ORIGINAL PAPER

Integrated approach for the molecular characterization of edited plants obtained via *Agrobacterium tumefaciens*-mediated gene transfer

Lorenza Dalla Costa[1] · Daniela Vinciguerra[2] · Lisa Giacomelli[1] · Umberto Salvagnin[1] · Stefano Piazza[1] · Katia Spinella[2] · Mickael Malnoy[1] · Claudio Moser[1] · Ugo Marchesi[2]

Abstract

Agrobacterium tumefaciens-mediated gene transfer—actually the most used method to engineer plants—may lead to integration of multiple copies of T-DNA in the plant genome, as well as to chimeric tissues composed of modified cells and wild type cells. A molecular characterization of the transformed lines is thus a good practice to select the best ones for further investigation. Nowadays, several quantitative and semi-quantitative techniques are available to estimate the copy number (CN) of the T-DNA in genetically modified plants. In this study, we compared three methods based on (1) real-time polymerase chain reaction (qPCR), (2) droplet digital PCR (ddPCR), and (3) next generation sequencing (NGS), to carry out a molecular characterization of grapevine edited lines. These lines contain a knock-out mutation, obtained via CRISPR/Cas9 technology, in genes involved in plant susceptibility to two important mildew diseases of grapevine. According to our results, qPCR and ddPCR outputs are largely in agreement in terms of accuracy, especially for low CN values, while ddPCR resulted more precise than qPCR. With regard to the NGS analysis, the CNs detected with this method were often not consistent with those calculated by qPCR and ddPCR, and NGS was not able to discriminate the integration points in three out of ten lines. Nevertheless, the NGS method can positively identify T-DNA truncations or the presence of tandem/inverted repeats, providing distinct and relevant information about the transgene integration asset. Moreover, the expression analysis of Cas9 and single guide RNA (sgRNA), and the sequencing of the target site added new information to be related to CN data. This work, by reporting a practical case-study on grapevine edited lines, explores pros and cons of the most advanced diagnostic techniques available for the precocious selection of the proper transgenic material. The results may be of interest both to scientists developing new transgenic lines, and to laboratories in charge of GMO control.

Keywords GMO · Genome editing · Copy number quantification · qPCR · ddPCR · NGS

✉ Lorenza Dalla Costa
lorenza.dallacosta@fmach.it

Daniela Vinciguerra
danielavinciguerra23@gmail.com

Lisa Giacomelli
lisa.giacomelli@fmach.it

Umberto Salvagnin
umberto.salvagnin@gmail.com

Stefano Piazza
stefano.piazza@fmach.it

Katia Spinella
katia.spinella-esterno@izslt.it

Mickael Malnoy
mickael.malnoy@fmach.it

Claudio Moser
claudio.moser@fmach.it

Ugo Marchesi
ugo.marchesi@izslt.it

[1] Research and Innovation Centre, Fondazione Edmund Mach, Via E. Mach 1, 38098 San Michele all'Adige, Italy

[2] Istituto Zooprofilattico Sperimentale del Lazio e della Toscana "M. Aleandri" Centro di Referenza Nazionale per la Ricerca di Organismi Geneticamente Modificati (CROGM), Rome, Italy

Introduction

The CRISPR-Cas9 technology has allowed public institutions and private companies to boost the development of crops more tolerant to diseases and environmental stresses, or with enhanced nutritional value in order to face global issues that are threatening agriculture and human life itself. Some genome edited plants, like the Calyxt Soybean with higher oleic acid content, is already cultivated and commercialized in the USA under conventional regulation [1, 2]. In fact, since 2016, the USA Department of Agriculture (USDA Animal and Plant Health Inspection Service-APHIS) has exempted more than 35 gene-edited products from GMO legislation. Moreover, in May 2020, USDA issued the SECURE rule, an updated revision of the biotechnology regulation, according to which crops whose genetic changes could have been produced with conventional breeding will be exempted from GMO legislation. Argentina, Brazil, Chile, Colombia, Honduras, Israel, Japan, and Australia have also decided to exempt edited plant varieties from GMO regulation, as long as no residual exogenous DNA remains integrated into the plant genome [3]. Conversely, the European Union has a more restrictive position towards genome editing and considers organisms obtained by such technology as GMOs even if they do not contain additional exogenous sequences in the genome [3]. For vegetatively propagated tree-species, the application of systems that either avoid the use of foreign genes or that remove them when no longer needed is not an easy task. On one hand, the use of protoplasts with RNPs, which avoids the introduction of exogenous DNA, suffers from the scarce availability of protocols for plant regeneration starting from single cells. On the other hand, the classical gene transfer mediated by *Agrobacterium* leads to T-DNA integration in the plant genome, and the strategies adopted so far failed to remove it entirely [4]. However, despite the research focused on obtaining transgene-free edited fruit-trees is ongoing [5–7], in Europe these crops are still considered GMO, regardless of the worldwide legislative landscape. The legal framework of the European Union (Directive 2001/18/EC, Regulation (EC) No 1829/2003; Regulation (EC) No 1830/2003) demands for event-specific methods for detection and quantification of GMO prior to their authorization and placing on the market. Nowadays, several techniques are available thanks to huge progresses in the field of molecular diagnostics. GMO analytical control is currently implemented with DNA-based methods, among which the real-time or quantitative PCR (qPCR) is the most used one for quantitative purposes because of its higher sensitivity and robustness even on complex and processed food matrices [8]. Many qPCR-based assays have been developed for the quantification of the transgene copy number (CN) in fruit crops [9–13]. However, qPCR can be affected by many factors, including the presence of inhibitors in the sample matrix and the reliability of the calibration system [14]. An alternative method is digital PCR [15, 16], in particular the droplet digital (ddPCR), a technique that relies on partitioning the sample into several thousands or even millions of individual droplets, and uses a flow-cytometry-like system to count positive PCR tests [17, 18]. ddPCR, unlike qPCR, does not require a calibration or a highly efficient amplification performance, and estimates the number of targeted copies per reaction using Poisson statistics. In addition, recent methods based on next generation sequencing have been used to detect exogenous genes in transgenic plants [19–21]. These were semi-quantitative techniques developed mainly for the safety assessment of GMO crops or to screen mutant collections in the framework of forward genetic studies.

In this work, qPCR, ddPCR, and NGS methods were assessed for their performances in the quantification of T-DNA integration copies and pros and cons of each technique are discussed. To operate under the best experimental conditions, we chose fresh plant tissue as testing material. In particular, we analyzed ten grapevine transgenic lines, which were transformed with *Agrobacterium tumefaciens* carrying a binary vector for genome editing. Three binary vectors were used for the transformation experiments, each carrying the CRISPR/SpCas9 system with a different sgRNA to target putative disease-susceptibility genes. One sgRNA was directed against *VviMLO7* (VIT_13s0019g04060), needed for susceptibility to powdery mildew, caused by the fungus *Erysiphe necator* [22], while the other two sgRNAs were directed against two different target sites (ts-1 and ts-2) of *VviDMR6* (VIT_16s0098g00860), a gene likely involved in the onset of downy mildew triggered by the oomycete *Plasmopara viticola* [23]. In addition, the expression of *SpCas9* and sgRNA and the mutation profile in the genomic target sites were evaluated to correlate the CN information with the activity of the sgRNA/Cas9 complex. Besides a technical dissertation about the most advanced diagnostic techniques nowadays available for CN quantification and food control, this study aims primarily to provide developers with tools for selecting the proper transgenic material at an early stage, to be propagated and maintained for research studies or commercial purposes.

Materials and methods

Plant material and binary vector for gene transfer

Gene transfer was performed via co-cultivation of embryogenic calli from the table grape varieties 'Crimson seedless'

(CS) and 'Sugraone' (S) with the *A. t.* strain EHA105 pCH32 containing the binary vector for expression of 35S::Cas9 and Ubi::sgRNA from Addgene (www.addgene. org). In particular, three binary vectors were used which differ for the 20-mer sgRNA present in the T-DNA cassette: (1) MLO7-sgRNA = ACTTGAAGAGCGTAGTTT GG; (2) DMR6-ts1-sgRNA = GCCGATGCTTGCAGG CTCTA and (3) DMR6-ts2-sgRNA = GTCCTTGCCGAG GTCGATTA. Actively growing Agrobacterium cultures, previously induced with 100 μM acetosyringone (AS) for 3 h, were re-suspended in liquid GSICA [24] with AS to an OD600 = 0.3–0.45, and then mixed with about 5 g of embryogenic callus in a volume of 20 mL. The co-culture was shook for 10 min at 25 °C (60 rpm), pelleted, blotted on sterile Whatman paper, transferred to GS1CA solid medium, and then incubated at 25 °C in the dark for 48 h. Embryogenic callus was then washed in liquid GS1CA supplemented with 1 g/L Timentin, spun and blotted on paper, transferred to solid GS1CA medium supplemented with 1 g/L Timentin, and maintained at 25 °C in the dark for 4 weeks. Thereafter, the callus was subcultured monthly on GS1CA medium supplemented with 1 g/L Timentin and 150 mg/L Kanamycin in the dark at 25 °C for 8 months. Developed embryos (torpedo stage) were transferred to NN medium [25] supplemented with 25 mg/L Kanamycin, in the light (16 h photoperiod) to induce embryo differentiation and germination. Transgenic shoots were transferred in WP medium [26] and subcultured every two months. 'CS' lines 1, 2, 4, 5, 6 were edited to knock-out *VviMLO7* gene while 'S' lines 3, 7, 8, 9, 10 were edited to knock-out *VviDMR6* gene, in particular the sgRNA was directed against target site 1 (ts-1) in lines 3, 7, 9, 10 and against target site 2 (ts-2) in line 8. For each transgenic lines, two in vitro biological replicates (A and B) were analyzed. For each sample, DNA was extracted from 3 leaves of a 6 cm-tall plantlet using Nucleospin Plant II (Macherey–Nagel, Düren, Germany). DNA concentrations were measured with the Picogreen dye (Thermo Scientific) in the NanoDrop 3300 Fluorospectrometer (Thermo Scientific) following the manufacturer protocol "PicoGreen Assay for dsDNA".

Real-time PCR quantification

The real-time PCR quantification of *nptII* or *SpCas9* CNs in grapevine lines was carried out according to the method developed by [10] and following the scheme reported in Online Resource 1. Reactions were performed in a 96-well plate on a C1000 thermal cycler (Bio-Rad, Hercules, USA) equipped with CFX96 real-time PCR detection system (Bio-Rad, Hercules, USA). The real-time PCR singleplex reaction was carried out in a 10 μl final volume containing 1×SsoAdvanced Universal Probes Supermix (Bio-Rad, Hercules, USA), 40 ng of genomic DNA, 0.3 μM primers (Sigma, Haverhill, UK)

and a 0.2 μM specific Taqman probe (Sigma, Haverhill, UK). The thermal protocol was as follows: polymerase activation for 3 min at 95 °C followed by 40 cycles of denaturation of 10 s at 95 °C, annealing of 5 s at 58 °C and 5 s at 60 °C and an elongation of 30 s at 72 °C. Primers and Taqman probes used to amplify grapevine endogenous gene *Chi* and exogenous genes *nptII* and *Cas9* are reported in Online Resource 2. The standard curves (four points, starting from 10^6 plasmid molecules and adopting a serial dilution of 1:5) were built with a plasmid pGEM-T easy (Promega, Madison, Wisconsin, USA) containing specific fragments of the genes to be quantified (*VviChi*, *nptII* and *SpCas9*). For each sample, the transgene (*nptII* or *SpCas9*) CN was calculated using the following formula: (transgene total copies/endogenous gene total copies)×2. The total copies of the transgene and of the endogenous gene were calculated on the basis of the mean values of the quantification cycles (Cq) of two technical replicates.

ddPCR quantification

The same DNA samples analyzed by real-time PCR were tested in droplet digital PCR (ddPCR). Each sample (10 lines, 2 biological replicates per line) was analyzed as detailed in Online Resource 1. The 2 technical replicates in a single plate were treated individually, being ddPCR independent from calibration. The reaction volume was 20 μL containing 1× ddPCR Supermix for probes (No dUTP) (Bio-Rad, Pleasanton, CA, USA), 300 nM of primers, 200 nM of probes and 6 ng of template DNA, for both the *VvChi* and *Cas9* assays. Droplet generation was carried out in DG8 cartridges (Bio-Rad) loaded on a QX100 droplet generator (Bio-Rad Laboratories, Inc.). Droplets were then transferred to 96-well plates amplified in a PCR thermocycler GeneAmp PCR System 9700 (Thermo Fisher Scientific, Waltham, MA, USA) with the following thermal profile: 10 min DNA polymerase activation at 95 °C, 45 cycles of a two-step thermal profile of 30 s at 94 °C for denaturation, and 60 s at 55 °C for annealing and extension, droplets stabilization at 98 °C for 10 min followed by an infinite hold at 4 °C. After thermal cycling, the 96-well plates were transferred to a QX100 droplet reader (Bio-Rad Laboratories, Inc.) and data were analyzed with QuantaSoft 1.7.4 software (Bio-Rad Laboratories, Inc.). The results obtained from each replicate had to fulfil the following acceptability criteria to be included in the subsequent elaborations: single peak of amplitude for the positive droplets; a number of accepted droplets above 10.000 per well; an amount of "rain" below 2.5% per well.

NGS method for the identification of T-DNA integration site and bioinformatics pipeline

The NGS method was performed according to [4]. Genomic DNA (1 μg) was reduced to fragments ranging

between 200 and 1000 bp using BIORUPTOR NextGen (Diagenode, Seraing, Belgium) with three cycles of 30 s at low intensity. The DNA fragmentation profile was checked on Tapestation 2200 (Agilent, Santa Clara, CA, USA) using D1000 or Genomic DNA ScreenTape. Fragmented DNA was purified with AMPure XP beads (Beckman, Brea, CA, USA) at $1.8 \times$ ratio, treated with NEBNext End Repair Module E6050S (New England Biolabs, Ipswich, MA, USA) and again purified with $1.8 \times$ AMPure XP beads. Purified end-repaired fragments were quantified with D1000 ScreenTape on Tapestation. Genomic blunt fragments were ligated to the Adaptors of the Universal GenomeWalker 2.0 kit (Takara Bio, Kusatsu, Japan) using T4 Ligase (Thermo Scientific, Waltham, MA, USA), and following the manufacturer's instructions. A PCR was performed in a 20 µl final volume containing $1 \times$ PCR BIO (Resnova, Rome, Italy), 0.25 µM of the primers ADAP_ill and PNOS_ill (see Online Resource 2) and 20 ng genomic DNA. PCR products were purified with $0.8 \times$ AMPure XP Beads to remove fragments smaller than 200 bp, primers and primer dimers. The library was sequenced by Illumina MiSeq (PE300) in house, at the FEM Sequencing Platform Facility (San Michele all'Adige, Italy). Approximately 100,000 reads per sample were produced. Two datasets of raw sequencing reads (fastQ files for both ends) were analyzed using VSEARCH 2.13.4 and BLAST 2.9.0 (https://blast.ncbi.nlm.nih.gov/Blast.cgi) software. Reads of Dataset 1, amplified with the primer ADAP_ill, were trimmed of 48 bp to remove the GenomeWalker adaptor sequence and then merged with the reads of dataset 2, amplified with the primer PNOS_ill (minimum overlapping $= 50$ bp); merged sequences were then clustered using an identity threshold (ID) minimum of 0.90. To identify exogenous sequences, clusters were mapped to the T-DNA vector sequence using BLAST and filtered for length of the alignment > 50 bp and e value (random background noise) below 0.01. Filtered sequences were then mapped against the reference genome, and hits with less than 10 mismatches and an e value above 10^{-6} were selected. According to the BLAST output, the T-DNA integration points were identified, and confirmed by PCR amplification of the junction regions between genomic DNA and T-DNA. PCR was performed in a 20 µl final volume containing $1 \times$ PCR BIO (Resnova, Rome, Italy), 50 ng of genomic DNA and 0.5 µM of the primers reported in Online Resource 2. Amplification products were checked on agarose gel, purified using PureLink Quick Gel Extraction (Invitrogen, Carlsbad, CA, USA) or PCR Purification Combo Kit (Thermo Scientific, Waltham, MA, USA) and sequenced by Sanger sequencing (FEM Sequencing Platform Facility). Sequencing outputs were analyzed with the BLAST online tool (blast.ncbi.nlm.nih.gov).

Gene expression analysis

For each line, the two in vitro biological replicates used for DNA measurements were micro-propagated (for a cycle of 1 month) and three leaves were collected from the respective daughter plants, and used for RNA extraction. Total RNA was isolated from grapevine leaves using the Spectrum Plant Total RNA Kit (Sigma–Aldrich, St. Louis, MO, USA). RNA was quantified with the spectrophotometer NanoDrop ND-8000 (NanoDrop Technologies, Wilmington, DE, USA) and by gel electrophoresis. Following DNase treatment, 1 µg of RNA was reverse-transcribed into cDNA with the SuperScript III Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) using random primers according to manufacturer's instructions. The real-time PCR was carried out on the CFX96 instrument (Biorad, Hercules, CA, USA) in a 12.5 µl volume containing SsoAdvanced Universal SYBR Green Supermix (Bio-rad, Hercules, CA, USA), 0.5 µM primers (Online Resource 2) and 1 µl of diluted cDNA (1:10). An initial denaturation step at 95 °C for 5 min was followed by 40 cycles at 95 °C for 10 s and 60 °C for 30 s. Finally, to detect non-specific amplification in cDNA samples a melting curve analysis was performed as follows: 95 °C for 10 s, 65 °C for 5 s and a stepwise T increase (0.5 °C/s) up to 95 °C with a continuous detection. Glyceraldehyde 3-phospate dehydrogenase (GAPDH) and Actin (Online Resource 2) were used as housekeeping genes. For each biological replicate of each plant line, three technical replicates were run in a single plate and data were elaborated using the software Bio-rad CFX Manager 3.0.

Editing evaluation on the CRISPR/SpCas9 target sites

Regions of the *VviMLO7* and *VviDMR6* genes containing the target sites were amplified with the primers reported in Online Resource 2. PCR was carried out in 20 µl final volume containing $1 \times$ PCR BIO (Resnova, Rome, Italy), 0.4 µM of each primer (both elongated with overhang Illumina adapters) and 50 ng of genomic DNA. The Illumina library was sequenced on an Illumina MiSeq (PE300) platform at the FEM Sequencing Platform Facility (San Michele all'Adige, Italy). Raw paired-end reads were processed by the CRISPResso pipeline (http://crispresso.rocks/) with default parameters.

Statistics

Statistical analyses were performed using R software (R Core Team 2020).

Results

Transgene copy number quantification: qPCR vs ddPCR

Edited grapevine plantlets were regenerated from single somatic embryos of cv 'Sugraone' and cv 'Crimson seedless' after *Agrobacterium* mediated transformation of embryogenic callus. *SpCas9* CN was quantified in two biological replicates for each line by qPCR and ddPCR. In addition, the selection marker gene *nptII* was quantified by qPCR (Fig. 1). As shown by statistical analysis (Table 1), the CN values of *SpCas9* and *nptII* measured with qPCR were not significantly different. Conversely, by comparing CN data quantified with the two techniques,

significant differences were found. As expected, these discrepancies were more pronounced when the comparison was done across different genes (*SpCas9*_ddPCR vs *nptII*_qPCR) rather than within the same gene (*SpCas9*_ddPCR vs *SpCas9*_qPCR) (Table 1). Being qPCR affected more than ddPCR in terms of amplification performances by the presence of single nucleotide polymorphisms occurring in the endogenous gene *VvChi* (e.g. due to somatic mutations caused by in vitro culturing of embryogenic callus cells), we checked the sequence of the amplicons in the edited lines. As reported in Online Resource 3, no differences were found in the sequence of *VvChi* recognized by the primer and probe set used for qPCR and ddPCR. In general, mean CN values ranged between one and two copies, with values higher than 2 in three lines (lines 3, 4, 7). For those lines, high standard errors were calculated, mainly

Fig. 1 *SpCas9* and *nptII* CN measured with real-time PCR (qPCR) and droplet digital PCR (ddPCR) in 10 grapevine edited lines (L1-L10). qPCR CN values are the average of six measures calculated by analyzing two biological replicates in three independent PCR sessions. ddPCR CN values are the average of 12 values calculated by analyzing two biological replicates in three independent PCR sessions. Bars represent standard deviation of the mean of the two biological replicates
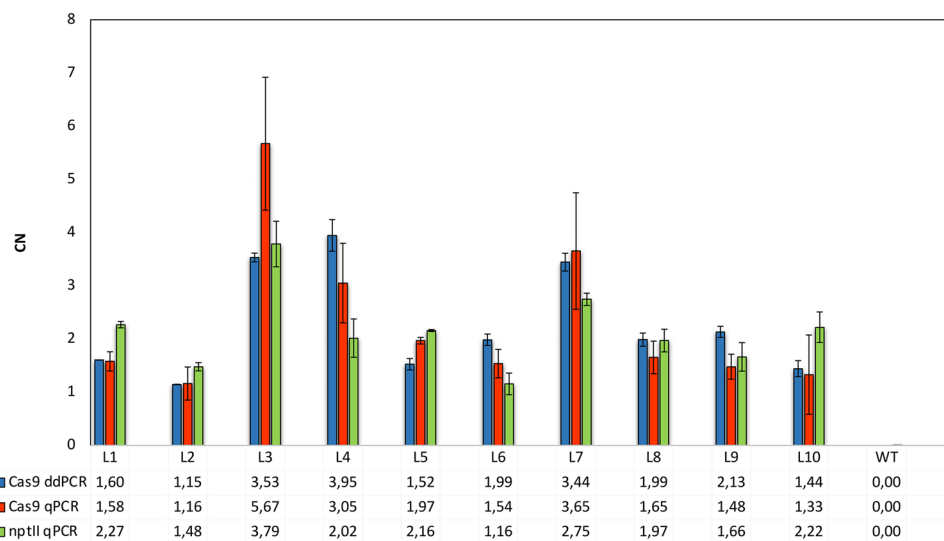

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	WT
■ Cas9 ddPCR	1,60	1,15	3,53	3,95	1,52	1,99	3,44	1,99	2,13	1,44	0,00
■ Cas9 qPCR	1,58	1,16	5,67	3,05	1,97	1,54	3,65	1,65	1,48	1,33	0,00
■ nptII qPCR	2,27	1,48	3,79	2,02	2,16	1,16	2,75	1,97	1,66	2,22	0,00

Table 1 Comparison of ddPCR and qPCR CN data by non-parametric analysis of variance (Kruskal–Wallis test) followed by a pairwise comparison (Wilcoxon test)

Line ID	ANOVA	Pairwise comparison		
	p value	*SpCas9*_ddPCR vs *SpCas9*_qPCR	*SpCas9*_ddPCR vs *nptII*_qPCR	*SpCas9*_qPCR vs *nptII*_qPCR
L1	0.00556**	0.96359	0.00032***	0.26948
L2	0.0738	0.892	0.055	0.269
L3	0.00429**	0.0039**	0.3355	0.3355
L4	0.000236***	0.03668*	0.00032***	0.09307
L5	0.00398**	0.0367*	0.0023**	0.6991
L6	0.00273**	0.3939	0.0013 **	0.3939
L7	0.191	0.38	0.36	0.36
L8	0.233	0.20	0.37	0.59
L9	0.00258**	0.0367*	0.0013**	0.9372
L10	0.0576	0.079	0.073	0.240

Each individual line was considered separately and for each of them, the CN data of the two biological replicates calculated with a specific method were grouped together. Asterisks represent the significance level of *p* values: *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$

associated to the *SpCas9* CN measured by qPCR (Fig. 1). During T-DNA integration in the plant genome, a frequent occurrence of tandem/inverted-repeated insertions of the entire T-DNA or of its fragments has been reported [4, 27]. In this scenario, ddPCR, unlike qPCR, would probably not distinguish the tandem/inverted multiple copies from a single copy. This occurrence was further investigated in line 3, for which the value of *SpCas9* CN calculated by ddPCR differed strongly from that of the qPCR output. Both biological replicates of line 3 were subject to genomic DNA digestion using the *BsmBI* restriction enzyme which cuts twice between *nptII* and *SpCas9* (Online Resource 4) separating putative multiple copies in tandem. On the digested gDNA (dDNA), ddPCR measured a *SpCas9* CN that is nearly double as compared to the value measured on not-digested gDNA (NdDNA). Mean *SpCas9* CN in NdDNA resulted equal to 3.47 in replicate 3A, and 3.59 in replicate B while mean *SpCas9* CN in dDNA resulted equal to 6.06 in replicate 3A, and 5.49 in replicate B. These latter values are not statistically different from the *SpCas9* CN mean value estimated by qPCR (i.e. 6.55 in replicate A and 4.79 in replicate B).

Precision of qPCR and ddPCR data

To compare the precision performances of qPCR and ddPCR, the standard deviations (SD) associated to the mean

CN values calculated for the two replicates of each line were plotted (Fig. 2). SDs of qPCR data were substantially more scattered than those of ddPCR, highlighting the higher precision of the ddPCR technique compared to qPCR. In addition, a bootstrap test on SD values was performed to statistically evaluate precision. The bootstrap output shown in Table 2 indicates that the 95% confidence intervals of ddPCR and qPCR did not overlap for line 1 through 6 and attested that the standard deviations of the two methods were significantly different, being the precision of ddPCR significantly higher than that of qPCR. For lines from 7 to 10 a small rate of overlap was found, which reduced the level of significance of the differences.

T-DNA integration point

T-DNA integration points were assessed in the ten grapevine lines with a method based on high throughput sequencing [4] and are described in Fig. 3. The genomic position of the T-DNA insertion was identified in 7 lines (lines 1, 2, 3, 4, 5, 9, 10), and in the case of line 9 two integration points were detected. In all cases integration occurred in intergenic regions. In addition, T-DNA rearrangements were detected. A tandem repeat T-DNA insertion with a LB-head preceded by a RB-tail was observed in lines 3, 7 and 10. Moreover, upstream the LB region, other unexpected DNA sequences were found: a fragment of the binary vector backbone in
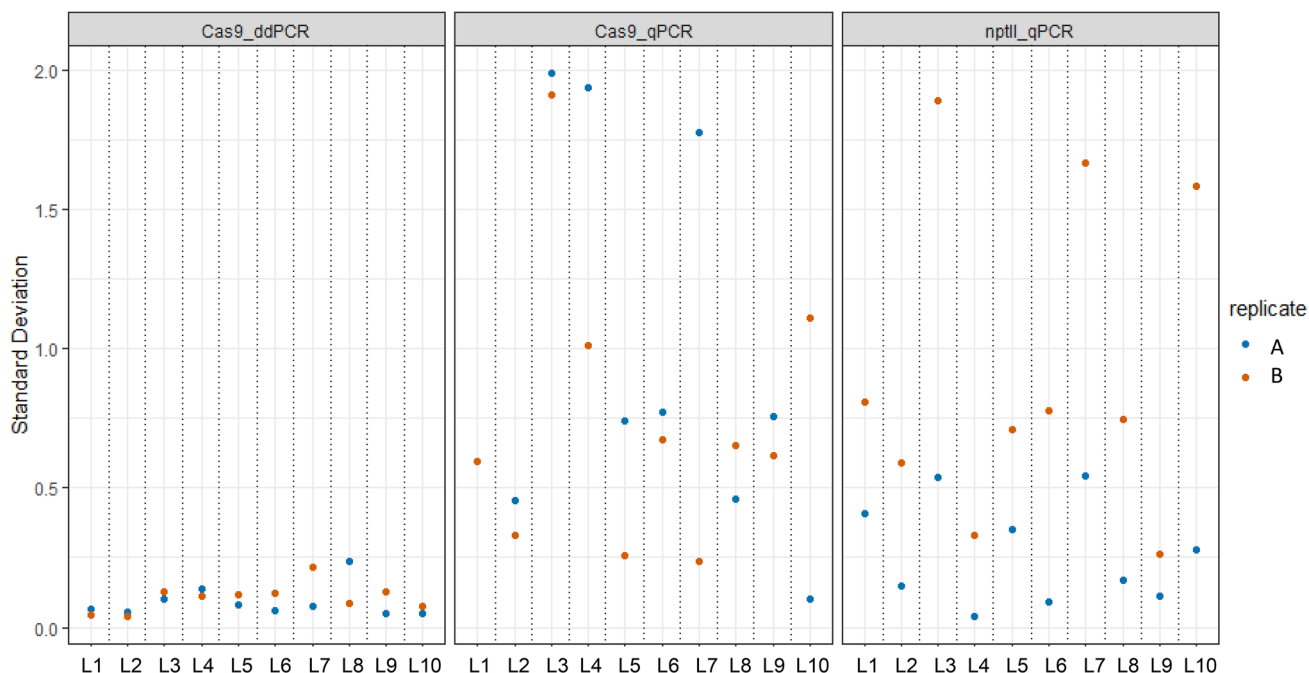

Fig. 2 Distribution of the standard deviations (SDs) associated to the mean CN values. For each replicate (a or b) of each transgenic line, mean values and SD were calculated on three measurements obtained in three independent qPCR sessions and on six measurements obtained in six independent ddPCR sessions

Table 2 Bootstrap estimation of 95% confidence interval (CI 2.5%–97.5%) applied to the dataset of standard deviations derived by qPCR and ddPCR CN quantifications (see Fig. 2)

Line ID	Method	2.5%	97.5%	Overlapping
L1	qPCR	0.2081556	0.672549	No
	ddPCR	0.02625254	0.075588	
L2	qPCR	0.09954535	0.592412	No
	ddPCR	0.01944881	0.063612	
L3	qPCR	0.8003703	2 .572481	No
	ddPCR	0.0355114	0.179767	
L4	qPCR	0.330765	2 .001029	No
	ddPCR	0.1138215	0.312139	
L5	qPCR	0.194971	0.660951	No
	ddPCR	0.05874009	0.158825	
L6	qPCR	0.1778733	0.83196	No
	ddPCR	0.05446206	0.160876	
L7	qPCR	0.2022415	1 .910962	0.0937
	ddPCR	0.05215993	0.295945	
L8	qPCR	0.1445879	0.763852	0.12509
	ddPCR	0.04901334	0.269677	
L9	qPCR	0.13675	0.716193	0.01387
	ddPCR	0.03477371	0.150617	
L10	qPCR	0.1214098	1 .224454	0.04611
	ddPCR	0.05714405	0.167515	

Using R software, 10 million permutations of random sampling from SD dataset were run, grouped by line and method. The absence of overlapping between the qPCR and ddPCR 95% CI indicates that the precision of qPCR method is significantly lower than the precision of ddPCR

lines 6 and 8 and a fragment of the *SpCas9* gene in lines 3 and 8. In line 3, a truncated copy of *SpCas9* upstream the T-DNA LB border was found, suggesting the probable loss of a substantial region of the T-DNA cassette containing the *nptII* gene. Such result may explain the quite divergent CN of *SpCas9* (i.e. 6.55 and 4.79 for replicates A and B, respectively) and *nptII* (i.e. 4.09 and 3.48 for replicates A and B, respectively) obtained by qPCR. Another interesting result was the detection of an identical genomic integration point in lines 1, 2 and 5 (Chr. 6, position 6,517,569). This outcome proved that the three plants that were believed to derive from independent transformation events, were actually clones, originated from the same transformation event.

SpCas9 and sgRNA expression

SpCas9 and sgRNA expressions were evaluated in two replicates for each plant line. According to the results showed in Fig. 4, a different *SpCas9* expression profile was observed among the 10 lines, with the lowest values for line 4 and the highest values for lines 1, 5, 8, 10. Regarding sgRNA expression, for MLO7-sgRNA the lowest expression was found in
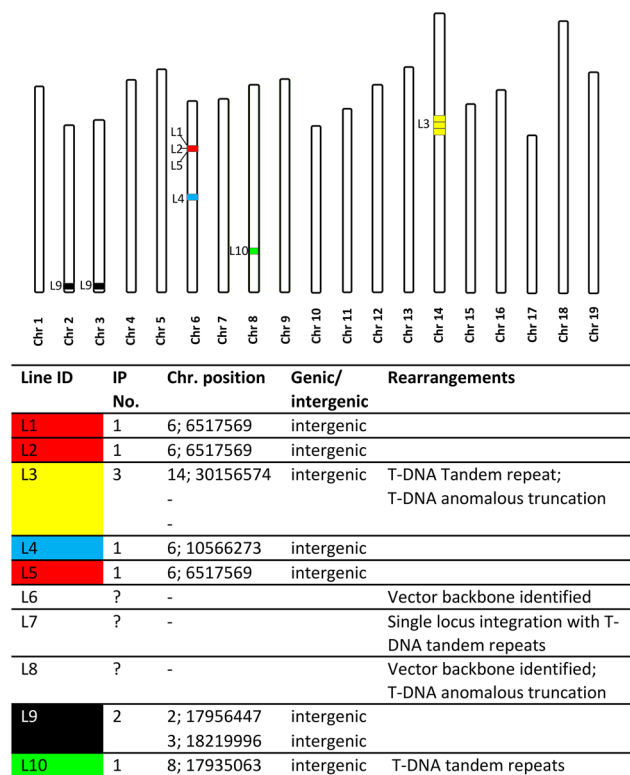
Fig. 3 Overview of the number and pattern of T-DNA integration in the ten edited lines. IP: integration point

line 4 (consistent with a similar lower value for *SpCas9*), while for DMR6-ts1-sgRNA in line 9 (the trend is similar for *SpCas9*). No comparison was possible for DMR6-ts2-sgRNA.

Editing profile in the target site

The editing profile of the three target sites was analyzed by Illumina sequencing (Fig. 5). A full mutated asset was found in all the lines with the exception of lines 4 and 6, which maintained a portion of non-mutated allele (wild type, wt) of 50% and 20% of the sequenced reads, respectively. The editing profiles of lines 1, 2, and 5, actually clones, were very similar. This result indicates that editing may have occurred prior to the division of the embryogenic unit into the progenitors of the three independent clones. Moreover, it is also worth noticing that the lines showing a complete and homogeneous mutation pattern (i.e. lines 3, 9, 10) were those edited with the same sgRNA (DMR6 ts-1).

Discussion

In the last few years gene editing technologies have been massively applied to unveil the function of candidate genes or to improve traits in crops [28–32]. In such experiments,
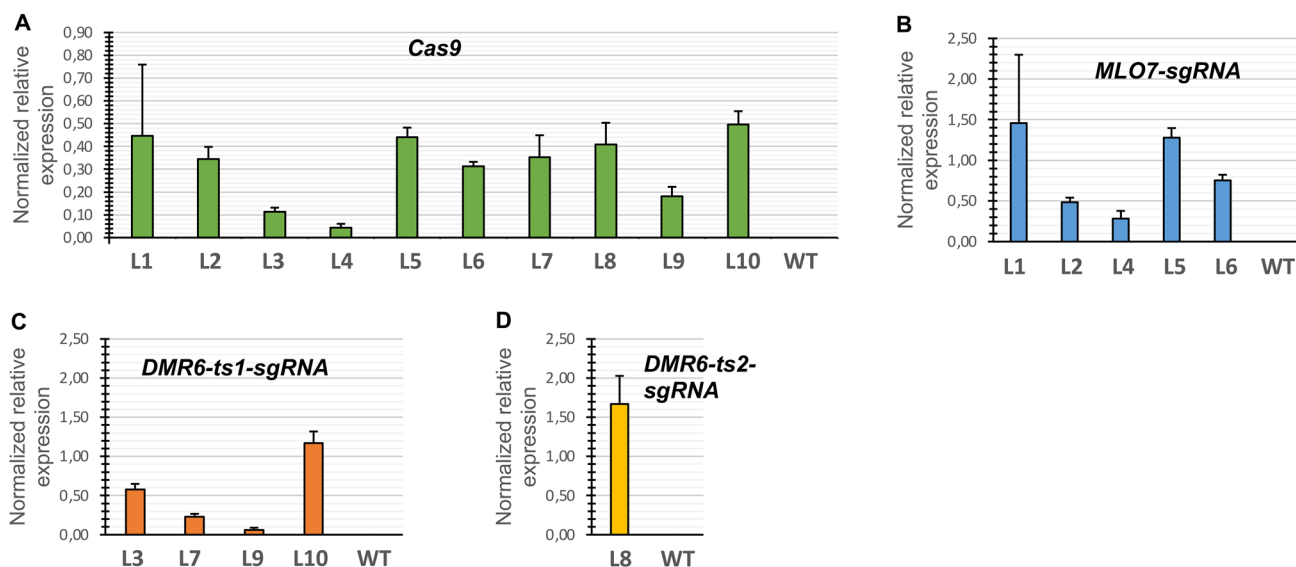
Fig. 4 Expression profile of *SpCas9* gene (**A**) and of sgRNA: MLO7 (**B**), DMR6-ts1 (**C**), DMR6-ts2 (**D**). Error bars indicated SD associated to the mean of three technical replicates

Fig. 5 Mutation profile of the target site detected by Illumina sequencing and CRISPResso software. For each line and biological replicate (**A** and **B**), the analyzed DNA was the same used for qPCR and ddPCR quantification
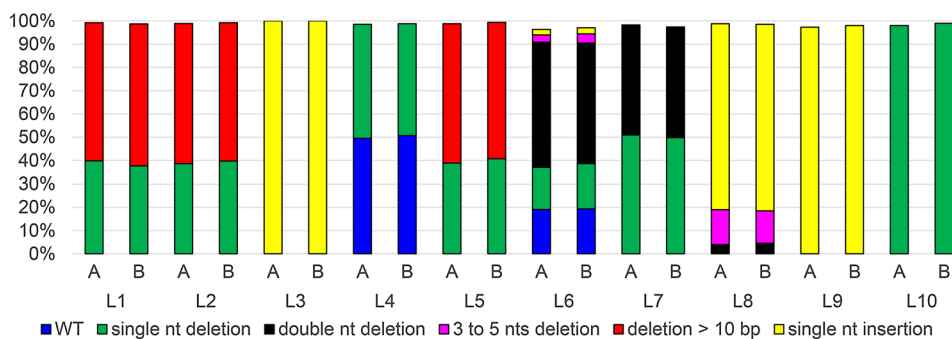
typically a large number of in vitro plants are generated, and, therefore, an early step of selection is required to maintain only the most interesting lines. Molecular characterization of the transformation events and phenotyping of the trait of interest are crucial for such selection process. In this study, we compared and evaluated a set of methods to carry out a molecular characterization of grapevine CRISPR/Cas9-edited lines, which contain inserted recombinant DNA and are, therefore, considered conventional GMOs. *Agrobacterium tumefaciens*-mediated transformation, actually the most common method used to engineer crops, produces lines with a number of integrated T-DNA ranging from one to several copies, as well as chimeric tissues with modified cells mixed to wt cells which results in fractional CN [33]. Both multiple copies and chimerism are undesired, the former because it is often associated with post-transcriptional silencing of the transgene [34, 35] and the latter because it can result in the loss of the trait after many cycles of plant propagation [11, 33, 36, 37]. Thus, an accurate and precise CN measurement is critical for lines selection, being

single or low CN generally desired [38]. Many studies have been carried out to compare the performances of qPCR and ddPCR in view of improving detection and quantification methods intended for those laboratories committed in official GMO control [15, 39–44]. In general, they argued in favor to ddPCR, being this technique insensitive to PCR inhibiting components often present in complex matrices and not dependent on calibration with standard curves obtained from certified reference materials for quantification. According to our results, as for accuracy qPCR and ddPCR outputs are broadly in agreement, especially for low CN values (Fig. 1). The most divergent values were those observed for line 3. In this case, however, ddPCR CN values measured on genomic DNA digested with a restriction enzyme, resulted closer to qPCR ones. This result pointed out that a T-DNA repeated integration pattern, which is a common outcome of *Agrobacterium tumefaciens* and biolistic-mediated transformation, cannot be resolved by ddPCR without separating tandem T-DNA cassettes. However, ddPCR resulted more precise than qPCR in all cases (Fig. 2). Besides qPCR and ddPCR,

a method to assess transgene CN is T-DNA integration site sequencing based on NGS that allows to identify and count unique insert-to-plant junctions at one end of the integration site [4]. A weak point of this method is its relatively low efficiency: it has been possible to discriminate the integration points only in 7 out of 10 lines. Moreover, according to our data, CN detected with NGS were often not consistent with those calculated by qPCR and ddPCR. According to [45], NGS (MiSeq) analysis is more sensitive than qPCR to measure high CNs but the presence of rearrangements may impair its accuracy. In fact, in case of tandem or inverted repeats, the NGS method cannot ensure resolution of multiple copies within a single integration point. Likewise, it cannot identify chimerical T-DNA integration. However, although unreliable for an accurate CN assessment, this method can clearly demonstrate the presence of T-DNA truncations (due to T-DNA border trimming or of tandem/inverted repeats), thus confirming potential hypothesis formulated on the basis of qPCR and ddPCR results, such as the case of line 3 where inconsistency in the CN of two exogenous genes was caused by a T-DNA truncation. These kinds of rearrangements may have a profound effect on inactivation or variable expression of the transgene [46, 47], and their detection is, therefore, crucial. Head-to-tail arrays of multiple copies frequently result in stable expression, whereas head-to-head or tail-to-tail arrangements generally result in silencing [48]. Moreover, the location of the integration point may also give indications about the possible influence of the surrounding genomic sequence on recombinant DNA expression, a feature that is known as position effect [49, 50]. In recent studies [47, 51], random transgene integration sites were analyzed in Chinese hamster ovary cells, mammalian cell lines used as biofactories for the production of therapeutic molecules. These authors found that transgene stability is ensured over time when integration occurred in genomic regions with high transcriptional activity and accessibility to transcription factors (not necessarily within highly transcribed genes). If transgene integration is directed to genomic "landing pads" with such features, also a multicopy integration pattern allowed for stable and high-rate expression of the recombinant DNA [52]. Mapping such regions in plant genomes would be of great importance to predict the behavior of a trait of interest over time in a modified plant. In addition, the available biotechnological tools for targeted gene insertion (i.e. the knock-in approach based on Site Directed Nucleases 3) [53] greatly encourage the identification of such regions in the genomes of crops. The empirical analysis of several events of transformation (i.e. assessment of the integration point and expression stability over long times together with bioinformatics predictions) may help to reach this goal. The knowledge of plant-T-DNA junctions is also very important to understand whether the integration could be detrimental to the plant such those occurring in important coding regions. Another interesting aspect is that NGS methods serves to identify lines that derive from the same transformation event and should be more correctly classified as clones (e.g. lines 1, 2, 5). Such information cannot be retrieved by qPCR and ddPCR, nor by the analysis of the editing profile, because a specific target site tends to be conservatively repaired, often resulting in the same mutation [54]. In the case of edited plants, the expression rate of *SpCas9* and sgRNA and the mutation profile of the target site allow to assess the activity of the CRISPR system integrated in the plant genome and to correlate the T-DNA integration pattern (genomic position, number of copies, presence of rearrangements) with the efficiency of the editing machinery. In our study, a multicopy integration correlates with a low expression of transgenes (in lines 3 and 4) and the lowest *SpCas9* and sgRNA expression is associated with a partial editing in the target site (line 4).

Conclusions

Based on the results obtained in this work, the integrated use of the three proposed techniques has proven to allow the characterization of transgenic plants already at an early stage. qPCR and ddPCR can be considered as alternative techniques to quantify the integration copy number of a transgene. However, their use in parallel can provide more complete information, especially in case of divergence between the results obtained. Indeed, while ddPCR is more precise and substantially unaffected by PCR inhibition, it is less accurate in quantifying tandemly repeated sequences unless digested with appropriate endonucleases. In parallel, the NGS method should be considered as a complementary technique to gather extensive knowledge about the transgene integration asset, which may be crucial for plant selection in the early stages of development.

Author contributions LDC conceived the project, performed qPCR analysis and statistics, and wrote the manuscript. DV performed ddPCR analysis. LG carried out the grapevine gene transfer experiments, regenerated edited lines, carried out Illumina sequencing of the target site in the edited lines, and revised the manuscript. US performed the gene expression analysis and revised the manuscript. SP carried out the analysis for the identification of T-DNA integration points and revised the manuscript. KS made the bibliography research, contributed in the writing of the introduction and revised the manuscript. MM supervised and revised the manuscript. CM supervised and revised the manuscript.

UM conceived the project, supervised and revised the manuscript. All the authors read and approved the final version of this manuscript.

Declarations

Conflict of interest The authors declare no competing interests.

Compliance with ethics requirements This study does not contain any studies with human participants or animals performed by any of the authors.

Availability of data and material The datasets generated during the current study are made available by the corresponding author on reasonable request.

Code availability Not applicable.

References

1. Menz J, Modrzejewski D, Hartung F et al (2020) Genome edited crops touch the market: a view on the global development and regulatory environment. Front Plant Sci 11:586027. https://doi.org/10.3389/fpls.2020.586027

2. Waltz E (2018) With a free pass, CRISPR-edited plants reach market in record time. Nat Biotechnol 36:6–7

3. Schmidt SM, Belisle M, Frommer WB (2020) The evolving landscape around genome editing in agriculture. EMBO Rep 21:19–22. https://doi.org/10.15252/embr.202050680

4. Dalla Costa L, Piazza S, Pompili V et al (2020) Strategies to produce T-DNA free CRISPRed fruit trees via *Agrobacterium tumefaciens* stable gene transfer. Sci Rep 10:20155. https://doi.org/10.1038/s41598-020-77110-1

5. Pompili V, Dalla Costa L, Piazza S et al (2020) Reduced fire blight susceptibility in apple cultivars using a high-efficiency CRISPR/Cas9-FLP/FRT-based gene editing system. Plant Biotechnol J. https://doi.org/10.1111/pbi.13253

6. Charrier A, Vergne E, Dousset N et al (2019) Efficient targeted mutagenesis in apple and first time edition of pear using the CRISPR-Cas9 system. Front Plant Sci 10:40. https://doi.org/10.3389/fpls.2019.00040

7. Metje-Sprink J, Menz J, Modrzejewski D, Sprink T (2019) DNA-Free genome editing: past, present and future. Front Plant Sci 9:1957. https://doi.org/10.3389/fpls.2018.01957

8. Shepherd C, Lauter ANM, Scott MP (2009) Determination of transgene copy number. Methods Mol Biol Transgenic Maize 526:129–134. https://doi.org/10.1007/978-1-59745-494-0_11

9. Chaouachi M, El Malki R, Berard A et al (2008) Development of a real-time PCR method for the differential detection and quantification of four solanaceae in GMO analysis: Potato (Solanum tuberosum), Tomato (Solanum lycopersicum), Eggplant (Solanum melongena), and Pepper (*Capsicum Annuum*). J Agric Food Chem 56:1818–1828. https://doi.org/10.1021/jf073313n

10. Dalla Costa L, Vaccari I, Mandolini M, Martinelli L (2009) Elaboration of a reliable strategy based on real-time PCR to characterize genetically modified plantlets and to evaluate the efficiency of a marker gene removal in grape (*Vitis spp.*). J Agric Food Chem 57:2668–2677. https://doi.org/10.1021/jf802740m

11. Faize M, Faize L, Burgos L (2010) Using quantitative real-time PCR to detect chimeras in transgenic tobacco and apricot and to monitor their dissociation. BMC Biotechnol 10:53

12. Biricolti S, Bogani P, Cerboneschi M, Gori M (2016) Inverse PCR and quantitative PCR as alternative methods to southern blotting analysis to assess transgene copy number and characterize the integration site in transgenic woody plants. Biochem Genet 54:291–305. https://doi.org/10.1007/s10528-016-9719-z

13. Dalla Costa L, Bozzoli M, Pompili V et al (2019) Development of a Taqman real-time PCR method to quantify nptII in apple lines obtained with 'established' or 'new breeding' techniques of genetic modification. Eur Food Res Technol. https://doi.org/10.1007/s00217-018-3187-0

14. Xu X, Peng C, Wang X et al (2016) Comparison of droplet digital PCR with quantitative real-time PCR for determination of zygosity in transgenic maize. Transgenic Res 25:855–864. https://doi.org/10.1007/s11248-016-9982-0

15. Bogožalec Košir A, Demšar T, Štebih D et al (2019) Digital PCR as an effective tool for GMO quantification in complex matrices. Food Chem 294:73–78. https://doi.org/10.1016/j.foodchem.2019.05.029

16. Morcia C, Ghizzoni R, Delogu C et al (2020) Digital PCR: what relevance to plant studies? Biology (Basel) 9:433. https://doi.org/10.3390/biology9120433

17. Sun Y, Joyce PA (2017) Application of droplet digital PCR to determine copy number of endogenous genes and transgenes in sugarcane. Plant Cell Rep 36:1775–1783. https://doi.org/10.1007/s00299-017-2193-1

18. Jacchia S, Kagkli DM, Lievens A et al (2018) Identification of single target taxon-specific reference assays for the most commonly genetically transformed crops using digital droplet PCR. Food Control 93:191–200. https://doi.org/10.1016/j.foodcont.2018.06.013

19. Lepage É, Zampini É, Boyle B, Brisson N (2013) Time- and cost-efficient identification of T-DNA insertion sites through targeted genomic sequencing. PLoS ONE 8:e70912. https://doi.org/10.1371/journal.pone.0070912

20. Guttikonda SK, Marri P, Mammadov J et al (2016) Molecular characterization of transgenic events using next generation sequencing approach. PLoS ONE 11:e0149515. https://doi.org/10.1371/journal.pone.0149515

21. Siddique K, Wei J, Li R et al (2019) Identification of T-DNA insertion site and flanking sequence of a genetically modified maize event IE09S034 using next-generation sequencing technology. Mol Biotechnol 61:694–702. https://doi.org/10.1007/s12033-019-00196-0

22. Pessina S, Lenzi L, Perazzolli M et al (2016) Knockdown of *MLO* genes reduces susceptibility to powdery mildew in grapevine. Hortic Res 3:16016. https://doi.org/10.1038/hortres.2016.16

23. Pirrello C, Zeilmaker T, Bianco L et al (2021) Mining grapevine downy mildew susceptibility genes: a resource for genomics-based breeding and tailored gene editing. Biomolecules 11:181. https://doi.org/10.3390/biom11020181

24. Franks T, Gang He D, Thomas M (1998) Regeneration of transgenic *Vitis vinifera* L. Sultana plants: genotypic and phenotypic analysis. Mol Breed 4:321–333

25. Nitsch JP, Nitsch C (1969) Haploid plants from pollen grains. Science 163:85–87. https://doi.org/10.1126/science.163.3862.85

26. McCown BH, Lloyd G (1981) Woody plant medium (WPM)–a mineral nutrient formulation for microculture of woody plant-species. Hortic Sci 16:453

27. Gambino G, Chitarra W, Maghuly F et al (2009) Characterization of T-DNA insertions in transgenic grapevines obtained by *Agrobacterium*-mediated transformation. Mol Breed 24:305–320. https://doi.org/10.1007/s11032-009-9293-8

28. Jacobs TB, Zhang N, Patel D, Martin GB (2017) Generation of a collection of mutant tomato lines using pooled CRISPR libraries. Plant Physiol 174:2023–2037. https://doi.org/10.1104/pp.17.00489

29. Ramadan M, Alariqi M, Ma Y et al (2021) Efficient CRISPR/Cas9 mediated Pooled-sgRNAs assembly accelerates targeting multiple genes related to male sterility in cotton. Plant Methods 17:1–13. https://doi.org/10.1186/s13007-021-00712-x

30. Meng X, Yu H, Zhang Y et al (2017) Construction of a genome-wide mutant library in rice using CRISPR/Cas9. Mol Plant 10:1238–1241. https://doi.org/10.1016/j.molp.2017.06.006

31. Bai M, Yuan J, Kuang H et al (2020) Generation of a multiplex mutagenesis population via pooled CRISPR-Cas9 in soya bean. Plant Biotechnol J 18:721–731. https://doi.org/10.1111/pbi.13239

32. Liu HJ, Jian L, Xu J et al (2020) High-throughput CRISPR/Cas9 mutagenesis streamlines trait gene identification in maize. Plant Cell 32:1397–1413. https://doi.org/10.1105/tpc.19.00934

33. Dalla Costa L, Pinto-Sintra AL, Campa M et al (2014) Development of analytical tools for evaluating the effect of T-DNA chimeric integration on transgene expression in vegetatively propagated plants. Plant Cell Tissue Organ Cult 118:471–484. https://doi.org/10.1007/s11240-014-0499-z

34. Lechtenberg B, Schubert D, Forsbach A et al (2003) Neither inverted repeat T-DNA configurations nor arrangements of tandemly repeated transgenes are sufficient to trigger transgene silencing. Plant J 34:507–517. https://doi.org/10.1046/j.1365-313X.2003.01746.x

35. Tang W, Newton RJ, Weidner DA (2007) Genetic transformation and gene silencing mediated by multiple copies of a transgene in eastern white pine. J Exp Bot 58:545–554. https://doi.org/10.1093/jxb/erl228

36. Dong J, Mchughen A (1993) Transgenic flax plants from *Agrobacterium* mediated transformation: incidence of chimeric regenerants and inheritance of transgenic plants. Plant Sci 91:139–148

37. Ding L, Chen Y, Ma Y et al (2020) Effective reduction in chimeric mutants of poplar trees produced by CRISPR/Cas9 through a second round of shoot regeneration. Plant Biotechnol Rep 14:549–558. https://doi.org/10.1007/s11816-020-00629-2

38. De Buck S, Windels P, De Loose M, Depicker A (2004) Single-copy T-DNAs integrated at different positions in the *Arabidopsis* genome display uniform and comparable beta-glucuronidase accumulation levels. Cell Mol Life Sci 61:2632–2645. https://doi.org/10.1007/s00018-004-4284-8

39. Gerdes L, Iwobi A, Busch U, Pecoraro S (2016) Optimization of digital droplet polymerase chain reaction for quantification of genetically modified organisms. Biomol Detect Quantif 7:9–20. https://doi.org/10.1016/j.bdq.2015.12.003

40. Iwobi A, Gerdes L, Busch U, Pecoraro S (2016) Droplet digital PCR for routine analysis of genetically modified foods (GMO)—a comparison with real-time quantitative PCR. Food Control 69:205–213. https://doi.org/10.1016/j.foodcont.2016.04.048

41. Collier R, Dasgupta K, Xing YP et al (2017) Accurate measurement of transgene copy number in crop plants using droplet digital PCR. Plant J 90:1014–1025. https://doi.org/10.1111/tpj.13517

42. Paternò A, Verginelli D, Bonini P et al (2018) In-House validation and comparison of two wheat (*Triticum aestivum*) taxon-specific Real-time PCR methods for GMO quantification supported by Droplet digital PCR. Food Anal Methods 11:1281–1290. https://doi.org/10.1007/s12161-017-1097-6

43. Jouanin A, Tenorio-Berrio R, Schaart JG et al (2020) Optimisation of droplet digital PCR for determining copy number variation of α-gliadin genes in mutant and gene-edited polyploid bread wheat. J Cereal Sci. https://doi.org/10.1016/j.jcs.2019.102903

44. Narancio R, John U, Mason J et al (2021) Digital PCR (dPCR) and qPCR mediated determination of transgene copy number in the forage legume white clover (*Trifolium repens*). Mol Biol Rep 48:3069–3077. https://doi.org/10.1007/s11033-021-06354-5

45. Yuen G, Khan FJ, Gao S et al (2017) CRISPR/Cas9-mediated gene knockout is insensitive to target copy number but is dependent on guide RNA potency and Cas9/sgRNA threshold expression level. Nucleic Acids Res 45:12039–12053. https://doi.org/10.1093/nar/gkx843

46. Kohli A, Gahakwa D, Vain F et al (1999) Transgene expression in rice engineered through particle bombardment: molecular factors controlling stable expression and transgene silencing. Planta 208:88–97

47. Dhiman H, Campbell M, Melcher M et al (2020) Predicting favorable landing pads for targeted integrations in Chinese hamster ovary cell lines by learning stability characteristics from random transgene integrations. Comput Struct Biotechnol J 18:3632–3648. https://doi.org/10.1016/j.csbj.2020.11.008

48. Kohli A, González-Melendi P, Abranches R et al (2006) The quest to understand the basis and mechanisms that control expression of introduced transgenes in crop plants. Plant Signal Behav 1:185–195. https://doi.org/10.4161/psb.1.4.3195

49. Singer SD, Liu Z, Cox KD (2012) Minimizing the unpredictability of transgene expression in plants: the role of genetic insulators. Plant Cell Rep 31:13–25. https://doi.org/10.1007/s00299-011-1167-y

50. Betts SD, Basu S, Bolar J et al (2019) Uniform expression and relatively small position effects characterize sister transformants in maize and soybean. Front Plant Sci 10:1209. https://doi.org/10.3389/fpls.2019.01209

51. O'Brien SAO, Lee K, Fu H et al (2018) Single copy transgene integration in a transcriptionally active site for recombinant protein synthesis. Biotechnol J 13:e1800226. https://doi.org/10.1002/biot.201800226.Single

52. Gaidukov L, Wroblewska L, Teague B et al (2018) A multi-landing pad DNA integration platform for mammalian cell engineering. Nucleic Acids Res 46:4072–4086. https://doi.org/10.1093/nar/gky216

53. Dong OX, Yu S, Jain R et al (2020) Marker-free carotenoid-enriched rice generated through targeted gene insertion using CRISPR-Cas9. Nat Commun 11:1178. https://doi.org/10.1038/s41467-020-14981-y

54. Lemos BR, Kaplan AC, Bae JE et al (2018) CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. Proc Natl Acad Sci USA 115:E2010–E2047. https://doi.org/10.1073/pnas.1716855115