MicroRNA annotation of plant genomes – Do it right or not at all

Richard S. Taylor*, James E. Tarver, Alireza Foroozani and Philip C. J. Donoghue*

MicroRNAs are non-coding regulators of gene expression and key factors in development, disease, and targets for bioengineering. Consequently, microRNAs have become essential elements of already burgeoning draft plant genome descriptions where their annotation is often particularly poor, contributing unduly to the corruption of public databases. Using the *Citrus sinensis* as an example, we highlight and review common failings of miRNAome annotations. Understanding and exploiting the role of miRNAs in plant biology will be stymied unless the research community acts decisively to improve the accuracy of miRNAome annotations. We encourage genome annotation teams to do it right or not at all.

Keywords:
- annotation; *Citrus*; evolution; genome; microRNA; miRBase; mis-annotation; plant

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Introduction

MicroRNAs (miRNAs) are short non-coding trans-acting regulators of gene expression in animals and plants, among other eukaryote kingdoms [1], that have been demonstrated to be important regulators of protein coding genes [2]. They are expressed in a great diversity of tissues and developmental stages [3], implicated as causal factors in macroevolutionary change [4], play fundamental roles in disease-resistance [5] and stress tolerance, and they are targeted for the bio-engineering of domesticated species [6]. Consequently, miRNA annotation has become an essential element of draft genome descriptions, and particularly so for plant species. Guidelines and recommendations for the accurate annotation of miRNAs have existed since the early days of miRNA research [7], and have been updated as technology and understanding have advanced. However, adherence to the established annotation practises varies from publication to publication and, unfortunately, while some of these are exemplars of best practise, the quality of miRNA annotation is generally very. While the problem of spurious miRNA annotation may be widely appreciated [8–13], plant genome annotation studies are contributing disproportionately to the corruption of public databases such as miRBase, serving to obscure rather than elucidate understanding of miRNA diversity, function and evolution. For example, though there are only 28,645 entries from across all of life in miRBase (v.21), 28,281 miRNA loci were recently described from soybean [14], and 98,068 from bread wheat [15]. It is extremely improbable that all but a small minority of these are valid.

Five principal factors contribute to poor miRNA annotation in draft genomes: (i) A failure to discriminate between miRNAs and other classes of small RNAs (sRNAs) [16]. Virtually all sRNAs derived from miRNA loci should be isoforms of one or two distinct products, but loci with heterogeneous processing, yielding large numbers of distinct products, are commonly mis-identified as miRNAs (e.g. [15, 17, 18–24]). In many of these cases, loci previously identified as producing multiple heterogeneous products, are annotated without any supporting evidence of processing precision. ii) A failure to provide sufficient evidence that a locus is a product of miRNA biogenesis, by demonstrating a coherent "hairpin" structure, which is necessary for processing by Dicer, and evidence of miRNA* expression, an artefact of processing by Dicer (e.g. [22–25]). (iii) Uncritical use of databases such as miRBase, that have been corrupted by the registration of many invalid miRNAs [9–12, 26], leading to the propagation of previous spurious annotations (e.g. [15, 17, 18–20, 23, 25, 27–30]). (iv) A general failure to take a predictive phylogenetic approach in evaluating the presence or absence of conserved miRNAs, which aids in the discrimination of

Think again

false negatives (e.g. [28, 31–33]) and false positives. For example, phylogenetically dubious annotations include the identification of animal miRNAs in a plant [20, 27, 34, 35], and algal miRNAs in an angiosperm [15, 36], among other unlikely scenarios (e.g. [15, 23, 25, 29, 30, 32, 34, 37]). (v) The incorrect and inconsistent naming of loci (e.g. [17, 22]).

Many of these problems of miRNA annotation are manifest in the draft genome of the orange, *Citrus sinensis* [22], where a large number of miRNAs were assigned to families identified elsewhere as exclusive to disparate lineages of plants and animals. For example, 38 novel families identified in *C. sinensis*, are considered exclusive to the European honeybee, *Apis mellifera*, in miRBase. If these annotations are correct, the implications are stunning − indicating either a common origin of animal and plant miRNAs, or evidence of horizontal transfer between plant and pollinator. The miRNA repertoire of *C. sinensis* is surprising in other ways, too, including the apparent absence of a number of ancient and otherwise highly conserved miRNA families common to most plant species, and the presence of many miRNA families listed on miRBase that are known to be spurious.

In an attempt to highlight the procedural errors that plague miRNA annotations in many draft genome publications, we undertook a thorough reanalysis of the miRNAome of *C. sinensis,* with use of novel sequencing data.

sRNA sequencing is integral to miRNAome annotation

Although the criteria for the identification of miRNAs are long established [7, 38], these have been updated over time to account for the depth of coverage of NGS data [39, 40]. The key features of a genuine miRNA are: (i) the canonical "hairpin" structure from which the functional "mature" strand and miRNA* strand are processed, (ii) a homogenous mode of processing by which few sequences other than the mature and miRNA* are processed, and (iii) the presence of the miRNA* strand

indicating processing by Dicer. Many software packages exist to identify miRNAs according to these rules, but in order to balance completeness of annotation with accuracy, it is not uncommon for there to be a high frequency of false positives, necessitating manual inspection of each annotated miRNA to ensure strict adherence to annotation criteria.

Approaches for the identification of miRNAs can be categorised as those based on conservation with other species, and de novo identification based on NGS data. Both methods rely on the identification of structural features that indicate processing by the canonical miRNA biogenesis pathways, but only small RNA sequencing allows de novo identification. The central tenets of accurate annotation are presented in Fig. 1; additional methodological details presented in Supplementary information.

The majority of annotations in the *C. sinensis* miRNAome are invalid

Following this pipeline, analysis of the same sRNA data set used for the *C. sinensis* draft genome [22] identified only 98 of the 227 miRNA loci previously described as fulfilling all of the annotation criteria: 77 loci exhibited no evidence of miRNA* expression; 54 loci exhibited heterogeneous processing of multiple products; 24 loci were classified in the incorrect family; 22 loci were duplicates; 20 loci had no evidence for any sRNAs being processed; and 10 loci had invalid hairpin structures (Fig. 2; Tables S1-2). Ninety-nine of the miRNAs described were attributed names (MIR6001-MIR6070) that miRBase has allocated to a wide range of other taxa including *Apis*, *Homo*, *Tribolium*, *Brassica*, *Nicotiana*, and *Solanum*. For example, the miR-6001 entry on miRBase is a miRNA that has only previously been described from honeybee. Finally, the *C. sinensis* draft genome [22] failed to identify 17 miRNA loci that had been described previously by members of the same team [41].

Our resequencing of *C. sinensis* small RNAs, combined with the data used in the draft genome, supports the annotation of 111 of the 227 loci (Tables S3 and Table S4), with invalid hairpin structure or imprecise processing invalidating the miRNA annotations of 62 of the 227 loci. Thirty-two loci lack sufficient evidence required for annotation, viz. expression of the miRNA* or, indeed, evidence of expression of the mature miRNA itself, along with the 22 duplicated sequences. The bulk of these spurious loci are among the putative novel families, leaving a residue of just nine valid miRNAs from among the 99 novel loci proposed. De novo identification of miRNAs using Mirdeep-P identified a further 11 miRNA loci and 21 additional valid loci were obtained from miRBase. In total, the known orange miRNA repertoire comprises 143 miRNAs attributable to 45 families, 10 of which are known only in *C. sinensis* (Fig. 2, Tables S5 and S6).

Over half of the miRNAs annotated in the *C. sinensis* draft genome failed to satisfy the established criteria required for annotation, with many loci exhibiting evidence that directly contradicts their identification as miRNAs (Fig. 2). Based solely on the data used in the original study, even fewer of these loci should have been annotated as miRNAs (only 98 of the 227 loci). The reasons for the large number of false positives are many and varied, but followed generally from a lack of adherence to the established guidelines for miRNA annotation that require strong evidence of processing by DCL-1. A two nucleotide offset between the mature and miRNA* strands provides strong evidence for DCL-1 processing. Without this evidence, the large number of unique reads and potential miRNA-like hairpin structures in the genome will, by chance, result in many reads in the 20-24 nt range being mapped to hairpin-like genome loci even though they are not processed by DCL-1. While both the 5` and 3` strands of miRNAs can be functional [42], it is common for one strand, the miRNA*, to degrade rapidly after disassociation from the mature strand and so the lack of miRNA* reads may be explained by low levels of expression within the cell, rather than evidence that the loci is not a miRNA. Consequently, the 25 loci in the orange genome that lack only this feature
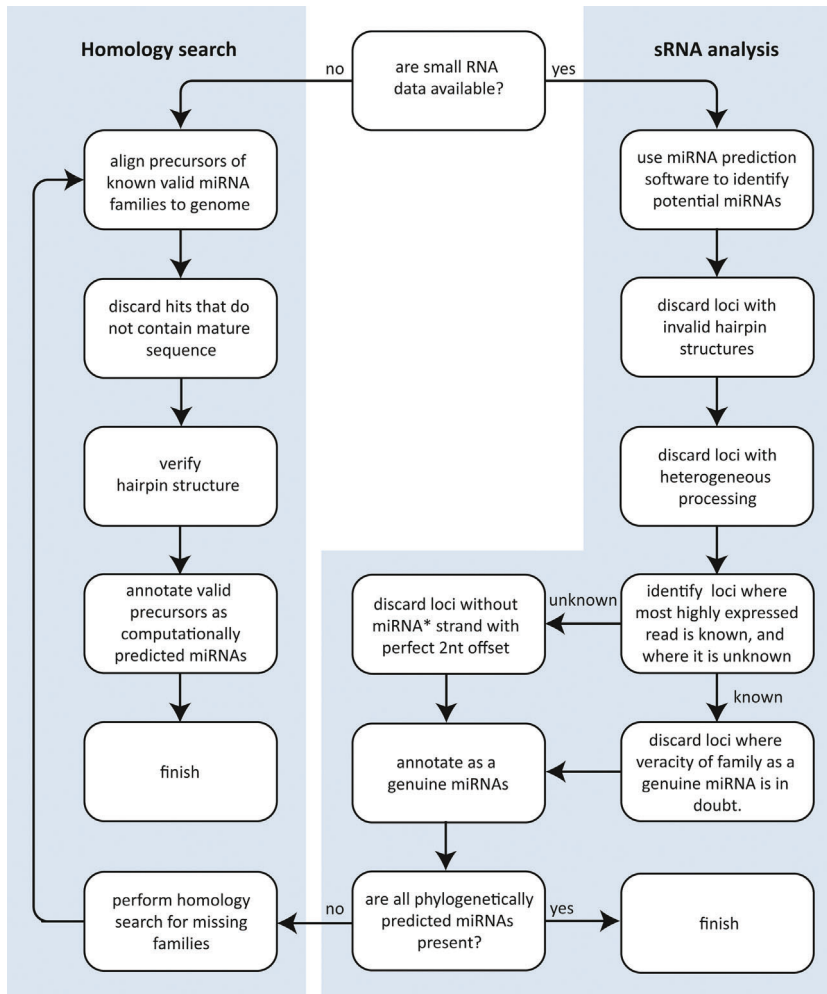
Figure 1. A pipeline for accurate miRNA annotation. This flowchart describes the steps that any microRNA annotation protocol should include in order to ensure accurate and complete annotation of new miRNA genes in whole genomes.

[45], Shortstack [44], and the UEA sRNA workbench [45]. All of these are excellent and extremely useful tools, but in each case the associated publications acknowledge that their software will yield false positives. The only way to avoid the misannotation of such loci as miRNAs is to manually verify that every locus satisfies the established rules of miRNA annotation. Secondly, due to an accumulation of previous incorrect annotations, both in miRBase and wider literature, any homology search must use a valid set of miRNAs to avoid the perpetuation of historical misannotations. Finally, the completeness of a miRNAome is rarely considered in relation to the suite of miRNAs that should be anticipated based on the phylogenetic position of a species. Hence, many miRNAome descriptions overlook the presence of ancient miRNA families. Ultimately, confidence in miRNAome annotation requires data from at least one small RNA library. Without this, the validation of novel loci is not possible, as the precision of processing and evidence of DCL-1 processing cannot be assessed without these data.

Is there a real and meaningful distinction between miRNAs and siRNAs?

The distinction between the tightly controlled processing of miRNA products and heterogeneous processing of siRNAs, is key to the correct identification of miRNA loci. The earliest miRNA families described were conserved across broad taxonomic groupings, exhibited extremely high levels of processing precision and, thus, were clearly distinct from siRNAs [46]. This distinction is often less clear among younger miRNAs, specific to young lineages, because of more heterogeneous processing [47, 48]. It has also been hypothesised that miRNAs can evolve from siRNA loci, suggesting that "proto" miRNAs exist that are in the process of evolving but that do not yet exhibit an unambiguous signature of canonical miRNA biogenesis [49]. Hence, some researchers have claimed that the distinction between miRNAs and other siRNAs is false or arbitrary, perhaps serving more to obscure than inform

can be described as candidate miRNAs, but they should not be annotated as miRNA loci until evidence of miRNA* expression is presented. In contrast, the 57 loci that exhibit heterogeneous processing across the precursor sequence, and the 10 loci that show insufficient complementarity in their precursor structure or mature sequence to allow DCL-1 processing, are certainly not miRNAs. Finally, the exciting possibility of miRNAs shared between orange and honeybee was revealed to be an artefact of inconsistency and inaccuracy in the naming of miRNAs; 22 miRNAs were synonyms of other loci and 23 miRNAs were assigned to the wrong family. All phylogenetically predicted miRNA families [9] are present except MIR828 and MIR3630, which are absent from the genome.

Avoiding the inclusion of spurious miRNAs in genome annotation

There appear to be three primary failings that result in the poor quality of genome annotations. The first of these, and the most significant, is the unquestioning inclusion of all miRNA loci predicted by bioinformatic software, without manual verification of the veracity of these loci. There are many high quality software packages available for the de novo prediction of miRNA loci, but in these there is always a trade off between producing a complete description of the miRNAome and reducing the number of false positives. Three of the leading packages for miRNA identification are miRdeep-P

Think again

Validity of annotation	Number of loci
Correctly annotated	111
Incorrectly annotated	116
Loci with evidence contradicting annotation	62
- heterogeneous processing	57
- invalid hairpin structure	10
Loci lacking sufficient evidence	32
- no evidence of miRNA* expression	25
- no sRNA data maps to given loci	7
Errors in nomenclature	45
- repeated loci	22
- incorrect family name	23
Inconsistency with miRBase nomenclature	99

Phylogenetic distribution						microRNA families in *C. sinensis*
Embryophyta					Embryophyta	156, 159, 160, 166, 167, 171, 390, 408, 477, 530, 535, 536
	Tracheophyta				Tracheophyta	396
		Euphyllophyta			Euphyllophyta	168, 169, 172
			Spermatophyta		Spermatophyta	162, 164, 393, 394, 395, 397, 398, 399, 482, 2950
				Angiospermae	Angiospermae	827, 2275, 1446
					Eudicotyledones	403, 1515, 2111, 3627, 4376, 9560
					Citrus sinensis	3951, 3952, 3954, 12105, 12106, 12107, 12108, 12109, 12110, 12111

Figure 2. Results of reanalysis of miRNAs in *C. sinensis*. (**a**) Summary of the reasons for rejection of miRNA loci annotated in the *C. sinensis* draft genome, based on a combined sRNA dataset of our sequencing data and the data used in the genome publication. (**b**) A summary of the miRNA families present in *C. sinensis*, and associated phylogenetic distribution. Many of the totals do not sum because some of the loci exhibit multiple errors.

the origin and function of the small RNA regulatory machinery [50].

There can be no doubt that both miRNAs and siRNAs are important functional classes of small RNAs and, indeed, there are examples in which miRNAs have evolved from and into siRNAs [51, 52]. However, in the vast majority of instances, miRNAs and siRNAs are readily distinguishable based on the established annotation criteria – principally, the consistency of their processing. In almost all miRNAs, consistency of processing of an evolutionarily conserved locus yields a distinct product that binds to a specific and equally conserved target sequence. siRNAs are not conserved in the same manner and the heterogenous nature of their processing bears this out. These factors make the distinction between miRNAs and siRNAs useful in the sense that it reflects functional, regulatory, as well as evolutionary differences between these two classes of small RNAs. At the very least, knowledge of the function of specific miRNAs can be translated between species in which they are conserved, in a manner that is not possible among siRNAs, except in a much more general sense. Thus, for both theoretical and functional reasons, we advocate a continued distinction between these two classes of small regulatory RNAs.

Conclusions and outlook

The annotation of the orange miR-NAome is particularly poor, but it is not otherwise unrepresentative of miRNA annotation in draft genomes. We have already highlighted the implausibly large repertoire of 98,068 miRNA loci in bread wheat [15] and 28,281 loci in soybean [14]. Similarly, yet less implausibly, 537 miRNA loci were recently described from mungbean [20], but over half are clearly spurious, assigned erroneously to miRNA families known only from animals. Even a cursory consideration of sequence homology, perhaps prompted by phylogenetic expectations of the mungbean miRNAome [9], could have avoided the annotation of yet more spurious miR-NAs. Neither the annotation criteria nor the pipeline we describe for their implementation are novel, they are simply not being observed. Consequently, spurious miRNAs are being described at an increasingly alarming rate, diminishing prospects for understanding, and exploiting the role of miRNAs in plant biology. This eventuality was foreseen by the original authors of the criteria for plant miRNA annotation who predicted that without adherence to a strict set of annotation criteria, there would be widespread "annotation of miRNAs that have a high likelihood of being siRNAs; many sequences of such uncertain provenance are likely to be identified for a broad range of plant species" [38]. This prophecy has now come to pass, and the research community must react decisively to improve the accuracy of miRNAome annotations and, particularly, the generally poor quality annotations associated with draft genomes

which contribute more than their fair share to the corruption of miRNA repositories. Undoubtedly, this is at least in part a consequence of the scope of plant genome descriptions, which has expanded with knowledge of the universe of coding and non-coding genes and regulatory elements. The same problem is not generally manifest in the annotation of animal genomes as miRNAomes are more usually described in distinct publications where their justification may achieve greater scrutiny. This may be the most practical solution to the problem of miRNA misannotation but, ultimately, we implore genome annotation teams that aim to tackle the miRNAome – do it right or not at all.

Acknowledgments

The authors have declared no conflict of interest.

References

1. **Tarver JE, Donoghue PCJ, Peterson KJ.** 2012. Do miRNAs have a deep evolutionary history?. *BioEssays* **34**: 857–66.
2. **Friedman RC, Farh KKH, Burge CB, Bartel DP.** 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**: 92–105.
3. **Rubio-Somoza I, Weigel D.** 2011. MicroRNA networks and developmental plasticity in plants. *Trends in Plant Sci* **16**: 258–64.
4. **Tarver JE, Sperling EA, Nailor A, Heimberg AM,** et al. 2013. MiRNAs: small genes with big potential in metazoan phylogenetics. *Mol Biol Evol* **30**: 2369–82.
5. **Alvarez-Garcia I, Miska EA.** 2005. MicroRNA functions in animal development and human disease. *Development* **132**: 4653–62.
6. **Franco-Zorrilla JM, Valli A, Todesco M, Mateos I,** et al. 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet* **39**: 1033–7.
7. **Ambros V, Bartel B, Bartel DP, Burge CB,** et al. 2003. *A Uniform System for MicroRNA Annotation*. New York, NY: RNA **9**: pp 277–9.
8. **Budak H, Bulut R, Kantar M, Alptekin B.** 2015. MicroRNA nomenclature and the need for a revised naming prescription. *Brief Funct Genomics* **15**: 65–71.
9. **Taylor RS, Tarver JE, Hiscock SJ, Donoghue PCJ.** 2014. Evolutionary history of plant microRNAs. *Trends in plant sci* **19**: 175–82.
10. **Meng Y, Shao C, Wang H, Chen M.** 2012. Are all the miRBase-registered microRNAs true?: a structure-and expression-based re-examination in plants. *RNA biology* **9**: 249–53.
11. **Hansen TB, Kjems J, Bramsen JB.** 2011. Enhancing miRNA annotation confidence in miRBase by continuous cross dataset analysis. *RNA Biol* **8**: 378–83.
12. **Jones-Rhoades MW.** 2012. Conservation and divergence in plant microRNAs. *Plant Mol Biol* **80**: 3–16.
13. **Fromm B, Billipp T, Peck LE, Johansen M,** et al. 2015. A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human micro-RNAome. *Annu Rev Genet* **49**: 213–42.
14. **Zhou ZK, Wang Z, Li WY, Fang C,** et al. 2013. Comprehensive analyses of microRNA gene evolution in paleopolyploid soybean genome. *Plant J* **76**: 332–44.
15. **Mayer KFX, Rogers J, Dolezel J, Pozniak C,** et al. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **345**: 1251788.
16. **Axtell MJ.** 2013. Classification and comparison of small RNAs from plants. *Annu Rev Plant Biol* **64**: 137–59.
17. **Jia JZ, Zhao SC, Kong XY, Li YR,** et al. 2013. *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* **496**: 91–5.
18. **Ling HQ, Zhao SC, Liu DC, Wang JY,** et al. 2013. Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* **496**: 87–90.
19. **Dohm JC, Minoche AE, Holtgrawe D, Capella-Gutierrez S,** et al. 2014. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature* **505**: 546–9.
20. **Kang YJ, Kim SK, Kim MY, Lestari P,** et al. 2014. Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nat Commun* **5**: 5443.
21. **Varshney RK, Chen WB, Li YP, Bharti AK,** et al. 2012. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat Biotechnol* **30**: 83—U128.
22. **Xu Q, Chen LL, Ruan XA, Chen DJ,** et al. 2013. The draft genome of sweet orange (*Citrus sinensis*). *Nat Genet* **45**: 59—U92.
23. **Yang RL, Jarvis DE, Chen H, Beilstein MA,** et al. 2013. The reference genome of the halophytic plant *Eutrema salsugineum*. *Front Plant Sci* **4**: 2013.00046.
24. **Nystedt B, Street NR, Wetterbom A, Zuccolo A,** et al. 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature* **497**: 579–84.
25. **Qin C, Yu CS, Shen YO, Fang XD,** et al. 2014. Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. *P Natl Acad Sci USA* **111**: 5135–40.
26. **Wang X, Liu XS.** 2011. Systematic curation of miRBase annotation using integrated small RNA high-throughput sequencing data for *C. elegans* and *Drosophila*. *Front Genet* **2**: 2011.00025.
27. **Wang KB, Wang ZW, Li FG, Ye WW,** et al. 2012. The draft genome of a diploid cotton *Gossypium raimondii*. *Nat Genet* **44**: 1098–103.
28. **Denoeud F, Carretero-Paulet L, Dereeper A, Droc G,** et al. 2014. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* **345**: 1181–4.
29. **Verde I, Abbott AG, Scalabrin S, Jung S,** et al. 2013. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat Genet* **45**: 487—U47.
30. **Olsen JL, Rouze P, Verhelst B, Lin YC,** et al. 2016. The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. *Nature* **530**: 331–5.
31. **Peng ZH, Lu Y, Li LB, Zhao Q,** et al. 2013. The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*). *Nat Genet* **45**: 456–61.
32. **Ibarra-Laclette E, Lyons E, Hernandez-Guzman G, Perez-Torres CA,** et al. 2013. Architecture and evolution of a minute plant genome. *Nature* **498**: 94–8.
33. **Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M,** et al. 2012. The genome of melon (*Cucumis melo L.*). *Proc Natl Acad Sci USA* **109**: 11872–7.
34. **Barvkar VT, Pardeshi VC, Kale SM, Qiu S,** et al. 2013. Genome-wide identification and characterization of microRNA genes and their targets in flax (*Linum usitatissimum*): characterization of flax miRNA genes. *Planta* **237**: 1149–61.
35. **Tang C, Yang M, Fang Y, Luo Y,** et al. 2016. The rubber tree genome reveals new insights into rubber production and species adaptation. *Nat Plants* **2**: 16073.
36. **Dassanayake M, Oh DH, Haas JS, Hernandez A,** et al. 2011. The genome of the extremophile crucifer *Thellungiella parvula*. *Nat Genet* **43**: 913—U137.
37. **Wang W, Haberer G, Gundlach H, Glasser C,** et al. 2014. The Spirodela polyrhiza genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. *Nat Commun* **5**: 3311.
38. **Meyers BC, Axtell MJ, Bartel B, Bartel DP,** et al. 2008. Criteria for annotation of plant MicroRNAs. *The Plant cell* **20**: 3186–90.
39. **Kozomara A, Griffiths-Jones S.** 2011. MiRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* **39**: D152–7.
40. **Kozomara A, Griffiths-Jones S.** 2014. MiRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**: D68–73.
41. **Xu Q, Liu Y, Zhu A, Wu X,** et al. 2010. Discovery and comparative profiling of microRNAs in a sweet orange red-flesh mutant and its wild type. *BMC Genom* **11**: 246.
42. **Okamura K, Phillips MD, Tyler DM, Duan H,** et al. 2008. The regulatory activity of micro-RNA star species has substantial influence on microRNA and 3 ' UTR evolution. *Nat Struct Mol Biol* **15**: 354–63.
43. **Yang XZ, Li L.** 2011. MiRDeep-P: a computational tool for analyzing the microRNA

Think again

transcriptome in plants. *Bioinformatics* **27**: 2614–5.

44. **Axtell MJ.** 2013. ShortStack: comprehensive annotation and quantification of small RNA genes. *RNA* **19**: 740–51.

45. **Stocks MB, Moxon S, Mapleson D, Woolfenden HC**, et al. 2012. The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. *Bioinformatics* **28**: 2059–61.

46. **Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B**, et al. 2002. MicroRNAs in plants. *Gene Dev* **16**: 1616–26.

47. **Ma Z, Coruh C, Axtell MJ.** 2010. Arabidopsis lyrata small RNAs: transient MIRNA and small interfering RNA loci within the *Arabidopsis* genus. *Plant Cell* **22**: 1090–103.

48. **Fahlgren N, Jogdeo S, Kasschau KD, Sullivan CM**, et al. 2010. MicroRNA gene evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. *Plant Cell* **22**: 1074–89.

49. **Cuperus JT, Fahlgren N, Carrington JC.** 2011. Evolution and functional diversification of MIRNA genes. *Plant Cell* **23**: 431–42.

50. **Desvignes T, Batzel P, Berezikov E, Eilbeck K**, et al. 2015. MiRNA Nomenclature: a view incorporating genetic origins, biosynthetic pathways, and sequence variants. *Trends Genet* **31**: 613–26.

51. **Rajagopalan R, Vaucheret H, Trejo J, Bartel DP.** 2006. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes and Development* **20**: 3407–25.

52. **Fahlgren N, Howell MD, Kasschau KD, Chapman EJ**, et al. 2007. High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of MIRNA genes. *PLoS ONE* **2**: e219.