Preprints.org

Article

Stepwise In Silico Specificity Approach for NGT PCR-Based Detection Methods in the Context of the EU GMO Regulations

Caroline Bedin Zanatta , Aline Martins Hoepers , Rubens Onofre Nodari , Sarah Zanon Agapito-Tenfen *

Article

Stepwise In Silico Specificity Approach for NGT PCR-Based Detection Methods in the Context of the EU GMO Regulations

Caroline Bedin Zanatta [1], Aline Martins Hoepers [1], Rubens Onofre Nodari [1] and Sarah Zanon Agapito-Tenfen [2,*]

[1] Departamento de Fitotecnia, Universidade Federal de Santa Catarina, Brazil

[2] Climate and Environmental Division, NORCE Norwegian Research AS, Norway

[*] Correspondence: saag@norceresearch.no

Abstract: The term new genomic techniques (NGTs), is an umbrella term to describe a variety of techniques that can alter the genetic material of an organism and that have emerged or have been developed since 2001, when the existing genetically modified organism (GMO) legislation was adopted. The analytical framework to detect GMOs in Europe is a established single harmonized procedure which is mandatory for authorization of GM food and feed, thus generating reliable, transparent and effective labelling scheme for GMO products. However, NGT products challenge the implementation and enforcement of the current regulatory system in the EU, relating in particular to the detection of NGT products that contain no foreign genetic material. Consequently, current detection methods will fail to meet minimum performance requirements. Although existing detection methods may be able to detect and quantify even small alterations in the genome, this does not necessarily confirm the distinction between products resulting from NGTs subject to the GMO legislation and other products. Therefore, this study provides a stepwise approach for the in silico prediction of PCR systems specificity by testing a bioinformatics pipeline for amplicon and primer set searches in current genomic databases. We have tested our methodology in two mutant genotypes produced by CRISPR/Cas9 in Arabidopsis thaliana. Overall, our results demonstrated that the single PCR system developed for the identification of a nucleotide insertion in the grf1-3 genotype has sufficient hits in the databases that would not allow discrimination of this mutated event. Differently, the second mutated genotype grf8-61 which contains a -3bp deletion, found no matches in the sequence variant database. Our approach is the first step in decision making on PCR system and modification of them for empirical testings.

Keywords: genetically modified organisms; site directed nucleases; CRISPR/Cas9; bioinformatics; detection; traceability

1. Introduction

Genetically modified organisms (GMOs) are subject to regulations in Europe and worldwide through domestic legislation and the Cartagena Protocol on Biosafety (CPB), an international GMO treaty by the United Nations. In Europe, prior to market approval of GMOs and derived food and feed products, GMO event-specific methods for their detection, identification, and quantification need to be in place according to European Commission (EC) Regulation No 1829/2003 on genetically modified food and feed. The analytical framework to detect GMOs in Europe relies on validated qualitative and quantitative real-time PCR methods, forming part of the single harmonized, time-limited, and transparent procedure for the authorization of GM food and feed in the European Union (EU) (EC,2021).

However, recent developments in the field of genetic engineering, known as 'new genomic techniques' (NGT), have raised concerns over the suitability of such detection methods and the possibility of implementing current GMO regulations due to small genomic alterations are an umbrella term used to describe various techniques that can alter the genetic material of an organism and have emerged or been developed since 2001 when the existing genetically modified organism

(GMO) legislation was adopted (EC, 2021). These techniques are based on site-directed nucleases (SDNs) (GROHMANN et al., 2019) and encompass different enzymes, such as meganucleases (LAGLIDADG endonucleases; EMN), zinc finger nucleases (ZNF), effector transcription factor nucleases (TALENs), and Clustered Regularly Interspaced Short Palindromic Repeats system (CRISPR-Cas9) (SPRINK et al., 2016) (BROLL; BRAEUNING; LAMPEN, 2019) (METJE-SPRINK et al., 2019) (GAJ; GERSBACH; BARBAS, 2013) (OSAKABE; OSAKABE, 2015).

In 2018, the Court of Justice of the European Union (CJEU) issued a significant decision stating that gene-editing techniques, including CRISPR-Cas9, should be classified and regulated as GMOs. As a result, any techniques that modify the genetic material of plants or animals using NGT should be subject to the same regulations applied to GMOs. In this context, the EU provide significant steps to implement legislation ensuring the traceability of GMO-containing foods and derivatives in the market, providing scientific guidelines and technical documents to empower consumers with informed choices through proper labeling (ENGL, 2008) (EUROPEAN LABORATORIES, 2015) (ARAKI; NOJIMA; ISHII, 2014) (FRAITURE et al., 2015).

In the context of detecting NGTs, two main steps are crucial for the development of PCR-based analysis methods: i) in-silico tests and ii) empirical tests. The in-silico computational assessments involve the verification of the compatibility of primer sequences and avoid unintended amplifications of other genomic regions or sequences from different organisms. Ensuring the specificity of primers designed for the edited organism is essential, taking into consideration any base differences. PCR relies on the complementarity between primers and the target DNA for successful amplification. The design of specific primers and probes involves predicting secondary structures to avoid interferences, such as primer dimers or hairpin formations, which can affect amplification specificity. Additionally, evaluating sequence similarity with other organisms ensures that the target sequences of interest do not share similarities with sequences from different organisms (ENGL, 2021). The technical complexity of designing a PCR system capable of distinguishing specific sequences with accuracy and reliability requires both in-silico and experimental testing. Some studies have successfully distinguished NGT canola and rice varieties using PCR (WEIDNER et al., 2022) (FRAITURE et al., 2022) (ZHANG et al., 2022).

To address the challenges associated with detecting edited organisms, our approach can be applied to explore and characterize NGT-modified organisms, as it adheres to the principles of open science, ensuring transparency in data and efforts for the detection, identification, and biosafety of these organisms. In this study, we present a step-by-step bioinformatics analysis to predict primer specificity requirements in the context of the EU GMO law and regulations, employing two NGT model sequences from Arabidopsis thaliana, each harboring distinct mutation patterns (single polymorphism for genotype 1 and a 3 bp deletion for genotype 2) in the target growth-regulating factor (GRF) gene. We expect our approach to offer a theoretical foundation for assessing primer specificity, complemented by empirical tests to support the hypothesis of selective target amplification.

2. Materials and Methods

2.1. Overall description of the stepwise approach

The theoretical basis of real-time polymerase chain reaction (RT-PCR) was used to screen mutations delivered by a CRISPR-Cas9 technique at the target site. The first step identified in our approach is the sanger sequencing of mutated alleles for confirmatory purposes. After confirmation of mutation pattern in selected genotypes, the next step is a comprehensive search on existing databases for sequence similarities and the natural variability around the mutated sequence. The outcomes of the search analysis were manually evaluated for the presence of mismatches in regions corresponding to the amplicon sequence. Lastly, an in-silico PCR testing is performed to estimate the amplification rates for each PCR system and a second confirmatory steps for primer and probe specificity. Each of these steps is described in detail in the next sections and a full picture of the overall approach is presented below (Figure 1). We performed all the steps in a model case for two gene-

edited genotypes of Arabidopsis. The description of the biological material used in this study is provided in the next section.
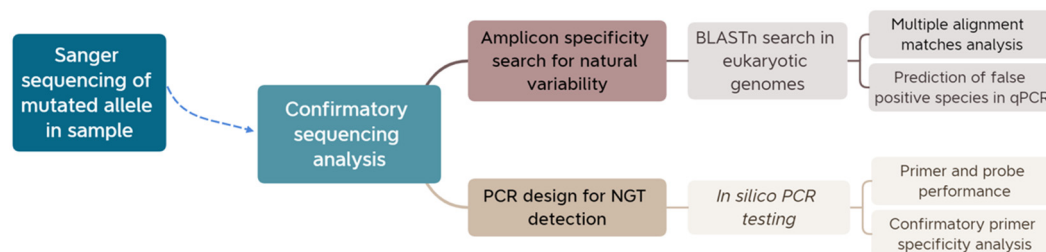
Figure 1. Screening approach for identifying material mutated by Site Directed Nucleases (SDN-1) technique. After the confirmation of sequencing, the predicted amplicon with suitable parameters was searched in the GenBank and was specificity evaluate as template in silico PCR.

2.2. Confirmatory sequencing analysis of CRISPR/Cas9 mutations

The seeds from two CRISPR-Cas9 genotypes, grf1-3 and grf8-61, as well as the control of Arabidopsis thaliana, were obtained from the Eurasian Arabidopsis Stock Center (NASC). The transformation of plants was carried out using the Agrobacterium strain GV3101 through the floral dip protocol. In this method, *Arabidopsis thaliana* with a Columbia background served as the wild-type strain for transformation and CRISPR/Cas9 mutagenesis. The plants were immersed in an Agrobacterium suspension to facilitate the transfer of CRISPR/Cas9 components into the plant cells (ANGULO et al., 2020). The grf1-3 genotype is a null mutant and transgene-free. The mutation is located in the GRF1 (Growth-Regulating Factor 1) gene, specifically in the locus AT2G22840. This mutation involves a single nucleotide polymorphism (SNP) with a Guanine insertion at position 10048382 on the positive strand of chromosome 2. On the other hand, the grf8-61 genotype carries a three-base pair (bp) deletion, resulting in the mutation of the GRF8 (Growth-Regulating Factor 8) gene located in the locus 2134986.Additionally, the Columbia ecotype, sequenced in the Arabidopsis Genome Initiative (https://www.ebi.ac.uk/ols/ontologies/efo/terms?short_form=EFO_0005147), was used in the study. Three individual plants per genotype were germinated in sterilized soil placed in pots. After 10 days of incubation at 4°C in the dark, the plants were transferred to a greenhouse with a light period of 8 hours/day and an intensity of approximately µmol m-2 s-1 at 25°C for growth. At the growth stage (G), ten rosette leaves per plant were used for DNA isolation using the Plant DNAzolTM Reagent from Invitrogen Co., US. The target regions from genotypes grf1-3, grf8-61, and Columbia were amplified through conventional PCR using Sanger primers with the following cycling conditions: 2 minutes at 94°C, followed by 35 cycles of 45 seconds at 94°C, 30 seconds at 60°C, 1 minute at 72°C, and a final extension of 10 minutes at 72°C (Applied Biosystems™ Veriti™ Thermal Cycler). The PCR products from each genotype were purified and sequenced using the Applied Biosystems Genetic Analyzer (3500×, São Paulo, Brazil). The obtained Sanger sequences from each genotype were checked for quality using Chromas version 2.6.6 (Technelysium Pty Ltd., Brisbane, Australia). The alignment comparison with default settings (MUSCLE) was performed to compare the sequences obtained from each Sanger fragment to the GRF1 (AT2G22840; ID: 816815; Chr2) and GRF8 (AT4G24150; ID: 828515; Chr4) genes of *Arabidopsis thaliana* (NCBI GenBank taxid:3702).

2.3. Amplicon sequence search for natural variants

The grf1-3 and grf8-61 amplicon sequences and their wild type of counterpart, Columbia (Col-01), were searched against the National Center for Biotechnology Information database (NCBI) using the Basic Local Alignment Search Tool (BLAST). The databases accessed were nucleotides (nr/nt)

available at GenBank + EMBL + DDBJ + PDB + RefSeq databases. Patent strings, phase strings 0, 1 and 2 HTGS, EST, STS, GSS, WGS, TSA and greater than 100MB were excluded. To be considered as valid hits, outcome sequences retrieved through BLAST needed to follow the 80% coverage parameter; 80 to 98% identity level, and sequence hits below 10 bp of mismatches were considered eligible. The presence of mismatches in regions corresponding to the primer and probe sets were manually verified using the Multiple Sequence Alignment Viewer (MSA) graphic displays for NCBI nucleotide alignment. The position of the single point mutation is indicated in Figure 3. A) and the positions of the exclusions, indicated (Figure 2. B) were used to search for position polymorphisms in a database composed of 1,001 variant genomes of *Arabidopsis thaliana* (https://1001genomes.org/). The reason behind the strategy to use amplicon, instead of classical primer and probe sets, was to avoid false positive hits when only one primer set, or probe matches the sequence, but no effective amplification is expected at similar amplicon size.

2.3. RT-PCR primer and probe design

The design of primers and probes for detecting CRISPR-Cas9 mutations in the grf1-3 and grf8-61 genotypes of Arabidopsis thaliana was performed using the Primer-3 Plus software (ROZEN; SKALETSKY, 2000) (UNTERGASSER et al., 2007). To ensure primer specificity and avoid PCR competition, the sequences of each primer set were manually verified in both DNA strand orientations to contain the CRISPR-Cas9 mutation. First, the size of the amplicon for the single-copy reference gene was taken as a basis to design the amplicons for each edited event (Figure 2). Subsequently, the selection of primer and probe sets was based on specific parameters. All oligonucleotides were checked for the GC content of the primers (between 45-55%) and the melting temperature (Tm), which was set 3 to 5°C below the Tm of the probe (between 62 and 68°C) (ROBLEDO; BEGGS; BENDER, 2005). Additionally, secondary structures and hairpins were considered (MIQE, 2009) (ENGL, 2021). The selected primers were also evaluated for similarity to the endogenous reference gene amplicon (AT1G03400) of Arabidopsis thaliana (YANG et al., 2018).

Figure 2. Schematic representation of the amplicon containing CRISPR-Cas9 mutations. In panel A, the 130 bp Amplicon for the grf1-3 event is illustrated, the InDel G insert is shown in the last base of the Forward target primer. Panel B, shows 127 bp amplicon from grf8-6. The asterisk represents the

probes in the amplicon. The arrow and box indicate the difference compared to the native primer sequence.

2.5. In-silico PCR testing

In addition, a second confirmatory step was performed for in-silico PCR testing using PrimerBLAST software (Ye et al., 2013). This analysis aims to screen for primer annealing in directions from 5' to 3' between all combinations, including forward-to-reverse, forward-to-forward and reverse-reverse. The in-silico PCR testing was performed in all available species in the database and no amplification was detected in data set species. For each amplicon template (grf1-3 and grf8-61) the PCR products were limited to the size of 100 to 300 bp, and the average melting temperature (Tm) (Supplementary File S2) was adjusted. The database analyzed was from genomes Eukariotic For the stringency of the primers, it was considered that the primer should contain at least 2 unintended targets, including at least 2 incompatibilities in the last 5 bps at the 3' end. The number of target sequences retrieved for each template was considered an indicative to the reproducibility and chances of finding amplification from different target sequences than the intended one.

3. Results

3.1. Sequence confirmation of mutated alleles

After sanger sequencing of mutated genotypes, and wild type non-mutated genotype, intend mutations were confirmed for each new allele obtained by CRISPR-Cas9 (Figure 2).
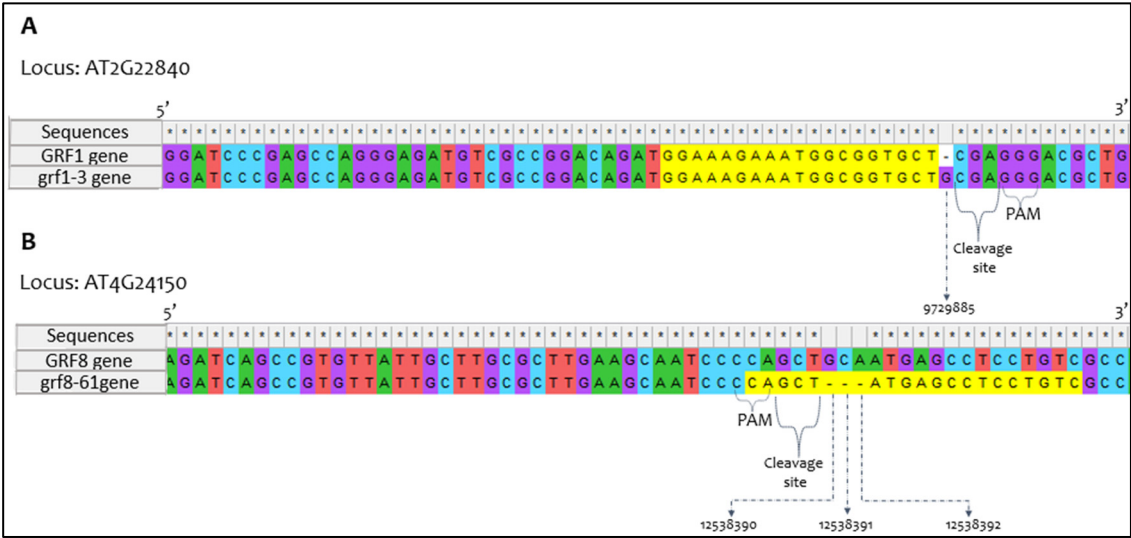
Figure 2. Alignment between the sequences obtained through the sequencing data. Panel A shows the comparison between the sequences obtained from the Columbia (Col-1) and the grf1-3 CRISPR-Cas9 allele. The location of the Guanine (G) insertion is demonstrated in the respective position of *Arabidopsis thaliana* genome. Panel B shows the comparison between sequences (Col-1) and the gfr8-61 CRISPR-Cas9 allele. The three bases deleted, Guanine (G) Cytosine (C) and Adenosine (A) are shown in the same respective genome. For both cases, the site at 3 bp after the PAM Cas9 was useful for mutations localization.

3.3. In-silico specificity assessment

The search for amplicon sequences for wild type GRF1 gene in the NCBI nucleotide collection (nr/nt) recovered a total of 120 hits/occurrences in 39 organisms. Of these, 78 hits and 15 organisms belong to the Brassicales Order (Supplementary Archive S3). We recovered 58 hits related to the genus Arabidopsis, of these, 56 accesses related to *Arabidopsis thaliana* (100% identity) and the other 2 hits belong to the species-genus *Arabidopsis lirata* (99.23% identity) and *Arabidopsis arenosa* (98.46% identity). In five hits, the sequences were 100% identical to the GRL1 gene of the growth-regulating

protein (B2CU94_ARATH). About the wild type GRF8 sequence, 57 hits were recovered with 100% coverage and 100% identity in *Arabidopsis thaliana* (Supplementary File S4). Of these, 25 accesses were 100% identical to the GRL8 gene of the protein growth regulatory factor in Arabidopsis thaliana (B2CUI8_ARATH). These results show that no significant homology is found for the two gene alleles outside the Arabidopsis genus (Supplementary File S5).

For amplicon grf1-3 CRISPR-Cas9 (mutated), 36 hits with sequences with 100% coverage were found. Two Arabidopsis species, *A. Lirata* and *A. Arenosa,* presented an identity of 99.71% and 98.47%, respectively to the mutated GRF1amplicon. Both species showed identical sequence with a SNP variation in position 15. Additionally, *A. arenosa* presents a mismatch at position 108. Other species, such as *Camelina sativa* and *Camelina hispida* showed 100% coverage but lower identity levels reaching 98.48% and 97.71%, respectively due to different mismatches. We also found 10 hits from 4 species from Brassica genus which presented 95.5% identity. *Raphanus sativus* species also showed 2 sequence hits with coverage and identity of 98% and 96.12%, respectively. No other polymorphisms were found in the database for *Arabidopsis thaliana* for the given amplicons. Matching results are presented in Table 1 for grf1-3.

Table 1. The recovered BLAST hits obtained from *grf1-3* CRISPR-Cas9 amplicon sequence.

Hits	Accession	Blast against the query grf1-3		(MSA) Mismatches (bp) against each accession		
		Quey cover	Per. Ident	Forward	Probe	Reverse
Arabidopsis thaliana genome assembly, chromosome: 2	LR782543.1	100%	99.24%	1	0	0
Arabidopsis thaliana genome assembly, chromosome: 2	LR699746.2	100%	99.24%	1	0	0
Arabidopsis thaliana genome assembly, chromosome: 2	LR699771.1	100%	99.24%	1	0	0
Arabidopsis thaliana genome assembly, chromosome: 2	LR699766.1	100%	99.24%	1	0	0
Arabidopsis thaliana genome assembly, chromosome: 2	LR699761.1	100%	99.24%	1	0	0
Arabidopsis thaliana genome assembly, chromosome: 2	LR699756.1	100%	99.24%	1	0	0
Arabidopsis thaliana genome assembly, chromosome: 2	LR699751.1	100%	99.24%	1	0	0
Arabidopsis thaliana genome assembly, chromosome: 2	LR215053.1	100%	99.24%	1	0	0
Arabidopsis thaliana growth-regulating factor 1 (GRF1), mRNA	NM_127849.4	100%	99.24%	1	0	0
Arabidopsis thaliana chromosome 2	CP116281.1	100%	99.24%	1	0	0
Arabidopsis thaliana genome assembly, chromosome: 2	OX298798.1	100%	99.24%	1	0	0
Arabidopsis thaliana genome assembly, chromosome: 2	OX298803.1	100%	99.24%	1	0	0
Arabidopsis thaliana ecotype 1254 chromosome 2 sequence	CP086755.1	100%	99.24%	1	0	0
Arabidopsis thaliana ecotype 5856 chromosome 2 sequence	CP086750.1	100%	99.24%	1	0	0

Arabidopsis thaliana ecotype 6021 chromosome 2 sequence	CP086745.1	100%	99.24%	1	0	0
Arabidopsis thaliana ecotype 6024 chromosome 2 sequence	CP086740.1	100%	99.24%	1	0	0
Arabidopsis thaliana ecotype 9412 chromosome 2 sequence	CP086735.1	100%	99.24%	1	0	0
Arabidopsis thaliana ecotype 9470 chromosome 2 sequence	CP086730.1	100%	99.24%	1	0	0
Arabidopsis thaliana chromosome 2	CP087127.2	100%	99.24%	1	0	0
Arabidopsis thaliana isolate t2t_salk_col chromosome 2	CP096025.1	100%	99.24%	1	0	0
Arabidopsis thaliana genome assembly, chromosome: 2	OW119597.1	100%	99.24%	1	0	0
Arabidopsis thaliana genome assembly, chromosome: 2	LR881467.1	100%	99.24%	1	0	0
Arabidopsis thaliana genome assembly, chromosome: 2	LR797808.1	100%	99.24%	1	0	0
Arabidopsis thaliana genome assembly, chromosome: 2	LR797803.1	100%	99.24%	1	0	0
Arabidopsis thaliana genome assembly, chromosome: 2	LR797798.1	100%	99.24%	1	0	0
Arabidopsis thaliana genome assembly, chromosome: 2	LR797793.1	100%	99.24%	1	0	0
Arabidopsis thaliana genome assembly, chromosome: 2	LR797788.1	100%	99.24%	1	0	0
Arabidopsis thaliana chromosome 2	CP002685.1	100%	99.24%	1	0	0
Arabidopsis thaliana At2g22840 mRNA for hypothetical protein, partial cds, clone: RAAt2g22840	AB493560.1	100%	99.24%	1	0	0
Arabidopsis thaliana isolate CS906 GRL1 (GRL1) gene, partial cds	EU550462.1	100%	99.24%	1	0	0
Arabidopsis thaliana isolate CS902 GRL1 (GRL1) gene, partial cds	EU550456.1	100%	99.24%	1	0	0
Arabidopsis thaliana isolate CS6799 GRL1 (GRL1) gene, partial cds	EU550455.1	100%	99.24%	1	0	0
Arabidopsis thaliana isolate CS901 GRL1 (GRL1) gene, partial cds	EU550445.1	100%	99.24%	1	0	0
Arabidopsis thaliana transcription activator (GRF1) mRNA, complete cds	AY102634.1	100%	99.24%	1	0	0
Arabidopsis thaliana chromosome 2 clone T20K9 map CIC06C07, complete sequence	AC004786.3	100%	99.24%	1	0	0
Arabidopsis thaliana Full-length cDNA Complete sequence from clone GSLTPGH12ZD08 of	BX820248.1	100%	99.24%	1	0	0

Hormone Treated Callus of strain col-0 of Arabidopsis thaliana (thale cress) PREDICTED: Arabidopsis xampl subsp. xampl growth-regulating factor 1 (LOC9316532), mRNA	XM_002878592.2	100%	98.47%	1	0	0
PREDICTED: Camelina sativa growth-regulating factor 1-like (LOC104713726), mRNA	XM_010430916.2	100%	98.47%	1	0	0
PREDICTED: Camelina sativa growth-regulating factor 1 (LOC104751923), mRNA	XM_010473979.2	100%	98.47%	1	0	0
PREDICTED: Camelina sativa growth-regulating factor 1-like (LOC104704976), mRNA	XM_010420970.1	100%	98.47%	1	0	0
Camelina hispida cultivar hispida voucher DAO 902780 chromosome 2	CP094632.1	100%	97.71%	1	0	1
Arabidopsis arenosa genome assembly, chromosome: 4	LR999454.1	100%	97.71%	1	0	1
Raphanus sativus genome assembly, chromosome: 6	LR778315.1	98%	96.12%	1	0	2 (gap)
PREDICTED: Raphanus sativus growth-regulating factor 1 (LOC108836427), mRNA	XM_018609585.1	98%	96.12%	1	0	2 (gap)
PREDICTED: Brassica rapa growth-regulating factor 1 (LOC103858395), mRNA	XM_009135745.3	100%	95.42%	1	0	1
Brassica oleracea HDEM genome, scaffold: C3	LR031872.1	100%	95.42%	1	0	1
Brassica rapa genome, scaffold: A03	LR031572.1	100%	95.42%	1	0	1
PREDICTED: Capsella rubella growth-regulating factor 1 (LOC17887921), mRNA	XM_006293922.2	100%	95.42%	1	1	1
PREDICTED: Brassica napus growth-regulating factor 1-like (LOC125584397), mRNA	XM_048752816.1	100%	95.42%	1	0	1
PREDICTED: Brassica napus growth-regulating factor 1 (LOC106389497), mRNA	XM_013829762.3	100%	95.42%	1	0	1
Brassica rapa genome assembly, chromosome: A03	LS974619.2	100%	95.42%	1	0	1
Brassica napus genome assembly, chromosome: C03	HG994367.1	100%	95.42%	1	0	1
Brassica napus genome assembly, chromosome: A03	HG994357.1	100%	95.42%	1	0	1
Brassica rapa subsp. Pekinensis growth-regulating xamp 1 mRNA, partial cds	JN698986.1	100%	95.42%	1	0	1

PREDICTED: Brassica oleracea var. oleracea growth-regulating factor 1 (LOC106328366), mRNA	XM_013766798.1	100%	94.66%	1	0	1
PREDICTED: Eutrema salsugineum growth-regulating factor 1 (LOC18021800), mRNA	XM_006404687.2	100%	93.89%	1	2	2
Arabis alpina genome assembly, chromosome: 6	LT669793.1	93%	95.90%	1	2	9 (gap)

For the grf8-1 amplicon, 26 hits were recovered with 96% coverage due to the presence of gaps in the initial bases, corresponding to (C, A, G, C, T) the forward primer. After the 6-mer position, the sequences showed 100% coverage. In two of those hits, the identity was 97.54% due to the presence of mismatches in the probe sequence. A 99.20% identity and 98% coverage was also found for two other hits that have a mismatch in the 5' 5-mer a transversion-type mutation in DNA (A) position 12538392 in the forward primer. No other sequences were found with 100% alignment for this amplicon. The variation across these hits indicates that this 5pb can be composed of polymorphic nucleotides. However, the three excluded bases were not found in the variant database for *Arabidopsis thaliana.* Table 3 complies the results obtained for the grf8-61 amplicon search against the database.

Table 2. The recovered BLAST hits obtained from *grf8-61* CRISPR-Cas9 amplicon sequence.In red illustrate the specie that is unreadable due mismatches in the amplicon. The displayed hits of data present 5pb or less of mismatches were collected in the NCBI Genbank.

Description	ccession	Blast against the query grf8-61		(MSA) Mismatches (bp) against each accession		
		Query Cover	Per. Ident	Forward	Probe	Reverse
Arabidopsis thaliana genome assembly, chromosome: 4	LR782545.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana genome assembly, chromosome: 4	LR699748.2	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana genome assembly, chromosome: 4	LR699773.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana genome assembly, chromosome: 4	LR699768.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana genome assembly, chromosome: 4	LR699758.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana genome assembly, chromosome: 4	LR699753.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana genome assembly, chromosome: 4	LR215055.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana chromosome 4	CP116283.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana genome assembly, chromosome: 4	OX298800.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana genome assembly, chromosome: 4	OX298805.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana ecotype 1254 chromosome 4 sequence	CP086757.1	98%	99.20%	0	0	4 (gaps) 1 mismach

Arabidopsis thaliana ecotype 5856 chromosome 4 sequence	CP086 752.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana ecotype 6021 chromosome 4 sequence	CP086 747.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana ecotype 6024 chromosome 4 sequence	CP086 742.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana ecotype 9412 chromosome 4 sequence	CP086 737.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana ecotype 9470 chromosome 4 sequence	CP086 732.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana chromosome 4	CP087 129.2	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana isolate t2t_salk_col chromosome 4	CP096 027.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana genome assembly, chromosome: 4	OW11 9599.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana genome assembly, chromosome: 4	LR881 469.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana genome assembly, chromosome: 4	LR797 810.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana genome assembly, chromosome: 4	LR797 805.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana genome assembly, chromosome: 4	LR797 800.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana genome assembly, chromosome: 4	LR797 795.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana chromosome 4	CP002 687.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana DNA chromosome 4, contig 10xample1010 No. 61	AL161 561.2	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana DNA chromosome 4, BAC clone T19F6, partial sequence (ESSA 10xample)	AL109 619.1	96%	100.00%	0	0	5 (gaps)
Arabidopsis thaliana chromosome IV BAC T19F6 genomic sequence, complete sequence	AC002 343.1	98%	99.20%	0	0	4 (gaps) 1 mismach
Arabidopsis thaliana genome assembly, chromosome: 4	LR699 763.1	96%	97.54%	0	2	5 (gaps)
Arabidopsis thaliana genome assembly, chromosome: 4	LR797 790.1	96%	97.54%	0	2	5 (gaps)
Arabidopsis thaliana growth-regulating factor 8 (GRF8), partial mRNA	NM_1 18547. 2	83%	100.00%			

3.4. In-silico PCR performance

A pair of primers should only amplify the target sequence. This is especially challenging when the target region differs by one or a few nucleotides from other potential targets. The in-silico PCR results presented 57 potential hits for the grf1-3 PCR system. Of this, 39 hits were observed for the forward primer containing a Cytosine (C) base inserted as a mismatch. The other 18 hits were distributed in *Arabidopsis hispida* (1 hit) and *Arabidopsis arenosa* (1) and Brassicas sp (16). Despite having mismatches in the amplicon against *Brassica napus, Brassica oleracea, Brassica Rapa*, these species were still considered a potential amplification sequence in our in-silico PCR prediction. These results suggest that the primer set designed to distinguish grf1-3 CRISPR-Cas9 genotype has the potential to amplify sequences in other species, such as Arabidopsis and Brassica.

Table 3. Summary of *in silico*-PCR amplification for the grf1-3 amplicon PCR in silico matrix for the grf1-3 amplicon. These theoric condition indicate unintended targets. Thus, primers may not be specific for the PCR template.

Total of mismatches	Number/hits analyzed	Sequences corresponding perfectly to the primer	Number of Blast hits recovered	Possible discrimination between the grf1-3 genotype and other lines
1	39	0		
2	17	0	57	**Low**
3	1	0		

The primer set design for the detection of 3bp deletion in the Growth Factor Regulator 8 gene showed a unique target in *Arabidopsis thaliana*. Therefore, the primer set designed can be considered specific for the grf8-61 genotype after consideration of the in silico PCR prediction.

Table 4. Summary of in-silico PCR amplification result for the grf8-61. The amplicon correctly identifies the new CRISPR-Cas9 grf8-61 genotype. The primers demonstrate specificity in-silico PCR.

Total of mismatches	Number/hits analyzed	Sequences corresponding perfectly to the primers	Number of Blast hits recovered	Possible discrimination between the grf8-3 genotype and other lines
1	1	1	1	**High**

4. Discussion

DNA breaks caused by site-directed nuclease, such as CRISPR-Cas, in genetically modified organisms (GMOs) can be repaired through two major pathways, one pathway joins the ends from a template (HR) and another pathway that joins the ends without homology (NHEJ). In plants produced by NBTs, predominantly, the NHEJ pathway is used to obtain knockout of genes and mutants with characteristics of interest to agriculture (YAN et al., 2020) (Rodríguez-Leal et al., 2017) (Sedeek et al., 2019).

Mutants generated by this repair mechanism have one or a few InDels other than their unmutated part, and these different nucleotides are used as identifiers to characterize the new GMO. In this study, two transcriptional growth regulatory genes (GRFs) regions in Arabidopsis thaliana were used as case studies to develop a stepwise approach for in silico prediction of specific PCR primer sets. The GRF protein gene family is involved in growth and development, and the stress response (Liu et al., 2016) (Huang et al., 2021, Li et al., 2021) (OMIDBAKHSHFARD et al., 2015).

The in silico screening of the amplicon sequences of the native genes of A. thaliana Columbia in the NCBI nucleotide database allowed us to identify that the primer sequence used in the GRF1 gene has an identity of 99% to Arabidopsis lyrate, 97.24% to Arabidopsis arenosa, and 94.6% to agricultural species Brassica napus, Brassica rapa, Brassica rapa subsp. Pekinensis and Brassica oleracea var oleracea. Another 15 organisms belonging to the order Brassicales were found with lower similarity, above 6 bp of mismatches. The amplicon sequence search for GRF8 gene showed only one hit in Arabidopsis thaliana with 100% identity, thus showing our own native sequence.

As expected, the BLAST of amplicon grf1-3 indicated that the single nucleotide polymorphism is the only difference in 35 sequence hits with 99.8% similarity. Similar results were found by (WEIDNER et al., 2022) analyzing amplicon of primer sets for the detection of gene-edited canola. In their study, the BLAST search yielded hits in B. oleracea and B. carinata, in which the only difference between the targets was the last nucleotide at position 22 – in the position of the complementary reverse primer to the base for detection (WEIDNER et al., 2022).

Most gene edited organisms will contain mutations in conserved regions and/or relevant exons for trait accomplishment. This is the case for gene edited Camelina sativa and Brassica napus obtained by CRISPR-Cas9 for high oleic acid content (Jiang et al., 2017; Morineau et al., 2017; Oku-Zaki et al., 2018), and the fungus Verticillium longisporum resistant Arabidopsis thaliana and Brassica napus which contain a unique base transition in the CRT1a gene (calreticulin) (PRÖBSTING et al., 2020). From a detection perspective, these GMOs need to show unique amplification products for which the method was designed (EFSA 2017). It should simultaneously distinguish products from naturally occurring variants, or from products obtained by conventional mutagenesis (ENGL, 2019). While real-time PCR amplification is influenced by the primer size, SNP position, hybridization, Taq polymerase fidelity and other PCR conditions; stable non-refracted thermodynamic energy to enhance specificity hybridization can be achieved by the blocking the last 2 pb of 3' end in strand in primer sequences (HUANG; ARNHEIM; GOODMAN, 1992) (AYYADEVARA; THADEN; SHMOOKLER REIS, 2000).

Whereas grf1-3 and grf8-61 genotypes were exclusive variants, meaning they do not have polymorphic position variants found in the Arabidopsis thaliana 1001 variants database, the results of the in-silico PCR were district for each genotype. The primer set designed for genotype grf1-3 does not distinguish the mutant event from 39 other sequence hits observed in Arabidopsis arenosa, Arabidopsis lyrata, Camelina sativa and Brassica sp. This means, for example, that samples from food mixtures of Arabidopsis thaliana and Camelina sativa, the primer set may not be able to differentiate the modified event of CRISPR-Cas9 for the grf1-3 event. For genotype grf8-61, the results show that the primer set is specific for the detection of the new genotype produced by CRISPR-Cas9 considering the NCBI genetic database search.

Quantitative real-time RT-qPCR method based on hydrolysis probe chemistry is the most widely used methods for GMO detection in Europe (ENGL 2015). In previous studies, amplicon and primer sets were used to identify whether the single point variation in a gene that carries a single nucleotide (adenosine) insertion in rice variety OsMADS26 (locus: Os08g02070). The specificity of the method in distinguishing the modification introduced by gene editing technology using its single variation point was achieved in a method based on 2-plex digital droplet PCR (FRAITURE et al., 2022). Similarly in rice, the identification of genotypes with InDels from 1 to 18 base pairs (bps) for the chlorophyll oxygenase 1 (CAO1) gene region was empirically demonstrated by real-time quantitative (qPCR) and digital droplet (ddPCR) PCR. In both PCR systems, it was possible to estimate genome-edited ingredient content (without in silico approach) in gene edited samples (ZHANG et al., 2021).

However, even when these PCR systems are able to distinguish gene-edited and unmodified parental lines empirically, there is a need to test whether the system is suitable in identifying such GMO events in food mixtures containing several other potential sequence variants. In this case, the value of an in silico approach is to anticipate potential sequence targets already available in genomic databases. A recent controversy shows how relevant it is to perform such prediction analysis. In an earlier study by Chhalliyil and co-authors (2022), empirical data showed specificity of a primer set based on locked nucleic acids (LNA) strategy for detecting and identifying the first commercialized genome-edited plant, the Cibus canola, containing two mutated genes AHAS1C and AHAS3A. Later that same year, Weidner and co-authors (2022) showed that the method might not be specific to the GMO event in case but could also show spurious amplification in other Brassica species.

Whereas the empirical PCR testing is the ultimate proof of method specificity, it is the in silico prediction that is able to anticipate failure to distinguish amplicons from different organisms. Therefore, our study provides a stepwise approach to search for amplicon and primer sets specificity in available databases. Additionally, it allows the elaboration of strategies to increase the amplification specificity in RT-qPCR and/or ddPCR.

Arabidopsis thaliana is a well-sequenced genomic plant model with a broad degree of knowledge of natural variants. However, the search and alignment approach proposed here, as well as in silico PCR testing, identified that other similar sequence occurrences are present at the NCBI-GenBank for the genotype by grf1-3. Therefore, wet lab laboratory testing will have to consider chemical modifications in the PCR system to increase specificity, such as locked nucleic acid

technology (LNA), RNAse-H systems, etc (BROCCANELLO et al., 2018) and/or ddPCR startegies to provide specificity in the detection of mutant and wild type variants simultaneously (Fraiture et al., 2022). Finally, considering that most of the genes of interest can be congruent in different species and that in the future NGT plants will be present in food and feed matrices, it is essential to generated information and also systematize and compile information to be included in GMO databases such as Eugenius, and/or JCR to promote the identification of specific sequences relating these GMO events in the future.

5. Conclusions

The findings suggest that primer set designed to identify the grf1-3 genotype containing a single nucleotide polymorphism might generate amplicons in other species. The DNA sequence mismatches that are present are not sufficient to be considered a event-specific PCR system. Under these conditions, the primer pair does not discriminate the grf1-3 genotype of the species of Arabidopsis thaliana itself, other species of the same genus (A. lirata and other phylogenetically close species (i.e., C. Sativa and Brassica sp). Differently, the primer set for grf8-61 genotype is shown to be sufficiently specific for identifying and discriminating the event modified by CRISPR-Cas9.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. Supplementary File 1. CRISPR-Cas9 amplicon sequences; Supplementary File 2. Oligonucleotides parameters; Supplementary File 3. Results from the GRF1 gene ecotype Columbia (A. thaliana) sequence searches against the forward reverse and probe sequences in the NCBI (nucleotide database).; Supplementary Table 4. Results from the GRF8 gene ecotype Columbia (A. thaliana) sequence searches against the forward reverse and probe sequences in the NCBI (nucleotide database);.

Author Contributions: CBZ worked on the conception of the study; designed the study; carried out sequence alignments; performed investigation; data acquisition and curation; wrote the draft; revised the article. SZA-T: contributed to the whole conception of the study; project supervision; revising the draft; funding acquisition. RON contributed to the conception of the study; funding acquisition; critically revised the draft.

Data Availability Statement: The original contributions presented in this study are publicly available.

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

References

1. AGAPITO-TENFEN, Sarah Z.; OKOLI, Arinze S.; BERNSTEIN, Michael J.; WIKMARK, Odd-gunnar; MYHR, Anne I. Revisiting Risk Governance of GM Plants : The Need to Consider New and Emerging Gene-Editing Techniques. [S. l.], v. 9, n. December, p. 1–16, 2018. DOI: 10.3389/fpls.2018.01874.
2. AHMAR, Sunny et al. A revolution toward gene-editing technology and its application to crop improvement. International Journal of Molecular Sciences, [S. l.], v. 21, n. 16, p. 1–28, 2020. DOI: 10.3390/ijms21165665.
3. ANGULO, Juan et al. Targeted mutagenesis of the Arabidopsis GROWTH-REGULATING FACTOR (GRF) gene family suggests competition of multiplexed sgRNAs for Cas9 apoprotein. bioRxiv, [S. l.], 2020.
4. AYYADEVARA, Srinivas; THADEN, John J.; SHMOOKLER REIS, Robert J. Discrimination of primer 3'-nucleotide mismatch by Taq DNA polymerase during polymerase chain reaction. Analytical Biochemistry, [S. l.], v. 284, n. 1, p. 11–18, 2000. DOI: 10.1006/abio.2000.4635.
5. BOYES, D. C.; ZAYED, A. M.; ASCENZI, R.; MCCASKILL, A. J.; HOFFMAN, N. E.; DAVIS, K. R.; GÖRLACH, J. Growth stage-based phenotypic analysis of Arabidopsis: A model for high throughput functional genomics in plants. Plant Cell, [S. l.], v. 13, n. 7, p. 1499–1510, 2001. DOI: 10.1105/tpc.13.7.1499.
6. BRASIL. DECRETO No 4.680, DE 24 DE ABRIL DE 2003. Brasília, 2003.
7. BROCCANELLO, Chiara; CHIODI, Claudia; FUNK, Andrew; MCGRATH, J. Mitchell; PANELLA, Lee; STEVANATO, Piergiorgio. Comparison of three PCR-based assays for SNP genotyping in plants. Plant

Methods, [S. l.], v. 14, n. 1, p. 1–8, 2018. DOI: 10.1186/s13007-018-0295-6. Disponível em: https://doi.org/10.1186/s13007-018-0295-6.

8. CAPECCHI, Mario R. The new mouse genetics: Altering the genome by gene targeting. Trends in Genetics, [S. l.], v. 5, n. C, p. 70–76, 1989. DOI: 10.1016/0168-9525(89)90029-2.

9. CBD. What is Agricultural Biodiversity? In: 2008, Anais [...]. : Convention on Biological Diversity, 2008. Disponível em: https://www.cbd.int/agro/whatis.shtml.

10. CHEN, Kunling; WANG, Yanpeng; ZHANG, Rui; ZHANG, Huawei; GAO, Caixia. CRISPR/Cas Genome Editing and Precision Plant Breeding in Agriculture. Annual Review of Plant Biology, [S. l.], v. 70, p. 667–697, 2019. DOI: 10.1146/annurev-arplant-050718-100049.

11. CHHALLIYIL, Pradheep; ILVES, Heini; KAZAKOV, Sergei A.; HOWARD, Stephanie J.; JOHNSTON, Brian H.; FAGAN, John. A real-time quantitative PCR method specific for detection and quantification of the first commercialized genome-edited plant. Foods, [S. l.], v. 9, n. 9, 2020. a. DOI: 10.3390/foods9091245.

12. CHHALLIYIL, Pradheep; ILVES, Heini; KAZAKOV, Sergei A.; HOWARD, Stephanie J.; JOHNSTON, Brian H.; FAGAN, John. A Real-Time Quantitative PCR Method Specific for Detection and Quantification of the First Commercialized Genome-Edited Plant. Foods, [S. l.], v. 9, n. 9, p. 1245, 2020. b. DOI: 10.3390/foods9091245. Disponível em: https://www.mdpi.com/2304-8158/9/9/1245.

13. ENGL. Detection of food and feed plant products obtained by new mutagenesis techniques. [S. l.], n. March, p. 21, 2019. Disponível em: https://gmo-crl.jrc.ec.europa.eu/doc/JRC116289-GE-report-ENGL.pdf.

14. EUROPEAN COMMISSION. EC, Science for Environment Policy Future Brief: Synthetic Biology and Biodiversity. [s.l: s.n.]. DOI: 10.2779/976543. Disponível em: http://ec.europa.eu/science-environment-policy.

15. EUROPEAN FOOD SAFETY AUTHORITY, Panel on Genetically Modified Organisms (GMO). Scientific opinion addressing the safety assessment of plants developed using Zinc Finger Nuclease 3 and other Site-Directed Nucleases with similar function. In: EFSA Journal. [s.l: s.n.]. v. 10. DOI: 10.2903/j.efsa.2012.2943.

16. EUROPEAN LABORATORIES. Definition of Minimum Performance Requirements for Analytical Methods of GMO Testing. [S. l.], 2015.

17. FRAITURE, Marie Alice; GUIDERDONI, Emmanuel; MEUNIER, Anne Cécile; PAPAZOVA, Nina; ROOSENS, Nancy H. C. ddPCR strategy to detect a gene-edited plant carrying a single variation point: Technical feasibility and interpretation issues. Food Control, [S. l.], v. 137, n. November 2021, 2022. DOI: 10.1016/j.foodcont.2022.108904.

18. GARNEAU, Josiane E. et al. The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature, [S. l.], v. 468, n. 7320, p. 67–71, 2010. DOI: 10.1038/nature09523.

19. HAURWITZ, Rachel E.; JINEK, Martin; WIEDENHEFT, Blake; ZHOU, Kaihong; DOUDNA, Jennifer A. Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease. Science, [S. l.], v. 329, n. 5997, p. 1355–1358, 2010. DOI: 10.1126/science.1192272. Disponível em: https://www.sciencemag.org/lookup/doi/10.1126/science.1192272.

20. HÖIJER, Ida et al. Amplification-free long-read sequencing reveals unforeseen CRISPR-Cas9 off-target activity. Genome Biology, [S. l.], v. 21, n. 1, p. 1–19, 2020. DOI: 10.1186/s13059-020-02206-w.

21. HOUGS, L. et al. Verification of analytical methods for GMO testing when implementing interlaboratory validated methods. version 2 ed. Luxembourg: JRC 109940, 2017. DOI: 10.2760/645114. Disponível em: https://ec.europa.eu/jrc.

22. HSU, Patrick D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nature Biotechnology, [S. l.], v. 31, n. 9, p. 827–832, 2013. DOI: 10.1038/nbt.2647.

23. HUANG, M. M.; ARNHEIM, N.; GOODMAN, M. F. Extension of base mispairs by Taq DNA polymerase: implications for single. Nucleic.Acids.Res., [S. l.], v. 20, n. 17, p. 4567–4573, 1992.

24. ICSWGSB, International Civil Society Working Group On Synthetic Biology. Synthetic Biology and the CBD Five key decisions for COP 13 & COP-MOP 8. [S. l.], p. 8pp, 2016. Disponível em: www.boell.de/en/2016/11/02/synthetic-biology-and-cbd-five-keydecisions- cop-13-cop-mop-8.

25. ISAAC, R. Stefan; JIANG, Fuguo; DOUDNA, Jennifer A.; LIM, Wendell A.; NARLIKAR, Geeta J.; ALMEIDA, Ricardo. Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. eLife, [S. l.], v. 5, n. APRIL2016, p. 1–14, 2016. DOI: 10.7554/eLife.13450.

26. JINEK, Martin; EAST, Alexandra; CHENG, Aaron; LIN, Steven; MA, Enbo; DOUDNA, Jennifer. RNA-programmed genome editing in human cells. eLife, [S. l.], v. 2013, n. 2, p. 1–9, 2013. DOI: 10.7554/eLife.00471.

27. JONES, Huw D. Regulatory uncertainty over genome editing. Nature Plants, [S. l.], v. 1, n. January, p. 2014–2016, 2015. DOI: 10.1038/nplants.2014.11.

28. KARGINOV, Fedor V.; HANNON, Gregory J. The CRISPR System: Small RNA-Guided Defense in Bacteria and Archaea. Molecular Cell, [S. l.], v. 37, n. 1, p. 7–19, 2010. DOI: 10.1016/j.molcel.2009.12.033. Disponível em: http://dx.doi.org/10.1016/j.molcel.2009.12.033.

29. KIM, Hyongbum; KIM, Jin Soo. A guide to genome engineering with programmable nucleases. Nature Reviews Genetics, [S. l.], v. 15, n. 5, p. 321–334, 2014. DOI: 10.1038/nrg3686. Disponível em: http://dx.doi.org/10.1038/nrg3686.

30. KOO, Taeyoung; LEE, Jungjoon; KIM, Jin-soo. 25. Koo et al. 2015 Measuring and Reducing Off-Target Activities of Programmable Nucleases Including CRISPR-Cas9. MC.pdf. [S. l.], v. 38, n. 6, p. 475–481, 2015.

31. KOONIN, Eugene V.; MAKAROVA, Kira S.; ZHANG, Feng. Diversity, classification and evolution of CRISPR-Cas systems. Current Opinion in Microbiology, [S. l.], v. 37, p. 67–78, 2017. DOI: 10.1016/j.mib.2017.05.008. Disponível em: http://dx.doi.org/10.1016/j.mib.2017.05.008.

32. LAWRENSON, Tom; SHORINOLA, Oluwaseyi; STACEY, Nicola; LI, Chengdao; ØSTERGAARD, Lars; PATRON, Nicola; UAUY, Cristobal; HARWOOD, Wendy. Induction of targeted, heritable mutations in barley and Brassica oleracea using RNA-guided Cas9 nuclease. Genome Biology, [S. l.], v. 16, n. 1, p. 1–13, 2015. DOI: 10.1186/s13059-015-0826-7. Disponível em: http://dx.doi.org/10.1186/s13059-015-0826-7.

33. LO, Yat Tung; SHAW, Pang Chui. DNA-based techniques for authentication of processed food and food supplements. Food Chemistry, [S. l.], v. 240, n. March 2017, p. 767–774, 2018. DOI: 10.1016/j.foodchem.2017.08.022. Disponível em: https://doi.org/10.1016/j.foodchem.2017.08.022.

34. MAKAROVA, Kira S.; GRISHIN, Nick V.; SHABALINA, Svetlana A.; WOLF, Yuri I.; KOONIN, Eugene V. A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biology Direct, [S. l.], v. 1, p. 1–26, 2006. DOI: 10.1186/1745-6150-1-7.

35. MANGHWAR, Hakim; LI, Bo; DING, Xiao; HUSSAIN, Amjad; LINDSEY, Keith; ZHANG, Xianlong; JIN, Shuangxia. CRISPR/Cas Systems in Genome Editing: Methodologies and Tools for sgRNA Design, Off-Target Evaluation, and Strategies to Mitigate Off-Target Effects. Advanced Science, [S. l.], v. 7, n. 6, 2020. DOI: 10.1002/advs.201902312.

36. MISHRA, Rukmini; ZHAO, Kaijun. Genome editing technologies and their applications in crop improvement. Plant Biotechnology Reports, [S. l.], v. 12, n. 2, p. 57–68, 2018. DOI: 10.1007/s11816-018-0472-0. Disponível em: http://dx.doi.org/10.1007/s11816-018-0472-0.

37. MODRZEJEWSKI, Dominik; HARTUNG, Frank; SPRINK, Thorben; KRAUSE, Dörthe; KOHL, Christian; WILHELM, Ralf. What is the available evidence for the range of applications of genome-editing as a new tool for plant trait modification and the potential occurrence of associated off-target effects: A systematic map. Environmental Evidence, [S. l.], v. 8, n. 1, p. 1–33, 2019. DOI: 10.1186/s13750-019-0171-5. Disponível em: https://doi.org/10.1186/s13750-019-0171-5.

38. MORENO-MATEOS, Miguel A.; VEJNAR, Charles E.; BEAUDOIN, Jean Denis; FERNANDEZ, Juan P.; MIS, Emily K.; KHOKHA, Mustafa K.; GIRALDEZ, Antonio J. CRISPRscan: Designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nature Methods, [S. l.], v. 12, n. 10, p. 982–988, 2015. DOI: 10.1038/nmeth.3543.

39. OSAKABE, Yuriko; OSAKABE, Keishi. Genome editing with engineered nucleases in plants. Plant and Cell Physiology, [S. l.], v. 56, n. 3, p. 389–400, 2015. DOI: 10.1093/pcp/pcu170.

40. PRÖBSTING, Michael et al. Loss of function of CRT1a (calreticulin) reduces plant susceptibility to Verticillium longisporum in both Arabidopsis thaliana and oilseed rape (Brassica napus). Plant Biotechnology Journal, [S. l.], v. 18, n. 11, p. 2328–2344, 2020. DOI: 10.1111/pbi.13394.

41. RIBARITS, Alexandra; ECKERSTORFER, Michael; SIMON, Samson; STEPANEK, Walter. Genome-Edited Plants: Opportunities and Challenges for an Anticipatory Detection and Identification Framework. Foods, [S. l.], v. 10, n. 2, p. 430, 2021. DOI: 10.3390/foods10020430. Disponível em: https://www.mdpi.com/2304-8158/10/2/430.

42. SALSMAN, Jayme; DELLAIRE, Graham. Precision genome editing in the CRISPR era. Biochemistry and Cell Biology, [S. l.], v. 95, n. 2, p. 187–201, 2017. DOI: 10.1139/bcb-2016-0137.

43. SHELAKE, Rahul Mahadev; PRAMANIK, Dibyajyoti; KIM, Jae Yean. Evolution of plant mutagenesis tools: a shifting paradigm from random to targeted genome editing. Plant Biotechnology Reports, [S. l.], v. 13, n. 5, p. 423–445, 2019. DOI: 10.1007/s11816-019-00562-z. Disponível em: https://doi.org/10.1007/s11816-019-00562-z.

44. SHILLITO, Raymond D.; WHITT, Sherry; ROSS, Margit; GHAVAMI, Farhad; DE VLEESSCHAUWER, David; D'HALLUIN, Katelijn; VAN HOECKE, Annelies; MEULEWAETER, Frank. Detection of genome edits in plants—from editing to seed. In Vitro Cellular and Developmental Biology - Plant, [S. l.], v. 57, n. 4, p. 595–608, 2021. DOI: 10.1007/s11627-021-10214-z.

45. SHRIVASTAV, Meena; DE HARO, Leyma P.; NICKOLOFF, Jac A. Regulation of DNA double-strand break repair pathway choice. Cell Research, [S. l.], v. 18, n. 1, p. 134–147, 2008. DOI: 10.1038/cr.2007.111.

46. TSAI, Shengdar Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nature Biotechnology, [S. l.], v. 33, n. 2, p. 187–198, 2015. a. DOI: 10.1038/nbt.3117.

47. TSAI, Shengdar Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nature Biotechnology, [S. l.], v. 33, n. 2, p. 187–198, 2015. b. DOI: 10.1038/nbt.3117.

48. VERKUIJL, Sebald AN; ROTS, Marianne G. The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies. Current Opinion in Biotechnology, [S. l.], v. 55, p. 68–73, 2019. DOI: 10.1016/j.copbio.2018.07.005. Disponível em: https://doi.org/10.1016/j.copbio.2018.07.005.

49. WAHLER, Daniela; SCHAUSER, Leif; BENDIEK, Joachim; GROHMANN, Lutz. 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, [S. l.], v. 6, n. 6, p. 1718–1727, 2013. DOI: 10.1007/s12161-013-9673-x.

50. WANG, Haifeng; LA RUSSA, Marie; QI, Lei S. CRISPR/Cas9 in Genome Editing and Beyond. Annual Review of Biochemistry, [S. l.], v. 85, n. 1, p. 227–264, 2016. DOI: 10.1146/annurev-biochem-060815-014607.

51. WEIDNER, Christopher et al. Assessment of the Real-Time PCR Method Claiming to be Specific for Detection and Quantification of the First Commercialised Genome-Edited Plant. Food Analytical Methods, [S. l.], v. 15, n. 8, p. 2107–2125, 2022. DOI: 10.1007/s12161-022-02237-y. Disponível em: https://doi.org/10.1007/s12161-022-02237-y.

52. WENDY, Cannon; PEDERSON, David. Mechanisms and conseqeunces of double-strand break formation in chromatin. Physiology & behavior, [S. l.], v. 176, n. 5, p. 139–148, 2017. DOI: 10.1002/jcp.25048.Mechanisms.

53. WIEDENHEFT, Blake; STERNBERG, Samuel H.; DOUDNA, Jennifer A. RNA-guided genetic silencing systems in bacteria and archaea. Nature, [S. l.], v. 482, n. 7385, p. 331–338, 2012. DOI: 10.1038/nature10886.

54. WILSON, Laurence O. W.; O'BRIEN, Aidan R.; BAUER, Denis C. The current state and future of CRISPR-Cas9 gRNA design tools. Frontiers in Pharmacology, [S. l.], v. 9, n. JUN, p. 1–6, 2018. DOI: 10.3389/fphar.2018.00749.

55. YAN, Jifang; XUE, Dongyu; CHUAI, Guohui; GAO, Yuli; ZHANG, Gongchen; LIU, Qi. Benchmarking and integrating genome-wide CRISPR off-target detection and prediction. Nucleic acids research, [S. l.], v. 48, n. 20, p. 11370–11379, 2020. DOI: 10.1093/nar/gkaa930.

56. ZHANG, Hongwen et al. An editing-site-specific pcr method for detection and quantification of cao1-edited rice. Foods, [S. l.], v. 10, n. 6, p. 1–14, 2021. DOI: 10.3390/foods10061209.

57. ZISCHEWSKI, Julia; FISCHER, Rainer; BORTESI, Luisa. Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. Biotechnology Advances, [S. l.], v. 35, n. 1, p. 95–104, 2017. DOI: 10.1016/j.biotechadv.2016.12.003. Disponível em: http://dx.doi.org/10.1016/j.biotechadv.2016.12.003.