CRISPR/Cas9-mediated mutagenesis of phytoene desaturase in diploid and octoploid strawberry

Fiona Wilson*, Kate Harrison, Andrew D. Armitage, Andrew J. Simkin and Richard J. Harrison

NIAB EMR, New Road, East Malling, Kent, ME19 6BJ, UK

***** Corresponding author. *Email address:* fiona.wilson@emr.ac.uk

ARTICLE INFO

Key words: Fragaria, strawberry, diploid, polyploid, CRISPR/Cas9, PDS

ABSTRACT

Gene editing using CRISPR/Cas9 is a simple and powerful tool for elucidating genetic controls and for crop improvement. We demonstrate use of CRISPR/Cas methodology in diploid *Fragaria vesca* ssp. *vesca* 'Hawaii 4' and octoploid *F. x ananassa* 'Calypso' by targeting the visible endogenous marker gene *PDS* (phytoene desaturase). *Agrobacterium*-mediated transformation of leaf and petiole explants was used for efficient stable integration of constructs expressing plant codon-optimised Cas9 and single guide sequences under control of either the *Arabidopsis U6-26* consensus promoter and terminator or *Fragaria vesca U6III* regulatory sequences. More than 80% ('Hawaii 4') and 50% ('Calypso') putative transgenic shoot lines exhibited mutant phenotypes. Of mutant shoot lines selected for molecular analysis, approximately 55% ('Calypso') and 75% ('Hawaii 4') included albino regenerants with bi-allelic target sequence variants. Our results indicate the *PDS* gene is functionally diploid in 'Calypso' and clearly demonstrate that CRISPR/Cas9 can be used to edit single copy genes at high frequency within the genome of the diploid and the same target in octoploid strawberry.

1. Introduction

The cultivated strawberry, *Fragaria* x *ananassa*, is one of the most important fruit crops in the family Rosaceae. It is among the most widely grown and consumed fruit throughout the world, with global production reaching 9.1M tonnes in 2016, an increase over the previous decade of more than 5% annually (www.freshfruitportal.com). In the UK alone, in 2015 115,000 tonnes were produced, with a market value of £253 million [1].

Traditional breeding is lengthy and difficult as *F.* x *ananassa* is an octoploid species with a complex genome, and is intolerant to inbreeding. Important for genetic improvement is the potential to enhance elite cultivars using genetic modification, first demonstrated with marker genes using *Agrobacterium*-mediated transformation in the 1990s [2, 3]. Since then a number of developments in transformation methodology and its application in *Fragaria* have been reported (see [4, 5]. The diploid wild strawberry *F. vesca* shares a high degree of sequence identity with the cultivated strawberry and is a model for genetic improvement in the genus due to its small genome size (240 mb), short generation time, and ease of genetic transformation [6, 7]. The availability of the complete genome for *F. vesca* [8, 9] and partial genome sequence of the octoploid genome [10] facilitates the identification and manipulation of genes controlling important traits.

From its first use in model species such as *Arabidopsis* [11, 12] and tobacco [12, 13], the rapid development of CRISPR/Cas9 technology has enabled widespread use of precise gene editing in plants [14-16]. Reports of successful gene editing in diploid crops include rice [17, 18], sorghum [19], tomato [20], maize [21, 22], *Populus* [23], apple [24], grapes [25] and cassava [26, 27]. In polyploid crops, multiple gene homoeoalleles have been simultaneously edited using CRISPR/Cas9 in wheat [28], potato [29] and *Brassica napus* [30, 31]. Multiple gene targets have been successfully targeted in Arabidopsis [32], maize [33] and rice [34, 35].

Defects in *PDS* gene function result in a distinctive albino phenotype [36], and recent studies in *Populus* [23], apple [24], grapes [25] and cassava [26] have targeted the *PDS* gene to demonstrate successful genome editing in these crops. Nishitani *et al.* [24] compared target sites in different exons of the apple *PDS* gene and showed that the greatest mutation efficiency was achieved by targeting exon 7. As strawberry is also a member of the family *Rosaceae*, we chose to target the same region in the *PDS* gene of diploid and octoploid strawberry.

Our results show the mutation of PDS in diploid and octoploid strawberry results in a clear albino phenotype at a high frequency. The data provided will inform future work in *Fragaria* using CRISPR/Cas9, both as a research tool and for modifying molecular mechanisms controlling traits of agronomic importance, where mutant phenotypes cannot be identified visually. The importance of this technology for use in creating new varieties in support of global food security has been recently reviewed [37].

2. Materials and Methods

2.1 Construction of CRISPR/Cas9 vectors

The binary vector pFGC-pcoCas9 (Addgene plasmid # 52256) was modified by insertion of pGreen plant marker constructs [38]: the 'Nos-Kan' cassette (comprising the *nptII* gene under control of nopaline synthase regulatory sequences) was cloned into the *Pme*1 site near the right border and the '35S-GFP' cassette (comprising the modified *GFP* gene, [39], under control of *35S* regulatory sequences) was cloned into the *Sma*1 site in the multiple cloning site to give pCas9-K-GFP (Fig. 1).

CRISPR cassettes incorporating the sgRNA sequence and regulatory elements (Fig. 1) were synthesised by Integrated DNA Technologies (IDT) Ltd and cloned into pCR®-Blunt II-TOPO® (Thermo Fisher Scientific, UK). Cassettes were inserted into the *Eco*R1 site of pCas9-K-GFP and the orientation confirmed by sequencing. The *PDS* coding sequence for *Malus domestica* (GenBank: KU508828.1) was BLAST aligned against the genomes for *F. vesca* (FvAssembly 1.1) and *F.* x *ananassa* (www.rosaceae.org). The CRISPR target sequence selected (ACCTGATCGAGTAACTACTGAGG) is common to both *Fragaria* genomes and is located on the sense strand in exon 7 of FvAssembly 4 [9] (Fig. 2). It corresponds almost exactly to the target sequence (ex7-20bp, ACCTGATCGAGTAACTACAGAGG) used successfully by Nishitani *et al.* [24]. The PAM sequences are underlined. The target guide RNA (gRNA) sequence (PDS74) comprises the first 20 base pairs of the target sequence preceded by an additional guanine base to maintain the native *U6* promoter start of transcription. This is placed under transcriptional control of the consensus sequence of the *Arabidopsis U6-26* promoter [13] or the corresponding sequence of the *F. vesca U6III* promoter and terminator sequences (Table S1).

2.2 In vitro culture of Fragaria stock cultures

In vitro shoot cultures of *F. vesca* 'Hawaii 4' and *F. x annanassa* 'Calypso' were maintained in a growth room at 21 ^{o}C with a 16/8 h light/dark photoperiod provided by fluorescent lamps (colour reference 835, colour temperature 3500K). Crowns were subcultured at 4-5 week intervals, 5 per honey jar containing 50 ml medium. Basal culture medium was Murashige and Skoog (MS) macro and micro elements and vitamins [40], supplemented with sucrose (30 g l^{-1}), solidified with Daishin agar (Duchefa D1004, 9 g l^{-1}). The pH was adjusted to 5.8 before autoclaving. Shoots were alternately cultured on basal medium supplemented with 6-benzylaminopurine (BAP) 0.1 mg l^{-1} and indole-3-butyric acid (IBA) 0.1 mg l^{-1} or BAP 0.5 mg l^{-1}, as described by Schaart [41].

2.3 Growth and transformation of diploid and octoploid strawberry

Transformation of 'Calypso' was performed essentially as Schaart [41], with minor modifications. *Agrobacterium tumefaciens* strain EHA105 [42] harbouring the binary vector was grown overnight, and pelleted at 2,000 xg for 10 minutes. The inoculum was prepared by re-suspending the overnight culture in filter-sterilised liquid MS-based medium supplemented with glucose (30 g l^{-1}) and acetosyringone (100 µM), pH 5.2, to give OD 600 nm 0.2 – 0.3. Petioles from apical leaves of *F. vesca* 'Hawaii 4' or leaflets from young expanding leaves of *F.* x *ananassa* 'Calypso' were harvested from shoots four weeks after subculture and submerged in the inoculum for approximately 10-15 minutes during explant preparation: petioles were cut into 4-5 mm pieces and leaflets were separated from each leaf before scoring transversely into 2 mm strips, leaving one leaf edge intact. Explants were then blotted to remove excess inoculum before transfer to Shoot Regeneration Medium (SRM): MS medium supplemented with α-naphthaleneacetic acid (NAA) (0.2 mg l^{-1}) and thidiazuron (TDZ) (1 mg l^{-1}). The pH was adjusted to 5.8 and Agargel ™ (A3301, Sigma) (5 g l^{-1}) was added before autoclaving. After autoclaving filter-sterilised glucose (300 g l^{-1}) was added to a final concentration of 30 g l^{-1}. After four days' dark incubation at 21 ^{o}C, explants were washed in a solution of filter-sterilised ticarcillin disodium/clavulanate potassium (TCA, Duchefa T0190) (400 mg l^{-1}) in water, blotted and transferred to SRM selection medium containing filter-sterilised TCA (400 mg l^{-1}) and kanamycin sulphate monohydrate (Kanamycin, Duchefa K0126) (100 mg l^{-1}). 'Hawaii 4' petioles were initially cultured in Sarstedt Cell Culture Flasks (T-25) containing 15 ml liquid selection medium, for four weeks, 30 - 50 explants per flask, shaking at 60 rpm at low light intensity, before transfer to semi-solid selection medium. Leaves were cultured throughout on semi-solid selection medium: approximately 10 leaves (abaxial side in contact with the medium) or 20 petiole pieces were cultured per 90mm triple-vent Petri dish containing approximately 30 ml medium. Dishes were sealed with Parafilm® M and incubated in the growth room as for shoot cultures. Subculture intervals were 4-6 weeks. At the second culture passage, half of the explants were transferred to SRM supplemented with gibberellic acid (GA_3) (1 mg l^{-1}) before autoclaving. Petioles were divided to separate calli developing at both petiole ends. Leaves were divided as expansion occurred to ensure contact with the selection medium, to facilitate subculturing and to separate regenerating calli. Calli with shoots were harvested over a 7-month period, and transferred to rooting medium: MS medium 2.2 g l^{-1}, supplemented with sucrose 20 g l^{-1}, BAP 0.1 mg l^{-1}, IBA 0.1 mg l^{-1}, solidified with 9 g l^{-1} Daishin agar, pH adjusted to 5.8 before autoclaving. Filter-sterilised kanamycin (100 mg l^{-1}) and TCA (400 mg l^{-1}) were added after autoclaving. For shoots with an obvious albino

phenotype, kanamycin was omitted from the medium. Albino and variegated shoots were subcultured onto rooting medium with TCA (400 mg l^{-1}).

2.4 Analysis of transgenic lines

DNA was extracted from leaf material using the DNeasy Plant Kit (Qiagen, Manchester, UK). Primers used in PCR and sequencing reactions (Table S2) were synthesised by IDT Ltd. PCR screening of putative transgenic lines for Nos-Kan cassette and sgRNA insertions was performed using PCRBIO Taq Mix Red (PCR Biosystems) and reaction conditions as detailed in Table S3. PCR amplicons spanning the target region or Left Border T-DNA-genomic DNA junction sites were sequenced by Illumina MiSeq: primers incorporated target-specific sequences at the 3' ends, and Illumina adapter overhang sequences (P5 forward overhang, P7 reverse overhang) at the 5' ends. A 401 bp genomic region was amplified using primer pair P5-300 F and P7-700F, binding 186 ->161 bp upstream and 195 -> 170 downstream of the target region, respectively. Amplicons were prepared for sequencing following Illumina guide 16S Metagenomic Sequencing Library Preparation (Part # 15044223 Rev. B). Template was amplified using Q5® Hot Start High-Fidelity 2X Master Mix (M0494, New England Biolabs Ltd.) and purified using Mag-Bind® TotalPure NGS (M1378, Omega Bio-tek Inc.). Dual indices and Illumina sequencing adapters were attached using Nextera XT Index Kit (15055290, Illumina) and 2x KAPA HiFi HotStart ReadyMix (KK2605, Kapa Biosystems). PhiX control V3 (15017660, Illumina) and the amplicon library were diluted to 4 pM concentration and the sample was spiked with 30% PhiX before loading onto a MiSeq Reagent Nano Kit v2 500 cartridge (MS-103-1003, Illumina).

Thermal Asymmetric Interlaced (TAIL) PCR was used to amplify T-DNA-genomic DNA junctions using PCRBIO Taq Mix Red (PCR Biosystems Ltd): first round PCR was performed using arbitrary degenerate primer AD3 [43] and T-DNA-specific primer TAIL R1 (binding 257 -> 228 downstream of the Left Border). Nested primer TAIL R2 (binding 209 -> 188 downstream of the left border) and AD3 were used for the second round PCR. PCR reactions were performed using a Verity 96-Well Thermal Cycler (#4375786, maximum block ramp of 3.9 ºC/Sec and a maximum sample ramp of 3.35 ºC/Sec). The PCR cycle was essentially as Liu et al., [44], with minor modifications (Table S3). Amplicons were purified using Mag-Bind® TotalPure NGS and Sanger-sequenced by Eurofins Genetic Services Ltd. The sequencing primer (TAIL SEQ) binds 177 -> 146 downstream of the T-DNA left border.

2.5 Sequencing analysis and software

Illumina sequencing reads were trimmed to remove low quality data and Illumina adapters with fastq-mcf (http://code.google.com/p/ea-utils) before being aligned to the reference *F. vesca* 'Hawaii 4' genome v4.0.a1 [9], using BWA v0.7.15 [45]. Variants in respect to the reference genome were predicted and quantified from aligned reads using the CrispRVariant package v1.9.2 [46]. Nucleotide and protein alignments were performed using Geneious version 10.0.2 (http://www.geneious.com) [47]. Schematic maps were prepared using IBS software [48].

3. Results and Discussion

3.1 Transformation and regeneration of mutant lines

We placed sgRNAs under transcriptional control of the *Arabidopsis U6-26* promoter (p*AtU6-26* vector) [13] or the corresponding sequence of the *F. vesca U6III* promoter (p*FvU6III* vector).

Following transformation, Calli, which produced shoots, were maintained on selective regeneration medium for up to seven months and shoots were successively harvested as they regenerated. Calli were regenerated on media with or without GA_3 to encourage rooting. Mutant shoots were efficiently obtained on medium without GA_3 and there is no clear indication of the effect of including GA_3 in the regeneration medium (Table 1). Up to 90% of 'Hawaii 4' and 'Calypso' shoot lines (shoots from isolated calli) formed roots on kanamycin or exhibited clear mutant phenotypes (Table 1 and Fig. 3 , 4).

Approximately 62% ('Hawaii 4' Fig 3) and 45% ('Calypso' Fig. 4) putative transgenic shoot lines exhibited mutant phenotypes (Table 1). Where shoot lines comprised a range of phenotypes, albino or variegated phenotypes were apparent in approximately 10% of tissues harvested. The proportions of 'Hawaii 4' and 'Calypso' mutant phenotypes derived from p*AtU6-26* are 25/47 lines (53%) and 19/38 lines (50%), respectively, and corresponding data for p*FvU6III* are 67/76 lines (88%) and 6/16 lines (38%). Native promoters are known to enhance mutation efficiency: for example, in soybean in three target genes, mutation efficiencies were increased by the use of the soybean U6-10 promoter compared to the Arabidopsis U6-26 promoter [49]. Relative mutation efficiencies in this experiment suggest that the FvU6III promoter may be more effective than the AtU6-26 consensus promoter in 'Hawaii 4' but not in 'Calypso', and that alternative U6 promoters native to the octoploid genome may enhance sgRNA expression in 'Calypso' and other commercial cultivars.

3.2 Analysis of transgenic lines

3.2.1 Types and frequency of target site mutations

Some 401 bp amplicons were generated for 96 samples, including wild-type 'Hawaii 4' and 'Calypso' (Fig. S1). Illumina MiSeq results were obtained for both wild types, 18 'Hawaii 4' and 7 'Calypso' transgenic shoot lines. Transgenic shoot lines exhibit a variety of mutant sequence variants at and around the target site in a highly conserved region of the *PDS* gene (Figs. 3, 4, S2, S3), resulting in phenotypes typical of defective *PDS* gene function, including a high proportion of completely albino phenotypes. Approximately 60-80% shoot lines regenerated albino shoots with 100% target site sequence variant reads.

The most common sequence mutations in ‘Hawaii 4’ (H4) and ‘Calypso’ are three variant types with 1, 2 or 3 base deletions at (or spanning) the cut site in exon 7 of the *PDS* gene (Table 2, Fig. 5): all are present in the data for the majority of ‘Hawaii 4’ lines and approximately half of ‘Calypso’ lines (Fig. 6A). Insertions and substitutions are uncommon and seen only in ‘Hawaii 4’: 3 of 18 lines have a 1-base insertion and one line has a 2-base insertion; and 2 lines have single A/C substitutions at or near the cut site: line 32 has both a single base substitution and a 1-base insertion; and line 46 has just a single base substitution. Large deletions (10 bases or more) are also uncommon, present in 4 of ‘Hawaii 4’ lines. Protein translation of in-frame mutations (deletions of 3, or multiples of 3, bases) shows loss of one or more amino acids around the cut site. Frameshift mutations (and mutation -44:45D) result in amino acid loss and generation of stop codons within exon 7, resulting in truncated PDS protein.

Table 1- Numbers and percentages of lines that regenerated in *Fragaria vesca* and *Fragaria x ananassa* depending on construct promoters and GA treatment

Promoter	Regeneration medium +/- GA_3	Number of lines harvested throughout the experiment				Total Lines harvested	Number of lines assessed for transformed phenotype	Putative transformed lines	% putative transformed lines	Lines with mutant phenotype	% putative transformants with mutant phenotype	Cultivar
		2 months	4 months	5 months	7 months							
At	+ GA	18	18	22	7	65	58	42	72	24	57	Hawaii-4
At	no GA	1	4	1	0	6	6	5	83	1	20	
Fv	+ GA	12	1	39	4	56	55	50	91	46	92	
Fv	no GA	15	8	15	0	38	37	26	70	21	81	
Average									**79**		**62**	
At	+ GA	0	11	10	21	42	30	22	73	8	36	Calypso
At	no GA	0	9	7	6	22	19	16	84	11	69	
Fv	+ GA	2	4	1	6	13	9	8	89	4	50	
Fv	no GA	2	3	4	3	12	10	8	80	2	25	
Average									**82**		**45**	

Table 2- Classification of mutants within regenerated lines (presence and absence)

Sequence variant	Hawaii 4	Calypso	Non-synonymous amino acid substitution	Frameshift & protein truncation	Amino acid deletion	Putative functional protein
-1:1D	17	4	-	+	+	-
-2:1D	2	-	-	+	+	-
-2:2D	10	3	-	+	+	-
-1:3D	11	3	-	-	+	-
-4:4D	-	1	-	+	+	-
-1:5D	1	3	-	+	+	-
-2:6D	1	-	-	-	+	-
-3:7D	-	1	-	+	+	-
-5:7D	-	1	-	+	+	-
-2:9D	1	-	-	-	+	-
-5:9D	1	1	-	-	+	-
-6:9D	1	-	-	-	+	-
-10:9D	1	-	-	-	+	-
-11:15D	1	-	-	-	+	-
-1:22D	1	-	-	+	+	-
-25:271D	1	-	-	-	+	-
-44:45D	1	-	-	+	+	-
1:3D	1	-	-	-	+	-
1:5D	1	-	-	+	+	-
1:1I	3	-	+	+	-	-
SNV:-3A & 1:2	1	-	+	+	-	-
SNV:1A	1	-	+	-	-	+

3.2.2 Types of allelic variant

Samples are either homozygous for one sequence variant type, heterozygous (having two sequence variant types), chimaeric (3 or more sequence variant types), or have a mixture of wild-type sequence and 1, 2 or more than 2 sequence variants: mono-WT, di-WT or multi-WT (Table S4, Fig. 6B). All four homozygous variants are of the most common variant type (-1:1D); and 34 of 35 heterozygous allelic variants are combinations of or include the most common 1-, 2- or 3-base sequence variants (-1:1D, -2:2D, -1:2D). Most homozygous and heterozygous allelic variants are of the most common sequence variant types and are likely to have arisen from simultaneous mutation of both alleles prior to cell division. Chimaeric allelic variants indicate that mutation occurred in one allele

either before cell division or at a later stage, followed by additional mutation events. Sequence data for multiple samples from single shoot lines share variant sequence types, and combinations of sequence types (Table S4, Fig. 3, 4, S2, S3). The timing and sequence of mutations will determine the mixture of the allelic variants within a shoot and shoot line.

Four of 7 'Calypso' shoot lines (lines 1, 5, 7, 100) and 14 of 18 'Hawaii 4' shoot lines (all lines except 9, 10, 32, 36) produced shoots which are biallelic (homozygous, heterozygous or chimaeric) mutants.

3.2.3 The evidence suggests that multiple shoots from individual callus lines will often be derived from the same transgenic event

For selected samples, sequences for left border integration sites were generated by TAIL PCR (Fig. S4). Putative plant chromosomal sequence (i.e. not matching vector sequence) at the integration site, often contiguous with partial or complete T-DNA left border sequence, was recovered from approximately 40% of shoot lines screened (7 of 17 'Hawaii 4' and 4 of 9 'Calypso' lines, Fig. 7A, Table S5). The T-DNA integration sequences are different for each shoot line (Table S4). For approximately 50% of shoot lines sequenced (9 'Hawaii 4' and 4 'Calypso' shoot lines) the sequence obtained beyond the T-DNA left border matches only binary vector sequence (Fig. 7B), indicating chromosomal integration of vector 'backbone' DNA, which was shown by De Buck *et al.* [49] to be a common occurrence in transgenic *Arabidopsis* and tobacco plants. Shoots of 'Hawaii 4' line 10 have a complex integration sequence comprising T-DNA, vector backbone sequence and sequence not matching vector DNA (Fig. S5).

The sequence data for multiple shoots from single lines (4 of 9 'Calypso' lines and 13 of 17 'Hawaii 4' lines) have either high homology to each other where putative genome integration sequences were obtained, or are all homologous to vector 'backbone' sequence (Fig. 7). The only line where shoots appear not to share a transgenic origin, is 'Hawaii 4' line 79: 79 C and 79 E TAIL sequences do not overlap but both align to the vector backbone (Fig. S5); 79E has a more complex pattern of mutation and does not align to vector sequence.

The data from both target site sequence analysis and TAIL PCR sequence analysis suggest that multiple shoots regenerated from individual callus lines are likely to share a transgenic integration origin.

3.2.4 Observed phenotype is frequently consistent with mutation sequence data obtained

Phenotypic observations of leaf samples of multiple shoots were compared to the mutant sequence data generated for each sample (7 of 10 'Calypso' and all 18 'Hawaii 4' transgenic shoot lines (Figs. 3, 4, S2, S3, Table S4). All shoot lines comprised a mixture of green, and/or variegated, pale green, white or albino shoots, except 'Hawaii 4' line 89, for which only albino callus and shoots were regenerated. A very small fraction of variant reads were identified in sequence data from wild-type samples: 0.6% and 0.02 % of total reads in the data for 'Hawaii 4' and 'Calypso' wild types, respectively. These reads were considered background 'noise' as a result of index-switching between Illumina barcodes, an event that occurs at low rates during Illumina sequencing of pooled samples on a single flow cell [50].

In most cases the observed phenotype reflects the sequence data obtained: sequence data for the majority of white samples in both cultivars (23 of 31 samples and 11 of 18 'Hawaii 4' lines; 10 of 14 samples and 4 of 7 'Calypso' lines) lack or have a negligible fraction of wild-type sequence reads (< 0.05%-0.1% of total reads), including some shoot lines for which the fraction of wild-type sequence data for all samples sequenced is negligible ('Hawaii 4' lines 12, 14, 79 and 89 and 'Calypso' line 1). The exception is 'Calypso' shoot 1C, which is probably not transformed (PCR data show absence of a product in reactions using Nos-Kan cassette and sgRNA insert site T-DNA primers. Sequence data for 'Calypso' shoot line 7D is anomalous as it was observed to have a variegated phenotype. The presence of a low fraction (<5%) of wild-type reads in data from white phenotypes in some cases may be attributed to subjectivity or difficulty in assessing phenotype (for example where the leaf samples are very small), and may reflect the normal limits of visual assessment for non-functional *PDS* alleles: 'Hawaii 4': 10 D1 (very small albino shoot); 28 A1 (white axillary on variegated shoot); 46 F (small albino shoot); 90 B (albino axillary shoot, apical expanded leaf, no green tissue obvious) (Fig. S2), 'Calypso': 100 2 (small albino shoot, Fig. S3); 5 Ci (albino leaf from variegated shoot, Fig. 4).

Only one shoot heterozygous for target sequence mutation lacks the expected albino phenotype: sequence data for a green 'Hawaii 4' transgenic 'Hawaii 4' 3A show that 50% of reads are deletion variants, but indicate that one allele has an amino acid substitution (threonine to asparagine) and that this encodes a functional PDS protein. Published PDS protein sequences [51] for maize and rice also encode asparagine at this position (Fig. S6).

The ranges of wild-type reads commonly observed for pale green and variegated samples are 7-19% (7 samples from 7 lines of 'Hawaii 4') and 6-36% (15 samples from 10

lines of 'Hawaii 4', 2 samples from 2 lines of 'Calypso'), which is taken to be due to the proportion of alleles and cells expressing functional PDS protein.

3.2.5 Possible variation in PDS gene allele functionality in Fragaria and gene redundancy in the octoploid genome

In lines of both cultivars, sequence data for some observed phenotypes show unexpected proportions of wild-type and mutated sequences (Table S4, Figs. 4, S2, S3): all 'Hawaii 4' transgenic samples with a green phenotype are single-allele sequence variants, with approximately only 20% ('Hawaii 4' 46A), and 30% to 60% wild-type reads ('Hawaii 4' lines 10 A2, 28 C2, 28 C3, 32 A1 32 A2, 74 A and 90C); also, white phenotypes with higher than expected wild-type sequence fractions (15-30%) include 'Hawaii 4' 9B, 28 - A, B, C2, C3, 32 A1, A2, B2, 39 B and 90 (Fig. S2). Nishitani *et al.* [24] observed green transformants in apple (another diploid member of the *Rosaceae*) with partially mutated sequences of the *PDS* gene. It is possible that one of the *PDS* alleles is silenced or non-functional, and the fraction of wild-type sequence reflects the proportion of non-functional *PDS* alleles. 'Hawaii 4' samples 36 A3 and 28 D1, which are largely white with green leaf edge sectors and a high fraction of wild-type reads (27% and 14%) are possibly periclinal chimaeras: a mosaic of white and green cells in the L2 layer would be manifested as small green flecks at the leaf edge, and would explain the high wild-type reads in comparison with samples which show small green sectors in the central part of the lamina (e.g. 'Hawaii 4' 2 Ci, 32 B1 and 39 C1, with 0.7-3.5% wild-type reads).

The phenotypic and sequence data for 'Calypso' suggest that in the octoploid, cultivated strawberry there is only one functional *PDS* gene allele: sequence data for albino shoot 94 includes 31% wild-type reads and a green transgenic shoot (59) has only 11% wild-type reads. Albino or variegated samples with unexpectedly high wild-type reads and complex chimaeric genotypes may include mutations of both functional and non-functional alleles (e.g. 'Hawaii 4' 28 A1, 46 F, 90 B, and 'Calypso' 20 B,100 2- see Supplementary Figures S2 and S3). Evidence for gene silencing and loss of redundant sequences in polyploids is well-documented [52], and gene copy loss and deletion in polyploid *Fragaria* has been described by Rousseau-Gueutin et al.[53]. The fact that a single wild-type amplicon was found is suggestive, but not conclusive that the PDS1 locus is functionally diploid and that complex variants observed in amplicon profiles are due to chimeric regeneration, rather than multiple edits of homeologs.

3.2.6 CRISPR/Cas9 gene edited shoots can be efficiently produced in diploid and octoploid Fragaria

The primary aim of this work was to generate a set of detailed data using a visible marker that will be of relevance for gene editing in *Fragaria* of single- and multi-copy genes for which there is no visible phenotype. We have shown that *via Agrobacterium*-mediated transformation of leaf and petiole explants it is possible to regenerate shoots, which are biallelic mutants for a single copy gene, at a high frequency in both the diploid and octoploid genomes, giving confidence that effective targeting of multiple alleles and gene copies may be achieved. Effective targeting by CRISPR-Cas9 mutagenesis of genes and targeting of multiple homoeologs has been demonstrated in other polyploid crops, such as wheat [28], potato [29] and oilseed rape [30, 31]. Additionally, efficient targeting of multiple gene copies has been shown in rice [32, 34, 35]. In the octoploid strawberry an alternative *U6III* promoter may enhance expression of guide sequences and increase mutagenesis. Zhou et al. [54] have recently reported successful gene editing (of auxin biosynthesis and auxin response factor genes) in wild strawberry, showing similar rates of mutation to those reported here, and demonstrated improved recovery of mutants by use of dual sgRNAs.

In conclusion, when combined with high-throughput screening [55] and high-throughput sequencing, we expect efficient isolation of non-visible mutants of single and multi-copy genes in both diploid and octoploid *Fragaria* to be achievable.

Author contributions

R.H. and F.W. devised the research plan. F.W. planned the experiments, designed and constructed the CRISPR cassettes and vectors, and performed phenotypic and molecular analysis of transgenic lines. K.H. produced transgenic plants. A.D.A. performed bioinformatic analysis of sequence data and generated variant allele summary plots. F.W. wrote the manuscript with input from R.J.H., A.D.A. and A.J.S.

Acknowledgements

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) (BB/K017071/2 and BB/N006682/2). The authors thank Maria Sobczyk for assistance with obtaining the *PDS* gene sequences for *F. vesca* and *F.* x *ananassa*, and Helen J Bates for assistance with Illumina sequencing. The binary vector pFGC-pcoCas9 was a gift from Jen Sheen (Addgene plasmid # 52256). pGreen plant selection

and marker cassettes were provided by Mark Smedley, John Innes Centre, Norwich, Norfolk, United Kingdom.

Figure Legends

Fig. 1. Schematic map of CRISPR cassette and T-DNA of pCas9-K-GFP. (A) Single guide RNA inserted into the multiple cloning site (MCS) of pCas9-K-GFP.
(B) pCas9-K-GFP T-DNA showing the position of marker/selection cassettes and pcoCas9 coding sequence between the left border (LB) and right border (RB), and the binding regions of primers used for TAIL-PCR (T1=Tail 1; T2=Tail 2), sgRNA insertion site PCR (mcs_F; mcs_R) and Nos-Kan cassette PCR (Nos-Kan_F; Nos-Kan_R).

Fig. 2. Schematic map of *Fragaria vesca PDS* gene. LG4-gene12690 (8,035 bp, FvAssembly v4) showing exons (yellow arrows) and the location of the CRISPR target sequence PDS74 (position 3'-5' = 3,873-3,892).

Fig. 3. Sequence mutations and corresponding phenotypes of CRISPR/Cas9 transgenic lines of 'Hawaii 4'. Examples of sequence mutations and allelic variants with corresponding phenotypes in transgenic shoot lines of *Fragaria* 'Hawaii 4' (line numbers 39 and 89). Each sequence variant type is identified by the location of the mutation relative to the cut site:number bases deleted (D). Deletions result in amino acid loss (mutations -1:3D and 1:3D) and also frameshift (mutation -1:1D) and generation of stop codons within exon 7, resulting in truncated PDS protein. The cut site is shown by a black vertical line; the target site and PAM site (AGG) are within the box in the reference sequence. WT=wild type. Scale bars are 5 mm.

Fig. 4. Sequence mutations and corresponding phenotypes of CRISPR/Cas9 transgenic lines of 'Calypso'. Examples of sequence mutations and allelic variants with corresponding phenotypes in transgenic shoot lines of *Fragaria* 'Hawaii 4' (line numbers 39 and 89). Each sequence variant type is identified by the location of the mutation relative to the cut site:number bases deleted (D). Deletions result in amino acid loss (mutations -1:3D and -5:9D) and also frameshift (mutation -1:1D, -1:5D, -4:4D and -5:7D) and generation of stop codons within exon 7, resulting in truncated PDS protein. The cut site is shown by a black vertical line; the target site and PAM site (AGG) are within the box in the reference sequence. WT=wild type. Scale bars are 5 mm.

Fig. 5. Mutations of the *PDS* gene in CRISPR/Cas9 transgenic lines of 'Hawaii 4' and 'Calypso'. (A) Summary of CRIPSR/Cas9 sequence variants the *Fragaria PDS* gene, aligned to the reference (wild-type) sequence for exon 7 (120 base pairs). The wild-type (WT) nucleotide sequence is shown at the top of the chart and the positions of the 20 base pair target region 'PDS74' and PAM site (AGG) are indicated above the chart by dotted and solid lines, respectively. The 'cut site' is 3 base pairs upstream of the PAM site (AGG) and is shown as a vertical line. Each sequence variant type is identified by the location of the mutation relative to the cut site:number of bases deleted (D), inserted (I), or substituted (SNV:location relative to cut site and base substitution). Deleted bases generating gaps within exon 7 are shown by dashes. Base insertions and substitutions and consequent misaligned sequences relative to the wild type are highlighted. Sequences are shown truncated where mutations cause a frameshift generating stop codons within exon 7.
(B) Amino acid sequences for each variant type aligned to the reference sequence (WT) for exon 7. Variant amino acids relative to the wild type are highlighted. Stop codons are shown in black and marked with an asterisk. Gaps in amino acid sequences within exon 7 are shown by a dashed line and partial amino acids are highlighted.

Fig. 6. Sequence mutations and allelic variants in 'Hawaii 4' and 'Calypso' CRISPR/Cas9 transformants. (A) Sequence mutations: location of deletions relative to the cut site:number base pairs deleted, insertions (1 or 2 base pairs) and single base pair substitutions. (B) Types of allelic variant (central ring = shoot lines, outer ring = individual samples). Frequencies of sequence mutations and types of allelic variant are shown as percentages of total shoot lines and samples sequenced. Numbers of shoot lines/samples sequenced for 'Hawaii 4' and 'Calypso' are 18/69 and 7/19, respectively.

Fig. 7. Left border T-DNA integration. TAIL PCR sequence data for transgenic 'Hawaii 4' (H4) and 'Calypso' (Cal) lines. (A) Assembly to the binary vector T-DNA. (B) Assembly to the binary vector pCas9-K-GFP. Where homologous data was obtained from more than one sample in shoot lines this is indicated (number of homologous shoots/number of shoots sequenced in shoot line) and sequence for only one (representative) sample is shown. Bases not matching the reference sequence (putative chromosomal sequence) are highlighted. S/SEQ=primer used for sequencing; LB=T-DNA left border.

References

[1] Defra, Horticultural Statistics, (2017). https://www.gov.uk/government/collections/horticultural-statistics

[2] D.J. James, A.J. Passey, D.J. Barbara, Agrobacterium-mediated transformation of the cultivated strawberry (Fragaria × anannassa duch.) using disarmed binary vectors, Plant Sci, 69 (1990) 79-94.

[3] N.S. Nehra, R.N. Chibbar, K.K. Kartha, R.S. Datla, W. Crosby, C. Stushnoff, Genetic transformation of strawberry by *Agrobacterium tumefaciens* using a leaf disk regeneration system, Plant Cell Rep, 9 (1990) 293.

[4] A.M. Husaini, J.A. Mercado, J.G. Schaart, J.A. Teixeira da Silva, Review of factors affecting organogenesis, somatic embryogenesis and *Agrobacterium tumefaciens*-mediated transformation of strawberry, Global Science Books, Japan, 2011.

[5] A.M. Husaini, A.M. Rafiqi, Plant role of osmotin in strawberry improvement, Molecular Biology Reports, 30 (2012) 1055.

[6] I. El Mansouri, J.A. Mercado, V. Valpuesta, J.M. Lopez-Aranda, F. Pliego-Alfaro, M.A. Quesada, Shoot regeneration and Agrobacterium-mediated transformation of Fragaria vesca L, Plant Cell Rep, 15 (1996) 642-646.

[7] T. Oosumi, H.A. Gruszewski, L.A. Blischak, A.J. Baxter, P.A. Wadl, J.L. Shuman, R.E. Veilleux, V. Shulaev, High-efficiency transformation of the diploid strawberry (Fragaria vesca) for functional genomics, Planta, 223 (2006) 1219-1230.

[8] V. Shulaev, D.J. Sargent, R.N. Crowhurst, T.C. Mockler, O. Folkerts, A.L. Delcher, P. Jaiswal, K. Mockaitis, A. Liston, S.P. Mane, P. Burns, T.M. Davis, J.P. Slovin, N. Bassil, R.P. Hellens, C. Evans, T. Harkins, C. Kodira, B. Desany, O.R. Crasta, R.V. Jensen, A.C. Allan, T.P. Michael, J.C. Setubal, J.M. Celton, D.J. Rees, K.P. Williams, S.H. Holt, J.J. Ruiz Rojas, M. Chatterjee, B. Liu, H. Silva, L. Meisel, A. Adato, S.A. Filichkin, M. Troggio, R. Viola, T.L. Ashman, H. Wang, P. Dharmawardhana, J. Elser, R. Raja, H.D. Priest, D.W. Bryant, Jr., S.E. Fox, S.A. Givan, L.J. Wilhelm, S. Naithani, A. Christoffels, D.Y. Salama, J. Carter, E. Lopez Girona, A. Zdepski, W. Wang, R.A. Kerstetter, W. Schwab, S.S. Korban, J. Davik, A. Monfort, B. Denoyes-Rothan, P. Arus, R. Mittler, B. Flinn, A. Aharoni, J.L. Bennetzen, S.L. Salzberg, A.W. Dickerman, R. Velasco, M. Borodovsky, R.E. Veilleux, K.M. Folta, The genome of woodland strawberry (Fragaria vesca), Nat Genet, 43 (2011) 109-116.

[9] P.P. Edger, R. VanBuren, M. Colle, T.J. Poorten, C.M. Wai, C.E. Niederhuth, E.I. Alger, S. Ou, C.B. Acharya, J. Wang, P. Callow, M.R. McKain, J. Shi, C. Collier, Z. Xiong, J.P. Mower, J.P. Slovin, T. Hytonen, N. Jiang, K.L. Childs, S.J. Knapp, Single-molecule

sequencing and optical mapping yields an improved genome of woodland strawberry (Fragaria vesca) with chromosome-scale contiguity, Gigascience, 7 (2018) 1-7.

[10] S.N. Isobe, H. Hirakawa, S. Sato, F. Maeda, M. Ishikawa, T. Mori, Y. Yamamoto, K. Shirasawa, M. Kimura, M. Fukami, F. Hashizume, T. Tsuji, S. Sasamoto, M. Kato, K. Nanri, H. Tsuruoka, C. Minami, C. Takahashi, T. Wada, A. Ono, K. Kawashima, N. Nakazaki, Y. Kishida, M. Kohara, S. Nakayama, M. Yamada, T. Fujishiro, A. Watanabe, S. Tabata, Construction of an integrated high density simple sequence repeat linkage map in cultivated strawberry (*Fragaria x ananassa*) and its applicability, DNA Res, 20 (2013) 79-92.

[11] Z. Feng, B. Zhang, W. Ding, X. Liu, D.L. Yang, P. Wei, F. Cao, S. Zhu, F. Zhang, Y. Mao, J.K. Zhu, Efficient genome editing in plants using a CRISPR/Cas system, Cell Res, 23 (2013) 1229-1232.

[12] J.F. Li, J.E. Norville, J. Aach, M. McCormack, D. Zhang, J. Bush, G.M. Church, J. Sheen, Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9, Nature Biotechnology, 31 (2013) 688-691.

[13] V. Nekrasov, B. Staskawicz, D. Weigel, J.D. Jones, S. Kamoun, Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease, Nature Biotechnology, 31 (2013) 691-693.

[14] L. Arora, A. Narula, Gene Editing and Crop Improvement Using CRISPR-Cas9 System, Frontiers in Plant Science, 8 (2017) 1932.

[15] C.L. Soyars, B.A. Peterson, C.A. Burr, Z.L. Nimchuk, Cutting Edge Genetics: CRISPR/Cas9 Editing of Plant Genomes, Plant & Cell Physiology, 59 (2018) 1608-1620.

[16] C. Jung, G. Capistrano-Gossmann, J. Braatz, N. Sashidhar, S. Melzer, Recent developments in genome editing and applications in plant breeding, Plant Breeding, 137 (2017) 1-9.

[17] J. Miao, D. Guo, J. Zhang, Q. Huang, G. Qin, X. Zhang, J. Wan, H. Gu, L.J. Qu, Targeted mutagenesis in rice using CRISPR-Cas system, Cell Res, 23 (2013) 1233-1236.

[18] H. Zhang, J. Zhang, P. Wei, B. Zhang, F. Gou, Z. Feng, Y. Mao, L. Yang, H. Zhang, N. Xu, J.K. Zhu, The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation, Plant Biotechnology Journal, 12 (2014) 797-807.

[19] W. Jiang, H. Zhou, H. Bi, M. Fromm, B. Yang, D.P. Weeks, Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice, Nucleic Acids Research, 41 (2013) e188.

[20] C. Brooks, V. Nekrasov, Z.B. Lippman, J. Van Eck, Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system, Plant Physiology, 166 (2014) 1292-1297.

[21] Z. Liang, K. Zhang, K. Chen, C. Gao, Targeted mutagenesis in Zea mays using TALENs and the CRISPR/Cas system, J Genet Genomics, 41 (2014) 63-68.

[22] J. Zhu, N. Song, S. Sun, W. Yang, H. Zhao, W. Song, J. Lai, Efficiency and Inheritance of Targeted Mutagenesis in Maize Using CRISPR-Cas9, J Genet Genomics, 43 (2016) 25-36.

[23] D. Fan, T. Liu, C. Li, B. Jiao, S. Li, Y. Hou, K. Luo, Efficient CRISPR/Cas9-mediated targeted mutagenesis in Populus in the first generation, Scientific Reports, 5 (2015) 12217.

[24] C. Nishitani, N. Hirai, S. Komori, M. Wada, K. Okada, K. Osakabe, T. Yamamoto, Y. Osakabe, Efficient Genome Editing in Apple Using a CRISPR/Cas9 system, Scientific Reports, 6 (2016) 31481.

[25] I. Nakajima, Y. Ban, A. Azuma, N. Onoue, T. Moriguchi, T. Yamamoto, S. Toki, M. Endo, CRISPR/Cas9-mediated targeted mutagenesis in grape, PLoS One, 12 (2017) e0177966.

[26] J. Odipio, T. Alicai, I. Ingelbrecht, D.A. Nusinow, R. Bart, N.J. Taylor, Efficient CRISPR/Cas9 Genome Editing of Phytoene desaturase in Cassava, Frontiers in Plant Science, 8 (2017) 1780.

[27] S.E. Bull, D. Seung, C. Chanez, D. Mehta, J.E. Kuon, E. Truernit, A. Hochmuth, I. Zurkirchen, S.C. Zeeman, W. Gruissem, H. Vanderschuren, Accelerated ex situ breeding of GBSS- and PTST1-edited cassava for modified starch, Sci Adv, 4 (2018) eaat6086.

[28] Y. Zhang, Z. Liang, Y. Zong, Y. Wang, J. Liu, K. Chen, J.L. Qiu, C. Gao, Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA, Nat Commun, 7 (2016) 12617.

[29] M. Andersson, H. Turesson, A. Nicolia, A.S. Falt, M. Samuelsson, P. Hofvander, Efficient targeted multiallelic mutagenesis in tetraploid potato (Solanum tuberosum) by transient CRISPR-Cas9 expression in protoplasts, Plant Cell Rep, 36 (2017) 117-128.

[30] J. Braatz, H.J. Harloff, M. Mascher, N. Stein, A. Himmelbach, C. Jung, CRISPR-Cas9 targeted mutagenesis leads to simultaneous modification of different homoeologous gene copies in polyploid oilseed rape (*Brassica napus*), Plant Physiology, 174 (2017) 935-942.

[31] H. Yang, J.J. Wu, T. Tang, K.D. Liu, C. Dai, CRISPR/Cas9-mediated genome editing efficiently creates specific mutations at multiple loci using one sgRNA in Brassica napus, Scientific Reports, 7 (2017) 7489.

[32] X. Ma, Q. Zhang, Q. Zhu, W. Liu, Y. Chen, R. Qiu, B. Wang, Z. Yang, H. Li, Y. Lin, Y. Xie, R. Shen, S. Chen, Z. Wang, Y. Chen, J. Guo, L. Chen, X. Zhao, Z. Dong, Y.G. Liu, A Robust CRISPR/Cas9 System for Convenient, High-Efficiency Multiplex Genome Editing in Monocot and Dicot Plants, Molecular Plant, 8 (2015) 1274-1284.

[33] W. Qi, T. Zhu, Z. Tian, C. Li, W. Zhang, R. Song, High-efficiency CRISPR/Cas9 multiplex gene editing using the glycine tRNA-processing system-based strategy in maize, BMC Biotechnol, 16 (2016) 58.

[34] G. Liang, H. Zhang, D. Lou, D. Yu, Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing, Scientific Reports, 6 (2016) 21451.

[35] K. Xie, B. Minkenberg, Y. Yang, Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system, Proc Natl Acad Sci U S A, 112 (2015) 3570-3575.

[36] G. Qin, H. Gu, L. Ma, Y. Peng, X.W. Deng, Z. Chen, L.J. Qu, Disruption of phytoene desaturase gene results in albino and dwarf phenotypes in Arabidopsis by impairing chlorophyll, carotenoid, and gibberellin biosynthesis, Cell Res, 17 (2007) 471-482.

[37] X. Ma, M. Mau, T.F. Sharbel, Genome Editing for Global Food Security, Trends Biotechnol, (2017).

[38] R.P. Hellens, E.A. Edwards, N.R. Leyland, S. Bean, P.M. Mullineaux, pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation, Plant Molecular Biology, 42 (2000) 819-832.

[39] J. Haselhoff, K.R. Siemering, D.C. Prasher, S. Hodge, Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly, Proc Natl Acad Sci U S A, 94 (1997) 2122-2127.

[40] T. Murashige, F. Skoog, A revised medium for rapid growth and bioassays with tobacco tissue culture, Physiol Plantarum, 15 (1962) 473-497.

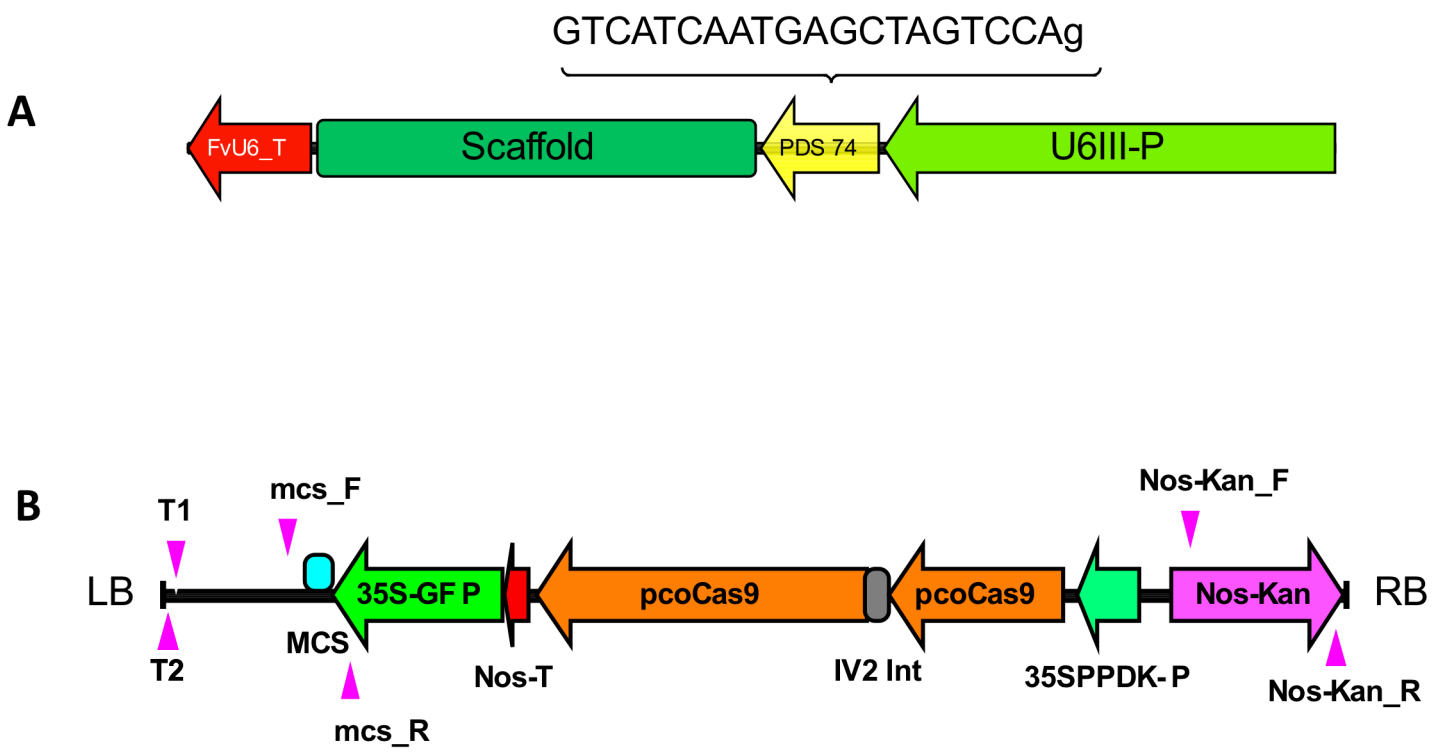
[41] J.G. Schaart, Agrobacterium-mediated transformation of strawberry, Bio Protoc, 4 (2014).

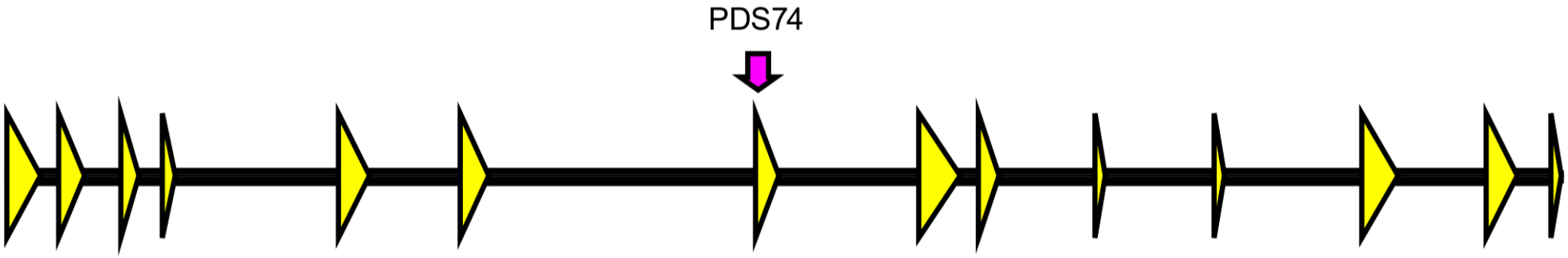
[42] E.E. Hood, S.B. Gelvin, L.S. Melchers, A. Hoekema, New Agrobacterium helper plasmids for gene transfer to plants, Transgenic Res, 2 (1993) 208-2018.

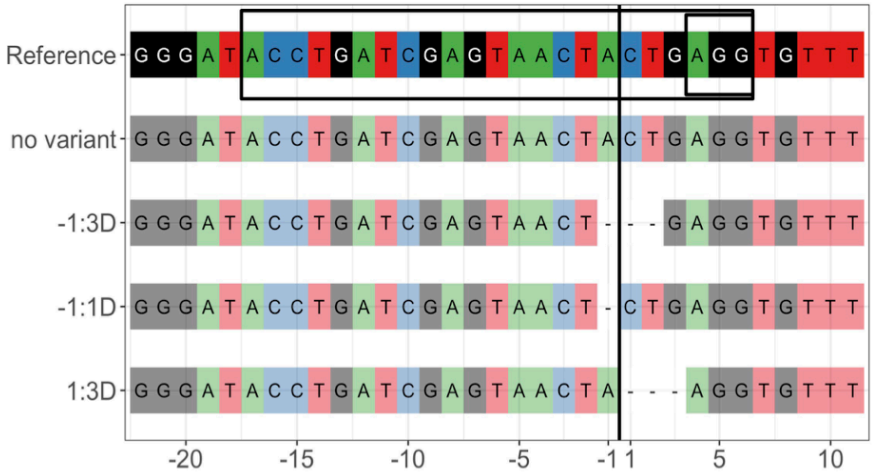
[43] Y.G. Liu, N. Mitsukawa, T. Oosumi, R.F. Whittier, Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR, The Plant Journal : For Cell and Molecular Biology, 8 (1995) 457-463.

[44] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics, 25 (2009) 1754-1760.

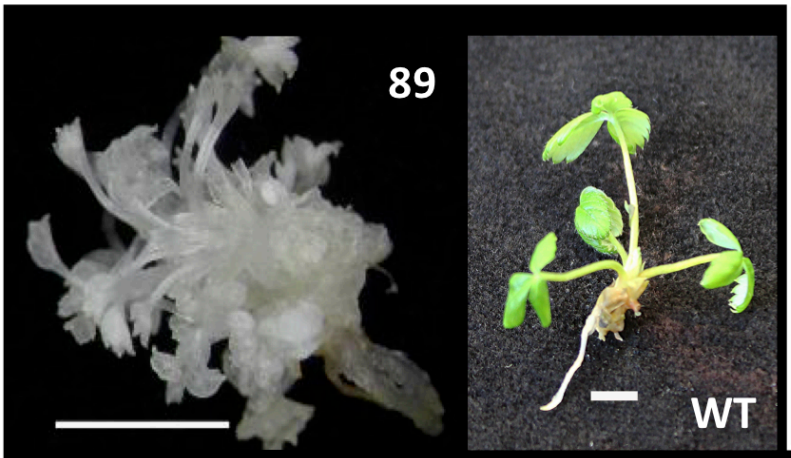
[45] H. Lindsay, A. Burger, B. Biyong, A. Felker, C. Hess, J. Zaugg, E. Chiavacci, C. Anders, M. Jinek, C. Mosimann, M.D. Robinson, CrispRVariants charts the mutation spectrum of genome engineering experiments, Nature Biotechnology, 34 (2016) 701-702.
[46] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond, Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, Bioinformatics, 28 (2012) 1647-1649.
[47] W. Liu, Y. Xie, J. Ma, X. Luo, P. Nie, Z. Zuo, U. Lahrmann, Q. Zhao, Y. Zheng, Y. Zhao, Y. Xue, J. Ren, IBS: an illustrator for the presentation and visualization of biological sequences, Bioinformatics, 31 (2015) 3359-3361.
[48] X. Sun, Z. Hu, R. Chen, Q. Jiang, G. Song, H. Zhang, Y. Xi, Targeted mutagenesis in soybean using the CRISPR-Cas9 system, Scientific Reports, 5 (2015) 10342.
[49] S. De Buck, C. De Wilde, M. Van Montagu, A. Depicker, T-DNA vector backbone sequences are frequently integrated into the genome of transgenic plants obtained by Agrobacterium-mediated transformation, Molecular Breeding, 6 (2000) 459-468.
[50] M. Costello, M. Fleharty, J. Abreu, Y. Farjoun, S. Ferriera, L. Holmes, B. Granger, L. Green, T. Howd, T. Mason, G. Vicente, M. Dasilva, W. Brodeur, T. DeSmet, S. Dodge, N.J. Lennon, S. Gabriel, Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms, BMC Genomics, 19 (2018) 332.
[51] C. The UniProt, UniProt: the universal protein knowledgebase, Nucleic acids research, 45 (2017) D158-D169.
[52] J.F. Wendel, D. Lisch, G. Hu, A.S. Mason, The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation, Curr Opin Genet Dev, 49 (2018) 1-7.
[53] M. Rousseau-Gueutin, A. Gaston, A. Ainouche, M.L. Ainouche, K. Olbricht, G. Staudt, L. Richard, B. Denoyes-Rothan, Tracking the evolutionary history of polyploidy in Fragaria L. (strawberry): new insights from phylogenetic analyses of low-copy nuclear genes, Mol Phylogenet Evol, 51 (2009) 515-530.
[54] J. Zhou, G. Wang, Z. Liu, Efficient genome editing of wild strawberry genes, vector development and validation, Plant Biotechnology Journal, 16 (2018) 1868-1877.
[55] L. Chen, W. Li, L. Katin-Grazzini, J. Ding, X. Gu, Y. Li, T. Gu, R. Wang, X. Lin, Z. Deng, R.J. McAvoy, F.G. Gmitter, Jr., Z. Deng, Y. Zhao, Y. Li, A method for the production and expedient screening of CRISPR/Cas9-mediated non-transgenic mutant plants, Hortic Res, 5 (2018) 13.


A
GTCATCAATGAGCTAGTCCAg
FvU6_T
Scaffold
PDS 74
U6III-P
B
T1
mcs_F
LB
T2
MCS
mcs_R
35S-GF P
Nos-T
pcoCas9
IV2 Int
pcoCas9
35SPPDK- P
Nos-Kan_F
Nos-Kan
Nos-Kan_R
RB

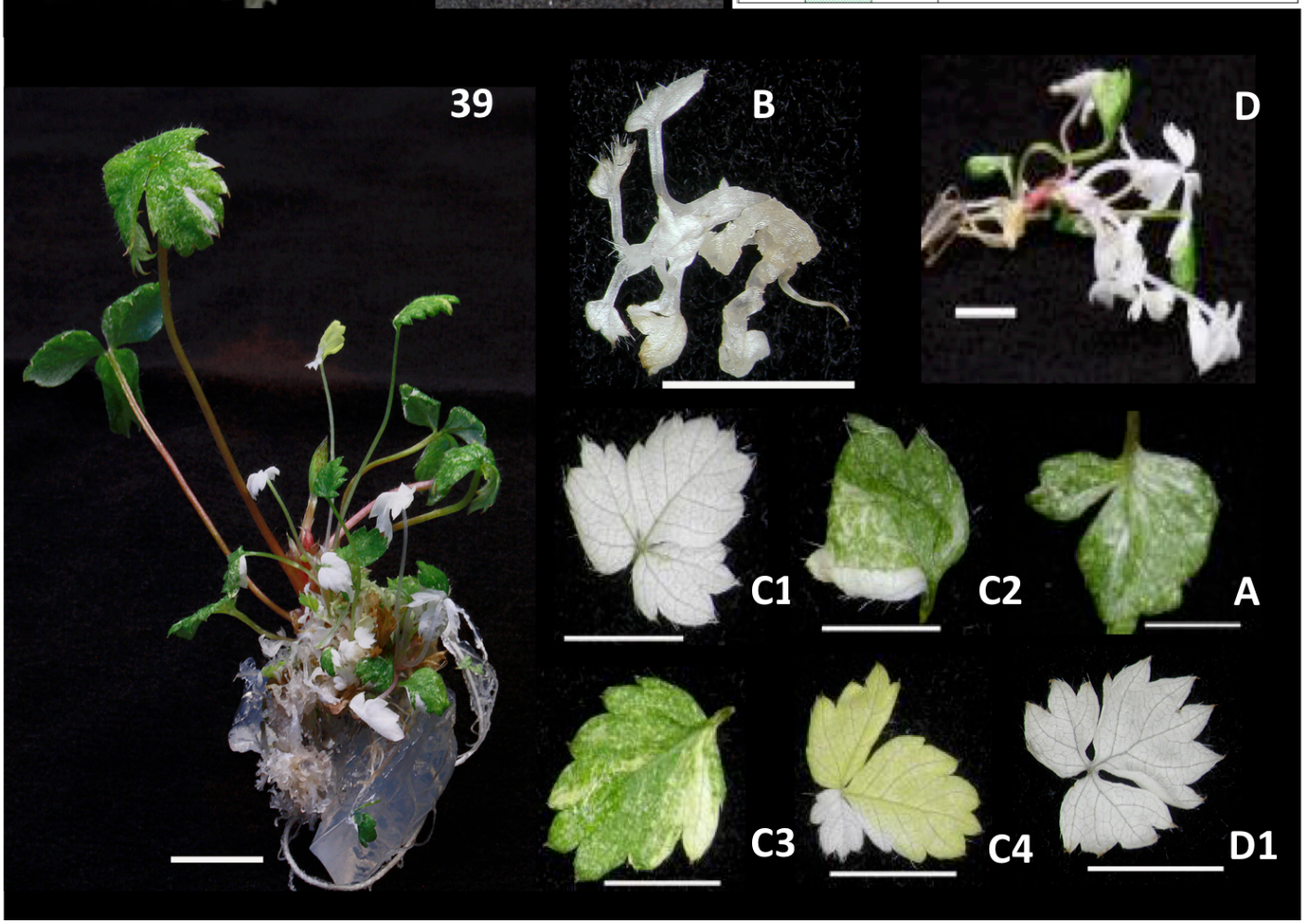

PDS74

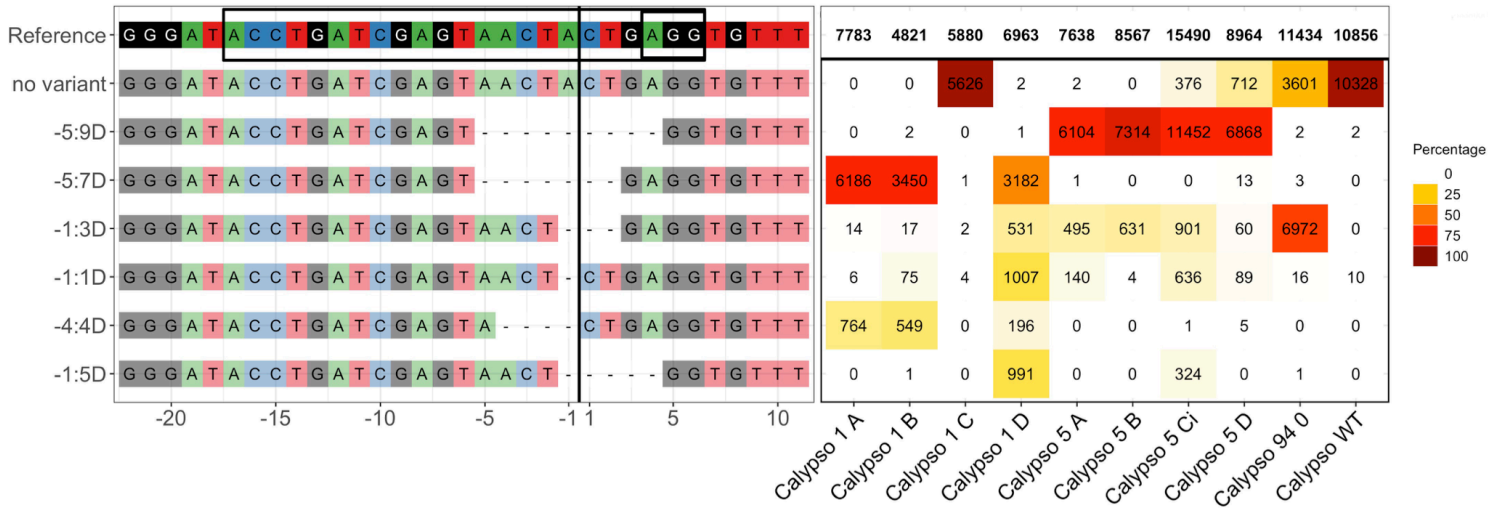


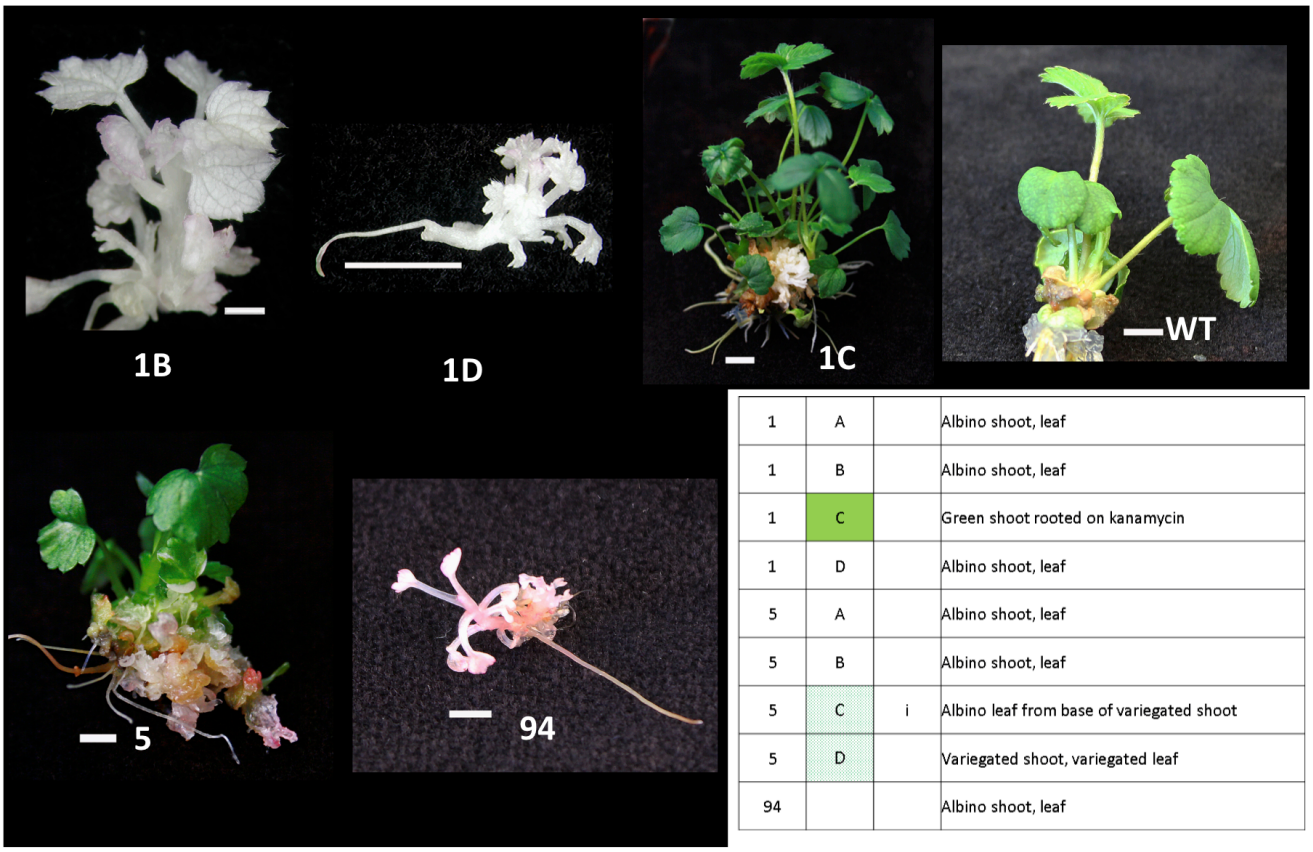
	13253	9669	5772	5337	8454	8095	9700	6369	7514
Hawaii 4 39 A	1295	2919	30	950	1623	1464	2	7	7068
Hawaii 4 39 B	8043	2670	3148	3105	4630	4547	9010	3399	25
Hawaii 4 39 C1	2932	2282	2509	976	1110	1466	595	2873	43
Hawaii 4 39 C2	17	1643	8	4	17	10	15	18	0
Hawaii 4 39 C3									
Hawaii 4 39 C4									
Hawaii 4 39 D1									
Hawaii 4 89 0									
Hawaii 4 WT									



89			Albino shoot from callus with only albino shoots, leaf
39	A		Variegated shoot, all leaves showing mottled variegation
39	B		White rooted shoot
39	C	1	Almost all white, green flecks
39	C	2	Variegated (green/white)
39	C	3	Variegated (green/yellow)
39	C	4	Variegated (yellow/white)
39	D	1	Albino (apical part of shoot)







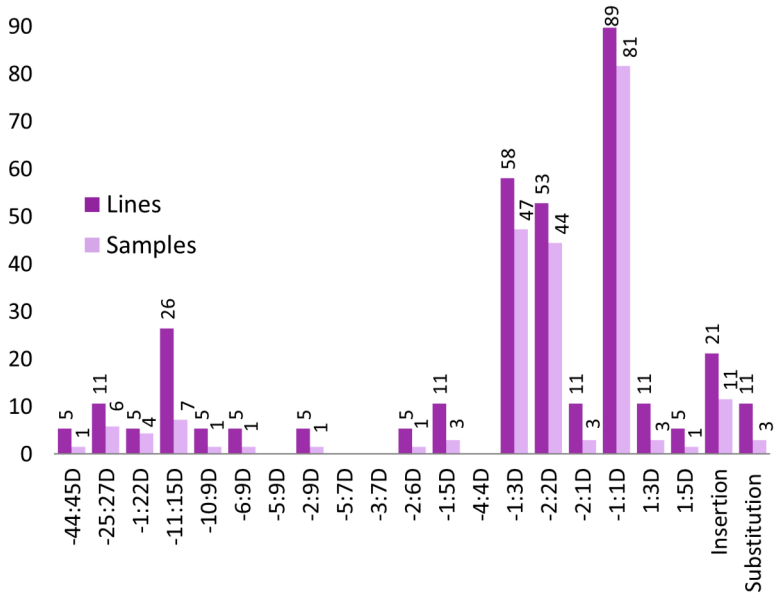
1	A		Albino shoot, leaf
1	B		Albino shoot, leaf
1	C		Green shoot rooted on kanamycin
1	D		Albino shoot, leaf
5	A		Albino shoot, leaf
5	B		Albino shoot, leaf
5	C	i	Albino leaf from base of variegated shoot
5	D		Variegated shoot, variegated leaf
94			Albino shoot, leaf

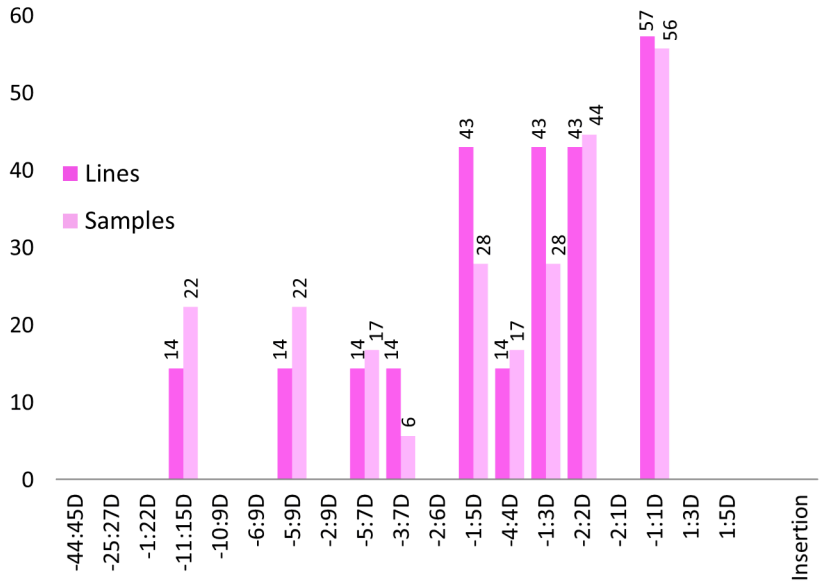
A

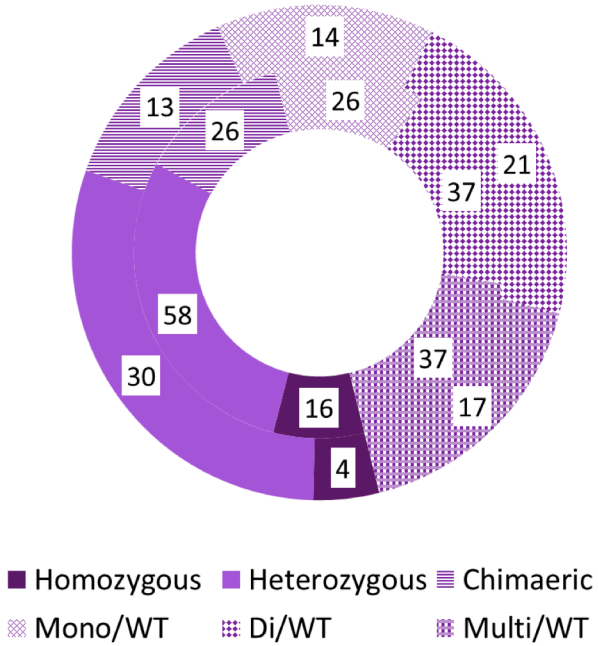
```
                1         10        20        30        40        50        60        70        80        90        100       110       120
0 WT            GGGATACCTGATCGAGTAACTACTGAGGTGTTTATTGCCATGTCAAAGGCCCTTAACTTTATTAATCCTGACGAGCTCTCCATGCAATGTATATTGATTGCTTTGAATCGATTTCTTCAG
-1:1D           GGGATACCTGATCGAGTAACT-CTGAGGTGTTTATTGCCATGTCAAAGGCCCTTAACTTTATTAATCCTGACGAGCTCTCCATGCAATGTATATTGA
-2:1D           GGGATACCTGATCGAGTAAC-ACTGAGGTGTTTATTGCCATGTCAAAGGCCCTTAACTTTATTAATCCTGACGAGCTCTCCATGCAATGTATATTGA
-2:2D           GGGATACCTGATCGAGTAAC--CTGA
-1:3D           GGGATACCTGATCGAGTAACT---GAGGTGTTTATTGCCATGTCAAAGGCCCTTAACTTTATTAATCCTGACGAGCTCTCCATGCAATGTATATTGATTGCTTTGAATCGATTTCTTCAG
-4:4D           GGGATACCTGATCGAGTA----CTGAGGTGTTTATTGCCATGTCAAAGGCCCTTAACTTTATTAATCCTGACGAGCTCTCCATGCAATGTATATTGA
-1:5D           GGGATACCTGATCGAGTAACT-----GGTGTTTATTGCCATGTCAAAGGCCCTTAA
-2:6D           GGGATACCTGATCGAGTAAC------GGTGTTTATTGCCATGTCAAAGGCCCTTAACTTTATTAATCCTGACGAGCTCTCCATGCAATGTATATTGATTGCTTTGAATCGATTTCTTCAG
-3:7D           GGGATACCTGATCGAGTAA-------GGTGTTTATTGCCATGTCAAAGGCCCTTAACTTTATTAATCCTGACGAGCTCTCCATGCAATGTATATTGA
-5:7D           GGGATACCTGATCGAGT-------GAGGTGTTTATTGCCATGTCAAAGGCCCTTAACTTTATTAATCCTGACGAGCTCTCCATGCAATGTATATTGA
-2:9D           GGGATACCTGATCGAGTAAC---------GTTTATTGCCATGTCAAAGGCCCTTAACTTTATTAATCCTGACGAGCTCTCCATGCAATGTATATTGATTGCTTTGAATCGATTTCTTCAG
-5:9D           GGGATACCTGATCGAGT---------GGTGTTTATTGCCATGTCAAAGGCCCTTAACTTTATTAATCCTGACGAGCTCTCCATGCAATGTATATTGATTGCTTTGAATCGATTTCTTCAG
-6:9D           GGGATACCTGATCGAG---------AGGTGTTTATTGCCATGTCAAAGGCCCTTAACTTTATTAATCCTGACGAGCTCTCCATGCAATGTATATTGATTGCTTTGAATCGATTTCTTCAG
-10:9D          GGGATACCTGAT---------ACTGAGGTGTTTATTGCCATGTCAAAGGCCCTTAACTTTATTAATCCTGACGAGCTCTCCATGCAATGTATATTGATTGCTTTGAATCGATTTCTTCAG
-11:15D         GGGATACCTGA---------------GGTGTTTATTGCCATGTCAAAGGCCCTTAACTTTATTAATCCTGACGAGCTCTCCATGCAATGTATATTGATTGCTTTGAATCGATTTCTTCAG
-1:22D          GGGATACCTGATCGAGTAACT----------------------CAAAGGCCCTTAACTTTATTAATCCTGACGAGCTCTCCATGCAATGTATATTGA
-25:27D                                 GAGGTGTTTATTGCCATGTCAAAGGCCCTTAACTTTATTAATCCTGACGAGCTCTCCATGCAATGTATATTGATTGCTTTGAATCGATTTCTTCAG
-44:45D                                TGA
1:3D            GGGATACCTGATCGAGTAACTA---AGGTGTTTATTGCCATGTCAAAGGCCCTTAACTTTATTAATCCTGACGAGCTCTCCATGCAATGTATATTGATTGCTTTGAATCGATTTCTTCAG
1:5D            GGGATACCTGATCGAGTAACT-----GGTGTTTATTGCCATGTCAAAGGCCCTTAA
1:1I            GGGATACCTGATCGAGTAACTAACTGA
SNV:-3A & 1:2I  GGGATACCTGATCGAGTAAATAACCTGAGGTGTTTATTGCCATGTCAAAGGCCCTTAACTTTATTAATCCTGACGAGCTCTCCATGCAATGTATATTGA
SNV:1A          GGGATACCTGATCGAGTAACTAATGAGGTGTTTATTGCCATGTCAAAGGCCCTTAACTTTATTAATCCTGACGAGCTCTCCATGCAATGTATATTGATTGCTTTGAATCGATTTCTTCAG
```

B

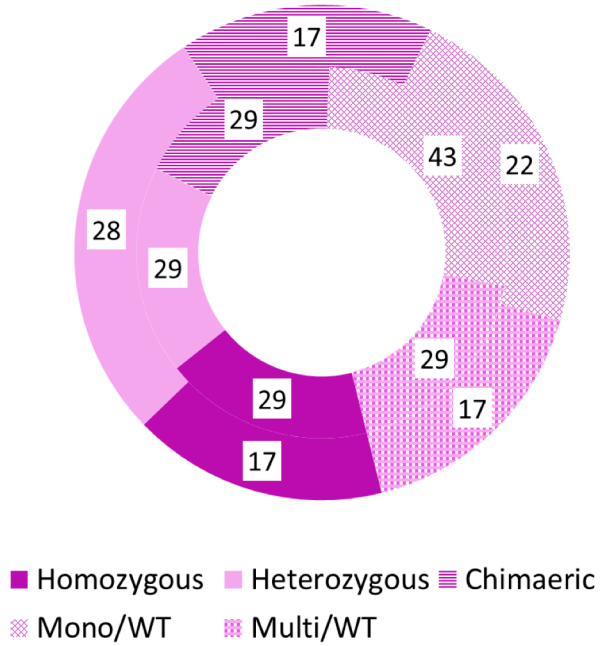
```
                1         10        20        30        40        50        60        70        80        90        100       110       120
0 WT            G I P D R V T T E V F I A M S K A L N F I N P D E L S M Q C I L I A L N R F L Q
-1:1D           G I P D R V T-L R C L L P C Q R P L T L L I L T S S P C N V Y *
-2:1D           G I P D R V T-L R C L L P C Q R P L T L L I L T S S P C N V Y *
-2:2D           G I P D R V T-- *
-1:3D           G I P D R V T-- E V F I A M S K A L N F I N P D E L S M Q C I L I A L N R F L Q
-4:4D           G I P D R V ---L R C L L P C Q R P L T L L I L T S S P C N V Y *
-1:5D           G I P D R V T-----G V Y C H V K G P *
-2:6D           G I P D R V T-----V F I A M S K A L N F I N P D E L S M Q C I L I A L N R F L Q
-3:7D           G I P D R V --------R C L L P C Q R P L T L L I L T S S P C N V Y *
-5:7D           G I P D R V-------- R C L L P C Q R P L T L L I L T S S P C N V Y *
-2:9D           G I P D R V T---------F I A M S K A L N F I N P D E L S M Q C I L I A L N R F L Q
-5:9D           G I P D R V-----------V F I A M S K A L N F I N P D E L S M Q C I L I A L N R F L Q
-6:9D           G I P D R ------------E V F I A M S K A L N F I N P D E L S M Q C I L I A L N R F L Q
-10:9D          G I P D ---------T E V F I A M S K A L N F I N P D E L S M Q C I L I A L N R F L Q
-11:15D         G I P E-----------V F I A M S K A L N F I N P D E L S M Q C I L I A L N R F L Q
-1:22D          G I P D R V T----------Q R P L T L L I L T S S P C N V Y *
-25:27D                                E V F I A M S K A L N F I N P D E L S M Q C I L I A L N R F L Q
-44:45D                                *
1:3D            G I P D R V T ---K V F I A M S K A L N F I N P D E L S M Q C I L I A L N R F L Q
1:5D            G I P D R V T-----G V Y C H V K G P *
1:1I            G I P D R V T N *
SNV:-3A & 1:2I  G I P D R V N N L R C L L P C Q R P L T L L I L T S S P C N V Y *
SNV:1A          G I P D R V T N E V F I A M S K A L N F I N P D E L S M Q C I L I A L N R F L Q
```



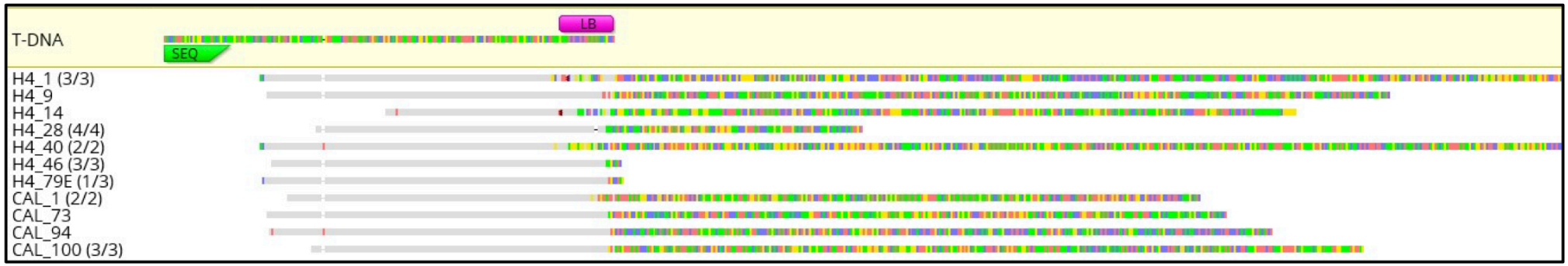




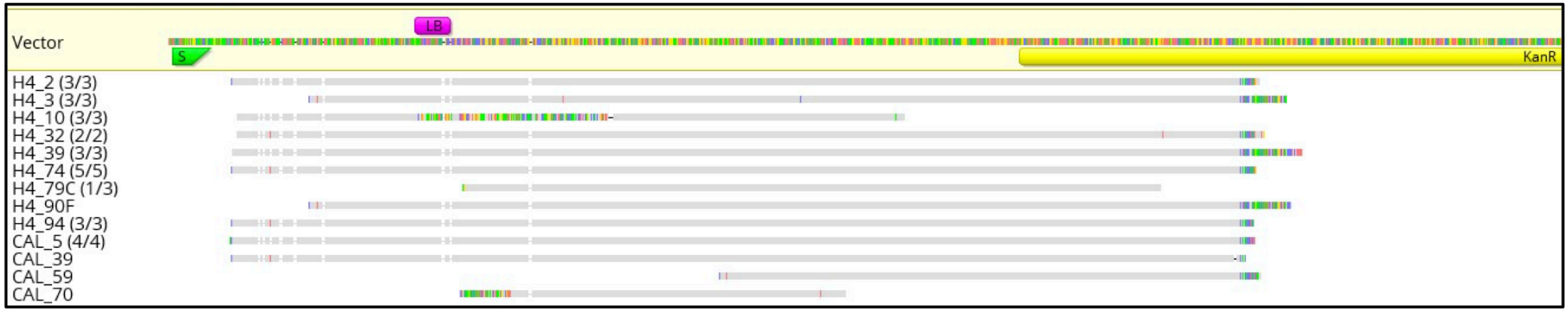
s in '
e pai
al rin
ercen



A


T-DNA
LB
SEQ
H4_1 (3/3)
H4_9
H4_14
H4_28 (4/4)
H4_40 (2/2)
H4_46 (3/3)
H4_79E (1/3)
CAL_1 (2/2)
CAL_73
CAL_94
CAL_100 (3/3)

B


Vector
LB
S
KanR
H4_2 (3/3)
H4_3 (3/3)
H4_10 (3/3)
H4_32 (2/2)
H4_39 (3/3)
H4_74 (5/5)
H4_79C (1/3)
H4_90F
H4_94 (3/3)
CAL_5 (4/4)
CAL_39
CAL_59
CAL_70