Do miRNAs have a deep evolutionary history?

James E. Tarver[1)2)], Philip C. J. Donoghue[1)] and Kevin J. Peterson[2)]*

The recent discovery of microRNAs (miRNAs) in unicellular eukaryotes, including miRNAs known previously only from animals or plants, implies that miRNAs have a deep evolutionary history among eukaryotes. This contrasts with the prevailing view that miRNAs evolved convergently in animals and plants. We re-evaluate the evidence and find that none of the 73 plant and animal miRNAs described from protists meet the required criteria for miRNA annotation and, by implication, animals and plants did not acquire any of their respective miRNA genes from the crown ancestor of eukaryotes. Furthermore, of the 159 novel miRNAs previously identified among the seven species of unicellular protists examined, only 28 from the algae *Ectocarpus* and *Chlamydomonas*, meet the criteria for miRNA annotation. Therefore, at present only five groups of eukaryotes are known to possess miRNAs, indicating that miRNAs have evolved independently within eukaryotes through exaptation of their shared inherited RNAi machinery.

Keywords:
animal; eukaryote; microRNA evolution; plant; protist

Introduction

microRNAs (miRNAs) are small (21–24 nucleotide) non-coding RNAs that post-transcriptionally regulate gene expression in myriad biological processes in both animals and plants [1–7]. Because of their importance in regulating so many different biological processes their mis-regulation is thought to be a major contributor to a variety of cancers and congenital diseases [8–10]. Furthermore, miRNAs have been implicated as important factors in organismal evolution [11, 12]. Animal and plant miRNAs are thought to have evolved independently [3, 13] given that they exhibit significant differences in their modes of biogenesis and mRNA target recognition [3–7, 14, 15], there are no miRNAs shared by animals and plants [16], and miRNAs have been perceived as absent from their phylogenetic intermediates [17, 18]. Although one study [19] has reported two miRNAs shared between plants and animals, three subsequent large sequencing studies [20–22] failed to recover the 21 nt miRNA that was initially reported, instead recovering a variety of longer fragments from the candidate pre-miR, indicative of a siRNA [23].

The absence of miRNAs in phylogenetic intermediates of animals and plants has recently been challenged by the description of 232 miRNAs from several protist lineages [24–36] (Fig. 1), including novel miRNAs specific to each protist lineage, as well as miRNAs known previously only from animals or plants. The discovery of miRNAs in protists is surprising in itself, but the presence of animal and plant miRNAs among their phylogenetic intermediates suggests that they have been inherited from the last common ancestor of eukaryotes (Fig. 1). By implication, it would appear that miRNAs are part of the ancestral eukaryotic gene regulatory apparatus, and that individual miRNAs are much more evolutionary labile than has been perceived previously [11]. According to this view, the distinctiveness of the miRNA repertoire of animals and plants is largely a consequence of loss, not innovation. For example, the presence of what is currently perceived to be a eutherian-mammal specific miRNA – miR-127 – in the unicellular organism *Giardia*, implies its loss in at least 21 different eukaryotic lineages ranging from plants to cnidarians to marsupial mammals (Fig. 1). Hence, these discoveries challenge not only our understanding of the evolution of plant and animal miRNAs [17, 18, 37–43], but also the utility of miRNAs as

DOI 10.1002/bies.201200055

[1)] School of Earth Sciences, University of Bristol, Bristol, UK
[2)] Department of Biological Sciences, Dartmouth College, Hanover, NH, USA

***Corresponding author:**
James E. Tarver
E-mail: james.tarver@bristol.ac.uk

Supporting information online

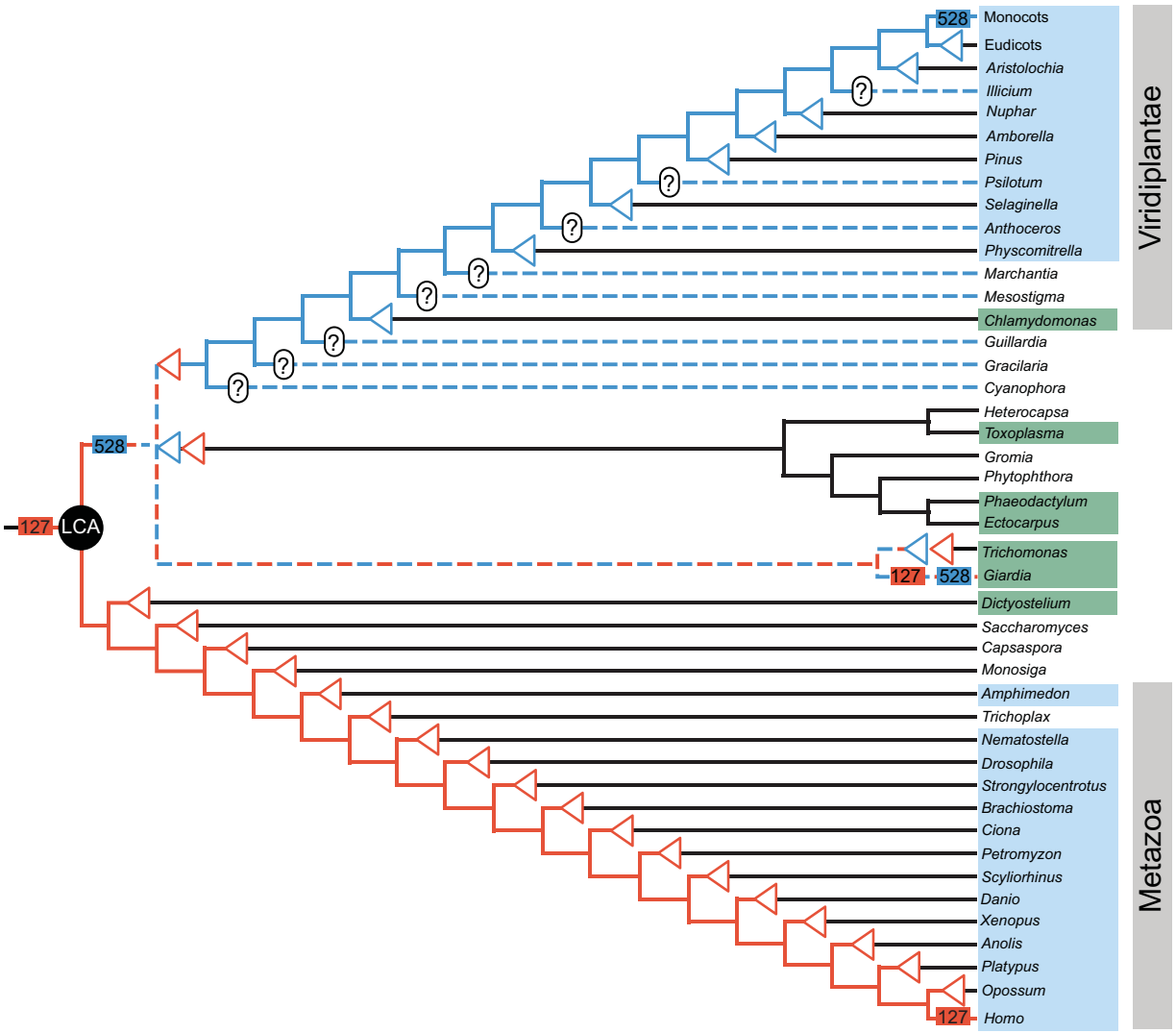

Figure 1. Distribution of miRNAs across the eukaryote tree of life. miRNAs are conventionally believed to have evolved twice, once in the animal and once in the plant lineages (blue boxes). Recent publications suggest that miRNAs also exist in many protist taxa (green boxes). This would imply that the last common ancestor (LCA) of crown eukaryotes possessed a complement of miRNAs. Moreover the presence of miR-127 and MIR-528 (previously only known in eutherians or monocots, respectively) within *Giardia* implies that either these two miRNAs evolved deeply within eukaryotes and were subsequently lost in all intervening lineages (triangles), or that extensive horizontal gene transfer has occurred. Those lineages from which small RNA libraries have not been sequenced are indicated with a ''?'' as the putative losses have not been experimentally confirmed.

phylogenetic markers [12, 44, 45], while questioning the view that miRNAs have an influential role in the evolution of organismal complexity [46–50].

Given the broad ranging and extraordinary implications of these previous results, we re-examined each of these putative protistan miRNAs to determine (1) whether these data indicate that miRNAs have a much deeper eukaryotic evolutionary history than has been perceived hitherto, (2) whether some animal and plant miRNAs have been inherited from an ancestral eukaryote, (3) if their distribution reflects horizontal transfer, or (4) whether these genes have merely been identified incorrectly among protists.

The processes of miRNA biogenesis

Identifying miRNAs in protists is problematic given the profound lack of knowledge on the RNAi machinery present within these groups. However, because miRNAs are defined by their mode of biogenesis, we begin by considering similarities and differences in the genesis of animal and plant miRNAs (Fig. 2), as a guide to annotating protist miRNAs. In animals, miRNA biogenesis begins with the transcription of DNA by either RNA polymerase II or RNA polymerase III [51], producing so-called pri-miRNAs (the ''pri'' signifying the ''primary'' transcript). The pri-miRNA is then processed within the nucleus by the RNase III enzyme Drosha and a double-stranded RNA binding protein (known as Pasha in flies and DGCR8

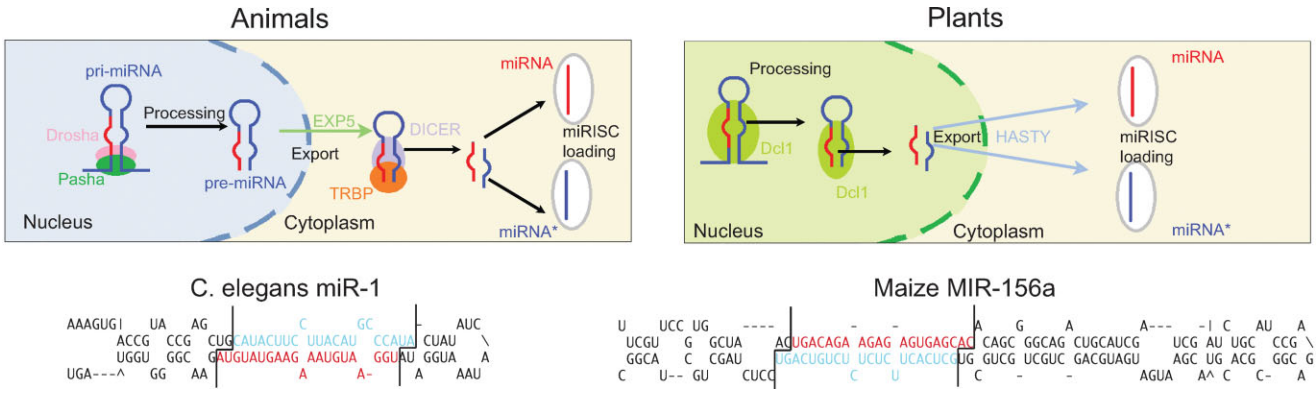

Figure 2. Biogenesis of miRNAs in animals and plants. At the top are the canonical pathways for animals (left) and plants (right). Note that the processing of animal miRNAs takes place in both the nucleus and the cytoplasm, whereas in plants the processing takes place entirely in the nucleus. Below are two representative miRNAs, miR-1 from *C. elegans* and MIR-156a from Maize. Both form the characteristic hairpin loop with no large internal bulges and sufficient complementary base-pairing between the miRNA and opposing arm. Note also the 2 nt overhang on the 3′ arm of both miRs, a hallmark of RNase III processing.

in humans) to form the characteristic small hairpin known as the pre-miRNA (the "pre" signifying the "precursor" transcript) [52, 53]. This pre-miRNA is then exported from the nucleus by Exportin-5-Ran-GTP (EXP5), a process that is independent of either the pre-miRNA loop structure or sequence identity, but is dependent upon the presence of the 3′ overhang and a double-stranded segment of RNA at least 16 nucleotides long [54, 55]. DICER, another RNAase III enzyme, recognizes the terminal loop region of the hairpin and cleaves the loop at a fixed distance leaving a 21–24 nucleotide RNA duplex now with two nucleotide overhangs at both 3′ ends (Fig. 2) [56]. Subsequently, the duplex separates into two separate strands – the guide stand (miRNA) and the passenger strand (miRNA*). The guide strand is loaded into an Argonaute (AGO) protein, which then regulates target mRNAs, while the miRNA* is usually degraded [1], although the miRNA* can also regulate target mRNAs [57]. Thus, in animals, miRNA biogenesis occurs in both the nucleus and in the cytoplasm, using two different RNase III enzyme complexes.

miRNA biogenesis in plants is similar to that of animals (Fig. 2). Nonetheless, many of the core pathways and processes are distinct and are less well understood [43]. A key difference between animal and plant biogenesis is that in plants miRNA biogenesis occurs entirely in the nucleus, where the gene DICER-LIKE1 (DCL1) is used both for pri- and pre-miRNA processing [37, 58, 59]. The resulting duplex is similar to the product of biogenesis in animals, with characteristic 3′ overhangs and well-defined 5′ ends [60] (Fig. 2). The processed miRNA duplex is exported from the nucleus to the cytoplasm by the Exportin-5 homologue HASTY and loaded onto an AGO protein to regulate target mRNAs [61]. A second key difference concerns the length of the pre structure. In most animals the pre structure is usually around 70–80 nucleotides long, while in plants and demosponges, the pre structures can be much longer, sometimes several hundred nucleotides long [3, 43]. Thus, both systems generate a structure with the mature gene product offset by two nucleotides from the miRNA*, but can differ in the relative distance this paired gene product is from the loop.

The criteria for miRNA annotation

The accurate identification and annotation of candidate miRNAs from deep-sequencing data is reliant on the use of established criteria [3, 16, 62, 63] summarized in Table 1, which differs slightly between plants and animals due to differences in their biogenesis (Fig. 2). In both plants and animals, the miRNA and miRNA* should be sequenced multiple times with homogeneity at the 5′ end, and a two nucleotide overhang on the 3′ end. Star sequences for all miRNAs may not be sequenced in a single experiment, but if the sequence is clearly similar to a known miRNA with an annotated star, the evolutionary conservation of the locus is considered sufficient even in the absence of a sequenced star sequence [52]. The 5′-end homogeneity is an important criterion [3, 16] as it clearly distinguishes fragments of longer RNA molecules, like rRNA from novel miRNAs. For example, Figure 3 shows three putative miRNAs derived from the guinea pig *Cavia porcellus* using the algorithm miRMiner [17]. The first is let-7c, which shows clear miRNA processing with 98.84% of the reads starting with the same nucleotide. Further, most of the star sequences also start with the same nucleotide. Importantly, there is no sequence overlap between these two sets of sequences. The middle panel, a putative novel miRNA from *Cavia*, shows a similar distribution of reads with 100% of the mature reads starting on the same nucleotide. In contrast the lower panel is a fragment of 28S rRNA. Based solely on structural considerations, this set of "mature" reads would be classified as a miRNA. However, there is no clear 5′ processing of the putative mature gene product, <50% of reads coming from the same start position, and there is continuous overlap among the reads.

Not only do eumetazoan (cnidarian + bilaterian) miRNAs differ in length from plant miRNAs, but they also differ in the anticipated extent of complementary

Table 1. Criteria for the identification of miRNAs following previous workers [3, 16, 56, 57]

Criteria that must be met for the accurate annotation of miRNAs.		Examples
Evidence of expression from small RNA libraries for both the candidate miRNA and miRNA*		Figure 3
The candidates cannot match an already identified fragment of the genome (e.g. rRNA, tRNA, snoRNA etc)		Figure 4G
The candidates cannot match to many different places throughout the genome		
The candidates must match exactly to an extended part of the genome to form the pre-miR		Figure 3
There must be consistent 5′ processing of both the candidate miRNA and miRNA*		Figure 3
The candidate miRNA and miRNA* should form a 2 nt overhang on the 3′ arm of the pre-miR		Figure 4A
Additional taxon specific criteria		
Animals	Must form a hairpin with 16 bp of the first 22 nt being complementary to the opposing arm with no large secondary structures	Figure 2
Plants	≤4 Mismatches of complementary base-pairing between miRNA and opposing arm	Figure 2

None of the above criteria are sufficient in their own right to identify a novel miRNA and all need to be observed. Examples of each are shown in other figures.

base pairing between the guide and star strand. In eumetazoan animals, 16 of the first 22 nucleotides of the identified mature sequence should exhibit complementary base pairing with the star sequence [52]. In plants, complementary base pairing between the mature and star sequences is more precise, fewer than four mismatches and any asymmetric bulges involving only one or two nucleotides [63]. In addition the increased length of plant miRNAs means that some secondary structures may form within the hairpin, something that does not occur in eumetazoan miRNAs. Finally, Ma et al. [64] suggested that in plants, miRNAs could be distinguished from siRNAs on the polarity of miRNA accumulation, a minimum requirement of 75% of reads coming from one arm. However, some animals use both the miRNA and miRNA*. For example, deep-sequencing data [65, 66] has confirmed a more even division of reads between both arms of miR-276 (66% vs. 34%) in *Drosophila* [57]. This criterion cannot, therefore, be applied more generally.

No specific criteria have been established for the annotation of protist miRNAs – instead researchers have largely based their identification of miRNAs on similarities to either plant or animal miRNAs. We apply the existing criteria for plant miRNAs where putative plant homologues have been identified in protists, and the less-stringent animal miRNA criteria were used in all other instances. More generally, if the candidate miRNA has a long hairpin with some secondary structure similar to plants, it must meet the complementary base-pairing requirements for plant miRNAs, rather than the weaker requirement for annotation of animal miRNAs. If it meets the weaker criteria for complementary base-pairing in animal miRNAs, it must then meet the stronger criteria regarding secondary structure.

Reanalysis of previously identified protist miRNAs

To assess the veracity of the novel and homologous candidate miRNAs previously described in protists [24–36], the miRNA and pre-miRNA sequences were recovered from the original publications and assessed according to established criteria [3, 16, 62, 63] (Table 1). For miRNAs previously identified in other organisms, the sequences were BLAST searched against both the organism's respective genome and miRBase [16] and folded using the default settings in mfold 3.2 [67]. Novel miRNAs were BLAST searched against GenBank to discriminate short fragments of longer degraded RNAs. One of the studies [24] provided neither pre-miRNA nor miRNA sequences; instead they provide a list of sequences annotated within miRBase to which their candidate miRNAs matched exactly (both in identity and length) and so we downloaded and used these sequences in this reanalysis. Of the nine protist taxa in which miRNAs have been identified previously, genome sequences are available for seven (*Ectocarpus siliculosus*, *Phaeodactylum tricornutum*, *Chlamydomonas reinhardtii*, *Dictyostelium discoideum*, *Toxoplasma gondii*, *Giardia lamblia* and *Trichomonas vaginalis*); genome sequences are not available for either *Trichomonas foetus* or *Pentatrichomonas hominis* and so the identification of any small RNA reads from a library as a miRNA cannot be substantiated. Thus, our analyses are confined to the seven taxa where the pre structure(s) can be evaluated properly, representing 232 putative protistan miRNAs, all of which are listed and described in the Supplementary Material.

Ectocarpus siliculosus

Cock et al. [33] described 26 miRNA genes from the brown alga *E. siliculosus*. Upon reanalysis 22 of the 26 miRNA genes appear to be robust, as they have both mature and star sequences and the requisite base-pairing rules with plant-like miRNA pre-structures (Fig. 4A; Supplementary Table 1). Two miRNAs – miR3463, miR3454d – do not show the requisite pre structure for a bona fide hairpin and, thus, are rejected. miR3460 and miR3465 show perfect complementary base-pairing (Fig. 4B), a characteristic of siRNAs (whereas miRNAs have a few mismatches) [68] and they should be considered as siRNAs unless additional supporting evidence is provided.

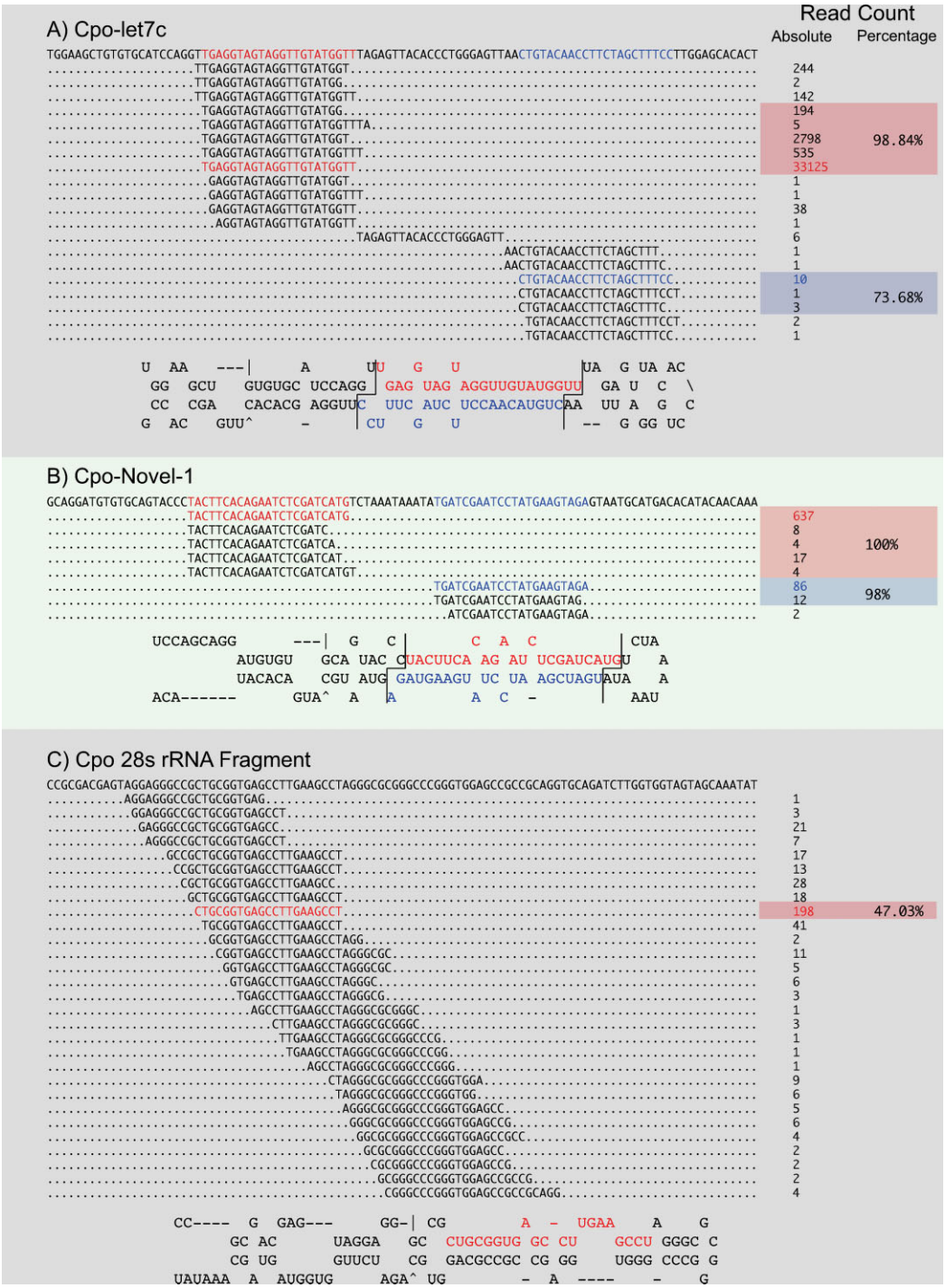

Figure 3. Read data from a small RNA library from the Guinea Pig (*Cavia porcellus*). RNA was extracted from a single individual of *C. porcellus* as described in Wheeler et al. [17]. Small RNA libraries were sequenced on the Illumina Genome Analyzer II platform at the University of Bristol Transcriptomics Facility. The number of reads sequenced was 13,495,328 and these were analysed using miRMiner [17] to discover previously identified miRNAs and novel miRNA genes specific to *C. porcellus*. These data show the homogeneity or lack thereof on the 5′ end of three candidate miRNAs in association with their structures. **A**: Read data for let-7c. Note that the read counts are dominated (>98%) by a specific 5′ end, and there is no overlap amongst the mature, loop and star reads (nor the 5′ and 3′ overlap reads [92], not shown). **B**: A putative novel miRNA with both mature and star regions sequenced and clear 5′ homogeneity at the 5′ end with 100% of the miRNA reads starting at the same position. **C**: In contrast this fragment of 28S rRNA, although passing the structural filters in miRMiner for miRNA annotation, does not express the appropriate star sequence, and shows 5′ end heterogeneity with the 21 bp fragment putative mature sequence accounting for less than <50% of reads. Note that due to space limitations, additional reads at both the 5′ and 3′ ends were not shown for this fragment of 28S rRNA, and if these were included the read count for the candidate miRNA drops to only 30.09% of the total reads across the putative 100mer.

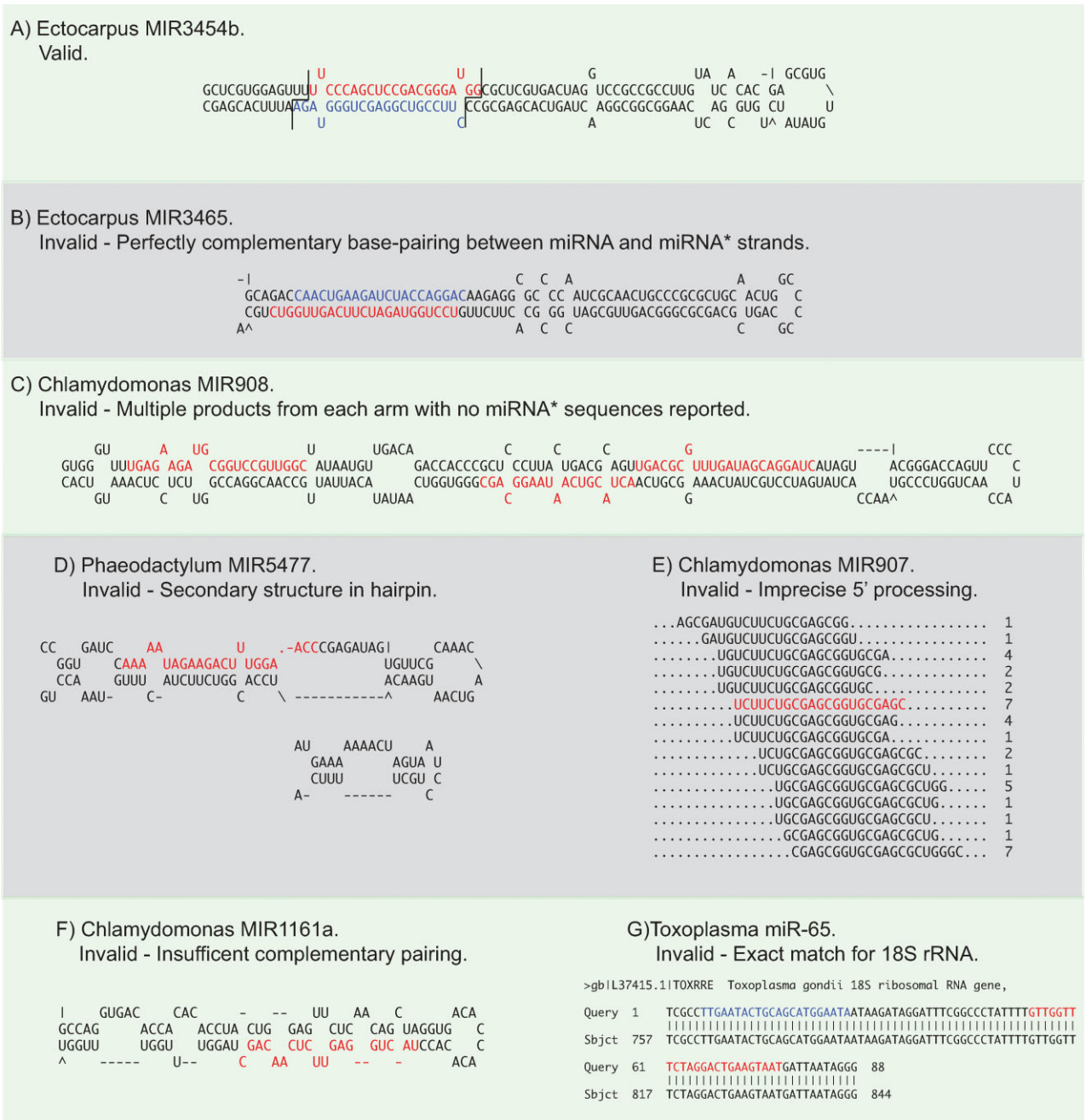

Figure 4. Examples of both real and misidentified miRNAs in a variety of protists, highlighting common reasons for their rejection in this study. **A**: A genuine miRNA from *Ectocarpus*. Note the 2nt overhang on the 3′ arm for the miRNA and miRNA* sequences. **B** and **C**: Two previously annotated miRNAs, which have characteristics such as perfectly complementary base-pairing (B) and multiple candidate miRNAs from the same arm without corresponding miRNA* sequences (C), which are characteristic of siRNAs. **D**: Secondary structure in the hairpin that overlaps with the miRNA, and so was rejected, while **E**: was rejected due to imprecise 5′ processing. **F**: Had only 14 bp complementary between the miRNA and opposing arm rather than the required minimum 16 bp, which is necessary for nuclear export by Exp-5 [55]. **G**: *Toxoplasma* miR-65, which was not only incorrectly named as it is not orthologous to miR-65, but also matched exactly to a fragment of 18S rRNA.

Chlamydomonas reinhardtii

Zhao et al. [36] and Molnar et al. [35] reported the presence of miRNAs from the unicellular chlorophyte green alga *C. reinhardtii*. To date, 50 lineage-specific miRNA sequences have been deposited in miRBase, but our reevaluation of these miRNAs suggests that only six are robust. The remainder either: (1) show multiple products from the same arm without corresponding star sequences and are thus more likely to consist of siRNA fragments rather than miRNA gene products (e.g. miR908; Fig. 4C); (2) have no reported star sequences (e.g. miR918); (3) do not show evidence of stable 5′

processing (e.g. miR907; Fig. 4E); or (4) do not show robust complementarity between the candidate mature and star sequences (e.g. miR1161a; Fig. 4F). This approach was corroborated by Nozawa et al. [69] who, in a recent reanalysis of plant miRNA genes, suggested that only five of the 50 currently recognized miRNAs in *Chlamydomonas* are genuine.

Phaeodactylum tricornutum

Huang et al. [32] reported the occurrence of 13 novel miRNAs in the diatom *P. tricornutum*, none of which withstand scrutiny, as 11 of the 13 putative miRNAs show extensive secondary structure in the pre-miR (Fig. 4D). The remaining two lack star sequences (as did all candidates) and, thus, it is not possible to confirm their identification as miRNA genes without additional sequencing.

Dictyostelium discoideum

Two novel miRNAs were described from the amoebozoan *D. discoideum* by Hinas et al. [34], but star sequences were not reported, making it impossible to evaluate their status as bona fide miRNAs. Although Hinas et al. [34] show that the biogenesis of these candidate miRNAs relies on the presence of DICER, this on its own is not sufficient, as DICER interacts with many different kinds of RNA molecules, including miRNAs. Thus, the presence of miRNAs in this taxon remains an open question.

Toxoplasma gondii

Braun et al. [30] reported 34 novel miRNA sequences from the alveolate *T. gondii*, again none of which withstand scrutiny. Over half of their candidate miRNAs (19) cannot be localized in the genome, and so cannot be considered genuine. An additional four candidates are fragments of other kinds of RNAs including mRNAs, rRNAs, and snoRNAs (Fig. 4G). Another four candidates do not show proper pre-miRNA structures, while the remaining seven candidates do not have star sequences. Thus, standard analyses cannot verify the presence of miRNAs within *Toxoplasma*.

Giardia lamblia and *Trichomonas vaginalis*

One hundred and seven miRNAs have been described from the excavate protozoans *G. lamblia* and *T. vaginalis*, including 15 miRNAs previously described from plants and 58 miRNAs described from animals [24, 25, 31]. Of the 73 candidate miRNAs previously identified in other taxa, 60 cannot be identified within the organisms' genomes. All of the remaining 13 fail to form the appropriate hairpin structure (e.g. Gim3 a purported homologue of gga-miR-202) or map to many places within the genome (e.g. Tvm2, a purported homologue of gma-MIR-1534, which maps to 836 different places within the genome). Sequences that map to many different places in a genome should be discarded and not annotated as miRNAs [16]. Notably none of the candidate miRNAs that map to multiple loci show evidence of canonical miRNA processing, such as reads supporting both strands of a miRNA duplex with a two-nucleotide overhang.

Of the 34 putative novel miRNAs, 23 do not form the requisite hairpin loop with sufficient complementary base-pairing, four map to numerous places within the genome, two have inconsistent 5′ processing and two cannot be located within the genomes of the respective taxa. The remaining three candidate miRNAs all lacked the requisite star sequences. Thus, there are no miRNAs currently known in these taxa.

Discussion

Of the 232 candidate miRNAs described from protists whose veracity could be assessed, including novels and homologues of plant and animal miRNAs, the vast majority fail to meet the community's established criteria for miRNA annotation (Table 2; Supplementary Table 1). Some of these studies have suggested a role in gene regulation for these putative miRNAs in *Giardia* [27–29]. However, none shows the requisite hairpin structure for miRNA annotation, which is not unexpected given that they are derived from snoRNAs [27, 29]. Nonetheless, this does not mean that these snoRNA fragments are not used in gene regulation in this taxon – but functional similarity does not constitute homology. Rather, if correctly interpreted this simply represents an exaptation of a snoRNA into a new role in gene regulation.

The indispensable miRNA* sequence

A large group of candidate miRNAs were classified as needing additional supporting evidence as they lack expression data of the miRNA* sequence. This could be rectified with additional sequencing. Indeed, as most studies investigating miRNAs are now using deep-sequencing technology, the presence of a miRNA* for the annotation of novel miRNAs should now be considered a requirement rather than a preference [3]. This is because, given the nature of deep-sequencing, millions of short sections of RNA will be sequenced, many of which will correspond to degraded fragments of longer RNAs and not miRNAs. Furthermore, given the large number of predicted hairpins (10^5–10^6) [70–72] that exist within eukaryotic genomes, many such fragments will correspond to these hairpins [3], not to miRNA loci. Such artifacts have hindered previous attempts to catalogue the miRNA repertoire of a variety of animals, with many of the miRNAs originally identified in *Drosophila* [73] and mice [74] having been reinterpreted subsequently [75, 76]. Thus the absence of star sequences for all these putative miRNAs may in itself be telling: even with deeper sequencing these may not be recovered in the future.

What's in a name?

Some of the studies [27–30] describing miRNAs in protists did not follow the established conventions for naming miRNAs [16, 62, 63]. Consequently, authors annotated novel miRNAs using names of already existing animal and plant miRNAs, implying homology, which can readily be misinterpreted [77, 78]. For example, Saraiya et al. [28] described miR-2 regulating the expression of 22 variant surface protein genes in *Giardia*. This would be a remarkable discovery providing fundamental insight into the evolution of

Table 2. Distribution of the 232 candidate miRNAs from the 13 studies, using established miRNA criteria.

Unable to find an exact match for the sequence in the genome	Pre-miRNA hairpin does not form correctly	miRNA matches to many places within the genome	Inconsistent 5′ processing	Multiple distinct products from the same arm	Exact BP matching between the miRNA and miRNA*	Fragment of rRNA, tRNA, mRNA, snoRNA	Further evidence needed to annotate as a miRNA/miRNA* duplex	miRNA miRNA*
83	52	8	6	14	8	4	29	28

miRNAs, given that miR-2 has otherwise been considered exclusive to protostome animals [17]. However, none of the candidate miRNAs identified in these studies [27–30] match any known sequence within miRBase, let alone those given the same name. Consequently, we reinterpreted all of these apparently homologous miRNAs as novel miRNAs unique to the taxon in question.

The multiple origins of miRNAs

Of all the putative miRNAs analyses herein, only 28 novel miRNAs from the algae *Ectocarpus* and *Chlamydomonas* passed the basic criteria for miRNA annotation. Curiously though none of the six miRNAs identified in *Chlamydomonas* are shared with any other taxon including land plants; likewise the 22 novel miRNAs identified in *Ectocarpus* are currently known only from this one species. This is similar to the situation in demosponges where none of the eight miRNAs identified thus far are found outside of silicisponges [79], and to the ever-expanding repertoire of eumetazoan and plant miRNAs that are taxon-specific. No miRNAs are shared between these five principal lineages, and so it is both the distinct process of biogenesis, and the pattern of gene regulation, that define miRNAs, rather than a particular nucleotide sequence. Thus, in order to identify homology it is necessary to examine the underlying processing machinery among the five lineages. Direct comparison between animals and plants suggests that the multi-domain protein DICER is the only conserved element in miRNA biogenesis, and that multiple DICERs exist within these separate linages [80–84]. The domain structure of plant, fungal and animal DICERs is very similar, which is indicative of homology across this vast phylogenetic expanse [82, 83] and, indeed, DICER has been reported from several different protist taxa [30, 85–87], albeit sometimes reduced in terms of its domain structure.

Because DICER is also integral to the biogenesis of small interfering RNAs (siRNAs), which are also present within unicellular protists, the presence of DICER cannot, on its own, constitute prima facie evidence for the existence of miRNAs. Indeed, given that both siRNAs and miRNAs undergo a very similar biogenetic process, siRNAs can be easily misidentified as miRNAs [68]. We rejected putative miRNAs in both *Ectocarpus* and *Chlamydomonas* because they exhibit many characteristics of siRNAs, including identical complementary base-pairing between both the mature and star products, and multiple distinct products from the same arm. One of the strongest criteria for the identification of miRNAs is precise 5′ editing, with greater heterogeneity at the 5′ end associated with siRNAs [68]. These data are lacking for *Ectocarpus*, while there are usually less than five reads per miRNA in the available data for *Chlamydomonas*, precluding the ability to assess the precision of the 5′ cut for most candidate miRNAs. Nevertheless, the clearest distinction between miRNAs and siRNAs is that miRNAs do not usually regulate their own expression, while both viral and endo siRNAs silence the same loci from which they are derived [68]. Unfortunately, no such experiments have been conducted on these algal miRNAs, raising the possibility that the currently recognized miRNAs in *Chlamydomonas* and *Ectocarpus* are also siRNAs.

Future directions

The continued sequencing of many small RNA libraries from protists will allow researchers to answer many of the questions that we raise. Currently only two clades, Eumetazoa and Plantae, have had their miRNA repertoires studied extensively. We recognize miRNAs within both *Chlamydomonas* and *Ectocarpus*, but how representative are these of green and brown algae more generally, and do other members of these clades share a subset of these miRNAs? And finally, do miRNAs in these clades function in the manner that we expect, given our knowledge of miRNA function in Eumetazoa and Plantae?

Although we show that there are currently no known miRNAs in other protists, absence of evidence does not constitute evidence of their absence. We anticipate that phylogenetically targeted sampling of these woefully understudied eukaryotic lineages is likely to reveal a broader diversity of miRNA synthesizing organisms. Only then will it be possible to determine whether miRNA synthesis is a shared primitive feature of eukaryotes or whether miRNA synthesis has evolved convergently among several different eukaryote lineages.

Conclusions

We set out to test an intriguing and revolutionary hypothesis, that miRNAs evolved once in the last common ancestor (LCA) of crown eukaryotes and that this ancestor shared miRNAs present today in plants and animals. Although this hypothesis is becoming widely accepted [77, 78, 88–91] our data clearly

refute it. Instead, our results show that of the 73 plant and animal miRNAs identified in protists, all fail to meet the criteria required for the identification of miRNAs. Of the 159 novel miRNAs identified, only 28 pass the current criteria for miRNA annotation, all of which belong to the green and brown algae. Therefore, to date only five eukaryotic clades are known to possess miRNAs, but in each case each clade possesses its own unique repertoire, animals possessing two non-overlapping complements, demosponge-specific and eumetazoan-specific miRNAs. Thus, to date no homologous miRNAs have been found that are shared between plants, animals and protists: each clade possesses its own unique complement of miRNAs, which indicates yet another example of molecular exaptation.

Acknowledgments

We would like to thank the editor Andrew Moore and two anonymous reviewers for thoughtful comments on an earlier draft of this manuscript. We acknowledge funding from the Marie Curie actions of EU FP7, the National Aeronautic and Space Agency, the National Science Foundation, the Natural Environmental Research Council, and the University of Bristol.

References

1. **Bartel D.** 2009. MicroRNAs: target recognition and regulatory functions. *Cell* **136**: 215–33.
2. **Berezikov E.** 2011. Evolution of microRNA diversity and regulation in animals. *Nat Rev Genet* **12**: 846–60.
3. **Axtell MJ**, **Westholm JO**, **Lai EC.** 2011. Vive la difference: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol* **12**: 221.
4. **Llave C**, **Xie Z**, **Kasschau K**, **Carrington J.** 2002. Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* **297**: 2053–6.
5. **Rhoades M**, **Reinhart B**, **Lim L**, **Burge C**, et al. 2002. Prediction of plant microRNA targets. *Cell* **110**: 513–20.
6. **Lee R**, **Feinbaum R**, **Ambros V.** 1993. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**: 843–54.
7. **Wightman B**, **Ha I**, **Ruvkun G.** 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* **75**: 855–62.
8. **Lee YS**, **Dutta A.** 2009. MicroRNAs in cancer. *Annu Rev Pathol Mech* **4**: 199–227.
9. **Lu J**, **Getz G**, **Miska EA**, **Alvarez-Saavedra E**, et al. 2005. MicroRNA expression profiles classify human cancers. *Nature* **435**: 834–8.
10. **Lu M**, **Zhang Q**, **Deng M**, **Miao J**, et al. 2008. An analysis of human microRNA and disease associations. *PLoS ONE* **3**: e3420.
11. **Peterson KJ**, **Dietrich MR**, **McPeek MA.** 2009. MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. *BioEssays* **31**: 736–47.
12. **Heimberg AM**, **Cowper-Sallari R**, **Semon M**, **Donoghue PCJ**, et al. 2010. microRNAs reveal the interrelationships of hagfish, lampreys, and gnathostomes and the nature of the ancestral vertebrate. *Proc Natl Acad Sci USA* **107**: 19379–83.
13. **Jones-Rhoades MW**, **Bartel DP**, **Bartel B.** 2006. MicroRNAs and their regulatory roles in plants. *Annu Rev Plant Biol* **57**: 19–53.
14. **Brennecke J**, **Stark A**, **Russell R**, **Cohen S.** 2005. Principles of microRNA-target recognition. *PLoS Biol* **3**: e85.
15. **Lewis B**, **Shih I**, **Jones-Rhoades M**, **Bartel D**, et al. 2003. Prediction of mammalian microRNA targets. *Cell* **115**: 787–98.
16. **Kozomara A**, **Griffiths-Jones S.** 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* **39**: D152–7.
17. **Wheeler BM**, **Heimberg AM**, **Moy VN**, **Sperling EA**, et al. 2009. The deep evolution of metazoan microRNAs. *Evol Dev* **11**: 50–68.
18. **Grimson A**, **Srivastava M**, **Fahey B**, **Woodcroft BJ**, et al. 2008. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455**: 1193–7.
19. **Arteaga-Vazquez M**, **Caballero-Perez J**, **Vielle-Calzada J-P.** 2006. A family of microRNAs present in plants and animals. *Plant Cell* **18**: 3355–69.
20. **Fahlgren N**, **Howell M**, **Kasschau K**, **Chapman E**, et al. 2007. High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of MIRNA genes. *PLoS ONE* **2**: e219.
21. **Mi S**, **Cai T**, **Hu Y**, **Chen Y**, et al. 2008. Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5′ terminal nucleotide. *Cell* **133**: 116–27.
22. **Rajagopalan R**, **Vaucheret H**, **Trejo J**, **Bartel D.** 2006. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev* **20**: 3407–25.
23. **Axtell MJ**, **Bowman JL.** 2008. Evolution of plant microRNAs and their targets. *Trends Plant Sci* **13**: 343–9.
24. **Huang P-J**, **Lin W-C**, **Chen S-C**, **Lin Y-H**, et al. 2012. Identification of putative miRNAs from the deep-branching unicellular flagellates. *Genomics* **99**: 101–7.
25. **Lin W-C**, **Huang K-Y**, **Chen S-C**, **Huang T-Y**, et al. 2009. Malate dehydrogenase is negatively regulated by miR-1 in *Trichomonas vaginalis*. *Parasitol Res* **105**: 1683–9.
26. **Lin W-C**, **Li S-C**, **Lin W-C**, **Shin J-W**, et al. 2009. Identification of microRNA in the protist *Trichomonas vaginalis*. *Genomics* **93**: 487–93.
27. **Li W**, **Saraiya AA**, **Wang CC.** 2011. Gene regulation in *Giardia lambia* involves a putative microRNA derived from a small nucleolar RNA. *PLoS Negl Trop Dis* **5**: e1338.
28. **Saraiya AA**, **Li W**, **Wang CC.** 2011. A microRNA derived from an apparent canonical biogenesis pathway regulates variant surface protein gene expression in *Giardia lamblia*. *RNA* **17**: 2152–64.
29. **Saraiya AA**, **Wang CC.** 2008. snoRNA, a novel precursor of microRNA in *Giardia lamblia*. *PLoS Pathog* **4**: e1000224.
30. **Braun L**, **Cannella D**, **Ortet P**, **Barakat M**, et al. 2010. A complex small RNA repertoire is generated by a plant/fungal-like machinery and effected by a metazoan-like argonaute in the single-cell human parasite *Toxoplasma gondii*. *PLoS Pathog* **6**: e1000920.
31. **Chen X**, **Collins LJ**, **Biggs PJ**, **Penny D.** 2009. High throughput genome-wide survey of small RNAs from the parasitic protists *Giardia intestinalis* and *Trichomonas vaginalis*. *Genome Biol Evol* **1**: 165–75.
32. **Huang A**, **He L**, **Wang G.** 2011. Identification and characterization of microRNAs from *Phaeodactylum tricornutum* by high-throughput sequencing and bioinformatics analysis. *BMC Genomics* **12**: 337.
33. **Cock JM**, **Sterck L**, **Rouze P**, **Scornet D**, et al. 2010. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* **465**: 617–21.
34. **Hinas A**, **Reimegard J**, **Wagner EGH**, **Nellen W**, et al. 2007. The small RNA repertoire of *Dictyostelium discoideum* and its regulation by components of the RNAi pathway. *Nucleic Acids Res* **35**: 6714–26.
35. **Molnar A**, **Schwach F**, **Studholme DJ**, **Thuenemann EC**, et al. 2007. miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. *Nature* **447**: 1126–9.
36. **Zhao T**, **Li G**, **Mi S**, **Li S**, et al. 2007. A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. *Genes Dev* **21**: 1190–203.
37. **Park W**, **Li J**, **Song R**, **Messing J**, et al. 2002. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr Biol* **12**: 1484–95.
38. **Reinhart BJ**, **Weinstein EG**, **Rhoades MW**, **Bartel B**, et al. 2002. MicroRNAs in plants. *Genes Dev* **16**: 1616–26.
39. **Lagos-Quintana M**, **Rauhut R**, **Lendeckel W**, **Tuschl T.** 2001. Identification of novel genes coding for small expressed RNAs. *Science* **294**: 853–8.
40. **Lau NC**, **Lim LP**, **Weinstein EG**, **Bartel DP.** 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**: 858–62.
41. **Lee RC**, **Ambros V.** 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**: 862–4.
42. **Mourelatos Z**, **Dostie J**, **Paushkin S**, **Sharma A**, et al. 2002. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev* **16**: 720–8.
43. **Cuperus JT**, **Fahlgren N**, **Carrington JC.** 2011. Evolution and functional diversification of MIRNA genes. *Plant Cell* **23**: 431–42.
44. **Lyson TR**, **Sperling EA**, **Heimberg AM**, **Gauthier JA**, et al. 2012. MicroRNAs support a turtle + lizard clade. *Biol Lett* **8**: 104–7.
45. **Campbell LI**, **Rota-Stabelli O**, **Edgecombe GD**, **Marchioro T**, et al. 2011. MicroRNAs and phylogenomics resolve the relationships of Tardigrada and suggest that velvet worms are the sister group of Arthropoda. *Proc Natl Acad Sci USA* **108**: 15920–4.
46. **Sempere LF**, **Cole CN**, **McPeek MA**, **Peterson KJ.** 2006. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *J Exp Zool B Mol Dev Evol* **306**: 575–88.

47. **Lee C-T**, **Risom T**, **Strauss WM.** 2007. Evolutionary conservation of microRNA regulatory circuits: an examination of microRNA gene complexity and conserved microRNA target interactions. *DNA Cell Biol* **26**: 209–18.
48. **Niwa R**, **Slack FJ.** 2007. The evolution of animal microRNA function. *Curr Opin Genet Dev* **17**: 145–50.
49. **Peterson KJ**, **Summons RE**, **Donoghue PCJ.** 2007. Molecular palaeobiology. *Palaeontology* **50**: 775–809.
50. **Heimberg AM**, **Sempere LF**, **Moy VN**, **Donoghue PCJ**, et al. 2008. MicroRNAs and the advent of vertebrate morphological complexity. *Proc Natl Acad Sci USA* **105**: 2946–50.
51. **Kim VN**, **Han J**, **Siomi MC.** 2009. Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol* **10**: 126–39.
52. **Han J**, **Pedersen J**, **Kwon S**, **Belair C**, et al. 2009. Posttranscriptional crossregulation between Drosha and DGCR8. *Cell* **136**: 75–84.
53. **Lee Y**, **Jeon K**, **Lee JT**, **Kim S**, et al. 2002. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J* **21**: 4663–70.
54. **Winter J**, **Jung S**, **Keller S**, **Gregory RI**, et al. 2009. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat Cell Biol* **11**: 228–34.
55. **Starega-Roslan J**, **Koscianska E**, **Kozlowski P**, **Krzyzosiak W.** 2011. The role of the precursor structure in the biogenesis of microRNA. *Cell Mol Life Sci* **68**: 2859–71.
56. **Tsutsumi A**, **Kawamata T**, **Izumi N**, **Seitz H**, et al. 2011. Recognition of the pre-miRNA structure by *Drosophila* Dicer-1. *Nat Struct Mol Biol* **18**: 1153–8.
57. **Okamura K**, **Phillips MD**, **Tyler DM**, **Duan H**, et al. 2008. The regulatory activity of microRNA star species has substantial influence on microRNA and 3′ UTR evolution. *Nat Struct Mol Biol* **15**: 354–63.
58. **Reinhart B**, **Weinstein E**, **Rhoades M**, **Bartel B**, et al. 2002. MicroRNAs in plants. *Genes Dev* **16**: 1616–26.
59. **Xie ZX**, **Kasschau KD**, **Carrington JC.** 2003. Negative feedback regulation of Dicer-Like1 in *Arabidopsis* by microRNA-guided mRNA degradation. *Curr Biol* **13**: 784–9.
60. **Kurihara Y**, **Takashi Y**, **Watanabe Y.** 2006. The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *RNA* **12**: 206–12.
61. **Park M**, **Wu G**, **Gonzalez-Sulser A**, **Vaucheret H**, et al. 2005. Nuclear processing and export of microRNAs in *Arabidopsis*. *Proc Natl Acad Sci USA* **102**: 3691–6.
62. **Ambros V**, **Bartel B**, **Bartel D**, **Burge C**, et al. 2003. A uniform system for microRNA annotation. *RNA* **9**: 277–9.
63. **Meyers B**, **Axtell M**, **Bartel B**, **Bartel D**, et al. 2008. Criteria for annotation of plant MicroRNAs. *Plant Cell* **20**: 3186–90.
64. **Ma Z**, **Coruh C**, **Axtell MJ.** 2010. *Arabidopsis lyrata* small RNAs: transient MIRNA and small interfering RNA loci within the *Arabidopsis* genus. *Plant Cell* **22**: 1090–103.
65. **Ruby JG**, **Stark A**, **Johnston WK**, **Kellis M**, et al. 2007. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res* **17**: 1850–64.
66. **Chung W-J**, **Okamura K**, **Martin R**, **Lai EC.** 2008. Endogenous RNA interference provides a somatic defense against *Drosophila* transposons. *Curr Biol* **18**: 795–802.
67. **Zucker M**, **Mathews DH**, **Turner DH.** 1999. *Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide*. In Barciszewski J, Clark BFC, eds; RNA Biochemistry and Biotechnology. Dordrecht: Kluwer Academic Publishers.
68. **Carthew RW**, **Sontheimer EJ.** 2009. Origins and mechanisms of miRNAs and siRNAs. *Cell* **136**: 642–55.
69. **Nozawa M**, **Miura S**, **Nei M.** 2012. Origins and evolution of microRNA genes in plant species. *Genome Biol Evol* **4**: 230–9.
70. **Bentwich I**, **Avniel A**, **Karov Y**, **Aharonov R**, et al. 2005. Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet* **37**: 766–70.
71. **Lim L**, **Glasner M**, **Yekta S**, **Burge C**, et al. 2003. Vertebrate microRNA genes. *Science* **299**: 1540.
72. **Lai E**, **Tomancak P**, **Williams R**, **Rubin G.** 2003. Computational identification of *Drosophila* microRNA genes. *Genome Biol* **4**: R42.
73. **Lu J**, **Shen Y**, **Wu Q**, **Kumar S**, et al. 2008. The birth and death of microRNA genes in *Drosophila*. *Nat Genet* **40**: 351–5.
74. **Sdassi N**, **Silveri L**, **Laubier J**, **Tilly G**, et al. 2009. Identification and characterization of new miRNAs cloned from normal mouse mammary gland. *BMC Genomics* **10**: 149.
75. **Berezikov E**, **Liu N**, **Flynt A**, **Hodges E**, et al. 2010. Evolutionary flux of canonical microRNAs and mirtrons in *Drosophila*. *Nat Genet* **42**: 6–9.
76. **Chiang HR**, **Schoenfeld LW**, **Ruby JG**, **Auyeung VC**, et al. 2010. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev* **24**: 992–1009.
77. **Atayde VD**, **Tschudi C**, **Ullu E.** 2011. The emerging world of small silencing RNAs in protozoan parasites. *Trends Parasitol* **27**: 321–7.
78. **Collins LJ.** 2011. Characterizing ncRNAs in human pathogenic protists using high-throughput sequencing technology. *Front Genet* **2**: 96.
79. **Sperling EA**, **Robinson JM**, **Pisani D**, **Peterson KJ.** 2010. Where's the glass? Biomarkers, molecular clocks, and microRNAs suggest a 200-Myr missing Precambrian fossil record of siliceous sponge spicules. *Geobiology* **8**: 24–36.
80. **Margis R**, **Fusaro AF**, **Smith NA**, **Curtin SJ**, et al. 2006. The evolution and diversification of Dicers in plants. *FEBS Lett* **580**: 2442–50.
81. **Murphy D**, **Dancis B**, **Brown J.** 2008. The evolution of core proteins involved in microRNA biogenesis. *BMC Evol Biol* **8**: 92.
82. **de Jong D**, **Eitel M**, **Jakob W**, **Osigus H-J.**, et al. 2009. Multiple Dicer genes in the early-diverging metazoa. *Mol Biol Evol* **26**: 1333–40.
83. **Shabalina SA**, **Koonin EV.** 2008. Origins and evolution of eukaryotic RNA interference. *Trends Ecol Evol* **23**: 578–87.
84. **Chapman EJ**, **Carrington JC.** 2007. Specialization and evolution of endogenous small RNA pathways. *Nat Rev Genet* **8**: 884–96.
85. **Gough J**, **Karplus K**, **Hughey R**, **Chothia C.** 2001. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol* **313**: 903–19.
86. **MacRae IJ**, **Zhou KH**, **Li F**, **Repic A**, et al. 2006. Structural basis for double-stranded RNA processing by dicer. *Science* **311**: 195–8.
87. **Prucca CG**, **Slavin I**, **Quiroga R**, **Elias EV**, et al. 2008. Antigenic variation in *Giardia lamblia* is regulated by RNA interference. *Nature* **456**: 750–4.
88. **Chen X**, **Penny D**, **Collins L.** 2011. Characterization of RNase MRP RNA and novel snoRNAs from *Giardia intestinalis* and *Trichomonas vaginalis*. *BMC Genomics* **12**: 550.
89. **Kolev NG**, **Tschudi C**, **Ullu E.** 2011. RNA interference in protozoan parasites: achievements and challenges. *Eukaryot Cell* **10**: 1156–163.
90. **Kolev NG**, **Ullu E.** 2009. snoRNAs in *Giardia lamblia*: a novel role in RNA silencing? *Trends Parasitol* **25**: 348–50.
91. **Liu Q**, **Tuo W**, **Gao H**, **Zhu X-Q.** 2010. MicroRNAs of parasites: current status and future perspectives. *Parasitol Res* **107**: 501–7.
92. **Berezikov E**, **Robine N**, **Samsonova A**, **Westholm J**, et al. 2011. Deep annotation of *Drosophila melanogaster* microRNAs yields insights into their processing, modification, and emergence. *Genome Res* **21**: 203–15.