

# Identification of a recently active *Prunus*-specific non-autonomous Mutator element with considerable genome shaping force

Júlia Halász<sup>1,†</sup>, Ossama Kodad<sup>2,†</sup> and Attila Hegedűs<sup>1,\*</sup>

<sup>1</sup>Department of Genetics and Plant Breeding, Corvinus University of Budapest, P.O. Box 53, Budapest, H-1518 Hungary, and

<sup>2</sup>Département d'Arboriculture, École Nationale d'Agriculture de Meknès, Meknès, BP S/40 Morocco

Received 12 January 2014; revised 24 April 2014; accepted 30 April 2014; published online 9 May 2014.

\*For correspondence (e-mail: hegedus.attila@uni-corvinus.hu).

<sup>†</sup>These authors contributed equally to this work.

## SUMMARY

Miniature inverted-repeat transposable elements (MITEs) are known to contribute to the evolution of plants, but only limited information is available for MITEs in the *Prunus* genome. We identified a MITE that has been named *Falling Stones*, *FaSt*. All structural features (349-bp size, 82-bp terminal inverted repeats and 9-bp target site duplications) are consistent with this MITE being a putative member of the Mutator transposase superfamily. *FaSt* showed a preferential accumulation in the short AT-rich segments of the euchromatin region of the peach genome. DNA sequencing and pollination experiments have been performed to confirm that the nested insertion of *FaSt* into the *S-haplotype-specific F-box* gene of apricot resulted in the breakdown of self-incompatibility (SI). A bioinformatics-based survey of the known *Rosaceae* and other genomes and a newly designed polymerase chain reaction (PCR) assay verified the *Prunoideae*-specific occurrence of *FaSt* elements. Phylogenetic analysis suggested a recent activity of *FaSt* in the *Prunus* genome. The occurrence of a nested insertion in the apricot genome further supports the recent activity of *FaSt* in response to abiotic stress conditions. This study reports on a presumably active non-autonomous Mutator element in *Prunus* that exhibits a major indirect genome shaping force through inducing loss-of-function mutation in the SI locus.

**Keywords:** transposition, Mutator transposon, miniature inverted-repeat transposable elements, *Prunus*, loss-of-function mutation, *S-haplotype-specific F-box*, self-compatibility, apricot.

## INTRODUCTION

Transposable elements (TE) contribute significantly to plant genomes (Bennetzen, 2000; Joly-Lopez *et al.*, 2012; Wendel *et al.*, 2012). TEs are classified into class I (retrotransposon) and class II (DNA transposon) elements. The first group transposes by a copy-and-paste mechanism using an RNA intermediate and subsequent reverse transcription, while class II elements are moving with a cut-and-paste mechanism. Both types are capable of attaining high copy numbers exhibiting major genome shaping forces in plants. Class I elements include long terminal repeat (LTR) retrotransposons, long and short interspersed nuclear elements, *Dictyostelium* intermediate repeat sequence (DIRS) and Penelope elements, most of which generate direct repeats of variable length upon insertion, called target site duplications (TSD). Diagnostic characteristics for most class II elements include terminal inverted

repeats (TIR) and TSDs. Both elements can be autonomous (encoding the enzymatic machinery required for transposition) or non-autonomous that may be mobilized *in trans* if they retain the capacity to be recognized by enzymes encoded by autonomous elements elsewhere in the genome (Kejnovsky *et al.*, 2012).

Miniature inverted-repeat transposable elements (MITEs) are non-autonomous members of DNA transposons in both plants and animals. These are characterized by small size (less than 600 bp), conserved TIRs, specific TSDs and have no coding potential of transposase enzyme (Wessler *et al.*, 1995; Feschotte *et al.*, 2002). As practically all DNA transposon superfamilies can produce non-autonomous deletion derivatives, they can be classified and named according to their structural features as members of the superfamily PIF/Harbinger, Tc1/Mariner, Mutator etc.

(Wicker *et al.*, 2007). Their high copy number might be explained by recognition of a related transposase and DNA repair through homologous recombination or moving from a post- to a pre-replicated region of chromatids during the S phase (Engels *et al.*, 1990; Kejnovsky *et al.*, 2012).

Apricot (*Prunus armeniaca* L.) belongs to the *Prunoideae* sub-family producing hermaphroditic flowers. Hence, a genetically controlled mechanism evolved in these species to prevent autogamous fruit/seed set, called self-incompatibility (SI). This mechanism is governed by the multiallelic *S*-locus having a pistil-expressed *S*-ribonuclease (*S*-*RNase*) gene and a pollen-expressed *S*-haplotype-specific *F*-box (*SFB*) gene. The intercellular reaction is expressed gametophytically and based on matching or differing alleles between the pollen tubes and the pistil. The rejection or acceptance depends on a not fully elucidated series of molecular events (Hegedűs *et al.*, 2012).

Although SI is predominant in *Prunus*, only self-compatibility (SC) forms have been identified in peach [*P. persica* (L.) Batsch.]. SC genotypes have also been identified in other *Prunus* species including cherries, almonds, plums and apricots. Most of these accessions have been intensively investigated and mutations have been described as responsible for either the loss-of-pistil function or pollen-*S*. Most of the natural mutations leading to SC in *Prunus* were clarified to be associated with TEs. *P. cerasus* *S*<sub>1</sub>' (insertion of a 615-bp *Ds*-like element into *SFB*), Hauck *et al.* (2006); *S*<sub>6m2</sub> (insertion of 2.6-kbp Mu element upstream of *S*-*RNase*), Yamane *et al.* (2003); *P. persica* *S*<sub>1</sub> (155-bp direct repeat in the middle of the HVb region) and *S*<sub>2</sub> (insertion of a 5-bp direct repeat between the V1 and V2 regions that resembles the footprint of a transposon and causes a frame-shift mutation), Tao *et al.* (2007). In Japanese apricot, a 6.8-kbp insertion disrupted the coding region of the *SFB*<sub>F</sub>-allele leading to truncated proteins and hence SC. The insertion was flanked by direct repeats and contained LTRs at the ends of the insertion similar to retrotransposons (Ushijima *et al.*, 2004).

In apricot, SC is predominant in European cultivars, while those of Chinese or Central Asian origins are self-incompatible. SC in apricot is due to a small insertion (358 bp) in the *SFB*<sub>C</sub> allele resulting in a premature stop codon before the hypervariable regions of the gene (Vilanova *et al.*, 2006). BLASTN analysis of the inserted element found similar sequences in other *Prunus* species and it contained inverted repeats at the ends of the inserted sequence, which are similar to the TIRs of transposable elements. Halász *et al.* (2007) identified the original, non-mutant version of the *S*<sub>C</sub>-allele and named it *S*<sub>8</sub>. Later, it was shown, that this mutation might have emerged in the south east of Turkey and disseminated through Mediterranean Europe and North Africa (Halász *et al.*, 2013; Kodad *et al.*, 2013b; Lachkar *et al.*, 2013). In Morocco, most

apricots grown in the oasis agro-ecosystem were confirmed to be self-compatible and carry the *S*<sub>C</sub>-haplotype (Kodad *et al.*, 2013a,b).

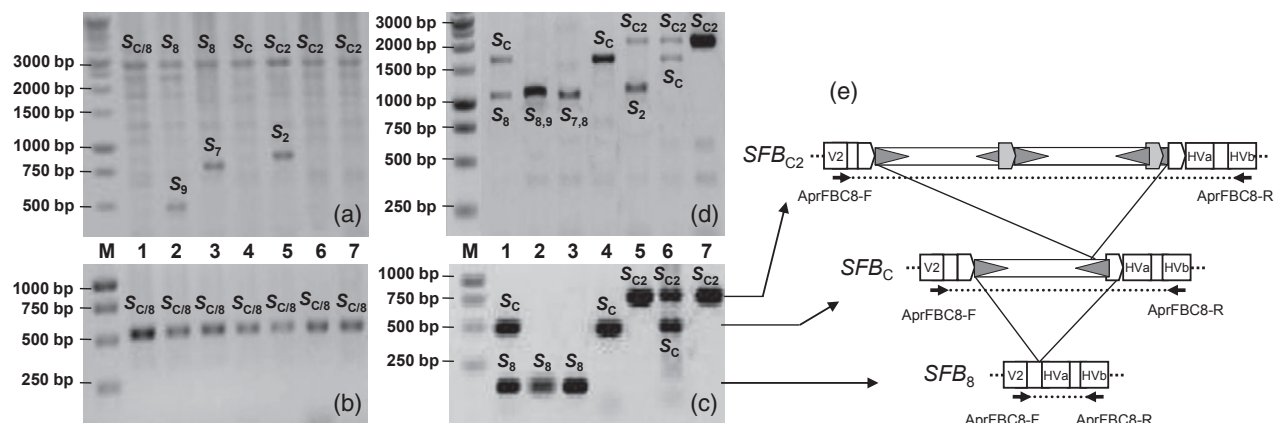
Despite their high copy numbers in genomes, only a limited number of active MITEs have been identified so far in species such as rice, maize and peanut (Zhang *et al.*, 2001; Kikuchi *et al.*, 2003; Shirasawa *et al.*, 2012). In the present study, we describe the identification and characterization of a non-autonomous Mutator element in *Prunus* that we have named *Falling Stones* (*FaSt*). We declare close association of *FaSt* with the coding part of the *Prunus* genome and confirm its role in the breakdown of SI in apricot. The detection of a nested pattern of *FaSt* non-autonomous Mutator elements in apricot *SFB* gene and microsatellite marker analysis were used to confirm recent activity of the element. A bioinformatics-based survey of the known *Rosaceae* and other genomes was carried out and a PCR assay was designed to verify the *Prunoideae*-specific occurrence of *FaSt* elements.

## RESULTS

### Identification and structural analysis of a MITE that induces self-compatibility in apricot

Self-compatibility was found to be predominant in apricots of Tighnit and Armed oases in the Draa Valley, southern Morocco (Kodad *et al.*, 2013a), and *S*-genotyping confirmed that this phenotype is attributed to the presence of the *S*<sub>C</sub>-allele (Kodad *et al.*, 2013b). While analyzing several more accessions from this region, another three selections (Tighnit-10, Armed-3 and Armed-7) were identified that carry the *S*<sub>C</sub>-*RNase* allele (Figure 1a,b). The sizes of the amplified fragments of the *S*-*RNase* first intron region were as follows 332 bp/355 bp for Tighnit-10 and 355 bp/355 bp for both Armed-3 and Armed-7. However, the AprFBC8-F and AprFBC8-R primer pair amplified a fragment of unexpected length, approx. 850 bp in each of the three accessions, clearly different from the 500 bp and 150 bp that are characteristic for the *SFB*<sub>C</sub> and non-mutated *SFB*-alleles (e.g., *SFB*<sub>2</sub> or *SFB*<sub>8</sub>), respectively (Figure 1c). It was further supported by another PCR assay amplifying a larger region of the *Prunus* *SFB*s (Figure 1d). This allele was labeled *SFB*<sub>C2</sub> and the *S*-genotypes of Tighnit-10, Armed-7 and Armed-3 were given as *S*<sub>2</sub>*S*<sub>C2</sub>, *S*<sub>C</sub>*S*<sub>C2</sub> and *S*<sub>C2</sub>*S*<sub>C2</sub>, respectively.

The SC of each accession was checked by following pollen tube growth and fruit set after self-pollination. Data confirmed that self-pollen tubes were able to enter the ovary and resulted in fruit set higher than 20% in the cases of Tighnit-10, Armed-3 and Armed-7. In contrast, after self-pollination of the self-incompatible Rich-1 accession, pollen tubes did not reach the ovary and failed to set any fruit (Table 1).



**Figure 1** PCR analysis of the *self-incompatibility ribonuclease* and *S-haplotype-specific F-box* genes in seven apricot cultivars.

Samples are (from left): 1-kb + ladder (M), Gönci magyarkajszai (*S<sub>C</sub>S<sub>9</sub>*) (1), Ceglédi óriás (*S<sub>9</sub>S<sub>9</sub>*) (2), Ait-Lahcen-3 (*S<sub>7</sub>S<sub>8</sub>*) (3), Armed-22 (*S<sub>C</sub>S<sub>C</sub>*) (4), Tighnit-10 (*S<sub>2</sub>S<sub>C2</sub>*) (5), Armed-7 (*S<sub>C</sub>S<sub>C2</sub>*) (6) and Armed-3 (*S<sub>C2</sub>S<sub>C2</sub>*) (7).

(a) Primers EM-PC2consFD and EM-PC3consRD were used for the detection of *S-RNase* second intron.

(b) Primers PaConsl F and AprSC8R were used for specific detection of *S<sub>C</sub>/S<sub>RNase</sub>* allele.

(c) A codominant primer pair, AprSFB-F1 and AprSFB-R, was used to amplify all *SFB* alleles.

(d) AprFBC8-F and AprFBC8-R primers were used to detect the *SFB<sub>C2</sub>* alleles.

(e) Schematic representation of the detected insertion mutation within the apricot *S-haplotype-specific F-box* gene. Narrower white squares with gray triangles pointing to each other indicate a non-autonomous class II transposable element (named *Falling Stones*, *FaSt*). White and gray pentagons facing the same direction refer to target site duplications generated from the *SFB* gene and *FaSt* sequences, respectively. Positions of the annealing site of AprFBC8-F and AprFBC8-R primers and relative length of the amplified fragments are shown by arrows and dotted lines. DNA regions were not drawn to scale.

**Table 1** Mean number of pollen tubes found at base of style and into the ovary, and fruit set percentage after self-pollination of Moroccan apricot genotypes

Genotype	Mean number of pollen tubes at the base of the style	Mean number of pollen tubes into the ovary	Fruit set (%)
Rich-1*	1.07 ± 2.31 <sup>d</sup>	0.00 ± 0.00 <sup>d</sup>	0.00 ± 0.00 <sup>c</sup>
Tighnit-10	8.47 ± 1.73 <sup>a</sup>	2.53 ± 1.16 <sup>b</sup>	23.73 ± 9.8 <sup>b</sup>
Armed-7	5.47 ± 2.43 <sup>c</sup>	2.09 ± 1.06 <sup>c</sup>	25.8 ± 2.4 <sup>b</sup>
Armed-3	7.28 ± 2.13 <sup>b</sup>	3.33 ± 1.28 <sup>a</sup>	30.51 ± 12.4 <sup>a</sup>

Mean values ( $n = 20$  pistils) ± standard error (SE) followed by different lower-case letters within a column are significantly different at  $P \leq 0.05$  according to Duncan's multiple range test.

\*Genotypes used as SI controls (Kodad *et al.*, 2013a,b).

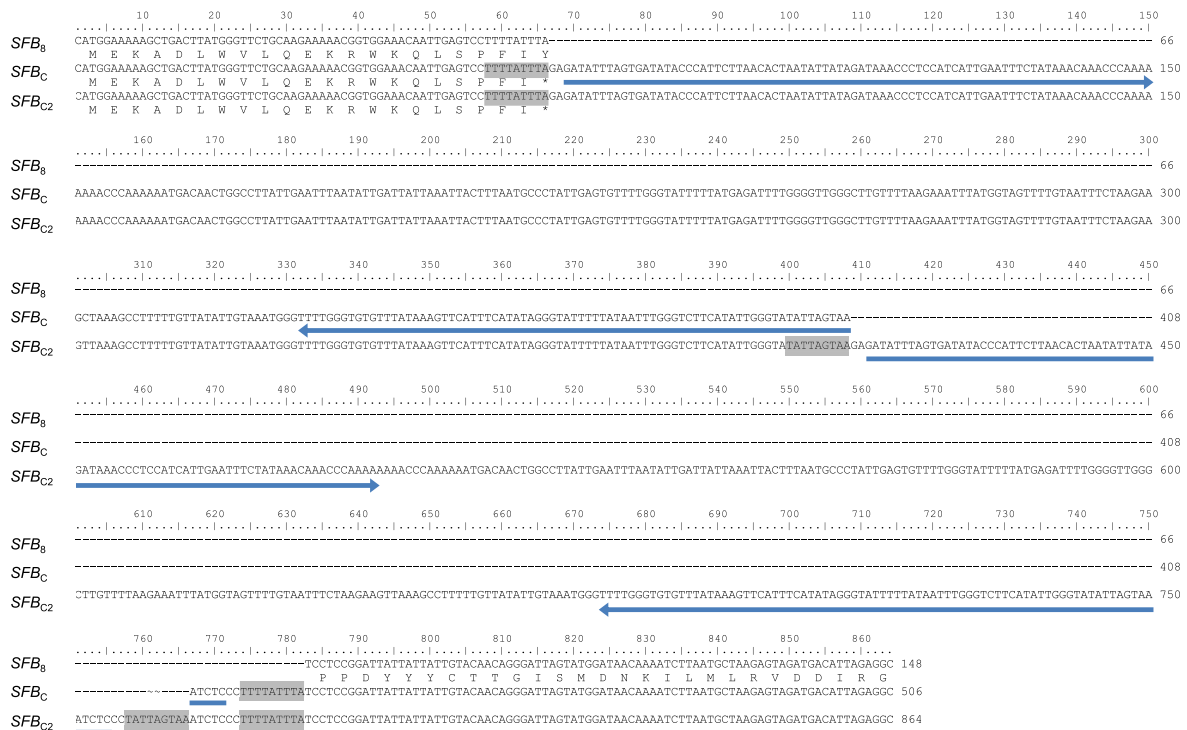
The 850-bp fragment was cloned and sequenced. The exact length of the insert was determined to be 864 bp. A 716-bp insertion disrupts the open reading frame in the *SFB<sub>C2</sub>* allele at an identical position to that of the 358-bp insertion in the *SFB<sub>C</sub>* allele (Figure 2). Moreover, the first 342 bp of the inserted sequence was also identical to that described from the *SFB<sub>C</sub>* allele. However, the most interesting feature of the 864-bp fragment was that the sequence interrupting the *SFB* open reading frame (ORF) contained two copies of the 358-bp element that has been described originally from *SFB<sub>C</sub>* (Vilanova *et al.*, 2006). The two copies occurred in a nested pattern with one element of 358-bp length inserted between nucleotides 342 and 343 of the other (Figures 1e and 2). Comparing the single elements of

the insertions revealed only three nucleotide substitutions at the 3' end of the sequences.

Features characteristic of DNA transposon insertion have been observed. Terminal inverted repeats of 82 bp flanked the two ends of a 181-bp loop. The nucleotide identity between the arms was approximately 75% (Figure S1). A predominant AT content was clear in the inserted sequence (Table 2). The doubled arrangement allowed us to reliably identify TSD generated upon transposon insertion. The first insertion (originally described from the *SFB<sub>C</sub>* allele in 'Currot' and 'Pannónia' apricot cultivars by Vilanova *et al.* (2006), and Halász *et al.* (2007), respectively) generated a 9-bp TSD of TTTTATTTA, while nested insertion of the same element is flanked by a TSD of TATTAGTAA sequence. This analysis clarified that the length of the inserted sequence without TSD was 349 bp, and differences between the original and nested copies occurred only in their TSDs while the inserted sequence elements themselves were identical. All these structural features are consistent with this element being a miniature inverted-repeat transposable element that we have named *Falling Stones*, *FaSt*.

#### Bioinformatic searches of *FaSt* in other genomes

A BLASTN analysis was carried out using the 349-bp inserted element as a query sequence and several homologs were found in the database (Table S1). *FaSt* showed significant homology with: (i) *Prunus persica* BAC clones 28F08 ( $E$ -value =  $8 \times 10^{-137}$ ) and several regions of JGI-BIXB-13B7 ( $E \leq 7 \times 10^{-68}$ ), (ii) a part of the intron in the



**Figure 2.** Alignment of partial nucleotide sequences of apricot *SFB*<sub>8</sub>, *SFB*<sub>c</sub> and *SFB*<sub>c2</sub> alleles. Target site duplications are shown in gray shading; terminal inverted repeats are marked by arrows.

**Table 2** Basic structural features of the *P. armeniaca S*-haplotype-specific *F*-box inserted *Falling Stones* element

	Left arm position	Right arm position	Arm length (bp)	Loop length (bp)	Identity (%)	AT content (%)
<i>FaSt</i>	3–84	266–347	82	181	74.69	73.17

*P. avium R2R3-MYB* transcription factor (*MYBF1*) gene ( $E = 2 \times 10^{-78}$ ), (iii) a part of the second intron of the *P. cerasifera S*<sub>6</sub> ( $E = 10^{-75}$ ) and *S*<sub>14-ribonuclease</sub> ( $E = 7 \times 10^{-73}$ ) alleles. Where flanking sequences were available, the presence of a characteristic 9-bp TSD was also evident.

#### *FaSt* copy number and distribution in the peach genome

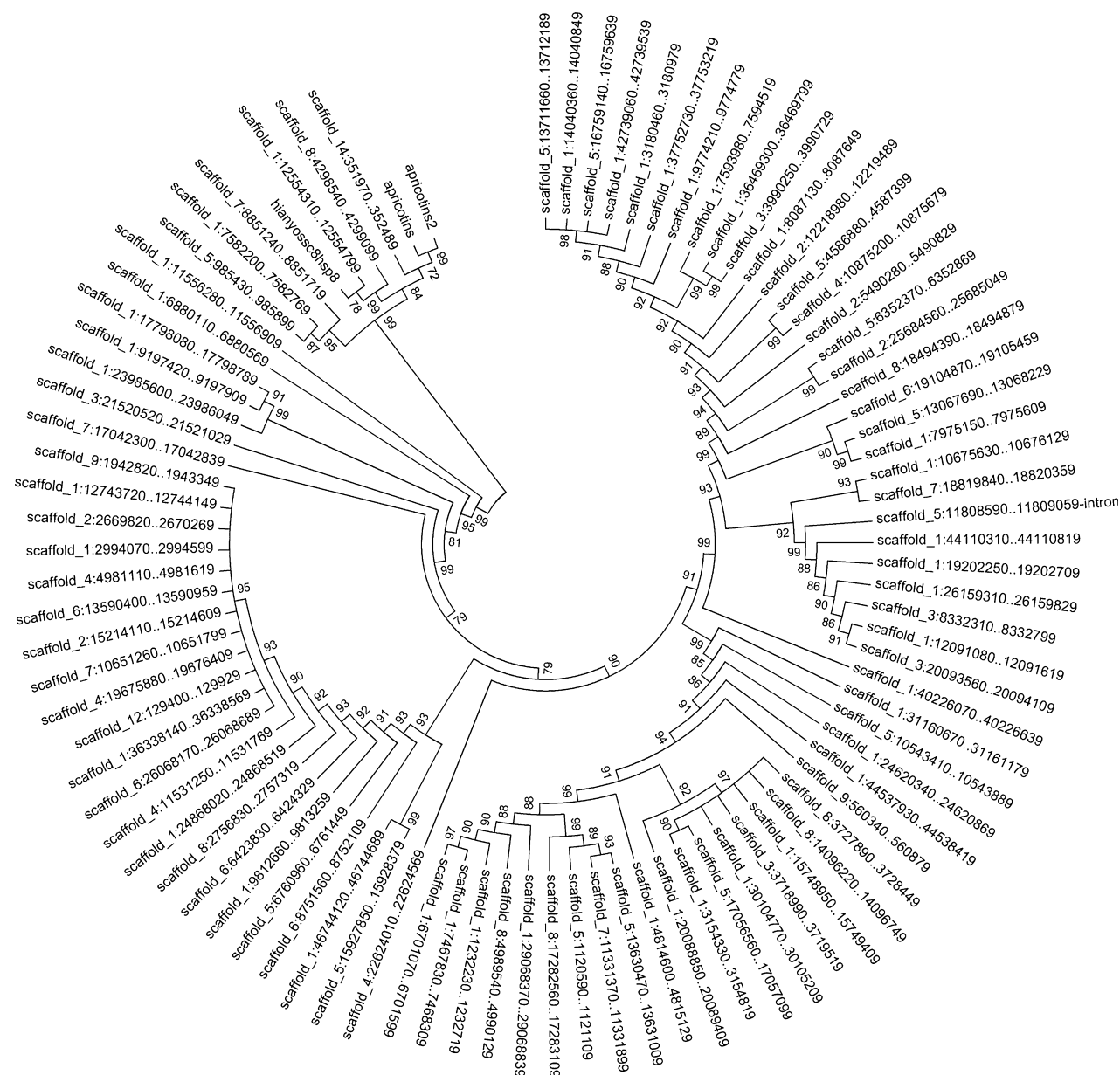
A BLAT (BLAST-like Alignment Tool) search was conducted on the peach genome using the 349-bp apricot *FaSt* as a query sequence. We detected 121 full copies and 60 additional copies having a fragment of more than 50% of the original length during the search (Table S2). All matches were characterized by  $E$ -values  $\leq 10^{-102}$ . There were several other high-scoring segment pairs throughout the peach genome but were neglected as having only a small fragment of the query sequence without TSD motifs. Sequences were downloaded from the peach genome database and compared with the apricot *FaSt* sequence. Their

identities ranged between 0.877 (scaffold\_1:11556280) and 0.925 (scaffold\_6:19104870, scaffold\_7:8851240 and scaffold\_8:18494390; Table S3). A majority-rule consensus tree of maximum likelihood analysis revealed no clustering according to chromosomes or chromosomal positions (Figure 3). One of the sequence variations occurred in 10 copies occurring on scaffolds 1, 2, 4, 6, 7 and 9; and another one in five copies on scaffolds 1, 3 and 8. Four copies of a variant were found on scaffolds 1 and 5; two identical sequences occurred on scaffold 1 and another pair of identical copies on scaffolds 5 and 7. In addition, those five bunches of identical sequences differed from each other only in some of the seven variable nucleotide positions (Figure S2 and Table S3).

*FaSt* elements were detected on each of the eight pseudochromosomes of peach (Figure 4). The most elements accumulated on chromosome 1 and the least occurred on chromosome 8 (Table S2). Interestingly, *FaSt* elements were less abundant in the regions of each putative centromere. *FaSt* elements also occurred in the subtelomeric region of chromosome 6, close to the *S*-locus (Figure 4).

#### Experimental verification of *FaSt* as a *Prunus*-specific transposable element

Flanking sequences of 90 full length *FaSt* insertions in the peach genome were analyzed for the presence of TSD. In



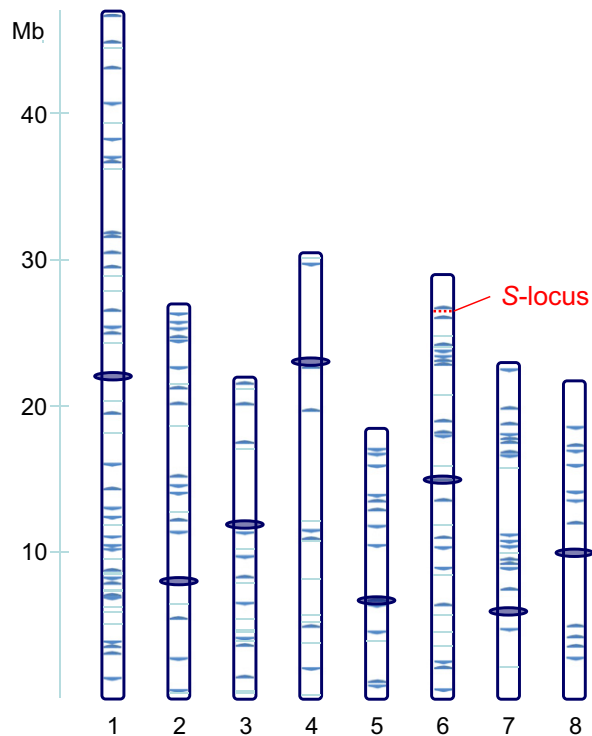
**Figure 3.** Majority-rule consensus tree of *Prunus* Falling Stones transposable elements using the maximum likelihood analysis and 1200 bootstrap simulations. Apricotins and apricotins2 refer to the *FaSt* elements identified in the apricot *SFB<sub>C</sub>* and *SFB<sub>C2</sub>* alleles.

all cases, 9-bp long TSDs flanked the inserted 349-bp elements. Sequences were downloaded and sites of insertion were analyzed together with those of the apricot *FaSt* elements. The 5' and 3' ends of the inserted sequence are largely conserved, while TSDs are predominantly AT content motifs (Figure 5).

All 41 plant genomes deposited in Phytozome v9.1 were screened for the presence of *FaSt* elements using BLAT searches but no match was found. The lack of homology was also evident in the cases of *Fragaria vesca* and *Malus × domestica* genomes and confirmed the results of BLAST analysis detecting homologous sequences only in

*Prunus* species, peach, cherry plum and sweet cherry (Table S1). Hence, conserved TIR regions of the apricot and peach *FaSt* elements were used to design PCR primers to amplify a 347-bp fragment of the *FaSt*. Massive amplification occurred in all 18 *Prunus* accessions encompassing apricot, peach, almond, cherry, diploid and hexaploid plums as well as wild species (Figure 6). However, none of the apple and quince belonging to the *Maloideae* sub-family of the *Rosaceae*, *Hippophae rhamnoides* (belonging to the *Elaeagnaceae* family of the *Rosales*) and *Cornus mas* (belonging to the *Cornales*, an order in the asterid clade) amplified the specific fragment.





**Figure 4.** Distribution of the identified *Prunus*-specific *Falling Stones* elements throughout the peach genome. Arrows indicate the orientation of the full length sequences and lines show fragmented copies. The approximate position of each centromere is indicated by an oval (not drawn to scale).

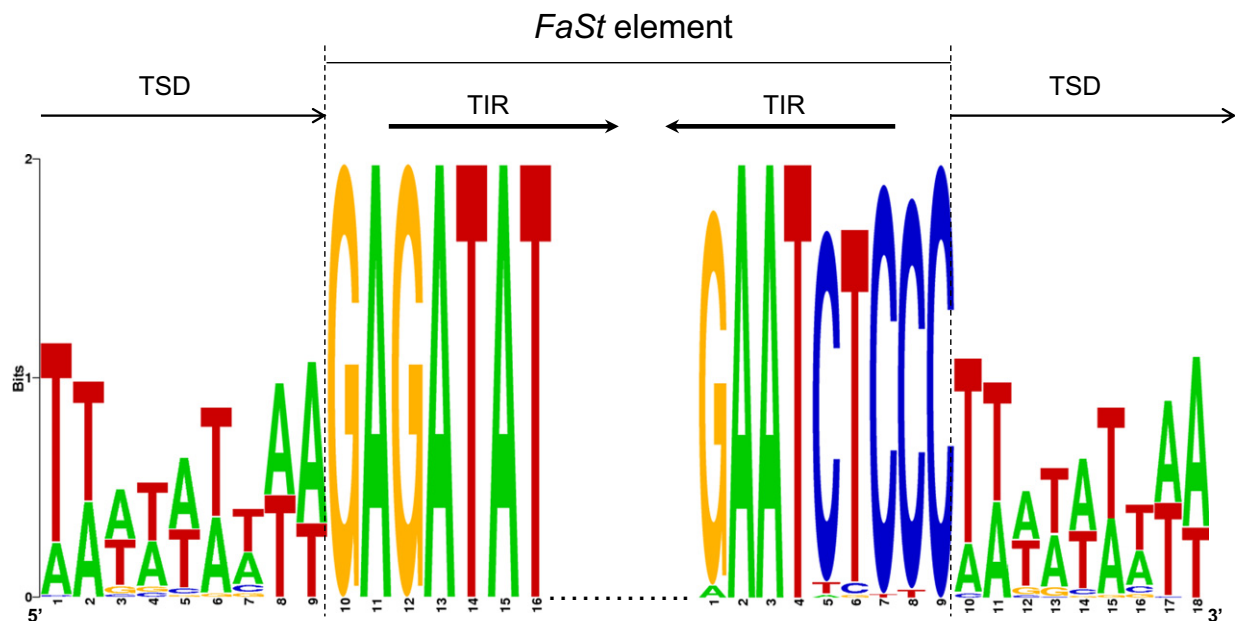
### Simple-sequence repeat (SSR) analysis of apricots to explore the origin of *SFB*<sub>C2</sub>-allele

Identical copies of the double inserted elements indicated that the second transposition event might have occurred relatively recently. We have performed an SSR analysis using 10 primer pairs to characterize genetic relationships among the three Moroccan apricot accessions carrying the *SFB*<sub>C2</sub>-allele with double insertion of the identified *FaSt* element. The self-incompatible cultivar 'Zard', a Central Asian genotype formed an outgroup, and the self-compatible Hungarian cultivar 'Ceglédi arany' was also far from all Moroccan accessions (Figure 7). Within Moroccan apricots, those originating in Gulmima formed a common cluster, while apricots originating in Armed, Tighnit, Ait-Lahcen, Skoura and the Atlas Mountain regions clustered intermixed. However, three genotypes (Armed-3, Armed-7 and Tighnit-10) carrying the *SFB*<sub>C2</sub>-allele formed a separate group and showed almost identical SSR profiles indicating their close genetic relationship.

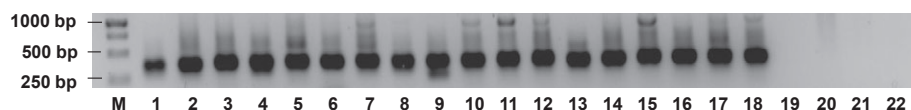
### DISCUSSION

#### Identification, structural and functional characterization of the *Prunus*-specific non-autonomous *Mutator* element, *Falling Stones* (*FaSt*)

Self-compatibility of apricot has been described as a consequence of a 358-bp inserted element interrupting the ORF of the *SFB* gene (Vilanova *et al.*, 2006). Furthermore,

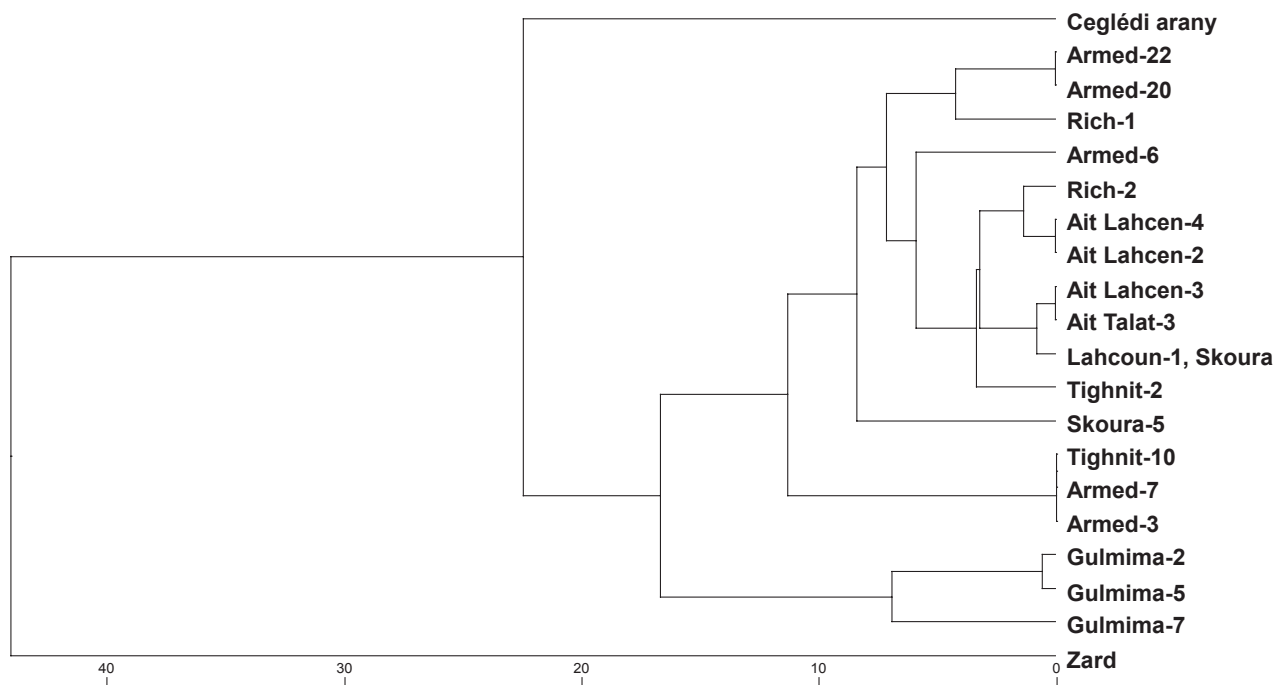


**Figure 5.** Sequence logo of the target site duplications at the insertion site of *Falling Stones* elements. Characteristic structural elements indicated include target site duplications (TSD), terminal inverted repeats (TIR) and edges of the inserted *FaSt* element.



**Figure 6.** Genomic PCR analysis (in negative) using the *FaSt*-specific primers.

Altogether, 18 different *Prunus* species and four species outside the *Prunus* genus were analyzed. M: 1 kb + DNA ladder, 1: *P. armeniaca* Armed-3 genotype; 2: *P. dulcis* 'Tétényi bőtermő'; 3: *P. persica* 'Babygold 7'; 4: *P. avium* 'Germersdorfi óriás'; 5: *P. cerasus* 'Érdi jubileum'; 6: *P. salicina* 'Santa rosa'; 7: *P. domestica* 'Tophit'; 8: *P. arabica*; 9: *P. tenella*; 10: *P. webbii*; 11: *P. salicina* × *P. armeniaca* hybrid; 12: *P. kansuensis*; 13: *P. kurdica*; 14: *P. sogdiana*; 15: *P. mume*; 16: *P. ferganensis*; 17: *P. cerasifera*; 18: *P. spinosa*; 19: *Malus* × *domestica* 'Jonagold'; 20: *Hippophae rhamnoides* 'Frugana'; 21: *Cydonia oblonga* 'Aromate'; and 22: *Cornus mas* 'Császlói'.



**Figure 7.** Dendrogram of the Moroccan and control apricot accessions.

An unweighted pair group method with arithmetic averages analysis was carried out based on the similarity matrix obtained using Nei's genetic distance after amplification with 10 pairs of microsatellite primers.

inverted repeats were shown at the ends of the inserted sequence that are similar to the TIRs of transposable elements. Our analysis indicated the presence of 82-bp TIRs at the ends of the inserted sequence. However, only the identification of *SFB*<sub>C2</sub>-allele containing two nested copies of the 358-bp sequence allowed the description of its basic structure. The first insertion is seemingly flanked by an 11-bp target site duplication of CCTTTTATTTA; however, the nesting insertion is only flanked by a 9-bp TSD. It was important to realize that the nesting element ends in a sequence of TCCC and hence the original insertion is also flanked by a 9-bp TSD (TTTTATTTA) and cytosines are not parts of the TSD. The correct identification of a 9-bp TSD places *FaSt* in the Mutator superfamily (Feschotte *et al.*, 2002).

In rice, among the 6600 identified MITEs, more than 10% were present as multimers (adjacent or nested insertions),

with *Tourist* elements having the highest self-insertion frequency (Jiang and Wessler, 2001). In addition, the fact that all 340 multimers in the rice genome were unique indicates that the multimers are not capable of further amplification. The situation is similar with the nested *FaSt* elements in apricot *SFB*<sub>C2</sub> since the second copy interrupts the TIR region at the 3' end of the first element. Hence, this TIR might not be recognized by the corresponding transposase enzyme to be cut and moved. If these elements are currently moving in the *Prunus* genome, the nested insertion will increase the stability of the self-compatible phenotype of apricot.

The accumulation of *FaSt* elements was shown to be associated with the gene-rich chromosomal regions of the peach genome that have been described by Verde *et al.* (2013). Similarly, the abundance of MITEs and of DNA transposons in general was clarified in the euchromatin

regions of genomes like *Arabidopsis* and rice (Feschotte *et al.*, 2002), *Sorghum* (Paterson *et al.*, 2009), *Brachypodium* (International Brachypodium Initiative, 2010) and barley (International Barley Genome Sequencing Consortium, 2012). The most important feature of the targeting mechanism might be the predominant AT content at the insertion site. Our analysis indicated that in the peach genome, approx. 30% of the *FaSt* elements were flanked by TSDs containing a C or G. This might be consistent with an insertion bias toward introns in the gene-rich region, as introns are AT-rich sequences in plants (Goodall and Filipowicz, 1989). Interestingly, we have found a *FaSt* copy in the second intron of the *P. cerasifera* *S-RNase* alleles, *S*<sub>6</sub> and *S*<sub>14</sub>. The abundance of MITEs in the euchromatin region may be explained by a more open conformation of the transcriptionally active DNA that is more accessible for transposases (Civán *et al.*, 2011). However, TSD with a predominant AT content, seems to outbalance this effect and target the transposition outside the protein coding region of genes. Very interestingly, nested insertions of DNA transposons were found to be the most frequent in exons when analyzing 802 TEs downloaded from eukaryotic genomes (Gao *et al.*, 2012). This might be associated with currently unknown mechanisms, but might also be explained by probabilistic principles. *FaSt* itself contains large AT-rich segments and hence provides the opportunity for nested insertion in a differing region of exon sequences. In addition, such events are free from the negative effect of purifying selection since the gene had been non-functional before insertion.

In apricot, a 358-bp insertion into the *S-haplotype-specific F-box* gene resulted in the breakdown of self-incompatibility (Vilanova *et al.*, 2006) as *SFB* is responsible for pollen function of the self/non-self recognition between pistil and pollen (Entani *et al.*, 2003; Romero *et al.*, 2004). In *SFB<sub>C</sub>*, loss-of-function (SC) was shown to be the consequence of a premature stop codon leading to the expression of a truncated protein (Vilanova *et al.*, 2006). The insertion preference for AT-rich sequences may also explain the nonsense mutation caused by the *FaSt* insertion into an apricot *SFB*. The insertion interrupted a Tyr codon (TAT) after its second nucleotide position. As the sequence of the inserted element starts with a guanine, a premature stop codon (TAG) was formed upon insertion into the ORF. *SFB<sub>C2</sub>* was presumed to be non-functional because the stop codon introduced by the original insertion is also present in this allele. Pollen tube growth assays and fruit set studies indeed confirmed the self-compatible phenotype of Armed-3 (*S<sub>C2</sub>S<sub>C2</sub>*). In addition, the homozygous *S*-genotype itself provides evidence of *SFB<sub>C2</sub>* being non-functional, as homozygosity of functional *S*-alleles cannot occur under the strict genetic control of SI (de Nettancourt, 2001).

### Evolutionary aspects of the *Prunus*-specific non-autonomous Mutator element, *Falling Stone*, *FaSt*

Among the 34 tested accessions we found only three apricot trees growing in Tighnit and Armed oases that carried the *SFB<sub>C2</sub>*-allele. Two accessions were heterozygous and one accession was homozygous for this allele. As *SFB<sub>C2</sub>* is non-functional and codes for SC this allele and *SFB<sub>C</sub>* must have had an equal probability of fixation over time. However, *SFB<sub>C</sub>* was the most frequent allele in these oases with a 95% frequency (Kodad *et al.*, 2013b). This discrepancy can only be explained by a more recent emergence of the *SFB<sub>C2</sub>* than that of *SFB<sub>C</sub>*. The distance between the two oases is approximately 4 km, and exchange of seed and propagation material between the places is very intensive. This was also confirmed by the close genetic relationship of the three accessions using microsatellite analysis. Hence, our results support a single and recent formation of the *SFB<sub>C2</sub>*-allele.

To answer the question why could we find signatures of recent activity of the *FaSt* elements only in Morocco, we have to consider some facts. First, in Tighnit and Armed, apricots have been intensively propagated, trees are younger and might have had more generations over time compared to other regions with less intensive propagation. This increases the chance of fixed mutations. In addition, apricots are predominantly self-compatible and hence they are characterized by small effective population size. Small population size will purge transposable element insertions less efficiently (Kejnovsky *et al.*, 2012). Another explanation might be that DNA transposons have been shown to specifically respond to various stress conditions (Casacuberta and González, 2013). For example, a MITE, *mPing*, in rice was able to confer a stress-inducible state to a set of nearby genes and acted as an enhancer element (Naito *et al.*, 2009). Transposon activity itself can be regulated by stress conditions through genetic or epigenetic regulation (Lisch, 2009). Dynamic DNA methylation changes in response to biotic stress (Downen *et al.*, 2014) and MITEs in the *Solanaceae* have been recently shown to generate small RNAs that have fundamental roles in eukaryotic gene regulation (Kuang *et al.*, 2009). This might also indicate a stress response in an arid and hot region like a Moroccan oasis.

Besides putative effects of *FaSt* amplification on the *Prunus* genome, its predominant insertion within the euchromatin regions resulted in a dramatic change in the sexual reproductive strategy of apricot. As most stone fruits are self-incompatible, fixation of such a mutation had been surely dependent on the avoidance of inbreeding effects. A shift from SI to SC might have brought about a homozygous state and homozygous regions had to be free from purifying selection (Halász *et al.*, 2013). However, the spread of the consequent self-compatible phenotype of



apricot produced an indirect genome shaping force leading to a serious loss of genetic variability in Eastern Europe and the Mediterranean region (Halász *et al.*, 2007; Bourguiba *et al.*, 2012).

We have mapped *FaSt* elements on the *P. persica* pseudochromosomes and found several copies in the subtelomeric region of chromosome 6 where the *S*-locus resides. We also found a complete and partial element within the intron of two *P. cerasifera* *S-RNase* alleles (Table S1). Since none of the 13 other *P. cerasifera* *S-RNase* alleles available in the database contained this element in their second intron, it supports a more recent insertion of *FaSt* elements compared to the time of *S* allele divergence and diversification. All these facts indicate that insertion of *FaSt* occurred relatively frequently around the *S*-locus of *Prunus*.

The identified apricot *FaSt* elements were clearly different from most of the peach *FaSt* sequences as shown by the high bootstrap value of their separate clade (Figure 3), but they had a relatively closer relationship with a subset of peach *FaSt* elements found on different chromosomes. Some peach *FaSt* elements of one chromosome were clustered with elements of another chromosome. Such was the case of sequences on scaffolds 1, 5, 7 and 8, and they were almost identical in DNA sequence (Figure S2). This pattern, consistent with the rapid evolution suggested by the sequence divergence of *FaSt* elements in different species like peach and apricot further supports recent activity of *FaSt*. Their relatively late emergence is also confirmed by the fact that orthologous regions in grape and poplar do not contain *FaSt* elements.

*FaSt* elements show considerable sequence variations in the peach genome and hence the identical nucleotide sequence of the nested elements in the apricot *SFB<sub>C2</sub>* allele might also be explained by a recent within-locus integration of *FaSt*. In an *S<sub>C</sub>S<sub>C</sub>* homozygote, the excision of the *FaSt* from one of the homologous chromosomes and its subsequent integration into the other might have occurred. Whether it happened or not, it is interesting to realize that this mutation has emerged in a region where most trees share the *S<sub>C</sub>S<sub>C</sub>* genotype which provides an opportunity for within-locus integration.

We propose naming the identified non-autonomous Mutator element as *Falling Stones*, both referring to its exclusive presence in the *Prunoideae* sub-family comprising stone fruits and their presumable current activity and an avalanche-like ongoing spread throughout the *Prunus* genome. Based on database searches and experimental data we confirmed its role in altering gene function. The breakdown of the SI barrier was of crucial importance during the evolution of apricot and since *FaSt*s accumulate predominantly in the gene-rich regions of the *Prunus* genome, further effects are likely to be identified in the future.

## EXPERIMENTAL PROCEDURES

### Plant material

Three apricot genotypes (Tighnit-10, Armed-3 and Armed-7) collected from Tighnit and Armed oases in the Draa Valley, Southern Morocco, were analysed in this study and compared with some previously tested genotypes (Kodad *et al.*, 2013b) from these locations. For microsatellite analysis, numbers of Moroccan accessions included in the study came from the following locations: Tighnit (1), Skoura (2), Atlas Mountain (2), Armed (3), Gulmima (3) and Kelaat M'Gouna (4). Some previously *S*-genotyped Hungarian cultivars ('Ceglédi óriás', *S<sub>8</sub>S<sub>9</sub>*, and 'Gönci magyarkajsi', *S<sub>C</sub>S<sub>8</sub>*) were used as controls (Halász *et al.*, 2005, 2007). For genomic PCR to analyse *Prunus*-specific occurrence of the *FaSt* sequences, six cultivated (*P. dulcis* 'Tétényi bőtermő', *P. persica* 'Babygold 7', *P. avium* 'Germersdorfi óriás', *P. cerasus* 'Érdi jubileum', *P. salicina* 'Santa rosa' and *P. domestica* 'Tophit') and 11 wild (*P. arabica*, *P. tenella*, *P. webbii*, a *P. salicina* × *P. armeniaca* hybrid, *P. kansuensis*, *P. kurdica*, *P. sogdiana*, *P. mume*, *P. ferganensis*, *P. cerasifera*, *P. spinosa*) *Prunus* species as well as *Malus* × *domestica* 'Jonagold', *Cydonia oblonga* 'Aromate', *Hippophae rhamnoides* 'Frugana' and *Cornus mas* 'Császlói' were used as negative controls.

### Self-compatibility evaluation

To assess the SC phenotype of the studied genotypes two tests were carried out: observation of pollen tube growth in the pistil and fruit set after self-pollination in the field. For pollen tube growth assessment, 100 flower buds at phenological stage 'D' (Baggiolini, 1952) were emasculated and self-pollinated 2 days later on a branch of the studied trees. In total, 20 pistils were collected 4 days after pollination and fixed in 1:1:18 (v/v/v) formaldehyde acetic acid: 70% (v/v) ethanol (FAA) in the laboratory. Prior to examination, FAA was removed from the pistils through rinsing in water, then autoclaved at 120°C in a 5% (w/v) solution of Na<sub>2</sub>SO<sub>3</sub> for 12 min at 1.2 kg cm<sup>-2</sup> pressure. Samples were maintained at 2–4°C until measurements were made. The pistils were prepared according to Socias i Company (1979), and the tube growth was assessed by observation in an Ortholux II microscope (Ernst Leitz GmbH, Wetzlar, Germany) with UV illumination provided by an Osram HBO 200 W/4 mercury lamp. In the present study, the number of pollen tubes in the ovary was assessed in order to evaluate the SC of each genotype.

Concerning the field test, several branches around each tree were selected and assigned at random to self-pollination to measure fruit set percentage. Flowers at phenological stage 'D' (Baggiolini, 1952) were emasculated and self-pollinated. Fruits were counted at the end of April and fruit set percentages were determined as the number of fruit set in relation to the number of pollinated pistils.

### DNA extraction

Genomic DNA was extracted from fully expanded young leaves using a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). The quantity and quality of DNA were analysed by NanoDrop™ ND-1000 spectrophotometer (Bio-Science, Budapest, Hungary).

### Genomic PCR with *S-RNase*, *SFB*- and *FaSt*-specific primers

PCR was conducted according to Sutherland *et al.* (2004) using the degenerate primers EM-PC2consFD and EM-PC3consRD for

amplification of the second intron region of the *S-RNase* gene. To amplify the first intron, fluorescently labeled (JOE) forward primer SRC-F (Romero *et al.*, 2004) was used in combination with reverse primer SRC-R (Vilanova *et al.*, 2005). For identification of *S<sub>C</sub>-RNase*, an allele-specific reverse primer, AprSC8R (Halász *et al.*, 2010) was used in combination with PaConsl F (Sonneveld *et al.*, 2003). A degenerate primer pair, AprSFB-F1 and AprSFB-R was used as a codominant marker to detect the *SFB<sub>C</sub>* allele as described by Halász *et al.* (2007). *SFB<sub>C</sub>/SFB<sub>B</sub>*-specific primers, AprFBC8-F and AprFBC8-R were used to discriminate between *SFB<sub>C</sub>* and *SFB<sub>B</sub>* alleles (Halász *et al.*, 2010). We used the program described for the primers (Sutherland *et al.*, 2004; Vilanova *et al.*, 2005; Halász *et al.*, 2007, 2010).

To amplify *FaSt* elements in *Prunus* species, a specific forward and a consensus reverse primer were designed based on the conserved terminal inverted-repeat arms of apricot and peach *FaSt* sequences: *FaSt*-F (5'-GAGATATTTAGTGATATACCAT-3') and *FaSt*-R (5'-GAGATTYAYTAWTATACC-3'). Reaction conditions were: initial denaturation at 94°C for 2 min and 35 cycles of 94°C for 10 sec, 55°C for 2 min and 72°C for 2 min. PCR was carried out in a PTC 200 thermocycler (MJ Research, Budapest, Hungary).

Approximately 20–80 ng of genomic DNA was used for PCR amplification in a 25-μl reaction volume containing 10× Taq™ buffer (Fermentas, Szeged, Hungary) with final concentrations of 4.5 mM MgCl<sub>2</sub>, 0.2 mM of dNTPs, 0.2 μM of the adequate primers and 0.75 U of Taq™ DNA polymerase (Fermentas).

### Microsatellite analysis

For microsatellite analysis, a set of eight genomic simple-sequence repeat (SSR) and two expressed sequence tag (EST) SSR primer pairs were selected on the basis of previous reports on *Prunus* species, covering different linkage groups (G1–G3, G5–G6): UDP 96005 (Cipriani *et al.*, 1999), BPPCT 025, BPPCT 038, BPPCT007, BPPCT 001 (Dirlewanger *et al.*, 2002), CPPCT 044 (Aranzana *et al.*, 2003), CPSC 021 (Mnejja *et al.*, 2004) as well as EPDCU 3083 and EPDCU 5100 (Howad *et al.*, 2005). The forward primers were labeled with 6-FAM fluorescent dye. PCR was carried out as described previously and according to the program reported for the primers.

### Data evaluation and bioinformatic analyses

The PCR products were separated on 2% TAE agarose gels at 100 V for 2 h and DNA bands were stained with ethidium bromide. Fragment sizes were estimated by comparison with the 1 kb + DNA ladder (Promega, Madison, WI, USA). For exact size determination of *S-RNase* first intron region fragments smaller than 500 bp and SSRs, the fluorescently labeled products were run in an automated sequencer ABI PRISM 3100 Genetic Analyzer using the GENOTYPER 3.7 software and GS500 LIZ size standard (Applied Biosystems, Budapest, Hungary).

DNA fragments to be sequenced were purified from agarose gel using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) and cloned into a pGEM-T Easy plasmid vector (Promega). Plasmid DNA was isolated with the Rapid Plasmid DNA Daily Miniprep Kit (V-gene, Szeged, Hungary) and sequenced in the above-described sequencer. For each fragment, the nucleotide sequences of four clones were determined in both directions.

Identification and characterization of the non-autonomous *Mutator* element were carried out using the Inverted Repeats Database 3.04 (Gelfand *et al.*, 2007). Repeat masker program (<http://repeat-masker.org>) was used to look for *FaSt* elements in plant and animal genomes. The identified *FaSt* sequence was also used as a query

sequence for BLASTN analysis (Altschul *et al.*, 1990) as well as BLAT (BLAST-like Alignment Tool) searches (Kent, 2002) on peach genome using Phytozome v9.1 (Goodstein *et al.*, 2014).

### Phylogenetic analyses

An alignment of 90 peach and two apricot *FaSt* sequences was carried out using the CLUSTAL W program (Thompson *et al.*, 1994) and the alignment was curated manually. Maximum likelihood analysis was carried out using MEGA5.1 (Tamura *et al.*, 2011) with 1200 bootstrap simulations and a majority-rule consensus tree was drawn.

Genetic relatedness among genotypes based on SSR analysis was studied by UPGMA (Unweighted Pair Group Method with Arithmetic averages) analysis using Popgene 1.32 (<http://www.ualberta.ca/~fyeh/>). A phylogenetic tree (dendrogram) was constructed using the TREEVIEW program (Page, 1996).

### Statistical analysis

Data were analyzed using the General Linear Model (GLM) procedure of the sas 2000 program (SAS Institute, Cary, NC, USA). Analysis of variance was performed using PROC GLM, with mean separations being conducted by Duncan's multiple range test at  $P \leq 0.05$ .

### Data access

Sequence data from this article have been submitted to the EMBL/GenBank databank under accession number KF956794.

### ACKNOWLEDGEMENTS

This work was supported by OTKA PD 78124 and the Moroccan-Hungarian Inter-Institutional S&T Cooperation Programme grant M/L n°3287 of the National Centre Research and Technology of Morocco (CNRST) and TÉT-10-1-2011-0505 of the National Development Agency of Hungary. J. Halász and A. Hegedűs are grateful for the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. The authors thank Gábor Tóth for his helpful comments and are grateful to Craig A. Ledbetter for critically revising the English of the manuscript.

### CONFLICTS OF INTEREST

The authors have no conflict of interest to declare.

### SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

**Figure S1.** Basic structure of the apricot *SFB<sub>C</sub>*-associated non-autonomous *Mutator* element, *Falling Stones*.

**Figure S2.** Alignment of the most similar peach *Falling Stones* DNA sequences.

**Table S1.** Sequences showing significant similarity with apricot *SFB<sub>C</sub>*-associated *Falling Stones* element revealed by BLASTN homology searches.

**Table S2.** Distribution of complete and partial *Falling Stones* elements in the peach genome.

**Table S3.** Sequence identity matrix of apricot and peach *Falling Stones* non-autonomous *Mutator* elements.

### REFERENCES

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

- Aranzana, M., Pineda, A., Cosson, P. et al. (2003) A set of simple-sequence repeat (SSR) markers covering the *Prunus* genome. *Theor. Appl. Genet.* **106**, 819–825.
- Baggiolini, M. (1952) Stades repères de l'abricotier. *Rev. Romande Agr. Viti-cult. Arboricult.* **8**, 28–29.
- Bennetzen, J.L. (2000) Transposable element contributions to plant gene and genome evolution. *Plant Mol. Biol.* **42**, 251–269.
- Bourguiba, H., Audergon, J.-M., Krichen, L., Trifi-Farah, N., Mamouni, A., Trabelsi, S., D'Onofrio, C., Asma, B., Santoni, S. and Khadari, B. (2012) Loss of genetic diversity as a signature of apricot domestication and diffusion into the Mediterranean Basin. *BMC Plant Biol.* **12**, 49.
- Casacuberta, E. and González, J. (2013) The impact of transposable elements in environmental adaptation. *Mol. Ecol.* **22**, 1503–1517.
- Cipriani, G., Lot, G., Huang, W.G., Marrazzo, M.T., Peterlunger, E. and Testolin, R. (1999) AC/GT and AG/CT microsatellite repeats in peach [*Prunus persica* (L.) Batsch]: isolation, characterisation and cross-species amplification in *Prunus*. *Theor. Appl. Genet.* **99**, 65–72.
- Civán, P., Svec, M. and Hauptvogel, P. (2011) On the coevolution of transposable elements and plant genomes. *J. Bot.* **2011**, Article ID 893546.
- Dirlwanger, E., Cosson, P., Tavaud, M., Aranzana, M., Poizat, C., Zanetto, A., Arús, P. and Laigret, F. (2002) Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). *Theor. Appl. Genet.* **105**, 127–138.
- Downen, R.H., Pelizzola, M., Schmitz, R.J., Lister, R., Downen, J.M., Nery, J.R., Dixon, J.E. and Ecker, J.R. (2014) Widespread dynamic DNA methylation in response to biotic stress. *Proc. Natl Acad. Sci. USA*, **109**, E2183–E2191.
- Engels, W.R., Johnson-Schlitz, D.M., Eggleston, W.B. and Sved, J. (1990) High-frequency P element loss in *Drosophila* is homolog dependent. *Cell*, **62**, 515–525.
- Entani, T., Iwano, M., Shiba, H., Che, S.F., Isogai, A. and Takayama, S. (2003) Comparative analysis of the self-incompatibility (S-) locus region of *Prunus mume*: identification of a pollen-expressed F-box gene with allelic-diversity. *Genes Cells*, **8**, 203–213.
- Feschotte, C., Zhang, X. and Wessler, S.R. (2002) Miniature inverted-repeat transposable elements (MITEs) and their relationship with established DNA transposons. In *Mobile DNA II*. (Craig, N., Craigie, R., Gellert, M. and Lambowitz, A., eds). Washington, DC: American Society of Microbiology Press, pp. 1147–1158.
- Gao, C., Xiao, M., Ren, X., Hayward, A., Yin, J., Wu, L., Fu, D. and Li, J. (2012) Characterization and functional annotation of nested transposable elements in eukaryotic genomes. *Genomics*, **100**, 222–230.
- Gelfand, Y., Rodriguez, A. and Benson, G. (2007) TRDB – the tandem repeats database. *Nucleic Acids Res.* **35**, D80–D87.
- Goodall, G.J. and Filipowicz, W. (1989) The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell*, **58**, 473–483.
- Goodstein, D.M., Shu, S., Howson, R. et al. (2014) *Phytozome: a comparative platform for green plant genomics*. *Nucleic Acids Res.* **40**, D1178–D1186.
- Halász, J., Hegedűs, A., Herman, R., Stefanovits-Banyai, E. and Pedryc, A. (2005) New self-incompatibility alleles in apricot (*Prunus armeniaca* L.) revealed by stylar ribonuclease assay and S-PCR analysis. *Euphytica*, **145**, 57–66.
- Halász, J., Pedryc, A. and Hegedűs, A. (2007) Origin and dissemination of the pollen-part mutated S<sub>p</sub>-haplotype which confers self-compatibility in apricot (*Prunus armeniaca*). *New Phytol.* **176**, 792–803.
- Halász, J., Pedryc, A., Ercisli, S., Yilmaz, K.U. and Hegedűs, A. (2010) S-genotyping supports the genetic relationships between Turkish and Hungarian apricot germplasm. *J. Am. Soc. Hortic. Sci.* **135**, 410–417.
- Halász, J., Hegedűs, A., Szikriszt, B., Ercisli, S., Orhan, E. and Ünü, H.M. (2013) The S-genotyping of wild-growing apricots reveals only self-incompatible accessions in the Erzincan region of Turkey. *Turk. J. Biol.* **37**, 733–740.
- Hauck, N.R., Ikeda, K., Tao, R. and Iezzoni, A.F. (2006) The mutated S<sub>1</sub>-haplotype in sour cherry has an altered S-haplotype-specific F-box protein gene. *J. Hered.* **97**, 514–520.
- Hegedűs, A., Lénárt, J. and Halász, J. (2012) Sexual incompatibility in *Rosa-ceae* fruit tree species: molecular interactions and evolutionary dynamics. *Biol. Plant.* **56**, 201–209.
- Howad, W., Yamamoto, T., Dirlwanger, E., Testolin, R., Cosson, P., Cipriani, G., Monforte, A.J., Georgi, L., Abbott, A.G. and Arús, P. (2005) Mapping with a few plants: using selective mapping for microsatellite saturation of the *Prunus* reference map. *Genetics*, **171**, 1305–1309.
- International Barley Genome Sequencing Consortium (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature*, **491**, 711–716.
- International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, **463**, 763–768.
- Jiang, N. and Wessler, S.R. (2001) Insertion preference of maize and rice miniature inverted repeat transposable elements as revealed by the analysis of nested elements. *Plant Cell*, **13**, 2553–2564.
- Joly-Lopez, Z., Forczek, E., Hoen, D.R., Juretic, N. and Bureau, T.E. (2012) A gene family derived from transposable elements during early angiosperm evolution has reproductive fitness benefits in *Arabidopsis thaliana*. *PLoS Genet.* **8**, e1002931.
- Kejnovsky, E., Hawkins, J.S. and Feschotte, C. (2012) Plant transposable elements: biology and evolution. In *Plant Genome Diversity*, vol 1 (Wendel, J.F., Greilhuber, J., Dolezel, J. and Leitch, I.J., eds). Vienna: Springer, pp. 17–34.
- Kent, W.J. (2002) BLAT – the BLAST-like alignment tool. *Genome Res.* **12**, 656–664.
- Kikuchi, K., Terauchi, K., Wada, M. and Hirano, H.-Y. (2003) The plant MITE mPing is mobilized in anther culture. *Nature*, **421**, 167–170.
- Kodad, O., Halász, J., Hegedűs, A., Messaoudi, Z., Pedryc, A. and Socias i Company, R. (2013a) Self-(in)compatibility and fruit set in 19 local Moroccan apricot (*Prunus armeniaca* L.) genotypes. *J. Hortic. Sci. Biotech.* **88**, 457–461.
- Kodad, O., Hegedűs, A., Socias i Company, R. and Halász, J. (2013b) Self-(in)compatibility genotypes of Moroccan apricots indicate differences and similarities in the crop history of European and North African apricot germplasm. *BMC Plant Biol.* **13**, 196.
- Kuang, H., Padmanabhan, C., Li, F., Kamei, A., Bhaskar, P.B., Ouyang, S., Jiang, J., Buell, C.R. and Baker, B. (2009) Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the *Solanaceae*: new functional implications for MITEs. *Genome Res.* **19**, 42–56.
- Lachkar, A., Fattouch, S., Ghazouani, T., Halász, J., Pedryc, A., Hegedűs, A. and Mars, M. (2013) Identification of self-(in)compatibility S-alleles and new cross-incompatibility groups in Tunisian apricot (*Prunus armeniaca* L.) cultivars. *J. Hortic. Sci. Biotech.* **88**, 497–501.
- Lisch, D. (2009) Epigenetic regulation of transposable elements in plants. *Annu. Rev. Plant Biol.* **60**, 43–66.
- Mnejja, M., Garcia-Mas, J., Howad, W., Badenes, M.L. and Arús, P. (2004) Simple-sequence repeat (SSR) markers of Japanese plum (*Prunus salicina* Lindl.) are highly polymorphic and transferable to peach and almond. *Mol. Ecol. Notes*, **4**, 163–166.
- Naito, K., Zhang, F., Tsukiyama, T., Saito, H., Hancock, C.N., Richardson, A.O., Okumoto, Y., Tanisaka, T. and Wessler, S.R. (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature*, **461**, 1130–1134.
- de Nettancourt, D. (2001) *Incompatibility and Incongruity in Wild and Cultivated Plants*, 2nd edn. New York, NY: Springer-Verlag.
- Page, R.D.M. (1996) Tree View: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12**, 357–358.
- Paterson, A.H., Bowers, J.E., Bruggmann, R. et al. (2009) *The Sorghum bicolor* genome and the diversification of grasses. *Nature*, **457**, 551–556.
- Romero, C., Vilanova, S., Burgos, L., Martínez-Calvo, J., Vicente, M., Llacer, G. and Badenes, M.L. (2004) Analysis of the S-locus structure in *Prunus armeniaca* L. Identification of S-haplotype specific S-RNase and F-box genes. *Plant Mol. Biol.* **56**, 145–157.
- Shirasawa, K., Hirakawa, H., Tabata, S., Hasegawa, M., Kiyoshima, H., Suzuki, S., Sasamoto, S., Watanabe, A., Fujishiro, T. and Isobe, S. (2012) Characterization of active miniature inverted-repeat transposable elements in the peanut genome. *Theor. Appl. Genet.* **124**, 1429–1438.
- Socias i Company, R. (1979) Aportación a las técnicas de observación de tubos polínicos. Caso del almendro. *An. Inst. Nac. Invest. Agrar. Ser. Prod. Veg.* **10**, 233–236.
- Sonneveld, T., Tobutt, K.R. and Robbins, T.P. (2003) Allele-specific PCR detection of sweet cherry self-incompatibility (S) alleles S1 to S16 using

- consensus and allele-specific primers. *Theor. Appl. Genet.* **107**, 1059–1070.
- Sutherland, B.G., Robbins, T.P. and Tobutt, K.R. (2004) Primers amplifying a range of *Prunus* S-alleles. *Plant Breeding*, **123**, 582–584.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.
- Tao, R., Watari, A., Hanada, T., Habu, T., Yaegaki, H., Yamaguchi, M. and Yamane, H. (2007) Self-compatible peach (*Prunus persica*) has mutant versions of the S haplotypes found in self-incompatible *Prunus* species. *Plant Mol. Biol.* **63**, 109–123.
- Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
- Ushijima, K., Yamane, H., Watari, A., Kakehi, E., Ikeda, K., Hauck, N.R., Iezzoni, A.F. and Tao, R. (2004) The S haplotype-specific F-box protein gene, *SFB*, is defective in self-compatible haplotypes of *Prunus avium* and *P. mume*. *Plant J.* **39**, 573–586.
- Verde, I., Abbott, A.G., Scalabrin, S. et al. (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat. Genet.* **45**, 487–494.
- Vilanova, S., Romero, C., Burgos, L., Llacer, G. and Badenes, M.L. (2005) Identification of self-(in)compatibility alleles in apricot (*Prunus armeniaca* L.) by PCR and sequence analysis. *J. Am. Soc. Hortic. Sci.* **130**, 893–898.
- Vilanova, S., Badenes, M.L., Burgos, L., Martinez-Calvo, J., Llacer, G. and Romero, C. (2006) Self-compatibility of two apricot selections is associated with two pollen-part mutations of different nature. *Plant Physiol.* **142**, 629–641.
- Wendel, J.F., Greilhuber, J., Dolezel, J., Leitch, I.J., Slotkin, R.K., Nuthikattu, S. and Jiang, N. (2012) The impact of transposable elements on gene and genome evolution. In *Plant Genome Diversity Volume 1* (Wendel, J.F., Greilhuber, J., Dolezel, J. and Leitch, I.J., eds). Vienna: Springer, pp. 35–58.
- Wessler, S.R., Bureau, T.E. and White, S.E. (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Curr. Opin. Genet. Dev.* **5**, 814–821.
- Wicker, T., Sabot, F., Hua-Van, A. et al. (2007) A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982.
- Yamane, H., Ikeda, K., Hauck, N.R., Iezzoni, A.F. and Tao, R. (2003) Self-incompatibility (S) locus region of the mutated S<sub>6</sub>-haplotype of sour cherry (*Prunus cerasus*) contains a functional pollen S allele and a non-functional pistil S allele. *J. Exp. Bot.* **54**, 2431–2437.
- Zhang, X., Feschotte, C., Zhang, Q., Jiang, N., Eggleston, W.B. and Wessler, S.R. (2001) P instability factor: an active maize transposon system associated with the amplification of Tourist-like MITEs and a new superfamily of transposases. *Proc. Natl Acad. Sci. USA*, **98**, 12572–12577.