Convert to 8-bit (Image → Type → 8-bit).
5. Identify the droplet edges (Process → Find edges).
6. Reduce the noise (Process → Noise → Despeckle).
7. Specify the width of the edges (Image → Adjust → Threshold).
8. Enable separation of the droplets (Process → Binary → Make binary).
9. Separate touching droplets (Process → Binary → Watershed).
10. Specify the width of the edges again (Process → Binary → Varonoi).
11. Identify the edges again (Image → Adjust → Threshold).
12. Measure (Analyse → Analyse Particles).
13. Set the parameters to only measure objects with circularity greater than 0.85 (Fig. M21), and hit OK.
14. A table of measurements, named 'Results', will open. Save this (File → Save as) for further Excel™ analysis. Also save the image after the analysis (Fig. M22), as this will help to determine which measurements to consider in the Excel™ analysis.

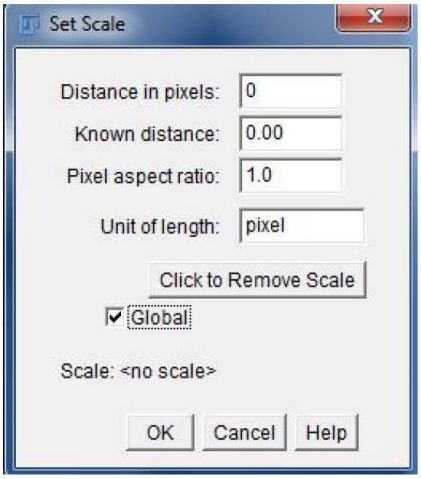


Fig M20 Setting the scale. Make sure to tick off 'Global'

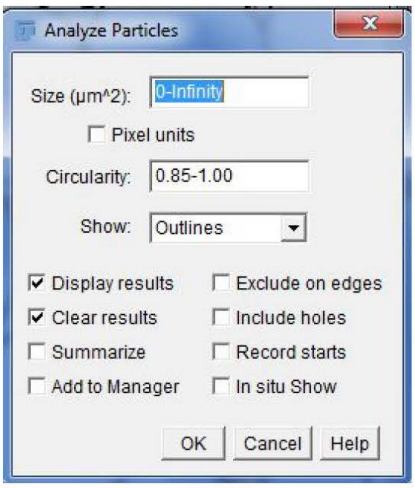


Fig M21 Measurement parameters

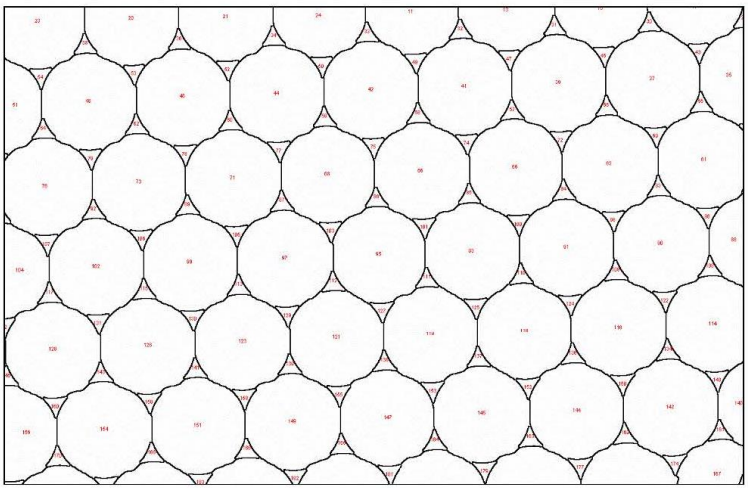


Fig M22 Droplets after the ImageJ™ analysis

Analysis in Excel™

The analysis in ImageJ™ was followed by analysis in Excel™. Figure M23 shows the output in Excel™ format. As there is a possibility that, even though the circularity has been set to higher than 0.85, not all measurements correspond to the droplets. To filter out unfit objects only those with roundness (round, column Q) above 0.95 were chosen. Even now all objects may not present the droplets, some objects with a very small area, may be bubbles left from the pipetting. To filter them out, only those measurements that have the area (column B), larger than 6000 (this correspond roughly to a diameter of 80 µm and the volume of 0.27 nL), were chosen. This threshold was chosen because there is an extremely low probability that an object of this size could represent a droplet, and even if this would be an actual droplet it would be excluded by the droplet reader.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	Area	Mean	Min	Max	Perim.	Major	Minor	Angle	Circ.	Feret	FeretX	FeretY	FeretAngle	MinFeret	AR	Round	Solidity
29	21443	0	0	0	562.985	165.270	165.197	97.243	0.85	172.003	1203	613	169.278	164.420	1.000	1.000	0.973
5	21275	0	0	0	560.357	164.701	164.469	69.623	0.851	171.234	1220	145	144.268	162.441	1.001	0.999	0.972
14	21341	0	0	0	556.257	165.374	164.307	49.407	0.867	171.304	1157	392	27.463	163.838	1.006	0.994	0.977
44	22091	0	0	0	564.742	168.303	167.122	92.625	0.87	173.589	352	978	28.189	166.583	1.007	0.993	0.976
7	21386	0	0	0	557.286	165.695	164.335	127.740	0.865	175.354	1047	262	27.880	163.000	1.008	0.992	0.976
17	21389	0	0	0	550.943	165.696	164.357	13.869	0.885	171.330	995	421	25.219	163.489	1.008	0.992	0.98
19	21783	0	0	0	563.470	167.185	165.894	87.112	0.862	173.727	719	354	113.762	166.039	1.008	0.992	0.973
26	21729	0	0	0	557.428	166.976	165.689	37.221	0.879	171.540	63	635	60.297	165.036	1.008	0.992	0.981
34	21807	0	0	0	566.056	167.409	165.855	86.719	0.855	174.072	799	807	88.354	165.052	1.009	0.991	0.975
37	23417	0	0	0	588.056	173.458	171.888	35.039	0.851	180.878	1141	775	171.736	171.141	1.009	0.991	0.973
40	22245	0	0	0	565.227	169.027	167.566	70.342	0.875	173.566	832	882	25.236	166.180	1.009	0.991	0.977
18	21136	0	0	0	543.772	164.861	163.235	88.223	0.898	169.000	836	440	22.620	162.311	1.010	0.99	0.984
27	21967	0	0	0	560.985	168.040	166.444	59.002	0.877	172.569	690	675	85.347	165.673	1.010	0.99	0.98
28	21493	0	0	0	557.227	166.302	164.554	82.396	0.87	171.759	498	545	115.521	163.575	1.011	0.989	0.979
32	20552	0	0	0	542.985	162.695	160.839	62.369	0.876	167.765	203	764	84.527	160.034	1.012	0.989	0.981
35	21956	0	0	0	560.399	168.099	166.302	83.433	0.879	173.528	604	670	115.974	165.721	1.011	0.989	0.979
30	21571	0	0	0	556.642	166.763	164.695	28.481	0.875	170.848	290	644	174.289	163.448	1.013	0.988	0.98
8	21543	0	0	0	560.742	166.716	164.527	116.056	0.861	172.325	203	156	111.801	164.304	1.013	0.987	0.975
45	21317	0	0	0	560.156	165.814	163.687	115.258	0.854	170.988	1097	930	171.254	163.475	1.013	0.987	0.975
23	21817	0	0	0	560.399	167.818	165.526	63.831	0.873	173.046	261	616	83.696	165.627	1.014	0.986	0.979
9	20942	0	0	0	542.701	164.532	162.060	60.623	0.894	170.488	891	282	24.611	162.000	1.015	0.985	0.983

Fig M23 Output of ImageJ™ in Excel™

Every object that was measured has its own number, which is in the first column of the Excel™ file (Fig. M23, column A). If a measurement appears to be an outlier, it is necessary to check if the object measured really corresponds to a droplet.

The object area was calculated using the major and minor axes of the best-fit ellipse according to equation M4. For this reason, only the data for area, major and minor axes, circularity and roundness are left. From this point on, the object is assumed to be a perfect sphere, thus radius and volume have been calculated using equations for a sphere, equations M5 and M6, respectively.

$$A = \frac{minor * major * \pi}{4} \quad \text{(M4)},$$

$$r = \sqrt{\frac{A}{\pi}} \quad \text{(M5)},$$

$$V = \frac{\frac{4}{3} * \pi * r^3}{10^6} \quad \text{(M6)}.$$

A.4 Supplementary Material for Publication 2.4

Table S1. Function and sequence of the primers used.

Oligo name	Oligo Sequence (5'-3')	Oligo function
p35S down_biotin	ATTGATGTGATATCTCCACTGACGT	Biotinylated primer used for linear enrichment
p35S up_biotin	CCTCTCCAAATGAAATGAACTTCCT	Biotinylated primer used for linear enrichment
tNOS down_biotin	GTCTTGCGATGATTATCATATAATTTCTG	Biotinylated primer used for linear enrichment
tNOS up_biotin	CGCTATATTTTGTTTTCTATCGCGT	Biotinylated primer used for linear enrichment
p35S down	AGGAAGTTCATTTCATTTGGAGAGG	Semi-nested primer p35S down
p35S up	GGTCTTGCGAAGGATAGTGGG	Semi-nested primer p35S up
tNOS down	AGATGGGTTTTTATGATTAGAG	Semi-nested primer tNOS down
tNOS up	CTCTAATCATAAAAACCCATCT	Semi-nested primer tNOS up
AAP	GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG	Universal primer complementary to poly dC tail

Table S2. Primer sequences and final concentrations used in the qPCR reaction used to determine size-dependent increase of specific target and loss of genomic background.

Target	F/ R/P**	Sequence	Final conc. in PCR [nM]
HMG*	F	TTGGACTAGAAATCTCGTGCTGA	400
	R	GCTACATAGGGAGCCTTGTCCT	400
	P	5'-FAM-CAATCCACACAAACGCACGCGTA-TAMRA-3'	200
cp4-epsps	F	GCCTCGTGTCGGAAAACCCT	400
	R	TTCGTATCGGAGAGTTCGATCTTC	400
	P	5'-FAM-TGCCACGATGATCGCCACGAGCTTCC-TAMRA-3'	200
I-rAct1	F	TCGTCAGGCTTAGATGTGCTAGA	400
	R	CTGCATTTGTCACAAATCATGAA	400
	P	5'-FAM-TTTGTGGGTAGAATTTGAATCCCTCAGC-TAMRA-3'	200
Cry3Bb1	F	CCGCCCAGGACTCCATCG	400
	R	GAGGCACCCGAGGACAGG	400
	P	5'-FAM-CTGCCGCCTGAGACCACTGACGAGC-TAMRA-3'	200
tNOS	F	GTCTTGCGATGATTATCATATAATTTCTG	400
	R	CGCTATATTTTGTTTTCTATCGCGT	400
	P	5'-FAM-AGATGGGTTTTTATGATTAGAGTCCCGCAA-TAMRA-3'	200
Ctp2-cp4epsps	F	GGGATGACGTTAATTGGCTCTG	375
	R	GGCTGCTTGCACCGTGAAG	375
	P	5'-FAM-CACGCCGTGGAAACAGAAGACATGACC-TAMRA-3'	150
P35S	F	ATTGATGTGATATCTCCACTGACGT	400
	R	CCTCTCCAAATGAAATGAACTTCCT	400
	P	5'-FAM-CCCACTATCCTTCGCAAGACCCTTCCT-TAMRA-3'	200

* maize endogenous gene

** forward / reverse / probe

Table S3. Tools used in building of the Galaxy workflow for UGMO detection.

Tool	Description and Tool Shed URL
Concatenate multiple datasets*	Combines multiple datasets tail-to-head, dataset collections can be merged to a single output file. (https://toolshed.g2.bx.psu.edu/repository?repository_id=12213e550a6c4a72&changeset_revision=1e06fa2771f9)
Convert characters[3–5]	All delimiters of the selected type are converted into TAB characters. (https://toolshed.g2.bx.psu.edu/repository/view_repository?changeset_revision=8a53d7f02ce4&id=84b1cc8e19e2e34c)
Sharplabtool* **Count** **Filter** **Reverse complement**	Tools count occurrences of each record, filter data on any column using simple expression and reverse complement of DNA/RNA sequences, respectively. (https://toolshed.g2.bx.psu.edu/repository?repository_id=ecc93bc8f0382e9e&changeset_revision=c2a356708570)
Cut columns[3–5]	Allows selection of specific columns from the input dataset. (https://toolshed.g2.bx.psu.edu/repository?repository_id=f11cf06cdcc13cfd&changeset_revision=842fab69c940)
Cutadapt[6] **1.8**	Finds and removes end sequences from high-throughput sequencing reads. (https://toolshed.g2.bx.psu.edu/view/iuc/package_cutadapt_1_8/980a47047f5 https://pypi.python.org/pypi/cutadapt/1.8)
FASTA-to-tabular[3–5]	FASTA formatted sequences are converted to TAB-delimited format. (https://toolshed.g2.bx.psu.edu/repository?repository_id=6fe870a3c7729070&changeset_revision=7e801ab2b70e)
FASTX-Toolkit* **FASTQ to FASTA**	Transforms FASTQ file format to a FASTA format. (https://toolshed.g2.bx.psu.edu/repository?repository_id=190c5b4338681335&changeset_revision=186b8d913e6c)
Filter sequences by ID*	Separates FASTA, FASTQ or SFF file on the basis of a specific ID present in a tabular file. (https://toolshed.g2.bx.psu.edu/repository?repository_id=4963efc937542d6d)
Bedtools[7]	A group of flexible tools for genome arithmetic and NGS analysis. (https://toolshed.g2.bx.psu.edu/repository?repository_id=1ec48b84b33d36d8&changeset_revision=41bba3e648d1)
Standalone NCBI BLAST+ tools[89]	Allows performing of BLAST on a local as well as public database. (https://toolshed.g2.bx.psu.edu/repository?repository_id=1d92ebdf7e8d466c&changeset_revision=7f3c448e119b)
Tabular-to-FASTA[3–5]	TAB-delimited format file is converted to FASTA formatted sequence. (https://toolshed.g2.bx.psu.edu/repository?repository_id=0b58502f66340f5d&changeset_revision=0b4e36026794)
Text processing **Text reformatting - awk** **Text transformation - sed**	Text processing tools using awk and sed programming language, respectively. (https://toolshed.g2.bx.psu.edu/repository?repository_id=2593fd36ae8011aa&changeset_revision=616efa22d193)
UPARSE[10]	Cluster OTUs - OTU clustering using the UPARSE-OTU algorithm. (https://toolshed.g2.bx.psu.edu/repository?repository_id=7edb1c97d9eb2336&changeset_revision=5d05b34a5fdf)
USEARCH[11]	Usearche_dereplication – eliminates replicate sequences. (https://toolshed.g2.bx.psu.edu/repository?repository_id=7666a94ede912d3f&changeset_revision=88fc52f1c5db USEARCH – search and clustering algorithms for sequence analysis. (https://toolshed.g2.bx.psu.edu/repository?repository_id=13df57e58dc7e075&changeset_revision=6f967c3a3f7b)

* Available in the Galaxy[3–5] ToolShed.

Table S4. BLAST output (first 5 hits) of the crCCS reads in the 'only unknown information bin'.

crCCS/crCCS length	Description	Query start-finish	Subject start-finish	Most specific subject annotation including the region of homology	Query cover [%]	E value	Ident [%]	Accession
120091/895	Zea mays subsp. mays genotype CMS-S mitochondrion, complete genome	11-895	92301-91416	1..557162/organism="Zea mays subsp. mays"/organelle="mitochondrion"/mol_type="genomic DNA"	98	0.0	99	DQ490951.2
	Zea mays subsp. parviglumis mitochondrion, complete genome	11-895	396277-397162	387242..442430/note="first copy of 55 kb R55"/rpt_type=inverted	98	0.0	99	DQ645539.1
	Zea mays subsp. mays genotype CMS-C mitochondrion, complete genome	11-895	134589-135474	125546..230101/note="first copy of 105 kb R105"/rpt_type=inverted	98	0.0	99	DQ645536.1
	Zea mays subsp. mays genotype CMS-T mitochondrion, complete genome	11-895	83514-84399	1..535825/organism="Zea mays subsp. mays"/organelle="mitochondrion"/mol_type="genomic DNA"	98	0.0	99	DQ490953.1
	Zea mays subsp. mays genotype male-fertile NA mitochondrion, complete genome	11-895	182577-181692	72199..191613/note="first copy of 120 kb R120"/rpt_type=direct	98	0.0	99	DQ490952.1
106520/971	Sequence 55213 from Patent WO2014036048	66-549	484-1	1..484/organism="Zea mays"/mol_type="unassigned DNA"	49	0.0	99	JC804838.1
	Sequence 58344 from Patent WO2014036048	436-715	436-1	1..653/organism="Zea mays"/mol_type="unassigned DNA"	45	6e-177	93	JC807969.1
	Sequence 11881 from Patent WO2014036048	164-354	228-39	1..577/organism="Zea mays"/mol_type="unassigned DNA"	19	2e-68	92	JC761506.1
	Sequence 52740 from patent US 7569389	870-960	177-267	1..689/organism="unknown"/mol_type="genomic DNA"	9	2e-28	93	GP689587.1
	Sequence 62396 from patent US 7569389	870-960	272-362	1..912/organism="unknown"/mol_type="genomic DNA"	9	4e-25	91	GP694415.1
43434/120	Compositions and methods for the therapy and diagnosis of ovarian cancer	27-99	233-161	1..256/organism="Homo sapiens"/mol_type="unassigned DNA"	60	7e-19	90	DL059073.1
	Compositions and methods for the therapy and diagnosis of ovarian cancer	27-99	5-77	1..246/organism="Homo sapiens"/mol_type="unassigned DNA"	60	7e-19	90	DL058560.1
	Sequence 9347 from Patent WO0192581	27-99	233-161	1..256/organism="Homo sapiens"/mol_type="unassigned DNA"	60	7e-19	90	CQ466569.1
	Sequence 8834 from Patent WO0192581	27-99	5-77	1..246/organism="Homo sapiens"/mol_type="unassigned DNA"	60	7e-19	90	CQ466056.1
	Compositions and methods for the therapy and diagnosis of ovarian cancer	38-99	56-117	1..298/organism="Homo sapiens"/mol_type="unassigned DNA"	51	3e-18	95	DL058704.1

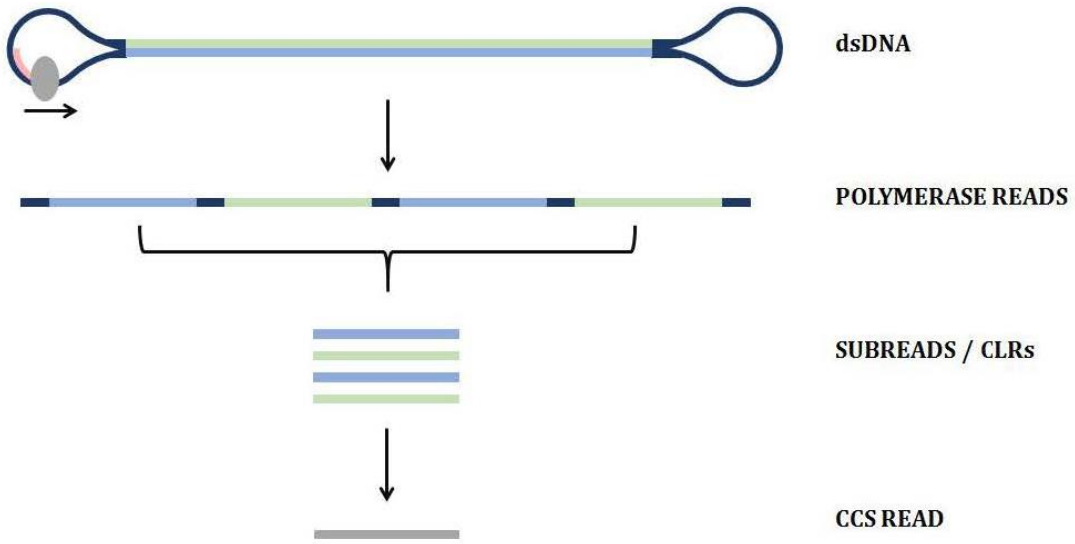


Figure S1. Formation of the circular consensus sequence (CCS) read. In PacBio sequencing a dsDNA molecule is sequenced. To do so, a single stranded loop is ligated to each side of the dsDNA (blue), a primer anneals to this loop (dark blue) and a strand displacement polymerase (grey) then passes the dsDNA. To gain a CCS reads the polymerase needs to pass the dsDNA template twice, the output is a so-called polymerase read, containing subreads interrupted by adapter sequences. The polymerase read is divided into subreads and the consensus of these subreads is a CCS read.

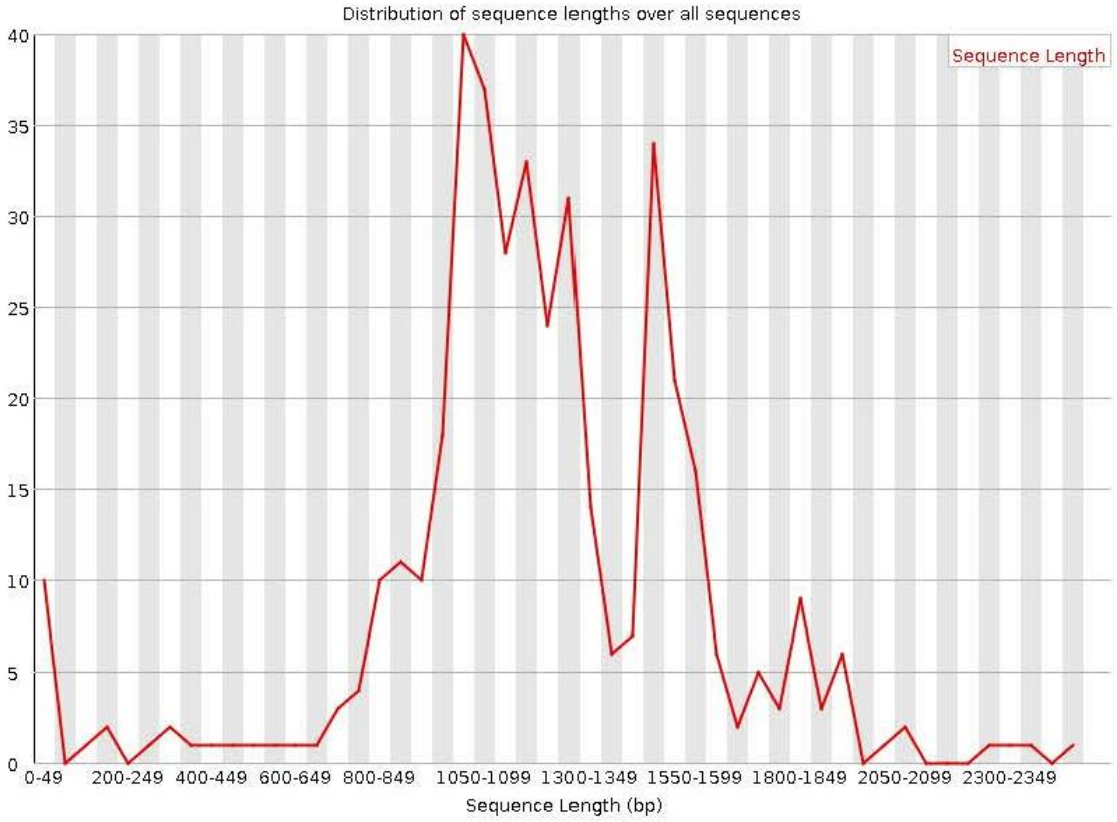


Figure S2. Length distribution of the 411 CCS reads as generated by FastQC.

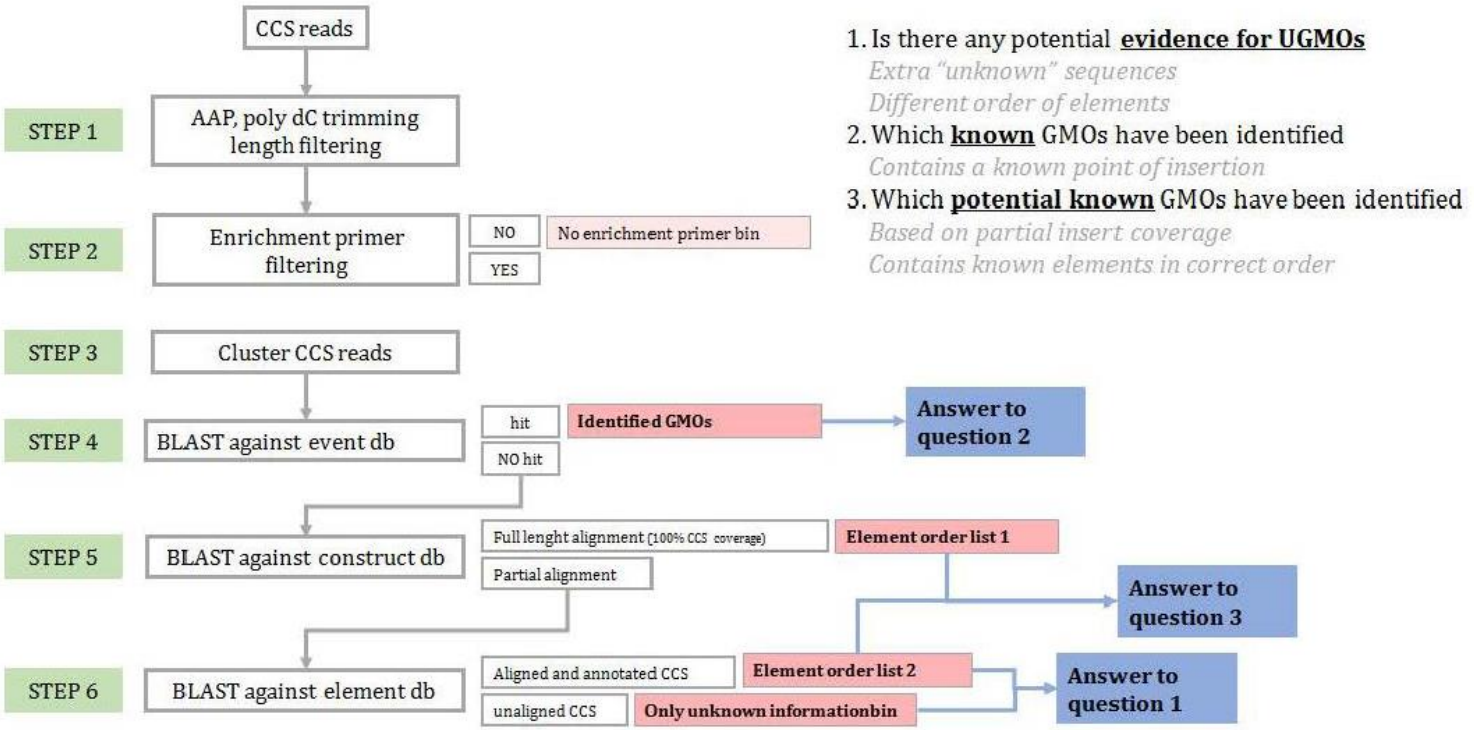


Figure S3. Building the workflow: in depth description. The workflow is divided in to two segments. First three steps prepare the reads for annotation in the following three steps. In linear enrichment (LE) process, a poly dC tail was added to 3' end of ssDNA acting as an annealing site for an abridged anchor primer or AAP, thus two-step trimming of poly dC tail and AAP sequence was performed with Cutadapt 1.8 in **Step 1.** First the AAP sequence was trimmed from 5' and a complement of AAP sequence from 3' prime end, followed by a removal of the poly dC and its complementary poly dG tail from 3' and 5' end,

respectively. For this Cutadapt 1.8 Galaxy wrapper was used, no parameters were changed for trimming of the AAP primers. To trim the poly dC and dG a string of 20 C/Gs was searched for at 3' and 5' end respectively, again no parameters of the trimming were changed. Input sequences were in FASTQ format. Because very short sequences are not informative, as the shortest transgenic element in our element database is 60 bp, a length filtering, using filter FASTQ v1.0.0, was also performed parallel to the second trimming in **Step 1**. In LE a semi-nested PCR (snPCR) was performed using downstream semi-nested primers and upstream semi-nested primers for both 35S promoter and NOS terminator. Therefor CCS reads had one of these sequences (or their complements) present either on 5' end (semi-nested primer) or on 3' end (complement of semi-nested primer). This enabled an enrichment primer filtering in **Step 2**, with the purpose to only keep the sequences enriched for, and minimise the contamination. Sequences not containing semi-nested primers were not discarded completely but rather put in a *no enrichment primer* bin, if they would be needed in final data assessment. Enrichment primer filtering was also done using Cutadapt 1.8 tool, parameters were set to minimum overlap length of 19 nucleotides, which is at least 75% of the primer length (primers are from 21 to 25 nt long). Output options were set to *do not trim adapters* and to show also the untrimmed reads, which in this case means the sequences that do not contain any of the primers. Later sequences were moved to *no enrichment primer* bin. For further steps sequences were first translated from FASTQ to FASTA format using FASTQ to FASTA tool. In **Step 3** the remaining sequences were clustered using cluster OTUs tool. Cluster OTUs uses UPARSE-OTU algorithm, a greedy algorithm where high-abundance reads are more likely to be correct amplicon sequences. Therefor the CCS reads were first dereplicated, with criteria for dereplication being set at prefix, and sorted by abundance. A minimum cluster radius was set at 0.97, corresponding to a minimum of 97% sequence identity. CCS reads representing the clusters were annotated in the next 3 steps. As the Cluster OTUs tool does not recognise +/- orientation of the sequences, a reverse complement of all sequences that have enrichment primers at the 5' end was made using the reverse complement tool, to further decrease the number of the clusters. In **Steps 4** cluster representing CCS (crCCS) reads were blasted against the event database, using blastn in BLAST+ tool. Descriptions of Blast hits were modified to include information on the number of times the sequence in the event database aligns to the CCS read, CCS read ID and the name of the event, and transferred to *confirmed events*. In **Step 5** the remaining CCS reads were blasted against CAF database. Hits in CAF database were transferred to *only known information*, with the additional rule that the complete CCS read needed to be fully covered by a database sequence. In the last step, **Step 6**, the remaining sequences were blasted against elements database, and hits in element database were named *unknown parts*, as for CCS reads that did not match any of the criteria, they were named *only unknown information* and needed to be manually further processed. As *Element order Data 1* and *2* were are not in a user-friendly form, they were further processed in an excel template.

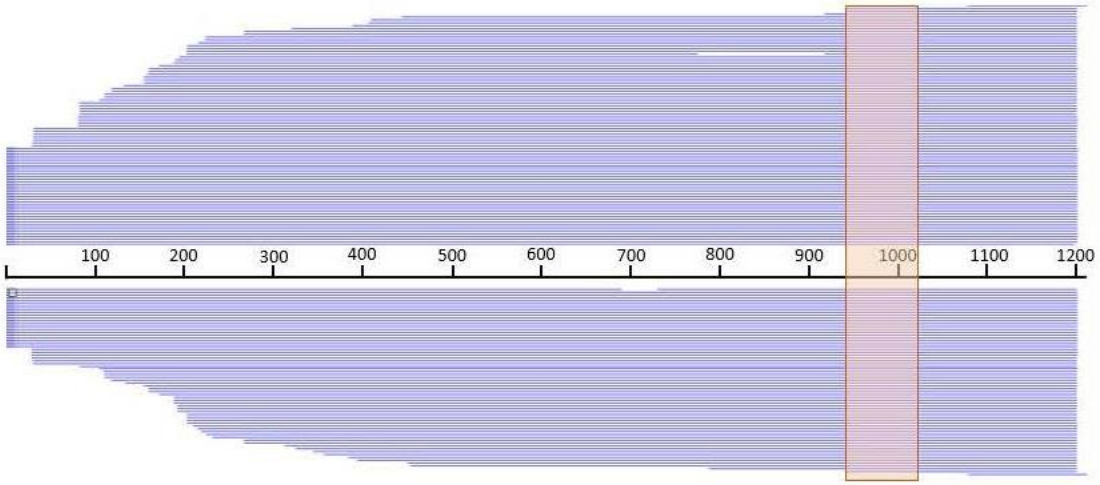


Figure S4. Bowtie2 mapping of CCS reads to RIKILT20151130 sequence. There is an overlap of 75 nucleotides when aligning sequences JQ406879 and AY326434. The overlap lies between nucleotides 943 and 1018 (orange) of the reference sequence RIKILT20151130. CCS reads were first filtered by quality, only those that had a quality score of 33 or more over at least 90% of the sequence were accepted. Using Bowtie2 mapping tool wrapper for Galaxy, with default settings, out of initial 411 sequences, 165 filtered sequences aligned to the reference, of those 133 CCS reads map to the overlap between sequences JQ406879 and AY326434 (orange).

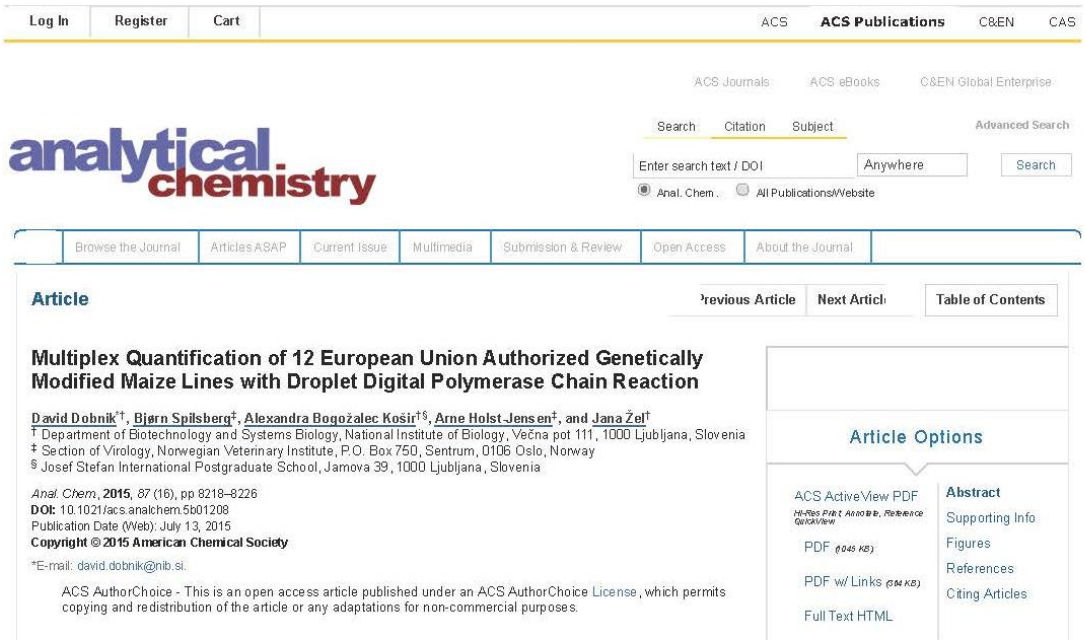
Data S1. Event database sequences.
Data S2. Reference sequences, CAF database.
Data S3. Annotations for reference sequences in CAF database.
Data S4. Element database - redundant, containing element sequences from all input GMOs.
Data S5. Element database , an un-redundant element database, containing only the longest of the element sequences.
Data S6. CCS reads with element orders tNOS – gap_- *Npt II* – gap_- p35S (133951, 45207), and p35S – gap_– tNOS – gap_– *uidA* (156962). p35S promoter is coloured yellow, tNOS terminator green *Npt II* gene red and *uidA* gene purple.

Sequences, and their annotations, populating all four databases used for building the pipeline, are originally part of the supplementary information. However, these sequences represent an extensive additional text; therefore, data S1-S6 are omitted here. Their titles are written above.

Appendix B

Permission for Reproduction of Included Publications

B.1 Permission for Reproduction of Publication 2.1



Log In | Register | Cart

ACS Publications
Most Trusted. Most Cited. Most Read.

Search

Authors & Reviewers | Librarians & Account Managers | ACS Members | Alerts | About Us | ACS &C

Standard *ACS AuthorsChoice/Editors' Choice* Usage Agreement

This ACS article is provided to You under the terms of this Standard *ACS AuthorChoice/Editors' Choice* usage agreement between You and the American Chemical Society ("ACS"), a federally-chartered nonprofit located at 1155 16th Street NW, Washington DC 20036. Your access and use of this ACS article means that you have accepted and agreed to the Terms and Conditions of this Agreement. ACS and You are collectively referred to in this Agreement as "the Parties").

1. SCOPE OF GRANT

ACS grants You non-exclusive and nontransferable permission to access and use this ACS article subject to the terms and conditions set forth in this Agreement.

2. PERMITTED USES

a. For non-commercial research and education purposes only, You may access, download, copy, display and redistribute articles as well as adapt, translate, text and data mine content contained in articles, subject to the following conditions:

i. The authors' moral right to the integrity of their work under the Berne Convention (Article 6bis) is not compromised.

ii. Where content in the article is identified as belonging to a third party, it is your responsibility to ensure that any reuse complies with copyright policies of the owner.

iii. Copyright notices or the display of unique Digital Object Identifiers (DOI's), ACS or journal logos, bibliographic (e.g. authors, journal, article title, volume, issue, page numbers) or other references to ACS journal titles, web links, and any other journal-specific "branding" or notices that are included in the article or that are provided by the ACS with instructions that such should accompany its display, should not be removed or tampered with in any way. The display of *ACS AuthorChoice* or *ACS Editors' Choice* articles on non-ACS websites must be accompanied by prominently displayed links to the definitive published versions of those articles on the ACS website.

iv. Any adaptations for non-commercial purposes must prominently link to the definitive published version on the ACS website and prominently display the statement: "This is an unofficial adaptation of an article that appeared in an ACS publication. ACS has not endorsed the content of this adaptation or the context of its use."

v. Any translations for non-commercial purposes, for which a prior translation agreement with ACS has not been established, must prominently link to the definitive published version on the ACS website and prominently display the statement: "This is an unofficial translation of an article that appeared in an ACS publication. ACS has not endorsed the content of this translation or the context of its use."

b. Each time You distribute this ACS article or an adaptation, ACS offers to the recipient a license to this ACS article on the same terms and conditions as the license granted to You under this License.

c. For permission to use ACS copyrighted articles beyond that permitted here, visit:
http://pubs.acs.org/copyright/permissions.html

3. PROHIBITED USES

a. Use of this ACS article for commercial purposes is prohibited. Examples of such prohibited commercial purposes include but are not limited to:

i. Copying or downloading of articles, or linking to such postings, for further distribution, sale or licensing, for a fee;

ii. Copying, downloading or posting by a site or service that incorporates advertising with such content;

iii. The inclusion or incorporation of article content in other works or services (other than normal quotations with an appropriate citation) that is then available for sale or licensing, for a fee;

iv. Use of articles or article content (other than normal quotations with appropriate citation) by a for-profit organizations for promotional purposes, whether for a fee or otherwise;

v. Sale of translated versions of the article that have not been authorized by license or other permission from the ACS

4. TERMINATION

ACS reserves the right to limit, suspend, or terminate your access to and use of the ACS Publications Division website and/or all ACS articles immediately upon detecting a breach of this License.

5. COPYRIGHTS; OTHER INTELLECTUAL PROPERTY RIGHTS

Except as otherwise specifically noted, ACS is the owner of all right, title and interest in the content of this ACS article, including, without limitations, graphs, charts, tables illustrations, and copyrightable supporting information. This ACS article is protected under the Copyright Laws of the United States Codified in Title 17 of the U.S. Code and subject to the Universal Copyright Convention and the Berne Copyright Convention. You agree not to remove or obscure copyright notices. You acknowledge that You have no claim to ownership of any part of this ACS article or other proprietary information accessed under this Agreement.

The names "American Chemical Society," "ACS" and the titles of the journals and other ACS products are trademarks of ACS.

6. DISCLAIMER OF WARRANTIES; LIMITATION OF LIABILITY

ACS warrants that it is entitled to grant this Agreement.

EXCEPT AS SET FORTH IN THE PRECEDING SENTENCE, ACS MAKES NO WARRANTY OR REPRESENTATION OF ANY KIND, EXPRESS OR IMPLIED, WITH RESPECT TO THIS ACS ARTICLE INCLUDING, BUT NOT LIMITED TO WARRANTIES AS TO THE ACCURACY OR COMPLETENESS OF THE ACS ARTICLE, ITS QUALITY, ORIGINALITY, SUITABILITY, SEARCHABILITY, OPERATION, PERFORMANCE, COMPLIANCE WITH ANY COMPUTATIONAL PROCESS, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

ACS SHALL NOT BE LIABLE FOR: EXEMPLARY, SPECIAL, INDIRECT, INCIDENTAL, CONSEQUENTIAL OR OTHER DAMAGES ARISING OUT OF OR IN CONNECTION WITH THE AGREEMENT GRANTED HEREUNDER, THE USE OR INABILITY TO USE ANY ACS PRODUCT, ACS'S PERFORMANCE UNDER THIS AGREEMENT, TERMINATION OF THIS AGREEMENT BY ACS OR THE LOSS OF DATA, BUSINESS OR GOODWILL EVEN IF ACS IS ADVISED OR AWARE OF THE POSSIBILITY OF SUCH DAMAGES. IN NO EVENT SHALL THE TOTAL AGGREGATE LIABILITY OF ACS OUT OF ANY BREACH OR TERMINATION OF THIS AGREEMENT EXCEED THE TOTAL AMOUNT PAID BY YOU TO ACS FOR ACCESS TO THIS ACS ARTICLE FOR THE CURRENT YEAR IN WHICH SUCH CLAIM, LOSS OR DAMAGE OCCURRED, WHETHER IN CONTRACT, TORT OR OTHERWISE, INCLUDING, WITHOUT LIMITATION, DUE TO NEGLIGENCE.

The foregoing limitations and exclusions of certain damages shall apply regardless of the success or effectiveness of other remedies. No claim may be made against ACS unless suit is filed within one (1) year after the event giving rise to the claim.

7. GENERAL

This Agreement sets forth the entire understanding of the Parties. The validity, construction and performance of this Agreement shall be governed by and construed in accordance with the laws of the District of Columbia, USA without reference to its conflicts of laws principles. You acknowledge that the delivery of the ACS article will occur in the District of Columbia, USA. You shall pay any taxes lawfully due from it, other than taxes on ACS's net income, arising out of your use of this ACS article and/or other rights granted under this Agreement. You may not assign or transfer its rights under this Agreement without the express written consent of ACS.

8. ACCEPTANCE

You warrant that You have read, understand, and accept the terms and conditions of this Agreement. ACS reserves the right to modify this Agreement at any time by posting the modified terms and conditions on the ACS Publications Web site. Any use of this ACS article after such posting shall constitute acceptance of the terms and conditions as modified.

Back to ACS Publishing Policies

Posted: 03/06/2014

B.2 Permission for Reproduction of Publication 2.2

nature.com

Licence to publish

Manuscript id:	SREP-16-47762A
Proposed title:	Development and inter-laboratory assessment of droplet digital PCR assays for multiplex quantification of 15 genetically modified soybean lines
The "Author(s)":	Alexandra Bogožalec Košir, Bjørn Splisberg, Arne Holst-Jensen, Jana Žel, David Dobnik
The "Article":	All content of the article, with the Proposed title listed above, including but not limited to all text, supplementary information, tables, graphs and images.
The "Journal":	Scientific Reports

Purpose

NPG will consider publishing the Contribution pursuant to the terms set forth herein, including granting readers rights to use the Contribution on an open access basis identified below.

Open access licence

CC BY*: This licence allows readers to copy, distribute and transmit the Contribution as long as it is attributed back to the author. Readers are permitted to alter, transform or build upon the Contribution, and to use the article for commercial purposes. Please read the full licence for further details at - http://creativecommons.org/licenses/by/4.0.

* The Creative Commons Attribution (CC BY) Licence is preferred by many research funding

bodies. We support use of this licence as it is recommended for maximum dissemination and use of open access materials.

Grant of rights

In consideration of NPG evaluating the Contribution for publication (and publishing the Contribution if NPG so decides) the Author(s) grant to NPG for the full term of copyright and any extensions thereto, except as set out in the Author rights below, the right and irrevocable licence:

a. to edit, adapt, publish, reproduce, distribute, display and store the Contribution in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world;
b. to translate the Contribution into other languages, create adaptations, summaries or extracts of the Contribution or other derivative works based on the Contribution and exercise all of the rights set forth (a) above in the translations, adaptations, summaries, extracts and derivative works;
c. to licence others to do any or all of the above, including but not limited to the right to grant readers the right to use the Contribution under the Creative Commons licence listed above; and
d. to re-licence article metadata without restriction (including but not limited to author name, title, abstract, citation, references, keywords and any additional information, as determined by NPG).

Author rights

Ownership of the copyright in the Contribution remains with the Author(s).

However, the Author(s)' re-use rights in the Contribution are subject to the rights and restrictions set forth in this Section, and in the Pre-print services and Warranties a) Section. After the Author(s) have submitted the Contribution to NPG hereunder, the Author(s)' rights to re-use the Contribution shall be the same as those set forth in the Creative Commons licence listed above, with the following additional re-use rights:

a. to reproduce the Contribution in whole or in part in any printed volume (book or thesis) of which they are the Author(s); and
b. to reuse figures or tables created by the Author(s) and contained in the Contribution in oral presentations and other works created by them.

Pre-print services

NPG acknowledges that an earlier version of the Contribution may have been submitted to a pre-print service (in accordance with that service's standard licence terms).

Warranties

The Author(s) warrant and represent that:

a. the Author(s) are the sole Author(s) of and sole owners of the copyright in the Contribution and the Contribution is the original work of the Author(s) and not copied (in whole or part) from another work. If however the

Contribution includes materials from other sources, the Author(s) warrant they have obtained the necessary rights from the owners of the copyright in all such materials and hereby license to NPG the rights to use such materials in accordance with the Grant of rights Section above. Copies of all such grant of rights from third parties have been submitted.

b. all of the facts contained in the Contribution are true and accurate;
c. the signatory (the Author or the employer) who has signed this Agreement below has full right, power and authority to enter into this Agreement and grant the rights herein on behalf of all of the Authors;
d. nothing in the Contribution is obscene, defamatory, libelous, violates any right of privacy or publicity, infringes any intellectual property rights (including without limitation copyright, patent, database or trademark rights) or any other human, personal or other rights of any person or entity or is otherwise unlawful; and
e. nothing in the Contribution infringes any duty of confidentiality which any of the Author(s) may owe to anyone else or violates any contract, express or implied, of any of the Author(s), and all of the institutions in which work recorded in the Contribution was created or carried out, have authorized such publication.

Dispute resolution

The Author(s) authorise NPG to take such steps as it considers necessary at its own expense in the Author(s) name and on their behalf if NPG believes that a third party is infringing or is likely to infringe copyright in the Contribution including but not limited to initiating legal proceedings.

The Author(s) grant NPG the perpetual right to edit, correct, or retract the Contribution if NPG considers (in its reasonable opinion) that such actions are required. The Author(s) hereby agree that they shall not object to NPG carrying out any such actions.

The Author(s) shall cooperate fully with NPG in relation to any legal action that might arise from the publication of the Contribution and the Author(s) shall give NPG access at reasonable times to any relevant accounts, documents and records within the power or control of the Author(s).

Reversion of rights

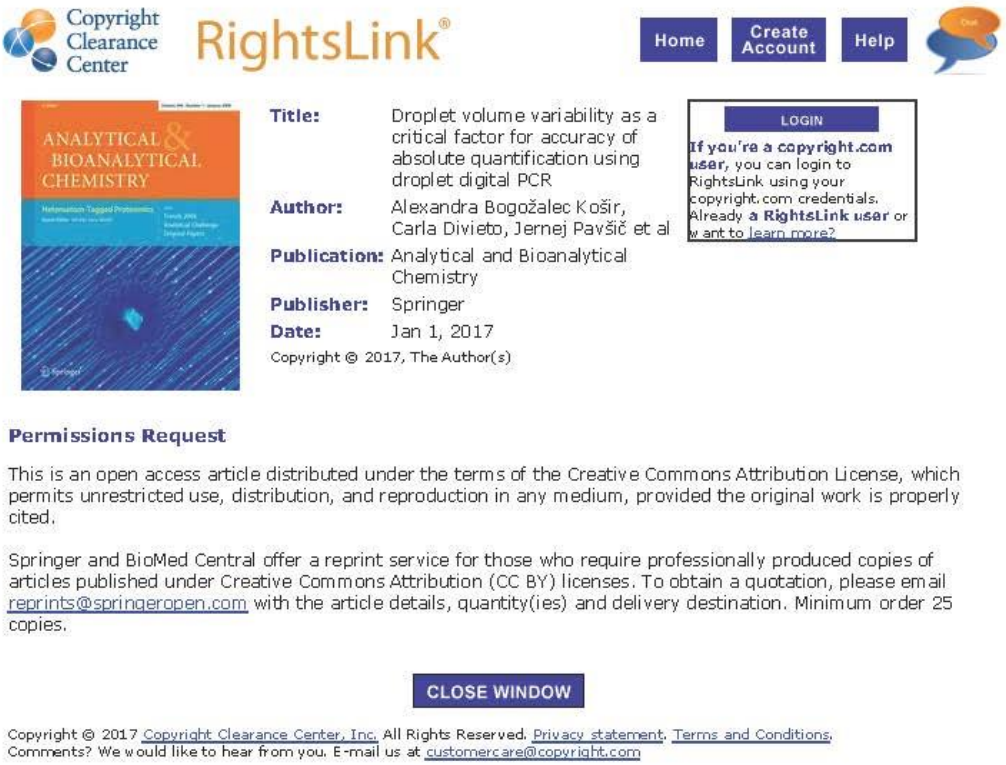
If the Contribution is rejected by NPG and not published, all rights under this licence shall revert to the Author(s).

Law & jurisdiction

This Agreement shall be governed by and construed in accordance with the laws of England and Wales. The parties irrevocably agree that the courts of England and Wales shall have exclusive jurisdiction to settle any dispute or claim that arises out of or in connection with this Agreement or its subject matter or formation.

Macmillan Publishers Limited (trading as Nature Publishing Group), registered office: Brunel Road, Houndmills, Basingstoke, Hampshire, RG21 6XS, UK. Company number 785998.

B.3 Permission for Reproduction of Publication 2.3



B.4 Permission for Reproduction of Publication 2.4

nature.com

Licence to publish

Manuscript id:	SREP-17-12867B
Proposed title:	ALF: a strategy for identification of unauthorized GMOs in complex mixtures by a GW-NGS method and dedicated bioinformatics analysis
The "Author(s)":	Alexandra Bogožalec Košir, Alfred Arulandhu, Marleen Voorhuijzen, Hongmei Xiao, Rico Hagelaar, Martijn Staats, Adalberto Costessi, Jana Žel, Esther Kok, Jeroen Peter van Dijk
The "Article":	All content of the article, with the Proposed title listed above, including but not limited to all text, supplementary information, tables, graphs and images.
The "Journal":	Scientific Reports

Purpose

NPG will consider publishing the Contribution pursuant to the terms set forth herein, including granting readers rights to use the Contribution on an open access basis identified below.

Open access licence

CC BY*: This licence allows readers to copy, distribute and transmit the Contribution as long as it is attributed back to the author. Readers are permitted to alter, transform or build upon the Contribution, and to use the article for commercial purposes. Please read the full licence for further details at - http://creativecommons.org/licenses/by/4.0.

* The Creative Commons Attribution (CC BY)

Licence is preferred by many research funding bodies. We support use of this licence as it is recommended for maximum dissemination and use of open access materials.

Grant of rights

In consideration of NPG evaluating the Contribution for publication (and publishing the Contribution if NPG so decides) the Author(s) grant to NPG for the full term of copyright and any extensions thereto, except as set out in the Author rights below, the right and irrevocable licence:

a. to edit, adapt, publish, reproduce, distribute, display and store the Contribution in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world;
b. to translate the Contribution into other languages, create adaptations, summaries or extracts of the Contribution or other derivative works based on the Contribution and exercise all of the rights set forth (a) above in the translations, adaptations, summaries, extracts and derivative works;
c. to licence others to do any or all of the above, including but not limited to the right to grant readers the right to use the Contribution under the Creative Commons licence listed above; and
d. to re-licence article metadata without restriction (including but not limited to author name, title, abstract, citation, references, keywords and any additional information, as determined by NPG).

Author rights

Ownership of the copyright in the Contribution remains with the Author(s).

However, the Author(s)' re-use rights in the Contribution are subject to the rights and restrictions set forth in this Section, and in the Pre-print services and Warranties a) Section. After the Author(s) have submitted the Contribution to NPG hereunder, the Author(s)' rights to re-use the Contribution shall be the same as those set forth in the Creative Commons licence listed above, with the following additional re-use rights:

a. to reproduce the Contribution in whole or in part in any printed volume (book or thesis) of which they are the Author(s); and
b. to reuse figures or tables created by the Author(s) and contained in the Contribution in oral presentations and other works created by them.

Pre-print services

NPG acknowledges that an earlier version of the Contribution may have been submitted to a pre-print service (in accordance with that service's standard licence terms).

Warranties

The Author(s) warrant and represent that:

a. the Author(s) are the sole Author(s) of and sole owners of the copyright in the Contribution and the Contribution is the original work of the Author(s) and not copied (in whole or part)

from another work. If however the Contribution includes materials from other sources, the Author(s) warrant they have obtained the necessary rights from the owners of the copyright in all such materials and hereby license to NPG the rights to use such materials in accordance with the Grant of rights Section above. Copies of all such grant of rights from third parties have been submitted.

b. all of the facts contained in the Contribution are true and accurate;
c. the signatory (the Author or the employer) who has signed this Agreement below has full right, power and authority to enter into this Agreement and grant the rights herein on behalf of all of the Authors;
d. nothing in the Contribution is obscene, defamatory, libelous, violates any right of privacy or publicity, infringes any intellectual property rights (including without limitation copyright, patent, database or trademark rights) or any other human, personal or other rights of any person or entity or is otherwise unlawful; and
e. nothing in the Contribution infringes any duty of confidentiality which any of the Author(s) may owe to anyone else or violates any contract, express or implied, of any of the Author(s), and all of the institutions in which work recorded in the Contribution was created or carried out, have authorized such publication.

Dispute resolution

The Author(s) authorise NPG to take such steps as it considers necessary at its own expense in the Author(s) name and on their behalf if NPG believes that a third party is infringing or is likely to infringe copyright in the Contribution including but not

limited to initiating legal proceedings.

The Author(s) grant NPG the perpetual right to edit, correct, or retract the Contribution if NPG considers (in its reasonable opinion) that such actions are required. The Author(s) hereby agree that they shall not object to NPG carrying out any such actions.

The Author(s) shall cooperate fully with NPG in relation to any legal action that might arise from the publication of the Contribution and the Author(s) shall give NPG access at reasonable times to any relevant accounts, documents and records within the power or control of the Author(s).

Reversion of rights

If the Contribution is rejected by NPG and not published, all rights under this licence shall revert to the Author(s).

Law & jurisdiction

This Agreement shall be governed by and construed in accordance with the laws of England and Wales. The parties irrevocably agree that the courts of England and Wales shall have exclusive jurisdiction to settle any dispute or claim that arises out of or in connection with this Agreement or its subject matter or formation.

Macmillan Publishers Limited (trading as Nature Publishing Group), registered office: Brunel Road, Houndmills, Basingstoke, Hampshire, RG21 6XS, UK. Company number 785998.

References

[1] European Commission, "Directive 2001/18/EC of the European parlament and the council of 12 march 2001 on the deliberate releas into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC," *Off. J. Eur. Communities*, vol. 2000, no. L106, pp. 1–38, 2001.

[2] U.S. Food and Drug administration, "Agency Summary Memorandum Re: Consultation with Calgene, Inc., Concerning FLAVR SAVR™ Tomatoes," 1994. .

[3] ISAAA, "ISAAA's GM Approval Database." [Online]. Available: http://www.isaaa.org/gmapprovaldatabase/. [Accessed: 20-Oct-2017].

[4] Biosafety Clearing House, "LMO quick-links," 2013. [Online]. Available: https://bch.cbd.int/resources/quicklinks.shtml. [Accessed: 08-Jun-2016].

[5] United States Department of Agriculture, "Petitions for Determination of Nonregulated Status," 2016. [Online]. Available: https://www.aphis.usda.gov/biotechnology/petitions_table_pending.shtml. [Accessed: 20-Oct-2017].

[6] U.S. Food and Drug Administration, "AquAdvantage Salmon Approval Letter and Appendix," 2015. .

[7] "Strategic Plan for the Cartagena Protocol on Biosafety for Period 2011-2020 (BS-V/16, Annex I)," 2014. .

[8] M. Vigani, V. Raimondi, and A. Olper, "International trade and endogenous standards: the case of GMO regulations," *World Trade Rev.*, vol. 11, no. 3, pp. 415–437, Jul. 2012.

[9] G. P. Gruère, "A Review of International Labeling Policies of Genetically Modified Food to Evaluate India ' s Proposed Rule," *AgBioforum*, vol. 10, no. 1, pp. 51–64, 2007.

[10] J. Žel, M. Milavec, D. Morisset, D. Plan, G. Van den Eede, and K. Gruden, *How to reliably test for GMOs*. Boston, MA, MA: Springer US, 2012.

[11] European Commission, "Regulation (EC) No 1829/2003 of the European parliament and of the council of 22 September 2003 on genetically modified food and feed," 2003.

[12] European Commission, "Commission Regulation (EU) No 619/2011 of 24 June 2011 laying down the methods of sampling and analysis for the official control of feed as regards presence of genetically modified material for which an authorisation procedure is pending or the authorisat," *Off. J. Eur. Union*, 2011.

[13] J. L. Neufeld, *The State of Food and Agriculture*, vol. 34, no. 4. 2014.

[14] "Rapid Alert System for Food and Feed (RASFF) Portal." [Online]. Available:

https://webgate.ec.europa.eu/rasff-window/portal/?event=SearchResultPrint&StartRow=0. [Accessed: 06-Nov-2017].

[15] A. Prusak, J. Strojny, and G. Rowe, "Is GMO 'sustainable'? A review of the environmental risk of GM plants in comparison with conventional andorganic crops," *Mod. Manag. Rev.*, no. January, 2014.

[16] "Genetically Engineered Crops: Experiences and prospectives," National Academies Press, Washington, D.C., Dec. 2016.

[17] M. Mazzara *et al.*, "Kernel Lot Distribution Assessment (KeLDA): A Comparative Study of Protein and DNA-Based Detection Methods for GMO Testing," *Food Anal. Methods*, vol. 6, no. 1, pp. 210–220, 2013.

[18] D. Morisset, P. Novak, D. Zupanič, K. Gruden, N. Lavrač, and J. Žel, "GMOseek: a user friendly tool for optimized GMO testing," *BMC Bioinformatics*, vol. 15, no. 1, p. 258, Jan. 2014.

[19] A. Block *et al.*, "The GMOseek matrix: a decision support tool for optimizing the detection of genetically modified plants," *BMC Bioinformatics*, vol. 14, no. 1, p. 256, Jan. 2013.

[20] A. Holst-Jensen *et al.*, "Detecting un-authorized genetically modified organisms (GMOs) and derived materials," *Biotechnol. Adv.*, vol. 30, no. 6, pp. 1318–1335, Nov. 2012.

[21] G. J. Randhawa, R. Chhabra, and M. Singh, "Multiplex PCR-Based Simultaneous Amplification of Selectable Marker and Reporter Genes for the Screening of Genetically Modified Crops," *J. Agric. Food Chem.*, vol. 57, no. 12, pp. 5167–5172, Jun. 2009.

[22] G. J. Randhawa, R. Chhabra, and M. Singh, "Decaplex and real-time PCR based detection of MON531 and MON15985 Bt cotton events," *J. Agric. Food Chem.*, vol. 58, no. 18, pp. 9875–9881, 2010.

[23] F. X. Luan, R. Tao, Y. G. Xu, J. Wu, and X. J. Guan, "High-throughput detection of genetically modified rice ingredients in foods using multiplex polymerase chain reaction coupled with high-performance liquid chromatography method," *Eur. Food Res. Technol.*, vol. 234, no. 4, pp. 649–654, Apr. 2012.

[24] A. Holck, B. O. Pedersen, and E. Heir, "Detection of five novel GMO maize events by qualitative, multiplex PCR and fluorescence capillary gel electrophoresis," *Eur. Food Res. Technol.*, vol. 231, no. 3, pp. 475–483, Jul. 2010.

[25] A. Germini *et al.*, "Development of a Seven-Target Multiplex PCR for the Simultaneous Detection of Transgenic Soybean and Maize in Feeds and Foods," *J. Agric. Food Chem.*, vol. 52, no. 11, pp. 3275–3280, Jun. 2004.

[26] H.-U. Waiblinger, B. Ernst, A. Anderson, and K. Pietsch, "Validation and collaborative study of a P35S and T-nos duplex real-time PCR screening method to detect genetically modified organisms in food products," *Eur. Food Res. Technol.*, vol. 226, no. 5, pp. 1221–1228, Mar. 2008.

[27] C. Bahrdt, A. B. Krech, A. Wurz, and D. Wulff, "Validation of a newly developed hexaplex real-time PCR assay for screening for presence of GMOs in food, feed and seed," *Anal. Bioanal. Chem.*, vol. 396, no. 6, pp. 2103–2112, Mar. 2010.

[28] H.-H. Dörries, I. Remus, A. Grönewald, C. Grönewald, and K. Berghof-Jäger,

"Development of a qualitative, multiplex real-time PCR kit for screening of genetically modified organisms (GMOs)," *Anal. Bioanal. Chem.*, vol. 396, no. 6, pp. 2043–2054, Mar. 2010.

[29] I. Huber *et al.*, "Development and Validation of Duplex, Triplex, and Pentaplex Real-Time PCR Screening Assays for the Detection of Genetically Modified Organisms in Food and Feed," *J. Agric. Food Chem.*, vol. 61, no. 43, pp. 10293–10301, Oct. 2013.

[30] H.-U. Waiblinger, L. Grohmann, J. Mankertz, D. Engelbert, and K. Pietsch, "A practical approach to screen for authorised and unauthorised genetically modified plants," *Anal. Bioanal. Chem.*, vol. 396, no. 6, pp. 2065–2072, Mar. 2010.

[31] T. W. Prins, R. A. van Hoof, I. M. J. Scholtens, and E. J. Kok, "Novel TaqMan PCR screening methods for element cry3A and construct gat/T-pinII to support detection of both known and unknown GMOs," *Eur. Food Res. Technol.*, 2017.

[32] I. M. J. J. Scholtens, B. Molenaar, R. A. van Hoof, S. Zaaijer, T. W. Prins, and E. J. Kok, "Semiautomated TaqMan PCR screening of GMO labelled samples for (unauthorised) GMOs," *Anal. Bioanal. Chem.*, vol. 409, no. 15, pp. 3877–3889, Jun. 2017.

[33] A. Z. Dinon, T. W. Prins, J. P. Van Dijk, A. C. M. Arisi, I. M. J. Scholtens, and E. J. Kok, "Development and validation of real-time PCR screening methods for detection of cry1A.105 and cry2Ab2 genes in genetically modified organisms," *Anal. Bioanal. Chem.*, vol. 400, pp. 1433–1442, 2011.

[34] G. J. Randhawa, D. Morisset, M. Singh, and J. Žel, "GMO matrix: A cost-effective approach for screening unauthorized genetically modified events in India," *Food Control*, vol. 38, no. 1, pp. 124–129, Apr. 2014.

[35] I. Scholtens *et al.*, "Practical experiences with an extended screening strategy for genetically modified organisms (GMOs) in real-life samples," *J. Agric. Food Chem.*, vol. 61, pp. 9097–9109, 2013.

[36] R. Spalinskas, M. Van den Bulcke, and A. Milcamps, "Efficient retrieval of recombinant sequences of GM plants by Cauliflower Mosaic Virus 35S promoter-based bidirectional LT-RADE," *Eur. Food Res. Technol.*, vol. 237, no. 6, pp. 1025–1031, Dec. 2013.

[37] G. Tan *et al.*, "SiteFinding-PCR: A simple and efficient PCR method for chromosome walking," *Nucleic Acids Res.*, vol. 33, no. 13, pp. 1–7, 2005.

[38] C. Liang *et al.*, "Detecting authorized and unauthorized genetically modified organisms containing vip3A by real-time PCR and next-generation sequencing," *Anal. Bioanal. Chem.*, vol. 406, no. 11, pp. 2603–2611, 2014.

[39] M. A. Fraiture, P. Herman, I. Taverniers, M. De Loose, D. Deforce, and N. H. Roosens, "An innovative and integrated approach based on DNA walking to identify unauthorised GMOs," *Food Chem.*, vol. 147, pp. 60–69, Mar. 2014.

[40] M. A. Fraiture *et al.*, "An integrated strategy combining DNA walking and NGS to detect GMOs," *Food Chem.*, 2017.

[41] M.-A. Fraiture *et al.*, "Validation of a sensitive DNA walking strategy to characterise unauthorised GMOs using model food matrices mimicking common rice products," *Food Chem.*, vol. 173, pp. 1259–1265, Apr. 2015.

[42] Q. Trinh, W. Xu, H. Shi, Y. Luo, and K. Huang, "An A-T linker adapter polymerase chain reaction method for chromosome walking without restriction site cloning bias," *Anal. Biochmeistry*, vol. 425, pp. 62–67, 2012.

[43] W. Xu *et al.*, "Randomly broken fragment PCR with 5′ end-directed adaptor for genome walking," *Sci. Rep.*, vol. 3, no. 3465, Dec. 2013.

[44] V. Thirulogachandar, P. Pandey, C. S. Vaishnavi, and M. K. Reddy, "An affinity-based genome walking method to find transgene integration loci in transgenic genome," *Anal. Biochem.*, vol. 416, no. 2, pp. 196–201, Sep. 2011.

[45] Q. Trinh, H. Shi, W. Xu, J. Hao, Y. Luo, and K. Huang, "Loop-linker PCR: An advanced PCR technique for genome walking," *IUBMB Life*, vol. 64, no. 10, pp. 841–845, Oct. 2012.

[46] S.-B. Park, H.-Y. Kim, and J.-H. Kim, "Multiplex PCR system to track authorized and unauthorized genetically modified soybean events in food and feed," *Food Control*, vol. 54, pp. 47–52, Aug. 2015.

[47] D. Dobnik, D. Štebih, A. Blejec, D. Morisset, and J. Žel, "Multiplex quantification of four DNA targets in one reaction with Bio-Rad droplet digital PCR system for GMO detection," *Sci. Rep.*, vol. 6, no. October, p. 35451, Oct. 2016.

[48] D. Morisset, D. Štebih, M. Milavec, K. Gruden, and J. Žel, "Quantitative analysis of food and feed samples with droplet digital PCR," *PLoS One*, vol. 8, no. 5, p. e62583, May 2013.

[49] P. Corbisier, S. Bhat, L. Partis, V. R. D. Xie, and K. R. Emslie, "Absolute quantification of genetically modified MON810 maize (Zea mays L.) by digital polymerase chain reaction.," *Anal. Bioanal. Chem.*, vol. 396, no. 6, pp. 2143–50, Mar. 2010.

[50] T. Demeke, T. Gräfenhan, M. Holigroski, U. Fernando, J. Bamforth, and S.-J. J. Lee, "Assessment of droplet digital PCR for absolute quantification of genetically engineered OXY235 canola and DP305423 soybean samples," *Food Control*, vol. 46, pp. 470–474, Dec. 2014.

[51] F.-U. Dalmira, P.-U. Melina, V. T. José-Benigno, L.-F. Josefina, G.-E. Raymundo, and A.-S. Abraham, "Development, Optimization, and Evaluation of a Duplex Droplet Digital PCR Assay To Quantify the T-nos / hmg Copy Number Ratio in Genetically Modified Maize," *Anal. Chem.*, vol. 88, no. 1, pp. 812–819, Jan. 2016.

[52] W. Fu *et al.*, "A highly sensitive and specific method for the screening detection of genetically modified organisms based on digital PCR without pretreatment," *Sci. Rep.*, vol. 5, no. 1, p. 12715, Oct. 2015.

[53] G. Kiddle *et al.*, "GMO detection using a bioluminescent real time reporter (BART) of loop mediated isothermal amplification (LAMP) suitable for field use," *BMC Biotechnol.*, vol. 12, no. 1, p. 15, 2012.

[54] G. J. Randhawa, M. Singh, D. Morisset, P. Sood, and J. Žel, "Loop-mediated isothermal amplification: Rapid visual and real-time methods for detection of genetically modified crops," *J. Agric. Food Chem.*, 2013.

[55] Y. Cheng *et al.*, "Loop-Mediated Isothermal Amplification for the Event-Specific Detection of Wheat B73-6-1," *Food Anal. Methods*, 2014.

[56] X. Huang, L. Chen, J. Xu, H. F. Ji, S. Zhu, and H. Chen, "Rapid visual detection

of phytase gene in genetically modified maize using loop-mediated isothermal amplification method," *Food Chem.*, 2014.

[57] M. Zhang, Z. Zhen, Y. Yu, X. Gao, and Y. Liu, "Development of a Rapid Event-Specific Loop-Mediated Isothermal Amplification Detection Method for Genetically Modified Maize NK603," *Food Anal. Methods*, 2016.

[58] Z. Zhen *et al.*, "Establishment of a loop-mediated isothermal amplification (LAMP) detection method for genetically modified maize MON88017," *Eur. Food Res. Technol.*, 2016.

[59] M. Singh, R. K. Bhoge, and G. Randhawa, "Real-time and visual loop-mediated isothermal amplification: Efficient GMO screening targeting pat and pmi marker genes," *Food Control*, vol. 71, pp. 248–254, Jan. 2017.

[60] X. Wang *et al.*, "Use of a novel metal indicator to judge loop-mediated isothermal amplification for detecting the 35S promoter," *Anal. Bioanal. Chem.*, vol. 409, no. 4, pp. 881–889, Feb. 2017.

[61] R. Li, J. Shi, B. Liu, D. Zhang, X. Zhao, and L. Yang, "International collaborative ring trial of four gene-specific loop-mediated isothermal amplification assays in GMO analysis," *Food Control*, vol. 84, pp. 278–283, Feb. 2018.

[62] A. Burrell, C. Foy, and M. Burns, "Applicability of Three Alternative Instruments for Food Authenticity Analysis:GMO Identification," *Biotechnol. Res. Int.*, vol. 2011, pp. 1–8, 2011.

[63] J. Guo *et al.*, "MPIC: A high-throughput analytical method for multiple DNA targets," *Anal. Chem.*, 2011.

[64] S. Basak, N. Z. Ehtesham, B. Sesikeran, and S. Ghosh, "Detection and identification of transgenic elements by fluorescent-PCR- based capillary gel electrophoresis in genetically modified cotton and soybean," *J. AOAC Int.*, 2014.

[65] E. D. Vega and M. L. Marina, "Characterization and Study of Transgenic Cultivars by Capillary and Microchip Electrophoresis," *Int. J. Mol. Sci.*, vol. 15, no. 12, pp. 23851–23877, Dec. 2014.

[66] F. von Götz, "See what you eat—broad GMO screening with microarrays," *Anal. Bioanal. Chem.*, vol. 396, no. 6, pp. 1961–1967, Mar. 2010.

[67] D. Dobnik, D. Morisset, and K. Gruden, "NAIMA as a solution for future GMO diagnostics challenges," *Anal. Bioanal. Chem.*, 2010.

[68] N. Shao *et al.*, "MACRO: A combined microchip-PCR and microarray system for high-throughput monitoring of genetically modified organisms," *Anal. Chem.*, vol. 86, no. 2, pp. 1269–1276, 2014.

[69] Y. Li, T. Xiong, H. Wu, and Y. Yang, "Visual DNA microarray coupled with multiplex-PCR for the rapid detection of twelve genetically modified maize," *BioChip J.*, vol. 10, no. 1, pp. 42–47, Mar. 2016.

[70] A. Turkec, S. J. Lucas, B. Karacanli, A. Baykut, and H. Yuksel, "Assessment of a direct hybridization microarray strategy for comprehensive monitoring of genetically modified organisms (GMOs)," *Food Chem.*, vol. 194, pp. 399–409, Mar. 2016.

[71] S. H. Choi, "Hexaplex PCR assay and liquid bead array for detection of stacked genetically modified cotton event 281-24-236??3006-210-23," *Anal. Bioanal. Chem.*, 2011.

[72] S. H. Choi, "Development of Detection System Using Multiplex PCR and Liquid Beadarray for Stacked Genetically Modified Rice Event (LS28×Cry1Ac)," *J. Korean Soc. Appl. Biol. Chem.*, vol. 53, no. 5, pp. 639–646, Oct. 2010.

[73] X. Han *et al.*, "Development and primary application of a fluorescent liquid bead array for the simultaneous identification of multiple genetically modified maize," *Biosens. Bioelectron.*, 2013.

[74] W. Fu *et al.*, "A high-throughput liquid bead array-based screening technology for Bt presence in GMO manipulation," *Biosens. Bioelectron.*, 2016.

[75] P. J. Sykes, S. H. Neoh, M. J. Brisco, E. Hughes, J. Condon, and A. A. Morley, "Quantitation of targets for PCR by use of limiting dilution," *Biotechniques*, 1992.

[76] B. Vogelstein and K. W. Kinzler, "Digital PCR," *Proc. Natl. Acad. Sci.*, vol. 96, no. 16, pp. 9236–9241, Aug. 1999.

[77] J. Pavšič *et al.*, "Standardization of nucleic acid tests for clinical measurements of bacteria and viruses," *J. Clin. Microbiol.*, 2015.

[78] "BioSitrace: Traceability for biologically relavant molecules and entities." [Online]. Available: http://biositrace.lgcgroup.com/. [Accessed: 24-Apr-2017].

[79] D. Wahler, L. Schauser, J. Bendiek, and L. Grohmann, "Next-Generation Sequencing as a Tool for Detailed Molecular Characterisation of Genomic Insertions and Flanking Regions in Genetically Modified Plants: a Pilot Study Using a Rice Event Unauthorised in the EU," *Food Anal. Methods*, vol. 6, no. 6, pp. 1718–1727, Dec. 2013.

[80] L. Young *et al.*, "Genetics, structure, and prevalence of FP967 (CDC Triffid) T-DNA in flax," 2011.

[81] L. Yang, C. Wang, A. Holst-Jensen, D. Morisset, Y. Lin, and D. Zhang, "Characterization of GM events by insert knowledge adapted re-sequencing approaches," *Sci. Rep.*, vol. 3, no. August 2015, Oct. 2013.

[82] K. Nakamura *et al.*, "Whole genome sequence analysis of unidentified genetically modified papaya for development of a specific detection method," *Food Chem.*, vol. 205, pp. 272–279, Aug. 2016.

[83] D. Kovalic *et al.*, "The Use of Next Generation Sequencing and Junction Sequence Analysis Bioinformatics to Achieve Molecular Characterization of Crops Improved Through Modern Biotechnology," *Plant Genome J.*, 2012.

[84] S. Willems *et al.*, "Statistical framework for detection of genetically modified organisms based on Next Generation Sequencing," *Food Chem.*, 2016.

[85] M. C. Kline and D. L. Duewer, "Evaluating Droplet Digital Polymerase Chain Reaction for the Quantification of Human Genomic DNA: Lifting the Traceability Fog," *Anal. Chem.*, vol. 89, no. 8, pp. 4648–4654, 2017.

[86] Decathlon Project, "Minimum Performance Parameters," 2015. [Online]. Available: http://www.decathlon-project.eu/sites/default/files/Deliverable6.1_MPP_Decathlon.pdf. [Accessed: 08-Jun-2016].

[87] U. Marchezi *et al.*, "Definition of Minimum Performance Requirements for Analytical Methods of GMO Testing," European Union Reference Laboratory for GM Food and Feed and European Network of GMO Laboratories, 2015.

[88] P. Corbisier *et al.*, "DNA copy number concentration measured by digital and droplet digital quantitative PCR using certified reference materials," *Anal. Bioanal. Chem.*, vol. 407, no. 7, pp. 1831–1840, Mar. 2015.

[89] J. a. Dagata, N. Farkas, and J. a. Kramer, "Method for Measuring the Volume of Nominally 100 μm Diameter Spherical Water-in-Oil Emulsion Droplets," Gaithersburg, MD, Feb. 2016.

[90] L. B. Pinheiro *et al.*, "Evaluation of a Droplet Digital Polymerase Chain Reaction Format for DNA Copy Number Quantification," *Anal. Chem.*, vol. 84, no. 2, pp. 1003–1011, Jan. 2012.

[91] J. Pavšič, J. Žel, and M. Milavec, "Digital PCR for direct quantification of viruses without DNA extraction," *Anal. Bioanal. Chem.*, vol. 408, no. 1, pp. 67–75, Jan. 2016.

[92] H.-B. Yoo *et al.*, "International Comparison of Enumeration-Based Quantification of DNA Copy-Concentration Using Flow Cytometric Counting and Digital Polymerase Chain Reaction," *Anal. Chem.*, vol. 88, no. 24, pp. 12169–12176, Dec. 2016.

[93] J. F. Huggett, S. Cowen, and C. A. Foy, "Considerations for digital PCR as an accurate molecular diagnostic tool," *Clin. Chem.*, vol. 61, no. 1, pp. 79–88, Jan. 2015.

[94] E. Afgan *et al.*, "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update," *Nucleic Acids Res.*, vol. 44, no. W1, pp. W3–W10, Jul. 2016.

Bibliography

Publications Related to the Thesis

Journal Articles

D. Dobnik, B. Spilsberg, A. B. Košir, A. Holst-Jensen and J. Žel , "Multiplex quantification of twelve EU authorised GM maize lines with droplet digital PCR," *Analytical Chemistry*, vol. 87, pp. 8218-8226, 2017. doi: 10.1021/acs.analchem.5b01208

A. B. Košir, B. Spilsberg, A. Holst-Jensen, J. Žel and D. Dobnik, "Development and inter-laboratory assessment of droplet digital PCR assays for multiplex qzuantification of 15 genetically modified soybean lines," *Scientific Reports*, 7, 2017. doi: 10.1038/s41598-017-09377-w

A. B. Košir, C. Divieto, J. Pavšič, S. Pavarelli, D. Dobnik, T. Dreo, R. Belloti, M. P. Sassi and J. Žel, "Droplet volume variabilitay as a critical factor for accuracy of absolute quantification using droplet digital PCR," *Analytical Bioanalytical Chemistry*, vol. 409, issue 28, pp 6689-6697, 2017. doi: 10.1007/s00216-017-0625-y

A. B. Košir, A. J. Arulandhu, M. M. Voorhuijzen, H. Xiao, R. Hagelaar, M. Staats, A. Costessi, J. Žel, E.J. Kok, J. P. van Dijk, "ALF: a strategy for identification of unauthorized GMOs in complex mixtures by a GW-NGS method and dedicated bioinformatics analysis," *Scientific Reports*, 7, 2017. doi: 10.1038/s41598-017-14469-8

Biography

Alexandra Bogožalec Košir was born on March 20, 1989 in Kranj, Slovenija. In 2013, she graduated from Biochemistry at the Faculty of Chemistry and Chemical Technology, University of Ljubljana, Slovenia, with the thesis entitled "*Selection, characterization and use of monoclonal antibodies against peptide 298-310 from primary structure of bilitranslocase*". The study for her diploma thesis was conducted at the Blood Transfusion Centre of Slovenia, Ljubljana, under the supervision of Prof. Dr. Vladka Čurin Šerbec, and Asst. prof. Dr. Uroš Rajčevič. The results of her studies were presented at two international congresses and published in Zdravniški vestnik.

In 2014, Alexandra became a young researcher at the Department of Biotechnology and Systems Biology, under the supervision of Prof. Dr. Jana Žel, PhD and Dr. David Dobnik. The same year she enrolled in a PhD study program of Nanosciences and Nanotechnologies at the Jožef Stefan International Postgraduate School, Ljubljana, Slovenia, also under the supervision of Prof. Dr. Jana Žel.

Her research focuses on the development and study of multi-target approaches for faster, more accurate detection and quantification of nucleic acids, using methods such as ddPCR and NGS. The subject of her work are genetically modified organisms. The majority of Alexandra's work was conducted under the European Union's Seventh Framework Programme project DECATHLON, within which she had a chance to cooperate with several researchers from diverse fields of research. In 2015, she visited RIKILT UR research institute in Wageningen, the Netherlands, where she trained and prepared a post NGS bioinformatics pipeline for detection of unauthorised/unknown genetically modified organisms in complex samples. She has attended workshops and congresses, where she presented her work, either in a form of a talk or a poster. Alexandra participated in organising two international hands-on workshops on dPCR, which were held at NIB in the fall of 2016 and 2017. She has also been active in NIB's efforts and activities for the promotion of science, especially by presenting the GMO field to the broader public.

As the National Institute of Biology is the National Reference Laboratory (NRL) in the field of GMO detection, Alexandra also gained a wide knowledge on GMO legislation in the European Union and other countries around the globe and issues on detection and quantification of nucleic acid in this context. She attended several NRL workshops and plenary meetings organised by The European Union Reference Laboratory for GM Food and Feed of the European Union Joint Research Centre. In the Chinese-Slovene bilateral project, she also visited Jiao Tong University in Shanghai, where she gained additional knowledge on molecular principles applied in the field of GMO detection and Chinese GMO legislation.