



Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

Detection of genetically modified organisms using highly multiplexed amplicon sequencing

C. Sarai Reyes-Avila^a, Dominique Waldvogel^b, Nicolas Pradervand^c, Sylvain Aubry^{d,e}, Daniel Croll^{a,*}

^a Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, CH-2000, Neuchâtel, Switzerland

^b Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland

^c Agroscope, Posieux, Switzerland

^d Department of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland

^e Federal Office for Agriculture, Bern, Switzerland

ARTICLE INFO

Keywords:

Genetically modified organisms

Crops

Multiplex PCR

NGS

Amplicon-sequencing

ABSTRACT

The circulation of products based on genetically modified (GM) organisms is highly regulated by some governments with strict implementation rules for the breeding, planting, marketing, labelling, and trading of such products. To ensure compliance, accurate detection methods for GM events are necessary, along with assurance that GM material falls within relevant threshold levels. The increasing complexity and potential of undocumented GM are a growing challenge for genetic screening. Here, we developed and assessed a highly multiplexed amplicon sequencing assay for the detection of GM events based on a microfluidics platform and next-generation sequencing (NGS). To probe GM events comprehensively, we designed a total of 230 new amplicons to cover flanking, promoter, junction, and coding sequences of GM sequences. In addition, we designed and implemented parallel amplification of ribosomal and chloroplast markers to define crop species identity from potentially mixed samples. Using reference GM material of 11 crop species and multiple amplicons, we successfully detected the presence of 10 known modifications per GM event. We also find that reported flanking sequences of GM events may not be all useful for diagnostic. We assessed the assay's potential to detect GM events in mixed samples as well as in highly diluted DNA. Finally, we performed a prospective search of potentially undocumented GM events in plant material. Our microfluidics-based amplicon GM detection approach fills important gaps in detecting potentially undocumented and complex GM events by recovering a wide range of specific amplicon sequences for evaluation. Integrating highly parallel amplicon assays in GM screening efforts should be an effective complement to aid post-market monitoring and regulatory compliance efforts.

1. Introduction

Genetically modified (GM) crops contain genes that have been artificially introduced conferring beneficial traits such as herbicide tolerance, and drought or pest resistance (Alasaad, Alzubi, and Kader 2016). The production of GM organisms has been rising and becoming more widely available commercially. Commercial applications of GM in crops have led to hundreds of variants of genetic modifications in dozens of crop species but mostly in the main cash crops (maize, soybean, canola and cotton). Since the '90s, the European Union (EU) implemented risk-based legislation governing the planting, marketing, labelling and trade of GMOs in Europe (Serageldin, 1999). To enforce legislation, it is

essential to develop accurate methods for the detection of GM and monitoring of threshold levels.

Any material deriving from GM crops might be identified by testing for the presence of introduced DNA. GM events or constructs that are often found in commercialized GM crops usually consist of several elements (promoters, terminators, genes, antibiotic resistance cassettes) that help to screen for their presence. There is a broad range of methods for GM DNA detection, and quantitative polymerase chain reaction (qPCR) approaches remain the most common (Aubry et al., 2021). qPCRs allow the detection, identification, and quantification of genetic modifications, based on PCR amplification recorded in real-time. As an example, a semiautomated TaqMan PCR screening of GMO-labelled

* Corresponding author.

E-mail address: daniel.croll@unine.ch (D. Croll).

<https://doi.org/10.1016/j.foodcont.2024.110670>

Received 19 March 2024; Received in revised form 13 June 2024; Accepted 20 June 2024

Available online 22 June 2024

0956-7135/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

samples was recently developed (Scholtens et al., 2017). Scholtens et al. semi-automated screening based on 32 primer and probe sequences derived from methods used in routine GMO feed sample screenings. Scholtens et al. verified the 32 primers in 59 different GMO reference materials. When a GMO element cannot be explained by any sample labelling, this is indicative of an unknown (*i.e.* unauthorized) GMO.

To facilitate the detection of several DNA targets in a single reaction, multiplex PCR-based methods were developed. qPCR-based multiplexing is constrained through costs and technical challenges to retain high reproducibility across laboratories. Official control laboratories are required certification to ensure high reproducibility standards. In addition, strategies have been developed to increase assay throughput without compromising sensitivity or specificity.

The introduction of microfluidic chips in multiplex PCR amplification offered significant gains in efficiency. Studies *e.g.* by Kuan-Lun et al. (2023), Nouwairi et al. (2022), Dong (2023), Yang et al. (2021), and Lehnert and Gijs (2024) showed that reaction times can be reduced with lower reagent volume requirements and robust target multiplexing. Nouwairi et al. achieved reduced PCR cycle time with a custom microfluidic chips (Nouwairi et al., 2022). Yang et al. showcased a novel circular array-shaped microfluidic chip that allows high-throughput detection of bacterial pathogens, significantly decreasing the time and spatial requirements typical of conventional PCR setups (Yang et al., 2023). Lehnert and Gijs (2024) showed that microfluidic systems can be useful for rapid diagnostics in clinical settings (Lehnert & Gijs, 2024). Microfluidics offers the most powerful options in combination with next-generation sequencing (NGS) of amplicons. NGS has been used extensively for GM detection and allows to simultaneously obtain multiple sequences and may facilitate the unambiguous detection of GMOs (Arulandhu et al., 2018). Arulandhu et al. tested five feed samples known to contain GMOs and screened 96 GMO-specific targets including endogenous, elements, constructs, and different events. Based on a related approach, Scholtens et al. targeted sequences from certified reference materials. Targeted sequencing was mainly applied to samples with multiple GM sequences of interest (Jagadeesan et al., 2019). DNA libraries were obtained from PCR-based amplification including several amplicons that can be sequenced using NGS technology. Standardized bioinformatics pipelines were applied to raw reads for filtering and assembly into contigs, which can be mapped to a reference sequence database (Willems et al., 2016). Establishing reference databases remains challenging as public disclosure of modified sequences can be incomplete (Moreira et al., 2017).

GMO detection traditionally was anchored in the simultaneous amplification of endogenous reference genes. Such genes typically amplify only in specific crop varieties where the GMO was initially developed, limiting how well an assay can be ported across diverse genetic backgrounds (Huang et al., 2013). Moreover, endogenous genes may vary significantly across different species, which reduces their utility for studies involving multiple species or agnostic assays. For example, in sugarcane, the selection of suitable endogenous reference genes is challenging due to the crop's complex polyploid and aneuploid genome structure. Such variability in traditional reference genes underscores the need for novel approaches. With the rapid increase in GMO testing needs, there is also a demand for standardization of GMO detection approaches to maintain versatility across the vast number of materials to be tested (Dong, 2023).

To address these challenges, we propose adopting a novel approach based on standardizing DNA-barcoding loci typically used for phylogenetic reconstruction. Furthermore, we avoided host-exogenous DNA boundaries in primer design to allow for cross-amplification of the same GM sequence in different backgrounds. This design strategy ensures that our assay can universally detect GM events without being limited by the host plant's genetic background, making it broadly applicable to various GM crops. To achieve these aims, we developed a highly multiplexed amplicon sequencing assay for the detection of GM events in crops. The assays was conceived by integrating a large set of available GM

sequences from public databases to design new, robust amplicons to amplify sequences linked to GM events in segments. Barcoding primers were designed to detect the species identity present in the material. Unlike traditional approaches that rely on endogenous reference genes as controls in GM crop detection, this study employs barcoded genes to identify the origin of tested samples. This innovative approach allows for the broad applicability of GMO detection across various plant species, overcoming the limitations of species-specific PCR which restricts analysis to detecting presence or absence of specific species. Amplifications were conducted in parallel using a microfluidics-based multiplex PCR and amplicons were sequenced as a pool using Illumina NGS. Filtered reads of positive and negative control samples were mapped to a compiled GM amplicon database. The performance of the assay, in particular specificity and sensitivity, in diluted and mixed samples were evaluated.

2. Materials and methods

2.1. Collection of samples

Genotyping was performed on 92 plant samples consisting primarily of GM-certified reference materials (Table S1). Samples covered eleven of the most widely cultivated plant species worldwide, such as maize (*Zea mays*), soybean (*Glycine max*), canola (*Brassica napus*), cotton (*Gossypium spp.*), alfalfa (*Medicago sativa*), potato (*Solanum tuberosum*), beetroot (*Beta vulgaris*), creeping bentgrass (*Agrostis stolonifera*), linseed (*Linum usitatissimum*), wheat (*Triticum aestivum*) and rice (*Oryza sativa*). The samples also included non-GMO crops used as negative controls. To assess the sensitivity of the assay, different concentrations of the same GM crops were used, typically ranging from 0.98 % to 100 % (ratio of the GM plant species in the total sample, expressed in mass/mass or copies/copies of haploid plant genomes). All our tested samples originated from certified reference materials (CRMs) provided under ISO17034 standards had been pre-validated for the presence of GM events and endogenous genes through qPCR testing. In addition to this certification, our laboratory also conducted extensive testing of these CRMs using official qPCR methods to confirm the presence of specified genes and ensure compliance with the reported standards. Detailed information about each sample and GMO events are provided in Table S1.

2.2. DNA extraction

DNA extractions were carried out by processing 200 mg of plant material with the NucleoSpin Plant II kit (Macherey-Nagel, GmbH, Germany) following the manufacturer's protocol. DNA concentrations of all samples were assessed using a NanoDrop One spectrophotometer and a Qubit (Thermo Scientific). Initial DNA concentration are reported in Table S1. We chose to standardize the DNA concentration of all our samples by diluting them in water to 50 ng μl^{-1} . To explore the effects of low DNA input, we diluted one of the GMO maize samples in a dilution series starting from 50 ng μl^{-1} down to 5 ng μl^{-1} (10-fold), 0.5 ng μl^{-1} (100-fold) and 0.05 ng μl^{-1} (1000-fold). For downstream applications, we performed two additional replicates for 41 samples, including replicates of the dilution series (Table S1).

2.3. Recovery of target sequences and barcoding plant species

We retrieved target sequences for amplicon design from the EUginius (European GMO Initiative for a Unified Database System) and the portugene (Moreira et al., 2017) GM sequence databases. Additionally, we added sequences encoding the dihydroflavonol 4-reductase gene from an unauthorized GM *Petunia* sequence (Fraiture et al., 2019). To avoid redundancy between the target sequences, we clustered nearly identical sequences and produced multiple alignments using Clustal Omega-v1.2.3 (Sievers & Higgins, 2014). Aligned sequences were used to produce a consensus, retaining ambiguous bases and yielding a total

of 115 unique sequences targeting specific GM events (Supplementary File S2). We also included primers for barcoding (*i.e.* species identification) purposes. For this, we added sets of primers for two widely used plant DNA barcoding loci (Kress, 2017): the chloroplast-encoded ribulose biphosphate carboxylase large chain gene (*rbcL*) and the nuclear ribosomal internal transcribed spacer (ITS).

2.4. Amplicon design

For the 115 unique target sequences, we segmented sequences exceeding 300 bp into multiple regions for individual amplicon design to improve amplification yields and coverage of the GM events. After the segmentation, we obtained 230 candidate loci for primer design according to Fluidigm Inc. recommendations. The targeted amplicon length ranged from 62 to 240 bp reflecting constraints in conserved sequences and base composition. We obtained a total of 230 pairs of primers corresponding to the GM target sequences (Supplementary File S1). In the case of the barcodes that amplify *rbcL* and the ITS, we compiled a set of 27 primers representing various amplicon designs covering the same loci corresponding to *rbcL* and ITS (Supplementary File S3).

2.5. DNA sequencing library preparation

Libraries were prepared following the manufacturer's protocol PN 101–0414 G1 for the Juno LP 192.24 integrated fluidic circuits plate (IFC; Fluidigm Corporation, San Francisco, CA, United States). After loading all reagents on the IFC, target amplicons were generated for each sample through PCR amplification on a specialized thermocycler (Juno system; Fluidigm). A total of 257 primers were subdivided into 10X assay pools, with each pool containing primers for various targets, ensuring universal applicability of the assay independent of the samples analyzed. This pooling strategy, which was identical for every sample, was partially determined by Fluidigm Inc. As an optimal strategy given the specific oligos. Custom designed primers were added randomly to the pools. The IFC with an inlet containing 2 µl of sample pre-mix, 2 µl genomic DNA (50 ng/µl) and 1 µl barcode primer mix consisting of a DNA sample and an individual barcode. High amplification efficiency of all 230 primer sets in the fluidic chip-based multiplex PCR was confirmed by the average sequencing depth of ~5000x, ensuring reliable and consistent target amplification across all samples. After amplification, samples were pooled in a single tube and purified. The first clean-up is double-sided (0.4X/0.9X Double-Sided SPRI), with first (0.4X) removing fragments that are bigger than the targets, then (0.9X) binding and selecting targets by washing off the smaller fragments. The second and third clean-ups were to remove excess primers (0.8X SPRI). Finally, sequencing adapters were added by PCR to the purified library followed by a final round of purification according to the manufacturer's protocol. The quantity and quality of the library were assessed using a Qubit fluorometer assay (ThermoFisher) and a 4200 TapeStation electrophoresis instrument (Agilent). The final library was sequenced on a single lane of a NextSeq 500 system (Illumina) in mid-output mode adding ~30% PhiX to reduce issues due to low sequence complexity.

2.6. Recovering of plant barcode sequences and GM targeted sequences by mapping

We obtained the *rbcL* sequence of maize from NCBI (NC_001666.2) and the ITS sequence from potato (CP046695.1). Matching *rbcL* and ITS sequences were then retrieved using BLAST (Altschul et al., 1990) to complete a library of barcoding sequences for all eleven included plant species. The crop barcode sequences were added to the GM reference sequences. We demultiplexed raw read data using bcl2fastq v-2.19.0.316 and used trimmomatic v-0.36 (Bolger et al., 2014) for quality trimming. The trimming parameters used were as follows: adapter clipping with ILLUMINACLIP:TruSeq3-PE.fa:2:30:10, removal

of leading low-quality bases with LEADING:15, removal of trailing low-quality bases with TRAILING:15, a sliding window trimming approach to cut when the average quality within a window of 5 bases falls below 15 (SLIDINGWINDOW:5:15), and dropping reads shorter than 50 bases (MINLEN:50). Forward and reverse reads were merged using flash v-1.2.11 (Magoč & Salzberg 2011). We aligned merged reads to the reference sequences using bowtie2 v-2.3.5 using the following settings: -very-sensitive-local -phred33 (Langmead et al., 2019). Based on the aligned reads, we calculated the depth per position ignoring locations outside of designed amplicons. We accessed all nucleotide sequences for *rbcL* and ITS available on NCBI for each of the eleven plant species generating eleven fasta files for *rbcL* and eleven fasta files for ITS. The fasta files were used as a reference to align the merged reads by bowtie2 v-2.3.5 using the following settings: -very-sensitive-local -phred33 (Langmead et al., 2019). From the aligned reads, we calculated the depth per position using samtools v-1.19. And we compared the depth to the previous depth of the *rbcL* sequence from maize (NC_001666.2) and the ITS sequence from potato (CP046695.1) with the retrieved BLAST sequences of the eleven crops. NC_001666.2 and CP046695.1 reference sequences were compared based on mapped reads per sample. We document our analyses pipeline in Supplementary File S4. All statistical analyses were performed with the R statistical software version 4.3.3.

3. Results

3.1. Sequence-guided design of the amplicon sequencing assay

In order to test the feasibility and assess the performances of the amplicon-sequencing assay to detect GM material in food, feed or seed matrices, we gathered a total of 115 consensus GM sequences from various sources: EUGenius (European GMO INITiative for a Unified Database System), and portuguese (Moreira et al., 2017) databases (Fig. 1A). In addition, we manually added three sequences from a gene encoding the dihydroflavonol 4-reductase from an unauthorized GM *Petunia* recently detected in the (Fraiture et al., 2019). For the design of amplicons, we retrieved the GM consensus sequences if multiple redundant sequences were found in databases. The lengths of these sequences ranged from 79 bp to 24,595 bp, with an average length of 1856 bp. Out of these, 99 sequences were longer than 300 bp. For sequences longer than 300 bp, multiple amplicons were designed to cover the entire target region. Specifically, the sequences longer than 300 bp were fragmented into smaller segments to ensure efficient amplification and coverage of the GM events, hence we designed multiple similarly spaced amplicons to cover the entire event (Fig. 1C). Based on this GM sequence set, 230 primer pairs for GM sequences that correspond to 82 unique GM sequences were designed (Fig. 1C) and 15 *rbcL* + 12 ITS primers were added for plant species identification (Fig. 1B). Following the manufacturer's instructions, we aimed for an amplicon length of ~200bp and we obtained a maximum length of 240 bp for 14 GM locus and the smallest amplicons with a length in the range of 69–199 bp for 21 GM loci (Fig. 2A). All 230 primer sets demonstrated high amplification efficiency in fluidic chip-based multiplex PCR, with an average sequencing depth of ~5000x consistent with robust amplification performance.

The amplicon sequencing was performed using a microfluidics platform (Fig. 1C). In a single flowcell run, we analyzed a total of 186 samples including replicates, representing 92 distinct samples of GMO and non-GMO control material. The species covered were alfalfa (n = 10), beetroot (n = 6), canola (n = 23), cotton (n = 26), creeping beetroot (n = 3), linseed (n = 3), maize (n = 66), potato (n = 6), rice (n = 3), soybean (n = 37) and wheat (n = 3) (Fig. 2B).

Illumina NextSeq 550 sequencing generated a total of 80,478,507 reads after trimming and merging read pairs per sample and locus. Furthermore, for every sample, the reads were aligned to the 115 target sequences and plant barcoding sequences *rbcL* and ITS sequences of the 11 species. One sample of maize obtained a maximum of 3,122,899

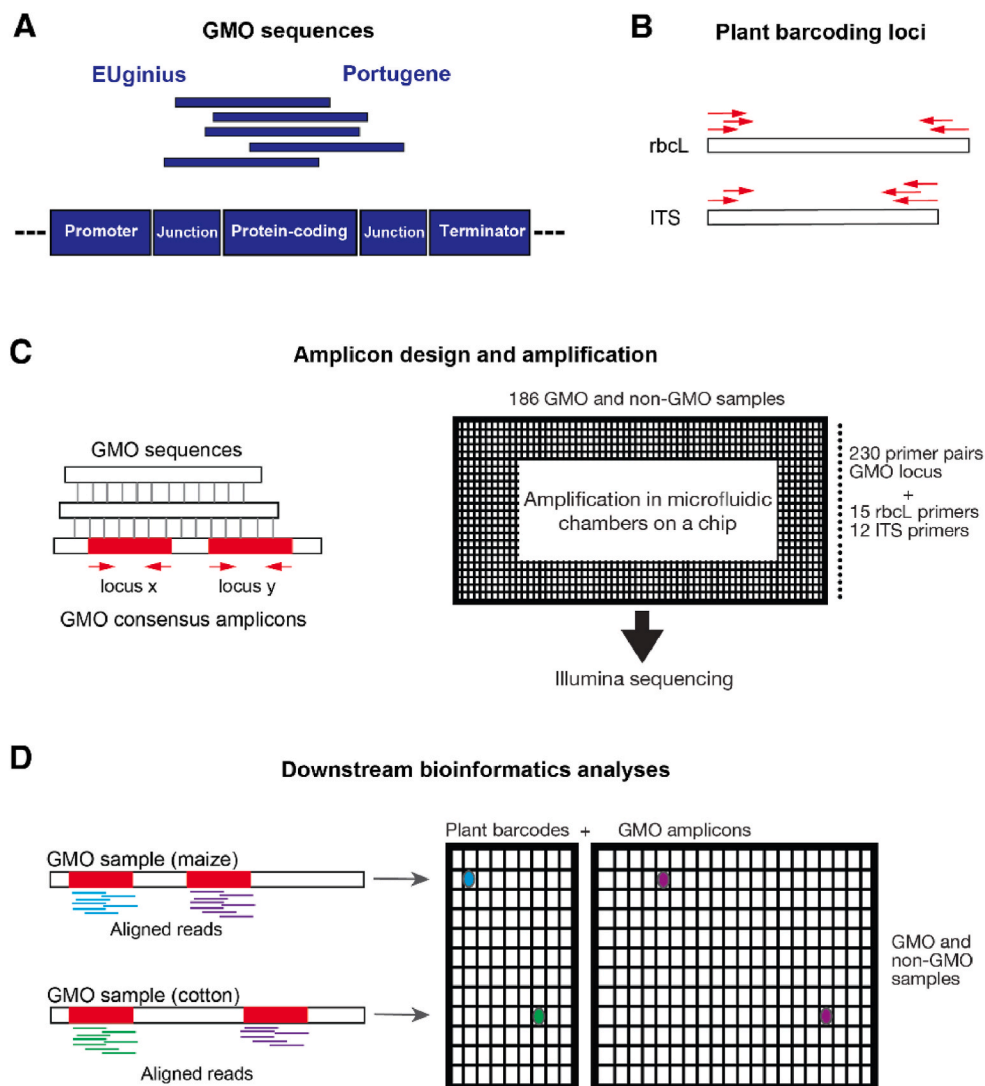


Fig. 1. Design and workflow of the targeted amplicon sequencing loci. (A) Recovery of GM sequence events from EUginius and Portugene databases, and the general structure of a GM event, consisting typically of promoter, junctions, protein-coding and terminator elements. (B) Multi-primer design for plant barcoding loci to cover sequence diversity among crop plants. (C) Multiple alignment of GMO sequences per event to obtain consensus sequences per locus in order to design consensus amplicons of around ~300bp. The microfluidics-based Juno system was used to amplify all amplicons across all samples in parallel. (D) Bioinformatics downstream analyses including alignment of reads against plant barcoding and GMO consensus sequences.

aligned reads (Fig. 2C). The minimum number was 229 aligned reads which corresponds to a sample of soybean (Fig. 2C). The other three minima of aligned reads ranged from 663 to 1418 and correspond to 3 replicates of DNA dilution (1000x) of a maize sample (Fig. 2B).

3.2. Identification of plant species by amplification of plant barcoding loci

An important step for the identification of GMO crops is to determine species identities. We incorporated in our targeting sequencing assay two sets of primers (Kress, 2017) previously recognized as plant DNA barcodes that amplify the *rbcl* and ITS loci. Given that our 15 *rbcl* + 12 ITS primers are amplifying in different regions of *rbcl* and ITS (Fig. 3A), we decided to work with the maximum number of mapped reads per barcode, which were correlated to the mean ($r = 0.897^{***}$, Pearson correlation coefficient), median ($r = 0.848^{***}$) and mode ($r = 0.848^{***}$). To evaluate the success of *rbcl* and ITS amplification, we extracted the maximum number of mapped reads for the 186 samples. The *rbcl* maximum count was significantly positively correlated to ITS maximum counts for alfalfa, canola, cotton, potato and soybean in the range of 0.983^{***} (p -value < 0.001) to 0.748^{***} (Fig. 3B). In the cases of

beetroot, creeping beetroot, rice and wheat, the correlation was not significant (p -values > 0.05), given that we have a low number of samples ($n = 3-6$; Fig. 3B). Maize samples showed also a positive correlation with 0.503^{***} (Fig. 3B). We also compared the maximum read count at barcodes against the total aligned counts to ensure that the general quality of the samples has been captured by the plant barcodes. For ITS, 10 plant species showed a significant and positive correlation from 0.88^{***} to 1^{***} , with the only exception of beetroot lacking a strong correlation (0.25^*). *RbcL* varies more between species with a range of -0.96 $p = 0.19$ to 0.97^{***} .

Every sample was mapped to the 11 crop species sequences of *rbcl* and ITS. We assigned species identities according to the highest read depth among *rbcl* and ITS reference sequences. As the species was known for all samples, we assessed the power of the assay to recover the true species. Only using the ITS primers we achieved 91 out of 186 correct predictions of the species identity corresponding to an accuracy of 48.9%. For the *rbcl* primers, we achieved 182 out of 186 correct predictions of the species identity corresponding to an accuracy of 97.8%. Hence, *rbcl* primers predicted more accurately the species and we used this barcoding locus for most species (Fig. 3C). The species that

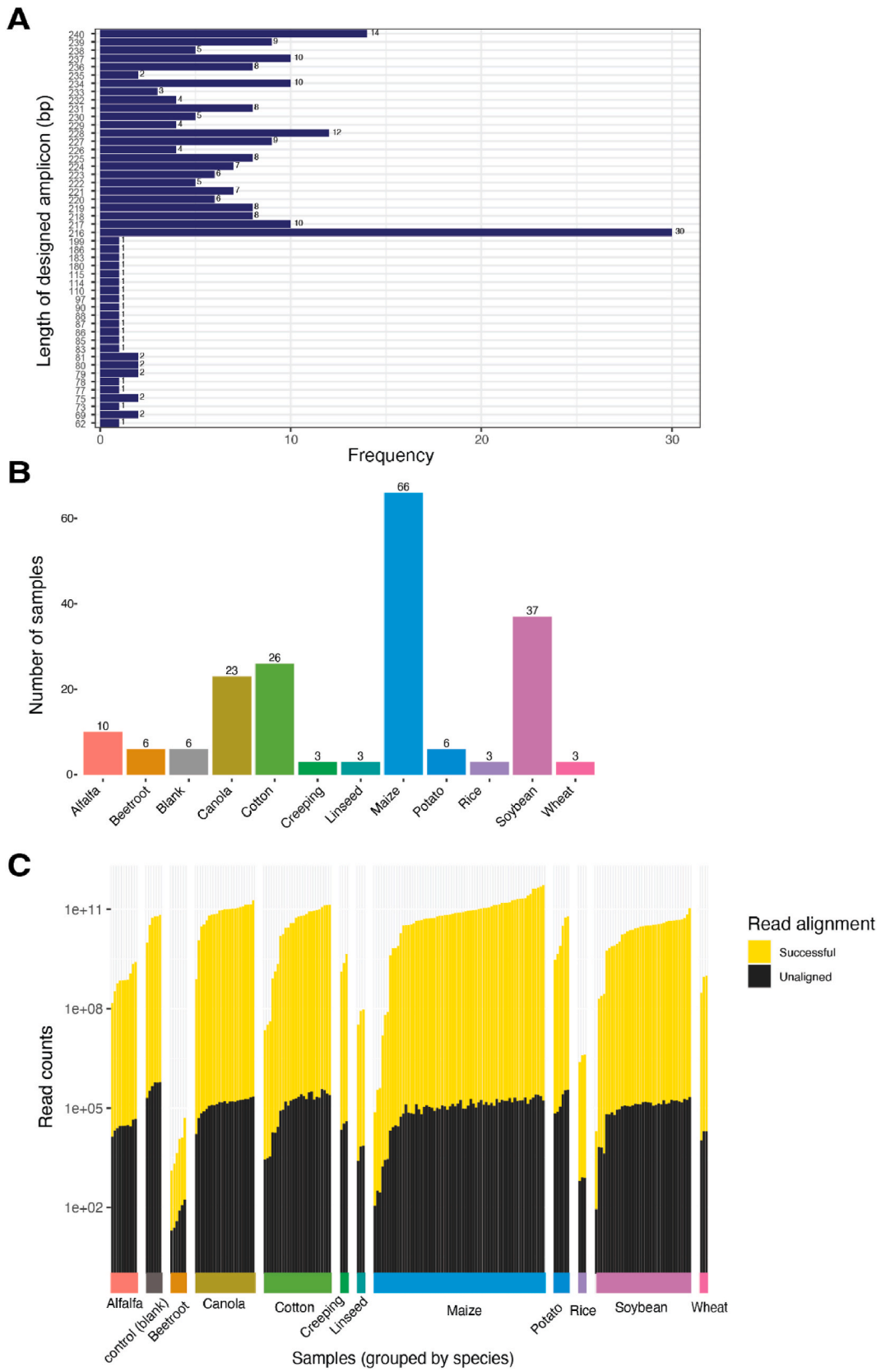


Fig. 2. Overview of the designed amplicons, samples used for testing and amplification. (A) Sequence length distribution among the 230 designed GM amplicons. (B) Number of samples per plant species covering 92 different GMO events across multiple samples. (c) Mapped and unmapped reads recovered per individual sample and species.

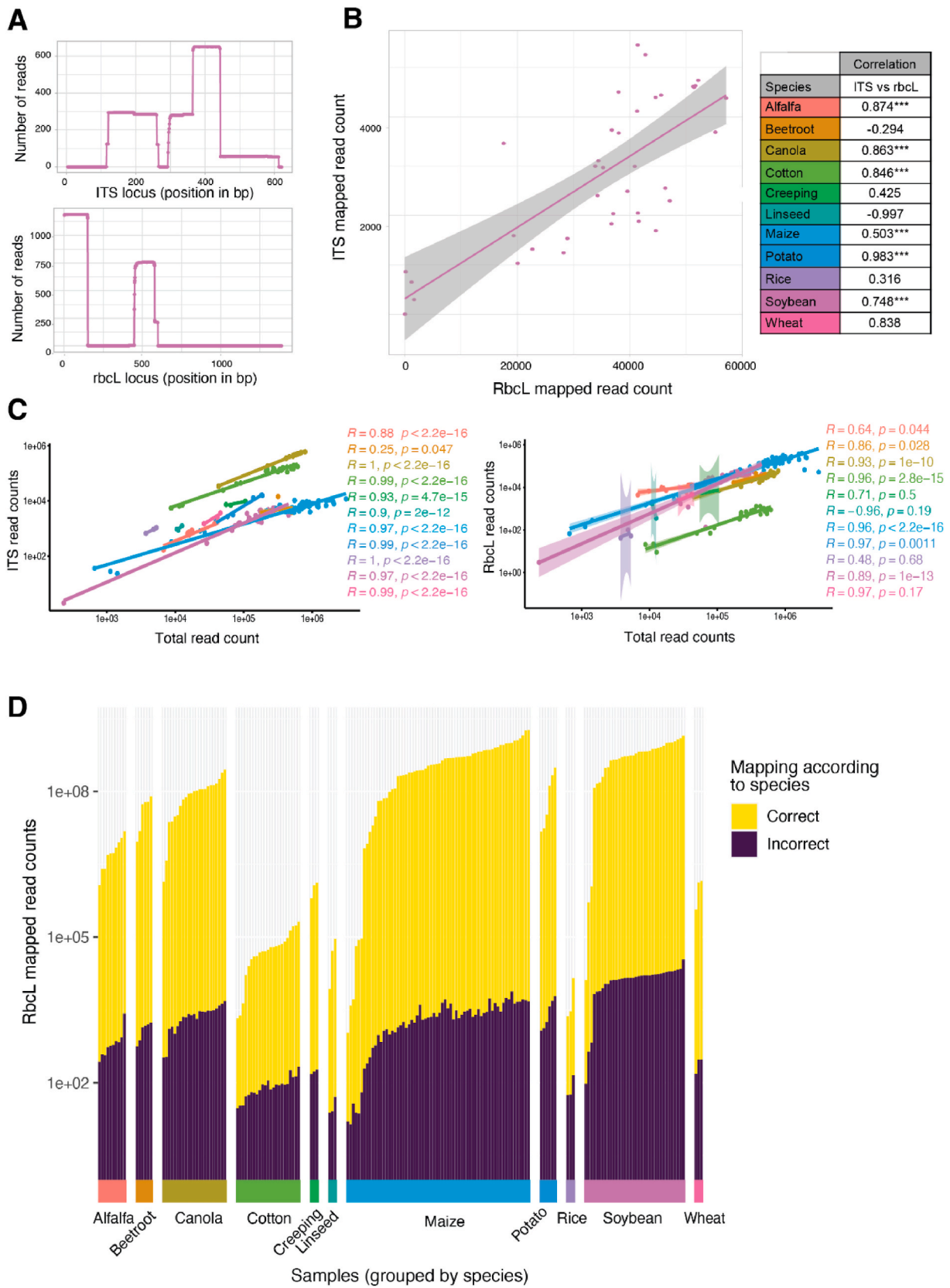


Fig. 3. Assessment of read mapping distribution on the plant barcoding reference sequences for the ITS and *rbcL* loci. (A) Depth of mapped reads at the ITS and *rbcL* loci for a soybean sample. (B) Correlations of the maximum mapped read count for ITS compared to *rbcL* for all soybean samples including a table of correlations for other plant species. (C) Correlation of the mapped read counts per plant barcoding locus compared to the total read counts. Correlation of total aligned read counts compared against the maximum count of *rbcL* reads showing *r* and *p*-values. (D) Counts of mapped reads for the *rbcL* locus mapped to the correct species reference sequence versus other reference sequences (note the logarithmic scale).

performed best was maize showing the highest correctly mapped reads ratio for 13 samples (79–172), meaning that the primers used to amplify *rbcl* and the *rbcl* sequence used to align are providing the best correct mapping. The worst correctly mapped reads ratios correspond to the three rice samples: 0.912–0.686.

3.3. Detection of GM events based on amplicon sequencing

The GM positive controls utilized in this study were sourced from CRM that have already been rigorously tested for endogenous reference gene amplification by the national certifying bodies. This pre-validation ensures the presence and accuracy of GM events and eliminating the need for repetitive endogenous gene testing in our experimental framework. Notably, each CRM used in this study has been verified for the accuracy of GM content by our laboratory using official qPCR methods. We analyzed 98 individual samples including 11 non-GMO controls. The samples were known to carry at least 10 sequence fragments from a specific GM event with 9 samples originating from maize and one from cotton (Supplementary Table S2). We analyzed the 10 GMO samples and compared these against the non-GMO controls. To account for variation in total reads among samples and the uneven presence of GM sequences, we normalized read counts using normalized mean depth by the maximum read depth at the barcoding locus. In the case of maize, we used *rbcl* because the read depth was higher in comparison to ITS (Fig. 3C). For cotton, we used ITS because the read depth was higher for ITS (Fig. 3C). From the 10 analyzed GM events, six events (MON89034, BT11, DAS59122, MIR162, MON88017 and MON87427) amplified well in the GMO sample and showed no meaningful amplification in the non-GMO control (Fig. 4A–F). The high amplification efficiency of all 230 primer sets in the fluidic chip-based multiplex PCR was evidenced by the consistent and robust detection of GM sequences by the average sequencing depth of ~5000x. For example, MON 89034 amplified well for 3'MON89034 (normalized count 0.094), 5'MON89034 (normalized count 0.004), hsp70-locus56 (normalized count 0.747), hsp70-locus307 (normalized count 0.285), Lta.lhcb1 (normalized count 0.002), tahsp17 (normalized count 0.331) and CTP2 (normalized count 0.405). For the three other GM events (MON810, SYN3272 and MON87460), the GMO maize also amplified more GM sequences compared to the non-GMO control (Fig. 4G–I). For example, the sample MON810 where the non-GMO control is amplified in 5' MON810-locus 59 (normalized count 0.383) vs the GM sample (normalized count: 0.526) (Fig. 4G). The last sample corresponds to MON15985 where the sequence labelled "Cotton_MON15985" amplified more in the GMO cotton (0.148) compared to non-GMO cotton (0.046 normalized counts; Fig. 4J). The most discriminant amplicon for detecting the MON15985 event was the Oriv amplicon.

3.4. Specificity and sensitivity of GMO detection

For monitoring purposes, mixed GMO samples need to be screened. To assess the power of our targeted sequencing assay, we tested three GMO samples diluted in control DNA at various ratios. For the GMO SYN3272 at the locus amy797E, the GMO (constituting 9.8%) amplified in the range of 0.01–0.002 normalized read counts contrasted with no recovered reads for the GMO mixed in at 0.98% (Fig. 5B). The sample DAS59122 amplified in both the 1% mixture in the range of 1.15×10^{-10} –0.0021 and the 10% mixture in the range of 0.00018–0.00796 (Fig. 5C). Unexpectedly, MON810 showed more amplification for GMO 1% compared to GMO 10% (Fig. 5A). For example, 5' MON810 locus 336 amplified for GMO 10% 1.2830 compared to 1.553 normalized read counts amplified for GMO 1%. The non-GMO control sample showed however 0.7572 normalized read counts. Next, we assessed the sensitivity of the amplicon assay to detect DNA diluted 10x, 100x, and 1000x. We tested the sample MON88017, however, the normalized counts mismatched the expected trend from the dilutions. The most likely explanation for this variability is the overall low number of reads likely

inducing noise in read counts (Fig. 5D). Using the mean depth of mapped reads (before normalization), we recovered the expected change in read depth along the dilution series (Fig. 5E).

3.5. Additional amplicons for exploratory detection of GM events

We recovered 73 target sequences from EUGenius, portogene and Fraiture et al. (2019) and amplified these in using the custom amplicon sequencing assay. We assessed amplification of the 10 GM sequences described above (Fig. 6). MON15985 cotton is amplifying the sequence Cotton_MON15985 as expected, and in addition OriV (Fig. 4J). Non-GMO controls showed amplification of multiple 3' and 5' GM flanking sequences as expected from the above findings. Three additional sequences including "FB707511.1 cry1A.105", "DL476427.1 CORN EVENT" and "aadA" showed amplification in the MON15985 GMO. Regarding the GMO maize samples, MON 89034 amplified the sequences 3'MON89034, hsp70, tahsp and CTP2 (Fig. 4A). Beyond amplified flanking sequences, three sequences amplified in the GMO including "FB707511.1 cry1A.105" and "DL476427.1 CORN EVENT", which the MON15985 GMO was not known to contain.

4. Discussion

Quantitative PCR has been for long the widely accepted standard for GMO detection in all matrices that could be circulating as food, feed, seeds and in the environment. qPCR is highly sensitivity but labour-intensive. Improving methods for detecting GMOs is therefore important. Here, we assessed the feasibility of a microfluidics-based approach for the detection of GM events. We designed a set of 230 amplicons that represent 82 unique GM events. In addition, 27 plant barcoding primers allowed to determine species contained in the samples. The advantage of using amplicon sequencing is the versatility and expandability to perform monitoring of samples without prior knowledge of the genetic modifications present.

Primers to amplify the barcoding genes *rbcl* and ITS allowed to identify most plant species contained in the sample. Samples originating from cotton were however not well identified with low read counts mapping to the *rbcl* gene. Low coverage may be due to selected primers performing poorly in those samples. The complexity of amplifying a barcoding locus using multiple overlapping pairs of primers was apparent in the read mapping patterns along the *rbcl* and ITS sequences. Replacing the pool of barcoding loci primers with pairs of custom-designed primers for a specific range of species would alleviate the complexity in amplification and likely increase consistency in amplification across the desired species. The detection of GM events in positive control samples was successful for a wide range of GM sequences. However, non-GMO controls amplified in some GM regions including 5' and 3' flanking sequences, the promoters, and terminators, which likely have homology to regions in the genome outside of the GM sequences. Such a lack of specificity in flanking and promoter sequences of GM events against regular plant DNA sequences can explain why GMO event promoters are also amplifying in non-GMO crops (Jores et al., 2021). The GM regions, which were amplified in non-GMO samples showed indeed sequence homology in genomes of non-GMO crops. This supports the lack of specificity in the 5' and 3' flanking sequences of GMO events. However, the recovered homologous sequences in crop genomes did not present 100% identity, which suggests that SNP calling could be used to differentiate between reads mapping to a GM event versus reads mapping to non-GMO sequences elsewhere in the genome.

Endogenous reference genes are conventionally used as controls in GM crop detection to ensure the presence of target DNA. Here, we opted for a different approach that extends beyond species-specific restrictions of amplifiable loci (Huang et al., 2013). Given the diverse nature of the plant materials to be tested, using typical endogenous reference genes would have constrained our ability to detect GM sequences in diverse monocot and dicot crops. Relying on barcoding loci such as chloroplast

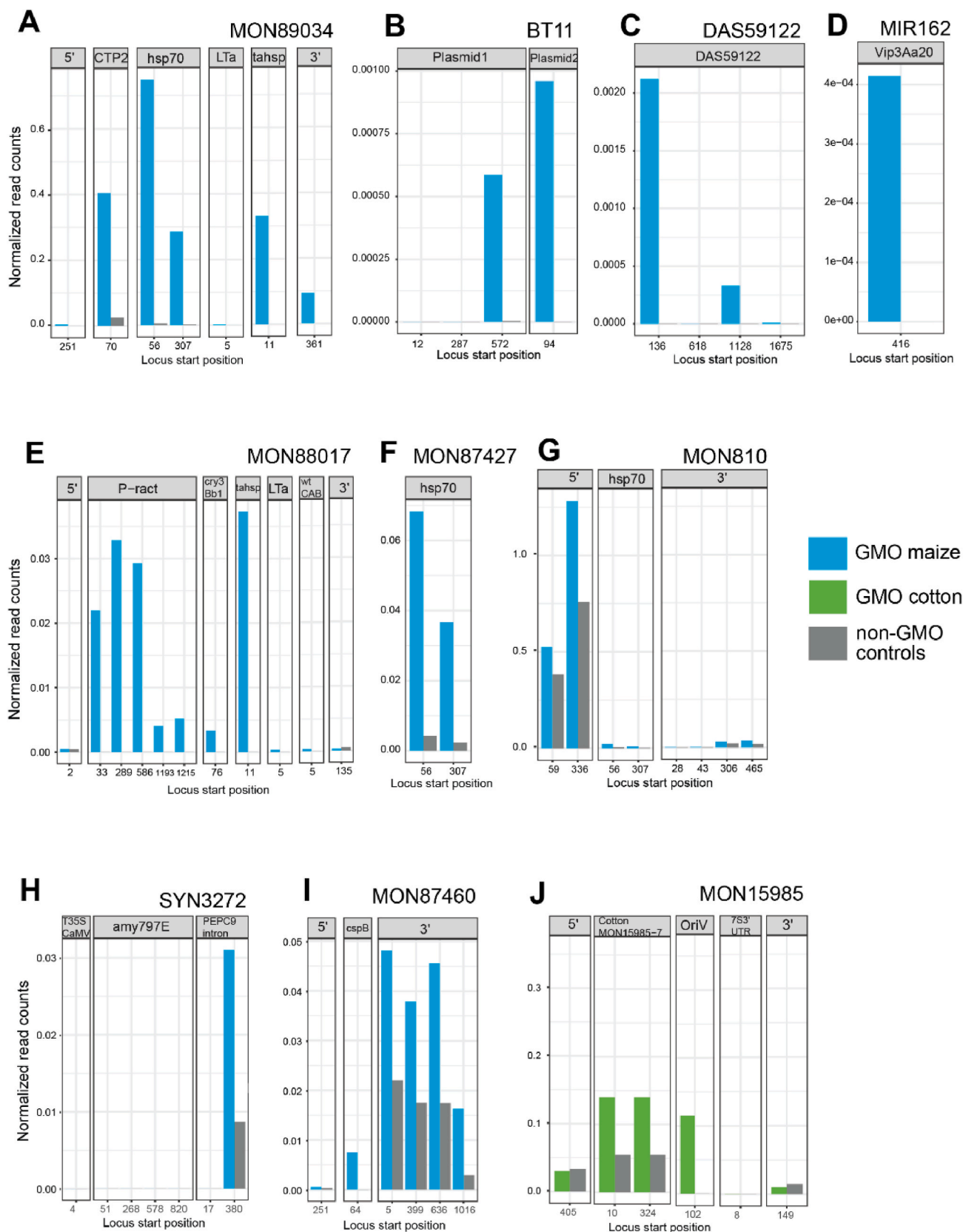


Fig. 4. Assessment of known GMO events by amplicon sequencing and contrast to non-GMO control samples. The following GM events were covered with multiple amplicons. Locus start positions indicate the amplicon location on the GMO consensus sequence. Events include (A) MON89034, (B) BT11, (C) DAS59122, (D) MON88017, (E) MON88017, (F) MON87427, (G) MON810, (H) SYN3272, (I) MON87460 and (J) MON15985.

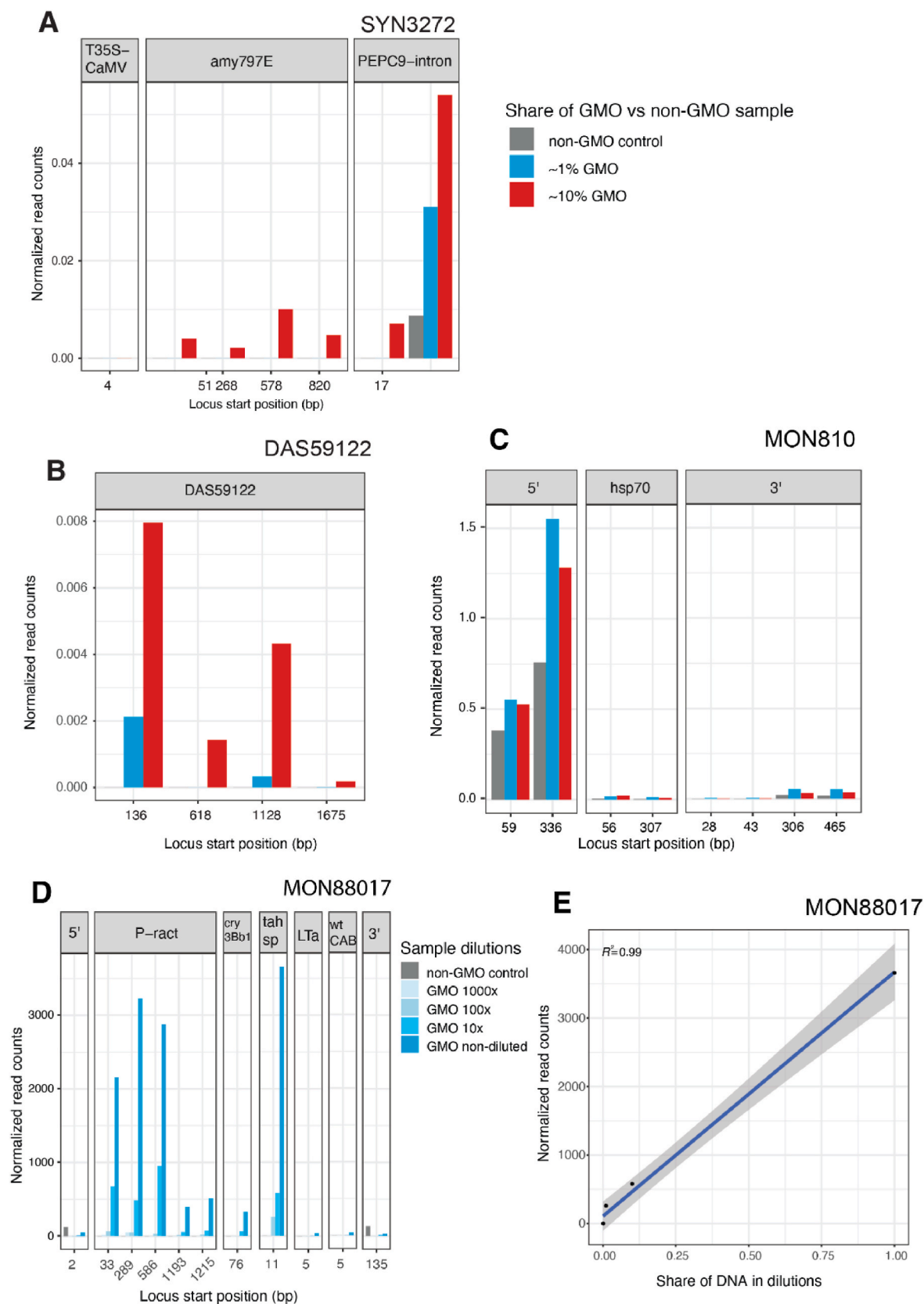


Fig. 5. Assessment of DNA concentration and sample dilution effects. (A) GM event called SYN3272, composed of three recovered regions (amy797E, PEPC9-intron, and T35S-CaMV) across different positions of the GM locus. Read counts were adjusted for each of the samples diluted to contain 0.98% concentration of a GMO sample in blue, the 9.8% concentration of GMO maize in red and the non-GMO maize in grey. (B) GM event DAS59122 was amplified at four different loci. (C) GM event MON810 spanning three assessed regions (3', 5' and hsp70). (D) Assessment of sample dilution effects (GMO samples mixed with sterile water) for GM event MON88017. (E) Correlation of GMO DNA share in diluted sampled with the normalized read counts.



Fig. 6. Amplifications of GM events included in the databases EUginius, Portugene and provided by Fraiture et al. (2019). Heatmaps for 11 samples covering nine different GMO events for maize with one non-GMO maize as control; one GMO event for cotton with one non-GMO cotton as control, aligned to the 82 reference sequences generated from the three databases (EUginius, portugene and Fraiture) and read counts were normalized using barcoding loci sequencing depth. The red stars are marking amplification in the samples for MON15985 GMO cotton where “FB707511.1 cry1A.105”, “DL476427.1 CORN EVENT” and “aadA” were amplified; for the MON 89034 “FB707511.1 cry1A.105” and “DL476427.1 CORN EVENT” and “dihydroflavonol4-reductase_MF521566.1”.

or ribosomal DNA markers makes amplification possible in principle across entire kingdoms. However, primer specificity and amplification biases need to be carefully evaluated. In our assessment, amplification of barcoding loci provided reasonable quantification accuracy. This is supported by the fact that our study relied on CRM, which had already undergone qPCR testing for endogenous gene presence. Our microfluidic approach could well identify species identities. However, we advocate for caution if even more diverse plant species are to be assayed. A preliminary trial run to compare amplification efficiency across species and taxonomic assignments is recommended. The microfluidic amplicons developed in this study diverge also in other ways from traditional GM detection approaches. Notably, we avoided host-exogenous DNA boundaries in our primer design. Such cross-boundary qPCR is typically employed to confirm the presence of a specific insertion event in a specific genetic background. However, designing amplicons only on exogenous DNA allows for universal amplification of specific GM sequences at any insertion locus or any genetic background. For specific host-exogenous DNA boundary detection our assay could either be expanded by additional amplicons or supplemented by qPCR validation.

Previous studies describing multiplex detection of GM events were largely based on amplicons stemming from known primer pairs used in routine GMO food and feed sample screening strategies and approved in GMO reference material (Scholtens et al., 2017; Arulandhu et al., 2018). In some jurisdictions including the European Union, there is an obligation to provide a detection method specific to the authorized GM event. As the common practice usually is centred on the use of qPCR, often only primer sequences are made available, limiting the investigation of GM events. The dependency on a certified set of primers restricts the completeness of the GM event sequences to be recovered, because amplicon sequences do not cover the complete GM event. The lack of long amplicon sequences and reference material accessibility are challenging to investigate GMO sequences comprehensively (Moreira et al., 2017). Our study shows that a *de novo* design of GM amplicons is feasible using public sequence information and that a broad range of potential GM sequences can be assessed in parallel. The approach of a microfluidics-based targeted amplicon sequencing assay enables to screen both hundreds of samples (or replicates) simultaneously but also allows for large sets of primers to be included in parallel. In principle, the microfluidics chips would allow the pooling of thousands of primer pairs for single-step amplifications. Given the uncertainty of amplifying specific sequences from unknown samples and modification events, the sequence information provides significantly greater certainty about the identity of an amplified sequence. This contrasts with qPCR approaches that lack direct validation capabilities under non-standard conditions. In contrast, targeted amplicon sequencing is less sensitive compared to qPCR. In our analyses, we found reliable amplification up to ~1:100 dilutions, at lower concentrations the detection is likely to be only poorly reproducible. Such detection limitations can be remedied partially by increasing the overall sequencing coverage of the amplicons as sensitivity is at least partially correlated with sequencing depth.

This study demonstrates the application of a comprehensive amplicon sequencing assay that leverages the parallelization offered by microfluidics platforms and the depth of NGS for the detection of GMOs. Our work fits into efforts to standardize and propose a statistical framework for the detection of GMOs based on the number of reads aligned per sample. Our workflow expands the capabilities by targeting many sequences of interest specifically and allowing for the efficient detection of plant species present in a sample. With 230 designed amplicons corresponding to GM events and additional primers for species-specific barcoding, our approach represents a significant advancement in GMO detection. The assay robustness was demonstrated by successfully identifying ten known genetic modifications across different crop species and showcasing the potential for uncovering undocumented genetic events. The integration of *rbcL* and ITS barcoding primers enhanced the assay's precision by enabling accurate species identification within mixed samples, an essential step in standardised GMO presence across

samples. Challenges remain in distinguishing between GM events and homologous non-GMO sequences. The use of sequence information provided by NGS offers a direct validation of the detected genetic material. The successful identification of GMOs in this study underscores the importance of developing advanced screening methods. Our study shows the relevance of amplicon sequencing that can be realistically implemented into GMO detection and efficiently analyzed using a structured bioinformatics pipeline.

Funding

DC received funding by the Swiss Federal Offices for Agriculture and for the Environment (FOEN).

CRediT authorship contribution statement

C. Sarai Reyes-Avila: Writing – original draft, Investigation, Formal analysis, Conceptualization. **Dominique Waldvogel:** Investigation. **Nicolas Pradervand:** Resources. **Sylvain Aubry:** Writing – review & editing, Methodology. **Daniel Croll:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

CSRA and DC have filed for a European patent for microfluidics-based GMO detection.

Data availability

Raw sequencing data is available from the NCBI Sequence Read Archive

Acknowledgments

Data was generated in collaboration with the Genetic Diversity Centre (GDC), ETH Zurich.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.foodcont.2024.110670>.

References

- Alasaad, N., Hussein, A., & Ahmad, A. K. (2016). Data in support of the detection of genetically modified organisms (GMOs) in food and feed samples. *Data in Brief*, 7 (June), 243–252. <https://doi.org/10.1016/j.dib.2016.02.035>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Arulandhu, A. J., van Dijk, J., Staats, M., Hagelaar, R., Voorhuijzen, M., Molenaar, B., van Hoof, R., et al. (2018). NGS-based amplicon sequencing approach; towards a new era in GMO screening and detection. *Food Control*, 93(November), 201–210. <https://doi.org/10.1016/j.foodcont.2018.06.014>
- Aubry, S., Avila, S. R., Croll, D., & Christ, B. (2021). Chapter 10. Omics-based detection, identification and quantification of GM food and feed: Current challenges and perspectives. In J. Barros-Velázquez (Ed.), *Food chemistry, function and analysis* (pp. 257–270). Cambridge: Royal Society of Chemistry. <https://doi.org/10.1039/9781839163005-00257>.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Dong, C., Li, F., Sun, Y., Long, D., Chen, C., Li, M., Wei, T., Martins, R. P., Chen, T., & Mak, P.-I. (2023). A syndromic diagnostic assay on a macrochannel-to-digital microfluidic platform for automatic identification of multiple respiratory pathogens. *Lab on a Chip*, November. <https://doi.org/10.1039/D3LC00728F>
- Fraiture, M.-A., Ujhelyi, G., Ovesná, J., Van Geel, D., De Keersmaecker, S., Saltykova, A., Papazova, N., & Roosens, N. H. C. (2019). MinION sequencing technology to

- characterize unauthorized GM *Petunia* plants circulating on the European union market. *Scientific Reports*, 9(1), 7141. <https://doi.org/10.1038/s41598-019-43463-5>
- GMO content - by analyte group - certified reference materials catalogue of the JRC. n.d. <https://crm.jrc.ec.europa.eu/c/By-analyte-group/GMO-content/40481/>. (Accessed 28 May 2024)
- Huang, H., Fang, C., Wang, R., Zhang, D., & Yang, L. (2013). Evaluation of four endogenous reference genes and their real-time PCR assays for common wheat quantification in GMOs detection. *PLoS One*, 8(9), Article e75850. <https://doi.org/10.1371/journal.pone.0075850>
- Jagadeesan, B., Gerner-Smidt, P., Allard, M. W., Leuillet, S., Winkler, A., Xiao, Y., Chaffron, S., et al. (2019). The use of next generation sequencing for improving food safety: Translation into practice. *Food Microbiology*, 79(June), 96–115. <https://doi.org/10.1016/j.fm.2018.11.005>
- Jores, T., Jackson, T., Wrightsman, T., Buckler, E. S., Cuperus, J. T., Fields, S., & Queitsch, C. (2021). Synthetic promoter designs enabled by a comprehensive analysis of plant core promoters. *Nature Plants*, 7(6), 842–855. <https://doi.org/10.1038/s41477-021-00932-y>
- Kress, W. J. (2017). Plant DNA barcodes: Applications today and in the future. *Journal of Systematics and Evolution*, 55(4), 291–307. <https://doi.org/10.1111/jse.12254>
- Langmead, B., Wilks, C., Antonescu, V., & Charles, R. (2019). Scaling read aligners to hundreds of threads on general-purpose processors. *John (Ed.) Bioinformatics*, 35(3), 421–432. <https://doi.org/10.1093/bioinformatics/bty648>
- Lehnert, T., & Gijs, M. A. M. (2024). Microfluidic systems for infectious disease diagnostics. *Lab on a Chip*, 24(5), 1441–1493. <https://doi.org/10.1039/D4LC00117F>
- Magoč, T., & Salzberg, S. L. (2011). Flash: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27(21), 2957–2963. <https://doi.org/10.1093/bioinformatics/btr507>
- Moreira, F., Carneiro, J., & Pereira, F. (2017a). A proposal for standardization of transgenic reference sequences used in food forensics. *Forensic Science International: Genetics*, 29(July), e26–e28. <https://doi.org/10.1016/j.fsigen.2017.04.022>
- Nouwairi, R. L., Cunha, L. L., Turiello, R., Scott, O., Hickey, J., Thomson, S., Knowles, S., Chapman, J. D., & Landers, J. P. (2022). Ultra-rapid real-time microfluidic RT-PCR instrument for nucleic acid analysis. *Lab on a Chip*, 22(18), 3424–3435. <https://doi.org/10.1039/D2LC00495J>
- Scholten, I. M. J., Molenaar, B., van Hoof, R. A., Zaaijer, S., Prins, T. W., & Kok, E. J. (2017). Semiautomated TaqMan PCR screening of GMO labelled samples for (unauthorised) GMOs. *Analytical and Bioanalytical Chemistry*, 409(15), 3877–3889. <https://doi.org/10.1007/s00216-017-0333-7>
- Serageldin, I. (1999). Biotechnology and food security in the 21st century. *Science*, 285(5426), 387–389. <https://doi.org/10.1126/science.285.5426.387>
- Sievers, F., & Higgins, D. G. (2014). Clustal Omega. *Current Protocols in Bioinformatics*, 48(December), 3.13.1–3.13.16. <https://doi.org/10.1002/0471250953.bi0313s48>
- Willems, S., Fraiture, M.-A., Deforce, D., De Keersmaecker, S. C. J., De Loose, M., Ruttink, T., ... Roosens, N. (2016). Statistical framework for detection of genetically modified organisms based on next generation sequencing. *Food Chemistry*, 192, 788–798. <https://doi.org/10.1016/j.foodchem.2015.07.074>
- Yang, B., Wang, P., Li, Z., You, Q., Sekine, S., Ma, J., Zhuang, S., Zhang, D., & Yamaguchi, Y. (2023). Simultaneous amplification of DNA in a multiplex circular array shaped continuous flow PCR microfluidic chip for on-site detection of bacterial. *Lab on a Chip*, 23(11), 2633–2639. <https://doi.org/10.1039/D3LC00274H>

Update

Food Control

Volume 168, Issue , February 2025, Page

DOI: <https://doi.org/10.1016/j.foodcont.2024.110917>



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Food Control

journal homepage: www.elsevier.com/locate/foodcont



Corrigendum to “Detection of genetically modified organisms using highly multiplexed amplicon sequencing” [Food Control 165 (2024) 110670]

C. Sarai Reyes-Avila^a, Dominique Waldvogel^b, Nicolas Pradervand^c, Sylvain Aubry^{d,e}, Daniel Croll^{a,*}

^a *Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, CH-2000, Neuchâtel, Switzerland*

^b *Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland*

^c *Agroscope, Posieux, Switzerland*

^d *Department of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland*

^e *Federal Office for Agriculture, Bern, Switzerland*

During copy-editing a reference was inadvertently deleted from the reference list. This reference has now been added again.

Reference to be added again:

Ho, K.-L., Ding, J., Fan, J.-S., Tsui, W. N. T., Bai, J., & Fan, S.-K.

(2023). Digital Microfluidic Multiplex RT-qPCR for SARS-CoV-2 Detection and Variants Discrimination. *Micromachines*, 14(8), 1627. <https://doi.org/10.3390/mi14081627>

DOI of original article: <https://doi.org/10.1016/j.foodcont.2024.110670>.

* Corresponding author.

E-mail address: daniel.croll@unine.ch (D. Croll).

<https://doi.org/10.1016/j.foodcont.2024.110917>

Available online 28 September 2024

0956-7135/© 2024 The Author(s). Published by Elsevier Ltd. All rights reserved, including those for text and data mining, AI training, and similar technologies.