Theor Appl Genet (2010) 121:449–463
DOI 10.1007/s00122-010-1322-9

ORIGINAL PAPER

SNP discovery and genetic mapping of T-DNA insertional mutants in *Fragaria vesca* L.

J. J. Ruiz-Rojas · D. J. Sargent · V. Shulaev · A. W. Dickerman · J. Pattison · S. H. Holt · A. Ciordia · Richard E. Veilleux

Received: 20 November 2009 / Accepted: 5 March 2010 / Published online: 27 March 2010

Abstract As part of a program to develop forward and reverse genetics platforms in the diploid strawberry [*Fragaria vesca* L.; ($2n = 2x = 14$)] we have generated insertional mutant lines by T-DNA mutagenesis using pCAMBIA vectors. To characterize the T-DNA insertion sites of a population of 108 unique single copy mutants, we utilized thermal asymmetric interlaced PCR (hiTAIL-PCR) to amplify the flanking region surrounding either the left or right border of the T-DNA. Bioinformatics analysis of flanking sequences revealed little preference for insertion site with regard to G/C content; left borders tended to retain more of the plasmid backbone than right borders. Primers were developed from *F. vesca* flanking sequences to attempt to amplify products from both parents of the reference *F. vesca* 815 × *F. bucharica* 601 mapping population. Polymorphism occurred as: presence/absence of an amplification product for 16 primer pairs and different size products for 12 primer pairs, For 46 mutants, where polymorphism was not found by PCR, the amplification products were sequenced to reveal SNP polymorphism. A cleaved amplified polymorphic sequence/derived cleaved amplified polymorphism sequence (CAPS/dCAPS) strategy was then applied to find restriction endonuclease recognition sites in one of the parental lines to map the SNP position of 74 of the T-DNA insertion lines. BLAST search of flanking regions against GenBank revealed that 46 of 108 flanking sequences were close to presumed strawberry genes related to annotated genes from other plants.

Communicated by A. Schulman.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1322-9) contains supplementary material, which is available to authorized users.

J. J. Ruiz-Rojas · V. Shulaev · S. H. Holt · R. E. Veilleux (✉)
Department of Horticulture, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA
e-mail: potato@vt.edu

D. J. Sargent · A. Ciordia
East Malling Research (EMR), New Road, East Malling, Kent ME19 6BJ, UK

J. Pattison
Department of Horticultural Science, North Carolina State University, 600 Laureate Way, Kannapolis, NC 28081, USA

V. Shulaev · A. W. Dickerman
Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

Introduction

The application of molecular techniques for genetic improvement of most rosaceous crops is limited by large genome size and complexity (Shulaev et al. 2008). The alpine strawberry, *Fragaria vesca* L., however, has several advantages for genomic investigation. *F. vesca* is the most closely related diploid strawberry species ($2n = 2x = 14$) to the cultivated octoploid strawberry, *Fragaria* × *ananassa* Duch. ($2n = 8x = 56$) (Potter et al. 2000; Rousseau-Gueutin et al. 2008), has a small genome estimated at 206 Mbp (Folta and Davis 2006), a short generation time (about 4 months), is easy to propagate from seed or clonally, and selected accessions can be easily transformed by *Agrobacterium tumefaciens* (Smith and Townsend 1907; Conn 1942) in tissue culture. With the growing use of *F. vesca* as a model rosaceous plant, we are working to develop forward and reverse genetic platforms in this diploid strawberry by implementing T-DNA mutagenesis (Oosumi et al. 2006).

T-DNA mutagenesis or 'gene tagging' is a technique used for generating loss-of-function mutations in genes by mobile or introduced DNA with a known sequence (T-DNA). The insertion of foreign T-DNA into a gene both mutates the gene and serves as a vehicle for the isolation of the unknown genomic DNA flanking the insertion site. The foreign DNA not only disrupts the expression of the gene into which it has been inserted, but also acts as a marker for subsequent identification of the mutation. An advantage of using T-DNA as the insertional mutagen, as opposed to transposon tagging, is that T-DNA insertions will not transpose subsequent to integration within the genome and are therefore chemically and physically stable through multiple generations (Krysan et al. 1999). In insertional mutant populations of Arabidopsis and rice, T-DNA has shown a preference for integration into gene-rich chromosomal regions rather than centromeric regions that contain fewer genes (Li et al. 2006; Shuangyan et al. 2003).

Characterization of the genomic DNA flanking the T-DNA insertion sites (IS) can be done by TAIL-PCR, originally described by Liu and Whittier (1995). This technique employs nested, specific primers designed outward from the T-DNA used sequentially in combination with arbitrary degenerate primers for the unknown genomic DNA region flanking the insertion site (Liu and Whittier 1995). As a result, priming creates specific and nonspecific products that are thermally controlled by PCR cycles carried out with alternating high and low annealing temperatures, allowing both types of primers to function at optimum efficiency. TAIL-PCR has been widely used to recover flanking sequences of T-DNA and/or transposon insertions for functional genomics, cloning such insertion-tagged genes, recovering upstream regulatory sequences of tagged genes, and cloning genes from large vectors such as P1 phage, yeast and bacterial artificial chromosomes (Liu and Whittier 1995; Liu and Chen 2007).

DNA marker analysis of genomes including those of rosaceous plants along with all other families of economically important plant species has revolutionized genetic studies in the past decade. Cleaved amplified polymorphic sequences (CAPS) and derived cleaved amplified polymorphism sequences (dCAPS) are PCR-based markers in which a restriction site is present or introduced in only one of two amplified sequences. This restriction site allows the use of restriction enzymes to cut one of two alleles and screen a mapping population easily and more economically than several other methods (Konieczny and Ausubel 1993; Neff et al. 1998; Neff et al. 2002). The CAPS method has been used to convert SNPs into PCR-based markers in rice (Komori and Nitta 2005). Genetic maps have been developed for all major crops including tomato (Liu 2005), sweet cherry (Olmstead 2008), rice (Murai et al. 2003) and apple (Sargent et al. 2009). In the diploid strawberry, the most complete genetic linkage map has been constructed from an F_2 population derived from an interspecific cross between two diploid *Fragaria* species, *F. vesca* 815 × *F. bucharica* 601 Losinsk. [formerly *F. nubicola* (Hook.f.) Lindl. ex Lacaita] (Sargent et al. 2004). The linkage map consists of a total of 348 molecular markers, which includes 272 microsatellite, or simple sequence repeat (SSR) loci, 35 gene-specific and STS markers, 40 RFLPs and 1 SCAR spanning 568.8 cM over seven linkage groups, as well as three morphological traits. Due to the abundance of transferable markers employed in the construction of this map, it has been adopted as the standard reference map for the genus, and has been used to build maps of the cultivated octoploid strawberry *F.* × *ananassa* (Rousseau-Gueutin et al. 2008; Sargent et al. 2009).

Insertional mutagenesis using *Agrobacterium* is an imprecise process. In addition to the desirable perfect single insertions of T-DNA into a gene to knock out its function, Krysan et al. (1999) describe a range of undesirable alternative integration patterns that may occur. The randomness of insertion has been questioned as T-DNA may show preference for integration into particular nucleotide regions. Our objective was to examine the range of insertion sites of a collection of strawberry transformants and position them on the strawberry genetic map. In this endeavor we recovered the genomic insertion sites of T-DNA strawberry insertional mutants using hiTAIL-PCR and examined the flanking sequences to develop allele-specific primers adjacent to T-DNA insertions. Single nucleotide polymorphisms (SNP) in these flanking regions between the parents of the reference mapping population (*F. vesca 815* and *F. bucharica 601*) were converted into CAPS/dCAPS markers that could be placed on the strawberry genetic map. In addition, we initiated functional annotation of our T-DNA insertional mutant collection.

Materials and methods

Isolation of unknown flanking sequences by hiTAIL-PCR

Independent T-DNA insertional mutant lines of *F. vesca* PI 551572 (FV10 or Hawaii 4) were generated using the *Agrobacterium*-mediated transformation protocol described by Oosumi et al. (2006). Transgenic lines carrying either of two binary vectors, pCAMBIA-1304 (GUS/GFP as selectable marker) or pCAMBIA 1302 (GFP selectable marker), were used. Genomic DNA was extracted from T_0 plants that were grown in soilless medium in a growth chamber following the previously described methods (Oosumi et al. 2006). The number of T-DNA insertions in each plant was estimated by Southern blot analysis (Lofstrand Labs

Limited, Gaithersburg, MD, USA). High-efficiency thermal asymmetric interlaced PCR (hiTAIL-PCR) according to Liu and Chen (2007) was performed on a subset of plants expected to carry a single copy of the T-DNA according to Southern blot results (data not shown). Initially, we used the right border nested primers of Liu and Chen (2007) on all the mutants. When right border reactions failed, we used left border nested primers (kindly provided by YG Liu, personal communication, Supplemental Table 1).

Primary and secondary reaction products were separated through a 1.2% agarose gel containing 0.1% ethidium bromide at 150 V for 45 min. Products were visualized over UV light and the largest band (not less than 600 bp) for each amplification was excised and purified using a QiaQuick gel extraction kit (Qiagen, Valencia, CA, USA) according to the manufacturer's recommendations and resuspended in 30 µl autoclaved distilled water. Products were sequenced at the Tufts University Core Facility (Boston, MA, USA).

Allele specific primer design from flanking sequences

Sequences generated from the hiTAIL-PCR fragments were analyzed using Vector NTI-10 software (http://www.invitrogen.com). Each sequence was aligned against the pCAMBIA vector to identify and remove any T-DNA or plasmid backbone sequence. Primers were designed from the remaining *Fragaria* DNA sequence to have a melting temperature between 55 and 65°C, a length between 20 and 30 nucleotides, and a GC content above 50%, using the primer design function of Vector NTI-10 software. Sequences were denoted VT (Virginia Tech) followed by the mutant number for each transgenic line.

Discovery of SNPs using PCR CAPS/dCAPS based markers

PCR was performed on genomic DNA of the parents of the diploid *Fragaria* reference linkage map (*F. vesca* 815 and *F. bucharica 601*) (Sargent et al. 2004) using primers designed from each T-DNA insertion flanking region in a final volume of 25 µl following the touchdown protocol of Sargent et al. (2003) with an annealing temperature from 55 to 50°C. Amplicons generated from each parent were sequenced twice to ensure fidelity at the Tufts University Core Facility and confirm the sequence polymorphism between the parental genotypes. Sequences were aligned using Vector NTI-10 to detect SNPs. CAPS/dCAPS markers were developed following the procedures of Konieczny and Ausubel (1993) and Michaels and Amasino (1998). Briefly, each potential SNP was isolated with a maximum of 25 flanking nucleotides on each side of a parental DNA sequence. Sequences were analyzed with the software dCAPS finder 2.0 (Neff et al. 2002), (http://helix.wustl.edu/dcaps/dcaps.html) to identify restriction enzymes that cleaved at the SNP position of one of the two parental sequences. If no restriction endonuclease recognition site was found, the program permits design of a PCR primer that introduces a mismatch into one of the two sequences flanking the SNP to create a recognition site and therefore allow allelic discrimination.

Genetic linkage mapping of SNPs in the diploid *Fragaria* reference map (*F. vesca* 815 × *F. bucharica* 601)

Primer pairs were used to generate polymorphic amplicons in the diploid *Fragaria* reference mapping population *F. vesca* 815 × *F. bucharica* 601. Polymorphisms were scored on 74 individuals as either presence or absence of an amplicon in the parents, or a length polymorphism created by a SNP site. PCR was performed in a final volume of 12.5 µl following the touchdown protocol of Sargent et al. (2003) using an annealing temperature from 55 to 50°C. Where required, products were digested for 2 h with the appropriate restriction endonuclease listed in Table 1 in a final volume of 20 µl containing 1× reaction buffer, 1 U restriction endonuclease and 1× BSA where required. Products were separated in a 1.2% agarose gel at 120 V for up to 3.5 h and visualized over UV light. Where polymorphisms between PCR products were too small to score reliably through agarose gel electrophoresis, the forward primer was fluorescently labeled with either 6-FAM or HEX fluorescent dyes (VHBio, Newcastle, UK) and products were scored after capillary electrophoresis as described by Sargent et al. (2003).

Chi-squared tests of goodness-of-fit to an expected segregation ratio of 1:2:1 or 3:1 were performed for all segregating T-DNA insertional mutant line markers using JOINMAP 4.0 software (Van Ooijen and Voorrips 2001). Marker placement was determined on the *F. vesca* 815 × *F. bucharica* 601 map using the mapping data of Sargent et al. (2008) following the procedures described therein. The map presented was constructed using MAPCHART 2.2 software for Windows (Voorrips 2002).

Functional annotation of the sequences derived from hiTAIL-PCR of T-DNA insertional mutant DNA sequences

Each *Fragaria* sequence flanking a T-DNA insertion site was blasted against the GenBank database (http://www.ncbi.nlm.nih.gov/) to identify proteins that may be homologous to the strawberry DNA sequences flanking the T-DNA insertions. Blastx [version 2.2.18 (Altschul et al. 1997)] was used to search flanking DNA sequences against the protein "nr" database downloaded on 15 May 2009

Table 1 Locus names, primer sequences, restriction enzymes, and product sizes of the 74 flanking sequence amplification products that were mapped by PCR CAPS/dCAPS-based markers from the *F. vesca* insertional mutants

Locus name	Forward primer (5′ to 3′)	Reverse primer (5′ to 3′)	Restriction enzyme	Product sizes (bp)	
				F. vesca 815	*F. bucharica* 601
VT-008a	aacttgacagctccgggattatgc	ctgcagttgtgcaacttaatcaacagt		650	
VT-008	tggtttgacctactgatcctctctgg	cactggattccaagtacttcctgat	*Mbo*I	550	680
VT-010	tgcccagtgctccctgtttgaaat	ctactgcatcaaactgacatgcgt	*Hpa*II	300/400	300/375
VT-021	agtggtgtaggctgtagctgaagt	cggcaataatcttgtccgtgtatgg	*Hae*III	200/250	500
VT-022	tcccaatccctcttattcttctgaaggc	ttgccagtcgggaattcctgttac	*Mbo*I	150	
VT-025	ccctttctgaacatgcatcccact	cagttgagtcactctaatttccttggg	*Mbo*I	550	680
VT-026	gcattctaagtcatggtcgacggg	gcaaccctagacataggaatatcgc		250	
VT-029	gccactttagaaatgcggct	tgttaatgtatgtgaatgatgtgacgc		600	750
VT-032	ttgtttcgcacgaggaagagggaa	aaaggatgggcgcagttgttcgat	*Mae*II	700	300
VT-034	gcaactgtccatgggcatctcaat	tgagcaccagcttgttgtatcagc	*Alu*I	740	300
VT-036	ccacaatgggtccacaattacaac	accatgtcggtaagcatcttct		200	230
VT-076	tgttgctgctcttagtgtgtgac	ggtccaaagagtctaaggtggtga		800	350
VT-081	tggaggcctttccataacccatga	accttctggcctctctcaatgaac	*Tsp*4CI	320	250
VT-087	cccaaagctctctctttccacaac	gatagtgcgtctgaggtgaatt	*Xap*I	150	190
VT-136	ttgccttgacctattccaggta	ccagagtaacaccaagttgaggca	*Rsa*I	150	200
VT-353	ttaacgtactaaaagatctccggtg	gggggatcggtatggaaatt	*ACi*I	300	450
VT-378	gcacgagtaaactttcccagttctg	acagaaggcttcctgaaactcagc		450	490
VT-382	ttaacgtactaaaagatctccggtg	tgttagctggaatttaatgtcaacc	*ACi*I	250	400
VT-393	ccacctcacccctctttttctctctctctt	ccagttcctgctcctcaacatccttaag		350	400
VT-398	acgttccaatcaaatttaaacatgt	gaagagaatggattgaccaaaaca		400	
VT-402	cccttggatcttagtttatcacatg	ttctaatacccatacgaaagggttt	Alu*I*	450	250
VT-408	atatgcatctggtcttaatcgatca	agtaatgaagctagcgtggtgtaga	*Tsp4CI*	450	380
VT-413	gtctaagcgtcatgatatccaacccc	aaggacaagatacgcgagacagtgag	*ApoI*	380	280
VT-423	agaacaaaaataaccgataccttgc	caaaatcatcaccacacccatcta		300	
VT-637	acaatattgcaaaagctcacccaag	ggtgccatggtacccctataattcat		1,100	
VT-643	aaatcgtaaaatttgggacgtgacat	attcttcattttctaatgct	*Tsp*4*CI*	350	280
VT-650	agggacacgtgtttgctaccagtt	gataaagcagtgaaaccgcatttta	*TspEI*	330	180
VT-656	aattactgtattagctccatcc	ccactttctcaatttctagt		300	
VT-659	aaccagggttttgtttcttgttg	gcaaacacaaaagaatgagactgaa	*Mbo*II	150	280
VT-661	accctcactacacatgcacttca	tgtcgtaagagggttcgtgtttaat	*Hin*dIII	150	320
VT-666	ttcacttctggtgtttctctttgtc	tggagagatggaattatagcagcta		300	
VT-673	tcatggtgtttgcactccacttt	aggcataagcagtgtgcgctatac	*Mae*II	180	300
VT-674	agtcaccaaaaccaccacgg	tgggcttagcgtatgtagtaggat	*Cvi*RI	250	280
VT-675	atatccaatacgacacagtcctgaa	gctttcatcctgttttggttcaaa	*Bsm*I	550	400
VT-679	aaagggcttttatatgggcttg	gcttaacctcaagcctcctaaatta		250	
VT-701	ttttggattggataagataaggcg	tcttctttgttagggtctgtgtgat	*Taq*I	300	420
VT-704	cgatagcatacacgcacccc	gcaccaaccaaagtgataatggatt	*Tsp*EI	180	250
VT-710A	aaaagagagtcctagaaaaacccagac	atttgctctgagtttgggctgctc		280	250
VT-710B	aaaagagagtcctagaaaaacccagac	atttgctctgagtttgggctgctc		280	250
VT-712	tagagcttagctgtcggtcattatg	ggttctggtgagaatgatttcatta	*Cvi*R1	230	280
VT-714	accaattgggtgttgccatg	ggagaggaaaccgatcaagcaatg		420	450
VT-720	accagtccctgacttcaaatatttt	agcaaaagtatgacacgaactgc	*Mbo*I	350	200
VT-723	cagaatcggacccccttgactgtt	ggaaccttgcacttgctgctttta	*Mn1*I	280	180
VT-725	catcttggatgacatagatgctctatccca	cagaatgtgtgggtaatagttggtctttga	*Eco*RI	680	450
VT-734	ataagggaattggtcgaaggtg	tgaacacaacacaatccaagctt	*Tsp*4CI	300	380
VT-735	ctgatcgagtttcagagagaatggc	gaatgagatagagagaaaagagaagaaaca		250	320

Table 1 continued

Locus name	Forward primer (5′ to 3′)	Reverse primer (5′ to 3′)	Restriction enzyme	Product sizes (bp)	
				F. vesca 815	*F. bucharica* 601
VT-737	ggctatcttggctagtgtgatgca	ttccaattcaggcctagtaggagc			270
VT-739	atcattctccttagcaatcgtgttcatctg	cgctaggcttgttaattactgcagttctat		1,000	
VT-743	ccaataacacaagaaggtatggca	caagagaggggattagaccaaa	*Mbo*II	150	200
VT-744	ctggttgttgtttgactagttgca	aactagtatttgagtccaaagccg		750	350
VT-748	ctggatttgggatcaacctcagac	ggaccaaggctaccagaatattcaa		550	
VT-752	ctctcaagggtacgtacaagtactatttgc	cgccatagtgtccctagattttaagtcttt	*Mbo*II	250	450
VT-753	attcaatccagttcagcattcaaaggg	gcaattaagtttggtcaatcatggagactc	*Cvi*RI	180	350
VT-756	tcagagccccaagatcgattt	cacacacagtagactccagtacaaccct	*Hpy*188I	250	210
VT-763	gggtgataatgatcagtagccgttgc	gcccttctcttacagacgtgtcttga			1,400
VT-773	gctgctccttgggaatgttacttg	agaagtcctaaataaccttaatgtcgaca	*Eco*RV	420	280
VT-792	gtttctaggccttaggctctttgg	cttcaggcatggcactctctc		400	550
VT-794	ataaacgcggggcaaaatcg	cagagatggcggaggtcatc	*Aci*I	240	280
VT-803	tatgctgaagtaacatatattctgctttcc	tcttgctacatccaaatgataactatcaa		500	400
VT-811	gtcaaattgatcctcggcagt	cggttgtttgatgcaatgcc		420	
VT-816	gataagtgggaagtcaagagtcagg	tggtcccttattaagtctatcggtc	*Mn1*I	350	250
VT-824	ggtatagacccaaattactgctaatctaca	aactatcctcccgagtcccaagta	*Hpy*188I	290	220
VT-825	tggtttttggatgatctatcaactagc	tgtgctgcactatacaaatatagttgtcct		450	550
VT-837	agtggtcactgtgctcttgtatcttcc	ggggtaggaagaattccaaaagttttg	*Apa*I	200	220
VT-838	ttattatgatgttgattgggctgctcg	ggatgatggatctttatataattggtgtt	*Hpy*188I	120	290
VT-846	atttcaatctagactgatacgaagtgagcg	ccaaatcccagacccaatcaaaa		1,200	1,400
VT-850	atatgtagcgtatgctcttacttaattggt	ggcagggaatgaagtcacctatgt	*Hpy*188I	300	400
VT-851	tactgagcacagggtgtataaaggga	ttcctaggagtgcagctttgttcc	*Bbv*I	300	200
VT-855	tttgaccatataataaatccttcactgtca	gaacaccaaatatgcataaaagttgtctat		400	
VT-861	cacaaattgctttggtcaaggttag	catagtacagctgccttccacaga	*Mse*I	400	280
VT-869	atatgtagcgtatgctcttacttaattggt	acaccatcatctccaccaccat	*Hpy*188I	300	350
VT-872	cacaaattgctttggtcaaggttag	aacagcaagacccatgagagca	*Mse*I	320	250
VT-873	cgggtcatttaacacgtcaagc	catgacatgtgacacgcgttgt	*Mbo*I	650	450
VT-877	aattggtcatcgtattggtaatttcc	ccaatctcaattcagaggtcaaaga	*Aci*I	500	600

with the expectation threshold set to $1e^{-20}$ and other parameters at default values. The top-scoring hit to each flanking sequence was used as evidence of proximity of a coding region.

Results

Isolation of unknown flanking sequences by hiTAIL-PCR

Flanking regions were recovered from either the left or right borders of the T-DNA insertion sites (Fig. 1). An average fragment size of 800 bp (ranging from 500 to 2,400 bp) was obtained. Only fragments greater than 600 bp were excised from the gel for sequencing. Specific products were not always observed in the primary reactions; however, once the primary reactions were diluted

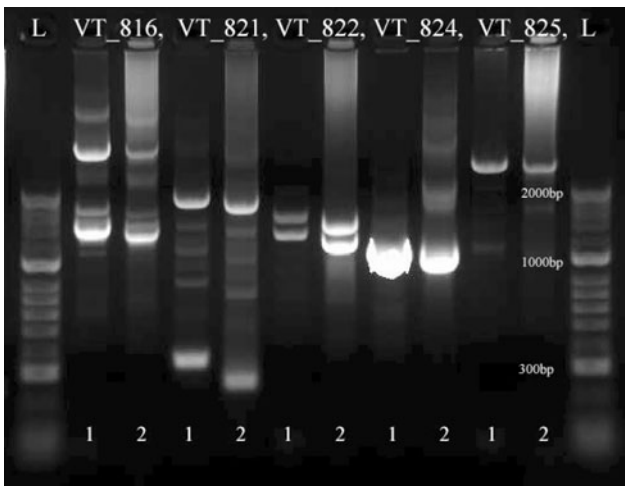

Fig. 1 Agarose gel electrophoresis of hiTAIL-PCR products of T-DNA insertional mutants of *F. vesca*. *L* 2,000 kb marker; under each of the five VT Mutant IDs, *lanes 1* and *2* represent the primary and secondary hiTAIL-PCR products, respectively

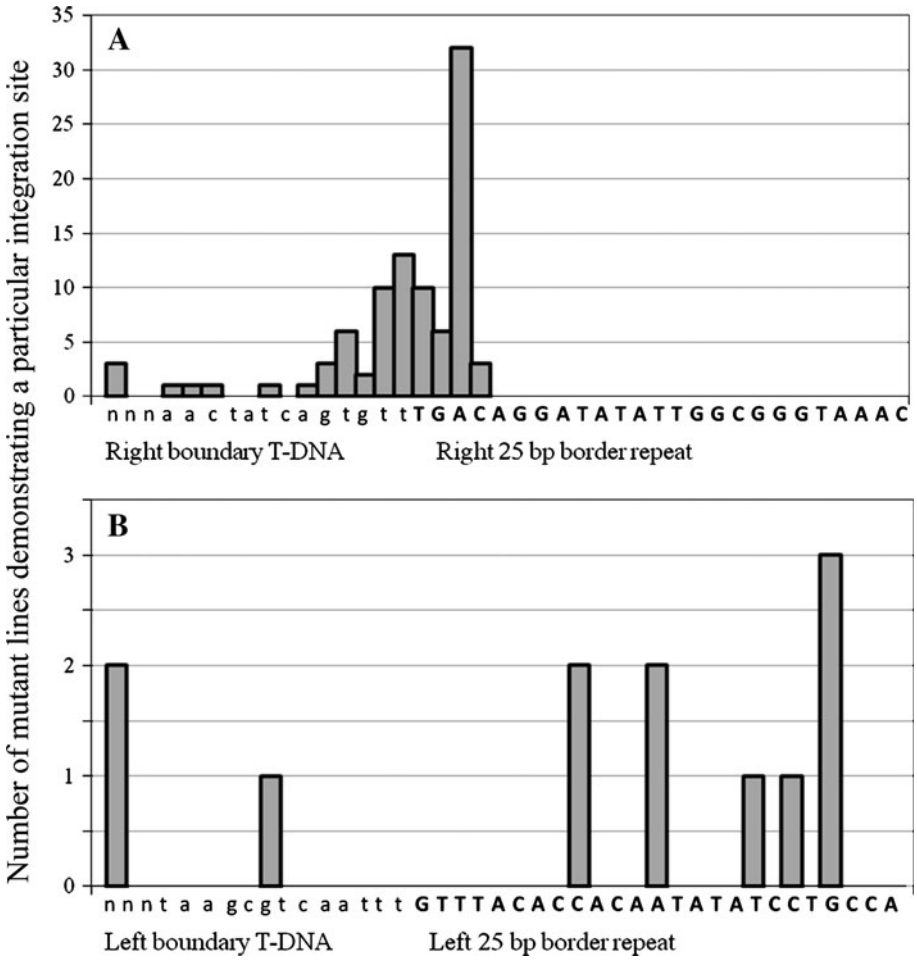

Fig. 2 Sequence analysis of pCAMBIA T-DNA insertion sites (IS) adjacent to the **a** right and **b** left border repeats. On the *x*-axis, nucleotides in lower case represent the T-DNA, those in bold and capital letters represent the 25 bp right and left imperfect repeat sequence adjacent to the T-DNA in the plasmid backbone. **a** Shows 96 mutants and **b** shows 12 mutants exhibiting each insertion. The *Y*-axis is the number of all mutants with each integration pattern

1:10 and used as a template for the secondary reactions, the specific products amplified during the subsequent secondary reactions were visible on a 1.2% agarose gel (Fig. 1). All sequences generated by hiTAIL-PCR contained approximately 128 bp of T-DNA right or left border sequence in addition to the flanking genomic sequence of the strawberry genome. In addition, the alignment of sequences revealed three duplicate sets of mutant lines (VT-861 and VT-872; VT-850 and VT-869; and VT-353 and VT-382), suggesting that more than one transgenic shoot that emerged from a callus representing the same transgenic event had been selected and transferred to the glasshouse.

Analysis of the boundaries of the T-DNA

In total, 108 transformation lines appeared to contain a single T-DNA insertion site that was stably integrated into the strawberry genome. Of these, 51 plants (53%) had T-DNA insertion sites that contained at least the first 4 bp of the 25 bp right border repeat sequence, while the remaining 45 (47%) lacked any of the right border repeat sequence. Nine (75%) of the sequences obtained with the left T-DNA border primers showed between 9 and 22 bp of the 25 bp left border repeat sequence, with the remaining three (25%) lacking any of the left border repeat sequence (Fig. 2).

Product amplification and scoring in the *F. vesca* 815 × *F. bucharica* 601 mapping progeny

All 108 primer pairs designed from transformation lines containing a single T-DNA insertion generated amplicons in one or both of the parents of the *F. vesca* 815 × *F. bucharica* 601 mapping population. Amplicons were obtained from just a single parent with 24 (22%) of the primer pairs: of these 20 (19%) amplified a product only in *F. vesca* 815 whereas four (2%) amplified a product only in *F. bucharica* 601. In subsequent analyses, 12 of these with only a *F. vesca* amplicon segregated 3:1 in the F_2 population as expected, two did not segregate as expected (assuming homozygosity in the parents), four displayed an amplification product for the other parent when the reaction was performed on the F_2, and two were monomorphic (*F. vesca* allele only) in the F_2. A further 15 (14%) primer pairs generated amplicons of different sizes between *F. vesca* 815 and *F. bucharica* 601 that could be resolved after agarose gel electrophoresis (Fig. 3); however, one of these could not be scored unambiguously in the

Fig. 3 Agarose gel demonstrating amplification patterns using primers developed from T-DNA flanking regions recovered from four different insertional mutants (VT) of *F. vesca.* Under each VT mutant number, *lane 1* is *F. vesca* 815, *lane 2* is *F. bucharica* 601 and *lane 3* is wild type *F. vesca* FV10, *L* 2,000 bp marker

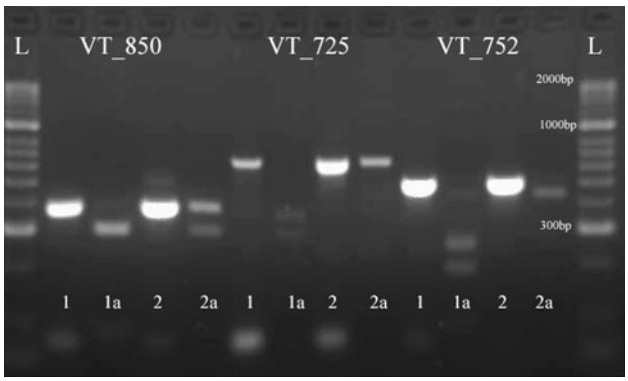

Fig. 4 Restriction enzyme digestion of PCR products generated from amplifying in T-DNA flanking sequences of three insertional mutants (VT) of *F. vesca* where the initial amplification products had been indistinct. Under each VT mutant number, *lanes 1* and *1a* are undigested and digested *F. vesca* 815, respectively, *lanes 2* and *2a* are undigested and digested *F. bucharica* 601, respectively; *L* 2,000 bp marker

F. vesca 815 × *F. bucharica* 601 progeny. A total of 79 (73.2%) primer pairs generated PCR products of the same size in the parents of the *F. vesca* 815 × *F. bucharica* 601 mapping population. Of these, 14 contained no SNP and thus were not polymorphic between *F. vesca* 815 and *F. bucharica* 601. Of the 65 (60%) sequences containing one or more SNPs, nine were heterozygous in the *F. bucharica* 601 parent but contained an allele common to the *F. vesca* 815 parent (*ab*×*aa*) and were subsequently homozygous in the F_1 plant (*aa*) from which the F_2 population derived, and thus did not segregate in the progeny. Ten primer pairs produced weak PCR amplicons and could not be scored following restriction enzyme digestion. The remaining 46 primer pairs yielded amplicons that could be cut with a restriction enzyme at a SNP position to reveal polymorphism between *F. vesca* 815 and *F. bucharica* 601 that segregated in the mapping population (Fig. 4). Table 1 lists locus name, forward and reverse primers, and the restriction endonucleases used to digest products containing SNPs, along with the amplicon sizes observed after digestion in the parents of the *F. vesca* 815 × *F. bucharica* 601 mapping population for the 74 mapped loci. Supplemental Table 2 lists similar information for the remaining 34 loci that could not be mapped using this strategy.

Thus, a total of 74 polymorphic markers segregated in the *F. vesca* 815 × *F. bucharica* 601 mapping population. Of these, 48 were co-dominant and segregated in an expected 1:2:1 Mendelian ratio in the progeny, 21 were dominant in *F. vesca* 815 and segregated in a 3:1 ratio in the F_2, and a further five were dominant in the *F. bucharica* 601 parent and approximated a 1:3 segregation in the progeny. Table 2 details the segregation data and chi-square values for goodness-of-fit to the expected and observed Mendelian segregation ratios for the 74 polymorphic markers segregating in the *F. vesca* 815 × *F. bucharica* 601 progeny. In total, 31 of the 74 markers (42%) exhibited significant deviation from the expected Mendelian segregation ratios (Table 2). In addition, the two markers in each of the three sets of duplicate (VT-861 and VT-872; VT-850 and VT-869; VT-353 and VT-382) T-DNA transgenic lines that were identified by hi-TAIL PCR produced identical segregation scored as expected.

Genetic linkage mapping of SNPs in the diploid *Fragaria* reference map (*F. vesca* 815 × *F. bucharica* 601)

After cosegregation analysis, all 74 polymorphic, segregating markers mapped to discrete loci on the *F. vesca* 815 × *F. bucharica* 601 reference map (Fig. 5). An average of 10.5 VT markers mapped to each of the seven *F. vesca* 815 × *F. bucharica* 601 linkage groups. The lowest number of markers (5) mapped to linkage group FG1, the greatest (16) mapped to linkage group FG5. Eight markers mapped to FG2, 9 to FG3, 13 to FG4, 14 to FG6, and 9 to FG7. Eight markers were located at the ends of six of the *F. vesca* 815 × *F. bucharica* 601 linkage groups, extending these groups by a total of 66.7 cM. The VT markers distributed across the seven linkage groups of the *F. vesca* 815 × *F. bucharica* 601 map and markers VT-861 and 872 (FG4) were located within a gap of more than 20 cM between markers BG-1 and CEL-1 on the previously published map of Sargent et al. (2008).

Functional annotation of the flanking regions for insertional mutant lines

All 108 T-DNA strawberry flanking sequences were blasted against the GenBank nr database (http://www.ncbi.nlm.

Table 2 Segregation analysis (chi-square values for goodness-of-fit to the expected Mendelian segregation ratios), linkage group (FG), map position, sequence characterization of T-DNA right (R) or left (L) border, and percentage of G/C content in each flanking sequence of 74 PCR markers derived from the flanking DNA sequence of T-DNA mutant strawberry lines in the diploid strawberry mapping population

Locus name	Classification	Expected	Observed	χ^2	*df*	Significance[a]	Linkage group	Map position	T-DNA border	% G/C content
VT-008	aa:ab:bb	1:2:1	19:42:13	2.32	2	–	FG6	42.7	RB	36.8
VT-008a	a_:bb	3:1	37:24	6.69	1	**	FG2	0.0	RB	38.0
VT-010	a_:bb	3:1	48:25	3.33	1	–	FG5	50.0	LB	33.1
VT-021	aa:ab:bb	1:2:1	10:30:32	15.44	2	***	FG4	32.1	LB	38.0
VT-022	aa:ab:bb	1:2:1	13:42:17	2.44	2	–	FG3	43.7	RB	37.6
VT-025	aa:ab:bb	1:2:1	20:42:13	2.39	2	–	FG6	42.4	RB	44.0
VT-026	a_:bb	3:1	57:17	0.16	1	–	FG1	31.9	RB	33.9
VT-029	aa:ab:bb	1:2:1	19:30:20	1.2	2	–	FG4	51.0	RB	32.9
VT-032	aa:ab:bb	1:2:1	10:33:29	10.53	2	**	FG4	31.0	LB	44.2
VT-034	aa:ab:bb	1:2:1	12:26:35	20.53	2	***	FG7	18.7	RB	33.8
VT-036	aa:ab:bb	1:2:1	3:30:39	38.0	2	***	FG2	18.0	RB	31.6
VT-076	a_:bb	3:1	52:23	1.28	1	–	FG6	72.8	RB	34.7
VT-081	aa:ab:bb	1:2:1	14:30:11	0.78	2	–	FG1	64.3	LB	38.6
VT-087	aa:ab:bb	1:2:1	10:37:29	9.55	2	**	FG5	51.8	RB	42.6
VT-136	aa:ab:bb	1:2:1	13:33:28	6.95	2	*	FG7	15.3	RB	38.5
VT-353	aa:ab:bb	1:2:1	12:17:25	13.67	2	***	FG7	5.2	RB	35.5
VT-378	aa:ab:bb	1:2:1	32:40:3	22.76	2	***	FG6	12.2	RB	36.5
VT-382	aa:ab:bb	1:2:1	12:17:25	13.67	2	***	FG7	5.2	RB	36.4
VT-393	aa:ab:bb	1:2:1	18:16:24	12.9	2	***	FG5	21.3	RB	40.4
VT-398	a_:bb	3:1	46:30	8.49	1	***	FG3	14.7	RB	32.7
VT-402	aa:ab:bb	1:2:1	13:48:6	14.01	2	***	FG4	0.0	RB	29.0
VT-408	aa:ab:bb	1:2:1	12:39:22	3.08	2	–	FG3	28.3	RB	36.0
VT-413	aa:ab:bb	1:2:1	9:35:26	8.26	2	–	FG5	50.6	RB	44.3
VT-423	a_:bb	3:1	44:32	11.86	1	**	FG7	22.3	RB	39.0
VT-637	a_:bb	3:1	38:38	25.33	1	**	FG2	22.3	RB	38.7
VT-643	a_:bb	3:1	45:30	9.0	1	**	FG5	55.0	RB	33.2
VT-650	aa:ab:bb	1:2:1	14:27:22	3.32	2	–	FG1	15.3	RB	42.6
VT-656	a_:bb	3:1	47:27	5.21	1	–	FG5	54.5	RB	29.1
VT-659	aa:ab:bb	1:2:1	15:39:22	1.34	2	–	FG7	63.7	RB	39.1
VT-661	aa:ab:bb	1:2:1	12:28:32	14.67	2	**	FG7	18.0	RB	36.6
VT-666	a_:bb	3:1	42:34	15.79	1	**	FG5	59.8	RB	41.6
VT-673	aa:ab:bb	1:2:1	11:36:25	5.44	2	–	FG3	20.8	RB	34.0
VT-674	aa:b_	1:3	17:56	0.11	1	–	FG6	50.0	RB	44.6
VT-675	aa:ab:bb	1:2:1	10:37:22	4.54	2	–	FG5	56.2	RB	40.5
VT-679	a_:bb	3:1	50:26	3.44	1	–	FG6	78.9	RB	42.3
VT-701	aa:ab:bb	1:2:1	15:40:19	0.92	2	–	FG4	76.3	RB	42.2
VT-704	aa:ab:bb	1:2:1	23:22:11	7.71	2	*	FG6	0.0	RB	41.7
VT-710A	aa:ab:bb	1:2:1	9:33:22	5.34	2	–	FG5	64.8	RB	46.2
VT-710B	aa:ab:bb	1:2:1	11:38:16	2.63	2	–	FG6	76.2	RB	46.2
VT-712	a_:bb	3:1	49:25	3.05	1	–	FG3	25.1	RB	39.7
VT-714	aa:ab:bb	1:2:1	5:30:34	25.55	2	***	FG2	26.0	RB	41.7
VT-720	aa:ab:bb	1:2:1	14:35:12	1.46	2	–	FG6	57.1	RB	34.2
VT-723	a_:bb	3:1	25:48	3.33	1	–	FG2	40.3	RB	41.8
VT-725	aa:ab:bb	1:2:1	12:32:16	0.8	2	–	FG3	52.5	RB	40.6
VT-734	aa:ab:bb	1:2:1	16:30:27	5.63	2	–	FG5	18.5	RB	46.9
VT-735	aa:ab:bb	1:2:1	24:18:12	11.33	2	***	FG5	81.3	RB	33.6

Table 2 continued

Locus name	Classification	Expected	Observed	χ^2	*df*	Significance[a]	Linkage group	Map position	T-DNA border	% G/C content
VT-737	a_:bb	3:1	40:35	32.11	1	***	FG2	16.7	RB	46.8
VT-739	a_:bb	3:1	65:6	10.37	1	***	FG6	35.9	RB	37.6
VT-743	aa:ab:bb	1:2:1	17:37:19	0.12	2	–	FG4	81.1	RB	39.9
VT-744	aa:ab:bb	1:2:1	16:36:26	5.35	2	–	FG3	24.3	RB	37.0
VT-748	a_:bb	3:1	56:20	0.07	1	–	FG6	56.1	RB	32.8
VT-752	aa:ab:bb	1:2:1	12:32:30	10.11	2	*	FG5	36.0	LB	36.0
VT-753	aa:ab:bb	1:2:1	20:29:19	1.5	2	–	FG4	85.9	LB	36.8
VT-756	aa:ab:bb	1:2:1	11:31:32	13.86	2	**	FG4	34.6	RB	37.9
VT-763	aa:ab:bb	1:2:1	11:41:20	3.64	2	–	FG3	27.5	RB	45.0
VT-773	aa:ab:bb	1:2:1	8:35:27	10.31	2	**	FG5	49.6	RB	38.1
VT-792	a_:bb	3:1	52:24	1.75	1	–	FG6	64.3	RB	37.0
VT-794	aa:b_	1:3	10:63	4.97	1	**	FG4	31.5	RB	45.5
VT-803	aa:ab:bb	1:2:1	10:32:11	2.32	2	–	FG7	77.8	RB	32.2
VT-811	a_:bb	3:1	45:31	10.11	1	***	FG5	53.7	RB	43.0
VT-816	aa:ab:bb	1:2:1	8:22:29	18.76	2	***	FG4	40.2	RB	41.4
VT-824	aa:ab:bb	1:2:1	17:42:11	3.83	2	–	FG6	43.5	RB	43.0
VT-825	aa:ab:bb	1:2:1	6:35:32	18.64	2	***	FG2	24.9	RB	52.2
VT-837	a_:bb	3:1	30:29	18.36	1	***	FG2	8.3	RB	36.1
VT-838	aa:b_	1:3	20:55	0.11	1	–	FG7	45.4	RB	34.9
VT-846	aa:b_	1:3	18:58	0.07	1	–	FG1	27.8	RB	37.2
VT-850	aa:ab:bb	1:2:1	13:30:30	10.23	2	**	FG5	34.7	RB	48.1
VT-851	a_:bb	3:1	39:26	7.8	1	**	FG4	20.2	RB	39.5
VT-855	a_:bb	3:1	52:24	1.75	1	–	FG3	34.1	RB	32.8
VT-861	aa:ab:bb	1:2:1	10:38:23	5.11	2	–	FG4	60.4	RB	35.8
VT-869	aa:b_	1:3	13:55	1.25	1	–	FG5	34.7	RB	45.9
VT-872	aa:ab:bb	1:2:1	11:40:23	4.38	2	–	FG4	60.4	RB	35.3
VT-873	aa:ab:bb	1:2:1	23:37:11	4.18	2	–	FG1	51.4	RB	33.7
VT-877	aa:ab:bb	1:2:1	5:41:11	12.23	2	***	FG6	57.1	RB	33.9

[a] *,**,***, significantly different from expected segregation at $p < 0.05$, 0.01, and 0.001 respectively; – not significantly different from expectation

nih.gov/). Of these, 49 (45%) showed high homology to proteins with a predicted or specific functionality. Table 3 lists the 49 VT loci and their putative protein function, GenBank accession number of the protein to which they were most closely matched and the species from which the protein was identified.

Discussion

Isolation of unknown flanking sequences by hiTAIL-PCR and analysis of the boundaries of the T-DNA

Agrobacterium-mediated transformation has many advantages over other transformation methods, including high efficiency, transfer of defined pieces of DNA (T-DNA), and the ability to transfer relatively large segments of DNA. This transformation process is controlled chromosomally and by plasmid-virulence encoded proteins. These virulence genes are located on the *Ti* plasmid and are responsible for the success of transfer of the T-DNA into the plant genome by illegitimate recombination, a process that joins two DNA molecules that do not share extensive homology. However the mechanism by which T-DNA integrates into the host genomes and the plant factors involved in the integration process are poorly understood (Brunaud et al. 2002; Bundock and Henry 2004; Dafny-Yelin et al. 2009; Windels et al. 2008). Finding the location of inserted T-DNA elements and characterizing the genomic DNA sequence into which they have been inserted using hiTAIL-PCR encompass a strategy for understanding gene function by forward genetics. We used hiTAIL-PCR to characterize strawberry T-DNA flanking sequences using nested specific primers from both right and left

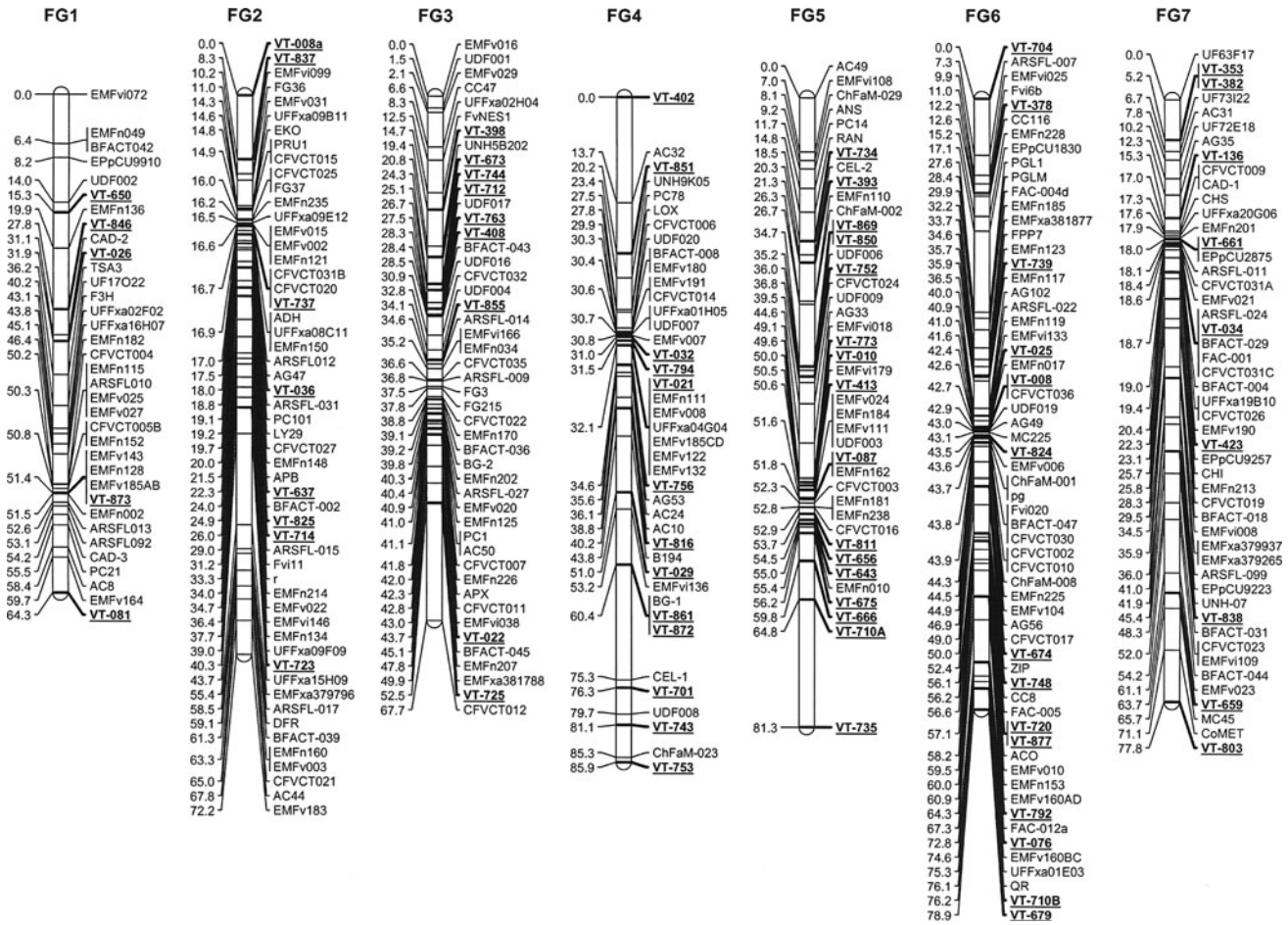

Fig. 5 Genetic linkage map of diploid strawberry demonstrating the positions of 74 insertional mutants of *F. vesca*. SNPs in the flanking sequences were used to develop CAPS/dCAPS markers. The new SNP-based markers as well as indel-based markers are identified by their VT mutant number and highlighted in bold. *Map distances* on the *left* are in cM

borders of the T-DNA, and LAD primer pools (Fig. 1). Although the LAD primers used in the pre-amplification reaction are nonspecific (Liu and Chen 2007), all four were able to amplify strawberry DNA to create one or more annealing site(s) for the primary and secondary hiTAIL-PCR reactions. Flanking sequences generated from more than 100 transgenic strawberry plants exhibited approximately 128 bp of the inserted T-DNA and then a flanking strawberry sequence. These results concur with Liu and Chen (2007), where all transgenic rice lines that were sequenced contained the inserted T-DNA as well as its flanking sequence.

Sequence analysis of both left and right borders of the T-DNA revealed variation of integration patterns of the T-DNA: the first 4 bp from the T-DNA 25-bp right border and up to 22 bp from the T-DNA 25-bp left border had integrated in different mutants. In rice, sequence analysis between the genomic DNA and T-DNA showed integration ranging from 1 to 3 bp of the 25-bp right border, and 6 up to 20-bp of the 25-bp left border (Hiei et al. 1994). Similar results have also been found in tobacco (Zambryski et al. 1982) and Arabidopsis (Brunaud et al. 2002).

We found that the flanking sequences analyzed showed an average GC content of 39.74% (SD = 0.053%). This deviates strongly from the GC content of 573 non-chloroplast protein coding regions found in GenBank (24 Feb 2010) with an average GC content of 44.6 ± 5.1% (*P* value from two-tailed *t* test = $1.8e^{-93}$). The background GC content of a typical rosaceous genome was estimated from a sample of 2,000 genome survey sequences from peach (*Prunus persica*) available in GenBank: mean = 38.14%, SD = 5.37%. To the extent that the *Prunus* genome approximates the *Fragaria* background GC content, tDNA insertion sites are not strongly biased by GC level (Fig. 6). Pan et al. (2005) and Brunaud et al. (2002) in analyses of T-DNA integration into the Arabidopsis genome found a preference for integration into regions ranging from 20 to 30% GC; however, the genomic GC content of Arabidopsis is about 36% and is low relative to other plant genomes.

Table 3 BLASTx analyses and predicted gene function of 49 T-DNA flanking strawberry sequences generated by hiTAIL-PCR from GenBank (http://www.ncbi.nlm.nih.gov/), as well as locus name, GenBank accession number, score, *E* value, and species where it was identified first

Locus name	GenBank accession No	Predicted gene function	GenBank NR protein data set result		Species
			Score	*E* value	
VT-008a	EEF41624.1	Peroxisome assembly factor-2, putative	113	$1e^{-22}$	*Ricinus communis*
VT-021	AAD49734.1	AF169795-1 glutamine synthetase precursor	130	$9e^{-28}$	*Juglans nigra*
VT-022	XP-002313228.1	Predicted protein	107	$8e^{-21}$	*Populus trichocarpa*
VT-025	EEF46369.1	Stem 28 kDa glycoprotein precursor, putative	186	$1e^{-44}$	*Ricinus communis*
VT-032	EEF41851.1	RNA binding motif protein, putative	273	$7e^{-71}$	*Ricinus communis*
VT-033	EEF36498.1	Protein pof4, putative	80	$8e^{-29}$	*Ricinus communis*
VT-054	XP-002313425.1	Cytochrome P450	245	$1e^{-62}$	*Populus trichocarpa*
VT-076	EEF49342.1	Lysosomal alpha-mannosidase, putative	102	$6e^{-57}$	*Ricinus communis*
VT-087	lNP-177103.1	AGO7 (ARGONAUTE7); nucleic acid binding	112	$5e^{-40}$	*Arabidopsis thaliana*
VT-136	EEF37316.1	Erythroblast macrophage protein emp	132	$5e^{-42}$	*Ricinus communis*
VT-393	EEF51349.1	Protein binding protein, putative	116	$1e^{-23}$	*Ricinus communis*
VT-413	EEF48596.1	Transcription factor, putative	310	$7e^{-82}$	*Ricinus communis*
VT-637	EEF41389.1	Interferon-induced guanylate-binding protein, putative	109	$9e^{-23}$	*Ricinus communis*
VT-640	EEF41389.1	Interferon-induced guanylate-binding protein, putative	109	$9e^{-23}$	*Ricinus communis*
VT-660**	CAA71490.1	Peroxidase	117	$4e^{-24}$	*Spinacia oleracea*
VT-666	EEF48571.1	Clathrin assembly protein, putative	157	$3e^{-36}$	*Ricinus communis*
VT-669**	EEF33958.1	Amino acid transporter, putative	169	$1e^{-39}$	*Ricinus communis*
VT-674	EEF51150.1	Leucine-rich repeat receptor protein kinase EXS precursor	213	$7e^{-53}$	*Ricinus communis*
VT-675	XP-002314629.1	SIN3 component, histone deacetylase complex	203	$9e^{-50}$	*Vitis vinifera*
VT-676**	EEF51150.1	Leucine-rich repeat receptor protein kinase EXS precursor, putative	714	0.0	*Ricinus communis*
VT-679	NP-198784.1	EMB2744 (EMBRYO DEFECTIVE 2744)	483	e^{-134}	*Arabidopsis thaliana*
VT-682**	lXP-002316906.1	Predicted protein	223	$5e^{-56}$	*Populus trichocarpa*
VT-691**	EEF39690.1	Leucoanthocyanidin dioxygenase, putative	226	$4e^{-72}$	*Ricinus communis*
VT-701	EEF36999.1	Polygalacturonase, putative	219	$1e^{-54}$	*Ricinus communis*
VT-714	XP-002264804.1	PREDICTED: hypothetical protein	114	$6e^{-23}$	*Vitis vinifera*
VT-715	EEF49033.1	UDP-glucosyltransferase, putative	321	$9e^{-86}$	*Ricinus communis*
VT-721	EEF46451.1	Cytoplasmic dynein light chain, putative	94	$1e^{-23}$	*Ricinus communis*
VT-723	EEF36880.1	Nucleic acid binding protein, putative	334	$2e^{-89}$	*Ricinus communis*
VT-725	EEF35141.1	DNA binding protein, putative	465	e^{-143}	*Ricinus communis*
VT-729	NP-190817.1	Heat shock protein-related	483	e^{-134}	*Arabidopsis thaliana*
VT-730	EEF29797.1	Cop9 signalosome complex subunit, putative	91	$2e^{-28}$	*Ricinus communis*
VT-734	NP-849842.1	AAA-type ATPase family protein	103	$3e^{-32}$	*Arabidopsis thaliana*
VT-735	NP-173839.1	AAA-type ATPase family protein	301	$3e^{-79}$	*Arabidopsis thaliana*
VT-739	XP-002271602.1	PREDICTED: hypothetical protein	161	$4e^{-37}$	*Vitis vinifera*
VT-750**	XP-002311853.1	Glycosyltransferase, CAZy family GT8	563	e^{-158}	*Populus trichocarpa*
VT-753	XP-002272777.1	PREDICTED: hypothetical protein	125	$2e^{-26}$	*Vitis vinifera*
VT-756	EEF50019.1	ATP-citrate synthase, putative	247	$5e^{-63}$	*Ricinus communis*
VT-763	XP-002300564.1	Predicted protein	131	$3e^{-28}$	*Populus trichocarpa*
VT-794	NP-194760.2	Ribitol kinase, putative proteína	119	$1e^{-24}$	*Arabidopsis thaliana*
VT-798	lNP-849972.1	AAA-type ATPase family protein	503	e^{-140}	*Arabidopsis thaliana*
VT-811	EEF48468.1	Pollen-specific protein C13 precursor, putative	150	$6e^{-34}$	*Ricinus communis*
VT-816	EEF47537.1	Anthranilate N-benzoyltransferase protein, putative	246	$7e^{-63}$	*Ricinus communis*

Table 3 continued

Locus name	GenBank accession No	Predicted gene function	GenBank NR protein data set result		Species
			Score	*E* value	
VT-824	ABQ53132.1	Putative CBF/DREB transcription factor	113	$5e^{-23}$	*Rosa chinensis*
VT-850	ACN38268.1	Flavonoid 3′ hydroxylase	201	$4e^{-49}$	*Vitis amurensis*
VT-851	CAN83125.1	PREDICTED: hypothetical protein	176	$2e^{-42}$	*Vitis vinifera*
VT-861	XP-002274118.1	PREDICTED: hypothetical protein	367	$1e^{-99}$	*Vitis vinifera*
VT-865	NP-001151166.1	Acetolactate synthase/amino acid binding protein	128	$3e^{-27}$	*Zea mays*
VT-869	ACN38268.1	Flavonoid 3′ hydroxylase	201	$4e^{-49}$	*Vitis amurensis*
VT-872	XP-002274118.1	PREDICTED: hypothetical protein	367	$1e^{-99}$	*Vitis vinifera*

** Monomorphic markers

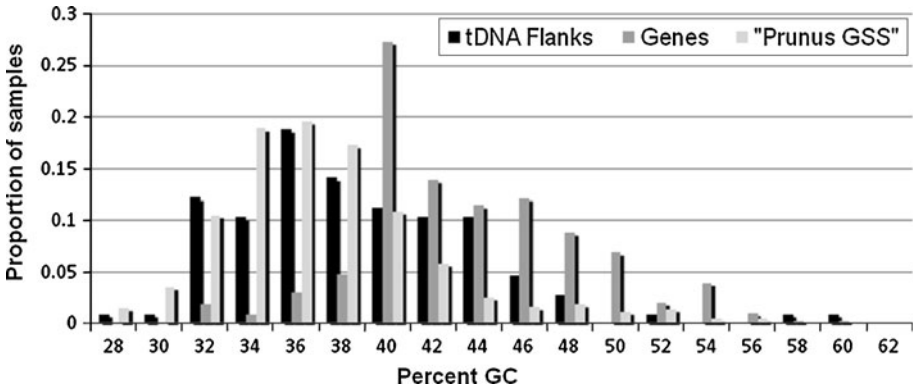

Fig. 6 Comparison of GC content among 108 tDNA flanking sequences (from hiTAIL-PCR), 573 *Fragaria* protein coding genes, and 2,000 *Prunus persica* genome survey sequences (GSS). The GSS from a closely related rosaceous genome estimate the background GC content of *Fragaria*

Amplification of products from strawberry flanking sequences and SNP discovery between *F. vesca* 815 and *F. bucharica* 601

After sequencing the flanking region of more than a 100 T-DNA strawberry mutants, all regenerated from *F. vesca* PI 551572, we designed and tested sequence-based primers on genomic DNA of the parental lines of the *F. vesca* 815 × *F. bucharica* 601 mapping population. For 39 (36%) of the primer pairs that revealed polymorphic products between the parental lines, we did not observe segregation in the F_2 mapping population due to the fact that the heterozygous *F. bucharica* 601 shared an allele in common with *F. vesca* 815 and this allele was transmitted to the F_1 plant used to raise the *F. vesca* 815 × *F. bucharica* 601 mapping population. However, 28 (26%) of the primer pairs revealed polymorphic products between the parental lines and segregated in the F_2 progeny allowing the markers to be placed on the *F. vesca* 815 × *F. bucharica* 601 map. The remaining 46 primer pairs amplified similar product size between the mapping parents. After sequencing those fragments we could design assays that utilized inherent differences between the parents in restriction sites (CAPS) or SNPs. Genotyping SNPs in a mapping population or among a panel of genotypes representative of a crop species can be accomplished using DNA sequencing platforms, microarrays, Taqman assays, high resolution melting analysis among others. Our approach was modeled after the efforts of Neff et al. (1998) who utilized incomplete restriction recognition sites in combination with the polymorphic SNP for allelic detection. The use of restriction enzymes with no more than 4 or 5 bp recognition sites decreased the cost of the procedure. For dCAPS, we designed primers around the incomplete recognition site with a mismatched nucleotide; when used to amplify the loci, a complete recognition site was created in one of the mapping parents thus allowing allelic discrimination. This approach has been used for SNP assaying to identify a functional polymorphism for the waxy character (*Wx*) of wheat (Yanagisawa et al. 2003), herbicide resistance via mutations in acetolactate-synthase (ALS) in ryegrass (Délye et al. 2009), and alternative marker developed for saturating the genetic linkage map of chickpea (Muehlbauer and Rajesh 2008). Our approach utilized the CAPS/dCAPS methodology to reveal the genetic location of the T-DNA insertion to characterize individual mutant lines.

SNPs represent the most common type of sequence polymorphism in plants, animals and humans (Brookes 1999; Henry 2008). In the present study, an average of one

SNP was found between *F. vesca* 815 and *F. bucharica* 601 every 60 bp, higher than the frequency previously observed between these two species accessions by Sargent et al. (2009). The abundance of nucleotide variation (SNPs, insertions and deletions) between *F. vesca* 815 and *F. bucharica* 601 facilitated the development of robust PCR markers that allowed us to place 74 new markers on the diploid strawberry genetic linkage map that correspond to the genetic location of T-DNA insertions in the respective transgenic strawberry lines. Flanking sequences that exhibited no SNP polymorphism were observed in nine cases, suggesting that these represented highly conserved regions. Some of these monomorphic sequences are shown in Table 3 with the predicted function of the interrupted gene as previously described in other crops such as coffee (Aggarwal et al. 2007).

Genetic linkage mapping of SNPs in the diploid *Fragaria* reference map (*F. vesca* 815 × *F. bucharica* 601)

The PCR CAPS/dCAPS-based markers showing polymorphism between the parents of the *F. vesca* 815 × *F. bucharica* 601 reference map were used to genotype the F_2 progeny analyzed in this investigation. We added and determined the location of 74 unique, independent single copy T-DNA insertion sites and increased the density of DNA markers saturating the reference *Fragaria* map. By comparing the position of our PCR CAPS/dCAPS-based markers within and between linkage groups, we deduced that eight VT markers were located at the distal regions of six linkage groups, thus elongating the linkage group beyond previously placed SSR markers (Fig. 5) (Hadonou et al. 2004; Sargent et al. 2004, 2006, 2007). We also found in linkage group FG3, FG4, and FG5 that four groups of three markers were closely linked to each other. One possible explanation is that these insertions occurred in genetic regions closely linked to the centromeres where recombination is generally repressed, thus giving the illusion of proximity. Alternatively, these genomic regions may harbor nucleotide sequences that are particularly susceptible to T-DNA integration.

Distorted segregation has been detected with almost all molecular and genetic markers due to factors like structure of the mapping population, genetic transmission, gametic and zygotic selection, non-homologous recombination, gene transfer, transposable elements and environmental agents (Jing et al. 2007; Knox and Ellis 2002; Zhao et al. 2006a). In our strawberry segregation data, we observed 42% of cases with significant deviation ($P \leq 0.05$) from the expected Mendelian segregation ratios (Table 2). Similar results were reported in the reference *Fragaria* map by Sargent et al. (2004) with 54% of SSR markers displaying distorted segregation. Molecular markers scored in an F_2 population of rice showed 33% genetic distortion (Zhao et al. 2006a), 54% in *Avena sativa* (Tanhuanpää et al. 2008), and 11% in sunflower (Lai et al. 2005). There were five cases (including 11 VT markers) where skewed markers grouped together, whereas the 21 remaining distorted VT markers mapped independently (Table 2; Fig. 5).

In two cases we observed an amplicon for *F. vesca* 815 but not *F. bucharica* 601 and expected a 3:1 segregation (presence:absence) in the F_2; however, we observed monomorphic F_2 with all genotypes exhibiting the *F. vesca* 815 amplicon. It is possible that this represents extreme distortion where the null allele of the *F. bucharica* 601 parent that was expected to be heterozygous in the F_1 could not be tolerated as homozygous in the F_2 with *F. vesca* 815 cytoplasm. The most likely map position for these markers would be at the proximal end of FG2 where the most extreme segregation distortion occurred on the *F. vesca* 815 × *F. bucharica* 601 map (Sargent et al. 2004).

Functional annotation of the hiTAIL PCR of T-DNA insertional mutant DNA sequences

To begin the functional characterization of T-DNA flanking strawberry sequences we used the GenBank NR protein data set to translate nucleotide sequences into encoded proteins. Almost 50% of the 108 T-DNA strawberry flanking sequences analyzed in this investigation were closely related to proteins with a predicted or specific functionality that have been previously characterized in other plant species. We found T-DNA mutants that likely alter transcription of genes affecting proteins related to: RNA binding motif, cytochrome P450, amino acid transport, cytoplasmic dynein light chain, ATPase, pollen-specific C13 precursor, flavonoid 3′ hydroxylase, among others (Table 3). Mutant VT_850 with a strong hit to a gene involved in flavonoid synthesis in *Vitis vinifera* is worthy of future investigation considering the important contribution of flavonoids to the health promoting properties of strawberry and other fruit crops (Shulaev et al. 2008). The remaining sequences (over half) had no predicted functionality. Mapped insertional mutants related to gene functions as described in this study provide an invaluable resource for forward genetics in fruit crops. Despite high transformation efficiency of *F. vesca* using the tissue culture methodology described by Oosumi et al. (2006), generating sufficient insertional mutants to saturate the strawberry genome would be prohibitively laborious to adopt a strategy of using a single transformation event to generate each insertional mutant. The *in planta* transformation protocol used to develop large knockout collections in Arabidopsis (Krysan et al. 1999; Robinson et al. 2009;

Woody et al. 2007) represents a much more efficient system but is unavailable for strawberry. Therefore, we are in the process of developing a transposon tagging strategy where unlimited transposants can be generated from a single launch pad transgenic plant carrying an *AcDs* construct for generating a more extensive population, as in rice and barley (Kolesnik et al. 2004; Upadhyaya et al. 2006; Zhao et al. 2006b).

Acknowledgments This work was supported by the USDA/NRI 2008-02284, Hatch VA-135816, the East Malling Trust for Horticulture Research, the BBSRC, and a Virginia Tech ASPIRES Grant.

References

Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Lalji S (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. Theor Appl Genet 114:359–372

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Brookes AJ (1999) The essence of SNPs. Gene 234:177–186

Brunaud V, Balzergue S, Dubreucq B, Aubourg S, Samson F, Chauvin S, Bechtold N, Cruaud C, DeRose R, Pelletier G, Lepiniec L, Caboche M, Lecharny A (2002) T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites. EMBO Rep 3:1152–1157

Bundock PC, Henry RJ (2004) Single nucleotide polymorphism, haplotype diversity and recombination in the *Isa* gene of barley. Theor Appl Genet 109:543–551

Conn HJ (1942) Validity of the genus *Alcaligenes*. J Bacteriol 44:353–360

Dafny-Yelin M, Tovkach A, Tzfira T (2009) Integration of *Agrobacterium* T-DNA in plant cells. In: Meier I (ed) Functional organization of the plant nucleus. Springer, Berlin, pp 157–186

Délye C, Boucansaud K, Pernin F, Le Corre V (2009) Variation in the gene encoding acetolactate-synthase in *Lolium* species and proactive detection of mutant, herbicide-resistant alleles. Weed Res 49:326–336

Folta KM, Davis TM (2006) Strawberry genes and genomics. Crit Rev Plant Sci 25:399–415

Hadonou AM, Sargent DJ, Wilson F, James CM, Simpson DW (2004) Development of microsatellite markers in *Fragaria*, their use in genetic diversity analysis, and their potential for genetic linkage mapping. Genome 47:429–438

Henry RJ (2008) Future prospects for plant genotyping. In: Henry RJ (ed) Plant genotyping II: SNP technology. CABI, Wallingford, pp 272–280

Hiei Y, Ohta S, Komari T, Kumashiro T (1994) Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. Plant J 6:271–282

Jing R, Johnson R, Seres A, Kiss G, Ambrose MJ, Knox MR, Ellis THN, Flavell AJ (2007) Gene-based sequence diversity analysis of field pea (*Pisum*). Genetics 177:2263–2275

Knox MR, Ellis THN (2002) Excess heterozygosity contributes to genetic map expansion in pea recombinant inbred populations. Genetics 162:861–873

Kolesnik T, Szeverenyi I, Bachmann D, Kumar CS, Jiang S, Ramamoorthy R, Cai M, Ma ZG, Sundaresan V, Ramachandran S (2004) Establishing an efficient *Ac/Ds* tagging system in rice: large-scale analysis of *Ds* flanking sequences. Plant J 37:301–314

Komori T, Nitta N (2005) Utilization of the CAPS/dCAPS method to convert rice SNPs into PCR-based markers. Breed Sci 55:93–98

Konieczny A, Ausubel FM (1993) A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. Plant J 4:403–410

Krysan PJ, Young JC, Sussman MR (1999) T-DNA as an insertional mutagen in *Arabidopsis*. Plant Cell 11:2283–2290

Lai Z, Livingstone K, Zou Y, Church SA, Knapp SJ, Andrews J, Rieseberg LH (2005) Identification and mapping of SNPs from ESTs in sunflower. Theor Appl Genet 111:1532–1544

Li Y, Rosso MG, Ülker B, Weisshaar B (2006) Analysis of T-DNA insertion site distribution patterns in *Arabidopsis thaliana* reveals special features of genes without insertions. Genomics 87:645–652

Liu Y (2005) Construction of a genetic map and location of quantitative trait loci for number of flowers per truss traits in tomato by SSR markers. J Nanjing Agric Univ 28:30–34

Liu YG, Chen Y (2007) High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. BioTechniques 43:649–656

Liu Y-G, Whittier RF (1995) Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics 25:674–681

Michaels SD, Amasino RM (1998) A robust method for detecting single-nucleotide changes as polymorphic markers by PCR. Plant J 14:381–385

Muehlbauer FJ, Rajesh PN (2008) Chickpea, a common source of protein and starch in the semi-arid tropics. In: Jorgensen RA (ed) Plant genetics and genomics: crops and models. Springer, Heidelberg, pp 171–186

Murai H, Sharma PN, Murata K, Hashimoto Z, Ketipearachi Y, Shimizu T, Takumi S, Mori N, Kawasaki S, Nakamura C (2003) Constructing linkage maps of brown planthopper resistance genes *Bph1*, *bph2*, and *Bph9* on rice chromosome 12. In: Khush GS, Brar DS, Hardy B (eds) Rice genetics collection. World Scientific Publishing, Singapore, pp 263–265

Neff MM, Neff JD, Chory J, Pepper AE (1998) dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. Plant J 14:387–392

Neff MM, Turk E, Kalishman M (2002) Web-based primer design for single nucleotide polymorphism analysis. Trends Genet 18:613–615

Olmstead JW (2008) Construction of an intra-specific sweet cherry (*Prunus avium* L.) genetic linkage map and synteny analysis with the *Prunus* reference map. Tree Genet Genom 4:897–910

Oosumi T, Gruszewski HA, Blischak LA, Baxter AJ, Wadl PA, Shuman JL, Veilleux RE, Shulaev V (2006) High-efficiency transformation of the diploid strawberry (*Fragaria vesca*) for functional genomics. Planta 223:1219–1230

Pan XK, Li Y, Stein L (2005) Site preferences of insertional mutagenesis agents in Arabidopsis. Plant Physiol 137:168–175

Potter D, Luby JJ, Harrison RE (2000) Phylogenetic relationships among species of *Fragaria* (Rosaceae) inferred from non-coding nuclear and chloroplast DNA sequences. Syst Bot 25:337–348

Robinson SJ, Tang LH, Mooney BAG, McKay SJ, Clarke WE, Links MG, Karcz S, Regan S, Wu YY, Gruber MY, Cui DJ, Yu M, Parkin IAP (2009) An archived activation tagged population of *Arabidopsis thaliana* to facilitate forward genetics approaches. BMC Plant Biol 9:101

Rousseau-Gueutin M, Lerceteau-Köhler E, Barrot L, Sargent DJ, Monfort A, Simpson D, Arús P, Guérin G, Denoyes-Rothan B (2008) Comparative genetic mapping between octoploid and diploid *Fragaria* species reveals a high level of colinearity between their genomes and the essentially disomic behavior of the cultivated octoploid strawberry. Genetics 179:2045–2060

Sargent DJ, Hadonou AM, Simpson DW (2003) Development and characterization of polymorphic microsatellite markers from *Fragaria viridis*, a wild diploid strawberry. Mol Ecol Notes 3:550–552

Sargent DJ, Davis TM, Tobutt KR, Wilkinson MJ, Battey NH, Simpson DW (2004) A genetic linkage map of microsatellite, gene-specific and morphological markers in diploid *Fragaria*. Theor Appl Genet 109:1385–1391

Sargent DJ, Clarke J, Simpson DW, Tobutt KR, Arus P, Monfort A, Vilanova S, Denoyes-Rothan B, Rousseau M, Folta KM, Bassil NV, Battey NH (2006) An enhanced microsatellite map of diploid *Fragaria*. Theor Appl Genet 112:1349–1359

Sargent DJ, Rys A, Nier S, Simpson DW, Tobutt KR (2007) The development and mapping of functional markers in *Fragaria* and their transferability and potential for mapping in other genera. Theor Appl Genet 114:373–384

Sargent DJ, Cipriani G, Vilanova S, Gil-Ariza D, Arús P, Simpson DW, Tobutt KR, Monfort A (2008) The development of a bin mapping population and the selective mapping of 103 markers in the diploid *Fragaria* reference map. Genome 51:120–127

Sargent DJ, Marchese A, Simpson DW, Howad W, Fernández-Fernández F, Monfort A, Arús P, Evans KM, Tobutt KR (2009) Development of "universal" gene-specific markers from *Malus* spp. cDNA sequences, their mapping and use in synteny studies within Rosaceae. Tree Genet Genom 5:133–145

Shuangyan C, Weizheng J, Mingyi W, Fan Z, Jie Z, Qiaojun J, Yunrong W, Feiyan L, Ping W (2003) Distribution and characterization of over 1000 T-DNA tags in rice genome. Plant J 36:105–113

Shulaev V, Korban SS, Sosinski B, Abbott AG, Aldwinckle HS, Folta KM, Iezzoni A, Main D, Arús P, Dandekar AM, Lewers K, Brown SK, Davis TM, Gardiner SE, Potter D, Veilleux RE (2008) Multiple models for Rosaceae genomics. Plant Physiol 147:985–1003

Smith EF, Townsend CO (1907) A plant tumor of bacterial origin. Science 25:671–673

Tanhuanpää P, Kalendar R, Schulman AH, Kiviharju E (2008) The first doubled haploid linkage map for cultivated oat. Genome 51:560–569

Upadhyaya NM, Zhu QH, Zhou XR, Eamens AL, Hoque MS, Ramm K, Shivakkumar R, Smith KF, Pan ST, Li SZ, Peng KF, Kim SJ, Dennis ES (2006) *Dissociation* (*Ds*) constructs, mapped *Ds* launch pads and a transiently-expressed transposase system suitable for localized insertional mutagenesis in rice. Theor Appl Genet 112:1326–1341

Van Ooijen JW, Voorrips R (2001) Joinmap 3.0: software for the calculation of genetic linkage maps. Plant Research International, Wageningen

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78

Windels P, De Buck S, Depicker A (2008) *Agrobacterium tumefaciens*-mediated transformation: Patterns of T-DNA integration into the host genome. In: Tzfira T, Citovsky V (eds) Agrobacterium: from biology to biotechnology. Springer, New York, pp 441–481

Woody ST, Austin-Phillips S, Amasino RM, Krysan PJ (2007) The WiscDsLox T-DNA collection: an Arabidopsis community resource generated by using an improved high-throughput T-DNA sequencing pipeline. J Plant Res 120:157–165

Yanagisawa T, Kiribuchi-Otobe C, Hirano H, Suzuki Y, Fujita M (2003) Detection of single nucleotide polymorphism (SNP) controlling the waxy character in wheat by using a derived cleaved amplified polymorphic sequence (dCAPS) marker. Theor Appl Genet 107:84–88

Zambryski P, Depicker A, Kruger K, Goodman HM (1982) Tumour induction by *Agrobacterium tumefaciens*: analysis of the boundaries of T-DNA. J Mol Appl Genet 1:361–370

Zhao B, Deng Q-M, Zhang Q-J, Li J-Q, Ye S-P, Liang Y-S, Peng Y, Li P (2006a) Analysis of segregation distortion of molecular markers in F_2 population of rice. Acta Genet Sin 33:449–457

Zhao T, Palotta M, Langridge P, Prasad M, Graner A, Schulze-Lefert P, Koprek T (2006b) Mapped *Ds*/T-DNA launch pads for functional genomics in barley. Plant J 47:811–826