Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening

Simon Moxon,[1,3] Runchun Jing,[2,3] Gyorgy Szittya,[2,3] Frank Schwach,[1] Rachel L. Rusholme Pilcher,[2] Vincent Moulton,[1] and Tamas Dalmay[2,4]

[1] School of Computing, University of East Anglia, Norwich NR4 7TJ, United Kingdom; [2] School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, United Kingdom

In plants there are several classes of 21–24-nt short RNAs that regulate gene expression. The most conserved class is the microRNAs (miRNAs), although some miRNAs are found only in specific species. We used high-throughput pyrosequencing to identify conserved and nonconserved miRNAs and other short RNAs in tomato fruit and leaf. Several conserved miRNAs showed tissue-specific expression, which, combined with target gene validation results, suggests that miRNAs may play a role in fleshy fruit development. We also identified four new nonconserved miRNAs. One of the validated targets of a novel miRNA is a member of the CTR family involved in fruit ripening. However, 62 predicted targets showing near perfect complementarity to potential new miRNAs did not validate experimentally. This suggests that target prediction of plant short RNAs could have a high false-positive rate and must therefore be validated experimentally. We also found short RNAs from a *Solanaceae*-specific foldback transposon, which showed a miRNA/miRNA*-like distribution, suggesting that this element may function as a miRNA gene progenitor. The other *Solanaceae*-specific class of short RNA was derived from an endogenous pararetrovirus sequence inserted into the tomato chromosomes. This study opens a new avenue in the field of fleshy fruit biology by raising the possibility that fruit development and ripening may be under miRNA regulation.

Gene expression is highly regulated in plants to ensure proper development and function of tissues and adequate responses to environmental changes. Since gene expression is a multistep process, it can be regulated at several levels. One of the most recently discovered regulatory mechanisms is post-transcriptional and involves 21–24-nt small RNA molecules (sRNA) (Phillips et al. 2007). The sRNA content of plant cells is surprisingly complex, suggesting an extensive regulatory role for these molecules (Lu et al. 2005a). All sRNAs are derived from double-stranded RNA (dsRNA), but dsRNA can be formed through different mechanisms. MicroRNAs (miRNAs) are generated from precursor RNA (pre-miRNA) with hairpin structures by DICER-LIKE 1 (DCL1) (Reinhart et al. 2002). Other sRNAs are produced from dsRNA synthesized by RNA dependent RNA polymerase 6 (RDR6) (*trans*-acting siRNAs) (Peragine et al. 2004; Vazquez et al. 2004), by RDR2 (heterochromatin siRNAs) (Lu et al. 2006), or by overlapping antisense mRNAs (natural antisense siRNAs) (Borsani et al. 2005). It is possible that there are other, unidentified mechanisms leading to dsRNA that could be sources of new classes of sRNAs.

The best-characterized class of plant sRNAs is miRNAs (Jones-Rhoades et al. 2006). The primary transcript (pri-miRNA) is transcribed by RNA polymerase II and contains an imperfect stem–loop secondary structure. DCL1 trims the hairpin structure (pre-miRNA), and then a further cleavage by the same enzyme

releases the miRNA/miRNA* duplex (Kurihara and Watanabe 2004). This duplex has a 2-nt 3'-overhang at each side and contains a few mismatches (Jones-Rhoades et al. 2006). One of the strands of the generated miRNA/miRNA* duplex is incorporated into the RNA-induced silencing complex (RISC). This strand is usually the mature miRNA strand, and the miRNA* strand gets degraded, although in some cases the miRNA* strand also accumulates at a lower level (Jones-Rhoades et al. 2006). The incorporated mature miRNA guides RISC to mRNAs containing a target site, and RISC down-regulates the expression of the mRNA. In plants the target site shows near perfect complementarity to the miRNA sequence, and, as a consequence, most target mRNAs are cleaved by RISC, although there are examples where the translation of the mRNA is suppressed without a cleavage (Chen 2004).

Most plant miRNAs have been identified by the traditional Sanger sequencing method in *Arabidopsis*, rice, and poplar, and comparison of miRNA sequences across plant families has shown that the majority of miRNAs are conserved (Axtell and Bartel 2005). However, some miRNAs appear to be species-specific, and Allen et al. (2004) suggested that these miRNAs have evolved recently ("young" miRNAs), in contrast to the conserved miRNAs ("old" miRNAs). Nonconserved miRNAs are often expressed at a lower level than conserved miRNAs, and this is one of the reasons why small-scale sequencing reveals mainly conserved miRNAs. Development of high-throughput pyrosequencing technology has allowed the discovery of several nonconserved or lowly expressed miRNAs through deep sequencing, for example, in *Arabidopsis* and wheat (Rajagopalan et al. 2006; Fahlgren et al. 2007; Yao et al. 2007). Since most plant developmental processes in-

[3] These authors contributed equally to this work.
[4] Corresponding author.
E-mail t.dalmay@uea.ac.uk; fax 0044-1603-592250.

volve miRNA regulation (Kidner and Martienssen 2005), the discovery of nonconserved miRNAs suggests that plant species/families with specific developmental features may contain nonconserved miRNAs that are involved in the regulation of gene expression specific to those features. We chose fleshy fruit formation and ripening as specific developmental features that are not characteristic of *Arabidopsis*, rice, or poplar. Therefore if miRNAs are involved in these processes, they should probably not be present in these species.

Here we describe the high-throughput sequencing analysis of tomato sRNAs from young fruits and leaves. The longer sRNAs (22, 23, and 24 nt) were found more frequently in fruit than the 21-nt class, but the most abundant class of leaf sRNAs was 21 nt. Most known conserved miRNAs were found in our sRNA libraries, and many of them showed differential expression between leaf and fruit with the accumulation of some of them changing rapidly during fruit development. Targets of known miRNAs were validated, and one of the miRNA-regulated genes that we discovered is a transcription factor involved in ripening. Interestingly, some miRNAs that had been suggested to be specific to *Arabidopsis*, poplar, or moss were also found in tomato. We also found novel tomato-specific miRNAs and validated their targets. One of these belongs to a gene family involved in fruit ripening. We also found a high copy number *Solanaceae*-specific foldback transposon associated with a miRNA/miRNA*-like sRNA pattern and identified sRNAs derived from an endogenous pararetroviral sequence.

Results

Deep sequencing of tomato short RNAs

Two separate sRNA libraries were generated from mixed-size (1–15 mm) green tomato fruits of MicroTom, a miniature rapid-cycling cherry tomato variety (Meissner et al. 1997). In addition, two sRNA libraries were prepared from tissue of young leaves of the same cultivar. The four libraries were sequenced by 454 Life Sciences using pyrosequencing technology that produced 721,874 reads yielding 402,197 and 168,570 sequences from fruits and leaves, respectively, with recognizable adaptor sequences (Table 1). These reads represented around 225,000 and 102,000 unique sRNA sequences in fruits and leaves, respectively. In both tissues, the 21-nt and 22-nt classes showed the highest degree of redundancy (Supplemental Fig. 1A,B), suggesting that sRNAs in these size classes are often produced from precursors from which clearly defined mature short sequences are excised. These sRNAs are often miRNAs and *trans*-acting siRNAs (ta-siRNAs) that are usually expressed at a high level (Vaucheret 2006). The 23-nt and 24-nt classes were much less redundant (Supplemental Fig. 1A,B), indicating that they derive from loci that produce heterogeneous sRNA populations, such as those found associated with RNA polymerase IV-dependent pathways in *Arabidopsis* that produce heterochromatin-related siRNAs. To compare sequence redundancy levels in samples of different sizes, we normalized the larger fruit sample to the number of reads in the leaf sample by extracting 1000 random subsets of 159,886 reads from the fruit sample. Supplemental Figure 1, C

Table 1. Statistics of sRNAs sequences from tomato fruit and leaf

	Reads	Matching BACv175	Unique sequences	Matching BACv175
Fruit				
Raw reads	537,036			
Adaptors removed	402,197	79,099	224,823	39,001
rRNA/tRNA exact matches removed	391,119	75,353	222,391	38,064
Match known miRNAs	14,536	3974	588	66
sRNAs mapping to predicted hairpins	3909	3909	409	409
Predicted hairpins sRNA abundance ≥ 3	3579	3579	128	128
Leaf				
Raw reads	184,838			
Adaptors removed	168,570	35,419	102,753	18,180
rRNA/tRNA exact matches removed	159,886	32,535	100,168	17,032
Match known miRNAs	18,352	5488	780	89
sRNAs mapping to predicted hairpins	4019	4019	235	235
Predicted hairpins sRNA abundance ≥ 3	3766	3766	30	30

and D, shows size distributions of the leaf sample in comparison to the random average of the normalized fruit samples. The distribution of redundant sequences for different size classes was similar in fruits and leaves (Supplemental Fig. 1C). However, the size distribution of nonredundant sRNAs was slightly different in the two tissues (Supplemental Fig. 1D). The nonredundant leaf sRNA distribution showed a peak at 21 nt, while there were more nonredundant fruit sRNAs of 22, 23, or 24 nt than of 21 nt. Assuming that the overall amount of 24-nt sRNA is related to the extent of transcriptional regulation, this observation suggests a more extensive regulation of gene expression by sRNAs at the transcriptional level in fruit than in leaf. This is probably because the longer sRNAs are often associated with DNA methylation and heterochromatin formation.

Known miRNAs

We searched for known miRNAs in our combined (fruit and leaf) tomato sRNA database and found 7912 redundant sequences matching 20 known miRNA families (Supplemental Tables 1,d 2; Supplemental Fig. 2). In addition, we identified 25,436 sequences that were either shorter/longer or contained up to two mismatches to the same 20 and another 10 known miRNA families. One of these, the algae-specific miR1151, gave a negative result by Northern blot analysis and was probably an artifact. However, we were able to confirm the expression of two miRNAs in tomato by Northern blot analysis that had previously been thought to be specific to *Arabidopsis* (miR858; Fahlgren et al. 2007) and moss (miR894; Fattash et al. 2007), respectively (Fig. 2). We also confirmed our previous observation (Pilcher et al. 2007) that miR482 (originally reported to be poplar-specific [Lu et al. 2005b]) is also expressed in tomato. These three examples show that miRNAs previously believed to be species- or family-specific can exist in several families. Data from more species are necessary to understand the evolution of these less conserved miRNAs.

We analyzed the expression levels of 13 additional known miRNAs that were present in our libraries and that had not been examined in our previous study (Pilcher et al. 2007) using Northern blot assays of samples from leaves, closed flower buds, and four different stages of fruits (Fig. 1). All tested miRNAs, except for miR165/166, miR403, and miR472, showed differential expression patterns in these tissues. Several miRNAs (miR156/157, miR164, miR408, miR858, and miR894) were more abundant in leaves and closed flowers than in fruits. In contrast, miR169 was expressed at a higher level in all fruit stages than in closed flowers, and it was almost undetectable in leaves. Intriguingly, two
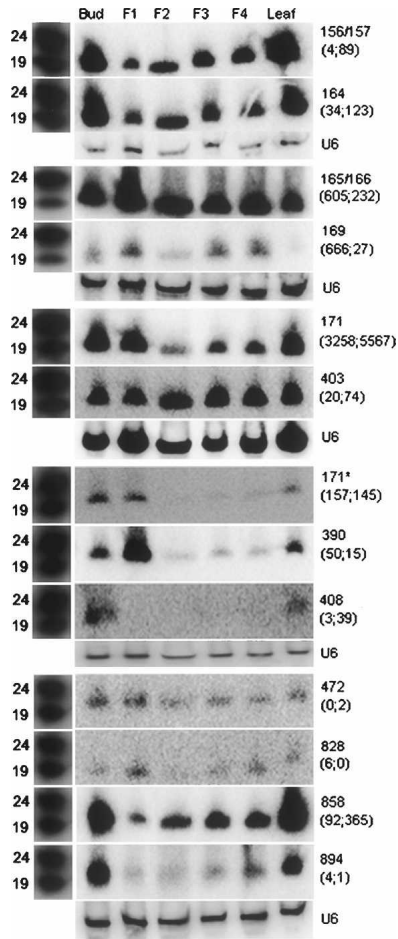
Figure 1. Expression of conserved tomato miRNAs. Total RNA from different tissues was extracted, separated, and transferred to membranes. The membranes were hybridized to miRNA-specific probes or a U6-specific probe (shown on the *right*) to demonstrate equal loading. Membranes were stripped and reprobed; equal loading is shown once for each membrane. Numbers between brackets indicate the number of sequences found in the fruit (*left*) and leaf (*right*) libraries for each miRNA. Different size fruits were used for RNA extraction: F1, 1–3 mm; F2, 5–7 mm; F3, 7–11 mm; F4, 11–14 mm. Size markers on the *left* are 24- and 19-nt RNA oligonucleotides.

known miRNAs showed differential expression between different fruit stages. miR171 (and miR171*) was as highly expressed in very small fruits (1–3 mm) as in leaves and closed flowers, and it accumulated at a lower level in more mature fruits. Interestingly, miR390 had much higher accumulation in very small fruits than in leaves and closed flowers, and it accumulated at a very low level in more mature fruits. This suggests that miR390 has a specific role in early fruit formation.

Several target genes of known miRNAs have been validated in *Arabidopsis*, rice, and poplar. However, it is not obvious which genes are targeted by these miRNAs in tomato because annotation of the partial genome sequence is not complete. In addition, Itaya et al. (2007) could only validate one out of three conserved miRNA target tomato genes (the miR172 targeted APETALA 2). We used the tomato Unigene EST database (http://sgn.cornell. edu/) to predict 12 targets that were all validated by 5′-RACE (rapid amplification of cDNA ends) assays (Fig. 2). Two targets are worth describing specifically—a MYB transcription factor

that is targeted by miR858 (originally thought to be *Arabidopsis*-specific) (Fahlgren et al. 2007) and colorless non-ripening (CNR), a member of the squamosa-promoter binding protein (SBP) family that was shown to be involved in fruit ripening (Manning et
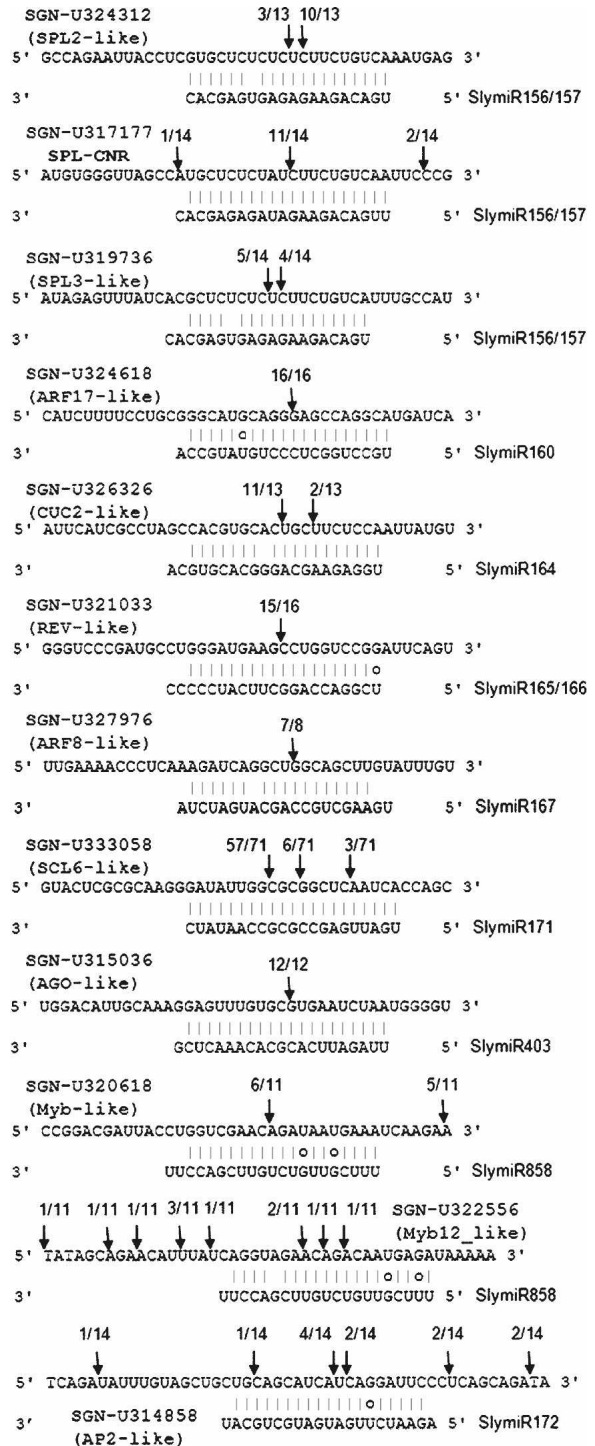

Figure 2. Target validation of conserved tomato miRNAs. 5′-RACE analysis was carried out for each predicted target gene. (Arrows) The 5′-ends of cleavage products. Cleavage sites outside of the displayed sequence are not shown. Target EST sequences are shown on *top* of the miRNA sequences.

al. 2006). *CNR* is targeted by miR156/157, which, for the first time, implies miRNA involvement in the maturation process of a commercially important fruit. We analyzed the expression pattern of miR157 during fruit ripening to test whether there was a negative correlation between the accumulation of *CNR* mRNA and miR157. *CNR* expression is up-regulated at the breaker stage (Manning et al. 2006), but miR157 accumulation did not show a significant change during fruit ripening (data not shown).

Novel miRNAs

As mentioned above, tomato genome sequencing is not yet complete, although many genomic BAC sequences are available (http://sgn.cornell.edu/about/tomato_sequencing.pl). We used version BACv175 (unfinished) for our analysis, which represents ~25% of the tomato genome. sRNA sequences that were not known miRNAs were mapped to BAC sequences. Secondary structures were predicted for each locus, and the ones that fulfilled the hairpin structure criteria described by Jones-Rhoades et al. (2006) were selected as candidate miRNAs. This analysis resulted in 219 (165 unique) candidates, 87 of which also had a predicted target in at least one tomato EST sequence. We also looked for sequenced miRNA* sequences, but most of the potential mature miRNAs were sequenced less than five times; therefore, no miRNA* sequences were found. (The average frequency of miRNA* is ~10% of the frequency of mature miRNA [Rajagopalan et al. 2006].) However, one sequence was found 19 times, and a potential miRNA* was sequenced nine times in our combined data sets. According to the criteria published by Rajagopalan et al. (2006), this sequence is the first novel bona fide miRNA (sly-miR1919) identified in tomato. The other miRNA candidates were further tested by Northern blot (miRNA) and 5′-RACE assay (target). Northern blot analysis was carried out for 92 candidates with hairpin structure but without sequenced miRNA*. Fifty-one were detected as discrete bands around 21 nt (Supplemental Fig. 3), and several showed differential expression in different tissues (Fig. 3). Three sRNAs showed very strong leaf specificity, one sRNA accumulated at higher level in closed flowers and all stages of fruits than in leaves, and seven sRNAs showed increasing levels of expression during fruit development, suggesting the possibility that several genes involved in fruit development are regulated by short RNAs.

Many miRNA candidates had predicted targets. We carried out 5′-RACE assays for 65 predicted targets, but most of them (62) could not be validated (Supplemental Table 3). Therefore, these sRNAs cannot be classified as miRNAs because, although they were sequenced, produced from stable hairpins, and accumulated as 21-nt RNAs, no miRNA* was sequenced, no target cleavage was shown, and their accumulation in a *dcl1* mutant could not be studied because of the lack of such a mutant in tomato. However, we strongly suspect that several of these potential miRNAs are bona fide miRNAs, although at least one of the abovementioned criteria would have to be shown to hold before they could be classified as miRNAs.

In addition to sly-miR1919, we found three new tomato miRNAs (secondary structures and accumulation patterns are shown in Supplemental Figs. 4 and 5, respectively). Although no miRNA* sequences were found for these three miRNAs, their predicted target genes were validated by 5′-RACE analysis (Fig. 4B). The target genes of sly-miR1917 are the splice variants of constitutive triple response 4 (*LeCTR4sv1* and *LeCTR4sv2*), which is a member of the CTR family, which are key negative regulators of ethylene responses (Adams-Phillips et al. 2004). Both sly-
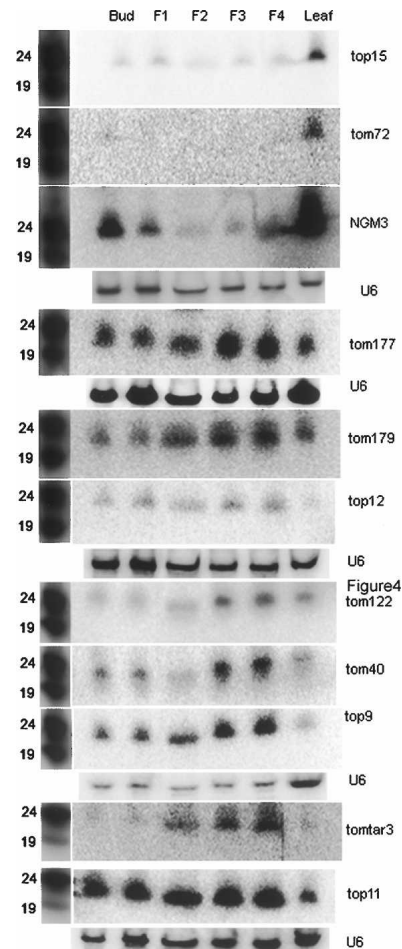

Figure 3. Differentially expressed tomato short RNAs. Probes specific to potential miRNAs (tom72, NGM3, tom177, tom179, tom122, tom40, and tomtar3) or short RNAs that could not be mapped to the available genome sequence but cloned many times (top15, top12, top9, and top11) were hybridized to the same membranes shown on Figure 2. Sequences of short RNAs are provided in Supplemental Figure 2. The U6-specific probe shows equal loading. Different size fruits were used for RNA extraction: F1, 1–3 mm; F2, 5–7 mm; F3, 7–11 mm; F4, 11–14 mm. Size markers on the *left* are 24- and 19-nt RNA oligonucleotides.

miR1918 and sly-miR1916 target unknown protein-expressing mRNAs (ESTs SGN-U326398 and SGN-U322371, respectively) that do not show any homology with annotated genes in the EMBL sequence database. Accumulation of the new miRNAs was analyzed by Northern blot, and sly-miR1918 and sly-miR1919 showed significantly stronger expression in fruit than in leaf or flower bud (Fig. 4A). In fact, these two miRNAs accumulated at a higher level in more mature fruit than in very young fruit. sly-miR1917 produced a consistently weak signal, and it was necessary to use an LNA (locked nucleic acid) probe to reveal stronger accumulation in more mature fruit. sly-miR1916 is expressed at a similar level in all analyzed tissues (Fig. 4A). The complete genome sequences of *Arabidopsis*, rice, and poplar were interrogated for miRNA genes homologous to the four new tomato miRNAs, but no perfect matches were found. A few loci were found with two mismatches, but none of these exhibited a hairpin structure with their flanking regions (data not shown). We therefore concluded that the new tomato miRNAs were not conserved in these species.
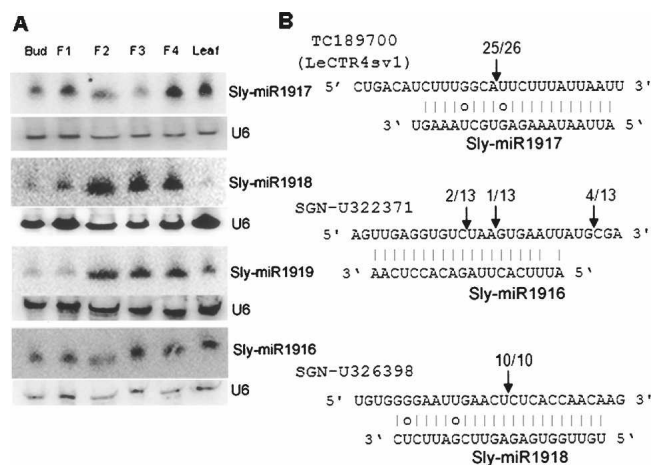
Figure 4. Expression and target validation of new nonconserved tomato miRNAs. (A) Northern blot analysis of new miRNAs showed that sly-miR1918 and sly-miR1919 accumulate preferentially in the fruit. The U6 probe was used to show equal loading. Different size fruits were analyzed: F1, 1–3 mm; F2, 5–7 mm; F3, 7–11 mm; F4, 11–14 mm. (B) The result of target validation for three new miRNAs. (Arrows) The 5′ ends of cleavage products mapped inside the displayed sequence. Target EST sequences are shown on *top* of the miRNA sequences.

Other tomato-specific sRNAs

Most sequenced *Arabidopsis* sRNAs are 24 nt and derived from transposons and other repeats (Rajagopalan et al. 2006; Fahlgren et al. 2007; Mosher et al. 2008). We found that the 24-nt class of sRNAs was also generally abundant in tomato, especially in fruit. However, the abundance of sRNA sequences from one particular class of transposons was exceptional: 9280 sequences were derived from type III foldback transposon tomato anionic peroxidase inverted repeat (TAPIR) (Hong and Tucker 1998). TAPIRs are ~280-nt-long inverted repeats, often located adjacent to genes (Mao et al. 2001). This element has a high copy number; we found 468 copies in the available genome sequence (25% of the genome) (Supplemental Fig. 6). Although sRNAs are usually well dispersed over transposon sequences, we found several sRNAs that mapped to TAPIRs derived from specific locations of the inverted repeat (Fig. 5A). In fact, most TAPIR loci produced sRNAs predominantly from two regions that were opposite to one another on the two stems of the hairpin structure (reminiscent of potential miRNA/miRNA* pairs). However, the most abundant sequences from the two regions did not form a precise miRNA/miRNA* duplex (Fig. 5A). We tried to compare the distribution of sRNAs derived from TAPIR to the accumulation pattern of sRNAs derived from a type III foldback transposons in *Arabidopsis* (hairpin elements) (Adé and Belzile 1999), but the published *Arabidopsis* sRNA databases contained almost no sRNA sequences derived from hairpin elements (Qi et al. 2006; Rajagopalan et al. 2006; Fahlgren et al. 2007; Mosher et al. 2008). Next we searched for hairpins in the *Arabidopsis* genome that are longer than 200 nt and produce sRNAs. The sRNA pattern of these loci was different from TAPIRs; sRNAs were scattered across the whole hairpin and were absent in *dcl4* plants. Next, we compared the positions of the most abundant sRNAs derived from different TAPIR loci. The most abundant sRNA sequence was different in some TAPIR elements, although some loci produced the same major sRNA. Northern blot analysis of the two most abundant TAPIR sRNAs gave a slightly different expression

pattern in spite of the fact that they were shifted only by 2 nt. TAPIR1 accumulated at a slightly higher level in leaf than fruit, and TAPIR2 was more abundant in fruit than in leaf (Fig. 5B).

The other new class of sRNAs, which was not found in libraries from other species, derived from endogenous pararetroviral (EPRV) sequences. Several DNA viruses were found integrated into the host genome, some of which can cause infection and some not (Harper et al. 2002). An EPRV was recently described in tomato that was proposed to be controlled by RNA silencing through sRNAs (Staginnus et al. 2007). Several sRNA sequences matched an integrated EPRV sequence, but, surprisingly, they were not randomly distributed. One particular sequence (EPRV1) was found with a very high frequency in all four libraries in addition to a few less abundant hotspots (Supplemental Fig. 7). Although Northern blot analysis confirmed the accumulation of EPRV1 and two other less abundant EPRV-specific siRNAs, the very high frequency of EPRV1 was not reflected by the signals (Supplemental Fig. 8). In fact, EPRV3 was easier to detect than EPRV1. Expression analysis of EPRV-specific siRNAs in different cultivars and wild species showed that their expression varies in different accessions, although integrated copies were detected in all of them (Staginnus et al. 2007).
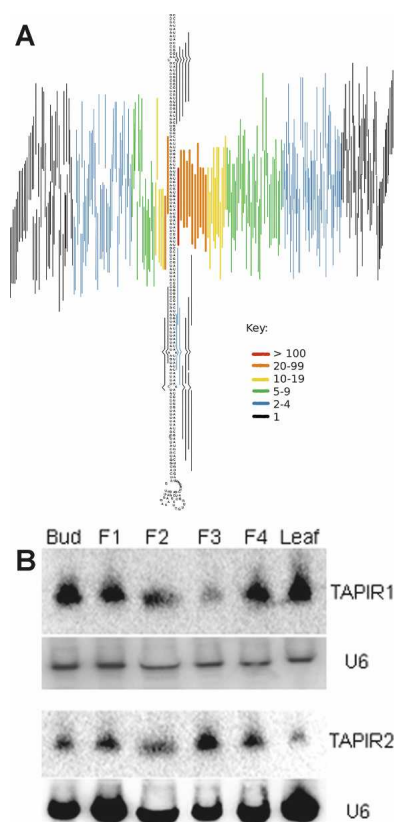

Figure 5. TAPIR-derived sRNAs. (A) Predicted secondary structure of one particular TAPIR element with lines representing the sRNA sequences mapping to the two arms of the hairpin. The color of the lines specifies the abundance of the sequences in the library. (B) Northern blot shows the accumulation of the two most abundant, overlapping sRNAs (exact positions are shown in Supplemental Fig. 6) from TAPIR elements. Membranes were stripped and reprobed for U6 to show equal loading. F1, 1–3 mm; F2, 5–7 mm; F3, 7–11 mm; F4, 11–14 mm.

Discussion

Conserved miRNAs in tomato

We generated sRNA libraries from fruit and leaf of tomato plants, and most conserved miRNA families were found in at least one of our sRNA libraries. Several conserved miRNAs showed differential expression in leaf, flowering bud, and fruits at different stages that could provide information about their function. sly-miR169 was preferentially accumulated in flower buds and fruits and was hardly detectable in leaves (Fig. 1). The only known target of this miRNA is a transcription factor of the CCAAT-binding family, HAP2, and this protein was shown to have an important role during nodule development in *Medicago truncatula* (Combier et al. 2006). HAP2 is also required for pollen tube guidance and fertilization (von Besser et al. 2006) and affects flowering time (Wenkel et al. 2006) in *Arabidopsis*. It remains to be seen whether HAP2 also plays a role in fruit development or if sly-miR169 can target other genes in tomato. Another miRNA highly expressed in fruits is sly-miR390. In fact, its expression sharply peaks in very young fruits (1–3 mm) (Fig. 1), suggesting that it plays a role in early fruit development. sly-miR390 is also expressed at a lower level in leaves, which is in line with its known function in *Arabidopsis*, where it controls leaf morphology through targeting *TAS3* (Adenot et al. 2006). Cleaved *TAS3* gives rise to ta-siRNAs targeting mRNAs in the AUXIN RESPONSE FACTOR (ARFs) family (Allen et al. 2005), and it will be interesting to see whether *TAS3* or other *TAS* gene-derived ta-siRNAs are involved in early fruit development (a *TAS3* homolog is present in the tomato genome). In contrast to sly-miR390 and sly-miR169, sly-miR894 was hardly detectable in fruits but accumulated at a high level in flower buds and leaves (Fig. 1). Interestingly, this miRNA was only found in moss previously, but no target gene was identified (Fattash et al. 2007). sly-miR408 was also absent in fruits, and it does not have any validated targets, although it was predicted to target plantacyanin genes (Sunkar and Zhu 2004). These examples show that target identification and validation of conserved miRNAs in different species is still important because miRNAs may have additional functions to the regulation of conserved target genes. Axtell et al. (2007) demonstrated that different sRNAs in different species can have similar functions, and it is also possible that conserved miRNAs would have different targets in different species. New targets for conserved miRNAs may be found in tomato after genome sequencing is completed, but at this time we could only use the available Unigene sequences downloaded from the SOL Genomics Network (Mueller et al. 2005). Twelve target genes were validated for nine conserved miRNAs (Fig. 2). One of these miRNAs, sly-miR858, is not present in the poplar and rice genome (Rajagopalan et al. 2006) but was found in *Arabidopsis*, where it targets the MYB12 transcription factor. We identified two MYB12-like genes in tomato, and both of them showed the same level of similarity to the *Arabidopsis* gene. One of them validated better (SGN-U320618) than the other (SGN-U322556), which showed a scattered cleavage pattern around the canonical cleavage site. Another transcription factor family regulated by miRNAs is the squamosa-promoter binding protein (SBP) family. The tomato members of this family are called SBP-like proteins (SPL). We validated three SPL genes targeted by sly-miR156/157; one of them is CNR, a key gene in fruit ripening (Manning et al. 2006). SBP genes are targeted by miR156/157 in *Arabidopsis*, but none of those is involved in fruit development since *Arabidopsis* does not have a

fleshy fruit. The precise biological role of miR156/157 in tomato needs further investigation. *CNR* mRNA is up-regulated in fruit at the breaker stage (Manning et al. 2006), but our Northern blot analysis did not show a decrease in the expression of miR157 at the breaker stage. It is possible that the expression of miR157 and *CNR* overlaps only partially and the function of the miRNA is to suppress *CNR* expression in specific cell types. Although the accumulation of a target mRNA and miRNA is expected to show a negative correlation, there are examples indicating that this is not always the case. Achard et al. (2004) demonstrated that miR159 targeted *MYB33*, but their accumulation did not show a negative correlation. Llave et al. (2002) reported a positive correlation between the levels of miR171 (called "miR39" in this study) and its target gene mRNA. The cleavage product of *CNR* shows that it is cleaved by miR156/157, and it remains to be seen what the consequences are if this cleavage does not take place.

Classification of nonconserved miRNAs

Recently various publications have reported high-throughput sequencing of sRNAs from *Arabidopsis* (Lu et al. 2005a, 2006; Rajagopalan et al. 2006; Fahlgren et al. 2007; Mosher et al. 2008) and other plant species (Axtell et al. 2007; Barakat et al. 2007a,b; Molnár et al. 2007; Yao et al. 2007; Morin et al. 2008). The common theme emerging from these reports is that the sRNA content of plants is very complex and, although a subset of sRNAs is conserved across different families, several sRNAs are specific to each species or family. The most conserved class of sRNAs is the miRNA class, but even these are not all conserved. These observations led to a change in the minimum criteria for classifying a sRNA as a miRNA initially set up by Ambros et al. (2003). On one hand, the conservation is not required anymore, but, on the other hand, it became apparent that a lot of loci that express sRNA can be folded into a stem–loop structure. This prompted Jones-Rhoades et al. (2006) to introduce new criteria to avoid the flooding miRBase with sequences that are not miRNAs. In particular, the conservation criterion was replaced with proof of biogenesis (demonstration of DCL1 dependency or cloning of perfect miRNA* sequences) or functional data (target validation by 5'-RACE). However, this criterion was not verified in several recent studies partly because a *dcl1* mutant was not available for species other than *Arabidopsis*. Since sequence complementarity between miRNAs and their target genes is very high in plants, target validation has been increasingly overlooked, and several recent studies have considered target prediction as sufficient functional data. This is probably due to the fact that, at least initially, all predicted targets that were experimentally tested proved to be real targets. However, most validated targets are recognized by conserved miRNAs and the predicted targets of most nonconserved miRNAs have never been tested experimentally. Here we show that most predicted targets of putative nonconserved miRNAs could not be validated experimentally, in contrast to the high validation rate of targets of conserved miRNAs. There are several possible explanations for negative 5'-RACE results, such as, the target genes are not expressed in the same cells as the putative miRNAs, or the cleavage product is not stable enough. However, it is more likely that many of the putative miRNAs are false-positive predictions and not true miRNAs. They are expressed and could derive from hairpin structure precursors, but it is now clear that these criteria hold for many thousands of loci in plant genomes, and it does not necessarily mean that they do derive from single-stranded stem–loop structures. In the absence of biogenesis data, it has to be shown that the po-

tential miRNAs mediate cleavage of mRNAs in order to classify them as miRNAs (Jones-Rhoades et al. 2006). Our observation suggests that some of the recently published nonconserved miRNAs predicted by high-throughput sequencing projects have to be considered cautiously. Many proposed nonconserved miRNAs, which are not supported by biogenesis data (demonstration of DCL1 dependency or cloning of perfect miRNA* sequences), could be siRNAs and not miRNAs.

We validated cleavage of three novel targets mediated by new nonconserved tomato miRNAs (Fig. 4), although one of them produced a less precise cleavage pattern. Sly-miR1917 cleaved at the twelfth position of the target site instead of the canonical tenth position. Since this miRNA was represented by only 11 (identical) sequences in our library, we cannot rule out that other forms of the miRNA also accumulate. We did not find a sequence shifted 2 nt upstream, but we did find a few sequences derived several nucleotides upstream or downstream of the sequence shown in Figure 4 (Supplemental Fig. 5), suggesting that different forms of the miRNA are produced from this locus. A larger-scale library generated in the future may find the form that would cause cleavage of this target gene at the position identified by our 5′-RACE. A fourth new miRNA was validated by cloning of the perfect miRNA* sequence. One of the novel targets is a member of the CTR gene family that suppresses ethylene response and is involved in fruit ripening (Adams-Phillips et al. 2004). This result, together with the regulation of CNR by sly-miR156/157, opens a new avenue in the field of gene expression regulation during fruit development and ripening.

Can some miRNA genes derive from transposons?

Transposon-specific sRNAs are usually abundant in sRNA libraries, but sRNAs derived from type III foldback transposon TAPIR sequences (Hong and Tucker 1998) were exceptionally highly represented in the two tomato sRNA libraries. TAPIR elements are flanked by nine nucleotide target site duplications (Supplemental Fig. 6), and they are mobile (Mao et al. 2001). However, sRNAs are not well dispersed over TAPIR elements like on other transposons. Instead, they map to specific positions that would be on opposite arms of a hairpin structure if the TAPIR element is expressed (Fig. 5). The most abundant sRNA species are similar to miRNA/miRNA* pairs, but they do not precisely pair with each other. Moreover, there are less frequently sequenced sRNAs around the most abundant species, although the pattern of sRNAs is likely to be less complex than it is shown in Figure 5. Some sRNAs may not be produced from the locus shown, but are instead produced from another TAPIR locus. Owing to the high degree of sequence similarity between TAPIR elements, the sRNA will map to both loci, resulting in a complex distribution of sRNAs on the hairpin. The sRNA pattern of TAPIRs is reminiscent of ath-miR822 and ath-miR839, two DCL4 processed nonconserved miRNAs (Rajagopalan et al. 2006). However, those miRNAs are single-locus genes in Arabidopsis, and we found 468 copies of TAPIR in the available 25% of the tomato genome. We tried to predict target genes for TAPIR-derived sRNAs, but our analysis only found other TAPIR elements (data not shown). It is tempting to propose that TAPIRs are potential progenitors of miRNA genes and that if a TAPIR-derived sRNA acquires a target gene and this regulation is beneficial for the plant, it could be fixed. Because of mutations, the miRNA-producing TAPIR element could lose mobility and eventually become a proper miRNA gene. It was shown recently that several human miRNAs

derive from transposable elements (Piriyapongsa and Jordan 2007; Piriyapongsa et al. 2007) that support our prediction. According to the current model, miRNA genes are derived from target genes through duplication, inversion, and mutations, and this is well supported by the fact that nonconserved miRNA genes often have low copy number and show extensive complementarity to the target gene beyond the mature miRNA sequence (Allen et al. 2004). Transposons could be an alternative source of miRNA genes, especially in plants with large genomes and high copy numbers of foldback transposons.

Methods

Cloning of small RNAs, Northern blot, and 5′-RACE analysis

Total RNA was extracted from tomato leaf, bud before flower blooming, and different developmental stages of whole fruits. Small RNA between 19 and 24 nt were cloned from leaf and fruit (mixture of different sizes between 1 and 15 mm) as described by Pilcher et al. (2007). Briefly, the sRNA fraction was purified and ligated to adaptors without dephosphorylating and rephosphorylating the sRNA. The RNA was converted to DNA by RT-PCR, and the DNA was sequenced by 454 Life Sciences. Twenty micrograms of each total RNA sample was used for Northern blot analysis as described by Pall et al. (2007). 5′-RACE analysis was carried out using poly(A) plus fraction and the GeneRacer kit (Invitrogen).

Bioinformatics analysis

Small RNA sequences were extracted from raw reads matching both the last 7 nt of the 5′-adaptor and the first 7 nt of the 3′-adaptor sequences. Sequences were then queried against ribosomal and transfer RNAs from Rfam (http://www.sanger.ac.uk/Software/Rfam/), the Arabidopsis tRNA database (http://lowelab.ucsc.edu/GtRNAdb/Athal/), and rRNA sequences obtained from EMBL using their SRS service (http://srs.ebi.ac.uk/). Any sRNAs having exact matches to these sequences were excluded from genomic mapping.

Reads of 18–30 nt were mapped to tomato BAC sequences (bacs.v175.seq) obtained from the SOL Genomics Network ftp://ftp.sgn.cornell.edu/tomato_genome/bacs using exact matching. sRNAs were then folded using RNAfold (http://www.tbi.univie.ac.at/~ivo/RNA/), and their structure was analyzed using miRCat (http://srna-tools.cmp.uea.ac.uk/mircat/). Target predictions were performed based on methods described by Allen et al. (2005). The sequences of all predicted targets are shown in Supplemental Figure 9.

Acknowledgments

References

Achard, P., Herr, A., Baulcombe, D.C., and Harberd, N.P. 2004. Modulation of floral development by a gibberellin-regulated microRNA. Development **131**: 3357–3365.
Adams-Phillips, L., Barry, C., Kannan, P., Leclercq, J., Bouzayen, M., and

Giovannoni, J. 2004. Evidence that CTR1-mediated ethylene signal transduction in tomato is encoded by a multigene family whose members display distinct regulatory features. *Plant Mol. Biol.* **54:** 387–404.

Adé, J. and Belzile, F.J. 1999. Hairpin elements, the first family of foldback transposons (FTs) in *Arabidopsis thaliana*. *Plant J.* **19:** 591–597.

Adenot, X., Elmayan, T., Lauressergues, D., Boutet, S., Bouche, N., Gasciolli, V., and Vaucheret, H. 2006. DRB4-dependent TAS3 *trans*-acting siRNAs control leaf morphology through AGO7. *Curr. Biol.* **16:** 927–932.

Allen, E., Xie, Z., Gustafson, A.M., Sung, G.H., Spatafora, J.W., and Carrington, J.C. 2004. Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat. Genet.* **36:** 1282–1290.

Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C. 2005. microRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. *Cell* **121:** 207–221.

Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., et al. 2003. A uniform system for microRNA annotation. *RNA* **9:** 277–279.

Axtell, M.J. and Bartel, D.P. 2005. Antiquity of microRNAs and their targets in land plants. *Plant Cell* **17:** 1658–1673.

Axtell, M.J., Snyder, J.A., and Bartel, D.P. 2007. Common functions for diverse small RNAs of land plants. *Plant Cell* **19:** 1750–1769.

Barakat, A., Wall, K., Leebens-Mack, J., Wang, Y.J., Carlson, J.E., and Depamphilis, C.W. 2007a. Large-scale identification of microRNAs from a basal eudicot (*Eschscholzia californica*) and conservation in flowering plants. *Plant J.* **51:** 991–1003.

Barakat, A., Wall, K.P., Diloretto, S., Depamphilis, C.W., and Carlson, J.E. 2007b. Conservation and divergence of microRNAs in *Populus*. *BMC Genomics* **8:** 481. doi: 10.1186/1471-2164-8-481.

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. 2005. Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* **123:** 1279–1291.

Chen, X. 2004. A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development. *Science* **303:** 2022–2025.

Combier, J.P., Frugier, F., de Billy, F., Boualem, A., El-Yahyaoui, F., Moreau, S., Vernié, T., Ott, T., Gamas, P., Crespi, M., et al. 2006. MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. *Genes & Dev.* **20:** 3084–3088.

Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Law, T.F., Grant, S.R., Dangl, J.L., et al. 2007. High-throughput sequencing of *Arabidopsis* microRNAs: Evidence for frequent birth and death of *MIRNA* genes. *PLoS ONE* **2:** e219. doi: 10.1371/journal.pone.0000219.

Fattash, I., Voss, B., Reski, R., Hess, W.R., and Frank, W. 2007. Evidence for the rapid expansion of microRNA-mediated regulation in early land plant evolution. *BMC Plant Biol.* **7:** 13. doi: 10.1186/1471-2229-7-13.

Harper, G., Hull, R., Lockhart, B., and Olszewski, N. 2002. Viral sequences integrated into plant genomes. *Annu. Rev. Phytopathol.* **40:** 119–136.

Hong, S.B. and Tucker, M.L. 1998. Genomic organization of six tomato polygalacturonases and 5′ upstream sequence identity with *tap1* and *win2* genes. *Mol. Gen. Genet.* **258:** 479–487.

Itaya, A., Bundschuh, R., Archual, A.J., Joung, J.G., Fei, Z., Dai, X., Zhao, P.X., Tang, Y., Nelson, R.S., and Ding, B. 2007. Small RNAs in tomato fruit and leaf development. *Biochim. Biophys. Acta* **1779:** 99–107.

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.

Kidner, C.A. and Martienssen, R.A. 2005. The developmental role of microRNA in plants. *Curr. Opin. Plant Biol.* **8:** 38–44.

Kurihara, Y. and Watanabe, Y. 2004. *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc. Natl. Acad. Sci.* **101:** 12753–12758.

Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. 2002. Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* **297:** 2053–2056.

Lu, C., Tej, S.S., Luo, S., Haudenschild, C.D., Meyers, B.C., and Green, P.J. 2005a. Elucidation of the small RNA component of the transcriptome. *Science* **309:** 1567–1569.

Lu, S., Sun, Y.H., Shi, R., Clark, C., Li, L., and Chiang, V.L. 2005b. Novel and mechanical stress-responsive microRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*. *Plant Cell* **17:** 2186–2203.

Lu, C., Kulkarni, K., Souret, F.F., MuthuValliappan, R., Tej, S.S., Poethig, R.S., Henderson, I.R., Jacobsen, S.E., Wang, W., Green, P.J., et al. 2006. MicroRNAs and other small RNAs enriched in the *Arabidopsis*

RNA-dependent RNA polymerase-2 mutant. *Genome Res.* **16:** 1276–1288.

Manning, K., Tör, M., Poole, M., Hong, Y., Thompson, A.J., King, G.J., Giovannoni, J.J., and Seymour, G.B. 2006. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat. Genet.* **38:** 948–952.

Mao, L., Begum, D., Goff, S.A., and Wing, R.A. 2001. Sequence and analysis of the tomato JOINTLESS locus. *Plant Physiol.* **126:** 1331–1340.

Meissner, R., Jacobson, Y., Melame, S., Levyatuv, S., Shalev, G., Ashri, A., Elkind, Y., and Levy, A. 1997. A new model system for tomato genetics. *Plant J.* **12:** 1465–1472.

Molnár, A., Schwach, F., Studholme, D.J., Thuenemann, E.C., and Baulcombe, D.C. 2007. miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. *Nature* **447:** 1126–1129.

Morin, R.D., Aksay, G., Dolgosheina, E., Ebhardt, H.A., Magrini, V., Mardis, E.R., Sahinalp, S.C., and Unrau, P.J. 2008. Comparative analysis of the small RNA transcriptomes of *Pinus contorta* and *Oryza sativa*. *Genome Res.* **18:** 571–584. doi: 10.1101/gr.6897308.

Mosher, R.A., Schwach, F., Studholme, D., and Baulcombe, D.C. 2008. PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc. Natl. Acad. Sci.* **105:** 3145–3150.

Mueller, L.A., Solow, T.H., Taylor, N., Skwarecki, B., Buels, R., Binns, J., Lin, C., Wright, M.H., Ahrens, R., Wang, Y., et al. 2005. The SOL Genomics Network: A comparative resource for Solanaceae biology and beyond. *Plant Physiol.* **138:** 1310–1317.

Pall, G.S., Codony-Servat, C., Byrne, J., Ritchie, L., and Hamilton, A. 2007. Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. *Nucleic Acids Res.* **35:** e60. doi: 10.1093/nar/gkm112.

Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L., and Poethig, R.S. 2004. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of *trans*-acting siRNA in *Arabidopsis*. *Genes & Dev.* **18:** 2368–2379.

Phillips, J., Dalmay, T., and Bartels, D. 2007. The role of small RNAs in abiotic stress. *FEBS Lett.* **581:** 3592–3597.

Pilcher, R.L., Moxon, S., Pakseresht, N., Moulton, V., Manning, K., Seymour, G., and Dalmay, T. 2007. Identification of novel small RNAs in tomato (*Solanum lycopersicum*). *Planta* **226:** 709–717.

Piriyapongsa, J. and Jordan, I.K. 2007. A family of human microRNA genes from miniature inverted-repeat transposable elements. *PLoS ONE* **2:** e203. doi: 10.1371/journal.pone.0000203.

Piriyapongsa, J., Mariño-Ramírez, L., and Jordan, I.K. 2007. Origin and evolution of human microRNAs from transposable elements. *Genetics* **176:** 1323–1337.

Qi, Y., He, X., Wang, X.J., Kohany, O., Jurka, J., and Hannon, G.J. 2006. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443:** 1008–1012.

Rajagopalan, R., Vaucheret, H., Trejo, J., and Bartel, D.P. 2006. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes & Dev.* **20:** 3407–3425.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. 2002. MicroRNAs in plants. *Genes & Dev.* **16:** 1616–1626.

Staginnus, C., Gregor, W., Mette, M.F., Teo, C.H., Borroto-Fernández, E.G., Machado, M.L., Matzke, M., and Schwarzacher, T. 2007. Endogenous pararetroviral sequences in tomato (*Solanum lycopersicum*) and related species. *BMC Plant Biol.* **7:** 24. doi: 10.1186/1471-2229-7-24.

Sunkar, R. and Zhu, J.K. 2004. Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. *Plant Cell* **16:** 2001–2019.

Vaucheret, H. 2006. Post-transcriptional small RNA pathways in plants: Mechanisms and regulations. *Genes & Dev.* **20:** 759–771.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C., Hilbert, J.L., Bartel, D.P., and Crete, P. 2004. Endogenous *trans*-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Mol. Cell* **16:** 69–79.

von Besser, K., Frank, A.C., Johnson, M.A., and Preuss, D. 2006. *Arabidopsis* HAP2 (GCS1) is a sperm-specific gene required for pollen tube guidance and fertilization. *Development* **133:** 4761–4769.

Wenkel, S., Turck, F., Singer, K., Gissot, L., Le Gourrierec, J., Samach, A., and Coupland, G. 2006. CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. *Plant Cell* **18:** 2971–2984.

Yao, Y., Guo, G., Ni, Z., Sunkar, R., Du, J., Zhu, J.K., and Sun, Q. 2007. Cloning and characterization of microRNAs from wheat (*Triticum aestivum* L.). *Genome Biol.* **8:** R96. doi: 10.1186/gb-2007-8-6-r96.