

COMMENTARY

Criteria for Annotation of Plant MicroRNAs

Blake C. Meyers,^{a,1} Michael J. Axtell,^{b,1} Bonnie Bartel,^c David P. Bartel,^d David Baulcombe,^e John L. Bowman,^f Xiaofeng Cao,^g James C. Carrington,^h Xuemei Chen,ⁱ Pamela J. Green,^a Sam Griffiths-Jones,^j Steven E. Jacobsen,^k Allison C. Mallory,^l Robert A. Martienssen,^m R. Scott Poethig,ⁿ Yijun Qi,^o Herve Vaucheret,^l Olivier Voinnet,^p Yuichiro Watanabe,^q Detlef Weigel,^r and Jian-Kang Zhuⁱ

^a Department of Plant and Soil Sciences, Delaware Biotechnology Institute, University of Delaware, Newark, Delaware 19711

^b Department of Biology and Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, Pennsylvania 16802

^c Department of Biochemistry and Cell Biology, Rice University, Houston, Texas 77005

^d Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Whitehead Institute, Cambridge, Massachusetts 02142

^e Plant Sciences Department, Cambridge University, Cambridge CB2 3EA, United Kingdom

^f School of Biological Sciences, Monash University, Melbourne VIC 3800, Australia

^g Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

^h Center for Genome Research and Biocomputing, Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon 97331

ⁱ Department of Botany and Plant Sciences, Institute of Integrative Genome Biology, University of California, Riverside, California 92521

^j Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, United Kingdom

^k Howard Hughes Medical Institute, Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, California 90095

^l Laboratoire de Biologie Cellulaire, Institut Jean-Pierre Bourgin, Institut National de la Recherche Agronomique, 78026 Versailles Cedex, France

^m Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724

ⁿ Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104

^o National Institute of Biological Sciences, Zhongguancun Life Science Park, Beijing 102206, China

^p Institut de Biologie Moleculaire des Plantes du Centre National de la Recherche Scientifique 12, 67084 Strasbourg Cedex, France

^q Department of Life Sciences, University of Tokyo, Meguro, Tokyo 153-8902, Japan

^r Department of Molecular Biology, Max Planck Institute for Developmental Biology, D-72076 Tübingen, Germany

MicroRNAs (miRNAs) are ~21 nucleotide noncoding RNAs produced by Dicer-catalyzed excision from stem-loop precursors. Many plant miRNAs play critical roles in development, nutrient homeostasis, abiotic stress responses, and pathogen responses via interactions with specific target mRNAs. miRNAs are not the only Dicer-derived small RNAs produced by plants: A substantial amount of the total small RNA abundance and an overwhelming amount of small RNA sequence diversity is contributed by distinct classes of 21- to 24-nucleotide short interfering RNAs. This fact, coupled with the rapidly increasing rate of plant small RNA discovery, demands an increased rigor in miRNA annotations. Herein, we update the specific criteria required for the annotation of plant miRNAs, including experimental and computational data, as well as refinements to standard nomenclature.

It has now been more than 5 years since the original guidelines for microRNA (miRNA) annotation in plants and animals were spelled out by a group of leading laboratories in the small RNA field (Ambros et al., 2003). To date, >180 genes encoding miRNAs have been annotated in *Arabi-*

dopsis thaliana, and >1000 have been annotated among all plants, as listed in miRBase, the home of miRNA data (Griffiths-Jones et al., 2008). During this time, a variety of methods and approaches have been adopted by labs working on diverse plant species, and this has led to occasional differences in the criteria and quality of data used to annotate miRNAs. The original guidelines stated that a small RNA could be designated as a miRNA if it

fulfilled a set of expression and biogenesis criteria (Ambros et al., 2003). The expression criteria included the identification of the small RNA by cloning and/or detection by hybridization, whereas the biogenesis criteria included a precursor transcript predicted to fold into a characteristic hairpin structure, phylogenetic conservation of the miRNA sequence and precursor secondary structure, and increased accumulation of a precursor when Dicer activity is

¹Address correspondence to meyers@dbi.udel.edu or mja18@psu.edu.
www.plantcell.org/cgi/doi/10.1105/tpc.108.064311

COMMENTARY

reduced (Ambros et al., 2003). However, in the past 5 years our understanding of small RNAs has grown, and the tools for analyzing them have improved. Many articles reporting new miRNAs in plants have assimilated the insights of recent years, but some have not. We believe that it is time to revise the minimal criteria for plant miRNA annotation with the goal of maintaining the highest quality data in the miRNA registry.

Herein, we refine the criteria for annotating miRNAs of plants. In addition to slightly different characteristics of plant versus animal miRNAs, the reason for this emphasis is that plants have relatively large and complex small RNA populations within which miRNAs are often a minority. By contrast, most of the small RNAs in the soma of vertebrates and flies are miRNAs. The more complex pool of plant small RNAs is largely due to the plant-specific RNA POLYMERASE IV/RNA POLYMERASE V (PolIV/PolV)-dependent short interfering RNAs (siRNAs) as well as secondary siRNAs, some of which are *trans*-acting. These diverse, endogenous siRNA populations make rigorous annotation of miRNAs more challenging. The enormous numbers of small RNAs being cataloged using next-generation DNA sequencing technologies demand a renewed uniformity in plant miRNA annotation to avoid misleading and inaccurate conflation of different classes of small RNAs.

PRIMARY CRITERION: PRECISE EXCISION FROM THE STEM OF A STEM-LOOP PRECURSOR

We submit that the fundamental defining feature of plant miRNAs is the precise excision of an ~21-nucleotide miRNA/miRNA* duplex from the stem of a single-stranded, stem-loop precursor. This duplex is an intermediate of miRNA biogenesis that is present after cleavage of the *MIRNA* stem-loop but before the mature miRNA enters the silencing complex. Stem-loop precursors are predicted using genomic DNA or known ESTs/transcripts as the input for RNA secondary structure prediction software. Plant *MIRNA* stem-loops are more variable in size and structural features

than those of animals, but confidently identified instances share the following characteristics: (1) The miRNA and miRNA* are derived from opposite stem-arms such that they form a duplex with two nucleotide, 3' overhangs; (2) base-pairing between the miRNA and the other arm of the hairpin, which includes the miRNA*, is extensive such that there are typically four or fewer mismatched miRNA bases; and (3) asymmetric bulges are minimal in size (one or two bases) and frequency (typically one or less), especially within the miRNA/miRNA* duplex. Small RNA-producing stem-loops that slightly violate one of these criteria could still be annotated as *MIRNAs*, provided that there is exceptional evidence of precise miRNA/miRNA* excision; but in general, those that violate these characteristics should not be classified as *MIRNAs*. Biogenesis from a stem-loop excludes endogenous siRNAs, as they generally arise from long, perfectly double-stranded RNA. The requirement for precision in biogenesis also excludes small RNAs that derive from arbitrary positions within otherwise acceptable stem-loop precursors as well as excluding randomly degraded mRNAs that have fortuitous overlap with predicted stem-loops. However, the requirement for precision does not imply that all small RNAs expressed from a stem-loop must be either the precise miRNA or precise miRNA*. Deep sequencing data clearly show that there are often, if not always, low frequency positional and length variants from all *MIRNA* stem-loops. In addition, some *MIRNA* precursors give rise to two or more distinct miRNA/miRNA* duplexes from different positions. The point at which the level of precision that defines a miRNA gives way to non-miRNA imprecision is somewhat subjective. However, as a general rule of thumb, stem-loops in which more than ~25% of observed small RNA abundance does not correspond to one (or more; see multifunctional stem loops below) distinct miRNA/miRNA* duplexes should be considered too imprecise to qualify as *MIRNAs*. Although ancillary criteria can enhance a miRNA annotation (see below), conclusive evidence of precise biogenesis from a qualifying stem-loop is the sole

criterion that is both necessary and sufficient for miRNA annotation.

The primary criterion is most readily satisfied by the sequencing of cDNAs derived from small RNA samples, coupled with analysis of the putative precursors of the small RNAs. In most cases, complete nuclear genome sequences are required to find all possible origins for sequenced small RNAs; however, species for which there is especially good EST coverage are also amenable to miRNA analysis. Isolation of only one or two small RNAs matching a predicted stem-loop does not suffice for a confident annotation; this is because cases of low coverage could miss small RNAs deriving from heterogeneous positions or from the opposite genomic strand, which indicate non-miRNA origins. Ideally, sequences representing both the miRNA and miRNA* would be used to satisfy the primary criterion. In the absence of miRNA* confirmation, a clear dominance of a specific small RNA sequence (the miRNA) from one arm of a predicted stem-loop is required. In these miRNA*-deficient cases, annotation is best supported by isolation and sequencing of the candidate miRNA from multiple, independent libraries. It should be noted that a very low abundance of just one or two sequencing reads of a putative miRNA, followed by detection via RNA gel blots, does not satisfy the primary criterion: Detection of a discrete small RNA by blot hybridization would be unexpected for a spurious decay product of a larger precursor, but a low sequencing depth cannot discriminate between siRNAs and miRNAs nor eliminate the possibility of heterogeneous processing. In such cases of low-depth sequencing, more extensive blot analysis with multiple probes would be needed to rule out small RNA accumulation from other positions within the putative *MIRNA* locus and to rule out small RNA accumulation from the opposite genomic strand. Similarly, annotation of miRNAs based solely on sequencing AGO-associated small RNAs (for instance, from immunoprecipitations) may be problematic because the 5' nucleotide specificity characteristic of many AGO proteins will prevent observation of the total population of small RNAs deriving from candidate stem-loops.

COMMENTARY

ANCILLARY CRITERIA

Other characteristics can be used to bolster plant miRNA annotations. However, these criteria are unnecessary and/or insufficient for miRNA annotation. This should not be taken to mean that these features are unworthy of study; on the contrary, investigation of the following aspects of miRNAs is critical for understanding the biological role and evolution of these molecules. Additionally, satisfaction of one or more of these ancillary criteria can significantly increase the confidence of a miRNA annotation. However, because they are individually either unnecessary, insufficient, or both, the following criteria are ancillary for the strict purpose of miRNA annotation:

Conservation

Conservation of miRNAs, assessed using either bioinformatics or direct experimentation, is a powerful indicator of their functional relevance and ancient origin. Preservation across lineages of a predicted stem-loop secondary structure along with the embedded miRNA sequence provides especially strong evidence in favor of a miRNA annotation. Nonetheless, many bona fide plant miRNAs lack readily detected homologs outside of the founding species. Thus, demonstration of conservation is not necessary for annotation of miRNAs. However, in contrast with the other ancillary criteria, clear evidence of conservation of both the stem-loop secondary structure and the mature miRNA sequence is by itself sufficient for confident annotation of orthologous miRNAs, provided that the precise stem-loop biogenesis criterion was experimentally satisfied in at least one species. In this respect, our guidelines retain the original miRNA criteria as described by Ambros et al. (2003).

Targets

Many currently known plant miRNAs mediate the regulation of specific target mRNAs by one or more molecular mechanisms, including target cleavage (Llave et al., 2002), translational repression

(Aukerman and Sakai, 2003; Chen, 2004), and pairing to non-protein-coding RNAs without cleavage (Axtell et al., 2006; Franco-Zorrilla et al., 2007). The biological roles of miRNAs seem to be restricted to target regulation or processing; however, it does not follow that identification of a target is necessary for miRNA annotation. Many less-conserved plant miRNAs have predicted targets that have proven difficult to confirm, while others appear to have no targets at all. Some of these may truly be target-less miRNAs that are evolutionarily transient, whereas others may have undiscovered targets that are not amenable to identification using current computational and experimental techniques. In either case, determining the function of a miRNA is not required for its annotation, just as protein-coding genes without known function can also be annotated. Small RNA-directed target regulation is also not sufficient for miRNA annotation. For example, *trans*-acting siRNAs also direct cleavage of target mRNAs, while heterochromatic siRNAs repress targets at the transcriptional level.

DCL1 Dependence

All plants have multiple *DICER-LIKE* (*DCL*) genes. Among these, *DCL1* appears to be largely specialized for miRNA production, whereas the others are specialized for the production of various siRNAs. In *Arabidopsis* and rice, hypomorphic *dcl1* mutations impact the accumulation of most, but not all, miRNAs. However, because null *dcl1* alleles are embryo lethal (at least in *Arabidopsis*), it is not possible to determine the exact dependence of every miRNA on *DCL1*. Because of the presence of several miRNAs whose accumulation depends upon *DCL4* (Rajagopalan et al., 2006), *DCL1* dependence is not necessary for miRNA annotation. In addition, the observation of *DCL1* dependence for accumulation of a small RNA is by itself insufficient to warrant annotation as a miRNA: Certain siRNAs, including many secondary siRNAs and some natural antisense siRNAs also require *DCL1* for their accumulation but do not derive from precisely processed stem-

loops. A strict requirement for evidence of relatively precise excision from qualifying stem-loops obviates the need for genetic analysis of *DCL* dependencies and enables miRNAs to be annotated in those species in which *dcl1* mutants are unavailable.

RDR and PolIV/PolV Independence

Plants also have one or more RNA-DEPENDENT RNA POLYMERASES (*RDRs*) thought to produce the dsRNA molecules from which many siRNAs derive. Thus, *RDR* dependencies are characteristic of many siRNAs. By contrast, miRNAs are not excised from *RDR*-derived dsRNAs. However, while *RDR* dependency defines many siRNAs, *RDR* independence is not restricted to miRNAs. For example, *RDR*-independent stem-loops will not satisfy the primary criterion if they are heterogeneously processed. Many siRNAs also depend on PolIV/PolV. But, as for *RDR* independence, PolIV/PolV independence is insufficient for miRNA annotation. *RDR* and PolIV/PolV dependence also does not necessarily preclude miRNA annotation: For instance, it is theoretically possible that *RDR* and/or PolIV/PolV-dependent siRNAs might impinge upon *MIRNA* transcription, thus causing an indirect genetic dependency. In summary, testing of *RDR* and PolIV/PolV dependencies is neither necessary nor sufficient for miRNA annotation. As with *dcl1* mutants, there are many species in which mutants of functionally characterized *RDRs* and/or the PolIV/PolV subunits are unavailable. Thus, this ancillary criterion is similarly limited in applicability.

REPEATS AND STRUCTURAL RNAs

Endogenous siRNAs derived from the PolIV/PolV pathway preferentially accumulate from repetitive regions, including tandem repeats and transposon-derived sequences. Because the diversity of PolIV/PolV-dependent small RNAs is much greater than that of miRNAs, special care should be taken when annotating *MIRNA* loci that overlap with repetitive sequences. In these cases, it becomes particularly

COMMENTARY

important to satisfy the primary criterion and to rule out the possibility of siRNA production from the locus of interest. This is not to say that repetitive regions in principle cannot harbor *MIRNA* loci. However, the burden of proof must be high, indeed, when describing *MIRNA* loci within repetitive regions to avoid misannotation of siRNAs as miRNAs.

Randomly sized fragments of non-*DCL*-dependent RNAs usually, if not always, contaminate libraries of small RNAs. Almost all of these are from the most abundant classes of cellular RNAs: rRNAs and tRNAs. Thus, any small RNAs that match the sense strand of mature rRNAs or tRNAs should be excluded a priori from consideration as miRNAs. In well-annotated genomes, such as *Arabidopsis*, excluding rRNAs and tRNAs is trivial. However, in other species, care should be taken to assemble and identify rRNA and tRNA sequences prior to annotation of novel miRNAs. Other classes of less abundant noncoding RNAs, such as small nuclear and small nucleolar RNAs, can also cause contamination issues.

ASSIGNING miRNAs TO DISTINCT FAMILIES

An additional complexity of miRNA annotation is the frequent presence within a genome of paralogous *MIRNA* loci producing identical or nearly identical mature miRNAs. Logically, these identical or nearly identical miRNAs have been grouped together into families. In a sequenced genome, once a single locus has been confirmed to produce a miRNA by the primary criterion, the initial identification of other potential family members is trivial; assessment of the potential stem-loop precursors at each site where perfect or near-perfect matches to the mature miRNA sequence occur yields a list of presumed paralogs that can be annotated as miRNAs. For species with unsequenced genomes, identifying candidate miRNA family members requires cDNA sequences or other experimental data. Candidate family members identified in this way may have variable degrees of experimental support.

For instance, both the initially confirmed *MIRNA* locus and a subsequently identified paralog could produce identical miRNA and miRNA* species, obscuring whether miRNA production derives from the initially identified locus, the paralog, or both. In some cases, sequenced miRNA* species or other rare miRNA variants provide locus-specific signatures proving expression from a specific paralog. Optimal annotation of paralogs would ideally include information on whether the annotation is based on ambiguous or unambiguous experimental data, or upon similarity alone.

The protocol of submission of novel miRNAs to the miRBase registry (Griffiths-Jones et al., 2008) after the acceptance of a manuscript has worked well to ensure an orderly assignment of miRNA numbers; we do not propose any changes to this protocol. Loci giving rise to identical or similar miRNAs are assigned the same number with sequential alphabetical suffixes (i.e., *MIR172a*, *MIR172b*, etc.). miRBase assigns three-letter prefixes based upon the genus and species from which the *MIRNA* is derived. When two or more miRNA-producing stem-loops are arrayed in tandem on a single precursor, each stem loop is treated as a distinct locus (for instance, *Zea mays MIR156b* and *MIR156c*; Chuck et al., 2007). However, there are some *MIRNA* genes that encode similar mature miRNA sequences that for historical reasons have been assigned distinct identifiers (for instance, the miR156/157, miR165/166, and miR170/171 families). The designation of miRNA names should take into consideration the number of mismatches compared with other named miRNAs, with from zero to two being typical, but up to four being acceptable, provided that the mature miRNAs derive from the same arm of the stem-loop in all cases. In addition, sequence-related miRNAs with different targets can be classified into different families at the author's discretion. For example, the sequence-related but functionally distinct *Arabidopsis* miR159 and miR319 families (which regulate MYB- and TCP-family targets, respectively; Palatnik et al., 2007) sometimes have been placed in separate families.

MULTIFUNCTIONAL STEM-LOOPS

There are several cases for which multiple miRNAs accumulate from the same precursor; these include cases where miRNA and miRNA* species accumulate to approximately equal levels (for instance, miR832; Rajagopalan et al., 2006), cases where overlapping but distinct miRNA species are produced from the same arm of a stem-loop (for instance, miR161; Allen et al., 2004), and cases where multiple miRNA/miRNA* duplexes are sequentially excised (for instance, miR163; Kurihara and Watanabe, 2004). In such cases, the naming procedure for miRNAs becomes a bit more complicated. When there are two miRNA species that accumulate in approximately equal proportions and derive from the same initial duplex, the miRNA/miRNA* nomenclature is omitted in favor of the suffixes 5p and 3p, which designate the miRNA species arising from the 5' and 3' arms, respectively, of the stem-loop. The distinct miRNA species resulting from overlapping or sequential processing should be distinguished by numerical suffixes (e.g., miR161.1 and miR161.2). In some cases, the numerical suffixes indicating the locus of origin must be combined with letter designations for locus-specific variants (for an example, see the rice MIR444d locus; Lu et al., 2008). Finally, recent data suggest that some miRNA*s could indeed be competent to direct target cleavage in plants (German et al., 2008), suggesting that revisions to the 5p/3p system of nomenclature might be in store for these previously annotated miRNA loci.

CONCLUSIONS

A problem in providing a definitive list of criteria for miRNA annotation is that this list is likely to evolve over time, much as standards have evolved over the last 5 years. Our intent with this communication is to delineate criteria that we believe uniquely characterize miRNAs and to provide guidelines for using these criteria to discriminate miRNAs from other small RNAs. One goal is to head off the future annotation of miRNAs that have a high

COMMENTARY

likelihood of being siRNAs; many sequences of such uncertain provenance are likely to be identified for a broad range of plant species now that next-generation sequencing of small RNAs has become commonplace. With the application of these renewed criteria, we hope to maintain a miRNA registry of the highest possible quality even at the risk of excluding some possibly valid entries that will require more support before inclusion. Authors must be self-policing in their annotations, but it also falls to reviewers to diligently flag questionable claims of miRNAs. With so many articles in so many journals, it may be difficult to effectively review each and every miRNA manuscript, but we hope that these criteria will be helpful to both authors and reviewers of articles describing new miRNAs.

REFERENCES

- 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.
- Ambros, V., et al. (2003). A uniform system for microRNA annotation. *RNA* **9**: 277–279.
- Aukerman, M.J., and Sakai, H. (2003). Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. *Plant Cell* **15**: 2730–2741.
- Axtell, M.J., Jan, C., Rajagopalan, R., and Bartel, D.P. (2006). A two-hit trigger for siRNA biogenesis in plants. *Cell* **127**: 565–577.
- Chen, X. (2004). A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development. *Science* **303**: 2022–2025.
- Chuck, G., Cigan, A.M., Saeteurn, K., and Hake, S. (2007). The heterochronic maize mutant *Corngrass1* results from overexpression of a tandem microRNA. *Nat. Genet.* **39**: 544–549.
- Franco-Zorrilla, J.M., Valli, A., Todesco, M., Mateos, I., Puga, M.I., Rubio-Somoza, I., Leyva, A., Weigel, D., Garcia, J.A., and Paz-Ares, J. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* **39**: 1033–1037.
- German, M.A., et al. (2008). Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat. Biotechnol.* **26**: 941–946.
- Griffiths-Jones, S., Saini, H.K., van Dongen, S., and Enright, A.J. (2008). miRBase: Tools for microRNA genomics. *Nucleic Acids Res.* **36**: D154–D158.
- Kurihara, Y., and Watanabe, Y. (2004). *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc. Natl. Acad. Sci. USA* **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., et al. (2008). Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). *Proc. Natl. Acad. Sci. USA* **105**: 4951–4956.
- Palatnik, J.F., Wollmann, H., Schommer, C., Schwab, R., Boisbouvier, J., Rodriguez, R., Warthmann, N., Allen, E., Dezulian, T., Huson, D., Carrington, J.C., and Weigel, D. (2007). Sequence and expression differences underlie functional specialization of *Arabidopsis* microRNAs miR159 and miR319. *Dev. Cell* **13**: 115–125.
- 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.