the plant journal

SEB Society for Experimental Biology

The Plant Journal (2016) **85,** 70–82 doi: 10.1111/tpj.13090

A highly specific microRNA-mediated mechanism silences LTR retrotransposons of strawberry

Nada Šurbanovski, Matteo Brilli, Mirko Moser and Azeddine Si-Ammour*

Functional Genomics, Department of Genomics and Biology of Fruit Crops, Research and Innovation Centre, Fondazione Edmund Mach, Via E. Mach 1, San Michele all' Adige 38010, Italy

Received 6 August 2015; revised 6 November 2015; accepted 20 November 2015; published online 27 November 2015.
*For correspondence (email azeddine.siammour@fmach.it).

SUMMARY

Small RNAs are involved in a plethora of functions in plant genomes. In general, transcriptional gene silencing is mediated by 24-nucleotide siRNAs and is required for maintaining transposable elements in a silenced state. However, microRNAs are not commonly associated with transposon silencing. In this study, we performed small RNA transcriptome and degradome analyses of the Rosaceae model plant *Fragaria vesca* (the woodland strawberry) at the genome-wide level, and identified miRNA families and their targets. We report a highly specific mechanism of LTR retrotransposon silencing mediated by an abundant, ubiquitously expressed miRNA (fve–miR1511) generated from a single locus. This miRNA specifically targets LTR retroelements, silencing them post-transcriptionally by perfectly pairing to the highly conserved primer binding site for methionyl initiator tRNA that is essential for reverse transcription. We investigated the possible origins of this miRNA, and present evidence that the pre-miR1511 hairpin structure probably derived from a locus coding for tRNA^{iMet} through a single microinversion event. Our study shows that this miRNA targets retrotransposons specifically and constitutively, and contributes to features such as genome stability, size and architecture in a far more direct way than previously thought.

Keywords: Degradome, miRNA, LTR retrotransposon, primer binding site, *Fragaria vesca*, tRNA.

INTRODUCTION

Several classes of small RNAs (sRNAs) with sizes ranging from 20–24 nucleotides are expressed in plants, with the 21 nt and 24 nt fractions being the most abundant (Axtell, 2013). The 21 nt sRNA class includes microRNAs (miRNAs), which have important roles in numerous biological functions and developmental pathways through post-transcriptional regulation of protein coding genes (Meins *et al.*, 2005; Jones-Rhoades *et al.*, 2006). miRNA duplexes are processed from hairpin-like precursor miRNA transcripts (pre-miRNAs) by the RNase III-like enzyme Dicer-like 1 (DCL1), leaving 2 nt overhangs at the 3′ termini. The mature strand is recruited by the RNA-induced silencing complex containing Argonaute 1 to mediate cleavage of cognate mRNAs or repress translation, but the miRNA on the opposite strand (miRNA*) is less stable (Rogers and Chen, 2013). In contrast, small interfering RNA (siRNAs) are typically 21–24 nt in length, are generated from double-stranded RNA of non-coding, viral or transposon transcripts, and are involved in DNA methylation, heterochromatin formation, and transcriptional and post-transcriptional gene silencing (Matzke and Mosher, 2014). In plants, a large proportion of 24 nt sRNAs originates from repeats and transposons. They direct *de novo* methylation of the transposable element from which they originate, in a sequence-specific manner through the RNA-directed DNA methylation pathway (Simon and Meyers, 2011). Therefore, transposable elements are maintained in a transcriptionally silent state by a self-reinforcing mechanism involving siRNA-directed *de novo* methylation and maintenance of the methylated state by DNA methyltransferases as well as post-translational histone modifications (Matzke and Mosher, 2014). Recently, miRNA families have also been implicated in transposon silencing, by a latent mechanism targeting transposon transcripts when they are epigenetically reactivated in Arabidopsis (Creasey *et al.*, 2014).

Transposons are mobile genetic elements that mobilize and multiply in the genome in a variety of ways (Grandbastien and Casacuberta, 2012; Bennetzen and Wang, 2014). Transposon activation may be detrimental for the host organism, and the increase in transposable element copy numbers in genomes has been shown to be

negatively correlated with host fitness (Slotkin and Martienssen, 2007). Nevertheless, they are also key players in generating genomic novelty, and are known to affect genome size, gene content, gene order and numerous other aspects of genome biology, through the genomic rearrangements that they cause (Slotkin and Martienssen, 2007; Lisch, 2013). Retroelements (class I transposable elements) are extremely abundant in genomes due to their mobilizing strategy, which involves transcription into RNA precursors and integration into new sites, and their transcriptional activity has been detected on a genome-wide scale (Jiao and Deng, 2007; Schulman, 2013). Because retroelements deteriorate over time, non-autonomous elements vastly outnumber autonomous elements in the genome. Genomic parasitism, whereby a non-autonomous retroelement engages the protein machinery of an autonomous element *in trans* and propagates through the genome, is quite common (Bennetzen and Wang, 2014).

Long terminal repeat (LTR) retrotransposons are high copy number elements transcribed from the promoter regions present in the direct repeat LTR sequences (Schulman, 2013). They are the most abundant class of transposable elements in plant genomes: mostly dormant but often reactivated under stress conditions (Alzohairy *et al.*, 2014). The general level of transcription of LTR retroelements varies between plant species, and they replicate via reverse transcription of their template RNA (Slotkin and Martienssen, 2007; Chang and Schulman, 2008). Most commonly, a host tRNA is co-opted as a primer for this purpose, and LTR retroelements harbor highly conserved sequences of 8-18 nucleotides found in close proximity to the 5′ LTR, known as primer binding sites (PBS) to which the tRNA molecules bind (Marquet *et al.*, 1995; Wilhelm and Wilhelm, 2001). Reverse transcription is a complex process, which has been studied in detail for Ty elements of yeast. These studies have shown that binding with full complementarity in the 10 nt PBS region is essential for the process, and even a single base pair mutation has severe effects on transposition (Keeney *et al.*, 1995). However, while complementarity to the PBS is necessary, it is not sufficient for tRNA priming of reverse transcription. In addition to the tRNA acceptor stem, whose 3′ end is complementary to the PBS sequence, the TψC and D arms are also involved in priming (Keeney *et al.*, 1995), and have been found to bind to other conserved stretches of nucleotides in the transposon sequence (Friant *et al.*, 1996). These extended interactions help stabilize the primer–template complex, and it has been shown *in vivo* that disruption of interactions between the retrotransposon and a $tRNA^{iMet}$ in regions outside the PBS also abolishes initiation of reverse transcription (Wilhelm and Wilhelm, 2001).

The genus *Fragaria* (strawberries) belongs to the family Rosaceae, which includes over 100 flowering plant genera. Many Rosaceous species are fruit crops of high nutritional value and economic importance. The woodland strawberry (*F. vesca* L.) is a model plant and a versatile experimental system. An interesting feature of the *F. vesca* genome is its small size, which has been attributed to the lack of highly abundant LTR retrotransposons (Shulaev *et al.*, 2011).

In this study, we investigated the *F. vesca* small RNA transcriptome with a focus on miRNAs. We identified known and previously undiscovered miRNA families, determined their tissue-specific expression, and identified their cleaved targets. We found that the most abundant miRNA of *F. vesca*, fve–miR1511, specifically cleaves LTR retrotransposon transcripts at the PBS site. We also present evidence that the miR1511 hairpin originated from a $tRNA^{iMet}$ locus and was probably generated by a single microinversion event.

RESULTS

Genome-wide analysis of the *F. vesca* sRNA transcriptome and degradome

Small RNAs of *F. vesca* 'Hawaii 4′ were extracted from leaves, stolons, flower buds, open flowers and fruits, and identified by deep sequencing. More than 80 million raw reads were obtained in total. After removing the adaptors, sequences with a length between 19 and 24 nucleotides were selected and filtered from rRNA, tRNA, snRNA and snoRNA (Table S1). A total of 66.8 million reads were obtained, of which 24.5 million were unique. The 24 nt class was prevalent in all tissues, and the second most abundant class was the 21 nt class, which showed high degree of redundancy (Figure S1). In order to determine the origin and distribution of the reads, sequences of 19–24 nt in length were mapped on the *Fragaria vesca* pseudochromosome genome version 1.1 (Figure 1). The genome was partitioned into non-overlapping windows of 100 kb, and the total normalized abundance of three classes of sRNAs was calculated and plotted. The 24 nt sRNA species had a fairly symmetrical distribution from the two strands, with abundance broadly coinciding with peaks of retrotransposon density and troughs of gene density, while the 21 nt class had an asymmetric distribution and was derived mostly from discrete loci across the genome. The coefficient of variation over the five tissues was calculated, and in the case of the 21 nt class, showed a number of hotspots, some of which may be attributed to individual 21 nt sequences that had similarity to known miRNA (Figure 1).

For the purpose of *F. vesca* miRNA identification, two slightly differing pipelines were used to analyze the sRNA sequencing (sRNA-seq) data. To identify known miRNAs, sRNA sequences that mapped to the *F. vesca* genome without mismatches were searched by alignment against the *Viridiplantae* pre-miRNA sequences in miRBase

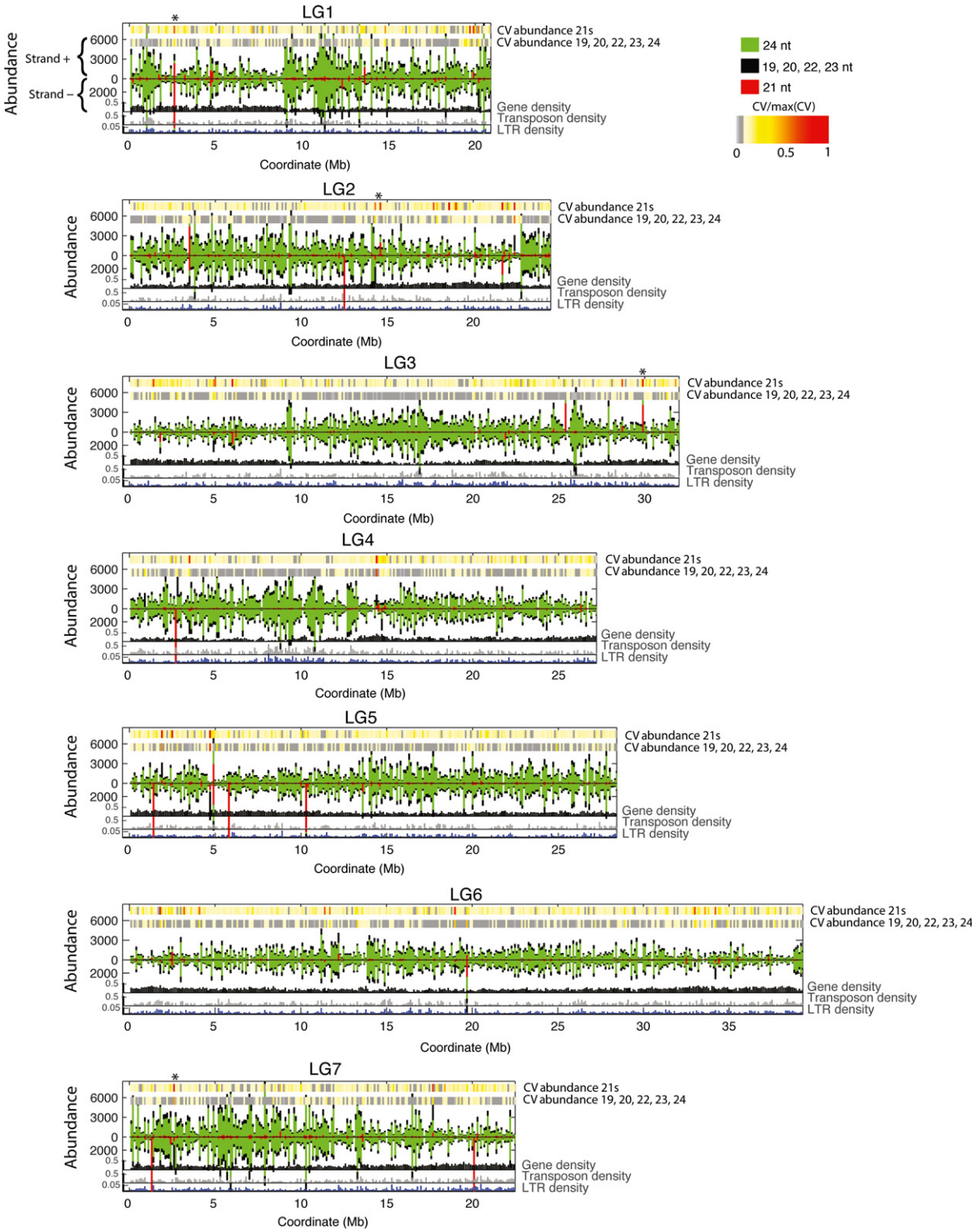

Figure 1. . Small RNAome of *F. vesca*.

Small RNA reads (19–24 nt) were mapped on linkage groups (LG) of *F. vesca* genome version 1.1. Colors in the stacked bar chart correspond to different lengths: 21 nt in red, 24 nt in light green, and 19, 20, 22 and 23 nt in black. Reads that map on the plus and minus strand were considered separately. The genomic features gene density, transposon density and LTR density are represented in the lower part of each panel by black, gray and blue bars, respectively. The coefficient of variation (CV) over the five tissues studied was calculated and is represented on the upper part of each panel. The color code for the CV tracks is also indicated. Asterisks indicate hotspots of variability that correspond to individual 21 nt sequences with similarity to known miRNAs: miR396 on LG1, miR167 on LG2, miR156/157 on LG3, and miR319 on LG7.

version 20.0 (www.mirbase.org). A subset of miRNAs matching the known sequences was further analyzed. The genome sequence spanning each perfect match was queried for the presence of a hairpin structure. To identify novel miRNAs, the already annotated known miRNAs were filtered out of the small RNA dataset, the remaining sequences were mapped on the *F. vesca* genome with no mismatches, and stem-loop like structures were identified using the miRDeep–P package (Yang and Li, 2011). The compulsory presence of an miRNA* in a predicted position was an additional criterion. miRNA sequences associated with loci that fulfilled all the above criteria were considered novel *F. vesca* miRNAs (Meyers *et al.*, 2008). The analysis identified 35 families of conserved or previously described miRNAs generated from 88 loci, and six novel miRNAs derived from six loci. Known and novel miRNAs, their abundances, foldback structures and position on *F. vesca* pseudochromosomes are given in Tables S2 and S3.

The abundance of miRNA families varied considerably (Figure 2 and Table S2). The most abundant family by far was fve–miR1511, which was derived from a single locus and exceeded the combined expression of the second and third most abundant miRNA families, fve–miR159/319 and fve–miR165/166, which were derived from nine and ten loci, respectively. fve–miR1511 showed ubiquitous expression and prevalence in all five tissues, with small variations in abundance ranging from 22 929 reads per million (RPM) in the fruit, to 44 674 RPM in the leaf, and yielded a total of 165 982 RPM for the mature miRNA and 1216 RPM for the miRNA* (Table S2). RNA blot hybridizations corroborated the abundance obtained by Illumina sequencing, showing that fve–miR1511 is strongly expressed in tissues at higher levels than fve–miR165/166 (Figure S2). In line with observations in other fruit species (Moxon *et al.*, 2008a; Pantaleo *et al.*, 2010; Xia *et al.*, 2012, 2015; Zhu *et al.*, 2012), most of the *Fragaria* miRNA families showed differential expression between the tissues analyzed (Figure 2 and Figure S2).

In order to experimentally identify transcripts targeted by *F. vesca* miRNAs, we constructed a PARE (parallel analysis of RNA ends) library from pooled RNAs of leaf, bud, flower and fruit tissues. This technique allows empirical analysis of the *F. vesca* degradome at the genome-wide level. A total of 98.8 million reads were obtained, processed and converged with the miRNA data using the

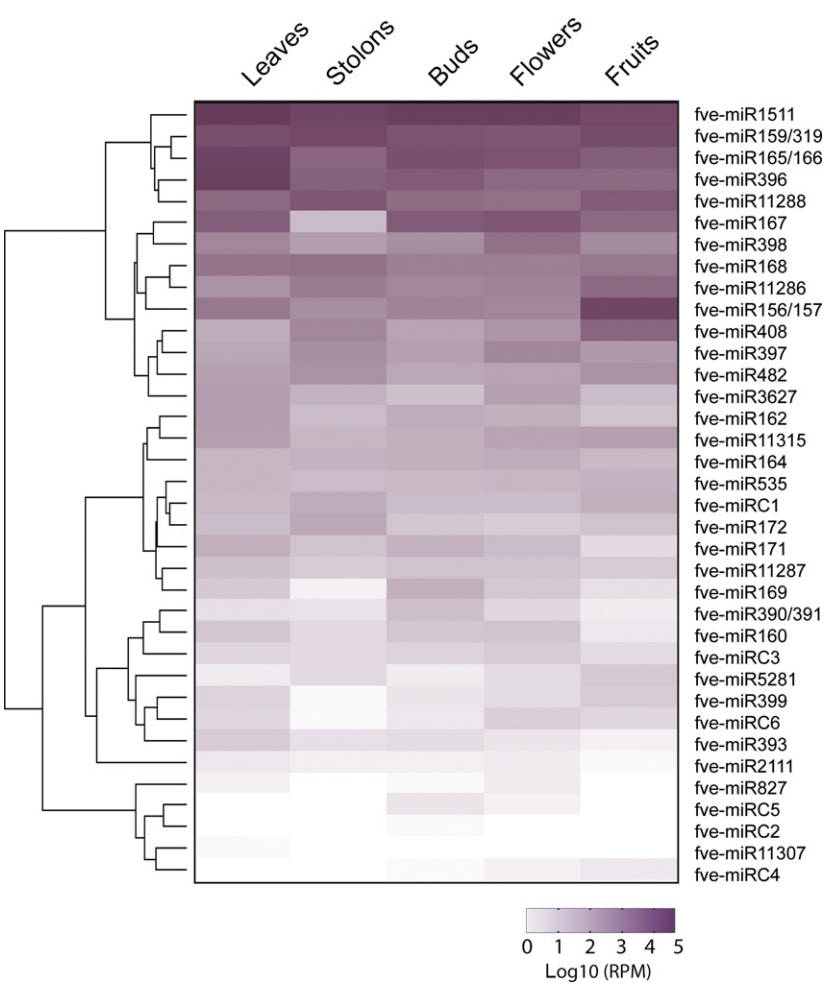

Figure 2. miRNA expression in *F. vesca* tissues. Heatmap representing hierarchical clustering of miRNA expression values in *F. vesca* tissues. The scale indicates reads per million (RPM) values expressed in $\log_{10}$.

SeqTar pipeline (Zheng *et al.*, 2011). In total, 86 genes targeted and cleaved by known miRNAs and six targeted and cleaved by novel miRNAs were identified using SeqTar under stringent criteria ($P_m < 0.1$, $P_v \leq 10^{-12}$, and a minimum of 5% reads in the valid peak). All targets identified in this analysis are listed in Table S4.

Cleaved targets of fve–miR1511

Fragaria vesca miR1511 was by far the most abundant miRNA identified in our study. The miR1511:miR1511* duplex contains two G:U wobble pairs and two mismatches, and is processed from the pre-miR1511 stem with 2 nt overhangs at the 3′ termini (Table S2). These features are typical of DCL1 products, suggesting that fve–miR1511 is a *bona fide* miRNA. Our pipeline, relying on an annotated transcriptome database for discovery of target genes, identified two cleaved targets of miR1511 (Table S4). A screen for putative domains in the coding regions of these potential targets revealed the presence of a plant mobile domain within gene03156 and a LTR-like sequence in the vicinity of the fve–miR1511 pairing site in the 4th exon of gene 15384 (Figure 3a,b). We therefore hypothesized that fve–miR1511 potentially targets additional transposon-containing loci in the *F. vesca* genome, and tested this hypothesis. The mature fve–miR1511 sequence was used to query the genome for potential targets using BLAST (http://blast.ncbi.nlm.nih.gov), resulting in 1950 matches. In order to identify possible cleavage sites in the matching sequences, regions 300 bp upstream and downstream of the fve–miR1511 pairing site were retrieved for each locus and converged with the degradome data in an additional SeqTar analysis. A total of 74 potential targets showing fve–miR1511-mediated cleavage were identified (Table S5 and Data S1) using stringent criteria as stated previously ($P_v \leq 10^{-12}$, $P_m < 0.1$, and a minimum of 5% reads in the valid peak). In order to determine their identity, the corresponding 74 sequences were used to perform a BLAST search against the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov) database, revealing that 40 of them had high similarity to *Cassandra* TRIM retroelements. A search for additional LTR sequences using *LTRharvest* software (Ellinghaus *et al.*, 2008) was performed, and 15 additional LTR retrotransposons were identified. In total, fve–miR1511 targets 55 LTR retrotrans-

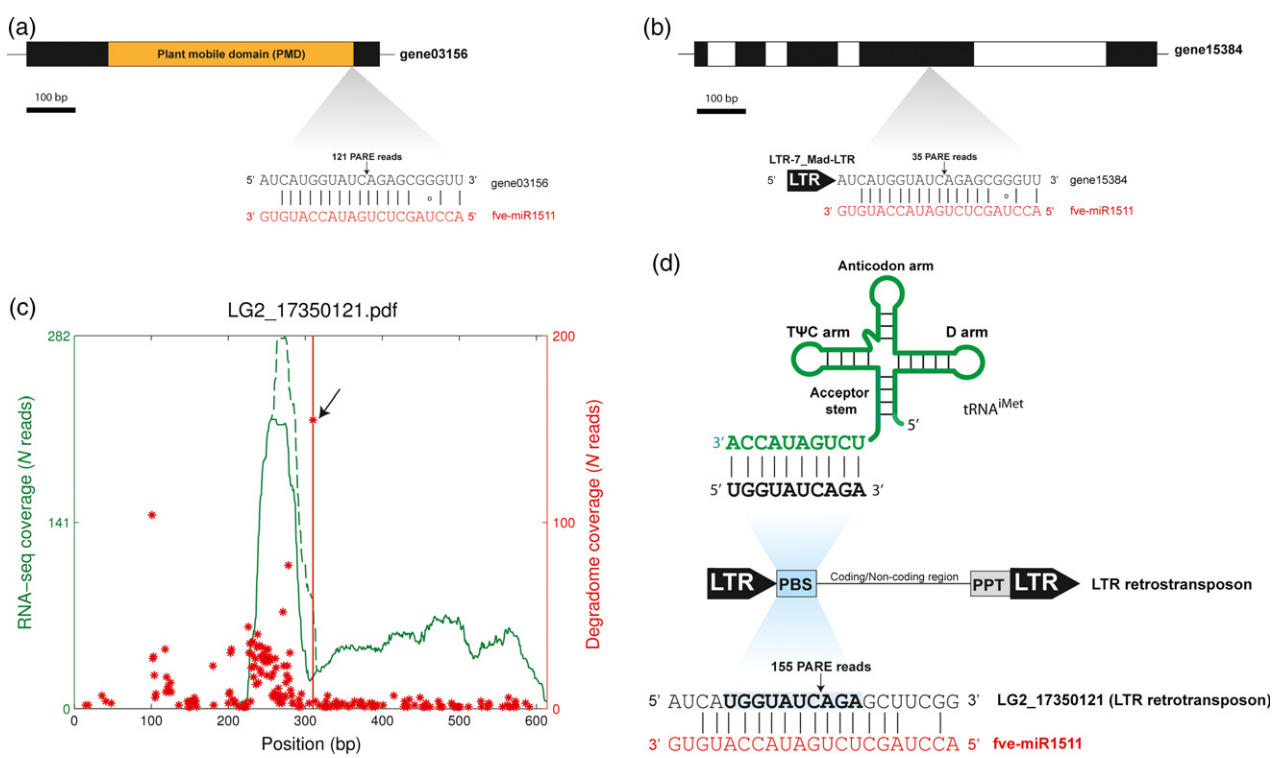

Figure 3. fve–miR1511 targets LTR retrotransposons at the PBS site.
(a,b) Pairing sites of fve–miR1511 on gene 03156 (a) and gene 15384 (b). Black and white boxes represent exons and introns, respectively. The plant mobile domain is indicated in yellow in (a). The LTR-7_Mad-LTR repeat motif inserted in the 4th exon of gene 15384 is indicated by a black arrow in (b).
(c) Example of a transcribed and cleaved LTR retrotransposon target with unique PARE reads only. The RNA-seq coverage is indicated in green; the dashed line represents total RNA-seq reads, and the full line represents RNA-seq reads mapping uniquely to the region. Red indicates mapping of the PARE degradome reads. The red line indicates the PBS site. The arrow indicates the valid cleavage peak.
(d) Schematic representation of PBS involvement in $tRNA^{iMet}$ priming to a LTR retrotransposon and in miR1511-mediated cleavage. The LTR sequences are indicated by black arrows. The PBS and the PPT (polypurine tract) site are represented by blue and gray boxes, respectively. fve–miR1511, indicated in red, targets the PBS sequence and mediates cleavage (indicated by a vertical arrow) of the LTR retrotransposon transcript.

Table 1 LTR retrotransposon targets of fve–miR1511

Coordinates[a]	Target binding site[b]	P_v[c]	Percentage[d]	Annotation[e]	PARE mapping[f]
LG1_10776535/[−]	CAAA**UGGUAUCAGA**GCCACUC	7.71E-84	16.31	TRIM *Cassandra*	N
LG1_16566742/[−]	CAAA**UGGUAUCAGA**GCCACUC	9.74E-85	16.93	TRIM *Cassandra*	N
LG1_20566502/[+]	CAAA**UGGUAUCAGA**GCCACUC	6.84E-85	17.03	TRIM *Cassandra*	N
LG1_7176090/[−]	CAAA**UGGUAUCAGA**GCCACUC	6.45E-69	8.68	TRIM *Cassandra*	N
LG2_13704440/[+]	CAAA**UGGUAUCAGA**GCCACUC	8.55E-68	8.27	TRIM *Cassandra*	N
LG2_14754740/[+]	AGCA**UGGUAUCAGA**GCCUCGGC	0.00E+00	6.48	LTR retrotransposon	U
LG2_17350121/[+]	AUCA**UGGUAUCAGA**GCUUCGGC	3.29E-194	7.46	LTR retrotransposon	U
LG2_18434401/[+]	CAAA**UGGUAUCAGA**GCCACUC	2.63E-75	11.42	TRIM *Cassandra*	N
LG2_21027920/[−]	AUCA**UGGUAUCAGA**GCUUUGGC	2.75E-32	31.37	LTR retrotransposon	N
LG2_9213093/[−]	CAAA**UGGUAUCAGA**GCCACUC	1.00E-72	10.23	LTR retrotransposon	N
LG3_14187734/[+]	CAAA**UGGUAUCAGA**GCCACUC	7.34E-74	10.74	TRIM *Cassandra*	N
LG3_16233332/[−]	CAAA**UGGUAUCAGA**GCCACUC	1.95E-85	17.42	TRIM *Cassandra*	N
LG3_17021806/[−]	CAAA**UGGUAUCAGA**GCCACUC	8.25E-76	11.66	TRIM *Cassandra*	N
LG3_17399929/[+]	AAAA**UGGUAUCAGA**GCCUAGGU	1.36E-55	65.71	LTR retrotransposon	N
LG3_17723328/[+]	CAGA**UGGUAUCAGA**GCCACUC	2.75E-72	10.04	LTR retrotransposon	N
LG3_17737303/[+]	CAAA**UGGUAUCAGA**GCGGGGA	8.21E-31	14.29	LTR retrotransposon	U
LG3_17930518/[+]	CAAA**UGGUAUCAGA**GCCACUC	4.53E-80	13.95	TRIM *Cassandra*	N
LG3_19151570/[+]	CAAA**UGGUAUCAGA**GCCACUC	2.04E-70	9.26	TRIM *Cassandra*	N
LG3_19462848/[−]	CAAA**UGGUAUCAGA**GCCACUC	8.08E-71	9.42	TRIM *Cassandra*	N
LG3_21220974/[−]	CAAA**UGGUAUCAGA**GCCACUC	1.65E-84	16.77	TRIM *Cassandra*	N
LG3_22974557/[−]	CAAA**UGGUAUCAGA**GCCACUC	5.51E-72	9.91	TRIM *Cassandra*	N
LG3_25800901/[−]	AGCA**UGGUAUCAGA**GCCUCGGC	0.00E+00	44.99	LTR retrotransposon	N
LG3_28258076/[−]	CAAA**UGGUAUCAGA**GCCACUC	4.40E-81	14.56	TRIM *Cassandra*	N
LG3_30669106/[+]	CAAA**UGGUAUCAGA**GCCACUC	1.09E-67	8.23	TRIM *Cassandra*	N
LG3_4290110/[+]	AUCA**UGGUAUCAGA**GCUUCGGC	1.49E-118	33.15	LTR retrotransposon Ty1-like	N
LG3_6263577/[+]	CAAA**UGGUAUCAGA**GCCACUC	2.24E-97	27.98	TRIM *Cassandra*	N
LG4_10311720/[−]	CAAA**UGGUAUCAGA**GCCACUC	5.02E-89	20.15	TRIM *Cassandra*	N
LG4_11167507/[+]	CAAA**UGGUAUCAGA**GCCACUC	8.18E-75	11.18	TRIM *Cassandra*	N
LG4_11247500/[+]	CAAA**UGGUAUCAGA**GCCACUC	2.97E-69	8.81	LTR retrotransposon	N
LG4_11253323/[+]	CAAA**UGGUAUCAGA**GCCACUC	3.86E-70	9.15	TRIM *Cassandra*	N
LG4_11559822/[−]	CAAA**UGGUAUCAGA**GCCACUC	1.01E-73	10.67	TRIM *Cassandra*	N
LG4_12056860/[+]	CAAA**UGGUAUCAGA**GCCACUC	7.88E-68	8.28	TRIM *Cassandra*	N
LG4_17710783/[−]	CAAA**UGGUAUCAGA**GCCACUC	5.92E-69	8.7	TRIM *Cassandra*	N
LG4_5896224/[−]	CAAA**UGGUAUCAGA**GCCACUC	8.17E-74	10.71	TRIM *Cassandra*	N
LG4_7577437/[+]	CAA**UGGUAUCAGA**GCCACUC	6.09E-82	15.08	TRIM *Cassandra*	N
LG5_12877757/[+]	CAAA**UGGUAUCAGA**GCCACUC	4.41E-73	10.38	TRIM *Cassandra*	N
LG5_13128742/[+]	CAAA**UGGUAUCAGA**GCCACUC	2.49E-87	18.82	TRIM *Cassandra*	N
LG5_13996276/[+]	CAA**UGGUAUCAGA**GCCACUC	2.21E-80	14.14	TRIM *Cassandra*	N
LG5_26315901/[−]	AGCA**UGGUAUCAGA**GCCUCGGC	0.00E+00	34.71	LTR retrotransposon	N
LG5_3406121/[−]	CAAA**UGGUAUCAGA**GCCACUC	1.82E-94	25	TRIM *Cassandra*	N
LG5_9134622/[+]	CAAA**UGGUAUCAGA**GCCACUC	1.47E-77	12.56	TRIM *Cassandra*	N
LG6_18169802/[−]	CAAA**UGGUAUCAGA**GCCACUC	2.23E-76	11.95	TRIM *Cassandra*	N
LG6_20289550/[−]	CAAA**UGGUAUCAGA**GCAGGGA	0.00E+00	36.33	LTR retrotransposon	U
LG6_21378406/[+]	CAAA**UGGUAUCAGA**GCCACUC	3.27E-89	20.3	TRIM *Cassandra*	N
LG6_22431886/[+]	CAAA**UGGUAUCAGA**GCCACUC	1.95E-85	17.42	TRIM *Cassandra*	N
LG6_28293794/[−]	AGCA**UGGUAUCAGA**GCUCAUG	0.00E+00	27.15	LTR retrotransposon	U
LG6_29476838/[+]	CAAA**UGGUAUCAGA**GCCACUC	4.53E-90	21.01	TRIM *Cassandra*	N
LG6_37315863/[−]	CAAA**UGGUAUCAGA**GCCACUC	3.22E-70	9.18	TRIM *Cassandra*	N
LG6_39107051/[−]	AAAA**UGGUAUCAGA**GCCUAGGU	5.39E-47	34.85	LTR retrotransposon	N
LG7_14424508/[+]	CAAA**UGGUAUCAGA**GCCACUC	5.05E-71	9.51	TRIM *Cassandra*	N
LG7_14424850/[+]	CAAA**UGGUAUCAGA**GCCACUC	7.24E-70	9.05	TRIM *Cassandra*	N
LG7_1669827/[+]	CAAA**UGGUAUCAGA**GCCACUC	1.55E-73	10.59	TRIM *Cassandra*	N
LG7_1907025/[−]	AUCA**UGGUAUCAGA**GCAACGGC	0.00E+00	12.8	LTR retrotransposon	U
LG7_2450827/[+]	CAAA**UGGUAUCAGA**GCCACUC	1.86E-75	11.49	TRIM *Cassandra*	N
LG7_6585340/[−]	CAAA**UGGUAUCAGA**GCCACUC	2.12E-89	20.45	TRIM *Cassandra*	N

[a]The strand orientation on the linkage group (LG) is indicated by [+] and [−] for the plus and minus strand, respectively.
[b]The intact PBS sequence is indicated in bold.
[c]A $P_v \leq 10^{-12}$ was used as a threshold.
[d]Percentage of reads in the valid peak.
[e]Targets were annotated using NCBI BLAST and *LTRharvest* tools on the strand equivalent to and including the binding site.
[f]The PARE degradome reads mapping on each miR1511 target are either unique (U) or found in related retrotransposons and thus not unique (N).

posons for cleavage (Table 1 and Data S1). The remaining 19 potential targets were similar to various uncharacterized sequences (Table S5).

Retrotransposons of the same family may be highly similar, and *Cassandra* retroelements showed high levels of similarity throughout the whole transposon sequence (Table S5). This feature, in combination with the short length of PARE reads, compromises our ability to distinguish between retroelements that are genuinely expressed and cleaved and those that are highly similar but are transcriptionally silenced or otherwise inactivated. We therefore imposed rigorous criteria in an attempt to obtain a subset of targets that are transcriptionally active, have PARE reads that map uniquely to them and are clearly cleaved. We used publically available *F. vesca* RNA-seq data (Kang *et al.*, 2013) to confirm expression indicated by our PARE analysis, and found that half of the 74 targets were indeed expressed in the leaf tissue alone. After removing all PARE reads that mapped to more than one target, we obtained six LTR retrotransposon transcripts with prominent cleavage and unique PARE signatures, all of which were highly expressed in the leaf tissue of strawberry, thus presenting unequivocal targets of miR1511 (Figure 3c, Table 1 and Data S2).

Close inspection of the 74 target binding sites identified by our analysis revealed the presence of a conserved core sequence in all of them. We found that this sequence was in fact the primer binding site (PBS) for tRNA^{iMet}, which is involved in initiation of reverse transcription of LTR retrotransposons (Figure 3d) for which miR1511 harbors a full complement. Comparison of fve–miR1511 with other miRNAs deposited in miRBase showed that a full complement to the PBS for tRNA^{iMet} is present in seven other miRNA families: miR829 (*Arabidopsis thaliana*), miR845 (*Arabidopsis thaliana*, *Arabidopsis lyrata*, *Vitis vinifera* and *Brachypodium distachyon*), miR1863 (*Oryza sativa*, *Cucumis melo* and *Picea abies*) miR5139 (*Rehmannia glutinosa*), miR7782 (*Brachypodium distachyon*), miR7783 (*Brachypodium distachyon*) and miR8155 (*Carica papaya*) (Figure 4). Most of these miRNA families belong to angiosperm species, but one was identified in a gymnosperm (*P. abies*). Interestingly, five of the miRNA sequences were 24 nt in length in Arabidopsis, grape (*Vitis vinifera*), *Brachypodium distachyon*, apple (*Malus domestica*) and melon (*Cucumis melo*). This implies that post-transcriptional regulation of retrotransposons may not be the only mechanism targeting the PBS sequence in these plant species.

tRNA^{iMet} as a putative ancestor of miR1511

We postulated that miR1511 may have originated from a tDNA^{iMet} locus, and assessed the evidence for this hypothesis. tRNA^{iMet} sequences are highly conserved even between distant species such as *Arabidopsis thaliana* and *Fragaria vesca.* We therefore assumed that the tRNA ancestor had an identical sequence to the present-day tRNA^{iMet}. This tRNA folds into a typical cloverleaf structure, whereby the complementary sequence to the PBS is located at the 3′ terminus, ending with CCA, which is added post-transcriptionally and is therefore not present in the genomic sequence. The tRNA extends for additional seven nucleotides (UCUGAUA) that are perfectly complementary to the core sequence of mature miR1511. The sequence at the 5′ end of the tRNA forming the stem is fully complementary to the 3′ end, and shows high similarity to miR1511* (Figure 5a,b). However, if the TψC arm is joined to the acceptor stem (e.g. after deletion of the D and anticodon arms), the similarity to both the mature miR1511 and miRNA* extends for a further six and five nucleotides, respectively (Figure 5b). We also observed that the sequence in the stem of the TψC arm (5′-CCAGG-3′) is virtually palindromic (5′-CCTGG-3′ in the reverse complement), and therefore a single microinversion event (Figure 5c) provides essentially the same result as the aforementioned deletion, while keeping the rest of the tRNA sequence in place to form a foldback structure of similar size to pre-miR1511 loci (Figure 5c). Moreover, the described microinversion creates a foldback sequence with extensive identity to *Prunus persica* pre-miR1511 (Zhu *et al.*, 2012), both in the PBS-related regions and outside them (Figure 5d,e). The tRNA sequences equivalent to the miRNA and miRNA* were compared to miR1511 from

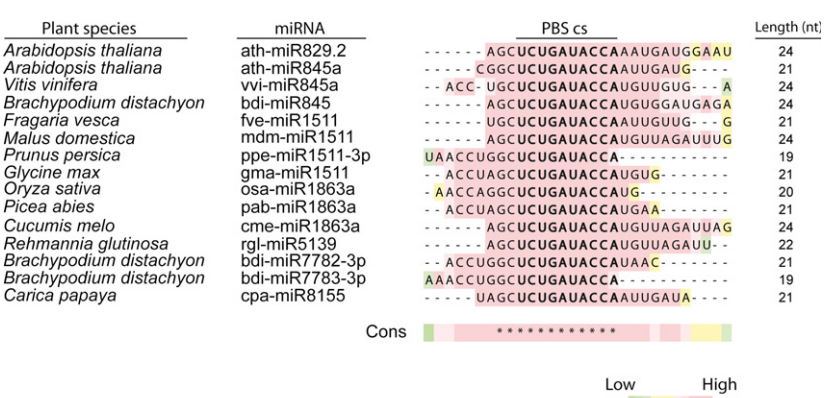

Figure 4. Plant miRNA families harboring a PBS complement.
The similarity between plant miRNAs harboring a PBS complement site (PBS cs, indicated in bold) present in miRBase version 20.0 was determined using M–Coffee. The color code for alignments represents the overall match quality at each nucleotide position. Conserved nucleotides across all miRNA sequences are indicated by asterisks in the consensus row (Cons).

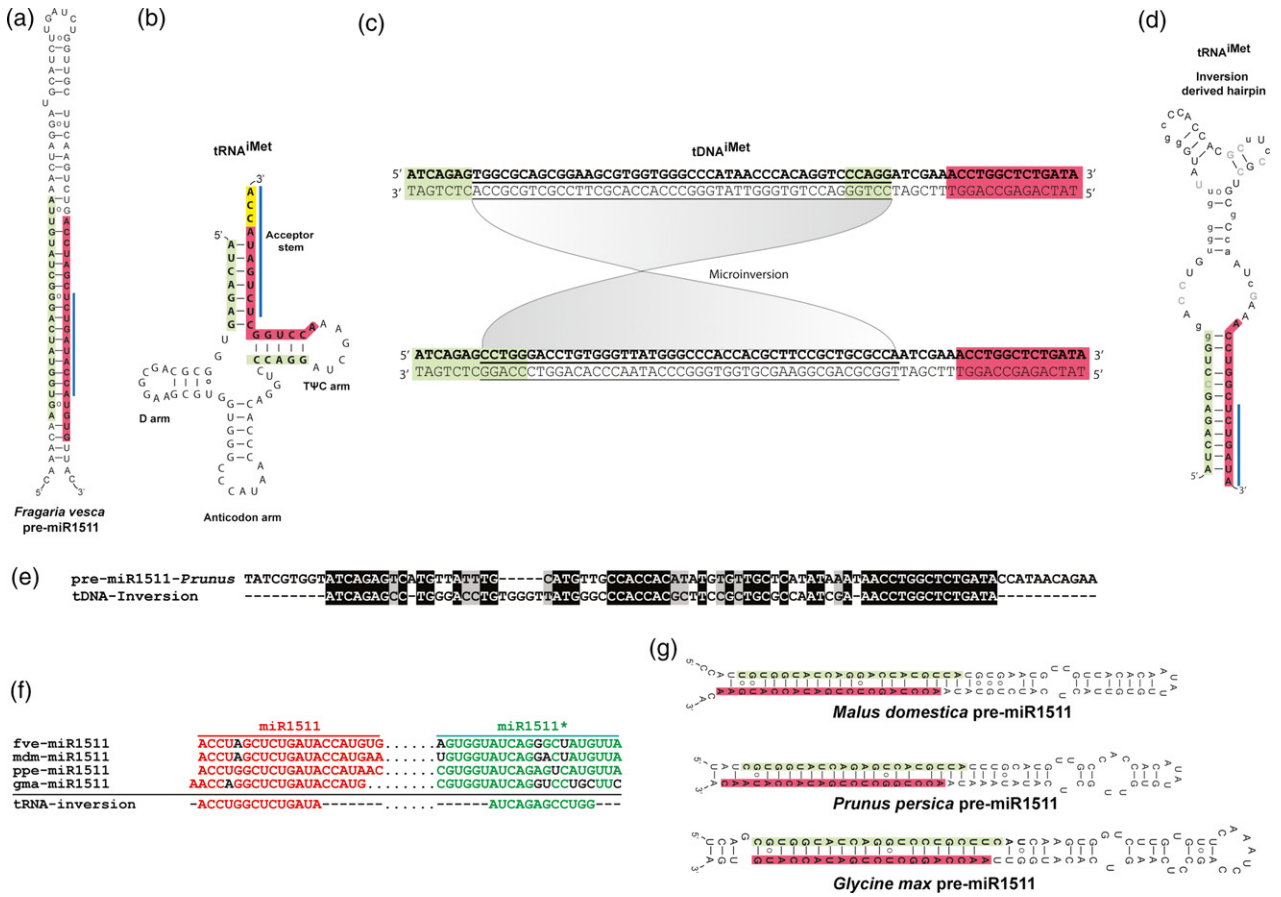

Figure 5. Proposed tRNA^{iMet} origin of miR1511.
(a) Secondary structure of fve–miR1511 pre-miRNA.
(b) Secondary structure of methionyl initiator tRNA (tRNA^{iMet}).
(c) Coding part of the methionyl initiator tDNA locus and the proposed microinversion.
(d) Secondary structure of the microinversion-derived tRNA sequence.
(e) Similarity between the microinversion-derived tDNA^{iMet} locus and *Prunus persica* pre-miR1511 locus.
(f) Alignments of reported miR1511 sequences and comparison to equivalent sequences from the proposed tRNA ancestor.
(g) Orientation of miRNA and miRNA* sequences in *M. domestica*, *P. persica* and *G. max* pre-miR1511.
Red/pink color indicates either mature miRNA in (a), (g) and (f) (above the line), or sequence similar to mature miRNA in (b), (c), (d) and (f) (under the line). Green color indicates either miRNA* in (a), (g) and (f) (above the line) or sequence similar to miRNA* in (b), (c), (d) and (f) (under the line). Yellow color in (b) indicates bases added post-transcriptionally. The blue bar indicates complementarity to the PBS site in (a), (b) and (d). The 45 bp inverted sequence is underlined in (c), and bold letters indicate the sense strand. In (e), black highlighting indicates identical bases, shown in bold black in (d); grey highlighting indicates transitions, shown in bold gray in (d); no highlighting indicates transversions, shown as lower-case letters in (d).

Malus domestica, *Prunus persica*, *Fragaria vesca* and *Glycine max*, and each position in the putative tRNA ancestor had a base that was present in the miR1511 sequence of at least one of the species (Figure 5f,g). We did not attempt to match the proposed tRNA ancestor to the consensus sequence because three of the sequences were more closely related (belonging to Rosaceae) and one was distantly related (belonging to Fabaceae).

DISCUSSION

Features of the *F. vesca* small RNAome

Mapping of the sRNAs on the *Fragaria* pseudochromosomes showed that 24 nt sRNAs had a symmetrical distribution and abundance that broadly correlated with peaks of retrotransposon density and troughs of gene density. In contrast, the 21 nt sRNA species displayed high redundancy and an asymmetrical distribution, as expected for sRNAs belonging to the miRNA class. During preparation of this paper, Xia *et al.* (2015) reported on *F. vesca* miRNAs and their targets; however, three additional known miRNA families and six families that were not previously reported were identified in our study (Table S3). The miRNA fve–miRC2, although not deposited in miRBase, was previously identified in *F.* × *ananassa* (Ge *et al.*, 2013; Xu *et al.*, 2013), but appears to be found only in the *Fragaria* genus. An interesting feature of *F. vesca* miRNA is the presence of several polycistronic pre-miRNA structures

that are rarely found in plants, such as pre-miR169 g/h, pre-miR482b/c and pre-miR11315 (Tables S2 and S3).

miR1511 targets LTR retrotransposon transcripts specifically at the PBS

The most abundant miRNA in all tissues was generated from a single locus, and did not belong to any of the usually highly abundant conserved miRNA families but was a member of the miR1511 family. This family was first reported in soybean (*Glycine max*, Fabaceae), but its target was not identified (Subramanian *et al.*, 2008). A subsequent study in soybean reported that a gene encoding a 60S ribosomal protein L4 is putatively cleaved by both miR1511 and miR1511* (Luo *et al.*, 2012). However, the suggested cleavage positions at the 7th and 20th nucleotide are not characteristic of Argonaute 1-mediated cleavage in plants. miR1511 was later identified in the Rosaceae species peach (*Prunus persica*) and apple (*Malus domestica*), where it was ubiquitously expressed and fairly abundant in all analyzed tissues, with the highest expression observed in the flower in both species, but no targets for miR1511 were described in these studies (Xia *et al.*, 2012; Zhu *et al.*, 2012). Finally, miR1511 was recently found in a different *F. vesca* cultivar, and was shown to be highly abundant in all analyzed tissues (leaf, seedling, flower bud, open flower, receptacle, ovary wall and seed) and at all time points; however, the targets of miR1511 were not identified (Xia *et al.*, 2015). All studies that attempted to find targets for miR1511 relied on annotated transcriptome databases, which do not contain transposable elements, and this may explain why the function of such a ubiquitous miRNA was obscure until now. Our initial analysis based on the annotated transcriptome uncovered two potential targets for fve–miR1511, both of which appeared to be retroelements. In an effort to fully understand the role of the most abundant miRNA in our study, we created an additional pipeline and discovered that it targets 74 putative transcripts, dispersed across all seven *F. vesca* chromosomes. A total of 55 of the 74 identified target sites were confirmed to be LTR retrotransposon sequences, and a minimum of six of these retrotransposons satisfy highly rigorous criteria by having a unique PARE signature, and being abundantly expressed as well as prominently cleaved, making them undisputable targets of fve–miR1511. An additional 19 potential targets could not be characterized; however, all 74 cleavage sites had an intact PBS sequence. It is well known that retroelements deteriorate over time and that their remnants make up the majority of plant DNA (Galun, 2003; Slotkin and Martienssen, 2007; Fernandez-Medina *et al.*, 2012; Lisch, 2013). It is therefore possible that some or all of the remaining 19 sequences in our analysis have a retrotransposon origin even though adequate homology to transposable elements deposited in the database was not found. Taking into consideration all our findings, we conclude that fve–miR1511 targets LTR retrotransposon transcripts specifically at the primer binding site complementary to methionyl initiator tRNA.

The PBS sequence for $tRNA^{iMet}$ is found in a variety of LTR retrotransposons; however, our analysis identified TRIM elements (terminal-repeat retrotransposons in miniature) as the most frequent target of miR1511. Genomes generally have a high abundance of highly reduced non-autonomous transposable elements but large autonomous elements are comparatively rare (Grandbastien and Casacuberta, 2012). This fact may account for both the lack of autonomous retrotransposons among the confirmed targets, and the prevalence of *Cassandra* elements. There may be an additional reason for the abundance of *Cassandra* elements, namely that this retroelement produces non-capped polyadenylated transcripts (Kalendar *et al.*, 2008), which may be readily captured and cloned into degradome libraries. Transcripts derived from LTR retrotranspons are rarely polyadenylated (Chang and Schulman, 2008). Non-polyadenylated transcripts are not generally captured by our approach, and it is likely that our degradome data show under-representation of these LTR retrotransposons. In addition, *Cassandra* elements have been thoroughly investigated, with many sequences deposited in the gene database, and they are fairly abundant in Rosaceae species, making their identification easier (Kalendar *et al.*, 2008; Yin *et al.*, 2014).

Members of the *Cassandra* family in *F. vesca* have high sequence similarity, making it difficult to know exactly which loci are expressed and cleaved. Nevertheless, our data suggest that members of the *Cassandra* family are regulated by miR1511, and we therefore compared the reported frequency of these well characterized retroelements in Rosaceae genomes with the reported abundance of miR1511 in the species that have been found to have this miRNA. miR1511 was the tenth most abundant miRNA species in apple (Xia *et al.*, 2012), the sixth most abundant in peach (Zhu *et al.*, 2012), and vastly outnumbered all other families in our study. Interestingly, the frequency of *Cassandra* elements in these species showed a reverse trend: apple, peach and strawberry had 2041, 667 and 132 copies of *Cassandra* in their genomes, accounting for 0.63%, 0.17% and 0.03% of their assembled genomic sequences, respectively (Yin *et al.*, 2014). These observations are in agreement with our findings for the role of miR1511.

Highly specific mechanism of LTR retrotransposon silencing

Keeping transposition under control is usually attributed to siRNA-directed DNA methylation or post-translational histone modifications (Matzke and Mosher, 2014). Recently, however, miRNAs have also been implicated in

retrotransposon silencing in a latent, epigenetically activated mechanism found in methylation mutants and pollen tissue of Arabidopsis (Creasey *et al.*, 2014). The study showed that Arabidopsis transposons were targeted by a number of miRNA species, many of them well known members of conserved miRNA families that are involved in other crucial functions in the plant. Sequence specificity of these diverse miRNAs for the transposons that they silence remains to be elucidated; however, the initial cleavage of transposon targets triggered generation of 21 nt easiRNA (epigenetically activated siRNA), similar to secondary siRNA biogenesis. We mapped the reads from our sRNA libraries onto the miR1511-cleaved transposons, and observed that 21 nt siRNAs generated from these transcripts are negligible in abundance (Data S3). Furthermore, and in contrast to the above mechanism, retrotransposon silencing by miR1511 is constitutive, ubiquitous and highly specific to the evolutionarily conserved PBS site that is necessary for transposition. Our results indicate that miR1511 counteracts a seemingly widespread expression of LTR retrotransposons in strawberry, and this appears to be the only function of miR1511.

Interestingly, miR1511 may not be the only miRNA targeting the PBS site for $tRNA^{iMet}$. In the miRNA database, there are seven additional miRNA families from angiosperms and gymnosperms that are fully complementary to this PBS sequence (Figure 4). Most of these families are poorly studied, and our findings indicate that they may constitute a plant miRNA superfamily that specifically targets the PBS of transposons. We propose that rare conserved regions such as the sites required for transposition may be of particular importance for host defense mechanisms.

Does miR1511 silence retrotransposons regardless of cleavage?

Due to complete complementarity in the core region, miR1511 has the capacity to bind to the PBS site without mismatches. This recognition results in cleavage of the transposon transcripts in strawberry, but, as retrotransposon transcripts undergo a different lifecycle to regular gene transcripts, the consequences of miRNA binding alone should be considered. Unlike regular gene transcripts, retrotransposons are reverse-transcribed, and the PBS has to be available for the tRNA molecule to bind to it. In order for reverse transcription to commence, the complex is stabilized by additional interactions of the TψC and D arms with the transposon transcript (Keeney *et al.*, 1995; Friant *et al.*, 1996; Schulman, 2013). Hence, binding of miR1511 to the PBS site effectively blocks priming of reverse transcription, regardless of the final destiny of the transcript such as cleavage by the RNA-induced silencing complex and/or other forms of degradation. Given the ubiquity of miR1511 in *F. vesca* tissues and the uncommonly high abundance of this miRNA, such a simple mechanism may be very efficient in preventing retrotransposition, and would also explain the small size of the *F. vesca* genome, which has already been attributed to a lack of LTR retrotransposons (Shulaev *et al.*, 2011). An additional point to consider is that the PBS–miRNA interaction may theoretically provide the elusive transposon-specific signal for RdRP (RNA dependent RNA polymerase), thereby converging this mechanism with the cellular siRNA response. While these avenues are speculative and beyond the scope of this study, we believe that they are worth exploring in the future.

Evolutionary origin of miR1511

From an evolutionary point of view, there are two main possibilities for the miR1511 origin. The first one involves retroelements themselves, and a number of studies have reported initial formation of various miRNA loci from transposable element sequences (Roberts *et al.*, 2014). According to this scenario, miR1511 may have originated as a result of insertion of a retroelement (or part of it) into the minus strand of another retroelement in the vicinity of the PBS site, providing the PBS complement on the plus strand. If the sequences of the retroelements were similar enough, they would also provide an miRNA foldback-like structure. However, reconstructing such an ancestor is a difficult and highly speculative task, as transposon sequences are diverse, numerous and prone to mutation. Nonetheless, this scenario cannot be excluded.

The second possibility for miR1511 origin arises from the fact that the PBS sequence is fully complementary to the 3′ end of the acceptor arm of $tRNA^{iMet}$. tRNA sequences are some of the most ancient and conserved RNA species known; they form secondary structures comprising short hairpins, and their tDNA is repeated multiple times, providing redundancy for the role of a single tDNA locus. It has also recently been suggested that the tRNA structure may be transformed into a foldback structure by very simple changes in the nucleotide sequence (Roberts *et al.*, 2013). Hence, we assessed the evidence for the hypothesis that miR1511 originated from a $tRNA^{iMet}$ locus. Strikingly, we found that a single microinversion event would create a foldback sequence with extensive similarity to *P. persica* pre-miR1511, both within and outside the regions associated with PBS function (Figure 5e). In the pre-miRNA sequence, miR1511* is located at the 5′ end, while the mature miR1511 is at the 3′ end in the loci of all four species (Figure 5a,g), suggesting that the same should apply to a putative ancestor, and this holds true in the case of the tRNA-derived hairpin.

The question of molecular ancestry for sequences that probably originated before the split between Rosaceae and Fabaceae more than a hundred million years ago (Crepet *et al.*, 2004) is a difficult one to answer beyond doubt;

however, the data available at present provide a number of clues in support of the hypothesis that the miR1511 hairpin derived from a $tRNA^{iMet}$ locus. Several studies in animals have already suggested a link between small non-coding RNAs and microRNA silencing mechanisms (Haussecker *et al.*, 2010; Maute *et al.*, 2013; Roberts *et al.*, 2013), and it has been shown that tRNA molecules are readily cleaved by the Dicer machinery in human cells (Cole *et al.*, 2009), implying that the transition between the functions of tRNA and an miRNA need not have been a long one.

EXPERIMENTAL PROCEDURES

Plant material

Fragaria vesca 'Hawaii 4' plants were planted in commercial soil and grown in the glasshouse at 25 ± 5°C without additional lighting. Plant tissue (leaves, stolons, flower buds, open flowers and fruits) was harvested from 8-week-old plants between June and August. Each tissue sample comprised a mixture of developmental stages collected on three different dates and at different times of day.

RNA isolation, library preparation and sequencing

For sequencing sRNA of *F. vesca*, a mirPremier microRNA kit (Sigma-Aldrich, http://www.sigmaaldrich.com) was used to extract sRNA individually from five tissues according to the manufacturer's instructions. For stolons, Plant RNA Isolation Aid (ThermoFisher Scientific, http://www.thermofisher.com) was used during the lysis step. sRNA quality was evaluated on an Agilent (http://www.agilent.com) RNA 6000 nanochip, and the quantity was assessed using Ribogreen dye (ThermoFisher scientific) and a Nanodrop 3300 fluorospectrometer (Thermo Scientific, http://www.thermofisher.com). The sRNAs were sequenced in multiplexing mode (1 × 50 bp) on an Illumina HiSeq 2500 (http://www.illumina.com) sequencing system at Fasteris (http://www.fasteris.com). Degradome libraries were constructed by Vertis Biotechnologie (http://www.vertis-biotech.com) using the parallel analysis of RNA ends (PARE) protocol described by German *et al.* (2009). Total RNA was isolated from a pooled sample of leaves, flower buds, open flowers and fruits, and the PARE libraries were sequenced on an Illumina HiSeq 2000 platform. The *F. vesca* sRNA and degradome raw reads were submitted to the NCBI database (http://www.ncbi.nlm.nih.gov) under accession number PRJNA282518.

Computational analysis of sRNAs, and miRNA identification

The raw sequence data were processed using CASAVA 1.8 software (Illumina) to remove the 5′ adapter sequence, while Cutadapt version 1.1 software (Martin, 2011) was used to remove the 3′ adapter sequence. Small RNA sequences between 19 and 24 nt long were then filtered from rRNA, tRNA, snRNA and snoRNA by comparison with the plant/fungal/microbial subset extracted from Rfam 11.0 databases (Burge *et al.*, 2013), and mapped on the *F. vesca* genome sequence version 1.1 (Sargent *et al.*, 2011) without mismatches using the Bowtie aligner version 1.0.0 (Langmead *et al.*, 2009). For genome-wide analysis of the origin and distribution of *F. vesca* sRNAs, the processed reads were mapped on the *F. vesca* pseudochromosomes using Bowtie 2 (Langmead and Salzberg, 2012). The genome was partitioned into 100 kb windows, and the counts for reads of various length within each window were recorded. The normalized abundance of reads was calculated as described in Methods S1. The additional tracks shown in Figure 1 representing genes, LTR transposons or LTR densities, and gene densities are from version 1.1 of the *F. vesca* genome available at the genome database for Rosaceae (www.rosaceae.org). Transposons and LTRs were predicted using LTR_FINDER with default parameters (Xu and Wang, 2007). In order to identify sRNAs corresponding to known and novel miRNAs, we ran the miRNA transcriptome analysis pipeline miRDeep–P (miRDP) (Yang and Li, 2011) with the filtered 19-24 nt sRNA dataset of each library separately. The *F. vesca* genome version 1.1 was used as reference in the pipeline, and the parameters were the default ones except for the length of the candidate precursors, which was set to 400 nt. For selection of known miRNAs and their pre-miRNAs in the miRDP output, the filtered 19-24 nt small RNA dataset was aligned using Bowtie version 1.0.0, with default parameters except for –v 0 as alignment type, to known plant pre-miRNA sequences retrieved from miRBase version 20.0 (www.mirbase.org). Finally, the candidate hairpin structures generated by miRDP were checked visually using the RNA hairpin folding and annotation tool (Moxon *et al.*, 2008b). *F. vesca* miRNA associated with hairpins that fulfilled all the selection criteria comprised the known miRNA dataset. DINAMelt software was used to obtain the pre-miRNA foldback structures (Zuker, 2003). To identify novel miRNA sequences, known miRNAs were first filtered out of the sRNA dataset, and the remaining sequences present in the miRDP output were mapped on *F. vesca* genome version 1.1. The presence of an miRNA* in a predicted position was an additional criterion. miRNA sequences associated with the loci that fulfilled all of the above criteria were considered as novel *F. vesca* miRNAs. Matlab (http://www.mathworks.com) was used to build and visualize the hierarchical clustering of miRNA expression in *F. vesca* tissues. Alignment of miRNAs harboring a PBS complement site was performed using M–coffee (Notredame *et al.*, 2000).

Target identification

The identified known and novel miRNA sequences were introduced into the SeqTar pipeline (Zheng *et al.*, 2011) to identify their potential targets. Raw PARE reads representing the *F. vesca* degradome were processed as described for sRNA reads to remove the adaptor sequences, and only reads of 20 nt in length were used for SeqTar analysis. The 'fvesca_v1.0_genemark_hybrid' annotated gene set predictions obtained from the genome database for Rosaceae (www.rosaceae.org) were used as the transcript database in the SeqTar pipeline. The output of the pipeline was filtered using the following parameters: a valid peak *P* value (P_v) $\leq 10^{-12}$, a mismatch *P* value (P_m) <0.1, and more than 5% reads in the valid peak (percentage >5%). To identify targets of miR1511, BLASTn analysis with parameters optimized for short queries was performed on the *F. vesca* genome version 1.1 with the mature miR1511 sequence. Regions of 300 bp upstream and downstream of each significant alignment at miR1511 pairing site were used for further investigation. These sequences served as putative transcripts in SeqTar analysis with PARE reads, using $P_v \leq 10^{-12}$, $P_m < 0.1$, and a percentage > 5%. An additional selection criterion for this analysis was that targets had to have more than ten PARE reads at the precise cleavage location. In order to annotate the targets, a BLAST search against the NCBI non-redundant database (http://blast.ncbi.nlm.nih.gov) was performed using the selected sequences. This showed that 40 of them had high similarity to LTR retroelements of the *Cassandra* family. The sequences were further analyzed using *LTRharvest* software (Ellinghaus *et al.*, 2008) with default parameters, allowing classifi-

cation of another 15 sequences as LTR retrotransposons. PARE reads mapping solely to individual miR1511 targets were considered unique, and were used in an additional SeqTar run with the same parameters as above. This stringent filtering allowed the valid peak to be described by only unique PARE reads. In a similar way, publically available RNA-seq reads from leaves of *F. vesca* (Kang *et al.*, 2013) were mapped on the 55 transcripts using Bowtie version 1.0.0 (Langmead *et al.*, 2009).

sRNA mapping on LTR retrotransposon targets

The 21 nt sRNA subset of each sRNA library was mapped separately on the 55 LTR target sequences using Bowtie version 1.0.0 with the parameters –v 0 and –m 1. The abundance of each read associated with the mapped positions was then normalized using the total number of 21 nt sRNAs present in that library, resulting in read per million (RPM) values. For each miR1511 target, the mean normalized read abundance was calculated to merge all libraries in the same plot. The same procedure was applied for sRNA mapping on fve–miR393 targets used as a positive control.

ACKNOWLEDGMENTS

The authors thank Claudia Kutter (University of Cambridge, Cambridge cancer centre, United Kingdom) and Dan Sargent for their critical comments. We are grateful to Dan Sargent for providing *F. vesca* Hawaii 4 seeds, Franca Valentini for plant care, and Elisa Asquini for technical support. We also thank Yun Zheng (Fudan University, Institute of Developmental Biology and Molecular Medicine, China) for providing the SeqTar software. Nada Šurbanovski dedicates the discussion of the paper to Professor Nikola Tucić. This work was supported by the Autonomous Province of Trento, Italy (TranscrApple, grandi progetti 2012, to A.S.A).

SUPPORTING INFORMATION

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

Data S1. Degradome t–plots of fve–miR1511 target transcripts.

Data S2. Degradome and RNA coverage on LTR retrotransposons.

Data S3. Abundance of the 21 nt sRNA reads mapping on miR1511 LTR retrotransposon targets.

Methods S1. Bioinformatic analysis of sRNA and miRNA identification and sRNA blot hybridization.

Figure S1. Size distribution of total and distinct reads.

Figure S2. RNA gel-blot hybridization of fve–miR1511 in various *F. vesca* tissues.

Table S1. Number of Illumina reads in *F. vesca* tissues.

Table S2. Abundance of known and novel *F. vesca* miRNAs, and their distribution on pre-miRNAs.

Table S3. Description of loci for known and novel *F. vesca* miRNAs.

Table S4. Target genes of known and novel *F. vesca* miRNAs.

Table S5. Coordinates and sequences of the 74 targets of fve–miR1511.

REFERENCES

Alzohairy, A.M., Sabir, J.S.M., Gyulai, G., Younis, R.A.A., Jansen, R.K. and Bahieldin, A. (2014) Environmental stress activation of plant long-terminal repeat retrotransposons. *Funct. Plant Biol.* **41**, 557–567.

Axtell, M.J. (2013) Classification and comparison of small RNAs from plants. *Annu. Rev. Plant Biol.* **64**, 137–159.

Bennetzen, J.L. and Wang, H. (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu. Rev. Plant Biol.* **65**, 505–530.

Burge, S.W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E.P., Eddy, S.R., Gardner, P.P. and Bateman, A. (2013) Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* **41**, D226–D232.

Chang, W. and Schulman, A.H. (2008) BARE retrotransposons produce multiple groups of rarely polyadenylated transcripts from two differentially regulated promoters. *Plant J.* **56**, 40–50.

Cole, C., Sobala, A., Lu, C., Thatcher, S.R., Bowman, A., Brown, J.W.S., Green, P.J., Barton, G.J. and Hutvagner, G. (2009) Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *RNA*, **15**, 2147–2160.

Creasey, K.M., Zhai, J., Borges, F., Van Ex, F., Regulski, M., Meyers, B.C. and Martienssen, R.A. (2014) miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. *Nature*, **508**, 411–415.

Crepet, W.L., Nixon, K.C. and Gandolfo, M.A. (2004) Fossil evidence and phylogeny: the age of major angiosperm clades based on mesofossil and macrofossil evidence from Cretaceous deposits. *Am. J. Bot.* **91**, 1666–1682.

Ellinghaus, D., Kurtz, S. and Willhoeft, U. (2008) *LTRharvest*, an efficient and flexible software for *de novo* detection of LTR retrotransposons. *BMC Bioinformatics*, **9**, 18.

Fernandez-Medina, R., Ribeiro, J.M., Carareto, C.M., Velasque, L. and Struchiner, C. (2012) Losing identity: structural diversity of transposable elements belonging to different classes in the genome of *Anopheles gambiae*. *BMC Genom.* **13**, 272.

Friant, S., Heyman, T., Wilhelm, M.L. and Wilhelm, F.X. (1996) Extended interactions between the primer $tRNAi^{Met}$ and genomic RNA of the yeast Ty1 retrotransposon. *Nucleic Acids Res.* **24**, 441–449.

Galun, E. (2003) *Transposable Elements. A Guide to the Perplexed and the Novice*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Ge, A., Shangguan, L., Zhang, X., Dong, Q., Han, J., Liu, H., Wang, X. and Fang, J. (2013) Deep sequencing discovery of novel and conserved microRNAs in strawberry (*Fragaria* × *ananassa*). *Physiol. Plant.* **148**, 387–396.

German, M., Luo, S., Schroth, G., Meyers, B. and Green, P. (2009) Construction of parallel analysis of RNA ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. *Nat Protocols*, **4**, 356–362.

Grandbastien, M.A. and Casacuberta, J. (2012) *Plant Transposable Elements*. Berlin/Heidelberg: Springer Verlag.

Haussecker, D., Huang, Y., Lau, A., Parameswaran, P., Fire, A.Z. and Kay, M.A. (2010) Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA*, **16**, 673–695.

Jiao, Y. and Deng, X. (2007) A genome-wide transcriptional activity survey of rice transposable element-related genes. *Genome Biol.* **8**, R28.

Jones-Rhoades, M.W., Bartel, D.P. and Bartel, B. (2006) MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* **57**, 19–53.

Kalendar, R., Tanskanen, J., Chang, W., Antonius, K., Sela, H., Peleg, O. and Schulman, A.H. (2008) Cassandra retrotransposons carry independently transcribed 5S RNA. *Proc. Natl Acad. Sci. U. S. A.* **105**, 5833–5838.

Kang, C., Darwish, O., Geretz, A., Shahan, R., Alkharouf, N. and Liu, Z. (2013) Genome-scale transcriptomic insights into early-stage fruit development in woodland strawberry *Fragaria vesca*. *Plant Cell*, **25**, 1960–1978.

Keeney, J.B., Chapman, K.B., Lauermann, V., Voytas, D.F., Aström, S.U., von Pawel-Rammingen, U., Byström, A. and Boeke, J.D. (1995) Multiple molecular determinants for retrotransposition in a primer tRNA. *Mol. Cell. Biol.* **15**, 217–226.

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25.

Lisch, D. (2013) How important are transposons for plant evolution? *Nat. Rev. Genet.* **14**, 49–61.

Luo, Z., Jin, L. and Qiu, L. (2012) MiR1511 co-regulates with miR1511* to cleave the GmRPL4a gene in soybean. *Chin. Sci. Bull.* **57**, 3804–3810.

Marquet, R., Isel, C., Ehresmann, C. and Ehresmann, B. (1995) tRNAs as primer of reverse transcriptases. *Biochimie*, **77**, 113–124.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 1.

Matzke, M.A. and Mosher, R.A. (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* **15**, 394–408.

Maute, R.L., Schneider, C., Sumazin, P., Holmes, A., Califano, A., Basso, K. and Dalla-Favera, R. (2013) tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. *Proc. Natl Acad. Sci. U. S. A.* **110**, 1404–1409.

Meins, F., Si-Ammour, A. and Blevins, T. (2005) RNA silencing systems and their relevance to plant development. *Annu. Rev. Cell Dev. Biol.* **21**, 297–318.

Meyers, B.C., Axtell, M.J., Bartel, B., Bartel, D.P., Baulcombe, D., Bowman, J.L., Cao, X., Carrington, J.C., Chen, X.M. and Green, P.J. (2008) Criteria for annotation of plant microRNAs. *Plant Cell*, **20**, 3186–3190.

Moxon, S., Jing, R., Szittya, G., Schwach, F., Rusholme Pilcher, R.L., Moulton, V. and Dalmay, T. (2008a) Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome Res.* **18**, 1602–1609.

Moxon, S., Schwach, F., Dalmay, T., MacLean, D., Studholme, D.J. and Moulton, V. (2008b) A toolkit for analysing large-scale plant small RNA datasets. *Bioinformatics*, **24**, 2252–2253.

Notredame, C., Higgins, D.G. and Heringa, J. (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217.

Pantaleo, V., Szittya, G., Moxon, S., Miozzi, L., Moulton, V., Dalmay, T. and Burgyan, J. (2010) Identification of grapevine microRNAs and their targets using high-throughput sequencing and degradome analysis. *Plant J.* **62**, 960–976.

Roberts, J.T., Cooper, E.A., Favreau, C.J. *et al.* (2013) Continuing analysis of microRNA origins. *Mobile Genetic Elements*, **3**, e27755.

Roberts, J.T., Cardin, S.E. and Borchert, G.M. (2014) Burgeoning evidence indicates that microRNAs were initially formed from transposable element sequences. *Mobile Genetic Elements*, **4**, e29255.

Rogers, K. and Chen, X. (2013) Biogenesis, turnover, and mode of action of plant microRNAs. *Plant Cell*, **25**, 2383–2399.

Sargent, D.J., Kuchta, P., Girona, E.L. *et al.* (2011) Simple sequence repeat marker development and mapping targeted to previously unmapped regions of the strawberry genome sequence. *Plant Genome*, **4**, 165–177.

Schulman, A.H. (2013) Retrotransposon replication in plants. *Curr. Opin. Virol.* **3**, 604–614.

Shulaev, V., Sargent, D.J., Crowhurst, R.N. *et al.* (2011) The genome of woodland strawberry (*Fragaria vesca*). *Nat. Genet.* **43**, 109–116.

Simon, S.A. and Meyers, B.C. (2011) Small RNA-mediated epigenetic modifications in plants. *Curr. Opin. Plant Biol.* **14**, 148–155.

Slotkin, R.K. and Martienssen, R. (2007) Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **8**, 272–285.

Subramanian, S., Fu, Y., Sunkar, R., Barbazuk, W.B., Zhu, J.-K. and Yu, O. (2008) Novel and nodulation-regulated microRNAs in soybean roots. *BMC Genom.* **9**, 160.

Wilhelm, M. and Wilhelm, F.X. (2001) Reverse transcription of retroviruses and LTR retrotransposons. *Cell. Mol. Life Sci.* **58**, 1246–1262.

Xia, R., Zhu, H., An, Y.-Q., Beers, E. and Liu, Z. (2012) Apple miRNAs and tasiRNAs with novel regulatory networks. *Genome Biol.* **13**, R47.

Xia, R., Ye, S., Liu, Z., Meyers, B. and Liu, Z. (2015) Novel and recently evolved miRNA clusters regulate expansive F–box gene networks through phasiRNAs in wild diploid strawberry. *Plant Physiol.* **169**, 594–610.

Xu, Z. and Wang, H. (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268.

Xu, X., Yin, L., Ying, Q., Song, H., Xue, D., Lai, T., Xu, M., Shen, B., Wang, H. and Shi, X. (2013) High-throughput sequencing and degradome analysis identify miRNAs and their targets involved in fruit senescence of *Fragaria ananassa*. *PLoS ONE*, **8**, e70959.

Yang, X. and Li, L. (2011) miRDeep–P: a computational tool for analyzing the microRNA transcriptome in plants. *Bioinformatics*, **27**, 2614–2615.

Yin, H., Du, J., Li, L., Jin, C., Fan, L., Li, M., Wu, J. and Zhang, S. (2014) Comparative genomic analysis reveals multiple long terminal repeats, lineage-specific amplification, and frequent interelement recombination for *Cassandra* retrotransposon in pear (*Pyrus bretschneideri* Rehd.). *Genome Biol. Evol.* **6**, 1423–1436.

Zheng, Y., Li, Y.-F., Sunkar, R. and Zhang, W. (2011) SeqTar: an effective method for identifying microRNA guided cleavage sites from degradome of polyadenylated transcripts in plants. *Nucleic Acids Res.* **40**, e28.

Zhu, H., Xia, R., Zhao, B., An, Y., Dardick, C., Callahan, A. and Liu, Z. (2012) Unique expression, processing regulation, and regulatory network of peach (*Prunus persica*) miRNAs. *BMC Plant Biol.* **12**, 149.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415.