Food Research International 221 (2025) 117218

Contents lists available at ScienceDirect

Food Research International

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

Genetic fingerprints derived from genome database mining allow accurate identification of genome-edited rice in the food chain via targeted high-throughput sequencing

Marie-Alice Fraiture [a,1], Jolien D'aes [a,1], Andrea Gobbo [a], Maud Delvoye [a], Anne-Cécile Meunier [b,c], Julien Frouin [b,c], Emmanuel Guiderdoni [b,c], Dieter Deforce [d], Charlotte De Vogelaere [a], Sigrid C.J. De Keersmaecker [a], Kevin Vanneste [a,2], Nancy H.C. Roosens [a,*,2]

[a] Sciensano, Transversal activities in Applied Genomics (TAG), rue Juliette Wytsman 14, 1050 Brussels, Belgium
[b] CIRAD, UMR AGAP institut, F-34398, Montpellier, France
[c] UMR AGAP institut, Université de Montpellier, CIRAD, INRAE, Institut Agro, F-34398 Montpellier, France
[d] Ghent University, Faculty of Pharmaceutical Sciences, Laboratory of Pharmaceutical Biotechnology, Ghent, Belgium

ARTICLE INFO

Keywords:
Genome-edited organism
Genetically modified organism
Genetic fingerprint
Machine learning
Multiplex PCR
High-throughput sequencing
Rice
CRISPR

ABSTRACT

Genome-edited (GE) organisms are currently classified as GMOs according to European legislation, requiring traceability and labelling in the food and feed supply chain. However, unambiguous identification of a specific GE organism with one or more induced single nucleotide variations (SNVs) dispersed across the genome remains challenging. This study explored whole-genome sequencing-based characterization, public genome databases, and machine learning tools to select key genetic elements and create a unique fingerprint for distinguishing a specific GE line. As a case study, a GE Nipponbare rice line containing a single CRISPR-Cas-induced SNV was used. To experimentally assess the detection of this fingerprint, a targeted high-throughput sequencing approach, including multiplex PCR-based enrichment of key genetic elements, was developed and successfully tested. This promising proof-of-concept demonstrates the potential of combining a unique genetic fingerprint with targeted high-throughput sequencing to facilitate the accurate detection of GE organisms, thereby supporting food traceability and regulatory compliance for the development of new GE lines, as well as protecting associated intellectual property.

1. Introduction

In food and feed, organisms generated through genome editing techniques, like CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) associated nuclease systems, are considered as genetically modified organisms (GMOs) according to a 2018 European Court of Justice ruling (European Commission Joint Research Centre & European Network of GMO Laboratories, 2023; Gelinsky & Hilbeck, 2018; Guertler et al., 2023). Consequently, in line with Directive 2001/18/EC, Regulation (EC) No 1829/2003 and Regulation (EC) No 1830/2003, the market authorization of such genome-edited (GE) organisms and derived products in the European food and feed chain is subject to labelling requirements. To ensure food and feed traceability and safeguard the consumers' freedom of choice, market controls are regularly organized by the European Competent Authorities, assisted by enforcement laboratories utilizing validated GMO detection methods (European Commission Joint Research Centre & European Network of GMO Laboratories, 2023). The development of detection methods for GE

* Corresponding author at: Sciensano, Transversal activities in Applied Genomics (TAG), rue Juliette Wytsman 14, 1050 Brussels, Belgium.
E-mail addresses: Marie-Alice.Fraiture@sciensano.be (M.-A. Fraiture), Jolien.Daes@sciensano.be (J. D'aes), Andrea.Gobbo@sciensano.be (A. Gobbo), Maud.Delvoye@sciensano.be (M. Delvoye), Anne-Cecile.Meunier@cirad.fr (A.-C. Meunier), Julien.Frouin@cirad.fr (J. Frouin), Emmanuel.Guiderdoni@cirad.fr (E. Guiderdoni), Dieter.Deforce@UGent.be (D. Deforce), Charlotte.DeVogelaere@sciensano.be (C. De Vogelaere), Sigrid.Dekeersmaecker@sciensano.be (S.C.J. De Keersmaecker), Kevin.Vanneste@sciensano.be (K. Vanneste), Nancy.Roosens@sciensano.be (N.H.C. Roosens).
[1] Equal first author contribution.
[2] Equal last author contribution.

https://doi.org/10.1016/j.foodres.2025.117218
Received 14 March 2025; Received in revised form 4 July 2025; Accepted 2 August 2025
Available online 5 August 2025

organisms carrying relatively large-scale nucleotide sequence variations at a certain genetic location is very similar to those used for GMOs created through classical genetic engineering, as both involve detecting an unnatural association of sequences (i.e. the junction between the host genome sequence and a newly introduced foreign sequence) at a certain genetic location. However, when the result of genome-editing comprises one or more induced single nucleotide variation(s) (SNVs) spread over the genome, the development of such methods is more complex since the detection method should be designed to target a combination of potentially multiple precisely defined and restricted nucleotide region (s). In addition, since these SNVs can potentially also occur naturally or through random mutagenesis techniques, detecting on-target SNVs, i.e. SNVs at the position of expected GE SNVs, may not be sufficient to confirm they are indeed the result of genome editing techniques (Bertheau, 2021; European Commission Joint Research Centre & European Network of GMO Laboratories, 2023; Fraiture et al., 2022; Fraiture et al., 2023; Guertler et al., 2023). It is therefore essential to assess, using available databases cataloguing known natural sequence variations, whether the on-target SNVs of interest in a particular GE line are unique compared to all other lines within the same species, variety or cultivar. Regarding crop species, the most extensive public database currently available is provided by the 3000 Rice Genomes (3KRG) Project (The 3,000 rice genomes project, 2014), representing the most comprehensive genomic dataset for *Oryza sativa*. The 3KRG were obtained through Illumina resequencing of 3024 germplasm accessions, encompassing rice genetic diversity from 89 countries, representing the five main subpopulations of rice (indica, aus, temperate japonica, tropical japonica and aromatic). Even though this database does not entirely reflect all existing genetic variation within the *O. sativa* species, identifying on-target SNVs as unique will be a valuable indicator for a high probability of inclusivity and exclusivity when identifying a GE rice line (European Commission Joint Research Centre & European Network of GMO Laboratories, 2023; Grohmann et al., 2019; Guertler et al., 2023; Ichihara et al., 2023). To address the challenges of unambiguously distinguishing one GE line from all other existing lines, targeting only the on-target SNV(s) introduced by genome editing techniques is thus generally insufficient, and the potential added value of supplementing this detection strategy by targeting additional key genetic elements needs to be considered. This can include different markers (European Commission Joint Research Centre & European Network of GMO Laboratories, 2023; Fraiture et al., 2023). Firstly, with genome editing techniques like CRISPR-Cas systems, the presence of a short Protospacer Adjacent Motif (PAM) near the on-target site is usually a prerequisite for recognizing the specific DNA region of interest and activating Cas nucleases for editing (Collias & Beisel, 2021; European Commission Joint Research Centre & European Network of GMO Laboratories, 2023). Secondly, off-target modifications, especially at locations with significant sequence similarity to the intended on-target site, as well as unintended modifications in the vicinity of the on-target site, may be introduced by genome editing techniques. Such observations can help to determine whether a certain organism line should be classified as GMO (Bertheau, 2021; Grohmann et al., 2019; Klees et al., 2022; Shillito et al., 2021; Sturme et al., 2022; Tang et al., 2018; Yang et al., 2022). Finally, identifying genetic targets associated with a particular genomic background, such as a specific set of SNVs unique to a species, variety or cultivar, is also essential. The development and public access to curated databases containing high-quality sequences that encompass the full diversity of genotypes, cultivars, and varieties for each species of interest are, however, crucial for comparing a specific line to reference genomes, including elite lines from breeders. However, generating such databases remains highly challenging. Moreover, new genetic variation may accumulate naturally in the offspring of GE lines during vegetative propagation or as result of crossing with other lines (Bertheau, 2021; European Commission Joint Research Centre & European Network of GMO Laboratories, 2023; Fraiture et al., 2023; Guertler et al., 2023; Shillito et al., 2021; Sturme et al., 2022; Zhu et al., 2016).

In this study, as a proof of concept, we investigated for the first time the potential of utilizing such key genetic elements to facilitate the detection of GE organisms. We aimed to identify key genetic elements and optimally combine them to create a unique genetic fingerprint capable of distinguishing a specific GE line from all other organism lines. As a case study, a GE rice line carrying one on-target SNV introduced by CRISPR-Cas technology was characterized by whole-genome sequencing to develop and assess the proposed approach. To efficiently generate a unique genetic fingerprint, we explored the CRISPR/Cas-associated PAM sequence, potential off-target modifications, as well as machine learning tools and algorithms to identify cultivar-specific SNVs, considering distance between the key genetic elements and the on-target site(s), as well as their position, frequency, distribution, ploidy, and allelic variation. Recent studies have previously highlighted the potential of machine learning for designing SNV-based markers to differentiate species or varieties, including within the plant kingdom (European Commission Joint Research Centre & European Network of GMO Laboratories, 2023; Fraiture et al., 2023; Riza et al., 2023; Yuan et al., 2022).

To experimentally evaluate if the detection of the generated unique fingerprint could comply with European standard requirements for GMO detection methods (Marchesi et al., 2015), a targeted high-throughput sequencing strategy was proposed and designed in this study, including prior PCR-based enrichment of key genetic elements, and tested on several rice samples containing the GE Nipponbare rice line at varying levels allowing to assess the method sensitivity. Resulting sequencing data were analysed using an in-house pipeline created to identify SNVs, both indels (small insertions and deletions) and substitutions (Fraiture et al., 2023). Although less commonly used by GMO enforcement laboratories compared to real-time PCR and digital PCR, the targeted high-throughput sequencing approach was selected for its two main advantages of providing detailed sequence information and enabling high multiplexing capacity allowing the simultaneous detection of multiple key genetic elements that comprise the unique genetic fingerprint. While the added value of targeted high-throughput sequencing has been demonstrated in various fields, including the detection of classical GMOs and, more recently the detection of on-target SNVs in GE lines, to the best of our knowledge, this approach has not yet been applied to identify and differentiate a GE organism line from other organism lines using a unique genetic fingerprint (Boutigny et al., 2020; Debode et al., 2019; Fraiture et al., 2018; Fraiture et al., 2023; Fraiture, Herman, De Loose, et al., 2017; Fraiture, Herman, Papazova, et al., 2017; Fraiture, Papazova, et al., 2019; Fraiture, Ujhelyi, et al., 2019; Holst-Jensen et al., 2016; Jo et al., 2021; Košir et al., 2017; Kovalic et al., 2012; Liang et al., 2014; Onda et al., 2018; Pallarz et al., 2023; Saltykova et al., 2022; Shillito et al., 2021; Shirasawa et al., 2016; Wahler et al., 2013; Willems et al., 2016). Based on the experimental results obtained in this study, the feasibility of the proposed strategy, regarding the minimum performance requirements for GMO detection methods were investigated (Marchesi et al., 2015). Additionally, its potential to support biotech breeders, the Competent Authorities and the GMO enforcement laboratories will be discussed.

2. Methods

2.1. Plant materials

Seeds from two different rice (*Oryza sativa* L.) lines were used, including a GE japonica rice line Nipponbare (NipGE) and its wild-type line (NipWT). NipWT seeds were originally kindly provided by Pr. M. Yano of the Rice Genome Sequencing Project in Tsukuba, Japan. The NipGE line, not commercially available, carries a homozygous single adenosine insertion in *OsMADS26* (Os08g02070) introduced into the coding region near the start codon, specifically between genomic positions 679,646 and 679,647 on chromosome VIII (AP014964.1), conferring putative biotic stress resistance and abiotic stress tolerance

(Fraiture et al., 2023). This NipGE line was generated using CRISPR/Cas9 technology with protospacer sequence 5'-GAACCGGGTTCTC-GATGCGA-3′, and PAM 5'-NGG-3′. The presence of the expected SNV at the on-target site has been previously confirmed in both NipGE and NipWT materials using ddPCR as well as conventional PCR followed by Sanger sequencing (Fraiture et al., 2022; Fraiture et al., 2023).

DNA from rice seeds was extracted using a CTAB-based procedure (International Standard ISO 21571, 2005) in combination with the Genomic-tip20/G kit (QIAGEN) as described previously (EURL, 2006; Fraiture et al., 2013). DNA concentration was measured by fluorometry using Qubit3.0 (ThermoFisher Scientific). DNA purity was assessed based on A260/A280 and A260/A230 ratios provided by spectrophotometry using NanoDrop2000 (ThermoFisher Scientific). DNA integrity was confirmed by electrophoresis using TapeStation4200 with associated Genomic DNA ScreenTapes and reagents (Agilent).

2.2. Generation of the unique genetic fingerprint

2.2.1. Whole-genome sequencing

Whole-genome sequencing was individually performed on the NipGE line and its NipWT line at NXTGNT (Belgium). For each rice line, 4 technical replicates were included, each with a standard DNA input of 1 μg. The whole-genome sequencing library was prepared PCR-free using the NEBNext Ultra II library prep kit (New England Biolabs), with size selection aiming at an average insert size of 450 bp, according to the manufacturer's instructions. Sequencing was performed on an Illumina NovaSeq6000 S4 system in 150 bp paired-end mode, according to the manufacturer's instructions. Libraries were sequenced accross two lanes of two independent sequencing runs, with four samples, two NipGE and two NipWT, per lane. Quality control of the data was performed with FastQC 0.11.7 (Table S1).

2.2.2. Germline variant calling

A bioinformatics pipeline was developed following the Genome Analysis Toolkit (GATK) version 4.1.9.0 best practices (Van der Auwera & O'Connor, 2020) to analyze the sequencing data from raw reads to a high-quality variant call-set. FASTQ files containing read data of individual samples were first aligned to NipRefSeq (IRGSP-1.0, GenBank Accession GCA_001433935.1), using BWA-MEM v. 0.7.17 (Li et al., 2009) with the option 'deterministic_aln' set to 100,000,000. Read groups required for downstream analyses were added to the reads by Picard AddOrReplaceReadGroups v. 2.23.3 (https://broadinstitute.github.io/picard) with the options RG_library: 'unknown', RG_platform: 'illumina', RG_center_name: 'unknown', create_index: 'true'. Duplicate reads were then marked with Picard MarkDuplicates v. 2.23.3 with the options validation_stringency: 'SILENT', optical_duplicate_pixel_distance: 2500, assume_sort_order: 'queryname', clear_dt: 'false', add_pg_tag_to_reads: 'false', max_records_in_ram: 10,000,000. During this step, the alignments were sorted by reads coordinate order and indexed by Picard SortSam v. 2.23.3 with the options sort_order: 'coordinate', create_index: 'true', create_md5_file: 'true', max_records_in_ram: 10,000,000. NM tags set by BWA-MEM were recalculated using Picard SetNmMdAndUqTags v. 2.23.3 with the option: create_index: 'true', for consistency with downstream tools. The depth and breadth of coverage for the final alignments were calculated based on the output of samtools 1.17 depth, including reads flagged as duplicates, while the evenness of coverage, expressed as the fold80, was calculated based on the output of picard/2.18.14 CollectRawWgsMetrics (Table S1). Variant calling was performed using GATK (v4.1.9.0) HaplotypeCaller with default settings. The variant profiles of the NipGE and NipWT samples were analysed and compared with vcftools 0.1.16 and bcftools 1.17, only taking into account the chromosome and plastid scaffolds while excluding unplaced contigs, to identify variants shared by the NipGE and NipWT lines, as well variants unique to either the NipGE or NipWT lines (Danecek et al., 2011; Danecek et al., 2021) (Table S2).

2.2.3. Screening for off-target modifications in the NipGE line

For the prediction of potential off-target recognition sites of the CRISPR/Cas9-derived nuclease, Cas-OFFinder 3.0.0b3 (Bae et al., 2014), was employed, with NipRefSeq (Genbank Accession GCA_001433935.1), the protospacer sequence, and the PAM as input, and allowing up to 2 RNA or 2 DNA bulges and up to 3 mismatches. The 287 predicted potential off-target sites (Table S3) were compared to the set of variants unique to the NipGE samples to identify variants that could potentially be the result of off-target modification due to the CRISPR/Cas9 procedure.

2.2.4. Design of Nipponbare cultivar-specific SNV-fingerprint

Genotypic data for *O. sativa* was derived from the 3,000 Rice Genomes (3KRG) project (The 3,000 rice genomes project, 2014), The full 32 million 3KRG SNP dataset was retrieved from the SNP-seek database (https://snpseekv3.irri-e-extension.com/v2/_download.zul, Accessed in Dec 2022), together with a previously selected subset of the 3KRG SNP dataset, comprising 160,000 high-quality SNP positions (Wang, Agosto-Pérez, et al., 2018; Wang, Mauleon, et al., 2018). In agreement with these findings, one 3KRG accession (IRIS_313–8921) was dropped from the dataset as it contained excessive missing data.

The full 3KRG SNP dataset was employed to investigate the presence of the on-target SNV, as well as the conservation level of the PAM next to the on-target SNV site. To explore the feasibility of the design of a single PCR assay targeting a region encompassing both the GE site and the cultivar-specific marker, the full 3KRG SNP dataset was screened with an in-house developed search algorithm to identify the minimal genomic region that included the GE site, while also encompassing sufficient genetic variation to discriminate the Nipponbare accessions from all the other germplasm accessions in the database. This search strategy included a sliding and iteratively increasing window approach, starting from the first SNP of CHR8 in the database until the last SNP preceding the GE SNV, and progressing by increments of 200 SNPs. For each starting point, an optimization strategy was followed consisting of multiple iterations of screening and eliminating accessions with a different genotype at the added SNP position, followed by increasing the window by 1 SNP, until only one accession, i.e. the Nipponbare accession, remained.

A second approach aimed to identify a minimal SNV fingerprint allowing to distinguish the Nipponbare cultivar accessions from all the other germplasm accessions in the database, without restrictions regarding the proximity of the selected SNVs to the GE site. A filtered dataset retaining the SNPs at the 160,000 high-quality positions was used as input to perform feature selection with the Python machine learning package scikit-learn v1.1.3. Missing data – including degenerate bases - in the filtered SNP dataset was handled conservatively by replacing these values with the genotype of the Nipponbare accession at the corresponding SNP index. An array of observation labels was generated by setting the label for the Nipponbare cultivar to "Y" and all other labels to "N". Following transformation of the features dataset and labels array, univariate feature selection was performed with the SelectKBest function, with chi2 as the scoring function. For the resulting set of 500 best scoring features (i.e. SNP positions) (Table S4), the genotype of the NipGE line and the NipWT line was checked to verify that they shared this genotype with the Nipponbare accession in the 3KRG database. The selected features were further investigated with an inhouse script to perform filtering of the 3,023 germplasm accessions in the dataset according to their genotype for each possible pair of the 500 SNP positions (Table S5).

2.3. GE line identification by high-throughput amplicon sequencing

2.3.1. Sample preparation

DNA from rice lines was used to prepare samples n°1–12 (Table S6). For samples n°1–7, DNA from NipGE was used for serial dilutions in DNA/RNA nuclease-free distilled water (ThermoFisher Scientific)

ranging from ~14,000 to ~14 estimated haploid genome copies. Different amounts of DNA from NipGE, ranging from ~12,600 to ~126 estimated haploid genome copies, were mixed, for sample n°8–11, with DNA from NipWT, ranging from ~13,874 to ~1400 estimated haploid genome copies. For sample n°12, only DNA from NipWT was used.

2.3.2. Conventional PCR

For the OsMADS26 method, a pair of previously designed primers was used (Table S7) (Fraiture et al., 2022, 2023). A set of new primers was manually designed, checking in parallel oligonucleotide parameters with IDT OligoAnalyzer tool, for the SNV1, SNV2, SNV3 and SNV4 methods (Table S7). For compatibility with Illumina sequencing, all these primers were then supplemented with Illumina overhang adapter sequences (Fraiture et al., 2023). The PCR amplification of each method, using TapeStation4200 with associated D1000 ScreenTapes and reagents (Agilent), and their sequence identity, using USB ExoSAP-IT PCR Product Cleanup (Affymetrix) and sequence determination on a Genetic Sequencer 3500 (ThermoFisher Scientific) as previously described (Fraiture et al., 2022), were confirmed (Tables S8–9).

A standard 25 µl reaction volume was used, including 1× KAPA HiFi HotStart ReadyMix (Roche) and 900 nM of each primer (Eurogentec). The PCR conditions were set as previously described (Fraiture et al., 2023). Using different combinations of methods, three different PCR assays were designed and tested: (i) Triplex-1 PCR, using the OsMADS26, SNV1 and SNV2 methods; (ii) Triplex-2 PCR, using the OsMADS26, SNV3 and SNV4 methods; and (iii) Pentaplex PCR, using the OsMADS26, SNV1, SNV2, SNV3 and SNV4 methods. The samples were tested in duplicate (n°1–4, 8–12) or in triplicate (n°5–7) (Table 1). Each PCR run included a NTC (No Template Control).

2.3.3. Library preparation, amplicon sequencing and data analysis

Each PCR product was purified by Agencourt Ampure XP (Beckman Coulter), using 50 µl of AMPure XP beads (Beckman Coulter) and 15 µl of elution UltraPure DNase/RNase-Free Distilled Water (Invitrogen). The amplicon sequencing libraries were prepared according to the manufacturer's instructions. Per sequencing run, a total of 12 samples, earlier individually barcoded, were pooled. Sequencing was carried out on an Illumina iSeq 100 system using the V2 chemistry (300 cycle), obtaining 150 bp paired-end reads.

The quality of raw sequencing data was evaluated using FastQC 0.11.7 (Andrews, 2010) with default settings. The allelic frequency (AF) of the target SNVs in the samples was estimated as previously described (Fraiture et al., 2023), with a number of modifications. Variants were called with respect to two references, NipRefSeq (Genbank Accession GCA_001433935.1), and the genome assembly of *O. sativa* cv. Kitaake (GenBank Accession GCA_009797565.1), to obtain AF estimates for the non-Nipponbare and Nipponbare genotypes, respectively. Additionally, the Kitaake reference sequence was modified in silico to introduce the nucleotide insertion of the NipGE line, thus allowing to obtain an AF estimate of the WT genotype. Both references were indexed with Samtools 1.17 (Li et al., 2009) and Bowtie2 2.3.4.3 (Langmead & Salzberg, 2012), while Picard 3.1.1 was used to generate a dictionary of the indexed reference FASTA file. The raw sequencing data was pre-processed using Trimmomatic 0.38 (Bolger et al., 2014) with the following settings: ILLUMI-NACLIP:NexteraPE-PE.fa:2:30:10:1:TRUE, LEADING:10, TRAILING:10, SLIDINGWIN-DOW:4:20, MINLEN: 40, after which the trimmed reads were aligned to the reference with Bowtie2 with the "–end-to-end" and "–very-sensitive" settings. The resulting alignments were further processed as previously described (Fraiture et al., 2023). Next, variants were called with LoFreq 2.1.3.1 (Wilm et al., 2012) using the options "–call-indels", and "–sig 0.001". The output calls were limited to the genomic regions targeted by the PCR assays (see section 2.3.2), excluding the primer sequences. The resulting VCF file was filtered with LoFreq, setting the minimal allelic frequency to 0.1 %, the minimum SNV and indel quality both to 1,000, and the minimum coverage to 500.

3. Results

3.1. GE rice line characterization by high-throughput whole-genome-sequencing

The model GE line for this study was a non-commercially available GE line of rice *O. sativa* (japonica group) cv. Nipponbare carrying a homozygous single adenosine insertion in *OsMADS26*, conferring putative biotic stress resistance and abiotic stress tolerance (Fraiture et al., 2023). The GE SNV is located after bp 679,646 in CHR8, based on the reference genome assembly of *O. sativa* cv. Nipponbare (Genbank Accession GCA_001433935.1), hereafter referred to as NipRefSeq.

To design a unique genetic fingerprint to identify and unambiguously distinguish the target GE line from any other line, an in-depth genomic characterization of the GE line and the wild-type (WT) cv. Nipponbare, hereafter referred to as NipGE and NipWT, respectively, was performed. Using high-throughput whole-genome sequencing with four technical replicates per rice line, high-quality and high-depth sequencing coverage of at least 300× per replicate was obtained for the NipGE and NipWT samples, providing an even and comprehensive coverage of the entire rice reference genome (Table S1).

Overall, the NipGE and NipWT samples shared 32,537 heterozygous as well as homozygous variants, including 7,599 indels, compared to NipRefSeq. 283 variants, including 180 indels, were shared by all the NipWT replicate samples but were absent from the NipGE replicate samples, while vice versa, 1,056 variants, of which 519 indels, including the GE SNV site, were shared by all the NipGE replicate samples but did not occur in the NipWT replicate samples (Table S2). The observed variant profiles are in line with findings of Tang et al. (Tang et al., 2018) concerning the frequency of spontaneous mutations occurring in rice. In general, around 30 to 50 spontaneous mutations are expected in the next rice generation due to seed propagation through harvesting, while this number increases to around 200 mutations per generation when the plant material is derived from tissue culture used for propagating the plant material or transferring DNA encoding the CRISPR-Cas9 machinery (Miyao et al., 2012; Tang et al., 2018).

3.2. Screening of CRISPR/Cas9-associated potential key genetic elements

Based on the obtained variant profiles, we assessed the potential of different genetic elements to serve as key element of the genetic fingerprint for the NipGE line. Firstly, the GE SNV site was screened in the full 3KRG SNP dataset, and found to be unique to the NipGE line, compared to the germplasm accessions in the 3KRG database (Fraiture et al., 2022, 2023), highlighting that the GE SNV itself could serve as a valuable key element of the genetic fingerprint. However, the PAM site associated with the GE SNV site was conserved across all the accessions in the 3KRG database, except for one single accession (IRBB60) in which a heterozygous genotype was reported for one of the three PAM-nucleotides. As the PAM site was highly conserved, it was not useful as a marker for the NipGE line in this case.

Next, the presence of potential off-target modifications resulting from the CRISPR/Cas9 gene editing procedure was investigated. None of the variants unique to the NipGE line overlapped with one of the predicted potential off-target sites (Table S3), indicating that off-target modifications had not occurred. Recent studies reported that off-target modifications only occurred in predicted off-target sites that had at most 2 or 3 mismatches compared to the target (protospacer) sequence, with a larger off-target reducing effect for mismatches close to the PAM (Sturme et al., 2022). In this study, the closest predicted off-target sites had at least 2 mismatches and 1 insertion or deletion compared to the target sequence. As no off-target modifications were observed, it was not useful as a marker for the NipGE line in this case.

3.3. In-silico design of a unique genetic fingerprint for the NipGE line

Given the absence of PAM or off-target modifications to use in the GE-specific fingerprint, we focused our efforts on involving the identification of the cultivar-specific genetic background. A first strategy was explored, based on the full 3KRG SNP dataset, to generate a unique genetic fingerprint for the NipGE line that contained, within close proximity of each other, both (i) the GE SNV site introduced by the CRISPR-Cas9 technology as well as (ii) a set of SNVs allowing to distinguish the Nipponbare cultivar from all other accessions in the database. However, for this strategy, it appeared that meeting both criteria required covering the genetic variation within a region of approximately 660 kbp in length, i.e. from bp 671,107 to 1,330,995 on CHR8 (AP014964.1). This strategy was therefore discarded in the present study as this sequence length far exceeds the technical limits for generating a single PCR amplicon.

An alternative strategy to generate a unique genetic fingerprint for the NipGE line was explored. In addition to the identification of the GE SNV site introduced by the CRISPR-Cas9 technology, the proposed strategy aimed to combine a minimal set of SNVs to constitute an unambiguous marker for the Nipponbare cultivar, without restrictions regarding proximity of the selected SNVs to the GE SNV site. With this aim, a feature selection based strategy was employed starting from a high quality dataset with 160,000 SNPs, not considering indels, by which a subset of 500 SNVs was retrieved (Table S4). The SNVs were ranked according to their ability to discriminate the Nipponbare cultivar from the other rice accessions in the database, expressed as the number of accessions sharing the SNV with the Nipponbare accession. Since rice is an inbreeding species and rice varieties typically display a high level of homozygosity, the genotype at these SNV positions was assumed to be homozygous for the reference allele (Nipponbare) as well as for the alternative allele.

For the 500 selected SNV positions, conservation of the allele was confirmed in NipRefSeq as well as in the NipGE and NipWT lines, verifying their potential as genetic markers for the cultivar background of these lines. However, not one of these SNVs could be used individually as a marker for the Nipponbare cultivar, since even for the highest ranking SNV at least 14 other rice accessions shared the genotype of Nipponbare, or had a missing genotype at this position (Table S4). Therefore, we screened all possible SNV-pair combinations from the 500 SNV-subset, resulting in 145 different 2-SNV barcodes that each uniquely identified the Nipponbare accession. These 145 barcodes included 99 different SNVs (i.e., a SNV could be employed in different barcodes) from almost all chromosomal arms of the rice genome, with

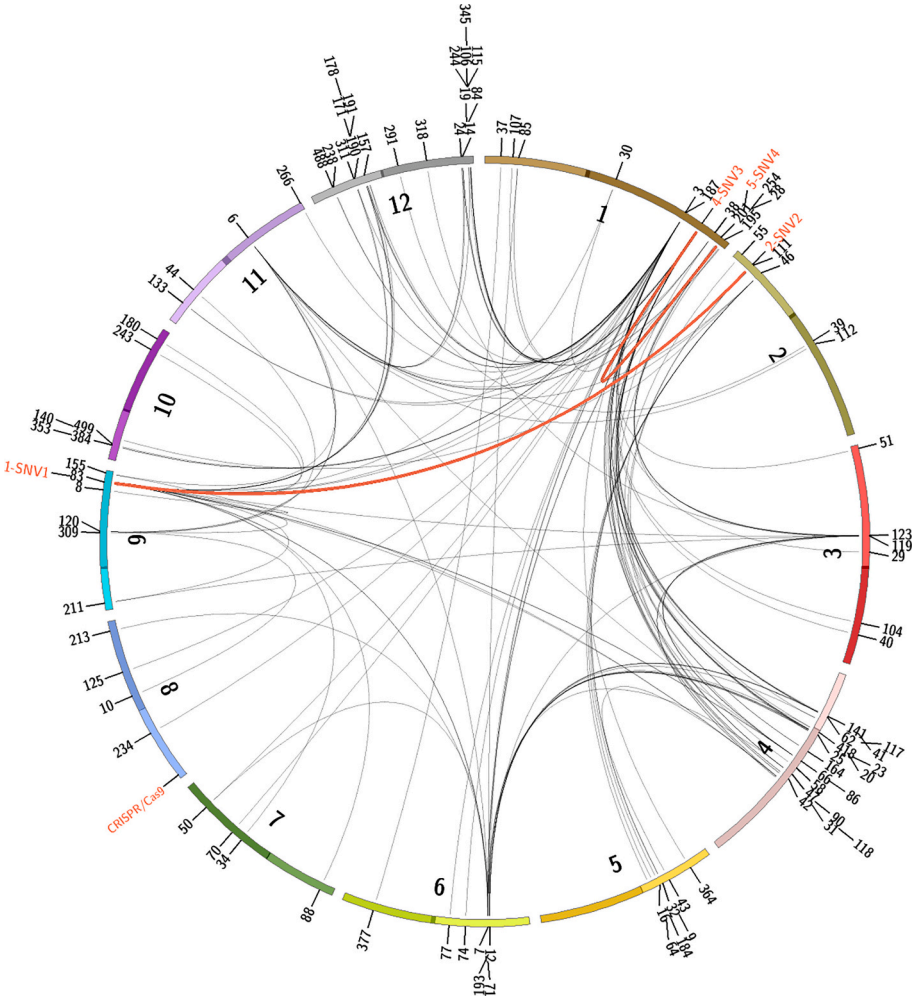

Fig. 1. Schematic representation of the genetic fingerprint for the NipGE line. Map of NipRefSeq (Genbank Accession GCA_001433935.1), depicting for each chromosome the approximate position of the centromere (Jiang et al., 2023) (darkest shade), with the upstream and downstream arm distinguished with a lighter and darker shade, respectively. Each connection on the plot represents one of the 145 cultivar-specific 2-SNV barcodes. The 99 involved SNVs are annotated with their rank according to their ability to discriminate the Nipponbare cultivar from the other rice accessions in the 3KRG database (Table S4). The two 2-SNV barcodes selected in this study for the PCR multiplex assays, and the position of the on-target SNV site (CRISPR/Cas9) of the NipGE line, are highlighted in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the exception of the long arm of CHR5. Nevertheless, Fig. 1 highlights considerable differences in the frequency of these SNVs depending on the chromosome, ranging from 4 SNVs on CHR7 to 18 SNVs on CHR12.

For each of these barcodes, the combined genotype at the two SNV positions was unique to the Nipponbare cultivar (Fig. 1, Table S5) compared to the other rice accessions in the 3KRG database. However, since this database does not contain all the existing rice varieties, in particular with respect to commercial rice cultivars, we envisage to combine several 2-SNV barcodes to obtain, in combination with the on-target SNV, a genetic fingerprint for the NipGE line that would be highly unlikely to be shared with any other existing rice variety.

3.4. Experimental evaluation of the unique genetic fingerprint using targeted high-throughput sequencing

To unambiguously detect and identify the NipGE line, we propose a targeted high-throughput sequencing approach, including a prior PCR enrichment step, allowing to simultaneously target all key genetic elements of the *in-silico* generated genetic fingerprint specific to the NipGE line, i.e., the GE position and a number of the cultivar-specific 2-SNV barcodes.

As a proof-of-concept, this approach was developed and experimentally evaluated using two of the 2-SNV barcodes. These two barcodes were selected based on the rank of the two SNVs they were composed of, prioritizing the highest ranked SNVs according to their ability to discriminate the Nipponbare cultivar from the other rice accessions in the 3KRG database (Fig. 1). However, all the other barcodes could in theory also provide an unambiguous identification of the Nipponbare cultivar, regardless of the rank of their SNVs.

Based on the two selected 2-SNV barcodes, three different multiplex PCR assays were designed and tested. The first assay, a triplex PCR (triplex-1), targeted a genomic region encompassing the GE SNV site (OsMADS26 method) along with two additional genomic regions encompassing a Nipponbare cultivar-specific SNV (SNV1 and SNV2 methods, targeting CHR9:21,176,062 and CHR2:3,247,472 on NipRefSeq, respectively), collectively forming a part of the unique genetic fingerprint of the NipGE line (subset fingerprint-1). The second assay, another triplex PCR (triplex-2), also targeted a genomic region encompassing the GE SNV site (OsMADS26 method) but included two other Nipponbare rice cultivar-specific SNVs (SNV3 and SNV4 methods, targeting CHR1:37,778,084 and CHR1:41,879,706 on NipRefSeq, respectively), together representing another part of the unique genetic fingerprint of the NipGE line (subset fingerprint-2). The third assay, a pentaplex PCR, combined the OsMADS26 method with the SNV1, SNV2, SNV3 and SNV4 methods (subset fingerprint-3), thereby integrating the Nipponbare cultivar-specific SNVs used in both triplex-1 and triplex-2. For method performance evaluation, all these multiplex PCR assays were applied on different rice samples in duplicate (n°1–4, 8–12) or in triplicate (n°5–7). Following high-throughput sequencing using an Illumina iSeq 100 system, the raw data were analysed through an in-house bioinformatic pipeline, providing the percentages of the observed allele frequency (AF) for each method composing each multiplex PCR assay (Table 1).

Firstly, method sensitivity was evaluated using serial dilutions from the NipGE line, ranging from 14,000 to 14 estimated haploid genome copies (samples n°1–7). For all these NipGE samples, the GE SNV site was detected using the OsMADS26 method in the triplex-1 (subset fingerprint-1), triplex-2 (subset fingerprint-2) and pentaplex (subset fingerprint-3) PCR assays (Table 1). Additionally, the targeted 2-SNV barcodes, part of the in-house generated Nipponbare cultivar-specific marker, were detected in all samples using both (i) SNV1 and SNV2 methods for the triplex-1 PCR assay (subset fingerprint-1), (ii) SNV3 and SNV4 methods for the triplex-2 PCR assay (subset fingerprint-2), or (iii) SNV1, SNV2 SNV3 and SNV4 methods for the pentaplex PCR assay (subset fingerprint-3) (Fig. 1, Table 1). According to these results, the NipGE line was successfully detected in all these samples using the three designed multiplex PCR assays. The proposed targeted high-throughput sequencing approach demonstrated also its compatibility with low levels of the target (i.e., 0.9 % and 0.1 %), including levels below 25 copies, in alignment with the minimum performance requirements for GMO detection methods (Marchesi et al., 2015). Moreover, the observed AF percentages for all PCR methods were generally close to 100 %, in line with the expected homozygosity of the cultivar-specific SNVs (Table 1).

Secondly, method performance was tested using rice samples composed of DNA from NipGE at high and low contamination levels mixed with NipWT (samples n°8–11). For each multiplex PCR assay, the GE SNV site as well as the SNV corresponding to the unmodified on-target site were both detected using the OsMADS26 method in all these mixture samples, as expected according to the sample composition (Table 1). However, using the pentaplex PCR assay, the GE SNV site was only observed in one of the two replicates for mixture sample n°10, highlighting the critical importance of analytical replicates for ensuring the reliability of the results, especially in the proposed approach (Table 1C). Regarding the Nipponbare 2-SNV barcodes using the SNV1-SNV4 methods, the expected SNVs corresponding to the Nipponbare and non-Nipponbare rice cultivars were observed in all these mixture samples with all multiplex PCR assays (Table 1). Based on these results, the presence of the NipGE line was detected, even at low contamination levels, emphasizing the applicability of the proposed targeted hight-throughput sequencing approach to target GE lines in food mixtures. In addition, in all samples, the observed AF percentages were in general closely matching with the expected values regarding the sample composition for all PCR methods investigated (Table 1).

Finally, a rice sample, containing only DNA from NipWT (sample n°12) was included in the analysis. Using the OsMADS26 method, only the SNV corresponding to the unmodified on-target sequence was detected with all multiplex PCR assays, as expected by the absence of the NipGE line in this sample (Table 1). Using the SNV1 and SNV2 methods within the triplex-1 and pentaplex PCR assays as well as the SNV3 and SNV4 methods within the triplex-2 and pentaplex PCR assays, the SNV combinations associated with the Nipponbare rice cultivars were observed in sample n°12 as expected. Based on these AF results, the absence of the NipGE line in the sample was demonstrated and the observed AF percentages were in general closely matching with the expected values regarding the sample composition for all PCR methods investigated (Table 1).

4. Discussion

Unambiguously identifying a specific GE line from among other lines is not trivial. To address this complex challenge, we utilized in-house whole-genome sequencing data, publicly available sequence databases, and machine learning tools to explore the possibility of generating a unique genetic fingerprint. As a case study, such fingerprints, composed of several key genetic elements, were evaluated and developed specifically in this study for the NipGE line (Fig. 2).

To construct a unique genetic fingerprint for the NipGE line, various types of key genetic elements were considered. As minimal requirement, the GE SNV site introduced by genome editing was first included. Since this GE SNV consists of an indel, it highlights the applicability of the proposed approach to both small insertions and deletions. From a comparative perspective, the presence of an inserted nucleotide in NipGE (e.g., +A) can be equivalently interpreted as a deletion in NipWT (e.g., −A), depending on the reference genome used for comparison. Such nucleotide variation harboured by the NipGE line was not observed in any other rice line within the full SNP dataset of natural rice diversity catalogued in the publicly available 3KRG database, which comprises over 3000 rice accessions (Fraiture et al., 2022, 2023). This finding at first glance could suggest that detecting this GE SNV site alone may be sufficient for the specific identification of the NipGE line. However, while the likelihood is low and no demonstrated evidence exists, the presence of this nucleotide variation in a rice line not included in the

3KRG database, despite the extensive size of included rice lines, cannot be entirely ruled out. This underscores the need to target additional key genetic elements for constructing the genetic fingerprint.

Therefore, the potential for inclusion of PAM and off-target sites in the genetic fingerprint was examined. However, the PAM site was highly conserved among rice varieties in the 3KRG. While PAMs could be considered as one potential type of key genetic element to strengthen GE organism detection, their use is also limited since PAMs are not required by all genome editing techniques. Additionally, even for initially PAM-dependent systems, an increasing number of PAM-free alternatives are emerging to expand the range of editable sites (Collias & Beisel, 2021; European Commission Joint Research Centre & European Network of GMO Laboratories, 2023). Furthermore, no off-target sites were detected in the NipGE line, consistent with the advancements in genome editing technologies that have drastically minimized or even eliminated unintended modifications. Ongoing technological developments continue to refine genome editing technologies and enhance precision, further reducing the likelihood of off-target effects substantially (Bertheau, 2021; Grohmann et al., 2019; Klees et al., 2022; Shillito et al., 2021; Sturme et al., 2022; Tang et al., 2018; Yang et al., 2022). Therefore, additional key genetic elements were investigated to establish the unique genetic fingerprint, in particular the potential use of SNVs specifically associated with the specific genetic background of the NipGE line.

With this goal, the possibility of amplifying a single large PCR amplicon carrying multiple key genetic elements was explored. By utilizing the genetic variation surrounding the GE SNV site, a single PCR amplicon could potentially serve as a unique genetic fingerprint for a specific GE line. This strategy was however not pursued in the present study due to the sequence length far exceeding the technical limitations for generating a single PCR amplicon. Nevertheless, even if the key genetic elements are distributed across different PCR fragments, as designed in this study, the analysis of mixture samples remains achievable. By leveraging machine learning tools, it is indeed possible to assess whether the detection of a particular combination of key genetic elements can be attributed to either a single line or a limited subset of lines, providing a valuable indicator for determining a high probability of exclusivity.

Considering the technical challenges associated with generating a single PCR amplicon carrying multiple key genetic elements, as an alternative approach to create a unique genetic fingerprint for the NipGE line, machine learning approaches and high-quality public databases

Table 1

Detection of the NipGE line in samples n°1–12 using triplex-1 (A), triplex-2 (B) and pentaplex (C) PCR assays combined with high-throughput sequencing. The composition of each sample, comprising NipGE and/or NipWT, is described, including the estimated genome copy number for each line. The samples were tested in duplicate (n°1–4, 8–12) or in triplicate (n°5–7) and the average amplicon sequencing depth is reported. For each subset fingerprint (1–3) tested, the observed allele frequency (AF) of each associated method (OsMADS26, SNV1, SNV2, SNV3 and/or SNV4) is indicated. For each method, the nucleotide variations expected for the NipGE line, the non-NipGE lines, the Nipponbare rice lines (Nip) or the non-Nipponbare rice lines (non-Nip) are indicated between brackets. The observed AF values at the GE SNV site specific to the NipGE line are marked in green while those corresponding to the WT nucleotide variation are highlighted in blue. The observed AF values related to the Nipponbare rice cultivar are shown in yellow. The presence of the Nipponbare rice cultivar is only marked as detected if the corresponding SNVs for both (or all four in case of the pentaplex) SNV methods are observed. Based on all these AF values, the detection or not of the NipGE line is respectively represented by "+" or "-". (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(A) Triplex-1 PCR assay followed by high-throughput sequencing

Samples			Subset fingerprint-1						
			OsMADS26 method		SNV1 method		SNV2 method		
N°	Composition	NipGE line detection ?	AF NipGE (+A)	AF non-NipGE (-A)	AF Nip (T)	AF non-Nip (G)	AF Nip (G)	AF non-Nip (C)	Aver. seq. depth
1	100 % NipGE ~14,000 copies	+ (2/2)	99.7%	0%	99.8%	0%	99.8%	0%	63,921
2	90% NipGE ~12,600 copies	+ (2/2)	99.6%	0%	99.8%	0%	99.9%	0%	61,186
3	50% NipGE ~7,000 copies	+ (2/2)	99.6%	0%	99.8%	0%	99.8%	0%	52,215
4	10% NipGE ~1,400 copies	+ (2/2)	99.6%	0%	99.8%	0%	99.7%	0%	52,138
5	5% NipGE ~700 copies	+ (3/3)	99.6%	0%	99.9%	0%	99.8%	0%	76,973
6	0.9% NipGE ~126 copies	+ (3/3)	99.2%	0%	99.8%	0%	99.8%	0%	33,883
7	0.1% NipGE ~14 copies	+ (3/3)	96.8%	0%	98.2%	0%	98.0%	0%	2,627
8	90% NipGE ~12,600 copies + 10% NipWT ~1,400 copies	+ (2/2)	93.9%	5.7%	99.6%	0%	99.7%	0%	63,528
9	50% NipGE ~7000 copies + 50% NipWT ~7000 copies	+ (2/2)	55.7%	43.5%	99.6%	0%	99.7%	0%	59,622
10	10% NipGE ~1,400 copies + 90% NipWT ~12,600 copies	+ (2/2)	8.8%	90.4%	99.7%	0%	99.7%	0%	56,547
11	0.9% NipGE ~126 copies + 99.1% NipWT 13,874 copies	+ (2/2)	0.7%	98.3%	100.0%	0%	99.9%	0%	139,898
12	100% NipWT ~14,000 copies	- (0/2)	0%	99.1%	99.6%	0%	99.6%	0%	59,100

(B) Triplex-2 PCR assay followed by high-throughput sequencing

(continued on next page)

Table 1 (*continued*)

Samples			Subset fingerprint-2						
			OsMADS26 method		SNV3 method		SNV4 method		
N°	Composition	NipGE line detection ?	AF NipGE (+A)	AF non-NipGE (-A)	AF Nip (T)	AF non-Nip (C)	AF Nip (G)	AF non-Nip (A)	Aver. seq. depth
1	100 % NipGE ~14,000 copies	+ (2/2)	99.6%	0%	100.0%	0%	99.8%	0%	107,370
2	90% NipGE ~12,600 copies	+ (2/2)	99.5%	0%	99.9%	0%	99.9%	0%	110,542
3	50% NipGE ~7,000 copies	+ (2/2)	99.4%	0%	100.0%	0%	99.9%	0%	102,889
4	10% NipGE ~1,400 copies	+ (2/2)	99.4%	0%	100.0%	0%	99.8%	0%	109,917
5	5% NipGE ~700 copies	+ (3/3)	99.3%	0%	99.9%	0%	99.9%	0%	117,501
6	0.9% NipGE ~126 copies	+ (3/3)	99.5%	0%	99.9%	0%	99.5%	0%	111,880
7	0.1% NipGE ~14 copies	+ (3/3)	97.2%	0%	93.3%	0%	90.3%	0%	21,366
8	90% NipGE ~12,600 copies + 10% NipWT ~1,400 copies	+ (2/2)	89.0%	10.4%	99.9%	0%	99.7%	0%	107,202
9	50% NipGE ~7000 copies + 50% NipWT ~7000 copies	+ (2/2)	48.2%	51.1%	99.9%	0%	99.7%	0%	109,827
10	10% NipGE ~1,400 copies + 90% NipWT ~12,600 copies	+ (2/2)	7.4%	91.8%	100.0%	0%	99.8%	0%	76,563
11	0.9% NipGE ~126 copies + 99.1% NipWT 13,874 copies	+ (2/2)	0.5%	98.4%	99.8%	0%	99.8%	0%	85,696
12	100% NipWT ~14,000 copies	- (0/2)	0%	98.8%	99.9%	0%	99.6%	0%	111,970

(C) Pentaplex PCR assay followed by high-throughput sequencing

Samples			Subset fingerprint-3										
			OsMADS26 method		SNV1 method		SNV2 method		SNV3 method		SNV4 method		
N°	Composition	NipGE line detection ?	AF NipGE (+A)	AF non-NipGE (-A)	AF Nip (T)	AF non-Nip (G)	AF Nip (G)	AF non-Nip (C)	AF Nip (T)	AF non-Nip (C)	AF Nip (G)	AF non-Nip (A)	Aver. seq. depth
1	100 % NipGE ~14,000 copies	+ (2/2)	99.9%	0%	99.9%	0%	99.8%	0%	99.9%	0%	99,8%	0%	65,534
2	90% NipGE ~12,600 copies	+ (2/2)	99.8%	0%	99.9%	0%	99.9%	0%	99.9%	0%	99.8%	0%	64,420
3	50% NipGE ~7,000 copies	+ (2/2)	99.7%	0%	99.9%	0%	99.8%	0%	99.9%	0%	99.8%	0%	65,866
4	10% NipGE ~1,400 copies	+ (2/2)	99.6%	0%	99.8%	0%	99.7%	0%	99.9%	0%	99.7%	0%	54,449
5	5% NipGE ~700 copies	+ (3/3)	99.0%	0%	99.9%	0%	99.3%	0%	99.9%	0%	99.7%	0%	54,337
6	0.9% NipGE ~126 copies	+ (3/3)	98.7%	0%	99.9%	0%	100%	0%	99.9%	0%	98.3%	0%	29,133
7	0.1% NipGE ~14 copies	+ (3/3)	89.0%	0%	98.0%	0%	96.7%	0%	92.8%	0%**	97.9%	0%	4,580
8	90% NipGE ~12,600 copies + 10% NipWT ~1,400 copies	+ (2/2)	94.1%	5.7%	99.8%	0%	99.8%	0%	99.9%	0%	99.7%	0%	74,812
9	50% NipGE ~7000 copies + 50% NipWT ~7000 copies	+ (2/2)	55.2%	44.2%	99.8%	0%	99.7%	0%	99.9%	0%	99.8%	0%	74,746
10	10% NipGE ~1,400 copies + 90% NipWT ~12,600 copies	(+) (1/2)	7.0%*	95.2%	99.7%	0%	99.7%	0%	99.8%	0%	98.7%	0%	56,442
11	0.9% NipGE ~126 copies + 99.1% NipWT 13,874 copies	+ (2/2)	0.6%	98.4%	100%	0%	99.9%	0%	99.8%	0%	99.9%	0%	65,997
12	100% NipWT ~14,000 copies	- (0/2)	0%	98.5%	99.5%	0%	99.6%	0%	99.8%	0%	98.7%	0%	55,515

* Not averaged as an AF value was observed in only one of the two replicates; ** Average of two replicates, as a false positive AF value (19.3 %) was observed in one of the three triplicates.

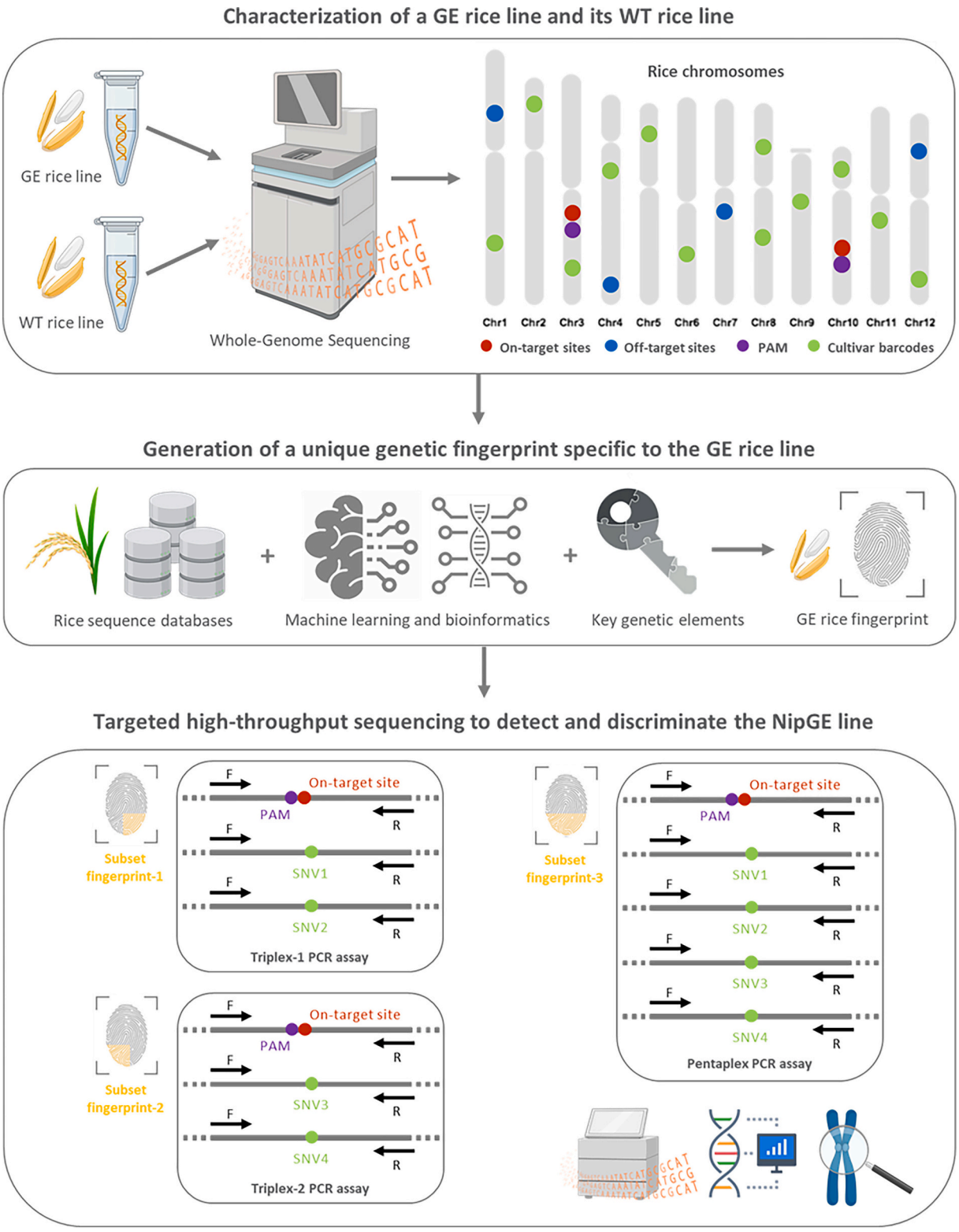

Fig. 2. Schematic representation of the proposed workflow applied in this study. First, a GE rice line and its WT rice line are whole-genome sequenced. Sequencing reads are used to compare and characterize at the genomic level the GE and WT rice lines, including the on-target SNV site(s), PAM(s), potential off-target sites and cultivar-specific barcodes composing the cultivar-specific marker. Second, using machine learning and bioinformatics tools, in-house characterized GE and WT rice lines and the publicly available 3KRG database, comprising over 3,000 rice accessions and encompassing the global natural rice diversity, are used to identify key genetic elements and subsequently to generate a unique genetic fingerprint specific to the GE rice line. On this basis, a targeted high-throughput sequencing approach, incorporating a prior PCR enrichment step, was proposed in this study to simultaneously target all key genetic elements comprising the unique genetic fingerprint specific to the NipGE line. As a proof-of-concept, this approach was developed to target a subset of these key genetic elements (the on-target SNV site and cultivar-specific barcodes). Three different multiplex PCR assays were designed and experimentally tested for their performance.

representing the global rice diversity (Mansueto et al., 2017; W. Wang et al., 2018) were exploited to identify a set of several independent Nipponbare cultivar-specific 2-SNV barcodes as being exclusively found in the Nipponbare rice lines (GE or not). Although computational approaches not based on machine learning, such as exhaustively considering all 2-SNV barcodes, could also have been applied in this study, machine learning offered substantial advantages. It provided high flexibility, allowing its application to both simple and complex cases. It also enabled the more efficient use of computational resources, which is an important consideration given the expanding size and complexity of genomic databases. Therefore, machine learning represents a flexible and powerful solution for detecting and identifying GE organisms.

Based on the genetic variation represented in the 3KRG database, each 2-SNV barcode could theoretically provide unambiguous identification of the Nipponbare cultivar. However, in practice potential misidentification is still possible, as commercial elite lines from breeders are not represented in 3KRG, although they are typically derived from crosses between different 3KRG cultivar lines (Bertheau, 2021). It is therefore not impossible that a single 2-SNV barcode might be present by chance, due to recombination or other events, in a commercial elite line, which would lead to a false positive identification of the Nipponbare cultivar genetic background. We therefore envisage to combine several 2-SNV barcodes in order to obtain an unambiguous marker for the Nipponbare cultivar, which would be very unlikely to occur through commercial breeding or natural events in any other rice line. As illustrated in Fig. 1, it would be feasible for instance to design a Nipponbare marker including SNVs distributed across nearly each arm of the rice chromosomes. Moreover, employing a redundant set of SNV-based cultivar-specific barcodes presents the advantage of addressing potential mutations in the target SNVs, which, although expected to be limited in elite lines from breeders, may appear in the offspring of GE lines through natural occurrence during vegetative propagation, self-pollination or crossing (Bertheau, 2021). Additionally, false negative results due to issues at the experimental level, leading to an incomplete fingerprint, could potentially be mitigated with a redundant set of barcodes.

Building on the designed fingerprints, we proposed a targeted high-throughput sequencing approach to identify and differentiate the NipGE line at the experimental level. This approach, which incorporates a prior PCR enrichment step, allows for the simultaneous targeting of all key genetic elements that constitute the designed unique genetic fingerprint (s) of the NipGE line. This PCR amplicon-based enrichment sequencing approach was chosen as it had been successfully applied previously to detect the on-target SNV site in the NipGE line, aligning with European standard requirements for GMO detection methods (Fraiture et al., 2023). As an alternative, a hybridization-based enrichment sequencing approach, using single-stranded nucleic probes complementary to the genomic regions of interest, may also be worth considering (Singh, 2022). As a proof-of-concept, the experimental feasibility of the proposed PCR amplicon-based enrichment sequencing approach was investigated using a subset of such key genetic elements. More precisely, three different multiplex PCR assays (triplex-1, triplex-2, pentaplex) were designed for the enrichment step, covering the GE SNV sites and Nipponbare cultivar-specific barcodes. All these PCR assays were applied on various samples containing high and low amounts of DNA from the NipGE line to assess the method sensitivity performance. These PCR assays were also applied on samples containing DNA from either NipWT only or mixed with high and low DNA amounts from NipGE. All final PCR products were subjected to high-throughput sequencing and then analysed using an in-house bioinformatic pipeline. The results indicated that the proposed targeted sequencing strategy was able to detect the NipGE line, even at trace level (i.e., 0.9 % and 0.1 %) as well as in mixtures with NipWT. The results also highlighted the flexibility of the proposed strategy, as different multiplex PCR assays compatible with the desired method performance may be created to target the same key genetic elements of interest. While the proposed targeted sequencing assays demonstrated satisfactory and promising performance in terms of specificity and sensitivity, this study remains however a proof of concept. Further experimental analyses of the method's performance will be necessary to evaluate whether it is entirely in line with minimum performance requirements for GMO detection methods (Marchesi et al., 2015). If the designed enrichment PCR assays do not meet the expected method performance standards, the proposed strategy, being a proof of concept, is not limited to the PCR assays developed in this study. It is a flexible approach, enabling the design of new PCR assays. Furthermore, this targeted high-throughput sequencing strategy is modular and can be supplemented with additional multiplex PCR assays targeting additional cultivar-specific barcodes. For instance, the successful implementation of a similar sequencing-based approach with high-multiplex PCR assays was recently illustrated in the public health sector for identifying SARS-CoV-2 variants (Ulhuq et al., 2023). Expanding the number of SNV-based cultivar-specific barcodes, distributed across different rice chromosomal arms, targeted by multiplex PCR assays is expected to further reduce the likelihood of false-positive or false-negative detection. The successful implementation of the triplex and pentaplex PCR assays in this study, covering up to five key genetic elements, highlights the feasibility of designing additional multiplex PCR assays to encompass a broader set of key genetic elements.

Regarding the application scope for the proposed strategy, although it was experimentally tested in this study using Nipponbare lines as case study, it is not expected to be exclusively limited to this specific rice cultivar. The proposed strategy is currently anticipated to be suitable for any known GE lines with a well-characterized and sequenced genetic background, like rice species, for which publicly available databases encompassing the main species diversity exist. The genetic fingerprints were generated using the full rice 3KRG database, a comprehensive and well-established genomic resource for *O. sativa* capturing the global natural rice diversity. This database is expected to expand to 15 K accessions in the near future and to 100 K in the long term, further reinforcing confidence in the proposed strategy for creating a unique genetic fingerprint. In parallel, additional efforts to capture global genetic variation of rice are underway, such as the generation of a variant map based on 10,548 accessions, including those from the 3KRG datadase, using publicly available WGS data from cultivated and wild rice (T. Wang et al., 2023). While the proposed approach is anticipated to have promising applicability, further validation using additional rice cultivars would be valuable as a future perspective. For species with insufficient knowledge regarding their genetic background, the applicability of this strategy remains limited. However, the spectrum of applicability of this strategy is expected to expand with ongoing advancements in the sequencing of key crop species, such as for tomato, maize and soybean (Della Coletta et al., 2021; European Commission Joint Research Centre & European Network of GMO Laboratories, 2023). While the availability of high-quality genetic data plays a crucial role, additional factors related to genetic complexity must also be considered, including ploidy and reproductive modes (i.e., outcrossing vs. self-fertilization). In particular, due to the high level of homozygosity typically observed for elite rice varieties, detection of genetic markers in pure lines can be expected at a clear-cut allelic frequency of either 100 % or 0 %. The broader implementation of the proposed strategy may be facilitated by initially focusing on species with genetic characteristics similar to rice, which serves as a well-established crop model species in this study. Tomato, for example, represents a particularly suitable candidate because, like rice, it has a diploid genome, a relatively small genome size and a predominantly self-pollinating reproductive system. Moreover, similar to rice, tomato is not only a highly significant agronomic crop but also a prime target for genome editing, making it an ideal starting point for further investigations aimed to progressively extend the proposed strategy to more genetically complex species (Ahmad, 2023; European Commission Joint Research Centre & European Network of GMO Laboratories, 2023; Tiwari et al., 2023; Zafar et al., 2020).

In conclusion, the proposed strategy allows to pave the way for

supporting the development of new GE lines, as their accurate identification may be essential for patenting such organisms produced by companies, while also strengthening the control conducted by the Competent Authorities and enforcement laboratories to ensure the traceability and authenticity of food and feed products, thereby reinforcing consumer trust.

CRediT authorship contribution statement

Marie-Alice Fraiture: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Jolien D'aes:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis. **Andrea Gobbo:** Writing – review & editing, Formal analysis. **Maud Delvoye:** Writing – review & editing, Formal analysis. **Anne-Cécile Meunier:** Writing – review & editing, Resources, Formal analysis. **Julien Frouin:** Writing – review & editing, Resources, Formal analysis. **Emmanuel Guiderdoni:** Writing – review & editing, Resources, Formal analysis. **Dieter Deforce:** Writing – review & editing, Formal analysis. **Charlotte De Vogelaere:** Writing – review & editing, Software. **Sigrid C.J. De Keersmaecker:** Writing – review & editing, Methodology. **Kevin Vanneste:** Writing – review & editing, Supervision, Project administration, Methodology. **Nancy H.C. Roosens:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The research that yielded these results was funded by Sciensano (Transversal activities in Applied Genomics Service) and by the European Union through the DARWIN project (Grant agreement ID: 101136462). The authors want to thank the technicians of the Transversal activities in Applied Genomics (TAG) service at Sciensano (Brussels, Belgium) for Illumina sequencing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodres.2025.117218.

Data availability

Raw whole-genome sequencing data and amplicon sequencing data is available in the European Nucleotide Archive under project accession number PRJEB84921.

References

Ahmad, M. (2023). Plant breeding advancements with "CRISPR-Cas" genome editing technologies will assist future food security. *Frontiers in Plant Science, 14*, Article 1133036. https://doi.org/10.3389/fpls.2023.1133036

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Bae, S., Park, J., & Kim, J.-S. (2014). Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics, 30*(10), 1473–1475. https://doi.org/10.1093/bioinformatics/btu048

Bertheau, Y. (2021). Advances in identifying GM plants: Toward the routine detection of 'hidden' and 'new' GMOs. In Royal Agricultural University, UK & L. manning (eds.), *Burleigh Dodds series in agricultural science* (pp. 87–150). Burleigh Dodds science publishing. Doi:10.19103/AS.2021.0097.22.

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

Boutigny, A.-L., Fioriti, F., & Rolland, M. (2020). Targeted MinION sequencing of transgenes. *Scientific Reports, 10*(1), 15144. https://doi.org/10.1038/s41598-020-71614-6

Collias, D., & Beisel, C. L. (2021). CRISPR technologies and the search for the PAM-free nuclease. *Nature Communications, 12*(1), 555. https://doi.org/10.1038/s41467-020-20633-y

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., … Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience, 10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

Debode, F., Hulin, J., Charloteaux, B., Coppieters, W., Hanikenne, M., Karim, L., & Berben, G. (2019). Detection and identification of transgenic events by next generation sequencing combined with enrichment technologies. *Scientific Reports, 9*(1), 15595. https://doi.org/10.1038/s41598-019-51668-x

Della Coletta, R., Qiu, Y., Ou, S., Hufford, M. B., & Hirsch, C. N. (2021). How the pan-genome is changing crop genomics and improvement. *Genome Biology, 22*(1), 3. https://doi.org/10.1186/s13059-020-02224-8

EURL. (2006). Sampling and DNA extraction of cotton seeds. Report from the validation of the "CTAB/genomic-tip 20" method for DNA extraction from ground cotton seeds. http://gmo-crl.jrc.ec.europa.eu/summaries/281-3006%20Cotton_DNAExtr.pdf.

European Commission Joint Research Centre & European Network of GMO Laboratories. (2023). *Detection of food and feed plant products obtained by targeted mutagenesis and cisgenesis*. Publications Office. https://data.europa.eu/doi/10.2760/007925.

Fraiture, M.-A., D'aes, J., Guiderdoni, E., Meunier, A.-C., Delcourt, T., Hoffman, S., … Roosens, N. H. C. (2023). Targeted high-throughput sequencing enables the detection of single nucleotide variations in CRISPR/Cas9 gene-edited organisms. *Foods, 12*(3), 455. https://doi.org/10.3390/foods12030455

Fraiture, M.-A., Guiderdoni, E., Meunier, A.-C., Papazova, N., & Roosens, N. H. C. (2022). ddPCR strategy to detect a gene-edited plant carrying a single variation point: Technical feasibility and interpretation issues. *Food Control, 137*, Article 108904. https://doi.org/10.1016/j.foodcont.2022.108904

Fraiture, M.-A., Herman, P., De Loose, M., Debode, F., & Roosens, N. H. (2017). How can we better detect unauthorized GMOs in food and feed chains? *Trends in Biotechnology, 35*(6), 508–517. https://doi.org/10.1016/j.tibtech.2017.03.002

Fraiture, M.-A., Herman, P., Papazova, N., De Loose, M., Deforce, D., Ruttink, T., & Roosens, N. H. (2017). An integrated strategy combining DNA walking and NGS to detect GMOs. *Food Chemistry, 232*, 351–358. https://doi.org/10.1016/j.foodchem.2017.03.067

Fraiture, M.-A., Herman, P., Taverniers, I., De Loose, M., Deforce, D., & Roosens, N. H. (2013). An innovative and integrated approach based on DNA walking to identify unauthorised GMOs. *Food Chemistry, 147*(4,072), 60–69. https://doi.org/10.1016/j.foodchem.2013.09.112

Fraiture, M.-A., Papazova, N., Vanneste, K., De Keersmaecker, S. C. J., & Roosens, N. H. (2019). Chapter 8. GMO detection and identification using next-generation sequencing. In M. Burns, L. Foster, & M. Walker (Eds.), *Food chemistry, function and analysis* (pp. 96–106). Royal Society of Chemistry. https://doi.org/10.1039/9781788016025-00096.

Fraiture, M.-A., Saltykova, A., Hoffman, S., Winand, R., Deforce, D., Vanneste, K., … Roosens, N. H. C. (2018). Nanopore sequencing technology: A new route for the fast detection of unauthorized GMO. *Scientific Reports, 8*(1), 7903. https://doi.org/10.1038/s41598-018-26259-x

Fraiture, M.-A., Ujhelyi, G., Ovesná, J., Van Geel, D., De Keersmaecker, S., Saltykova, A., … 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

Gelinsky, E., & Hilbeck, A. (2018). European court of justice ruling regarding new genetic engineering methods scientifically justified: A commentary on the biased reporting about the recent ruling. *Environmental Sciences Europe, 30*(1), 52. https://doi.org/10.1186/s12302-018-0182-9

Grohmann, L., Keilwagen, J., Duensing, N., Dagand, E., Hartung, F., Wilhelm, R., Bendiek, J., & Sprink, T. (2019). Detection and identification of genome editing in plants: Challenges and opportunities. *Frontiers in Plant Science, 10*, 236. https://doi.org/10.3389/fpls.2019.00236

Guertler, P., Pallarz, S., Belter, A., Eckermann, K. N., & Grohmann, L. (2023). Detection of commercialized plant products derived from new genomic techniques (NGT)—Practical examples and current perspectives. *Food Control, 152*, Article 109869. https://doi.org/10.1016/j.foodcont.2023.109869

Holst-Jensen, A., Spilsberg, B., Arulandhu, A. J., Kok, E., Shi, J., & Zel, J. (2016). Application of whole genome shotgun sequencing for detection and characterization of genetically modified organisms and derived products. *Analytical and Bioanalytical Chemistry, 408*(17), 4595–4614. https://doi.org/10.1007/s00216-016-9549-1

Ichihara, H., Yamada, M., Kohara, M., Hirakawa, H., Ghelfi, A., Tamura, T., … Isobe, S. N. (2023). Plant GARDEN: A portal website for cross-searching between different types of genomic and genetic resources in a wide variety of plant species. *BMC Plant Biology, 23*(1), 391. https://doi.org/10.1186/s12870-023-04392-8

International Standard ISO 21571. (2005). *Foodstuffs—Methods of analysis for the detection of genetically modified organisms and derived products—Nucleic acid extraction*. International Organisation for Standardisation.

Jiang, S., Zhang, X., Yang, X., Liu, C., Wang, L., Ma, B., … Wang, J. (2023). A chromosome-level genome assembly of an early matured aromatic japonica rice variety Qigeng10 to accelerate rice breeding for high grain quality in Northeast

China. *Frontiers in Plant Science, 14*, 1134308. https://doi.org/10.3389/fpls.2023.1134308
Jo, J., Kim, Y., Kim, G. W., Kwon, J.-K., & Kang, B.-C. (2021). Development of a panel of genotyping-in-thousands by sequencing in capsicum. *Frontiers in Plant Science, 12*, Article 769473. https://doi.org/10.3389/fpls.2021.769473
Klees, S., Heinrich, F., Schmitt, A. O., & Gültas, M. (2022). agReg-SNPdb-plants: A database of regulatory SNPs for agricultural plant species. *Biology, 11*(5), 684. https://doi.org/10.3390/biology11050684
Košir, A. B., Arulandhu, A. J., Voorhuijzen, M. M., Xiao, H., Hagelaar, R., Staats, M., … Dijk, J. P. V. (2017). ALF: A strategy for identification of unauthorized GMOs in complex mixtures by a GW-NGS method and dedicated bioinformatics analysis. *Scientific Reports, 7*(1), 14155. https://doi.org/10.1038/s41598-017-14469-8
Kovalic, D., Garnaat, C., Guo, L., Yan, Y., Groat, J., Silvanovich, A., … Bannon, G. (2012). The use of next generation sequencing and junction sequence analysis bioinformatics to achieve molecular characterization of crops improved through modern biotechnology. *The plant Genome, 5*(3). https://doi.org/10.3835/plantgenome2012.10.0026. plantgenome2012.10.0026.
Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nature Methods, 9*(4), 357–359. https://doi.org/10.1038/nmeth.1923
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352
Liang, C., Van Dijk, J. P., Scholtens, I. M. J., Staats, M., Prins, T. W., Voorhuijzen, M. M., … Kok, E. J. (2014). Detecting authorized and unauthorized genetically modified organisms containing vip3A by real-time PCR and next-generation sequencing. *Analytical and Bioanalytical Chemistry, 406*(11), 2603–2611. https://doi.org/10.1007/s00216-014-7667-1
Mansueto, L., Fuentes, R. R., Borja, F. N., Detras, J., Abriol-Santos, J. M., Chebotarov, D., … Alexandrov, N. (2017). Rice SNP-seek database update: New SNPs, indels, and queries. *Nucleic Acids Research, 45*(D1), D1075–D1081. https://doi.org/10.1093/nar/gkw1135
Marchesi, U., Mazzara, M., Broll, H., Giacomo, M. D., Grohmann, L., Herau, V., … Woll, K. (2015). European Network of GMO Laboratories (ENGL)—Definition of Minimum Performance Requirements for Analytical Methods of GMO Testing. *JRC Report.* , Article JRC95544. https://doi.org/10.13140/RG.2.1.2060.5608
Miyao, A., Nakagome, M., Ohnuma, T., Yamagata, H., Kanamori, H., Katayose, Y., Takahashi, A., Matsumoto, T., & Hirochika, H. (2012). Molecular spectrum of somaclonal variation in regenerated rice revealed by whole-genome sequencing. *Plant and Cell Physiology, 53*(1), 256–264. https://doi.org/10.1093/pcp/pcr172
Onda, Y., Takahagi, K., Shimizu, M., Inoue, K., & Mochida, K. (2018). Multiplex PCR targeted amplicon sequencing (MTA-Seq): Simple, flexible, and versatile SNP genotyping by highly multiplexed PCR amplicon sequencing. *Frontiers in Plant Science, 9*, 201. https://doi.org/10.3389/fpls.2018.00201
Pallarz, S., Fiedler, S., Wahler, D., Lämke, J., & Grohmann, L. (2023). Reproducibility of next-generation-sequencing-based analysis of a CRISPR/Cas9 genome edited oil seed rape. *Food Chemistry: Molecular Sciences, 7*, Article 100182. https://doi.org/10.1016/j.fochms.2023.100182
Riza, L. S., Zain, M. I., Izzuddin, A., Prasetyo, Y., Hidayat, T., & Abu Samah, K. A. F. (2023). Implementation of machine learning in DNA barcoding for determining the plant family taxonomy. *Heliyon, 9*(10), Article e20161. https://doi.org/10.1016/j.heliyon.2023.e20161
Saltykova, A., Van Braekel, J., Papazova, N., Fraiture, M. A., Deforce, D., Vanneste, K., … Roosens, N. H. C. (2022). Detection and identification of authorized and unauthorized GMOs using high-throughput sequencing with the support of a sequence-based GMO database. *Food Chemistry: Molecular Sciences, 4*, Article 100096. https://doi.org/10.1016/j.fochms.2022.100096
Shillito, R. D., Whitt, S., Ross, M., Ghavami, F., De Vleesschauwer, D., D'Halluin, K., … Meulewaeter, F. (2021). Detection of genome edits in plants—From editing to seed. *In Vitro Cellular & Developmental Biology - Plant, 57*(4), 595–608. https://doi.org/10.1007/s11627-021-10214-z
Shirasawa, K., Kuwata, C., Watanabe, M., Fukami, M., Hirakawa, H., & Isobe, S. (2016). Target amplicon sequencing for genotyping genome-wide single nucleotide polymorphisms identified by whole-genome resequencing in peanut. *The plant Genome, 9*(3). https://doi.org/10.3835/plantgenome2016.06.0052. plantgenome2016.06.0052.
Singh, R. R. (2022). Target enrichment approaches for next-generation sequencing applications in oncology. *Diagnostics, 12*(7), 1539. https://doi.org/10.3390/diagnostics12071539
Sturme, M. H. J., Van Der Berg, J. P., Bouwman, L. M. S., De Schrijver, A., De Maagd, R. A., Kleter, G. A., & Battaglia-de Wilde, E. (2022). Occurrence and nature of off-target modifications by CRISPR-Cas genome editing in plants. *ACS Agricultural Science & Technology, 2*(2), 192–201. https://doi.org/10.1021/acsagscitech.1c00270
Tang, X., Liu, G., Zhou, J., Ren, Q., You, Q., Tian, L., Xin, X., Zhong, Z., Liu, B., Zheng, X., Zhang, D., Malzahn, A., Gong, Z., Qi, Y., Zhang, T., & Zhang, Y. (2018). A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice. *Genome Biology, 19*(1), 84. https://doi.org/10.1186/s13059-018-1458-5
The 3,000 rice genomes project. (2014). The 3,000 rice genomes project. *Gigascience, 3*(1). https://doi.org/10.1186/2047-217X-3-7, 2047-217X-3–7.
Tiwari, J. K., Singh, A. K., & Behera, T. K. (2023). CRISPR/Cas genome editing in tomato improvement: Advances and applications. *Frontiers in Plant Science, 14*, Article 1121209. https://doi.org/10.3389/fpls.2023.1121209
Ulhuq, F. R., Barge, M., Falconer, K., Wild, J., Fernandes, G., Gallagher, A., … McHugh, M. P. (2023). Analysis of the ARTIC V4 and V4.1 SARS-CoV-2 primers and their impact on the detection of omicron BA.1 and BA.2 lineage-defining mutations. *Microbial Genomics, 9*(4). https://doi.org/10.1099/mgen.0.000991
Van der Auwera, G. A., & O'Connor, B. D. (2020). *Genomics in the cloud: Using docker, GATK, and WDL in terra* (1st ed.). Sebastopol, CA: O'Reilly Media.
Wahler, D., Schauser, L., Bendiek, J., & Grohmann, L. (2013). 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 Analytical Methods, 6*(6), 1718–1727. https://doi.org/10.1007/s12161-013-9673-x
Wang, D. R., Agosto-Pérez, F. J., Chebotarov, D., Shi, Y., Marchini, J., Fitzgerald, M., … McCouch, S. R. (2018). An imputation platform to enhance integration of rice genetic resources. *Nature Communications, 9*(1), 3519. https://doi.org/10.1038/s41467-018-05538-1
Wang, T., He, W., Li, X., Zhang, C., He, H., Yuan, Q., Zhang, B., Zhang, H., Leng, Y., Wei, H., Xu, Q., Shi, C., Liu, X., Guo, M., Wang, X., Chen, W., Zhang, Z., Yang, L., Lv, Y., & Shang, L. (2023). A rice variation map derived from 10 548 rice accessions reveals the importance of rare variants. *Nucleic Acids Research, 51*(20), 10924–10933. https://doi.org/10.1093/nar/gkad840
Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., … Leung, H. (2018). Genomic variation in 3,010 diverse accessions of asian cultivated rice. *Nature, 557*(7703), 43–49. https://doi.org/10.1038/s41586-018-0063-9
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
Wilm, A., Aw, P. P. K., Bertrand, D., Yeo, G. H. T., Ong, S. H., Wong, C. H., … Nagarajan, N. (2012). LoFreq: A sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Research, 40*(22), 11189–11201. https://doi.org/10.1093/nar/gks918
Yang, J., Zhang, J., Du, H., Zhao, H., Li, H., Xu, Y., Mao, A., Zhang, X., Fu, Y., Xia, Y., & Wen, C. (2022). The vegetable SNP database: An integrated resource for plant breeders and scientists. *Genomics, 114*(3), Article 110348. https://doi.org/10.1016/j.ygeno.2022.110348
Yuan, X., Li, Z., Xiong, L., Song, S., Zheng, X., Tang, Z., Yuan, Z., & Li, L. (2022). Effective identification of varieties by nucleotide polymorphisms and its application for essentially derived variety identification in rice. *BMC Bioinformatics, 23*(1), 30. https://doi.org/10.1186/s12859-022-04562-9
Zafar, K., Sedeek, K. E. M., Rao, G. S., Khan, M. Z., Amin, I., Kamel, R., … Mahfouz, M. M. (2020). Genome editing technologies for rice improvement: Progress, prospects, and safety concerns. *Frontiers in Genome Editing, 2*, 5. https://doi.org/10.3389/fgeed.2020.00005
Zhu, H., Misel, L., Graham, M., Robinson, M. L., & Liang, C. (2016). CT-finder: A web service for CRISPR optimal target prediction and visualization. *Scientific Reports, 6*(1), Article 25516. https://doi.org/10.1038/srep25516