Briefings in Bioinformatics, 19(6), 2018, 1273–1289

doi: 10.1093/bib/bbx058
Advance Access Publication Date: 30 May 2017
Paper

ceRNAs in plants: computational approaches and associated challenges for target mimic research

Alexandre Rossi Paschoal*, Irma Lozada-Chávez*, Douglas Silva Domingues, and Peter F. Stadler

Corresponding authors: Alexandre Rossi Paschoal, UTFPR - CÂMPUS CORNÉLIO PROCÓPIO, 1640 Avenue Alberto Carazzai, 86300-000, Cornélio Procópio, PR - Brazil. Tel.: +55 (43) 3520-4055; Fax: +55 (43) 3520-4010; E-mail: paschoal@utfpr.edu.br; Peter F. Stadler, Bioinformatics Group, Department of Computer Science and Interdisciplinary, Center for Bioinformatics, University Leipzig, Härtelstraße 16-18, 04107, Leipzig, Germany; E-mail: studla@bioinf.uni-leipzig.de
*These authors contributed equally to this work.

Abstract

The competing endogenous RNA hypothesis has gained increasing attention as a potential global regulatory mechanism of microRNAs (miRNAs), and as a powerful tool to predict the function of many noncoding RNAs, including miRNAs themselves. Most studies have been focused on animals, although target mimic (TMs) discovery as well as important computational and experimental advances has been developed in plants over the past decade. Thus, our contribution summarizes recent progresses in computational approaches for research of miRNA:TM interactions. We divided this article in three main contributions. First, a general overview of research on TMs in plants is presented with practical descriptions of the available literature, tools, data, databases and computational reports. Second, we describe a common protocol for the computational and experimental analyses of TM. Third, we provide a bioinformatics approach for the prediction of TM motifs potentially cross-targeting both members within the same or from different miRNA families, based on the identification of consensus miRNA-binding sites from known TMs across sequenced genomes, transcriptomes and known miRNAs. This computational approach is promising because, in contrast to animals, miRNA families in plants are large with identical or similar members, several of which are also highly conserved. From the three consensus TM motifs found with our approach: MIM166, MIM171 and MIM159/319, the last one has found strong support on the recent experimental work by Reichel and Millar [Specificity of plant microRNA TMs: cross-targeting of mir159 and mir319. J Plant Physiol 2015;180:45–8]. Finally, we stress the discussion on the major computational and associated experimental challenges that have to be faced in future ceRNA studies.

Key words: target mimics; ceRNA; consensus ceRNA motif; computational method; noncoding RNA; bioinformatics

Alexandre R. Paschoal is an adjunct professor at the Federal University of Technology, Paraná (UTFPR), Brazil. He received his PhD in Bioinformatics at the University of São Paulo (2012). His research includes noncoding RNAs, bioinformatics and pattern recognition approaches, especially for miRNA studies.
Irma Lozada Chávez is a PhD researcher at the Interdisciplinary Center for Bioinformatics, University of Leipzig, Germany. She received her bachelor (2004) and master (2006) degrees in Biology and Biochemical Sciences at the National Autonomous University of Mexico (UNAM). She uses transdisciplinary approaches from bioinformatics and astrobiology to understand the origin and evolution of life on the Earth and elsewhere in the Universe. She is a founder member of the Mexican Society of Astrobiology (SOMA).
Douglas S. Domingues is a Research Professor at the Department of Botany, Institute of Biosciences, São Paulo State University (UNESP) in Rio Claro, Brazil. He uses genomics and transcriptomics approaches in non-model plants to understand gene function, evolution of gene families and genome components, as well as molecular responses to environmental constraints.
Peter F. Stadler received his PhD in Chemistry from the University of Vienna in 1990 and then worked as Assistant and Associate Professor for Theoretical Chemistry at the same School. In 2002, he moved to Leipzig as Full Professor for Bioinformatics. Since 1994, he is External Professor at the Santa Fe Institute. He is also an External Scientific Member of the Max Planck Society since 2009 and a Corresponding Member Abroad of the Austrian Academy of Sciences since 2010.
Submitted: 7 March 2017; **Received (in revised form):** 18 April 2017. **Accepted:** 27 April 2017

© The Author 2017. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/bib/article-abstract/19/6/1273/3858378 by UB Leipzig user on 15 January 2020

Introduction

MicroRNAs (miRNAs) are one of the most studied noncoding RNAs (ncRNAs) so far [1]. miRNAs are small and endogenous RNAs, whose main function has been identified as negative regulators of specific messenger RNAs (mRNAs) in eukaryotes [2] (see definitions in Box 1). Typically, miRNAs act in a sequence-specific manner to interact with their targets by base pairing to complementary sites in either the untranslated regions (3′ or 5′ UTRs) or the open reading frames (ORFs). The miRNA:target interactions produce, thus, the silencing of specific protein-coding genes at the posttranscriptional level, by inducing mRNA cleavage and/or decreasing protein translation (Figure 1A). Through this mechanism, miRNAs act as key regulators of diverse developmental processes, stress response, metabolism and apoptosis [3, 4].

Although miRNAs have been identified in fungi [6, 7], plants [8–10] and animals [11], their biogenesis and mechanisms of action are distinct among the major eukaryotic supergroups [3, 11, 12]. For instance, miRNA:target pairing is almost perfectly complementary and extends over twice as many nucleotides in plants, when compared with animals and fungi [2, 13–15] (see Box 1). These empirical observations clearly suggest that miRNAs have arisen multiple times as endogenous components from a more general RNA interference machinery [2, 3].

It is becoming clear that miRNA interactions are not restricted to the 'classical' protein-coding mRNAs, but they extend to several non-protein-coding RNA classes containing competing miRNA-binding sites (Figure 1B). In 2007, Franco-Zorrilla and coauthors [5] found that miR399 can be sequestered by the ncRNA IPS1 in *Arabidopsis thaliana*. The transcripts of IPS1 and PHO2 mRNAs—the latter, a key gene involved in phosphorous homeostasis [16]—have a similar 23 nucleotide (nt) binding site that competes for miR399. In contrast to the miRNA-binding sites of PHO2, the IPS1 binding sites form a 3 nt mismatch loop or bulge opposite the miR399 cleavage site on miRNA binding (Figure 1C). This loop avoids the cleavage of IPS1 and sequesters miR399 instead, and as a consequence, the original PHO2 targets are deregulated. This mechanism describing the inhibition of miRNA activity is defined as 'target mimicry', and the transcripts sequestering or competing for a miRNA are globally known as 'competing endogenous RNAs' (ceRNAs), although they are also known as 'target mimics' (TMs) in plants and 'miRNA decoys' or 'microRNA sponges' in mammals [17, 18] (see Box 1). In the present review, we will use the term TMs when referring to experimentally and computationally identified TMs in plants, whereas the term ceRNAs will be used to refer those TMs involved in the general mechanism of miRNA regulation by target mimicry in eukaryotes. Table 1 summarizes the information available for ceRNA across several taxa. We particularly refer to Le and coauthors [18] for a comprehensive review in the animal kingdom.

ceRNAs include protein-coding RNAs, long ncRNAs, viral RNAs, pseudogenes, circular RNA (circRNA) species and artificial RNAs [15, 24, 25] (see definitions in Box 1). miRNA targets and ceRNAs have highly similar miRNA-binding sites, often called miRNA recognition elements (MREs) (Figure 1). The competition for these binding sites may differ among cell types because of differential expression of the decoy transcripts, and therefore may have a strong impact on the regulation of gene expression, as reviewed in [15, 28]. Thus, ceRNA networks can be formed when a collective pool of transcripts are synergistically sequestering the miRNA activity through competition for one or several miRNAs [15, 23]. Furthermore, ceRNA might not be restricted to eukaryotes, as similar mechanisms have been found in bacterial genomes, where RNAs may compete for small RNA regulators or regulatory proteins [27].

Because miRNAs have the potential to module the regulation of at least two-thirds of an eukaryotic transcriptome [15], ceRNAs have generated substantial interest as a relevant regulatory mechanism to modulated miRNA activity [2, 15, 23, 29]. Moreover, ceRNAs provide a powerful methodology to assess the potential function of any miRNA and uncharacterized RNA transcripts through the use of both naturally occurring endogenous target mimics (eTMs) and engineered artificial target mimics (aTMs). In plants, for instance, several aTMs have been engineered via the short-tandem target mimicry (STTM) technology [30, 31] (see Box 1) and viral vectors [32–34]. Few computational approaches have been also used to complement the prediction and design of potential TMs [35–39]. Several studies and discussions highlight the powerful use of aTMs in plant engineering and biotechnology [2, 19, 21, 22, 26]. There are, however, several major computational and experimental challenges to identify and validate targets mimics in plants.

Here, we provide an overview of the state of art in ceRNA research on plants along two major sections. In the first section, we emphasize on computational aspects behind the construction of databases, data availability as well as on the prediction and analysis of TMs. We further describe a general *in silico* procedure for TM prediction and their experimental validation. In the second section, we introduce a useful bioinfomatics approach for the prediction of TMs motifs that have the potential of cross-targeting both members within the same or from different miRNA families, based on the identification of consensus miRNA-binding sites from known long intergenic noncoding RNA:miRNA–TM interactions. Finally, highlights and challenges are exhibited at the end of this article to discuss the need for novel bioinformatics and biological applications in future studies.

A comprehensive overview of the TM literature and research

Literature survey: time line for the research on TM

As described in Supplementary Figure S1, we manually mined the literature in PubMed and applied several filters to collect those publications involved in the computational development of ceRNA research in plants. Although the number of publications has been growing each year, computational approaches focused on ceRNA research are still restricted. From the 107 papers obtained, only 20 were related to computational analyses. Supplementary Figures S2 and S3 show the cumulative number of these publications and their distribution by topics, half of them are related to specific analyses on ceRNA research (see 'Computational analysis' section), while the remaining publications involve the discovery of TMs and the development of tools and databases. Not surprisingly, almost half of the publications are concentrated on two genome models: *A. thaliana* and/or *Oryza sativa*. Figure 2 shows a time line of these publications. Most studies are based on the complementary-pairing pattern found for the classical miR399:IPS1 system in *A. thaliana* by Franco-Zorrilla [5]. It has provided a model for the development of the following rules for TMs identification (Figure 1C): (i) there is a central bulge of length 3 nts in the target sequence and between the positions 10 and 11 of the miRNA sequence (green or diagonal lines in the background positions at the

Downloaded from https://academic.oup.com/bib/article-abstract/19/6/1273/3858378 by UB Leipzig user on 15 January 2020

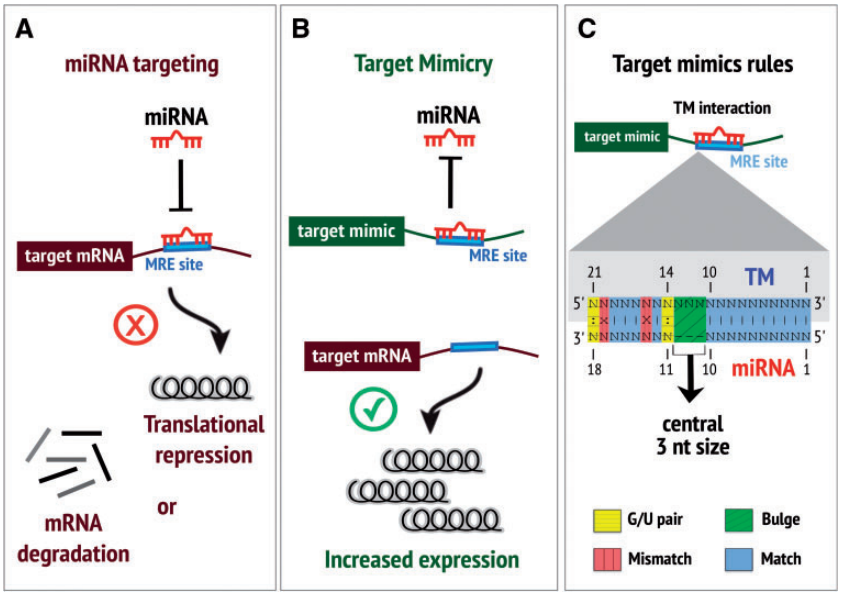

Figure 1. The target mimicry mechanism. (**A**) The general mechanism for miRNA:target-mRNA binding. In plants, MREs are located in several regions and also at the 3′ UTR of a target mRNA (in burgundy/black) and exhibit a more perfect complementarity binding with the 5′ region of the miRNA. The miRNA:target-mRNA binding leads to either the translational repression of the target mRNA or the target mRNA degradation. (**B**) The general mechanism for miRNA:TM binding. Any other transcript (rather than the target mRNA) is called TMs (in green/heavy grey) if it possesses MREs competing for a miRNA. Once the miRNA:target-mimics binding occurs, the sequestration of the miRNA leads to the derepression of the target mRNAs, and thus, to the increase of their translational expression. (**C**) A representation of the 'Target Mimic rules' derived from the classical TM model miR399:IPS1 discovered by Franco-Zorrilla [5] in *A. thaliana*. The rules state that: (i) there is central bulge of length 3 nts in the TMs, and between positions 10 and 11 of the miRNA (in green or diagonal lines in background); (ii) G/U pairs may appear only after position 10 of the miRNA and at the end of the interaction (in yellow or horizontal lines in background); (iii) one or two mismatches are tolerated after position 11 of the miRNA (in red or vertical lines in background).

Table 1. Summary of recent literature on ceRNA research

Article type	Main subject	Model species	Summary	Year	References
Review	ceRNA technologies	Plants	Description of main strategies to inhibit sponge–ceRNA and artificial miRNAs activities	2011	[19]
Review	RNA-based technologies	Plants	Discussion on the strengths and weaknesses in inhibiting the activity of miRNAs via target mimicry, anti-miRNA technology and transcriptional gene silencing of MIR gene promoters	2011	[20]
Commentary	ceRNA overview	Plants	Summary of the evidence for ceRNA in plants and animals, highlighting the pioneer discovery of ceRNA in plants	2011	[17]
Review	ceRNA engineering	Plants	Potential applications of artificial ceRNA for plant genetic engineering	2013	[21]
Research	ceRNA engineering	Plants	A comparison of different miRNA decoy technologies demonstrates that no one guarantees the strongest miRNA inhibition, but rather miRNA families respond differently to each approach	2015	[22]
Review	ceRNA engineering	Crops	Major strategies for developing miRNA/ceRNA-based manipulation of miRNAs and TMs for the potential improvement of crop programs	2015	[23]
Review	lncRNAs	Plants	A brief description of lncRNAs acting as TMs	2015	[24]
Review	circRNA sponge	Animals	Description of the biogenesis, properties and functions of endogenous circRNAs sponges, with special focus on human cancer. Information of tools and databases is also provided	2016	[25]
Review	RNA-based technologies	Plants and animals	A brief introduction on RNA-based technologies, including ceRNA and STceRNA technology (see Box 1.)	2016	[26]
Review	miRNA World	Animals	Deep description of bioinformatic methods and experimental techniques for exploratory studies of miRNAs and their targets	2016	[2]
Review	miRNA sponge	Animals	Description of *in silico* methods and databases for the identification, classification and analysis of miRNA sponge	2016	[18]
Review	ceRNA hypothesis	Animals	A critic evaluation of the evidence for and against the ceRNA hypothesis, with a focus to discuss the validation and cellular relevance of mimic–targets interactions	2016	[15]
Perspective	ceRNAs overview	Bacteria	Discussion about the similarities and differences between the ceRNA hypothesis and bacterial small RNAs regulation competing for RNA-binding proteins	2016	[27]

Downloaded from https://academic.oup.com/bib/article-abstract/19/6/1273/3858378 by UB Leipzig user on 15 January 2020

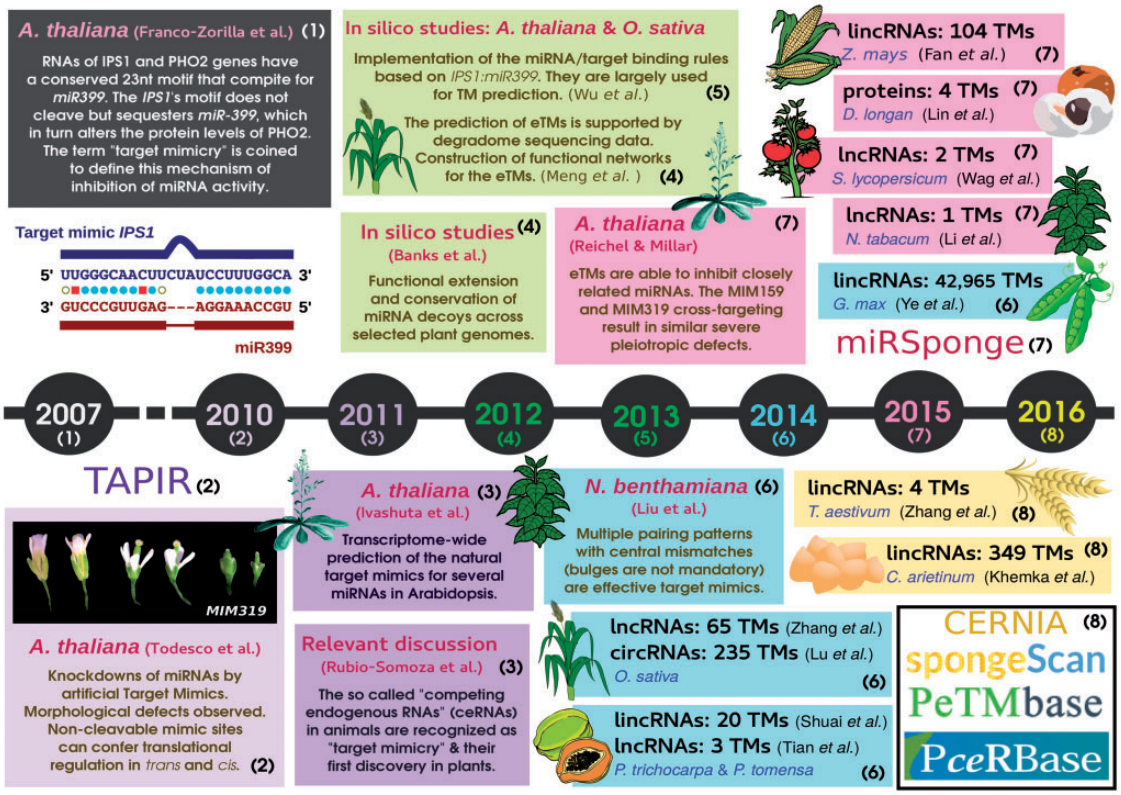

Figure 2. Time line for relevant publications on ceRNA research in plants since its discovery in 2007. The color/number of each box (also for tools/databases names) matches the color/number of the year in which the study was published. A brief description of the most remarkable finding(s) for each publication is provided. The number of TMs displayed for some studies is estimated as those miRNA:transcript interacting pairs that are reported as computationally predicted or experimentally validated TMs (either aTMs or eTMs). A representation of the classical TM model *miR399:IPS1* discovered by Franco-Zorrilla [5] in *A. thaliana* is provided in the corresponding publication. The figure depicting the morphological effects (smaller flowers) of the TM MIM319 in *A. thaliana* is partially reproduced from the Figure 3A by Todesco *et al.* [30] with the kind permission of *PLoS Sites* (www.PLOS.org) in accordance with the Creative Commons Attribution (CC BY) license. The images for model species were kindly provided by clipartkid.com, openclipart.org, cliparts.co and clipartbro.com.

Figure 1C); (ii) the G/U pairs may only appear after position 10 of the miRNA and at the end of the interaction (yellow or vertical lines in the background position at the Figure 1C); (iii) one or two mismatches are tolerated in the second stem, i.e. after position 11 of the miRNA. As we describe in next sections, these rules have been largely used in most computational studies and tools with some modifications.

Tools

The only tool available for the TMs prediction in plants is the TAPIR Web server [40], which is used to predict miRNA targets, but it also uses the classical rules to computationally predict TMs (Figure 3 and Table 2). Currently, TAPIR provides precomputed data for 10 plant genomes: *A. thaliana*, *Chlamydomonas reinhardtii*, *Glycine max*, *Medicago truncatula*, *O. sativa*, *Physcomitrella patens*, *Populus trichocarpa*, *Sorghum bicolor*, *Vitis vinifera* and *Zea mays*. TAPIR uses RNAhybrid [47] to compute potential interaction regions based on the thermodynamics RNA structure formation, and it also applies the 'Franco-Zorrilla rules' as a template to identify likely binding sites. Modifications of these classical rules, as described by Wu *et al.* [38], were further implemented in TAPIR; however, these rules are still based on a small set of pairing patterns derived from the first aTMs discovered in plants (Figure 3A). Yet, TAPIR is still the first step to facilitate computational analyses of miRNA:target interactions.

The programs SSearch from the FASTA3 package [48] and psRNATarget [49] are also largely used to search site complementary of the candidate targets to miRNAs. On the one hand, SSearch does a rigorous Smith–Waterman search for similarity between a miRNA and a group of potential target sequences [48]. On the second hand, the psRNATarget program predicts potential TM sites according to both approaches, the reverse complementary matching between the miRNA and target sequences, and the target-site accessibility calculated by the unpaired energy (UPE) required to 'open' secondary structure around the small RNA's target site on the mRNA [49].

Two new tools for TMs prediction could be also adapted for ceRNA analyses in plants (Table 2). spongeScan is a Web application tool for prediction of MREs in long noncoding RNAs (lncRNAs) acting as sponges [44]. As it accepts genome coordinates, this Web application could be used to make TM analysis in plants, but all current examples involve human lncRNA data. On the other hand, CERNIA is a computational method for inferring competing endogenous interactions [46]. It uses miRNA-binding sites to infer ceRNA networks in UTR and coding sequence (CDS) interactions across different ceRNA classes (mRNA, lncRNA and pseudogenes). Again, these applications are so far restricted to human genomic data.

Databases

A recent review by Le and coauthors [18] covered six main database resources for miRNA sponges with a focus on animals. From this review, only the miRSponge database [41] contains

Downloaded from https://academic.oup.com/bib/article-abstract/19/6/1273/3858378 by UB Leipzig user on 15 January 2020

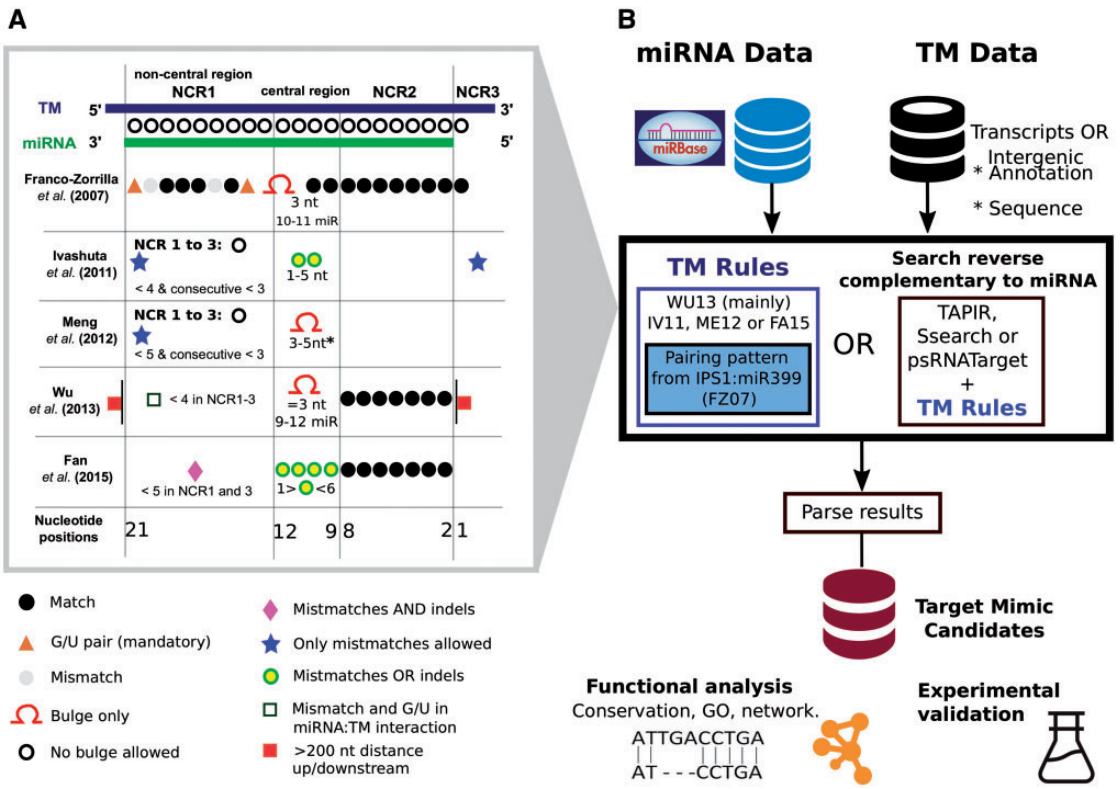

Figure 3. General computational workflow and rules for TMs identification. (**A**) The modifications of the TM rules according to different publications. A general miRNA:target-mimic model is provided at the top of the table, with marked nucleotide positions (from 1 to 21) and pairing regions (one central and three noncentral regions: NCR1, NCR2 and NCR3); the blue/black and green/gray lines represent the TM and miRNA sequences, respectively. The modifications of the TM rules follow the symbology provided below the comparative table. Important to note that Meng *et al.* [37] consider the position of the central regions according to the miRNA length: 9–11th of 19 nt miRNA size, 10–11th of 20 nt size, 10–12 of 21 nt, 11–12th of 22 nt, 11–13th of 23 nt and 12–13th of 24 nt. (**B**) Four major steps outline the process to identify TMs: (i) collection of known/predicted TM data set and a set of miRNAs of interest, (ii) the computational prediction of TMs, (iii) the functional annotation and (iv) experimental evidence for TMs. See text for more details.

Table 2. Tools and databases available for TM research

Type	References	Name	TM type	Species	Data source	Evidence	Year	Data	Site
DB	[41]	miRSponge	lncRNA and artificial ncRNAs	Ath and 10 non-plants	1200 articles	ML	2015	DW	www.bio-bigdata.net/miRSponge
DB	[42]	PceRBase	mRNA, lncRNA and other transcripts	26 plants	Phytozome v10, TAIR10, MSU RGCP7, miRBase, literature	C	2016	SM and DW	bis.zju.edu.cn/pcernadb
DB	[43]	PeTMbase	lncRNA	10 plants	GreeNC, PNRD, miRBase and NCBI SRA	C	2016	DW	petmbase.org
Tool	[40]	TAPIR	Any	10 plants	RNAhybrid and miRBase	C/TMR	2010	DW	bioinformatics.psb.ugent.be/webtools/tapir
Tool	[44]	spongeScan	lncRNA	Any (human)	Ensembl, miRBase and literature [45]	C	2016	SM	spongescan.rc.ufl.edu
Tool	[46]	CERNIA	mRNA, lncRNA, pseudogene	Cancer data	miRanda, DT-Hybrid and miRNA–target interactions	C/L	2016	SM	github.com/dsardina/cernia

Note: The number of TMs displayed for a particular study is estimated as those pairs miRNA:transcript that are computationally predicted or experimentally validated as TM interactions. DB, Database; C, computational evidence; ML, manual literature curation; TMR, target mimics rules; L, literature; DW, available to download from the Web site; SM, available to download from supplementary material; TAIR, The Arabidopsis Information Resource; TIGR, The Institute for Genome Research; NCBI SRA, NCBI Sequence Read Archive; MSU RGCP, MSU Rice Genome Annotation Project Database; GReeNC, the Green Non-Coding Database; PNRD, Plant Non-coding RNA Database; GtRNAdb, Genomic tRNA Database; NONCODE, noncoding RNA Database; RGAP, Rice Genome Annotation Project; RMD, Rice Mutant Database; ITAG = International Tomato Annotation Group.

Downloaded from https://academic.oup.com/bib/article-abstract/19/6/1273/3858378 by UB Leipzig user on 15 January 2020

experimentally validated miRNA targets and sponges, basically obtained from literature mining. As most of these data (90.8%) are related to human and mouse genomes, only a small fraction (1.6%) of the data is from plants.

Yet, two new databases are available for ceRNA research in plants (Table 2). The recently released PceRBase (Plant Competing Endogenous RNA dataBase) [42] is the first database specifically devoted to describe TMs in plants. This database provides TM target–target and TM target-mimic pairs for 26 plants species, which provides 167 608 TMs. The reported TM interactions are based on predictions with the TAPIR tool [40] and follow the TM rules described in 'Literature survey: time line for the research on TM' section. In addition, PceRBase provides functional annotation such as gene ontology (GO) and TMs networks.

The second ceRNA database is PeTMbase (Plant Endogenous Target Mimics database) [43], which contains 2728 TMs predicted for 11 species: *A. thaliana*, *Brachypodium distachyon*, *G. max*, *Hordeum vulgare*, *M. truncatula*, *O. sativa*, *P. trichocarpa*, *S. bicolor*, *Solanum lycopersicum*, *Triticum aestivum* and *Z. mays*. PeTMbase uses a lncRNA database in plants and transcriptome leaf library to characterize lncRNAs transcripts (Table 3). It also uses the TM rules implemented by Wu *et al.* [38] for TM identification (Figure 3A).

A protocol for the computational analysis of TMs

The computational studies described in next section use complementary strategies to predict, annotate and validate TMs. It seems worthwhile, therefore, to first introduce a consensus protocol for this task in four steps, as observed in Figure 3B: (i) the collection of miRNA and putative TM data sets, (ii) the computational prediction of TMs, (iii) functional annotation and (iv) experimental evidence for TMs.

Data sources

An initial data set consists of two components: (i) a set of miRNAs of interest, and (ii) a collection of TMs candidates. MiRNAs are taken from the miRBase [60], or derived from in-house experimental data or are extracted from specific publications. In particular, for poorly annotated non-model organisms, it may be necessary to first perform a homology search for miRNAs using miRBase entries of related species as starting point. Collections of RNA-seq-derived transcripts or genomic sequence data may serve as initial sources to predict potential TMs. One may also resort to catalogs of lncRNAs from third-party data, in particular for *A. thaliana* and *O. sativa*, for which the most extensive annotation is available by far. Transcript annotations with reliable UTRs and including lncRNAs may not be available for less-studied non-model organism. In general, non-model organism's data are based on computational analysis when this is available. In this case, genomic sequence is the only viable option. Of course, this has the disadvantage of a much larger search space and thus a lower accuracy for any TM prediction method.

Computational TM discovery

As represented in Figure 3, the computational approach for TM prediction can be carried out through two strategies, either independently or complementary. One approach involves the creation of in-house scripts to search for the complementarity-pairing patterns between selected miRNAs and transcripts through the direct use of either the original or modified TM rules. This approach provides of a precise filter criteria to the cost of a limited scope in the number of predicted TMs. The second approach involves the use of some tools implementing or complementing the TM rules, by searching, for instance, for complementary regions between the miRNA and its target mRNA with the SSearch [48] or the psRNATarget programs [49], or by searching base-pairing sites with the RNAhybrid program [61]. As described previously, the Web server TAPIR is also largely used in this case. After this step, a list of computational TM candidates is obtained.

Functional annotation

Usually, customized in-house scripts and generic tools are used interactively to analyze TM conservation and functional annotation. For instance, the assessment of GO ontologies can be carried out with the AgriGO [62] software, and some analyses of functional networks among TMs can be performed with the Cytoscape software [63, 64].

Experimental validation

Computational predicted TMs can be tested for experimental evidence of, at least, two key attributes: (1) inducing the inhibition of miRNA activity via target mimicry, and (2) its naturally occurring endogenous co-expression with the targeted miRNA. At present, the classical *miR399:IPS1* model, discovered by Franco-Zorrilla [5], is the only known system in plants that has robust evidence for these two conditions. Given that a detailed review of all experimental approaches used to valid TMs is beyond the scope of this contribution, we only outline here the main experimental principles and the most common methods used to achieve them. We refer to the references provided in Table 2 for accessing to complete reviews on these topics.

Regarding the first attribute, the designed experiments should demonstrate, or at least provide enough evidence, that the potential TM is not cleaved by the miRNA, but instead, that the potential TM inhibits the repressive effect of the miRNA on their mRNA targets. Thus, analyses on transcript expression and/or degradome sequencing data are often used to proof one or both of the following observations. First, degraded mRNA targets, but nondegradation of the potential TM, are observed after the miRNA is overexpressed; and second, the expression of the miRNA is not detected, but an overexpression of their mRNA targets is, after overexpressing the potential TM. Any engineered transcript or computationally predicted TM fulfilling these conditions is considered a validated aTM, but it is not necessarily an eTM, as this TM may not function as such under endogenous conditions. Several methodological approaches can be used to measure levels of expression, including sRNA northern blotting, quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) [35, 55, 59], RNA ligase-mediated (RLM)-rapid amplification of complementary DNA ends (RACE) [54], poly(A) polymerase-mediated-RACE, among others.

The second attribute is by far the most difficult to address, as it involves the co-expression of both the miRNA and the aTM under naturally occurring conditions in the plant. In several cases, the identification of the miRNA and aTM in publicly available expression data under the same or similar conditions is used to assume co-expression of both RNAs. The aTMs covering this step could already be considered as potential eTMs. However, measuring miRNA levels in either *in vivo* conditions or transgenic plants expressing the desired aTM is by far the best way to assess co-expression. Those TMs being functionally analyzed under their native promoter, such as IPS1 [5], are robustly validated as 'naturally ocurring eTMs'. However, those aTMs and eTMs overexpressed through strong constitutive

Downloaded from https://academic.oup.com/bib/article-abstract/19/6/1273/3858378 by UB Leipzig user on 15 January 2020

Table 3. Summary of computational analysis and data availability of TMs in plants

Scale	Species	TM transcript	TMs number	Data sources[a]	Computational evidence	Experimental analysis[b]	Reference
Large	*A. thaliana*	cDNA (protein coding, short peptide, pseudogene, TE, lncRNA and other RNA)	324	miRBase v15, TAIR v9 (cDNA)	SSearch (FASTA3) and in-house scripts based on modified TM rules from [5, 30]	miRNA, phenotypic analysis and transgenic analysis (qRT-PCR; northern blot)	Ivashuta *et al.* [35]
Large	*A. thaliana and O. sativa*	cDNA	300 and 260	miRBase v17, TAIR v10, TIGR v6.1, GtRNAdb, NONCODE v3, agriGO and Degradome (target)	SSearch (FASTA3) and in-house scripts based on TM rules [5, 30, 35]	Network and GO analysis	Meng *et al.* [37]
Large	*A. thaliana and O. sativa*	lncRNAs	36 and 189	miRBase v18, TAIR v10, RGAP v7, PlantGDB (cDNA, EST and transcripts) and RNA-seq	Prediction tools and in-house scripts based on TM rules [5, 30, 35]	Expression (10 transcripts) and co-expression of lncRNAs and miR160/miR166	Wu *et al.* [38]
Large	*G. max*	Intergenic	42 965	miRBase v20, Phytozome, NCBI [Transcriptome (mRNA and Small RNA) and Degradome (Target)]	psRNATarget, psRobot, BLAST and in-house scripts based on TM rules [38]	Expression and no degradome evidence in 457 TM (transcripts)	Ye *et al.* [39]
Large	*Z. mays*	lincRNA	104	miRBase v21, Phytozome v9.1, MaizeGDB (cDNA), NCBI GEO and AgriGO	BLAST, MAFFT v6.864b and GSTAr.pl script based on a slight modification of TM rules [35, 38]	Conservation and network/GO analysis; seven lincRNAs are potential TMs in drought-stressed leaves	Fan *et al.* [50]
Small	*P. trichocarpa*	lincRNA	20		psRNATarget and in-house scripts based on TM rules [38]	Expression of lincRNA20 and lincRNA2752 in drought response; conservation of potential TMs (lincRNA432 and lincRNA1174 for miR160), as in [38]	Shuai *et al.* [51]
Small	*O. sativa*	lncRNA	65	RGAP v7, RMD database (mutant)	Prediction tools and in-house scripts based on TM rules [38]	Transient transformation (XLOC_0063639, XLOC_007072) and co-expression (lncRNA and mRNA) for potential TMs of miR160 and miR164	Zhang *et al.* [52]
Small	*D. longan*	Protein coding	4	Longan EC transcriptome database (SRA050205)	TAPIR	Co-expression of two TMs (transcript and miRNA)	Lin *et al.* [53]
Small	*N. tabacum*	lncRNA	1	Plant Genome Database and small-RNA data	In-house scripts based on TM rules [39]	No degradation evidence (degradome and RLM-RACE); overexpressed/silenced transgenic nta-eTMX27	Li *et al.* [54]
Small	*S. lycopersicum*	lncRNA	2	miRBase v21, ITAG2.4	psMimic and psRobot tool rules based on [38]	Co-expression of lncRNA/miRNA	Wang *et al.* [55]
Small	*O. sativa*	circRNA	235 MREs	miRBase v21	In-house scripts based on TM rules [37]	Transgenic overexpression (Os08circ16564 and miR172), but no evidence for TM action	Lu *et al.* [56]
Small	*C. arietinum*	lincRNA	349	NCBI GEO (RNA-seq), miRNAs from and GO terms annotation	psRNATarget and network analysis	Conservation analysis	Khemka *et al.* [57]
Small	*T. aestivum*	lincRNA	4		TAPIR	Co-expression of miRNA and lincRNA	Zhang *et al.* [58]
Small	*P. tomentosa*	lncRNA	3	miRBase v20	Predicted rules based on [38]	Co-expression analysis	Tian *et al.* [59]

Note: See Table 3 for name descriptions of databases and genome projects.
aMain data sources used in the article.
bAny experimental evidence to support TM predictions.

Downloaded from https://academic.oup.com/bib/article-abstract/19/6/1273/3858378 by UB Leipzig user on 15 January 2020

promoters via viral vectors (as it usually occurs in the remaining cases reported so far) may misrepresent their endogenous role under naturally (*in vivo*) occurring conditions.

Notable changes in plant phenotypes and/or the examination of binding complementarity (as in [65, 66]) might further be assessed for both aTMs and eTMs. For the former, some tests for transgenic expression of aTMs under designed or *in vivo* conditions are performed to generate the loss-of-function outcome for the desired miRNAs [30, 35, 38, 65, 66]. As a pioneer study, for instance, Todesco and coauthors [30] engineered the IPS1 gene to generate 73 aTMs to test their inhibit effect on the majority of known miRNA families in *A. thaliana*. Several morphological defects in the aerial part of the plant were observed as a consequence of 15 aTMs targeting ~20% of the analyzed miRNA families. Another interesting case is represented by the study of Li and coauthors (2015), who found that the lncRNA nta-eTMX27 is likely responsible for the upregulation of the *QPT2* gene (a key regulator of nicotine content) in tobacco plants by inhibiting the action of nta-miRX27 under *in vivo* topping conditions. Transgenic expression of nta-eTMX27 under a constitutive promoter in *Nicotiana tabacum* and the absence of its degradation products based on degradome and RLM-RACE data was reported as functional evidence [54].

Computational analysis

Here, we summarize 14 computational analyses related to the identification of TMs in plants (Table 3). Seven of them report large-scale systematic TM predictions over four species: *A. thaliana* [35, 37, 38], *G. max* [39], *O. sativa* [37, 38] and *Z. mays* [50]. In total, these computational studies report 44 861 putative TMs (Tables 2 and 3), 95.77% of which are derived only from the *G. max* report [39]. The remaining publications are case-specific analyses. We briefly describe these studies next.

Ivashuta *et al.* [35] implemented the first computational analysis to systematically search and predict TMs in the *Arabidopsis* transcriptome based on the miRNAs reported in miRBase (Table 3). Additionally, to the position of the 3 nt bulge structure already shown to be critical to inhibit miRNA activity [5, 30], the authors also allowed miRNA-binding sites to form 1–5 nts bulges between positions 10 and 11 or with mismatches at positions 10 and 11 (Figure 3). Based on the study of Todesco *et al.* [30], the search parameters also allowed for the miRNA decoys to be located into the UTR of protein-coding transcripts. Thus, MREs following these rules were searched against complementary DNAs (cDNAs) for protein coding (CDS), short peptides, pseudogenes, transposable elements (TEs), lncRNAs and other RNA transcripts. They identified 324 putative decoy sites in 317 transcripts belonging to 260 loci, including the locus of the IPS1 decoy. The authors found that most predicted TMs are mainly located within CDS and as abundantly in UTR regions as ORF regions. Furthermore, most of the TMs exhibited only one or two mismatches in their MREs, while a minor number possess mainly a 1 or 3 nt bulge. Importantly, expression analyses (e.g. northern blot and qRT-PCR) of miRNAs and TMs in transgenic plants support these predictions.

Soon after, Meng *et al.* [37] carried out a genome-wide computational prediction of TM across *A. thaliana* and *O. sativa*. cDNA data were used to search TMs, and according to a slight modification of the TM rules by Franco-Zorrilla *et al.* and Ivashuta *et al.* [35]. As shown in Figure 3, the presence of 3–5 nt bulges in the central region—but no mismatches and indels—was mandatory. Thus, 300 and 260 TMs were predicted in *Arabidopsis* and rice, respectively. In contrast to Ivashuta *et al.* [35], the densities of the TMs sites were found to be higher within the UTRs rather than within CDS. After degradome sequencing, data were used to identify miRNA targets; furthermore, the authors constructed and investigated mimic–miRNA–target regulatory networks.

Wu *et al.* [38] developed a computational methodology to predict TMs in plants based on 20 miRNAs conserved between *A. thaliana* and *O. sativa*. As shown in Figure 3, this study implemented major modifications to the TM rules, particularly the upstream/downstream distance (>200 nt) of the MRE within the TM sequence. A total of 36 TM candidates were identified for 11 miRNAs in *A. thaliana*, while 189 TMs were predicted for 19 miRNAs in rice. Around 40% of the predicted TMs show transcriptional evidence based on RNA-seq data, as well as the expression of miRNAs and some putative targets was evaluated by semiquantitative RT-PCR and northern blot analysis in transgenic plants. Based on this evidence, the authors suggest that these eTMs could be potentially involved in the regulation of plant development by competing for *miR160* and *miR166*. Finally, overexpression analyses in leaves of transgenic *Nicotiana benthamiana* were also carried out for miR156, miR159 and miR172. This study demonstrates that, regardless the similarity of the targets, different aTMs can compete for the same miRNA through well-conserved MREs.

Two additional studies have used a further modification of the TM rules implemented by Wu *et al.* [38] (Figure 3). On one hand, Ye *et al.* [39] performed a computational genome-wide analysis in *G. max* and predicted 42 965 TMs candidates. Only 457 aTMs (~1%) have expression evidence, and 119 of these aTMs were found to target 33 miRNAs and be potentially involved in the lipid metabolism of soybean and other four crop genomes. On the other hand, Fan *et al.* [50] extended the prediction of 104 miRNA-long intergenic noncoding RNA (lincRNA) duplexes across 86 lincRNAs in *Z. mays* and found 58 TM candidates. A conservation analysis of the potential TMs was performed using BLASTN and MAFFT. Functional miRNA-lincRNA-mRNA networks were explored to infer putative lincRNAs acting as decoys. Only seven linRNAs:miRNA candidates were supported by co-expression analyses, and their activity is suggested to be potentially associated with drought stress.

Nine additional publications focused on more specific systems. circRNA-miRNA interactions were investigated in *O. sativa* [56]. A total of 235 TM candidates were found, and co-expression analyses in transgenic plants provided evidence for two of them: Os08circ16564:miR172 and Os08circ16564:miR810). The remaining eight studies used transcriptome data to evaluate the impact of stresses or development on aTM activity. As part of the computational identification of lncRNAs expressed in tomato in response to the infection by yellow leaf curl virus, the co-expression of two lncRNAs (slylnc0195 and slylnc1077) and their corresponding targeted miRNAs (*miR166* and *miR399*) was analyzed by qRT-PCR [41, 55].

A study in rice identified 65 lncRNAs as TM candidates to compete for conserved miRNAs, by using transcriptome data in reproductive tissues [52]. Transient transformation of XLOC_0063639 and XLOC_007072, as well as their co-expression with *miR160* and *miR164*, was provided as evidence for TMs. Additionally, three TMs were identified in *Populus tomentosa* related to the responses to gibberellin treatment: TCONS_00264314:ptc-miR6459b, TCONS_00015177:ptc-miR169y and TCONS_00013311:ptc-miR169y [59]. The authors also reported expression changes of the miRNA target Potri.002G059900 as a consequence of the TCONS_00264314:ptc-miR6459b interaction. In *N. tabacum*, nta-miRX27 was found to

Downloaded from https://academic.oup.com/bib/article-abstract/19/6/1273/3858378 by UB Leipzig user on 15 January 2020

be sequestered by the lncRNA nta-eTMX27, which modulates the expression of the *QPT2* gene in the nicotine biosynthesis pathway [54]. RT-PCR and functional analyses using transgenic plants were conducted to provide experimental support for the regulatory nta-miRX27:nta-eTMX27 circuit. This study is the first to depict ceRNA function in plant secondary metabolism.

Khemka *et al.* [57] used RNA-seq data from eight successive stages of flower development and three vegetative tissues in *Cicer arietinum* to predict TMs. They found 349 lincRNA-miRNA pairs; however, the authors do not clarify how many of these pairs are considered as aTMs and/or eTMs. Functional annotations of the TMs were inferred from network interactions. Zhang *et al.* used TAPIR to survey intergenic and noncoding regions of *T. aestivum* based on transcript data produced during the response to fungal infections [58]. The authors reported four lincRNAs as potential aTMs. On the other hand, Shuai *et al.* [51] reported 8 sense and 12 antisense aTMs during a drought state in *P. trichocarpa*. Finally, four putative protein-coding TMs were identified in *Dimocarpus longan* [53], as part of an analysis of the transcriptomic response to the miR160 family. The co-expression patterns of two TM candidates (the transcripts Unigene64827 and Unigene66891) and of the targeted miRNA (dlo-miR160) provided support for their potential role during somatic embryogenesis in *D. longan*.

Summary of available data

We estimated 215 207 candidate TMs derived from 17 computational studies and three databases (Tables 2 and 3). The vast majority of these predictions correspond to the reports of *G. max* (19.96%) [39] and the PceRBase (77.88%) [42]. The overlap between these two resources remains unclear. PceRbase contains TM data from 26 plant species, where almost 87% of then come from only 4 species: *A. thaliana* (∼45%), *S. bicolor* (15.2%), *O. sativa japonica* (13.9%) and *G. max* (12.7%). The remaining 4634 potential TMs are reported in literature for 15 species; however, most of these TMs candidates are from studies in *O. sativa* (25.4%), *A. thaliana* (23.3%), *Z. mays* (12.4%), *G. max* (10.5%) and *C. arietinum* (7.5%), while the other 10 plant species represent 20.7% of the data. This second set of available TMs candidates is distributed among six categories: artificial RNAs, cDNAs, circRNAs, lincRNAs, lncRNAs and other RNAs and protein coding.

The ceRNA motif approach

Many plant miRNA families are large and comprise many loci encoding identical (or similar) miRNAs [67], with individual family members often expressed both spatially and temporally [68, 69]. We therefore asked whether TM-binding sites are diverse enough within the same miRNA family members and across different miRNA families to share a consensus TM motif. To explore this idea, we developed here a useful computational approach to predict candidate TMs with a potential to cross-target several members of the same or different miRNA families. Our approach is based on the identification of consensus miRNA-binding sites from known TMs across sequenced genomes, transcriptomes and known miRNAs in plants. Two hypothesis are key to justify this approach. First, miRNAs within the same or different families might share mRNA targets through the same or similar MRE motifs within an organism. Second, some of these TM networks might be extended among other plant species owing to the conservation of miRNA families or the expression of similar mRNAs and transcripts. Accordingly, our approach encompasses the following five steps (Figure 4).

Method

Collection and filtering of TM data

We first extracted miRNA:TM-binding sequences from the supplementary material of four studies [50, 51, 57, 58]. We limited the analysis to 40 nonredundant miRNA:lincRNAs-binding sequences with available experimental evidence for target mimicry, co-expression data or sequence conservation analysis (Supplementary Data D1). It is important to note that the miRNA:lincRNA-binding sequences selected for this study might or might not represent aTMs and eTMs. Also, the genome sequences of *C. arietinum*, *P. trichocarpa*, *Z. mays*, *A. tauschii*, *T. aestivum* and *A. thaliana* were retrieved from the Phytozome, Ensembl Plants and GigaDB databases (Supplementary Table S1).

Building of consensus TM motifs

We then applied a direct clustering of MRE sequences with the MAFFT alignment program [70] and the following parameters –retree 1000 –maxiterate 1000 –localpair –clustalout. Manual curation was performed to define the clusters of potential MRE-binding motifs; a minimum of three sequences is required to build a consensus motif. We further used the MEME program [71], with a maximum motif size of 20 nts, to obtain a consensus sequence for each MRE cluster. The final consensus TM motifs were obtained by complementing the results from MAFFT and MEME approaches (Figure 4).

Localization of TM motifs across genomic features

To identify the potential features harboring TM motifs, we performed a homology-based search of all consensus MRE motifs obtained previously against the complete genomes of the six species with the BLAST program [72] and the parameters blastn -word_size 6 -dust yes. All those hits with 95% of the alignment coverage and 100% of the identity are regarded as potential MRE sequences. The genomic location of each MRE hit was assigned to the coding (i.e. CDS), intronic or intergenic regions according to the gene annotation file of the corresponding genome. If multiple MRE hits are found for the same genome location, we considered only the hit with the lowest *E*-value and the highest BLAST score.

Mapping of TM motifs across transcriptomic data

To evaluate the potential of the consensus TM motifs to be part of transcripts potentially acting as miRNA decoys, the consensus MRE-binding motifs were searched across the reference transcripts of the six species with two methods. We used the psRNATarget program with the parameters: maximum expectation: 3.0, hspsize: 18, number of targets: 100, UPE: 25.0, flanking length: 17 to 13 nts, and central mismatch: 9–11 nts. We restricted the search of MRE motifs to 18-mers owing to the high rate of spurious occurrences on a genome/transcriptome-wide analysis. We also used TAPIR with the default parameters to identify more conserved and significant predicted TM sites. The use of both approaches allowed us to identify potentially relevant species-specific TM interactions and to restrict the number of false positives. We took into account TM matches located in different transcripts, although some of them might be different isoforms from the same transcript.

Downloaded from https://academic.oup.com/bib/article-abstract/19/6/1273/3858378 by UB Leipzig user on 15 January 2020

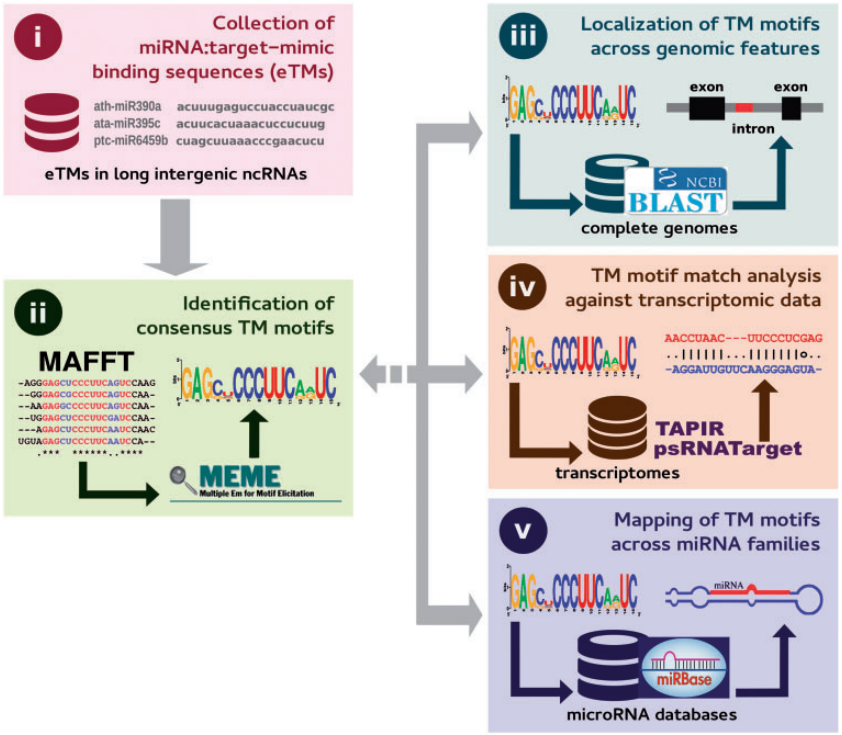

Figure 4. The ceRNA motif discovery approach. The workflow is developed in five major steps: (i) collection of known TMs from databases and/or published data, 40 TMs from lncRNAs were selected for this study; (ii) clustering of TMs with the MAFFT program and identification of consensus TMs based on sequence similarity with the MEME program; (iii) localization of TM motifs across genomic features with the BLAST program; (iv) TM motif match analysis against transcriptomic data with both TAPIR and psRNATarget programs; and (iv) identification of miRNAs potentially harboring such consensus TM motifs by searching against known miRNAs, we used here the miRBase, SSearch and BLAST programs for this purpose. Note that the depuration of the consensus TM motifs can be done at any step of the approach.

Identification of miRNAs potentially harboring the consensus TM motifs

Finally, the identification of the original and potentially new miRNAs harboring the consensus TM motifs was assessed for each of the species. Each consensus TM motif was compared against the mature sequences of known miRNAs families from the miRBase version 21 with the Ssearch and BLAST programs [48]. The mature miRNA sequences for *C. arietinum* were obtained from the report by Jain *et al.* [73]. In addition to the six initial species, we also included in the analysis: eight eudicots species, four monocots species, *Amborella trichopoda* and *Selaginella moellendorffii*. Those hits with an $e-value < 1$ were taken as potential targeted miRNAs.

Results

We carried out a five-step approach to detect consensus TM motifs potentially cross-targeting family members and/or different miRNA families across plants. We originally created six consensus TM motifs, dominated by MRE sequences from *C. arietinum*, the plant species with the larger number of MRE:lincRNA-binding sequences publicly available (Figure 5A and Supplementary Figure S4). However, half of these TM motifs were eliminated after poor performance was found with the further steps of our approach (Supplementary Table S3).

Figure 5A shows the three consensus TM motifs that performed best in our approach. They include 18 original MRE sequences from four main miRNA families: miR159, miR166, miR171 and miR319. First, we observe that members of miR159 are shared between TM motifs 1 and 3, and also that TM motifs 1 and 2 are *C. arietinum* specific. However, TM motif 2 (MIM171) is specific for members of the miR171a, d, e, f family, whereas TM motif 1 (MIM159/319) is formed by members of miR319a, f and miR159a families. In contrast, TM motif 3 (MIM159/166) is formed by two different miRNA families (miR166b, d, e and each miR159d, e) from two different species, *C. arietinum* and *Z. mays*, respectively. As shown in Supplementary Table S2, several BLAST hits for the three candidate TM motifs were found in all species: MIM159/319 (33 candidates), MIM171 (42 candidates) and MIM159/166 (73 candidates). These candidate TM motifs are particularly located in intergenic regions (81.9%), while minor fractions are observed in exons (15.4%) and introns (2.7%).

Figure 5B shows the distribution of 1095 and 783 matches found for the three TM motifs with the psRNATarget and TAPIR programs, respectively (Supplementary Table S3). Differences in the outcome of both programs are expected, given that psRNATarget allows more relaxed complementarity patterns, while TAPIR restricts the complementarity parameters to the presence of central bulges. Robust analyses are further necessary to measure false positives.

Finally, Figure 5C shows the distribution of known miRNA families potentially able to match any of the three consensus TM motifs across selected plant species. The three TM motifs recovered their corresponding miRNA families in *C. arietinum*, but not in *Z. mays*. As observed in the alignment of the TM motif 3 in Figure 5A, the failed recovery of zma-miR159d, e by the MIM159/166 motif can be explained by its high divergence at those variable nucleotide positions from the two maize sequences. Thus, we suggest that this TM motif is restricted to members of the miRNA166 family, that is, a consensus MIM166 motif. Although the performance of each TM motif is different within each miRNA family and across species, their distribution

Downloaded from https://academic.oup.com/bib/article-abstract/19/6/1273/3858378 by UB Leipzig user on 15 January 2020

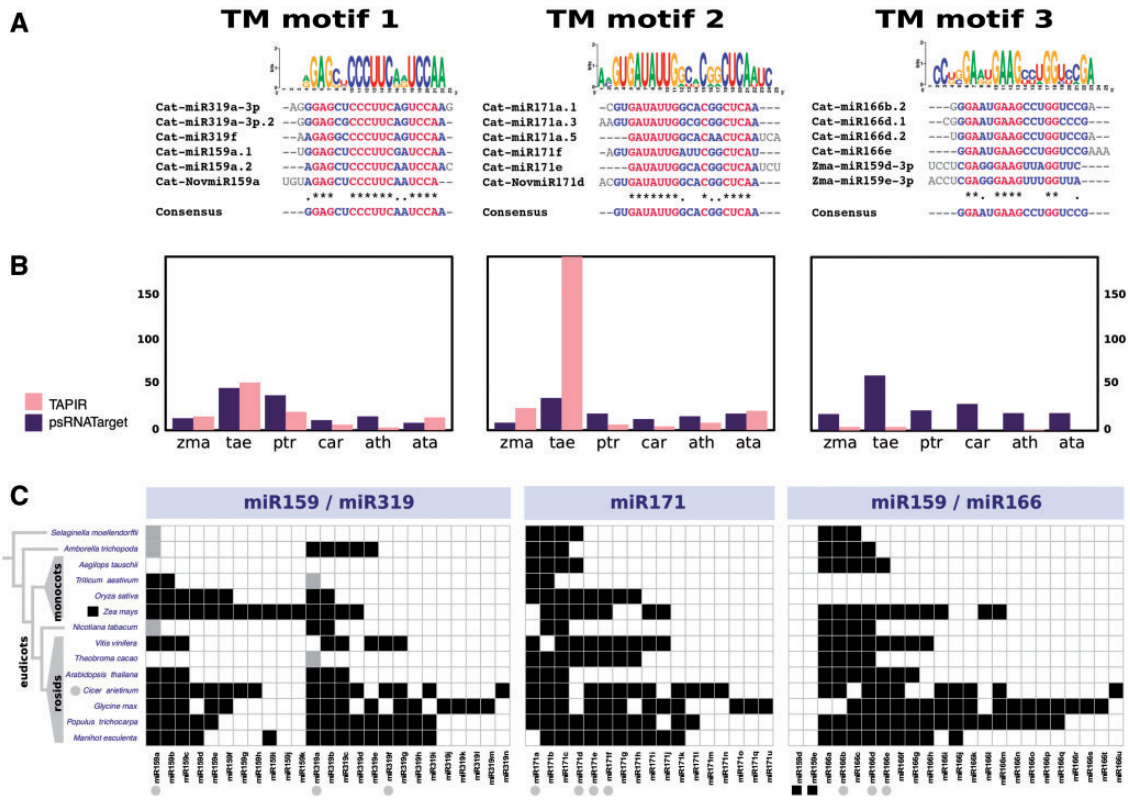

Figure 5. Performance of the three consensus TM motifs identified by our approach. (**A**) The three consensus TM motifs found with our approach are shown. TM motif 1 is formed by two miRNA families (miR159 and miR319) and TM motif 2 represents only members from the miR171 family in *C. arietinum*, whereas TM motif 3 includes members of two miRNA families (miR159 and miR166) from *C. arietinum* and *Z. mays*. A WebLogo was used to visualize the three consensus TM motifs (http://weblogo.berkeley.edu/). The logo represents the sequence conservation for every nucleotide on each position of the alignment. (**B**) The number of transcripts that matched every consensus TM motif, according to psRNATarget and TAPIR, is shown for the six species from which the original TM data were collected. (**C**) The distribution of the miRNA families potentially matching any of the consensus TM motifs (black filled squares) is shown for a selected collection of species. Black squares and gray circles are highlighting the miRNA members and species forming the corresponding consensus TM motif. Note that non-filled squares (or white squares) denote the absence of a candidate miRNA:TM-motif owing to either the absence of that miRNA (family or member) in the genome or the absence of that TM-binding site on the miRNA (family or member). Also, gray filled squares denote the presence of the candidate miRNA:TM-motif, but because no particular family member (isoform) is annotated, the match was assigned directly to the first member of the miRNA family.

patterns agree with the evolutionary proximity of the species to *C. arietinum*. Accordingly, *C. arietinum*, *P. trichocarpa*, *G. max* and *Morchella esculenta* exhibit a consensus MIM159/319 motif potentially conserved between Fabales and Malpighiales. Furthermore, consensus MIM171 and MIM166 motifs are not only potentially conserved in the previous species but also throughout the Rosids clade. Noteworthy, the evolutionary extent of these consensus TM motifs might be biased by an incomplete data set of known miRNA families contained in miRBase.

Discussion

We report here three consensus TM motifs: MIM159/319, MIM166 and MIM171. Interestingly, the TM motif MIM159/319 is supported by previous studies suggesting or demonstrating the cross-targeting mimics of both miR319 and miR159 under non-endogenous conditions [5, 30, 57, 66]. It is well known that miR159 and miR319 are related and ancient miRNA families essential for plant development and fertility, whose their 21 nt mature sequences still share 17 identical nucleotides [8, 107, 108]. However, miR159 and miR319 have different sequence specificity and expression patterns *in vivo*, which leads to the regulation of different developmental processes through distinct targets [16, 109–111] (Table 4). While miR159 limits the expression of the mRNAs encoding MYB transcription factors (TFs), miR319 moderates the activity of the mRNAs encoding TCP-TFs [16, 109, 111]. Under artificial conditions, however, Franco-Zorrilla *et al.* [5] observed the deregulation of the MYB genes by an artificial MIM319 in *A. thaliana*, whereas Todesco *et al.* [30] found similar severe pleiotropic defects provoked independently by MIM159 and MIM319.

Only recently, Reichel and Millar [66] demonstrated that the artificial MIM319 and MIM159 are indeed able to inhibit these two closely related miRNAs and generate highly similar developmental defects. Most importantly, the authors found that the central 3 nt bulge is not mandatory for inhibition of miRNAs, but rather poor central complementarity between miR319/159 and MIM319/159 is sufficient to confer an inhibitory effect [66]. This is agreement with the finding by Liu and coauthors [65] that multiple pairing patterns with central mismatches are effective aTMs of miR164 in *N. benthamiana*, regardless of whether efficacy was measured at the mRNA or protein level. In addition, a potential cross-targeting mimics of miR319 and miR159 has been also predicted by the lincRNAs Ca_linc_1728 and Ca_linc_1666 in *C. arietinum* [57]. These findings provide support for the existence of a consensus MIM159/319 motif potentially able to inhibit two close-related miRNA families across several

Downloaded from https://academic.oup.com/bib/article-abstract/19/6/1273/3858378 by UB Leipzig user on 15 January 2020

Table 4. Summary of biological information for the miRNAs analyzed with our 'consensus ceRNA motif' approach

miRNA	Function	Targets	Species	References
miR159	Flowering time control in short-day photoperiods. Leaf, anther and receptacle development. Seed size and shape determination	MYB33, MYB65, MYB101, GAMYB and GAMYB1	Ath, Hvu, Lot, Asa, Osa and Fra	[69, 74–84]
miR166	Embryo development. Floral organ polarity. Leaf primordial, shoot apical meristem and floral stem cell development. Root and nodule development	HD-ZIP III, PHB, PHV, REV, RDL, HB and HB8,10 and15	Ath, Mtr, Gma, Osa, Zma and Ppa	[85–95]
miR171	Negative regulation on shoot branching. Root colonization regulation	SCL6-II, III and IV and NSP2	Ath and Mtr	[96, 97]
miR319	Leaf morphogenesis, complexity and senescence. Jasmonate synthesis. Petal growth and development. Male and female gametophyte development	TCPs	Ath, Ama and Sly	[76, 98–106]

Note: Ath: *A. thaliana*, Hvu: *H. vulgare*, Lot: *Lolium termulentum*, Asa: *Avena sativa*, Osa: *O. sativa*, Fra: *Fragaria*, Mtr: *M. truncatula*, Gma: *G. max*, Zma: *Z. mays*, Ppa: *P. patens*, Ama: *Antirrhinum majus*, Sly: *S. lycopersicum*.

plant species, if and only if the sufficient levels, timing and spatial expression patterns of the TMs and the miRNA families on each plant occur naturally. In agreement with Wu *et al.* [38], the consensus TM motifs MIM171 and MIM166 could also be part of TMs potentially regulating several members of the same miRNA family and their corresponding mRNA targets (Table 4).

It is important to note that the robustness of our approach is based on the sequential implementation of these five steps because they are complementary. For instance, the implementation of the BLAST filtering Step (iii) and the miRNA target Step (iv) allows us to robustly remove artifacts or less likely TM motifs, and to further depurate/incorporate sequences in the consensus TM motifs. We consider that the 'ceRNA motif approach' is useful to design potential regulatory circuits and networks within and across plant species, not only because it allows the extension of TMs between members of the same and different miRNA families but also because it retrieves information to evaluate the evolutionary and regulatory divergence of sequences harboring TMs. For instance, Ivashuta *et al.* [35] have reported variations in the efficiency of regulation derived from the artificial design of TMs with different sequences. Furthermore, the design of aTMs with consensus TM motifs might help to provide a single non-cleavable target site, and hence to engineer a single plant expression vector, that simultaneously reduces the expression of all members of a miRNA family sharing such TM motif under the same or different expression conditions throughout the plant [20]. Thus, individual members of the same miRNA family harboring distinct TM motifs can then be identified in advance and tested independently for differential expression and function. In *A. thaliana*, some individual family members have shown indeed different phenotypes after disrupting their expression, such as miR164c (floral phenotype) [112] and miR164a, b (vegetative phenotype) [68, 113]. It would seem logical thus to extend the 'ceRNA motif approach' to a large-scale analysis.

Concluding remarks and perspectives

The ceRNA hypothesis suggests that, under certain conditions, highly abundant MRE-bearing RNAs have the potential to titrate away miRNAs from their other targets, and thus to compete for posttranscriptional control [2, 15, 29]. Nevertheless, the scope of ceRNAs as a global mechanism regulating miRNA activity, and thus their role as ubiquitous components of a self-regulatory RNA network, is still under debate. As summarized in this work, the understanding of ceRNAs in plants has been promoted by the increased number of studies on experimental and computational analyses, tools and databases published during the past decade. Nevertheless, major computational and experimental challenges remain.

To start with, the current collection of experimentally validated TMs could be computationally expanded to more genomes and transcript classes, such as plant pseudogenes and viral RNAs. There are 215 207 TMs predicted in >26 plant species. According to the statistics from the PceRBase Web site, however, 77.8% of the available TM candidates are mainly derived from four species: *A. thalinana*, *G. max*, *O. sativa* and *S. bicolor*. Likewise, only 4 of 17 studies described TMs that are not derived from lncRNAs. The next logic step is, therefore, to assess the evolutionary conservation and expansion of the available TMs across at least 80 completely sequenced plant genomes available, mostly at the Ensembl Plants and Phytozome databases. Computational approaches such as the 'consensus ceRNA motif' developed here might substantially contribute to fulfill this task. Likewise, it is necessary to build more repositories, not only to integrate different data sets but also to offer further exploratory analysis regarding the possible functions, regulatory networks and evolutionary conservation of TMs. This integration of data is particularly important to compensate the absence of experimental miRNA-biding interactions in plants. In contrast to the research on animals, PeTMbase and PceRBase are the only two databases devoted to ceRNA research in plants.

As reviewed here, the development of new computational approaches to identify likely TMs is key for the design of functional experiments in plants. Most studies and tools replicate the simplified rules to represent mimic targeting interactions, based on the classical TM model discovered by Franco-Zorrilla *et al.* [5] and commonly modified in computational studies [35, 37, 38]. As a consequence, the few tools available to predict TMs tend to produce results with few common predictions, which in turn represents a major obstacle for experimental validation. Rather than relying merely on MRE conservation and the position of the 3 nt bulge [5, 29], thus, computational modeling and new algorithms are necessary to incorporate alternative modes of miRNA/TM complementarity patterns based on the already known diversity

Downloaded from https://academic.oup.com/bib/article-abstract/19/6/1273/3858378 by UB Leipzig user on 15 January 2020

of canonical and non-canonical binding mechanisms [2, 15, 65, 4, 114, 115]. This is particularly important after Reicher and Millar [66], as well as Liu *et al.* [65] found that the central 3 nt bulge characteristic of most reported TMs is not mandatory for inhibition of miRNAs, but rather poor central complementarity is sufficient. Computational analyses are also necessary to investigate the sequence and interaction context of the MRE-binding sites located within the TMs. For instance, previous studies have reported contrasting findings regarding the preferential location of TM sites within protein-coding regions [36], at the UTRs [37] or similarly distributed along both genome features [35].

Yet, the most critical challenge on ceRNA research is the experimental validation of the TMs identified by *in silico* studies as well as the availability of experimental TM data. For instance, binding interaction data (e.g. RNA:RNA and Protein:RNA) generated with cross-linking immunoprecipitation (CLIP)-based techniques (see Box 1) is extremely rare for plants. Indeed, there is just one CLIP-Seq data set available in GEO NCBI (GSE39051). To compensate this, several TM studies have used degradome data to validate miRNA–target candidates and to test the absence of TM degradation [37]. A major experimental challenge involves the identification of the subcellular structures where ceRNAs are able to approach the concentration of the mRNA targets abundance and of the corresponding miRNA to inhibit miRNA activity and derepress the miRNA targets [66, 116, 117]. Such ceRNA abundances are not observed under real physiological conditions, not even for highly expressed miRNA families in animals [28, 116, 118]. Likewise, Reichel and Millar [66] have recently found that measuring miRNA levels in plants expressing aTMs is not necessarily an absolute indicator of miRNA activity, as observed in the cases of MIM159 [66] and MIM156 [30] in *A. thaliana*. This observation may be explained, in part, because none of the current technological approaches to analyze TM activity guarantees the strongest miRNA inhibition, but rather distinct miRNA families respond differently to them [22]. Accordingly, it has been extensively discussed that the selection of the appropriated technology to carry out the experimental TM validation is key to generate the desired loss-of-function outcome [2, 15, 20, 22]. Clearly, these challenges offer an open opportunity to expand the experimental research of ceRNA in plants.

This review has summarized the major studies behind the emergence and development of ceRNA research in plants after a decade of its discovery. In an attempt to contribute to face some of the challenges in this field, we have also proposed a useful bioinformatics approach for the prediction of consensus TMs motifs that have the potential of cross-targeting either several members within a miRNA family or closely related miRNA families in plants. Future studies should focus on assessing some of these challenges to produce a clear picture about the evolutionary and biotechnological extent of ceRNAs.

BOX 1. Glossary of main terms

Protein-coding sequence
The portion of a gene, composed of exons, coding for protein. The region is bounded close to the 5′ end by a start codon and close to the 3′ end with a stop codon.

Messenger RNA
A RNA harboring genetic information to encode a protein or a functional RNA.

Untranslated region
The terminal nucleotide regions, one on each side of a CDS, on a strand of mRNA.

Noncoding RNA
A RNA molecule with little or no protein-coding capacity.

microRNA
A small ncRNA that inhibits gene expression by guiding the RNA-induced silencing complex (RISC) to target mRNAs. In plants, the length of miRNAs is 21 nts, while it ranges from 18 to 25 nts in animals.

miRNA recognition elements
miRNA-binding sites. In animals, MREs are canonically located at the 3′ UTR of a target mRNA and have sequence similarity to the 5′ seed region of the miRNA. In plants, MREs are located at the 3′ UTR and other regions. Also, MREs are called ‘canonical’ when they exhibit a perfect complementarity with the miRNA; whereas MREs are named as ‘non-canonical’ when they exhibit a less perfect complementary binding owing to the presence of mismatches and bulges.

Target mRNA
A mRNA whose expression is suppressed by miRNA in a specific manner using MREs.

Target mimicry
A phenomenon involving the sequestering of miRNAs by transcripts harboring MREs similar to the corresponding target mRNA, which in turn leads to changes in the expression of the miRNA target gene, rather than its repression by the miRNA.

Competing endogenous RNA
A coding or ncRNA that sequesters miRNA by competing for MREs against target mRNAs. Same as TM or miRNA sponge/decoy.

ceRNA networks
Networks of interactions that represent a collective pool of transcripts (coding, noncoding and aTM) synergically competing for miRNA binding using MREs.

Target mimics
A RNA exhibiting target mimicry.

microRNA sponge/decoy
Another term used for TM. If these molecules are endogenous (but not artificially introduced), then they are equivalent to competitive endogenous RNAs. miRNA sponge is the term most commonly used in animals, while miRNA decoy and TM are broadly used terms.

Long noncoding RNA
Transcripts longer than 200 nts in length and with little or no protein-coding capacity.

Long intergenic noncoding RNA
lncRNA transcripts originated from the DNA segment present between the protein-coding genes.

Pseudogene
A disabled gene copy, identified through DNA homology to its parent gene, with evolutionarily acquired mutations that lead (in principle) to the loss of its original functional capacity. It is originated from either gene duplication (duplicated pseudogene) or reverse transcription (processed pseudogene).

Circular RNAs
A class of transcribed RNA, derived mostly from non-canonical splicing, with its exon ends joined forming a loop.

Short-tandem target mimicry technology
A technology for target mimicry to silence miRNA, where two short sequences mimic small RNA target sites

Downloaded from https://academic.oup.com/bib/article-abstract/19/6/1273/3858378 by UB Leipzig user on 15 January 2020

separated by a linker that lead to degradation of the targeted miRNA [23].

Individual-nucleotide resolution CLIP

A cross-linking immunoprecipitation (CLIP) technique for identifying protein–RNA interactions, in which the direct cross-linking site can be identified where the reverse transcribed cDNA is truncated [15].

Key Points

- TMs appear to play a critical role in the regulation of miRNA activity in animals and plants. However, major computational and experimental challenges are faced, particularly in plants.
- The 'Target Mimic rules' implemented from the classical TM model miR399:IPS1 have been largely used for computational prediction of TMs since its discovery by Franco-Zorrilla and coauthors in 2007. Thus, new algorithms are needed to model the great diversity of known canonical and non-canonical miRNA:TM interactions.
- Comprehensive descriptions and data availability are summarized for 20 computational studies, 2 databases, and 1 tool for the identification of TMs in plants.
- A useful bioinformatics approach for plants is provided to predict TM motifs that have the potential of cross-targeting either several members within a miRNA family or closely related miRNA families, based on the identification of consensus miRNA-binding sites from known TMs.
- From our bioinformatics approach, three consensus TM motifs are provided, MIM166, MIM171 and MIM159/319, to explore experimental validation. The consensus MIM159/319 motif shows already strong experimental support by Reichel and Millar [66].

Supplementary Data

Supplementary data are available online at http://bib.oxfordjournals.org/.

Funding

The National Council of Technological and Scientific Development (CNPq) of Brazil (grant number 00454505/2014-0) and also UTFPR The German Academic Exchange Service (DAAD) of Germany (grant number 57314018) (to A.R.P.). The National Council for Science and Technology in Mexico (CONACyT's support number 185993) and the Max Planck Institute (MPI) for Mathematics in the Sciences (to I.L.C.). A CNPq research fellowship (to D.S.D.).

References

1. Paschoal AR, Maracaja-Coutinho V, Setubal JC, *et al*. Non-coding transcription characterization and annotation: a guide and web resource for non-coding RNA databases. *RNA Biol* 2012;**9**(3):274–82.
2. Steinkraus BR, Toegel M, Fulga TA. Tiny giants of gene regulation: experimental strategies for microRNA functional studies. *Wiley Interdiscip Rev Dev Biol* 2016;**5**(3):311–62.
3. Lozada-Chávez I, Stadler PF, Prohaska SJ. " Hypothesis for the modern RNA world": a pervasive non-coding RNA-based genetic regulation is a prerequisite for the emergence of multicellular complexity. *Orig Life Evol Biosph* 2011;**41**(6):587–607.
4. Ameres SL, Zamore PD. Diversifying microRNA sequence and function. *Nat Rev Mol Cell Biol* 2013;**14**(8):475–88.
5. Franco-Zorrilla JM, Valli A, Todesco M, *et al*. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet* 2007;**39**(8):1033–7.
6. Kang K, Zhong J, Jiang L, *et al*. Identification of microRNA-like RNAs in the filamentous fungus *Trichoderma reesei* by Solexa sequencing. *PLoS One* 2013;**8**(10):e76288.
7. Chen R, Jiang N, Jiang Q, *et al*. Exploring microRNA-like small RNAs in the filamentous fungus *Fusarium oxysporum*. *PLoS One* 2014;**9**(8):e104956.
8. Reinhart BJ, Weinstein EG, Rhoades MW, *et al*. MicroRNAs in plants. *Genes Dev* 2002;**16**(13):1616–26.
9. Jones-Rhoades MW, Bartel DP, Bartel B. MicroRNAs and their regulatory roles in plants. *Annu Rev Plant Biol* 2006;**57**:19–53.
10. Zhang B, Pan X, Cobb GP, *et al*. Plant microRNA: a small regulatory molecule with big impact. *Dev Biol* 2006;**289**(1):3–16.
11. Ha M, Kim VN. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* 2014;**15**(8):509–24.
12. Millar AA, Waterhouse PM. Plant and animal microRNAs: similarities and differences. *Funct Integr Genomics* 2005;**5**(3):129–35.
13. Liu N, Okamura K, Tyler DM, *et al*. The evolution and functional diversification of animal microRNA genes. *Cell Res* 2008;**18**(10):985–96.
14. Lee HC, Li L, Gu W, *et al*. Diverse pathways generate microRNA-like RNAs and dicer-independent small interfering RNAs in fungi. *Mol Cell* 2010;**38**(6):803–14.
15. Thomson DW, Dinger ME. Endogenous microRNA sponges: evidence and controversy. *Nat. Rev. Genet* 2016;**17**(5):272–83. [DOI:10.1038/nrg.2016.20] [PubMed:27040487].
16. Jin D, Wang Y, Zhao Y, *et al*. MicroRNAs and their cross-talks in plant development. *J Genet Genomics* 2013;**40**(4):161–70.
17. Rubio-Somoza I, Weigel D, Franco-Zorilla JM, *et al*. ceRNAs: miRNA target mimic mimics. *Cell* 2011;**147**(7):1431–2.
18. Le T, Zhang D, Liu JL, *et al*. Computational methods for identifying miRNA sponge interactions. *Brief Bioinform* 2016, doi: 10.1093/bib/bbw042.
19. Reichel M, Li J, Millar AA. Silencing the silencer: strategies to inhibit microRNA activity. *Biotechnol Lett* 2011;**33**(7):1285–92.
20. Eamens AL, Wang MB. Alternate approaches to repress endogenous microRNA activity in *Arabidopsis thaliana*. *Plant Signal Behav* 2011;**6**(3):349–59.
21. Zhou M, Luo H. MicroRNA-mediated gene regulation: potential applications for plant genetic engineering. *Plant Mol Biol* 2013;**83**(1–2):59–75.
22. Reichel M, Li Y, Li J, *et al*. Inhibiting plant microRNA activity: molecular sponges, target mimics and sttms all display variable efficacies against target microRNAs. *Plant Biotechnol J* 2015;**13**(7):915–26.
23. Gupta PK. MicroRNAs and target mimics for crop improvement. *Curr Sci India* 2015;**108**:1624–33.
24. Liu X, Hao L, Li D, *et al*. Long non-coding RNAs and their biological roles in plants. *Genomics Proteomics Bioinformatics* 2015;**13**(3):137–47.
25. Kulcheski FR, Christoff AP, Margis R. Circular RNAs are miRNA sponges and can be used as a new class of biomarker. *J Biotechnol* 2016;**238**:42–51.

Downloaded from https://academic.oup.com/bib/article-abstract/19/6/1273/3858378 by UB Leipzig user on 15 January 2020

26. Teotia S, Singh D, Tang X, *et al*. Essential RNA-based technologies and their applications in plant functional genomics. *Trends Biotechnol* 2016;**34**(2):106–23.
27. Bossi L, Figueroa-Bossi N. Competing endogenous RNAs: a target-centric view of small RNA regulation in bacteria. *Nat Rev Microbiol* 2016;**14**(12):775–84.
28. Jens M, Rajewsky N. Competition between target sites of regulators shapes post-transcriptional gene regulation. *Nat Rev Genet* 2015;**16**(2):113–26.
29. Salmena L, Poliseno L, Tay Y, *et al*. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 2011;**146**(3):353–8.
30. Todesco M, Rubio-Somoza I, Paz-Ares J, *et al*. A collection of target mimics for comprehensive analysis of microRNA function in *Arabidopsis thaliana*. *PLoS Genet* 2010;**6**(7):e1001031.
31. Yan J, Gu Y, Jia X, *et al*. Effective small RNA destruction by the expression of a short tandem target mimic in arabidopsis. *Plant Cell* 2012;**24**(2):415–27.
32. Yan F, Guo W, Wu G, *et al*. A virus-based miRNA suppression (VBMS) system for mirna loss-of-function analysis in plants. *Biotechnol J* 2014;**9**(5):702–8.
33. Zhao J, Liu Q, Hu P, *et al*. An efficient potato virus x -based microRNA silencing in *Nicotiana benthamiana*. *Sci Rep* 2016;**6**:20573.
34. Sha A, Zhao J, Yin K, *et al*. Virus-based microRNA silencing in plants. *Plant Physiol* 2014;**164**(1):36–47.
35. Ivashuta S, Banks IR, Wiggins BE, *et al*. Regulation of gene expression in plants through mirna inactivation. *PLoS One* 2011;**6**(6):e21330.
36. Banks IR, Zhang Y, Wiggins BE, *et al*. RNA decoys: an emerging component of plant regulatory networks? *Plant Signal Behav* 2012;**7**(9):1188–93.
37. Meng Y, Shao C, Wang H, *et al*. Target mimics: an embedded layer of microRNA-involved gene regulatory networks in plants. *BMC Genomics* 2012;**13**:197.
38. Wu HJ, Wang ZM, Wang M, *et al*. Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants. *Plant Physiol* 2013;**161**(4):1875–84.
39. Ye CY, Xu H, Shen E, *et al*. Genome-wide identification of non-coding RNAs interacted with microRNAs in soybean. *Front Plant Sci* 2014;**5**:743.
40. Bonnet E, He Y, Billiau K, *et al*. Tapir, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinformatics* 2010;**26**(12):1566–8.
41. Wang P, Zhi H, Zhang Y, *et al*. miRSponge: a manually curated database for experimentally supported miRNA sponges and ceRNAs. *Database* 2015; doi: 10.1093/database/bav098.
42. Yuan C, Meng X, Li X, *et al*. PceRBase: a database of plant competing endogenous RNA. *Nucleic Acids Res* 2017;**45**(D1):D1009–14.
43. Karakulah G, Ycebilgili Kurtoğlu K, Unver T. Petmbase: a database of plant endogenous target mimics (ETMS). *PLoS One* 2016;**11**(12):e0167698.
44. Furio-Tari P, Tarazona S, Gabaldn T, *et al*. Spongescan: a web for detecting microRNA binding elements in lncrna sequences. *Nucleic Acids Res* 2016;**44**(W1):W176–80.
45. Memczak S, Jens M, Elefsinioti A, *et al*. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 2013;**495**(7441):333–8.
46. Sardina D, Alaimo S, Ferro SA, *et al*. A novel computational method for inferring competing endogenous interactions. *Brief Bioinform* 2016; doi: 10.1093/bib/bbw084.
47. Rehmsmeier M, Steffen P, Hochsmann M, *et al*. Fast and effective prediction of microRNA/target duplexes. *RNA* 2004;**10**(10):1507–17.
48. Pearson WR. Flexible sequence similarity searching with the fasta3 program package. *Methods Mol Biol* 2000;**132**:185–219.
49. Dai X, Zhao PX. psrnatarget: a plant small RNA target analysis server. *Nucleic Acids Res* 2011;**39**:W155–9.
50. Fan C, Hao Z, Yan J, *et al*. Genome-wide identification and functional analysis of lincRNAs acting as miRNA targets or decoys in maize. *BMC Genomics* 2015;**16**:793.
51. Shuai P, Liang D, Tang S, *et al*. Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in *Populus trichocarpa*. *J Exp Bot* 2014;**65**(17):4975–83.
52. Zhang YC, Liao JY, Li ZY, *et al*. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol* 2014;**15**(12):512.
53. Lin Y, Lai Z, Tian Q, *et al*. Endogenous target mimics down-regulate mir160 mediation of arf10, -16, and -17 cleavage during somatic embryogenesis in *Dimocarpus longan* lour. *Front Plant Sci* 2015;**6**:956.
54. Li F, Wang W, Zhao N, *et al*. Regulation of nicotine biosynthesis by an endogenous target mimicry of microRNA in tobacco. *Plant Physiol* 2015;**169**(2):1062–71.
55. Wang J, Yu W, Yang Y, *et al*. Corrigendum: genome-wide analysis of tomato long non-coding RNAs and identification as endogenous target mimic for microRNA in response to TYLCV infection. *Sci Rep* 2016;**6**:32828.
56. Lu T, Cui L, Zhou Y, *et al*. Transcriptome-wide investigation of circular RNAs in rice. *RNA* 2015;**21**(12):2076–87.
57. Khemka N, Singh VK, Garg R, *et al*. Genome-wide analysis of long intergenic non-coding RNAs in chickpea and their potential role in flower development. *Sci Rep* 2016;**6**:33297–Sep.
58. Zhang H, Hu W, Hao J, *et al*. Genome-wide identification and functional prediction of novel and fungi-responsive lincRNAs in *Triticum aestivum*. *BMC Genomics* 2016;**17**:238–Mar.
59. Tian J, Song Y, Du Q, *et al*. Population genomic analysis of gibberellin-responsive long non-coding RNAs in populus. *J Exp Bot* 2016;**67**(8):2467–82.
60. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 2014;**42**:68–73.
61. Kruger J, Rehmsmeier M. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res* 2006;**34**:W451–4.
62. Du Z, Zhou X, Ling Y, *et al*. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res* 2010;**38**:W64–70.
63. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 2005;**21**(16):3448–9.
64. Shannon P, Markiel A, Ozier O, *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;**13**(11):2498–504.
65. Liu Q, Wang F, Axtell MJ. Analysis of complementarity requirements for plant microRNA targeting using a *Nicotiana benthamiana* quantitative transient assay. *Plant Cell* 2014;**26**(2):741–53.
66. Reichel M, Millar AA. Specificity of plant microRNA target mimics: cross-targeting of mir159 and mir319. *J Plant Physiol* 2015;**180**:45–8.
67. Chitwood DH, Timmermans MC. Target mimics modulate miRNAs. *Nat Genet* 2007;**39**(8):935–6.

Downloaded from https://academic.oup.com/bib/article-abstract/19/6/1273/3858378 by UB Leipzig user on 15 January 2020

68. Nikovics K, Blein T, Peaucelle A, *et al*. The balance between the mir164a and cuc2 genes controls leaf margin serration in arabidopsis. *Plant Cell* 2006;**18**(11):2929–45.
69. Allen RS, Li J, Stahle MI, *et al*. Genetic analysis reveals functional redundancy and the major target genes of the Arabidopsis mir159 family. *Proc Natl Acad Sci USA* 2007;**104**(41):16371–6.
70. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;**30**(4):772–80.
71. Bailey TL, Boden M, Buske FA, *et al*. Meme suite: tools for motif discovery and searching. *Nucleic Acids Res* 2009;**37**:W202–8.
72. Altschul SF, Gish W, Miller W, *et al*. Basic local alignment search tool. *J Mol Biol* 1990;**215**(3):403–10.
73. Jain M, Chevala VV, Garg R. Genome-wide discovery and differential regulation of conserved and novel microRNAs in chickpea via deep sequencing. *J Exp Bot* 2014;**65**(20):5945–58.
74. Achard P, Herr A, Baulcombe DC, *et al*. Modulation of floral development by a gibberellin-regulated microRNA. *Development* 2004;**131**(14):3357–65.
75. Blazquez MA, Green R, Nilsson O, *et al*. Gibberellins promote flowering of Arabidopsis by activating the leafy promoter. *Plant Cell* 1998;**10**(5):791–800.
76. Palatnik JF, Allen E, Wu X, *et al*. Control of leaf morphogenesis by microRNAs. *Nature* 2003;**425**(6955):257–63.
77. Murray F, Kalla R, Jacobsen J, *et al*. A role for HvGAMYB in anther development. *Plant J* 2003;**33**(3):481–91.
78. Kaneko M, Inukai Y, Ueguchi-Tanaka M, *et al*. Loss-of-function mutations of the rice GAMYB gene impair alpha-amylase expression in aleurone and flower development. *Plant Cell* 2004;**16**(1):33–44.
79. Csukasi F, Donaire L, Casañal A, *et al*. Two strawberry mir159 family members display developmental-specific expression patterns in the fruit receptacle and cooperatively regulate fa-GAMYB. *New Phytologist* 2012;**195**(1):47–57.
80. Woodger FJ, Millar A, Murray F, *et al*. The role of GAMYB transcription factors in GA-regulated gene expression. *J Plant Growth Regul* 2003;**22**(2):176–84.
81. Tsuji H, Aya K, Ueguchi-Tanaka M, *et al*. GAMYB controls different sets of genes and is differentially regulated by microRNA in aleurone cells and anthers. *The Plant Journal* 2006;**47**(3):427–44.
82. Gocal GFW, Sheldon CC, Gubler F, *et al*. GAMYB-like genes, flowering, and gibberellin signaling in *Arabidopsis*. *Plant Physiology* 2001;**127**(4):1682–93.
83. Millar AA, Gubler F. The *Arabidopsis* GAMYB-like genes, myb33 and myb65, are microRNA-regulated genes that redundantly facilitate anther development. *Plant Cell* 2005;**17**(3):705–21.
84. Reyes JL, Chua NH. Aba induction of mir159 controls transcript levels of two MYB factors during *Arabidopsis* seed germination. *Plant J* 2007;**49**(4):592–606.
85. Williams L, Grigg SP, Xie M, *et al*. Regulation of *Arabidopsis* shoot apical meristem and lateral organ formation by microRNA mir166g and its AtHD-ZIP target genes. *Development* 2005;**132**(16):3657–68.
86. Jung JH, Seo PJ, Ahn JH, *et al*. Arabidopsis RNA-binding protein fca regulates microrna172 processing in thermosensory flowering. *J Biol Chem* 2012;**287**(19):16007–16.
87. Kim J, Jung JH, Reyes JL, *et al*. microRNA-directed cleavage of athb15 mRNA regulates vascular development in Arabidopsis inflorescence stems. *Plant J* 2005;**42**(1):84–94.
88. Nogueira FT, Madi S, Chitwood DH, *et al*. Two small regulatory RNAs establish opposing fates of a developmental axis. *Genes Dev* 2007;**21**(7):750–5.
89. Nagasaki H, Itoh J, Hayashi K, *et al*. The small interfering RNA production pathway is required for shoot meristem initiation in rice. *Proc Natl Acad Sci USA* 2007;**104**(37):14867–71.
90. Prigge MJ, Clark SE. Evolution of the class III hd-zip gene family in land plants. *Evol Dev* 2006;**8**(4):350–61.
91. Zhou GK, Kubo M, Zhong R, *et al*. Overexpression of mir165 affects apical meristem formation, organ polarity establishment and vascular development in Arabidopsis. *Plant Cell Physiol* 2007;**48**(3):391–404.
92. Boualem A, Laporte P, Jovanovic M, *et al*. Microrna166 controls root and nodule development in *Medicago truncatula*. *Plant J* 2008;**54**(5):876–87.
93. Subramanian S, Fu Y, Sunkar R, *et al*. Novel and nodulation-regulated microRNAs in soybean roots. *BMC Genomics* 2008;**9**:160.
94. Carlsbecker A, Lee JY, Roberts CJ, *et al*. Cell signalling by microrna165/6 directs gene dose-dependent root cell fate. *Nature* 2010;**465**(7296):316–21.
95. Miyashima S, Koi S, Hashimoto T, *et al*. Non-cell-autonomous microrna165 acts in a dose-dependent manner to regulate multiple differentiation status in the *Arabidopsis* root. *Development* 2011;**138**(11):2303–13.
96. Wang L, Mai YX, Zhang YC, *et al*. Microrna171c-targeted scl6-ii, scl6-iii, and scl6-iv genes regulate shoot branching in arabidopsis. *Mol Plant* 2010;**3**(5):794–806.
97. Lauressergues D, Delaux PM, Formey D, *et al*. The microRNA mir171h modulates arbuscular mycorrhizal colonization of *Medicago truncatula* by targeting nsp2. *Plant J* 2012;**72**(3):512–22.
98. Nag A, King S, Jack T. mir319a targeting of tcp4 is critical for petal growth and development in arabidopsis. *Proc Natl Acad Sci USA* 2009;**106**(52):22534–9.
99. Cubas P, Lauter N, Doebley J, *et al*. The TCP domain: a motif found in proteins regulating plant growth and development. *Plant J* 1999;**18**(2):215–22.
100. Nath U, Crawford BC, Carpenter R, *et al*. Genetic control of surface curvature. *Science* 2003;**299**(5611):1404–7.
101. Mallory AC, Dugas DV, Bartel DP, *et al*. MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs. *Curr Biol* 2004;**14**(12):1035–46.
102. Koyama T, Furutani M, Tasaka M, *et al*. TCP transcription factors control the morphology of shoot lateral organs via negative regulation of the expression of boundary-specific genes in arabidopsis. *Plant Cell* 2007;**19**(2):473–84.
103. Schommer C, Palatnik JF, Aggarwal P, *et al*. Control of jasmonate biosynthesis and senescence by mir319 targets. *PLoS Biol* 2008;**6**(9):e230.
104. Shleizer-Burko S, Burko Y, Ben-Herzel O, *et al*. Dynamic growth program regulated by lanceolate enables flexible leaf patterning. *Development* 2011;**138**(4):695–704.
105. Thiebaut F, Grativol C, Carnavale-Bottino M, *et al*. Computational identification and analysis of novel sugarcane microRNAs. *BMC Genomics* 2012;**13**:290.
106. Ori N, Cohen AR, Etzioni A, *et al*. Regulation of lanceolate by mir319 is required for compound-leaf development in tomato. *Nat Genet* 2007;**39**(6):787–91.
107. Li Y, Li C, Ding G, Jin Y. Evolution of mir159/319 microRNA genes and their post-transcriptional regulatory link to siRNA pathways. *BMC Evol Biol* 2011;**11**:122.
108. Chávez Montes RA, de Ftima Rosas-Cárdenas F, De Paoli E, *et al*. Sample sequencing of vascular plants demonstrates

Downloaded from https://academic.oup.com/bib/article-abstract/19/6/1273/3858378 by UB Leipzig user on 15 January 2020

widespread conservation and divergence of microRNAs. *Nat Commun* 2014;**5**:3722.
109. Axtell MJ, Bartel DP. Antiquity of microRNAs and their targets in land plants. *Plant Cell* 2005;**17**(6):1658–73.
110. Palatnik JF, Wollmann H, Schommer C, *et al*. Sequence and expression differences underlie functional specialization of Arabidopsis microRNAs mir159 and mir319. *Dev Cell* 2007;**13**(1):115–25.
111. Luo Y, Guo Z, Li L. Evolutionary conservation of microRNA regulatory programs in plant flower development. *Dev Biol* 2013;**380**(2):133–44.
112. Baker CC, Sieber P, Wellmer F, Meyerowitz EM. The early extra petals1 mutant uncovers a role for microRNA mir164c in regulating petal number in Arabidopsis. *Curr Biol* 2005;**15**(4):303–15.
113. Laufs P, Peaucelle A, Morin H, Traas J. MicroRNA regulation of the cuc genes is required for boundary size control in Arabidopsis meristems. *Development* 2004;**131**(17):4311–22.
114. Shin C, Nam JW, Farh KK, *et al*. Expanding the microRNA targeting code: functional sites with centered pairing. *Mol Cell* 2010;**38**(6):789–802.
115. Chi SW, Hannon GJ, Darnell RB. An alternative mode of microRNA target recognition. *Nat Struct Mol Biol* 2012;**19**(3):321–7.
116. R, Denzler V, Agarwal J, *et al*. Assessing the cerna hypothesis with quantitative measurements of mirna and target abundance. *Mol Cell* 2014;**54**(5):766–76.
117. Yuan Y, Liu B, Xie P, *et al*. Model-guided quantitative analysis of microRNA-mediated regulation on competing endogenous RNAs using a synthetic gene circuit. *Proc Natl Acad Sci USA* 2015;**112**(10):3158–63.
118. Bosson AD, Zamudio JR, Sharp PA. Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. *Mol Cell* 2014;**56**(3):347–59.

Downloaded from https://academic.oup.com/bib/article-abstract/19/6/1273/3858378 by UB Leipzig user on 15 January 2020