CHAPTER 1

EVOLUTIONARY GENOMICS OF microRNAs AND THEIR RELATIVES

Andrea Tanzer[a,d] Markus Riester,[a] Jana Hertel,[a] Clara Isabel Bermudez-Santana,[a,b] Jan Gorodkin,[c] Ivo L. Hofacker,[d] Peter F. Stadler,[a,d,e,f]

[a] Bioinformatics Group, Department of Computer Science; and Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstrasse 16-18, D-01407, Leipzig, Germany
[b] Department of Biology, Universidad Nacional de Colombia, Tr. 38 40-01, Bogotá, D.C., Colombia
[c] Division of Genetics and Bioinformatics, IBHV, University of Copenhagen, Grønnegårdsvej 3, DK-1870 Frederiksberg, Denmark
[d] Institute for Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria
[e] Fraunhofer Institute for Cell Therapy und Immunology, Deutscher Platz 5e, D-04103 Leipzig, Germany
[f] Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA

1.1 INTRODUCTION

MicroRNAs (miRNAs) are an abundant class of small noncoding RNA (ncRNA) genes which were found in eukaryotes, in particular in metazoans and plants, and in their viruses. MicroRNA research has come a long way, since the first discoveries of lin-4 [102] and let-7 [147] in *Caenorhabditis elegans*. The turn of the century brought the realization that miRNAs form a large new class of ncRNAs [101, 97, 95] that provide a ubiquitous and powerful mechanism for RNA-mediated control of gene expression. The `miRBase` [63], a comprehensive database collecting published miRNAs as well as assigning unique names [6] to novel ones, started with only 218 sequences (v1.0, December 2002) and now lists 6396 entries in the current version 11.0 (April 2008). Today (08/05/2008) there are 3233 publications about miRNAs in PubMed of which 755 are reviews. These numbers might ilustrate the impact of this field of research on our understanding of the infomation encoded by the fast majority of genomic sequences and transcribed units.

(Title, Edition). By (Author)

MicroRNAs were the first *small regulatory RNAs* found in animals, but turned out not to be the only ones. During the past few years, a variety of additional classes were detected, many of which share functional properties and processing machinery. In the following section we will address those similarities as well as differences by outlining biogenesis and function.

1.2 THE SMALL RNA ZOO

Size and final destination of the RNA classes addressed in this chapter define them as a reasonably homogeneous group of functional RNAs: They are about 20-30nt in length, and they guide large protein complexes to their targets, thus comprising the 'RNA sensor' allowing sequence specific binding of the proteins. Both miRNAs and siRNAs form subclasses of this large class of small ncRNAs. Like miRNAs, many other small RNAs are involved in gene silencing. Whereas miRNAs function post-transcriptionally, others are involved other types of function. MiRNAs stand out from the other small RNAs in many ways, in particular by their energetically stable precursor hairpin, which have been a key component in computational search methods.

While most of this contribution deals with microRNAs, in this section we attempt to compile the related small RNAs that got into the focus of RNA research. Given that about 1% of the human genome contains protein coding genes, it is likely that only a fraction of the regulatory RNome has been discovered so far. New insights constantly require regrouping of classes of small RNAs, such that our list can only provide a snapshot of the current knowledge.

1.2.1 Endogenous siRNAs

The term siRNAs (siRNAs) is often used for ∼20nt long regulatory RNAs and thus summarizes members of classes introduced in this section. However, the original meaning of the term siRNAs stems from Hamilton and Baulcombe [67], who discovered ∼25nt long RNA intermediates in either transgene-induced post-transcriptional gene silencing (PTGS) or virus-induced PTGS in plants. Meanwhile, siRNAs were detected in numerous eukaryotes across kingdoms [39]. They all originate from endogenous or exogenous (viral) transcripts, which are turned into double stranded RNA by RNA-dependent RNA polymerase (RdRP), show high complementarity to their target mRNAs and induce degradation of their targets. Endogenous siRNAs have also been found in most major eukaryote lineages, including animals (*Caenorhabditis elegans* [7], *Drosophila melanogaster* [9], and mouse [170, 189]), fungi (*Schizosaccharomyces pombe* [146]), amoebozoa (*Dictyostelium* [94]), plants (*Arabidopsis thaliana*, [195], and kinetoplastids (*Trypanosoma brucei* [43]). On the other hand, several lineages have lost the entire RNA interference (RNAi) machinery, including budding yeasts and Leishmanias, see [179] for a review of RNA interference in protozoan parasites.

Recently, endogenous siRNA were detected in higher eukaryotes that lack RdRP. A novel class of short interfering RNAs in *Drosophila melanogaster* was found to be excised from hairpins longer than animal miRNAs and in several instances longer than plant miRNAs [127]. These hairpins, termed *hpRNA* are located in regions of limited coding potential and were found by searching for inverted repeats resulting from inverted terminal repeats of transposons or tandem invertions of transposable elements and mRNAs. The siRNAs of size ∼21 are processed from the hairpin by known components of both the siRNA and miRNA pathways However, due to 5' methylation and their dependence on Dicer-2 and

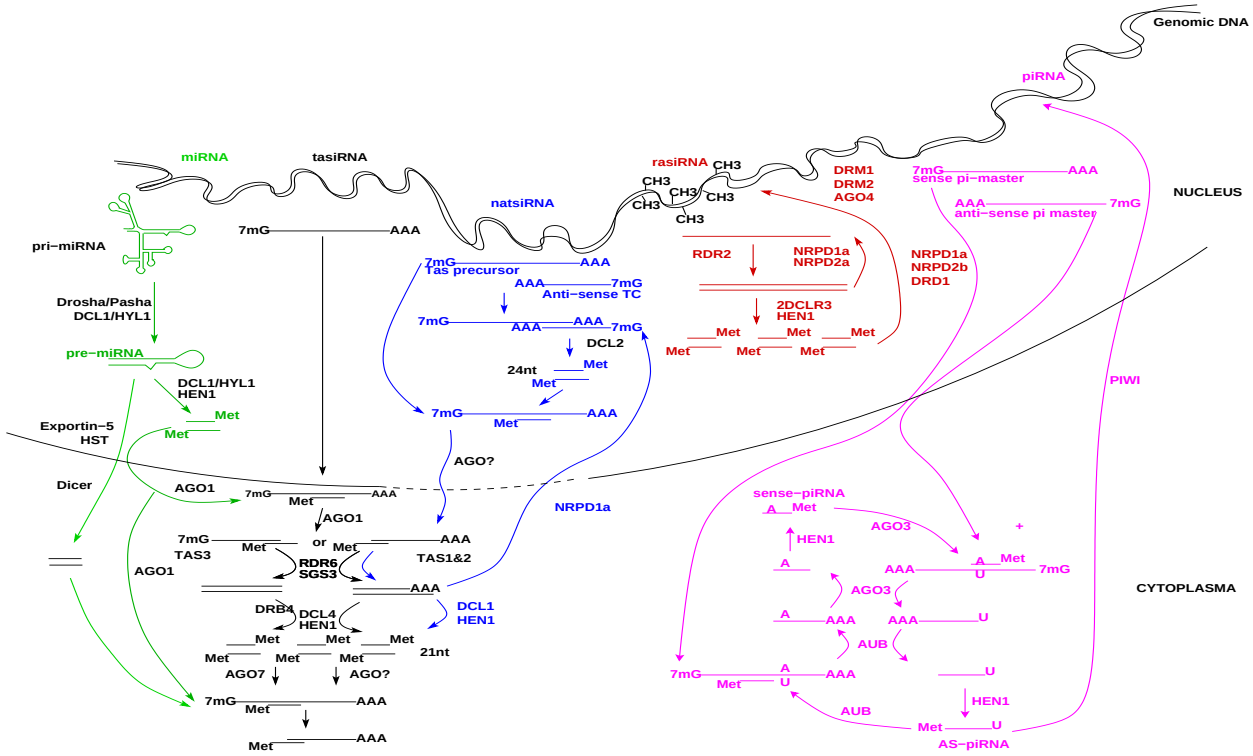

Figure 1.1 Biogenesis of major small RNA families. (green) miRNAs are transcribed as long primary transcripts, which are processed by the nuclear RNase III Drosha and its cofactor Pasha (DCL1/HYL1 in plants). In Vertebrates, these stem-loop structures are exported to the cytoplasm by means of the exportin-5 pathway, where the mature miRNA is cleaved by Dicer. In plants, the second cleavage step also takes place in the nucleus and short methylated dsRNAs are exported by HST. (black) plant tasiRNAs are processed in the cytoplasm. Tas precursors use the same export mechanism as protein coding mRNAs. miRNA primed synthesis of dsRNA is followed by DCL4 mediated dicing and HEN1 methylation. (blue) natsi RNAs in plants might use a mechanism similar to tasiRNAs. Cis anti-sense transcripts bind the sense RNA and serve as primers for RdRP (RNA dependent RNA polymerases). (red) rasi RNAs in plants never leave the nucleus. Primary transcripts are converted into dsRNA by RDP2 (an RdRP) and diced by DCL3. The resulting small RNAs guide DNA methylation. (magenta) A not yet complete model outlining piRNA processing (*ping-pong mechanism*). Transcription of piRNA clusters results in mature piRNAs antisense to their target transposon. Upon binding, an “antisense piRNA” is processed and interacts further with the transposon. Only weak sequence constraints are required. The process does not require any Dicer homolog. Cleavage is mediated by Ago3, Aubergine acts as a cofactor. [91]. (Figure based on drawings in refs. [87, 183])

AGO2 it was concluded that the short RNAs derived from hpRNAs are siRNAs and not miRNAs. In mice, pseudogenes and transposons were also shown to serve as source for potential siRNAs [170, 189]. So-called *tasiRNAs* (trans acting endogenous siRNAs) in plants are transcribed in trans to their target mRNAs and lead to mRNA degradation [87].

MicroRNAs and siRNAs share several components and processing steps in each of their maturation pathways. However, there are a number of differences. For instance, siRNAs show a high degree of sequence complementarity to their target sites compaired to miRNAs.

1.2.2 piRNA

Another class of small RNAs that was discovered in the attempt to find miRNAs are the germline specific piRNAs (Piwi interacting RNA) of length 25–32nt [98, 89, 10]. In *Drosophila melanogaster* piRNAs are involved in repression of transposons in the germline. In contrast to rasiRNAs (see below), piRNAs are restricted to specific genomic loci and are organized in a limited number of large clusters of noncoding transcripts.

PiRNAs were found to be expressed in two meiotic stages in spermatocytes. Pachytene piRNAs are depleted of repeats. Pre-pachytene piRNAs in contrast depend on Mili proteins, show similarity to repeat sequences, and mediate DNA methylation of transposable elements such as L1 elements [11].

In nematodes, the **21U** RNAs [149] are characterized by an initial uridine 5'-monophosphate, and a chemical modification at either the 2' or 3' oxygen of this nucloetide, as reported for small RNAs in plants and rasiRNAs in flies [106, 180]. A recent study identified them as the piRNAs of *Caenorhabditis elegans* by virtue of their association with Piwi-Argonaute [20]. They are far more diverse than miRNAs, and unlike siRNAs and piRNAs in other organisms, which are expressed in tight clusters, the 21U-RNAs appear to be autonomously expressed.

1.2.3 rasiRNAs

Repeat associated RNAs in animals[9] and plants [66] both lead to silencing of repeat regions by DNA methylation. However, they show certain differences in biogenesis. In plants, transcripts from transposons are turned into dsRNA by means of RdRP (RNA-dependent RNA Polymerase). In *D. melanogaster*, rasi RNAs were discovered in a genome wide screen for small RNAs and found to be expressed in testes and early embryos and were later shown to interact with Piwi proteins [10]. Thus rasiRNAs might be another family or subgroup of piRNAs.

1.2.4 "Exotic" small RNA species

Mouse Meryl RNA is transcribed during meiosis as a several kb long polyadenylated primary transcript and is then processed by Drosha into ∼80nt long fragments. Dicer products were found *in vitro* but not *in vivo*. The function of mrhl remains elusive. A homologous sequence was so far only found in rat [56]. The ciliate protozoan *Tetrahymena* undergoes a complicated mechanism of macro- and micronuclei formation during sexual reproduction. In the course of this process, DNA is removed from the macronuclei. Small scan RNAs (scnRNAs) originate from, possibly repeat or transposon containing, regions in the micronucleus, and guide histone methylation which in turn recruits proteins facilitating DNA excision. This process might help to prevent propagation of transposons onto the next *Tetrahymena* generation [87].

Another class of small RNAs whose function is not yet well understood are the 20–200nt log PASRs (promoter-associated small RNAs) and TASRs (termini-associated small RNAs). They associate with about 50% of mammalian protein coding genes in promoter and termini regions respectively and the PASRs also correlate with the expression of proteins [82]. It remains unclear at present, whether PASRs and TASRs are related to siRNAs in function and biogenesis, or whether they belong to an entirely distinct part of the cells' regulation system.

1.3 SMALL RNA BIOGENESIS

1.3.1 Components of the small RNA processing machinery

Type III RNases. RNase III type enzymes bind and cleave dsRNAs and are divided into 3 families. Besides the cleavage domain, they all contain an dsRNA binding domain. In small RNA pathways, we find members of class I and II.

Drosha, a class II enzyme, resides in the nucleus and requires Pasha (*H.s.* DGCR8 [DiGeorge syndrome critical region 8, a homolog to the *Drosophila melanogaster* Pasha]) as co-factor. It cleaves pre-miRNAs from longer precursors, which are then further processed by Dicer. So far, Drosha homologs were exclusively found in animals. Drosophilids and possibly all arthropods harbour two homologs, whereas all other metazoans have only a single copy [122].

Dicer, a Class III enzyme, has an N-terminal DExD/H-box helicase and a PAZ (Piwi/Argonaute/Zwille) domains. It 'dices' long dsRNA into $\sim$ 20nt long duplexes with a typical 2nt overhang at the 3' end. In contrast to Drosha, it is found in all organisms using small RNA pathways described here. The number of homologs within a genome varies greatly by organism. *Drosophila* has two (Dcr1 and Dcr2), all other metazoans and protists one and plants even four (DCL1-4) homologs involved in different small RNA pathways. [87, 122, 114].

Piwi Proteins and Argonautes. The family of Argonaute proteins (AGOs) comprises a multitude of different members of various functions[75]. AGOs consist of an N-terminal PAZ domain, also found in Dicer, and the C-terminal PIWI domain. The exact functions of the domains remain unresolved. However, the PIWI domain seems to bind to the 5' seed region of miRNAs, whereas the PAZ domain interacts with the 3'-OH. Vertebrates have four AGOs (Ago1-4, also known as eIFC1-4). Ago2 is required for RNAi, whereas Ago1 acts in translational inhibition. Both interact with Dicer [122]. For a detailed review of the numerous members of the Argonaut family we refer to [133]. Detailed studies in *Drosophila* were described in [176, 54].

Piwi proteins are predominantly expressed in the germ line. They contain the characteristic Piwi domain and were found to associate with piRNAs. In vertebrates, 3 Piwis were found so far: Mouse and zebrafish homologs are termed Mili, Miwi,Miwi2 and Zili, Ziwi, Ziwi2, respectively. Even though Mili is expressed in ovaries, Piwis seem to promote male germ line specific functions [10].

Polymerases. When it comes to transcription, small RNAs behave just like ordinary protein coding genes. Expression of miRNAs for instance has been studied in great detail. The primary transcripts originate either from introns (although often driven by an intronic promoter) or mlncRNAs (mRNA like noncoding RNAs). Most of them are transcribed

by DNA polymerase II and show alternative start and splice sites, are 5' caped and 3' polyadenylated.

Organisms with strong siRNA activity require another enzyme in order to multiply their response to parasitic RNAs. In plants, protozoans and lower metazoans, RdRP (RNA dependent RNA polymerases) performs siRNAs primed synthesis of dsRNA, which is then cleaved by RISC (RNA induced silecing complex) and Dicer homologs. In the case of plant rasiRNAs, the resulting small RNAs mediate silencing of the genomic loci of the parasitic sequences (transposable elements).

Even though endogenous siRNAs were found, Drosophilids and vertebrates lack endogenous RdRP homologs. Exogenous (transposon, viral encoded) RdRPs are not required for siRNA function [164]. It is tempting to speculate that this lack of RdRP in vertebrates might have led to the emergence of new defence mechanisms in order to respond to viral and other infections, e.g. the acquired immune system.

1.3.2 MicroRNA Biogenesis

Unless stated otherwise, we outline here miRNA biogenesis in the mammalian genome. (For miRNAs in introns, see below.) The process of miRNAs in intergenic regions, is that a primary pol II transcript is 5' capped and added a polyA tail [30]. Some of these transcripts can also function as protein coding mRNAs [30].

The primary transcript (*pri-miRNA*) is then further processed in the nucleus by the *microprocessor complex*, consisting of endonucleases Pasha and Drosha [62], resulting in a characteristic hairpin of length 60–120 nucleotides (nts). In plants, which do not contain Drosha, its function is carried out by the homologs DCL1 and HYL1 (reviewed in [79]).

The resulting stem-loop precursor, also referred to as the *pre-miRNA* is transported into the cytoplasm by Exportin-5 [113]. In the cytoplasm the pre-miRNA becomes processed further and is both sliced and diced. Dicer associates with TRBP (trans-activator RNA binding protein) and process the hairpin into a double stranded RNA (dsRNA) of length $\sim$22nt with a 2nt 3'-overhang.

In general, only one strand of the duplex termed mature miRNA will be incorporated into RISC to guide it to the target mRNA. The other strand (miRNA*) becomes degraded. However, recent results in *Drosophila melanogaster* revealed that a number of miRNA* sequences might be functional, since they are expressed above backgound signal and show higher conservation than expected of a non-functional sequence in a pri-miRNA helix [129]. The molecular machinery determines which of the two strands gets loaded into the RISC complex by sensing the strand which 5' end is less stable bound compared to the 3' end in the miRNA:miRNA* duplex [153, 93, 86, 177]. In addition, short conserved sequence motifs within the mature miRNA might serve as signals in both asymmentric processing and strand selection [60].

1.3.3 Biogenesis of other small RNAs

Only miRNAs are generated without the help of RdRP in both plants and animals, see fig. 1.1. In plants, the primary transcripts of other small RNAs are converted to dsRNA, which in turn is cleaved by Dicer homologs. The resulting small RNAs are often 3'-methylated by HEN1. In contrast, higher metazoans use small RNAs (most of them of unknown origin as in the case of piRNAs) to slice primary transcripts. In both cases, each RNA family requires a distinctive set of Ago, Piwi and Dicer homologs. Depending on the subcellular localization of pathway components and targets, small RNAs shuttle between nucleus and

cytoplasm. Exportin-5 is the only export pathway so far described in detail, but there are speculations about an piRNA specific transport mechanism.

For a more detailed description see fig. 1.1.

1.3.4 Three main mechanism, same global effect on gene expression

Originally, RNAi (RNA interference) described a variety of gene silencing processes which require small RNAs mediating site specificity. RNAi was discovered in *Caenorhabditis elegans* [52] and can be induced in a number of eukaryotes, as *Drosophila melanogaster* [84], vertebrates [48], and many protozoans [179]. In plants, co-suppression or transcriptional gene silencing (PTGS) was first described in petunia [124, 181] RNAi also refers to an efficient technology to knock down expression of specific genes [52] for which Craig C. Mello and Andrew Fire were awarded the Nobel Price for Medicine in 2006 [53]. (reviewed in [88])

1.3.4.1 Translational Inhibition *Classes: miRNA*

The small RNA binds to an mRNA and causes translational inhibition. The degree of base-pairing between RNA and target sequence as well as protein components in the miRNPs (Ago1) determines the mode of function. The so called seed region ($\sim$7nt on 5'end of RNA) mediates sequence specificity. RNA degradation requires almost perfect complementarity, whereas translational inhibition allows a certain number of unpaired bases. The actual mechanism behind translational repression has not been resolved yet. MicroRNAs were isolated from RNPs containing ribosomes, RISC, mRNA and miRNAs [45] suggesting that miRNA binding blocks transcriptional elongation by stalling ribosomes leading to release of the nascent transcript. In contrast, more recent studies showed that at least some miRNAs are able to inhibit the formation of the translational initiation complex [117]. Efficient miRNA repression in metazoan seem to be governed by multiple targets residing in the 3' UTR of the messenger, that is the same or different miRNAs target the same mRNA simultaneously. MicroRNA functions were reviewed in detail in [29].

1.3.4.2 RNAi: mRNA degradation *Classes: miRNA, siRNA, tasiRNA, natsiRNA, piRNA*

In contrast to translational repression, RNAi causes degradation of the target by RISC. Two factors determine this mode: the composition of the RISC complex and the small RNA:mRNA binding pattern. RNAi requires the presence of Ago2 and nearly perfect complementarity between small RNA and its target. Whereas metazoan miRNAs target the 3' end of the mRNA and by some not yet fully understood mechanism cause blocking of translation, miRNAs in plants target the coding region and cause degradation by an siRNA like pathway (reviewed in [51]).

1.3.4.3 Transcriptional gene silencing and Imprinting *Class: miRNA, rasiRNA, piRNA*

Small RNAs were shown to promote *de novo* methylation as well as maintenance of DNA methylation [13] in plants and animals. Several studies also gave rise to the idea that histone methylation of specific loci might be guided by small RNAs. MicroRNAs target promoter regions of genes, whereas rasiRNAs shut down repeat rich regions in the genome.

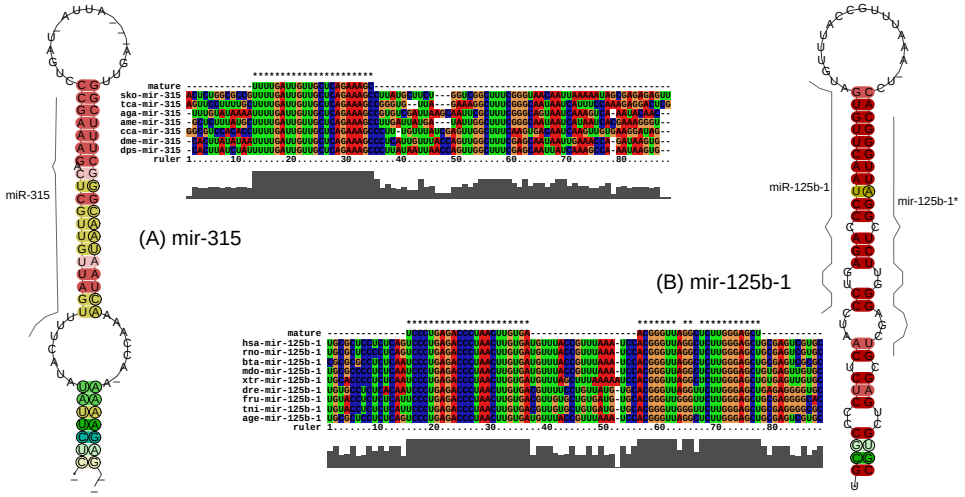

Figure 1.2 MicroRNA sequence and structure features illustrated by *mir-125b-1* and *mir-315*. The $\sim$ 85nt precursor folds into the typical hairpin structure (secondary structure predicted with `RNAfold`), which is cleaved by *Dicer* resulting in the mature miRNA ($\sim$ 20nt) indicated by a line. In case of *mir-125b-1*, the mature miR and miR* are both well conserved. For `mir-315` only one miR is expressed, which is much better conserved then the opposite side of the stem. The `ClustalW` multiple sequence alignment of the precursor sequences emphasize the conservation pattern. The colors of the base pair encode the number of consistent and compensatory mutations supporting that pair: Red marks pairs with no sequence variation; ochre, green and turquoise mark pairs with 2,3,4 different types of pairs, respectively.

1.4 COMPUTATIONAL MICRORNA PREDICTION

There are two basic strategies to detect novel miRNAs. The simpler one uses sequence homology to experimentally known miRNAs as well as the characteristic hairpin structure of the pre-miRNA [190, 104, 72, 41].

The *de novo* computational prediction of miRNAs primarily relies on the thermodynamically stable pre-miRNA hairpin and on the characteristic pattern of sequence conservation. Conservation is high at both sides of the stem region and is decreasing towards the unpaired region of the apical loop. If only one mature miRNA is produced from the precursor, the region encoding the mature sequences is best conserved. In some cases both sides of the hairpin produce mature sequences, usually labeled miR and miR*. In this case both mature loci are conserved nearly equally, as in the case of *mir-125b-1*, fig. 1.2,

Several software tools have been designed to utilize this information for miRNA gene finding: `miRscan` [109], `miRseeker` [96], and `miralign` [186], `RNAmicro` [73] all have lead to the discovery of a large number of animal microRNAs. For closely related species, phylogenetic shadowing can be used to identify regions that are under stabilizing selection and exhibit the characteristic variations in sequence conservation between stems, loop, and mature miRNA [24]. Genomic context also can give additional information: `Mirscan-II`, for example, takes conservation of surrounding genes into account [126], while the propensity of microRNAs to appear in genomic clusters is used as an additional selection criterion in [4].

On the other hand, there is the `miRank` tool [197], that is independent of genomic annotation and cross-species conservation. This is, in particular, important due to the quality of many sequenced genomes and the lack of well annotated related species.

MicroRNA detection without the aid of comparative sequence analysis is a very hard task but unavoidable when species-specific miRNAs are of prime interest. The `miR-abela` approach first searches for hairpins that are robust against changes in the folding windows (and also thermodynamically stabilized) and then uses a support vector machine (SVM) to identify microRNAs among these candidates [158]. A related technique is described by [198].

Conclusively, computational prediction of novel miRNAs can be roughly categorized into the following types: Straightforward sequence and/or structure homology search, the characterisation of candidates based on scored sequence and/or structural properties, machine learning techniques and the prediction of novel miRNAs in combination with putative targets, compare [199].

Plant miRNA precursors show much more variations in lengths and secondary structures and therefore filters must be less restrictive in this context. On the other hand, plant miRNA targets display complementary sites with near-perfect base-pairings. Tools like `findMIRNA` [1] thus predict miRNAs and their targets simultaneously and ignore candidate miRNA genes without putative targets.

1.5 MICRORNA TARGETS

Since microRNAs act as guide molecules that program the RISC complex to recognize a target mRNA, it is essential to understand the mechanism by which miRNAs recognize their targets and to predict target mRNAs for a given miRNA sequence.

To date the number of verified miRNA-mRNA interactions is still small. The `Tarbase` [157] database currently lists only 570 mRNAs targeted by 123 animal miRNAs. These known interactions have been used heavily to derive rules of miRNA-mRNA interactions. However, only a few guiding principles have emerged: (i) Perfect complementarity between miRNA and target is not required; in fact, most miRNA:mRNA complexes form imperfect duplexes containing mismatches as well as bulges. (ii) miRNA:mRNA duplexes are asymmetric; the 5' end of the miRNA (3' end of the target) binding more strongly than the 3' side of the miRNA. (iii) Base pairing at positions 9-11 triggers mRNA degradation, whereas mismatches at this positions lead to translational inhibition leaving mRNA merely intact. The region comprising positions 2-8 on the miRNA often exhibits perfect complementarity and is therefore referred to as the *seed* region [44, 5]. There are at best weak sequence signals associated with either miRNA or target sites. Target sites with evolutionarily conserved seed regions, however, show stronger regulatory impact than non-conserevd ones [16, 155]. Proteins from non-conserevd targets, however, outnumber those with conserved ones 6:1.

The context of the target site also influences protein response: an AU rich local neighborhood significantly increases the effect on protein expression [16]. Cooperative effects caused by additional target sites within 40nt can enhance PTGS. While the effect of multiple seed regions in the 3'-UTR is cummulative for translational repression, this is not the case for mRNA cleavage. For mir-223 [16], the majority of experientally verified targets with 7-8mer seed regions lead to mRNA mRNA destabilisation, while only a small fraction of mRNA remained stable and was downregulated vie translational repression.

1.5.1 How many targets?

Since miRNAs are short and need not match perfectly, it should come as no surprise that a single miRNA can regulate several targets. How many targets a typical miRNA might have is still open to debate. This also reflected in the widely fluctuating number of targets returned by the various target prediction approaches. For example Robins *et al.* [148] estimate less than 30 targets per miRNA, while Miranda *et al.* [119], based on their `rna22` method, suggest that a single miRNA may have several thousand targets. SILAC analysis (**s**table **i**sotope **l**abelling with **a**mino acids in **c**ell culture) [130] of mir-233 in neutrophils showed that 78 out of 3819 proteins investigated were direct targets. Since only a third of the proteom was quantified, mir-223 might have $\sim$ 200 targets in neutrophils and possibly even more targets specifically present in other celltypes and processes [16].

In part these diverging numbers may be due to the fact that it is not clear what constitutes a functional target site. Some targets of a miRNA might lead to only slightly lower protein expression levels, or may become functional only at elevated miRNA concentrations. It is clear that a large fraction of human mRNAs are under miRNA control. However, the more generous estimates for the number of miRNAs and the number of targets per miRNA suggest a picture where every mRNA is subject to regulation by a large ensemble of miRNAs from the cells miRNA milieu. In such a scenario any mutation in a 3' UTR would be expected to influence expression patterns. The observation that housekeeping genes seem to avoid miRNA regulation through the use of very short 3'UTRs [162] is consistent with this view.

MicroRNAs preferentially target mRNAs whose protein-products also have regulatory functions. Overrepresented groups include transcription factors, components of the miRNA machinery, and other proteins involved in translational regulation, as well as components of the ubiquitin machinery [77]. This points at an intricately interwoven network of transcriptional and post-transcriptional regulation [205].

The average number of targets per plant miRNA is low due to their high similarity to the target site and comprise mostly closely related genes [79]. One rare example for a miRNA with unrelated targets is *Arabidopis* mir395 [2], regulating an ATP sulfurylase and a sulfate transporter.

1.5.2 Target prediction

Over the past years a plethora of new methods have been proposed to predict microRNA targets, see [142] for a recent review. In most cases the initial search for candidate sites is purely sequence based. An often used approach, exemplified by the `miRanda` [77] and `PicTar` [92] programs, is to equip a standard local alignment algorithm with a scoring system that favours base complementarity, using separate scores for G-C, A-U, G-U pairs and mismatches. A similar effect can be achieved by training hidden Markov models [163], or even by pattern search using on sequence patterns that are overrepresented in a database of known miRNAs [119]. The resulting scores provide a measure for the thermodynamic stability of the miRNA:mRNA duplex. The sequence based methods can be substituted with a direct search for the most stably interacting sites under the standard energy model for RNA structures. The first such approach was implemented in `RNAhybrid` [145] and is slower than sequence alignment only by a constant factor. Alternatively, some methods, such as `TargetScan` [105] immediately start with a search for (near-)perfect seed matches which are then extended towards the 3' side.

In plants, approximate matching of the whole miRNA sequence is typically used and empirical scoring rules later penalize mismatches in the seed region [78, 152, 201]. There

are however recent findings that more extensively mismatched targets also exist in plants, which are missed with this approaches [28].

In any case, the initial phase tends to generate a large number of candidate sites that have to be further filtered and ranked in order to produce predictions with reasonable confidence. The most important features used to rank targets are (i) quality of the seed match; (ii) conservation of target site in related species; (iii) existence of multiple target sites in a single 3'UTR; (iv) sequence composition around the mRNA target site [64]; (v) hybridization energy of the miRNA:mRNA duplex; (vi) structural accessibility of the target site.

All of these criteria imply some kind of balance between sensitivity and specificity, i.e. the ability to predict as many target sites as possible while avoiding false predictions. For example, restricting oneself to targets with perfect seed complementarity significantly reduces the false positive rate but will exclude many valid targets. As a compromise some methods allow G-U base pairs or maybe a single mismatch or bulge within the seed region. Yet some validated targets have poorly matched seed regions that will defy almost any seed based approach [184, 42, 119].

Similarly, the introduction of evolutionary conservation led to a marked improvement in prediction accuracy [105, 92]. Many methods rely on conservation either by demanding that target sites for a particular miRNA occur in homologous genes from several species or, more strongly, that these target sites occur at homologous positions of the aligned mRNAs.

The work of [192] follows an alternative route by first determining conserved regions in 3'UTRs of mammalian mRNAs to determine more than 100 candidate motifs like involved in posttranscriptional regulation. More than half of them were then identified as a putative targets for known microRNAs.

Presumably, however, many microRNAs are evolutionary young or even species specific [22], and in this case evolutionary conservation is of little help.

Since secondary structure of the mRNA might interfere with miRNA binding, a few recent methods have tried to improve target predictions by including the effect of target site accessibility [112, 85]. Accessibility is usually expressed as the probability that the target site is free of secondary structure (and thus available for binding) or equivalently the free energy needed to break any existing structure. The total binding energy of the miRNA can than be expressed as the sum of the free energy gained from forming the hetero-duplex and the breaking energy expended to make the site accessible [121]. Including the breaking energy gives a significant improvement over using the interaction energy alone, as done e.g. in `RNAhybrid`, and may yield comparable performance with conservation based methods

Current target prediction methods are still burdened with a significant false positive rate. Presumably this is not because some vital ingredient is missing in current methods, but simply because the set of known validated targets (as well as known non-functional sites) is currently too small to allow optimizing the relative weight of the features discussed above. This situation may well change soon as significant experimental effort is expended for the large scale identification of miRNA targets, e.g. by immuno-precipitation of mRNAs with components of the RISC complex [47].

Comparing target prediction with experimental proteome analysis revealed that predictions from `Target Scan` and `Pictar`, which are both looking for seed matches, gave the most accurate results [16, 155].

1.5.3 Targets and polymorphisms

Single nucleotide polymorphisms (SNPs) can destroy miRNA targets sites or inactivate the miRNA itself [58]. In fact even a single substitution can have a dramatic effect [27]. Natural variation by SNPs not only disrupts miRNA-mRNA interactions but can also give rise to novel miRNA targets. A prime example is the Belgian Texel sheep, famous for their hyper-developed muscles. A QTL study of the phenotype [37, 57] uncovered a SNP in the the 3' UTR of the *myostation* gene (*gdf8*), which is involved in limiting the growth of muscle tissue. The G→A SNP creates target sites for mir-1 and mir-206, which result in down regulation of *myostation* and thereby of higher muscle growth. In a similar vein, a G→A SNP (var321) in the 3' UTR of SLITRK1, which is associated with Tourette's Syndrome, tightens the binding with miR-189. Recent work [185] reports a link between miR-433 and the SNPs in the FGF20 (*fibroblast growth factor 20*) gene, which is expressed in the brain a has been shown to be associated with Parkinson Disease. A more systematic study [151] identified about 400 SNPs in target sites and reported SNPs that give rise to ∼250 putative novel target sites.

SNP data were used to estimate that approximately 30%–50% of the non-conserved miRNA targets in 3' UTRs are functional when the transcript and miRNA are coexpressed [36]. Databases collecting disease-relevant miRNA-related SNPs are emerging: examples are `www.patrocles.org` by Georges and coworkers and PolymiRTS (`compbio.utmem.edu/miRSNP`) [17].

1.6 EVOLUTION OF MICRORNAs

1.6.1 Animal microRNAs

The numbers of annotated microRNAs collected in the `MiRBase` [1] [63] varies greatly between different animals. For instance, it currently lists 670 human and 184 frog miRNAs, but only 34 in the tuinucate *Ciona intestinalis* and 63 in the planarian *Schmidtea mediterranea*. A few microRNA families, such as let-7 [134], mir-1, and mir-124 [72] are well-conserved among most animal clades. On the other hand, many other families are evolutionary very young, some even specific to primates and possibly to human [24, 22].

Members of a given miRNA family can be fairly reliably recovered from genomic DNA sequences due to the extreme sequence conservation of the mature miR and the characteristic stable hairpin structure of the precursor. Such a systematic search for miRNA homologs can be used to determine first the phylogenetic distribution of a family and then to infer the likely time of evolutionary origin which must predate the last common ancestor of all extant family members. Pioneered in [72] and subsequently extended to increasingly larger data sets and complemented by experimental verification of predicted miRNAs [156, 140, 70], the analysis (see fig. 1.3 for updated data) reveals striking patterns in miRNA evolution and suggests that miRNAs have huge impacts on animal phylogeny. The dramatically expanding repertoire of both miRNA genes and their putative targets [100] appears to be correlated with major body-plan innovations. On the other hand, lineage specific microRNAs may account for phenotypic variation in closely related species.

A survey of the literature reported in [100] concludes that the diversity of the microRNA repertoire, the complexity of their expression patterns, and the diversity of the miRNA

[1] `MiRBase` 11.0 (April 2008), `http://microrna.sanger.ac.uk/sequences/`

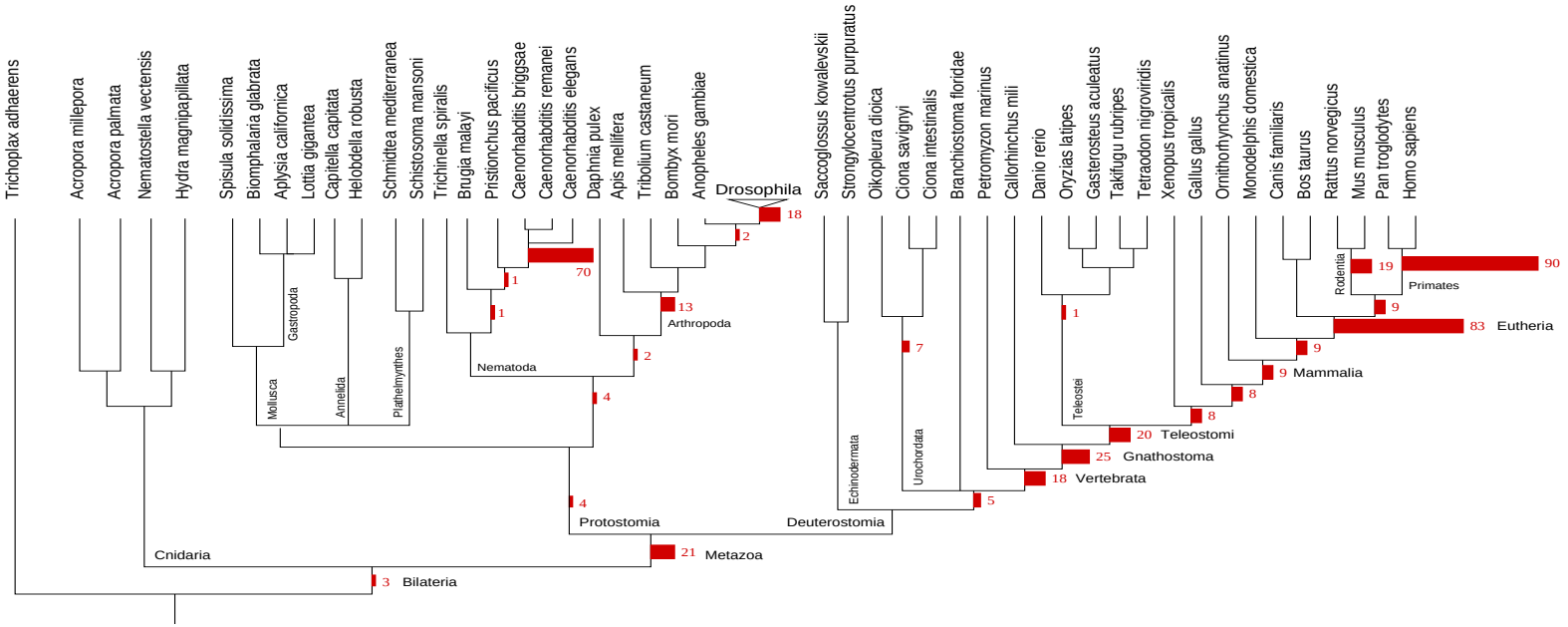

Figure 1.3 Evolution of animal miRNAs. Starting from `mirBase` 11.0 (April 2008), a comprehensive homology search in all genomes shown in the tree was conducted. Each microRNA family is mapped to the branch leading to the last common ancestor of the computationally identified homologs (for technical details we refer to [72]). Innovation of new miRNA families is clearly an on-going process in metazoan evolution. Due to the incomplete genomes of the lamprey *Petromyzon marinus* and the shark *Callorhinchus milii*, the assignment of innovations around the root of vertebrates is uncertain in details, and more complete data might shift some innovations back to the gnathostome and/or the vertebrate root. Taken together, there is, however, a clear increase microRNA innovation between the vertebrate ancestor and the split of the teleost and tetrapod lineages. The most striking burst of innovations, however, is observed in the eutherian ancestor. Note that the data are biased by the fact that independent surveys for miRNAs have been conducted only for a few model organisms, thus the lack of innovations along many of the invertebrate lineages might be due to missing data.

targets, are correlated with the animal's morphological complexity. Mechanistically, this is more than plausible since the miRNA pathway can influence large gene networks in a coordinated manner and miRNAs are known to be involved in the regulation of nearly all cellular processes.

The evolution of microRNAs is characterized not only by the continuing innovation of novel families but also by the diversification of established families spawning additional paralogous family members. Animal miRNAs are often organized in genomic clusters, usually indicating a single polycistronic primary precursor transcript, which may carry members of several distinct microRNA families. Like protein-coding gene families, the miRNA families evolve through gene duplications and gene loss, fig 1.7 [173, 172, 72]. Two distinct types of duplication events can be distinguished: (a) local duplications leading to additional copies on the same primary transcript, and (b) non-local duplications which eventually place the paralogs under different transcriptional control. The cause for non-local duplications are mostly the whole-genome duplications in early vertebrate and in the teleost evolution [161], while only a few individual duplications of primary miRNA precursor genes have been described [72]. In contrast to the typical mode in protein evolution, mature miRNA paralogs usually acquire no or only minimal substitutions, suggesting that functional differences between paralogs are predominantly caused by differences in the regulation of their expression and processing rather than by changes in the portfolio of their potential targets.

The continuing innovation of miRNAs is also highlighted by the presence of a large number of evolutionarily very young and sometimes even species-specific miRNAs. A pipeline designed to find miRNAs without enforcing initial constraints of evolutionary conservation [22] discovered 89 novel miRNAs of which 53 are primate specific. These finding partially overlap similar results from other groups [24, 25, 193].

In a high-throughput sequencing study [25] of human and chimpanzee small RNAs, Plasterk and co-workers found 447 miRNA that were novel at the time, of which 244 were expressed in human and 230 in chimpanzee with an overlap of only 27. Of novel these human miRNAs, more than 50 are specific to primates and 8% specific to human according to sequence conservation. The same study also shows that some miRNA families apparently expand in a species-specific fashion.

The general trend of expanding the microRNA repertoire in most lineages appears to correlate with increasing morpological complexity [72, 156, 140, 70]. The morphological simplification in urochordates, on the other hand, is accompanied by the loss of numerous highly conserved bilaterian miRNAs and a reorganisation of their miRNAs that clearly sets them apart from the other chordate lineages [55]. In *Oikopleura*, the effect is particularly striking. In urochordates, a large number of introns have been eliminated due to the strong pressure towards genome compression, explaining the reduction of the fraction of intronic microRNAs from $\sim$ 80% in vertebrates to less than 30%. The need to reduce genome size may also explain why the majority of urochordate miRNAs is located antisense to their target gene [55].

Another example for the opposite trend of minimizing or reorganising the miRNA repertoire can be found in the flatworm lineage. The planarian *Schmidtea mediterranea* encodes 71 miRNAs [132]. In contrast to other protostome lineages, most of their precursor sequences cannot be faithfully aligned with family members in other phyla. In the trematode flatworm *Schistosoma mansoni*, the closest sequenced relative of *Schmidtea*, we could recognize less than 10 microRNAs in a computational survey of its genome (unpublished data).

1.6.2 Plant microRNAs

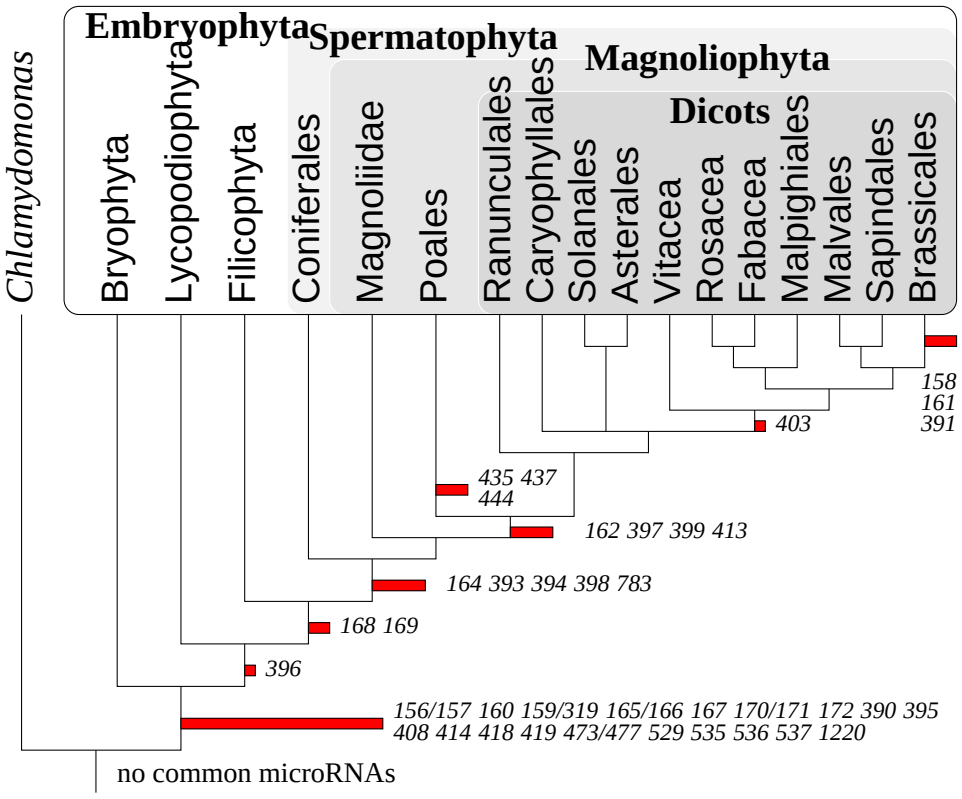

Figure 1.4 Phylogenetic distribution of plant microRNA families. As in fig. 1.3, microRNA families are mapped to the branch leading to the last common ancestor of anotated family members. The figure combines the data listed in `mirbase` 11 (April 2008) and in refs. [15, 18, 167, 191, 200].

As in animals, microRNA innovation is an ongoing process in plant evolution, fig. 1.4. Interestingly, there are much fewer distinct families of conserved miRNAs, many of which are evolutionarily very old, see e.g. [200, 15, 167]. At least 16 families date back to the last common ancestor of bryophytes and angiosperms. At the same time many plants studies so far exhibit large diverse sets of species-specific miRNAs that often outnumbering the conserved miRNAs. [15, 18, 49, 50, 110, 169]. Many of these species-specific miRNAs are single-copy genes and show significant sequence similarity with their putative targets, supporting the view that these miRNAs are indeed evolutionary very recent. Conceivably, some of the species-specific miRNAs may be misclassified members of other siRNA families.

Non-local events can be detected and dated by examining conservation patterns of protein coding genes flanking individual miRNA family members allowing calculation of phylogenetic trees of miRNA families [115]. About 67% of all *Arabidopsis* multifamily miRNA genes, for instance, emerged from local duplications.

1.6.3 MicroRNAs and Viruses

MicroRNAs regulate host-pathogen interactions in different directions (virus → virus, virus → host, host → virus) and stages of the viral life cycle (infectious, latent) and therefore pathways (replication, apoptosis, infection). The mode of interaction also depends on the subcellular localization of the virus within the host cell.

Since the first cleavage step of the pre-miRNA from the primary transcript takes place in the nucleus, viruses encoding their own miRNAs have to be able to cross the nuclear membrane. This is the case in particular for retroviruses, which even integrate into host genomes, and DNA viruses. RNA viruses remaining in the cytoplasm require either an transport mechanism shuttling their mRNA into the nucleus, or some alternative miRNA maturation pathway.

EBV (Epstein Barr virus also called human herpesvirus-4 (HHV-4)) was the first virus shown to encode several microRNAs [136] located in introns and UTRs. Typically, viral miRNAs are conserved only in closely related species or not at all, making their computational prediction a difficult task. A machine-learning approach using a set of properties of stem-loop structures such as free energy of folding, length and base pair compositions [135] nevertheless lead to the discovery of miRNAs from a diverse array of DNA and Retroviruses including *Herpesviruses*, *Polyomaviruses* and *Adenoviridae*. One microRNA each was found in Measle virus (*Paramyxoviridae*) and yellow fever virus (YFV, *Flaviviridae*) [135]. Drosophila C virus, a *Picornavirus* naturally associated with *Drosophila melanogaster*, also expresses small RNAs [9]. In the following, we briefly introduce a few well-studied cases; for more detailed reviews of microRNAs in viruses we refer to [123, 165, 150, 38, 46].

DNA Viruses Encode Their Own miRNAs. The latency-associated transcript (LAT) in Herpes simplex virus-1 (HSV-1) inhibits apoptosis and helps the virus to remain undetected in the infected hosts cells. A miRNA encoded in this transcript targets components (TGF-beta, SMAD3) of the TGF-beta pathway, which induces apoptosis, and thus sustain viability of the host. [65]

In Simian virus 40 (SV40), a member of the polyomavirus family, sv40-miR-S1 is processed from the 3' UTR of the late pre-mRNA. It targets T-antigen, one of the viral early genes, which in turn is recognized by cytotoxic T lymphocytes (CTLs). The mechanism enables the virus to escape the host defense response. The miRNA is highly conserved across related primate polyomaviruses [166].

Epstein-Barr virus (EBV) encodes several miRNA. One of them (miR-BART2) was predicted to target viral DNA polymerase BALF5 and was recently shown to inhibit transition from latent to lytic viral replication [19]. miR-K12-10 from Kaposi's sarcoma associated virus is encoded in the ORF of the karposi gene. Excision of the miRNA caused cleavage of the mRNA. In addition, this miRNA provides an editing site leading to a glycine to serine change in the kaposin protein [135].

RNA viruses regulated by host miRNAs. Human mir-122 leads to accumulation of viral RNA during Hepatitis C virus (HCV) infection, and was therefore suggested to positively interfere with viral replication. This explains why successful HCV infection depends on the presence of mir-122 [80]. Knocking down components of the miRNA Pathways [144] or mir-122 [81] leads to reduced HCV replication. Other RNA viruses are sensitive to host miRNAs. Mouse mir-24 and mir-93 serve as host defense by targeting large protein

(L protein) and phosphoprotein (P protein) genes of rhabdovirus vesicular stomatitis virus (VSV) [131].

Retroviral RNAs blocking host miRNA pathways. Several retroviral RNAs have been shown to enter or interfere with the miRNA pathways and thus cause inhibition of the host machinery. Adenoviral VA RNA*I I* is processed by dicer. It is highly abundant in late-cells and blocks the host miRNA machinery by saturating the various protein components. For instance, 60% of the small RNAs incorporated into the RISC complex resemble viral VA RNA*I I* products. [8, 196]. Two miRNAs of the mir-17 cluster (mir-17 and mir-20a) target histone acetyltransferase Tat cofactor PCAF, an important factor HIV-1 replication. The expression of these miRNAs was found to be actively suppressed by HIV-1 [178]. Human miR-32, finally, targets Tas, a gene of primate foamy virus type 1 (PFV-1), that suppresses the microRNA pathways [99].

Plant viruses. As to-day, no miRNAs in plant virus genomes have been reported and although exogenous RNAi plays a central role in fighting viruses in plants [67], there is no evidence that miRNAs are directly involved in responses to viral infections. High mutation rates allow viruses to escape miRNA cleavage by quickley altering sequences of putativee miRNAs target sites [159]. Furthermore, almost every plant virus encodes suppressors of the siRNA-mediated host response to infections and some of these inhibit steps that are shared with the miRNA pathway [83, 35]. It is however reported that some viruses without such PTGS suppressors may also exploit the miRNA pathway [21]. An example for a plant miRNA with a probably regulatory role in an infection response has been observed in *Brassica rapa*. Here, an evolutionary young, *Turnip mosaic virus* (TuMV) induced miRNA cleaves specific disease-resistance genes of the TIR-NBS-LRR class [69].

1.6.4 Mirtrons

Mirtrons are alternative precursors for microRNAs that employ the splicing machinery for the first steps of their processing, thereby bypassing Drosha cleavage. This alternative processing pathway was recently described in mammals, *Drosophila*, and *Caenorhabditis* [23, 128, 61], fig. 1.5, and even in rice a candidate mirtron has been reported recently [206].

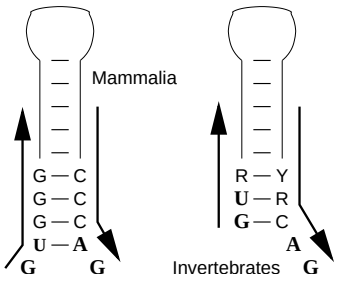

Figure 1.5 Mirtrons exhibit characteristic sequence patterns just inside the exon/intron boundaries that differ significantly between vertebrates and invertebrates [23]. The splice-donor **GU** and the splice acceptor **AG** are shown in bold. Arrows indicate the mature microRNAs, which can be located on both arms. While their 5' end is well defined, there is some variation at their 3'end.

While mirtrons are often well-conserved within nematodes, insects, and vertebrates, none of the known mirtrons is shared between these clades. Since vertebrate and invertebrate mirtrons exhibit several differences, fig. 1.5, Berezikov *et al.* [23] suggested that the mirtron pathway evolved independently in several clades. Alternatively, mirtron sequences might not be sufficiently well-conserved in order to unambiguously establish homology between phyla.

1.7 ORIGIN(S) OF microRNA FAMILIES

1.7.1 Metazoa

Since almost the entire eukaryotic genome is transcribed [175], there is no shortage in RNAs that can potentially enter the microRNA processing cascade. In fact, stem-loop structures of the approximate size of microRNA precursors are a highly abundant feature of random RNA sequences. It stands to reason that a sizeable fraction of these is sufficiently stable and symmetric to be processed. Indeed, a computational approach that started with an initial search for all hairpins in the human genome and subsequently employed stringent computational and experimental filtering [22] identified 53 primate-specific microRNAs. Similarly, in [24] several lineage-specific miRNAs are listed, some of them exhibiting rapid evolution. This picture is reinforced by high-throughput sequencing [25], which found hundreds of specific miRNAs in human and chimp brain, respectively. This lead us to conclude [173] that novel metazoan microRNA families constantly arise from expressed transcripts that are currently not unter strong selection. Hairpins formed by these precursor RNAs are then processed with a non-negligible probability to novel microRNAs, which are retained and rapidly optimized if they provide a beneficial regulatory impact.

In a some cases, the precursor transcript can be identified either as a repetitive element (see Section 1.7.3) or a pseudogene. The latter are good candidate for ancestors of novel miRNA-bearing transcripts, because expressed pseudogenes are found in a reasonable number in many genomes, often arising from strongly expressed genes such as housekeeping genes. Examples of observed miRNAs in pseudogenes are the primate specific *mir-220* and *mir-492* [40].

1.7.2 Mechanisms in plants

Evolutionary young, species specific plant miRNAs often show high sequence similarities to their target genes even beyond the mature miRNA sequence. For example, both arms of *miR822* show extended similarity with DC1 domain containing genes [2], and a similar pattern was reported for *mir161*, *mir163* [3], *miR826* and *miR841* [141] and their predicted targets. In some cases, the sequence similarities also include promoter regions [187]. This observations lead to the *inverted duplication hypothesis* [3], which postulates that miRNA genes arise from local inverted duplications of their target genes. A variant of this mechanism has been proposed for *miR842* and *miR846* in *Arabidopsis* [141], where miRNA and miRNA* likely arose by an early duplication event within their targets. Later duplications then generated this miRNA loci. Transcription of such young miRNA genes produces foldback structures that are probably processed by *DCL4* and aquired mutations then may lead to a switch to *DCL1* processing [14].

1.7.3 microRNAs and Transposable Elements

A subset of the mammalian miRNAs are derived from Transposable elements (TEs), tab. 1.1. This phenomenon appears to be associated with the expansion of TEs in mammalian genomes, since no repeat-related miRNA precursors have been reported in chicken or *Drosophila*. The single example in *Caenorhabditis elegans*, *cel-mir-69*, was later reclassified as siRNA [108].

Overall, TE-derived miRNAs are significantly less conserved than non-TE derived ones [139], and the list include several lineage-specific miRNAs (e.g. rno-mir-333 and hsa-mir-

Table 1.1 microRNAs derived from Transposable Elements

Repeat class	Mammalian microRNAs [a]
LINE	{(hsa, mmu, rno) 28, 151/151*, 325}, {(hsa, mmu) 374, 421, 493}, {(hsa) 95, 545, 552, 562, 571, 576, 578, 579, 582, 588, 606, 616, 619, 625, 626, 634, 644, 648, 649}
MITE	{(hsa) 361, 513-[a-1, a-2], 544, 548-[a-1, a-2, a-3, b, c, d-1, d-2], 570, 579, 584, 587, 603, 645, 652}
SINE	{(hsa, mmu) 130-b, 330, 345, 370, 378}, {(hsa) 422-a, 566, 575, 607, 619, 633, 640, 649}, {(rno) 333}
LTR	{(hsa) 548-a-3, 558}, {(mmu) 297}, {(rno) 327}
DNA(mariner)	{(hsa, mmu, rno) 340}
Other (Arthur1)	{(hsa) 659}
	Plant microRNAs [b]
DNA	{(ath) 416},{(osa) 439-[a, b], 817, 821-[a, b, c]}
LTR	{(ath) 401, 854-[a, b, c, d], 855}, {(osa) 416, 420, 531}
MITE	{(ath) 405 [a, b, d]}{(osa) 442, 443, 445-a, 806-[b, g] 807-[b, c] 809-h, 811-[a, b, c], 812-[a, b, c, d, e], 813, 814-[a, b, c], 816, 818-[b, e], 819-[a, d, f, g, h], 821-[a, b, c]}

Abbreviations. hsa: human, mmu: mouse, rno :rat, ath: arabidopsis, osa: rice. LINE: long interspersed element; MITE: miniature inverted-repeat transposable element; SINE: short interspersed element; LTR: long terminal repeat retrotransposons; MIR: mammalian interspersed repeat

References. a compiled from [160, 139, 137] and *b* from [12, 138].

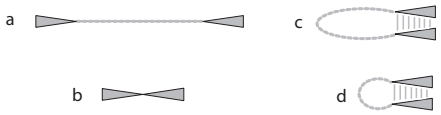

Figure 1.6 Transition from a full-length DNA element (a) with terminal inverted repeats (black triangles) enclosing an ORF to a MITE (b) which consists of the inverted repeats only. Transcripts with a large internal region (c) give rise to siRNAs, while short hairpin RNAs arising from MITEs (d) are the first step towards generating microRNAs from TEs. Adapted from [137, 138].

95). The better conserved ones mostly stem from L2 and MIR elements [160], while mariner derived elements MADE1 and other miniature inverted-repeat transposable elements (MITEs) are a major source of human-specific microRNAs (fig. 1.6 and [204]).

Several genomic loci in plants have been reported to encode both siRNAs and miRNAs. Comparative analysis revealed that these are repeat-derived [138]: While long nearly exact double-strands, including those formed by the terminal inverted repeats of full-length DNA elements, produce siRNAs, miRNAs are derived from short imperfect hairpin structures. The latter may arise from MITEs, which consist of two terminal inverted repeats with little intervening sequence (fig. 1.6).

1.7.4 Are Animal and Plant microRNAs homologous?

Until very recently, endogeneous miRNAs were known only in multicellular organism: landplants and metazoans. This picture changed with the discovery of miRNAs in the

green algae *Chlamydomonas reinhardtii* [202, 120] and in the slime mold *Dictyostelium discoideum* [74]. A computational study presents evidence for miRNAs in Trypanosomes [116], although these reports have not yet been verified experimentally. There is no convincing evidence that any of the known microRNA families dates back to the last common ancestor of plants and animals. The only published candidate, mir854/855 [12], cannot be traced consistently through either plant or animal phylogeny; the low-complexity sequence is most likely an analogous invention. In the same vein, none of the seed plant microRNAs is related to microRNAs of the green algae *Chlamydomonas reinhardtii* [202, 120]. *Chlamydomonas* miRNAs differ in serveral respects from microRNAs of landplants. In particular, multiple mature miRs are processed from a single stem loop. The slime mold miRNAs [74] also show no homology to either plant, *Chlamydomonas*, or animal miRNAs.

The small RNA processing pathways and the RNAi machinery in particular are evolutionarily very old [179], presumably dating back to ancestral eukaryote since its components are present in the most basal lineages [31]. For the origin of the miRNA processing machinery there are two possible scenaria between which we cannot distinguish based on the available evidence:

(1) It arose once, rather early in eukaryote evolution. In this case, the ancestral microRNAs have then long been replaced by more modern innovations in the different kingdoms, while the protein components of the microRNA processing machinery have been retained.

(2) The endogeneous production of specific miRNAs has evolved multiple times with different requirements on the RNAs to be processed. Thus, not only the microRNAs arose independently but the processing machinery was also derived multiple times from ancestral siRNA pathway(s).

Chlamydomonas, for instance, has undergone extensive duplications of Dicer and Argonaute proteins after the divergence of the green algae and land plant lineages leading to a diversification of the core RNAi machinery [32].

1.8 GENOMIC ORGANIZATION

1.8.1 Clusters and Families

Mammalian genomes contain two distinct types of microRNA clusters. In the first type, groups of microRNAs expressed from polycistronic primary precursors are easily recognized by the syntenic conservation of their genomic location over long evolutionary times [72]. Such clusters typically contain only a few miRNA precursor hairpins, the largest and most impressive example being the *mir-17* clusters, whose evolution is summarized in fig. 1.7 above. The largest cluster of this type in vertebrates is the mir-379/mir-656 cluster, located in human within the imprinted *DLK-DIO3* region on chromosome 14 [33]. This cluster is present in the genomes of all sequenced placental mammals [59]. `MiRBase` [63] lists 42 miRNAs in human and 37 in mouse located in this cluster. Its members are produced from a large non-coding RNA [154].

The second type of clusters consists of large numbers of miRNAs which are transcribed independently or possibly in small groups. An example is the C19MC cluster [26], whose members are individually transcribed by pol-III utilizing the promoters of Alu elements.

In constrast, miRNA clusters are not frequently observed in plant genomes. One of these exceptional case is the miRNA-395 family. Clusters of various sizes and intergenic distances

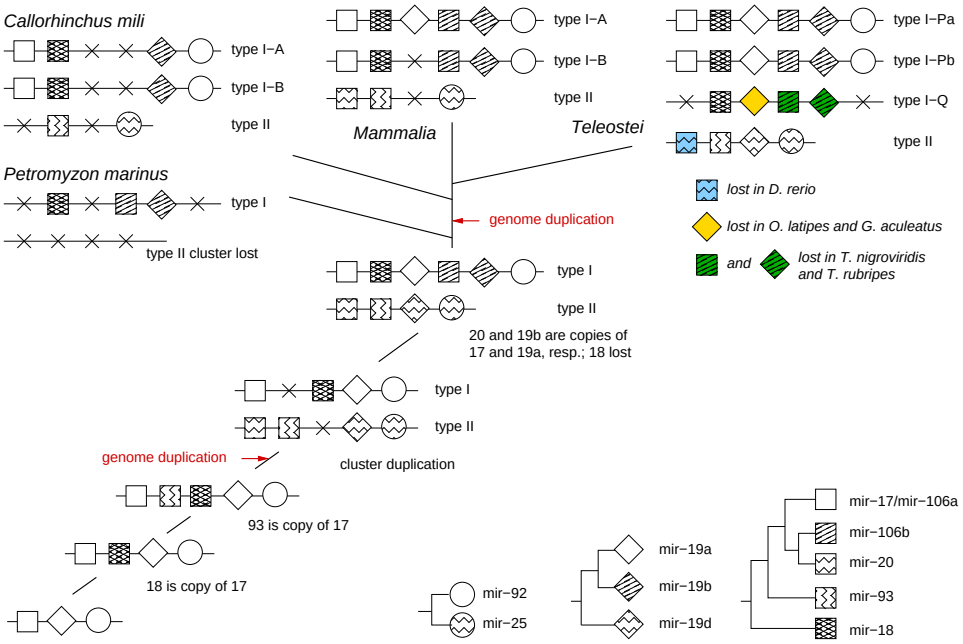

Figure 1.7 The evolution of the mir-17 clusters is governed by a complex history of duplications and loss of individual members as well as duplications of entire clusters. The extant clusters consist of members of three non-homologous groups of miRNAs, namely the *mir-17*, *mir-19*, and *mir-92* groups each of which is composed of several subfamilies with different `mirbase` names (lower right insets). Only *mir-92* pre-dates the origin of vertebrates, which is the earliest evidence for clusters stemming from lamprey and shark. The formation of the ancestral cluster, and the divergence of both *mir-18* and *mir-93* from the *mir-17* group appears to have pre-dated the first round of genome duplication in the ancestral vertebrate. Differential loss of one of the *mir-93* and *mir-18* paralogs apparently followed the first duplication. The two clusters then evolved independently: The type-I cluster was extended by a duplication of the entire region from *mir-17* over *mir-18* to *mir-19a*, immediately behind *mir-19a* and a secondary loss of the *mir-18* copy. MiccroRNAs of the type-II cluster evolved independently in their sequence, resulting in homologous miRNAs *mir-106a, 19d* and *mir-25*. Only a single cluster was found in the genome of the lamprey *Petromyzon marinus* which contains both a *mir-20* and *mir-19b* homolog, suggesting that it shares the first genome duplication. A second round of genome duplication results in two copies of type-I clusters while the type-II cluster was not duplicated at all. In elephant fish (*Callorhinchus milii*) as an early representative whose genome was exposed to 2 genome duplications, *mir-19a* and *mir-20* were lost in both type-I clusters and *mir-106a* and *mir-19d* were lost in the single type-II cluster. In mammals, the homologous miRNAs *mir-19a* and *mir-19d* were lost in the second copy of type-I cluster and the type-II cluster, respectively, while the first copy of the type-I cluster remained complete. In teleost fishes, which underwent a third whole genome duplication, the two copies of type-I clusters were duplicated and one of these duplicated clusters was lost subsequently, resulting in 3 type-I clusters (type-I-A,B,C). Again, only one copy of the type-II cluster was retained. In zebrafish (*Danio rerio*) the first gene of the type-II cluster (*mir-106b*) and the first (*mir-17*) and last (*mir-92*) ones of the third copy of type-I cluster were lost. While *mir-19a* is absent in medaka (*Oryzias latipes*) and in stickleback (*Gasterosteus aculeatus*), the pufferfishes *Takifugu rubripes* and *Tetraodon nigrovirides* lost *mir-20* and *mir-19b*.
The figure is based on a re-evaluation and extension of earlier studies of the mir-17 cluster [173, 174].

have been reported for several genomes and rice EST data indicate that at least some of them are expressed as single, policistronic transcript [78]. A MIR156 tandem cluster has been reported both in several monocots and in the dicotylednous plant *Ipomea nil*; MIR169 and MIR1219 are also observed as clusters in distantly related plants [169, 168]. In all these cases, the clusters contain only members of the same family.

In contrast to all other landplants investigated to-date, about a quarter of the miRNAs of the moss *Physcomitrella patens* are located in clusters [169, 15]. The exceptional microRNAs of the green alga *Chlamydomonas reinhardtii* are also partially clustered. In particular, several members of the MIR918 family are potentially derived from a single stem-loop [202, 120].

1.8.2 Regulation of microRNA Expression

Most pri-miRNAs are transcribed by RNA polymerase II since these transcripts contain cap structures as well as poly(A) tails [103, 30]. Core promoters have been characterized in both animals and plants [204]. Transcriptional regulation of pri-miRNAs does not seem to differ substantially from potein-coding pol-II transcripts, although only a few examples have been analyzed in detail. Expression of the human mir-21 gene, for example, depends on the transcription factor Stat3 due to two Stat3 binding sites in an upstream enhancer region that is strictly conserved since the first observed evolutionary appearance of mir-21 and Stat3 [111]. This connection between microRNA and transcription factor is highly conserved in evolution, fig. 1.8. Recently, the involement of Stat3 on mir-21 expression was also observed in a teleost [143]. MicroRNA misregulation by the oncogenic transcription factor Myc [34] and the tumor suppressor p53 [71] contributes to tumorigenesis. Phylogenetic footprinting, furthermore, revealed that transcription factors that play essential roles in development preferentially regulate miRNA genes in *Drosophila* [185]. A recent analysis of the primary precursors of mouse microRNAs uncovered a conserved sequence element that might be involved in post-transcriptional regulation of microRNAs [68].

Some microRNAs, most notably the members of the C19MC cluster on human Chr.19, are transcribed by RNA polymerase III, in this case utilizing Alu repeats to recruit the polymerase [26].

In general, it not clear whether transposable elements provide the transcriptional starts for adjacent microRNAs [160], although human MITEs are transcribed as read-through transcripts initiated from adjacent genomic positions and not by means of a strand-specific promoter provided be the transposable element itself [137]. Transposable elements that dually code for microRNAs and siRNAs can be expressed as readthrough transcripts from intronic regions of spliced RNA messages [138].

About a quarter of the human microRNAs is located in introns of protein-coding genes [107], in *Xenopus* intronic locations are predominant [171]. Contrary to intronic snoRNAs, the majority of intronic miRNAs are processed from unspliced intronic regions before the catalysis of splicing in vertebrates [90] and show a bias towards large 5' introns [203]. However, recently discovered mirtrons are diced from unbranched introns (see 1.6.4).

Differential miRNA precursor processing in both the nucleus and cytoplasm may lead to distinct expression profiles of both miRNA precursors and their mature microRNAs, indicating that post-transcriptional processing plays an important role in regulating miRNA expression [125, 171]. MicroRNA specific sequence motifs located within a few hundred nucleotides upstream of the pre-mir hairpain in both nematodes and vertebrates [126, 76] may well be involved in these processes.

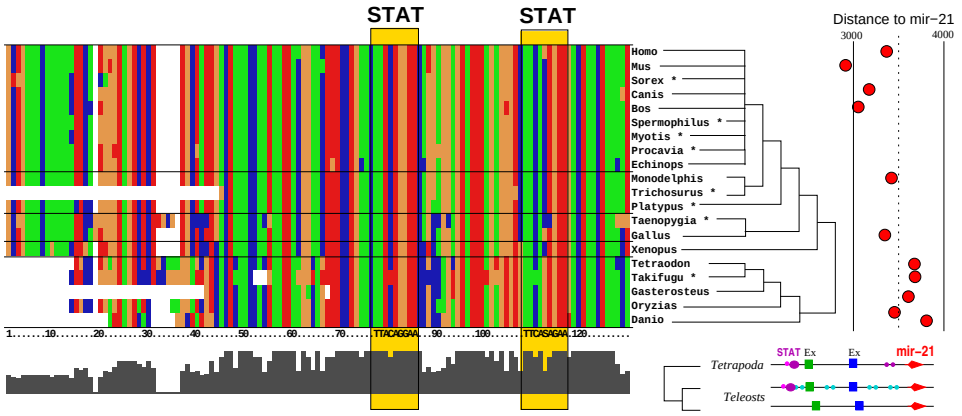

Figure 1.8 Evolutionarily conserved regulation of mir-21 by Stat3. A highly conserved enhancer featuring two Stat3 binding sites (l.h.s.: schematic sequence alignment) is located between 3 and 4kb upstream of the pre-mir-21 hairpin (r.h.s.). Although the enhancer is located in an intron of the adjacent TMEM49 gene, this distance does not correlate with genome size suggesting that the enhancer region is associated with mir-21 and independent of the TMEM49 gene. Adapted from [111].

Plant pri-miRNAs are probably also transcribed by RNA Polymerase II [193]. Differential pri-miRNA processing is also observed in plants and splicing variants of several pri-miRNAs have been reported (e.g. [193, 188]). In constrast to animals, the overwhelming majority of plant miRNA genes are located in regions between annotated genes [146] and only a few miRNA loci have been reported to overlap with protein-coding genes [141, 15]. In general, little is known about the regulation of miRNA expression in plants. MicroRNAs miR162 and miRNA168, for example, were shown to be negative regulators of miRNA pathways in plants by targeting *DCL1* [194] and *AGO1* [182], respectively, both central genes of the plant miRNA machinery. In *Arabidopsis*, the *DCL1* pre-mRNA level is also self-regulated with the help of mir838, which is located in intron 14 of *DCL1*. High levels of *DCL1* proteins lead to a competition with the splicing machinery and *DCL1* processed *DCL1* primary transcripts are non-functional and subject to degradation [141].

Analyses of the upstream regions of known miRNAs revealed a TATA box sequence motif in the promotor region [193, 188]. Furthermore, binding sites of the transcription factors AtMYC2, ARF, SORLREP3, and LFY are overrepresented in comparison to protein-coding genes, indicating an important role of these transcription factors in miRNA regulation [118]. The miRNA319a locus has been investigated in different species from *Brassicaceae* [188] and a strongly conserved upstream region has shown to be essential for transcription.

1.9 SUMMARY AND OUTLOOK

We are just about to understand the signifficance of small RNAs as regulators in eukaryotes. Starting from miRNAs, researcher from a variety of disciplines have set out for the quest for other small noncoding transcripts by exploring the RNome. The impact of the findings was remarkable. Not only have we found that transcription goes beyond/extends regions of protein coding genes, but that these noncoding regions are of signifficant information content. Intergenic DNA, often described as 'playground of evolution', turned out to harbour

a plethora of cis and trans regulatory elements, many of them in the form of noncoding RNAs. In this chapter we try to provide a comprehensive view of one such class of ncRNAs, but also underscored the importance of other RNA regulators. To-date, microRNAs are one of the best described classes of small ncRNAs.

Less than a decade of microRNA research has profoundly changed the perceptions of the role of RNAs from rather uninteresting carriers of coding information to key players in cellular regulation. Indeed, microRNAs affect gene expression on multiple levels: specific histone methylation patterns alter the accessibility of genomic regions, activation or silencing of promoters defines transcriprtional activity of genes, and finally post-transcriptional gene silencing (PTGS) of mRNAs serves as another checkpoint before energy consuming translation into protein takes place. In all these processes, miRNAs serve as the exchangeable RNA module in large protein complexes and sign responsible for specific interactions with the target sequences.

The consequences of this novel picture of eukaroytic regulation need to be explored in more detail, using also approaches from systems biology. Studying the evolutionary history of genes and targets revealed an RNA-based gene regulatory layer, implying an additional source for genome plasticity. Question on how novel RNAs contribute to an increase in genome complexity and how they lead to the emergence of novel traits remain largely unanswered. Tracing back the ancestor(s) of recent small ncRNAs seems a promising approach towards undestanding whether small RNAs in plants and metazoans are analogies or homologies. The protein machinery that facilitates processing as well as functionality clearly shares main features, but at the same time posesses enough flexibility to allow the aquisition of novel RNA substrates.

Even the epigenome turned out to be under the influence of RNA control. The discovery of rasiRNAs and piRNAs, for instance, reveals a flux of information between generations that goes beyond the “programs” hard coded in our genomes. Are we indeed ribo-organsims? If so, we need to be careful in designing RNA based drugs. Shortly after siRNAs were introduced as the ultimate tool for transient knock-down expriments, researchers found themselves dealing with cross-reactivity and other unexplainable side effects or even no effects at all. Nevertheless, these new technologies seem to be key in the development of new lab technologies and medical applications. A variety of diseases, foremost cancer, were linked to RNA misexpression, once again pointing out that small can be mighty.

REFERENCES

1. A. Adai, C. Johnson, S. Mlotshwa, S. Archer-Evans, V. Manocha, V. Vance, and V. Sundaresan. Computational prediction of miRNAs in *Arabidopsis thaliana*. *Genome Res.*, 15:78–91, 2005.
2. E. Allen, Z. Xie, A. M. Gustafson, and J. C. Carrington. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell*, 121:207–221, 2005.
3. E. Allen, Z. Xie, A. M. Gustafson, G. H. Sung, J. W. Spatafora, and J. C. Carrington. Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat. Genet.*, 36:1282–1290, 2004.
4. Y. Altuvia, P. Landgraf, G. Lithwick, N. Elefant, S. Pfeffer, A. Aravin, M. J. Brownstein, T. Tuschl, and H. Margalith. Clustering and conservation patterns of human microRNAs. *Nucleic Acids Res.*, 33:2697–2706, 2005.
5. V. Ambros. The functions of animal microRNAs. *Nature*, 431:350–355, 2004.

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

7. V. Ambros, R. C. Lee, A. Lavanway, P. T. Williams, and D. Jewell. MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr. Biol.*, 13:807–818, 2003.

8. O. Aparicio, N. Razquin, M. Zaratiegui, I. Narvaiza, and P. Fortes. Adenovirus virus-associated RNA is processed to functional interfering RNAs involved in virus production. *J Virol*, 80:1376–1384, 2006.

9. A. Aravin, M. Lagos-Quintana, A. Yalcin, M. Zavolan, D. Marks, B. Snyder, T. Gaasterland, J. Meyer, and T. Tuschl. The small RNA profile during *Drosophila melanogaster* development. *Dev Cell*, 5:337–350, 2003.

10. A. A. Aravin, G. J. Hannon, and J. Brennecke. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science*, 318:761–764, 2007.

11. A. A. Aravin, R. Sachidanandam, A. Girard, K. Fejes-Toth, and G. J. Hannon. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science*, 316:744–747, 2007.

12. M. Arteaga-Vázquez, J. Caballero-Pérez, and J. P. Vielle-Calzada. A family of microRNAs present in plants and animals. *Plant Cell*, 18:3355–3369, 2006.

13. W. Aufsatz, M. F. Mette, J. van der Winden, A. J. Matzke, and M. Matzke. RNA-directed DNA methylation in *Arabidopsis*. *Proc Natl Acad Sci U S A*, 99 Suppl 4:16499–16506, 2002.

14. M. Axtell and J. Bowman. Evolution of plant microRNAs and their targets. *Trends Plant Sci.*, 13:343–349, Jul 2008.

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

16. D. Baek, J. Villén, C. Shin, F. D. Camargo, S. P. Gygi, and D. P. Bartel. The impact of micrornas on protein output. *Nature*, 2008. epub ahead of print.

17. L. Bao, M. Zhou, L. Wu, L. Lu, D. Goldowitz, R. W. Williams, and Y. Cui. PolymiRTS database: linking polymorphisms in microRNA target sites with complex traits. *Nucleic Acids Res*, 35:51–54, 2007.

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

19. S. Barth, T. Pfuhl, A. Mamiani, C. Ehses, K. Roemer, E. Kremmer, C. Jäker, J. Höck, G. Meister, and F. A. Grässer. Epstein-barr virus-encoded microRNA miR-BART2 down-regulates the viral DNA polymerase BALF5. *Nucleic Acids Res*, 36:666–675, 2008.

20. P. J. Batista, J. G. Ruby, J. M. Claycomb, R. Chiang, N. Fahlgren, K. D. Kasschau, D. A. Chaves, W. Gu, J. J. Vasale, S. Duan, D. Conte Jr., S. Luo, G. P. Schroth, J. C. Carrington, D. P. Bartel, and C. C. Mello. PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Mol Cell.*, 31:67–78, 2008.

21. A. Bazzini, H. Hopp, R. Beachy, and S. Asurmendi. Infection and coaccumulation of tobacco mosaic virus proteins alter microRNA levels, correlating with symptom and plant development. *Proc. Natl. Acad. Sci. U.S.A.*, 104:12157–12162, 2007.

22. I. Bentwich, A. A. Avniel, Y. Karov, R. Aharonov, S. Gilad, O. Barad, A. Barzilai, P. Einat, U. Einav, E. Meiri, E. Sharon, Y. Spector, and Z. Bentwich. Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.*, 37:766–770, 2005.

23. E. Berezikov, W. J. Chung, J. Willis, E. Cuppen, and E. C. Lai. Mammalian mirtron genes. *Mol Cell*, 28:328–336, 2007.

24. E. Berezikov, V. Guryev, J. van de Belt, E. Wienholds, and H. A. Ronald Plasterk. Phylogenetic shadowing and computational identification of human microRNA genes. *Cell*, 120:21–24, 2005.

25. E. Berezikov, F. Thuemmler, L. W. van Laake, I. Kondova, R. Bontrop, E. Cuppen, and R. H. Plasterk. Diversity of microRNAs in human and chimpanzee brain. *Nat Genet.*, 38:1375–1377, 2006.

26. G. M. Borchert, W. Lanier, and B. L. Davidson. RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol.*, 13:1097–1101, 2006.

27. J. Brennecke, A. Stark, R. B. Russell, and S. M. Cohen. Principles of microRNA-target recognition. *PLoS Biol*, 3, 2005.

28. P. Brodersen, L. Sakvarelidze-Achard, M. Bruun-Rasmussen, P. Dunoyer, Y. Yamamoto, L. Sieburth, and O. Voinnet. Widespread translational inhibition by plant miRNAs and siRNAs. *Science*, 320:1185–1190, 2008.

29. N. Bushati and S. M. Cohen. microRNA functions. *Annu Rev Cell Dev Biol*, 23:175–205, 2007.

30. X. Cai, C. H. Hagedorn, and B. R. Cullen. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA*, 10:1957–1966, 2004.

31. J. M. Carlton, R. P. Hirt, J. C. Silva, A. L. Delcher, M. Schatz, Q. Zhao, J. R. Wortman, S. L. Bidwell, U. C. M. Alsmark, S. Besteiro, T. Sicheritz-Ponten, C. J. Noel, J. B. Dacks, P. G. Foster, C. Simillion, Y. Van de Peer, D. Miranda-Saavedra, G. J. Barton, G. D. Westrop, S. Müller, D. Dessi, P. L. Fiori, Q. Ren, I. Paulsen, H. Zhang, F. D. Bastida-Corcuera, A. Simoes-Barbosa, M. T. Brown, R. D. Hayes, M. Mukherjee, C. Y. Okumura, R. Schneider, A. J. Smith, S. Vanacova, M. Villalvazo, B. J. Haas, M. Pertea, T. V. Feldblyum, T. R. Utterback, C.-L. Shu, K. Osoegawa, P. J. de Jong, I. Hrdy, L. Horvathova, Z. Zubacova, P. Dolezal, S.-B. Malik, J. M. Logsdon, K. Henze, A. Gupta, C. C. Wang, R. L. Dunne, J. A. Upcroft, P. Upcroft, O. White, S. L. Salzberg, P. Tang, C.-H. Chiu, Y.-S. Lee, T. M. Embley, G. H. Coombs, J. C. Mottram, J. Tachezy, C. M. Fraser-Liggett, and P. J. Johnson. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science*, 315:207–212, 2007.

32. J. Casas-Mollano, J. Rohr, E. Kim, E. Balassa, K. van Dijk, and H. Cerutti. Diversification of the core RNA interference machinery in *Chlamydomonas reinhardtii* and the role of DCL1 in transposon silencing. *Genetics*, 179:69–81, 2008.

33. J. Cavaillé, H. Seitz, M. Paulsen, A. Ferguson-Smith, and J. Bachellerie. Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Hum. Mol. Genet.*, 11:1527–1538, 2002.

34. T. Chang, D. Yu, Y. S. Lee, E. A. Wentzel, D. E. Arking, K. M. West, C. V. Dang, A. Thomas-Tikhonenko, and J. T. Mendell. Widespread microRNA repression by Myc contributes to tumorigenesis. *Nat. Genet.*, 40:43–50, 2008.

35. P. Chellappan, R. Vanitharani, and C. Fauquet. MicroRNA-binding viral protein interferes with *Arabidopsis* development. *Proc. Natl. Acad. Sci. U.S.A.*, 102:10381–10386, 2005.

36. K. Chen and N. Rajewsky. The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet*, 8:93–103, 2007.

37. A. Clop, F. Marcq, H. Takeda, D. Pirottin, X. Tordoir, B. Bibé, J. Bouix, F. Caiment, J. M. Elsen, F. Eychenne, C. Larzul, E. Laville, F. Meish, D. Milenkovic, J. Tobin, C. Charlier, and M. Georges. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat Genet*, 38:813–818, 2006.

38. B. R. Cullen. Viruses and microRNAs. *Nat Genet*, 38 Suppl:25–30, 2006.

39. B. Czech, C. D. Malone, R. Zhou, A. Stark, C. Schlingeheyde, M. Dus, N. Perrimon, M. Kellis, J. A. Wohlschlegel, R. Sachidanandam, G. J. Hannon, and J. Brennecke. An endogenous small interfering RNA pathway in *Drosophila*. *Nature*, 2008.

40. E. J. Devor. Primate microRNAs miR-220 and miR-492 lie within processed pseudogenes. *J. Heridity*, 97:186–190, 2006.

41. T. Dezulian, M. Remmert, J. F. Palatnik, D. Weigel, and D. H. Huson. Identification of plant microRNA homologs. *Bioinformatics*, 22:359–360, 2006.

42. D. Didiano and O. Hobert. Perfect seed pairing is not a generally reliable predictor for microRNA-target interactions. *Nat Struct Mol Biol*, 13:849–851, 2006.

43. A. Djikeng, H. Shi, C. Tschudi, and E. Ullu. RNA interference in *Trypanosoma brucei*: cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs. *RNA*, 7:1522–1530, 2001.

44. J. Doench and P. Sharp. Specificity of microRNA target selection in translational repression. *Genes Dev*, 18:504–511, 2004.

45. J. Dostie, Z. Mourelatos, M. Yang, A. Sharma, and G. Dreyfuss. Numerous microRNPs in neuronal cells containing novel microRNAs. *RNA*, 9:180–186, 2003.

46. D. M. Dykxhoorn. MicroRNAs in viral replication and pathogenesis. *DNA Cell Biol*, 26:239–249, 2007.

47. G. Easow, A. Teleman, and S. Cohen. Isolation of microRNA targets by miRNP immunopurification. *RNA*, 13:1198–1204, 2007.

48. S. M. Elbashir, W. Lendeckel, and T. Tuschl. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.*, 15:188–200, 2001.

49. N. Fahlgren, M. D. Howell, K. D. Kasschau, E. J. Chapman, C. M. Sullivan, J. S. Cumbie, S. A. Givan, T. F. Law, S. R. Grant, J. L. Dangl, and J. C. Carrington. High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of MIRNA genes. *PLoS ONE*, 2:e219, 2007.

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

51. W. Filipowicz, S. N. Bhattacharyya, and N. Sonenberg. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet*, 9:102–114, 2008.

52. A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391:806–811, 1998.

53. A. Z. Fire. Gene silencing by double-stranded RNA. *Cell Death Differ*, 14:1998–2012, 2007.

54. K. Förstemann, M. D. Horwich, L. Wee, Y. Tomari, and P. D. Zamore. *Drosophila* microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. *Cell*, 130:287–297, 2007.

55. X. Fu, M. Adamski, and E. M. Thompson. Altered miRNA repertoire in the simplified chordate, *Oikopleura dioica*. *Mol. Biol. Evol.*, 25:1067–1080, 2008.

56. G. Ganesan and S. M. Rao. A novel noncoding RNA processed by Drosha is restricted to nucleus in mouse. *RNA*, 2008.

57. M. Georges, A. Clop, F. Marcq, H. Takeda, D. Pirottin, S. Hiard, X. Tordoir, F. Caiment, F. Meish, B. Bibé, J. Bouix, J. M. Elsen, F. Eychenne, E. Laville, C. Larzul, D. Milenkovic, J. Tobin, and A. C. Charlier. Polymorphic microRNA-target interactions: a novel source of phenotypic variation. *Cold Spring Harb Symp Quant Biol*, 71:343–350, 2006.

58. M. Georges, W. Coppieters, and C. Charlier. Polymorphic miRNA-mediated gene regulation: contribution to phenotypic variation and disease. *Curr Opin Genet Dev*, 17:166–176, 2007.

59. E. Glazov, S. McWilliam, W. Barris, and B. Dalrymple. Origin, evolution, and biological role of miRNA cluster in DLK-DIO3 genomic region in placental mammals. *Mol. Biol. Evol.*, 25:939–948, 2008.

60. J. Gorodkin, J. Havgaard, M. Ensterö, M. Sawera, P. Jensen, M. Ohman, and M. Fredholm. MicroRNA sequence motifs reveal asymmetry between the stem arms. *Comput Biol Chem*, 30:249–254, 2006.

61. J. G. Graham Ruby, C. H. Jan, and D. P. Bartell. Intronic microRNA precursors that bypass Drosha processing. *Nature*, 48:83–86, 2007.

62. R. I. Gregory, K. P. Yan, G. Amuthan, T. Chendrimada, B. Doratotaj, N. Cooch, and R. Shiekhattar. The microprocessor complex mediates the genesis of microRNAs. *Nature*, 432:235–240, 2004.

63. S. Griffiths-Jones, H. K. Saini, S. van Dongen, and A. J. Enright. `miRBase`: tools for microRNA genomics. *Nucleic Acids Res.*, 36:D154–D158, 2008.

64. A. Grimson, K. Farh, W. Johnston, P. Garrett-Engele, L. Lim, and D. Bartel. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, 27:91–105, 2007.

65. A. Gupta, J. J. Gartner, P. Sethupathy, A. G. Hatzigeorgiou, and N. W. Fraser. Anti-apoptotic function of a microRNA encoded by the HSV-1 latency-associated transcript. *Nature*, 442:82–85, 2006.

66. A. Hamilton, O. Voinnet, L. Chappell, and D. Baulcombe. Two classes of short interfering RNA in RNA silencing. *EMBO J*, 21:4671–4679, 2002.

67. A. J. Hamilton and D. C. Baulcombe. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, 286:950–952, 1999.

68. S. He, H. Su, C. Liu, G. Skogerbø, H. He, D. He, X. Zhu, T. Liu, Y. Zhao, and R. Chen. MicroRNA-encoding long non-coding RNAs. *BMC Genomics*, 21:236, 2008.

69. X. He, Y. Fang, L. Feng, and H. Guo. Characterization of conserved and novel microRNAs and their targets, including a TuMV-induced TIR-NBS-LRR class R gene-derived novel miRNA in *Brassica*. *FEBS Lett.*, 582:2445–2452, 2008.

70. A. M. Heimberg, L. F. Sempere, V. N. Moy, P. C. J. Donoghue, and K. Peterson. MicroRNAs and the advent of vertebrate morphological complexity. *Proc. Natl. Acad. Sci. USA*, 105:2946–2950, 2007.

71. H. Hermeking. p53 enters the microRNA world. *Cancer Cell*, 12:414–418, 2007.

72. J. Hertel, M. Lindemeyer, K. Missal, C. Fried, A. Tanzer, C. Flamm, I. L. Hofacker, P. F. Stadler, and The Students of Bioinformatics Computer Labs 2004 and 2005. The expansion of the metazoan microRNA repertoire. *BMC Genomics*, 7:15 [epub], 2006.

73. J. Hertel and P. F. Stadler. Hairpins in a haystack: Recognizing microRNA precursors in comparative genomics data. *Bioinformatics*, 22:e197–e202, 2006. ISMB 2006 contribution.

74. A. Hinas, J. Reimegård, E. G. Wagner, W. Nellen, V. Ambros, and F. Söderbom. The small RNA repertoire of *Dictyostelium discoideum* and its regulation by components of the RNAi pathway. *Nucleic Acids Res.*, 6714-6726:35, 2007.

75. G. Hutvagner and M. J. Simard. Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol*, 9:22–32, 2008.

76. A. Inouchi, S. Shinohara, H. Inoue, K. Kita, and M. Itakura. Identification of specific sequence motifs in the upstream region of 242 human miRNA genes. *Comput Biol Chem.*, 31:207–214, 2007.

77. B. John, A. J. Enright, A. Aravin, T. Tuschl, C. Sander, and D. S. Marks. Human microRNA targets. *PLoS Biol*, 2:e363, 2004.

78. M. W. Jones-Rhoades and D. P. Bartel. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell*, 14:787–799, 2004.

79. M. W. Jones-Rhoades, D. P. Bartel, and B. Bartel. MicroRNAs and their regulatory roles in plants. *Annu Rev Plant Biol*, 57:19–53, 2006.

80. C. Jopling, M. Yi, A. Lancaster, S. Lemon, and P. Sarnow. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science*, 309:1577–1581, 2005.

81. C. L. Jopling, K. L. Norman, and P. Sarnow. Positive and negative modulation of viral and cellular mRNAs by liver-specific microRNA mir-122. *Cold Spring Harb Symp Quant Biol*, 71:369–376, 2006.

82. P. Kapranov, J. Cheng, S. Dike, D. A. Nix, R. Duttagupta, A. T. Willingham, P. F. Stadler, J. Hertel, J. Hackermüller, I. L. Hofacker, I. Bell, E. Cheung, J. Drenkow, E. Dumais, S. Patel, G. Helt, M. Ganesh, S. Ghosh, A. Piccolboni, V. Sementchenko, H. Tammana, and T. R. Gingeras. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, 316:1484–1488, 2007.

83. K. Kasschau, Z. Xie, E. Allen, C. Llave, E. Chapman, K. Krizan, and J. Carrington. P1/HC-Pro, a viral suppressor of RNA silencing, interferes with *Arabidopsis* development and miRNA function. *Dev. Cell*, 4:205–217, 2003.

84. J. R. Kennerdell and R. W. Carthew. Use of dsRNA-mediated genetic interference to demonstrate that *frizzled* and *frizzled-2* act in the wingless pathway. *Cell*, 95:1017–1026, 1998.

85. M. Kertesz, N. Iovino, U. Unnerstall, U. Gaul, and E. Segal. The role of site accessibility in microRNA target recognition. *Nat. Genet.*, 39:1278–84, 2007.

86. A. Khvorova, A. Reynolds, and S. Jayasena. Functional siRNAs and miRNAs exhibit strand bias. *Cell*, 115:209–216, 2003.

87. V. N. Kim. MicroRNA biogenesis: coordinated cropping and dicing. *Nat Rev Mol Cell Biol*, 6:376–385, 2005.

88. V. N. Kim. Small RNAs: classification, biogenesis, and function. *Mol Cells*, 19:1–15, 2005.

89. V. N. Kim. Small RNAs just got bigger: Piwi-interacting RNAs (piRNAs) in mammalian testes. *Genes Dev*, 20:1993–1997, 2006.

90. Y. K. Kim and V. N. Kim. Processing of intronic microRNAs. *EMBO J.*, 26:775–783, 2007.

91. C. Klattenhoff and W. Theurkauf. Biogenesis and germline functions of piRNAs. *Development*, 135:3–9, 2008.

92. A. Krek, D. Grün, M. N. Poy, R. Wolf, L. Rosenberg, E. J. Epstein, P. MacMenamin, I. da Piedade, K. C. Gunsalus, M. Stoffel, and N. Rajewsky. Combinatorial microRNA target predictions. *Nat Genet*, 37:495–500, 2005.

93. J. Krol, K. Sobczak, U. Wilczynska, M. Drath, A. Jasinska, D. Kaczynska, and W. J. Krzyzosiak. Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. *J Biol Chem*, 279:42230–42239, 2004.

94. M. Kuhlmann, B. Popova, and W. Nellen. RNA interference and antisense-mediated gene silencing in *Dictyostelium*. *Methods Mol Biol.*, 346:211–226, 2006.

95. M. Lagos-Quintana, R. Rauhut, W. Lendeckel, and T. Tuschl. Identification of novel genes coding for small expressed RNAs. *Science*, 294:853–858, 2001.

96. E. C. Lai, P. Tomancak, R. W. Williams, and G. M. Rubin. Computational identification of *Drosophila* microRNA genes. *Genome Biol.*, 4:R42 [Epub], 2003.

97. N. C. Lau, L. P. Lim, E. G. Weinstein, and D. P. Bartel. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, 294:858–862, 2001.

98. N. C. Lau, A. G. Seto, J. Kim, S. Kuramochi-Miyagawa, T. Nakano, D. P. Bartel, and R. E. Kingston. Characterization of the piRNA complex from rat testes. *Science*, 313:363–367, 2006.

99. C. H. Lecellier, P. Dunoyer, K. Arar, J. Lehmann-Che, S. Eyquem, C. Himber, A. Saïb, and O. Voinnet. A cellular microRNA mediates antiviral defense in human cells. *Science*, 308:557–560, 2005.

100. C. T. Lee, T. Risom, and W. M. Strauss. Evolutionary conservation of microRNA regulatory circuits: an examination of microRNA gene complexity and conserved microRNA-target interactions through metazoan phylogeny. *DNA Cell Biol.*, 26:209–218, 2007.

101. R. C. Lee and V. Ambros. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*, 294:862–864, 2001.

102. R. C. Lee, R. L. Feinbaum, and V. Ambros. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, 75:843–854, 1993.

103. Y. Lee, M. Kim, J. Han, K. H. Yeom, S. Lee, S. H. Baek, and V. N. Kim. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J*, 23:4051–4060, 2004.

104. M. Legendre, A. Lambert, and D. Gautheret. Profile-based detection of microRNA precursors in animal genomes. *Bioinformatics*, 21:841–845, 2005.

105. B. P. Lewis, C. B. Burge, and D. P. Bartel. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, 120:15–20, 2005.

106. J. Li, Z. Yang, B. Yu, J. Liu, and X. Chen. Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in *Arabidopsis*. *Curr. Biol.*, 15:1501–1507, 2005.

107. S. C. Li, P. Tang, and W. C. Lin. Intronic microRNA: discovery and biological implications. *DNA Cell Biol.*, 26:195–207, 2007.

108. L. P. Lim, N. C. Lau, E. G. Weinstein, A. Abdelhakim, S. Yekta, M. W. Rhoades, C. B. Burge, and D. P. Bartel. The microRNAs of *Caenorhabditis elegans*. *Genes Dev.*, 17:991–1008, 2003.

109. L. P. Lim, N. C. Lau, E. G. Weinstein, A. Abdelhakim, S. Yekta, M. W. Rhoades, C. B. Burge, and P. B. Bartel. The microRNAs of *Caenorhabditis elegans*. *Genes & Development*, 17:991–1008, 2003.

110. M. Lindow and A. Krogh. Computational evidence for hundreds of non-conserved plant microRNAs. *BMC Genomics*, 6:119, 2005.

111. D. Löffler, K. Brocke-Heidrich, G. Pfeifer, C. Fried, J. Hackermüller, A. Kretzschmar, R. Burger, M. Gramatzki, C. Blumert, K. Bauer, H. Cvijic, P. F. Stadler, and F. Horn. Induction of microRNA-21 contributes to the oncogenic potential of Stat3. *Blood*, 110:1330–1333, 2007.

112. D. Long, R. Lee, P. Williams, C. Y. Chan, V. Ambros, and Y. Ding. Potent effect of target structure on microRNA function. *Nat. Struct. Mol. Bio.*, 14:287–294, 2007.

113. E. Lund, S. Güttinger, A. Calado, J. Dahlberg, and U. Kutay. Nuclear export of microRNA precursors. *Science*, 303:95–98, 2004.

114. I. J. MacRae and J. A. Doudna. Ribonuclease revisited: structural insights into ribonuclease iii family enzymes. *Curr Opin Struct Biol*, 17:138–145, 2007.

115. C. Maher, L. Stein, and D. Ware. Evolution of *Arabidopsis* microRNA families through duplication events. *Genome Res.*, 16:510–519, 2006.

116. B. Mallick, Z. Ghosh, and J. Chakrabarti. MicroRNA switches in *Trypanosoma brucei*. *Biochem Biophys Res Commun.*, 372:459–463, 2008.

117. G. Mathonnet, M. R. Fabian, Y. V. Svitkin, A. Parsyan, L. Huck, T. Murata, S. Biffo, W. C. Merrick, E. Darzynkiewicz, R. S. Pillai, W. Filipowicz, T. F. Duchaine, and N. Sonenberg. MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F. *Science*, 317:1764–1767, 2007.

118. M. Megraw, V. Baev, V. Rusinov, S. Jensen, K. Kalantidis, and A. Hatzigeorgiou. MicroRNA promoter element discovery in *Arabidopsis*. *RNA*, 12:1612–1619, 2006.

119. K. C. Miranda, T. Huynh, Y. Tay, Y. S. Ang, W. L. Tam, A. M. Thomson, B. Lim, and I. Rigoutsos. A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. *Cell*, 126:1203–17, 2006.

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

121. U. Mückstein, H. Tafer, J. Hackermüller, S. H. Bernhart, P. F. Stadler, and I. L. Hofacker. Thermodynamics of RNA-RNA binding. *Bioinformatics*, 22:1177–1182, 2006.

122. D. Murphy, B. Dancis, and J. R. Brown. The evolution of core proteins involved in microRNA biogenesis. *BMC Evol Biol*, 8:92–92, 2008.

123. V. Nair and M. Zavolan. Virus-encoded microRNAs: novel regulators of gene expression. *Trends Microbiol*, 14:169–175, 2006.

124. C. Napoli, C. Lemieux, and R. Jorgensen. Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *Plant Cell*, 2:279–289, 1990.

125. G. Obernosterer, P. J. Leuschner, M. Alenius, and J. Martinez. Post-transcriptional regulation of microRNA expression. *RNA*, 12:1161–1167, 2006.

126. U. Ohler, S. Yekta, L. P. Lim, D. P. Bartel, and C. B. Burge. Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. *RNA*, 10:1309–1322, 2004.

127. K. Okamura, W. J. Chung, J. G. Ruby, H. Guo, D. P. Bartel, and E. C. Lai. The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature*, 453:803–806, 2008.

128. K. Okamura, J. W. Hagen, H. Duan, D. M. Tyler, and E. C. Lai. The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell*, 130:89–100, 2007.

129. K. Okamura, M. D. Phillips, D. M. Tyler, H. Duan, Y. T. Chou, and E. C. Lai. The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution. *Nat Struct Mol Biol*, 15:354–363, 2008.

130. S. E. Ong, B. Blagoev, I. Kratchmarova, D. B. Kristensen, H. Steen, A. Pandey, and M. Mann. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell Proteomics*, 1:376âŁ“386, 2002.

131. M. Otsuka, Q. Jing, P. Georgel, L. New, J. Chen, J. Mols, Y. J. Kang, Z. Jiang, X. Du, R. Cook, S. C. Das, A. K. Pattnaik, B. Beutler, and J. Han. Hypersusceptibility to vesicular stomatitis virus infection in dicer1-deficient mice is due to impaired mir24 and mir93 expression. *Immunity*, 27:123–134, 2007.

132. D. Palakodeti, M. Smielewska, and B. R. Graveley. MicroRNAs from the planarian *Schmidtea mediterranea*: a model system for stem cell biology. *RNA*, 12:1640–1649, 2006.

133. J. S. Parker and D. Barford. Argonaute: A scaffold for the function of short regulatory RNAs. *Trends Biochem Sci*, 31:622–630, 2006.

134. A. E. Pasquinelli, B. J. Reinhart, F. Slack, M. Q. Martindale, M. I. Kurodak, B. Maller, D. C. Hayward, E. E. Ball, B. Degnan, P. Müller, A. Spring, Jürg Srinivasan, M. Fishman, J. Finnerty, J. Corbo, M. Levine, P. Leahy, E. Davidson, and G. Ruvkun. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, 408:86–89, 2000.

135. S. Pfeffer, A. Sewer, M. Lagos-Quintana, R. Sheridan, C. Sander, F. A. Grässer, L. F. van Dyk, C. K. Ho, S. Shuman, M. Chien, J. J. Russo, J. Ju, G. Randall, B. D. Lindenbach, C. M. Rice, V. Simon, D. D. Ho, M. Zavolan, and T. Tuschl. Identification of microRNAs of the herpesvirus family. *Nat Methods*, 2:269–276, 2005.

136. S. Pfeffer, M. Zavolan, F. A. Grässer, M. Chien, J. J. Russo, J. Ju, B. John, A. J. Enright, D. Marks, C. Sander, and T. Tuschl. Identification of virus-encoded microRNAs. *Science*, 304:734–736, 2004.

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

138. J. Piriyapongsa and I. K. Jordan. Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA*, 14:814–821, 2008.

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

140. S. E. Prochnik, D. S. Rokhsar, and A. A. Aboobaker. Evidence for a microRNA expansion in the bilaterian ancestor. *Dev Genes Evol.*, 217:73–77, 2007.

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

142. N. Rajewsky. microRNA target predictions in animals. *Nat. Genet.*, 38:S8–s13, 2006.

143. R. K. Ramachandra, M. Salem, S. Gahr, C. E. Rexroad III, and J. Y. Yao. Cloning and characterization of microRNAs from rainbow trout (*Oncorhynchus mykiss*): their expression during early embryonic development. *BMC Dev. Biol.*, 8:41, 2008.

144. G. Randall, M. Panis, J. D. Cooper, T. L. Tellinghuisen, K. E. Sukhodolets, S. Pfeffer, M. Landthaler, P. Landgraf, S. Kan, B. D. Lindenbach, M. Chien, D. B. Weir, J. J. Russo, J. Ju, M. J. Brownstein, R. Sheridan, C. Sander, M. Zavolan, T. Tuschl, and C. M. Rice. Cellular cofactors affecting hepatitis c virus infection and replication. *Proc Natl Acad Sci U S A*, 104:12884–12889, 2007.

145. M. Rehmsmeier, P. Steffen, M. Höchsmann, and R. Giegerich. Fast and effective prediction of microRNA/target duplexes. *RNA*, 10:1507–1517, 2004.

146. B. Reinhart, E. Weinstein, M. Rhoades, B. Bartel, and D. Bartel. MicroRNAs in plants. *Genes Dev.*, 16:1616–1626, 2002.

147. B. J. Reinhart, F. J. Slack, M. Basson, A. E. Pasquinelli, J. C. Bettinger, A. E. Rougvie, H. R. Horvitz, and G. Ruvkun. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, 403:901–906, 2000.

148. H. Robins, Y. Li, and R. W. Padgett. Incorporating structure to predict microRNA targets. *Proc Natl Acad Sci U S A*, 102:4006–9, 2005.

149. J. G. Ruby, C. Jan, C. Player, M. J. Axtell, W. Lee, C. Nusbaum, H. Ge, and D. P. Bartel. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell*, 127:1193–1207, 2006.

150. P. Sarnow, C. L. Jopling, K. L. Norman, S. Schütz, and K. A. Wehner. MicroRNAs: expression, avoidance and subversion by vertebrate viruses. *Nat Rev Microbiol*, 4:651–659, 2006.

151. M. A. Saunders, H. Liang, and W. H. Li. Human polymorphism at microRNAs and microRNA target sites. *Proc Natl Acad Sci U S A*, 104:3300–3305, 2007.

152. R. Schwab, J. Palatnik, M. Riester, C. Schommer, M. Schmid, and D. Weigel. Specific effects of microRNAs on the plant transcriptome. *Dev. Cell*, 8:517–527, 2005.

153. D. Schwarz, G. Hutvágner, T. Du, Z. Xu, N. Aronin, and P. Zamore. Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, 115:199–208, 2003.

154. H. Seitz, H. Royo, M. Bortolin, S. Lin, A. Ferguson-Smith, and J. Cavaillé. A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. *Genome Res.*, 14:1741–1748, 2004.

155. M. Selbach, B. Schwanhäusser, N. Thierfelder, Z. Fang, R. Khanin, and N. Rajewsky. Widespread changes in protein synthesis induced by microrna. *Nature*, 2008. epub ahead of print.

156. L. F. Sempere, C. N. Cole, M. A. McPeek, and K. J. Peterson. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *J Exp Zoolog B Mol Dev Evol.*, 306:575–588, 2006.

157. P. Sethupathy, B. Corda, and A. G. Hatzigeorgiou. Tarbase: A comprehensive database of experimentally supported animal microRNA targets. *RNA*, 12:192–7, 2006.

158. A. Sewer, N. Paul, P. Landgraf, A. Aravin, S. Pfeffer, M. J. Brownstein, T. Tuschl, E. van Nimwegen, and M. Zavolan. Identification of clustered microRNAs using an *ab initio* prediction method. *BMC Bioinformatics*, 6:267 [epub], 2005.

159. C. Simón-Mateo and J. García. MicroRNA-guided processing impairs Plum pox virus replication, but the virus readily evolves to escape this silencing mechanism. *J. Virol.*, 80:2429–2436, 2006.

160. N. R. Smalheiser and V. I. Torvik. Mammalian microRNAs derived from genomic repeats. *Trends Genet.*, 21:322–326, 2005.

161. J. Spring. Genome duplication strikes back. *Nat. Genet.*, 31:128–129, 2002.

162. A. Stark, J. Brennecke, N. Bushati, R. B. Russell, and S. M. Cohen. Animal microRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell*, 123:1133–1146, 2005.

163. A. Stark, J. Brennecke, R. B. Russell, and S. M. Cohen. Identification of *Drosophila* microRNA targets. *PLoS Biol*, 1:e60, 2003.

164. P. Stein, P. Svoboda, M. Anger, and R. M. Schultz. RNAi: mammalian oocytes do it without RNA-dependent RNA polymerase. *RNA*, 9:187–192, 2003.

165. C. S. Sullivan and D. Ganem. MicroRNAs and viral infection. *Mol Cell*, 20:3–7, 2005.

166. C. S. Sullivan, A. T. Grundhoff, S. Tevethia, J. M. Pipas, and D. Ganem. Sv40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic t cells. *Nature*, 435:682–686, 2005.

167. R. Sunkar and G. Jagadeeswaran. *In silico* identification of conserved microRNAs in large number of diverse plant species. *BMC Plant Biol.*, 8:37, 2008.

168. R. Sunkar, X. Zhou, Y. Zheng, W. Zhang, and J. K. Zhu. Identification of novel and candidate miRNAs in rice by high throughput sequencing. *BMC Plant Biol.*, 8:25, 2008.

169. M. Talmor-Neiman, R. Stav, W. Frank, B. Voss, and T. Arazi. Novel micro-RNAs and intermediates of micro-RNA biogenesis from moss. *Plant J.*, 47:25–37, 2006.

170. O. H. Tam, A. A. Aravin, P. Stein, A. Girard, E. P. Murchison, S. Cheloufi, E. Hodges, M. Anger, R. Sachidanandam, R. M. Schultz, and G. J. Hannon. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature*, 453:534–538, 2008.

171. G. Q. Tang and E. S. Maxwell. *Xenopus* microRNA genes are predominantly located within introns and are differentially expressed in adult frog tissues via post-transcriptional regulation. *Genome Res.*, 18:104–112, 2008.

172. A. Tanzer, C. T. Amemiya, C.-B. Kim, and P. F. Stadler. Evolution of microRNAs located within *Hox* gene clusters. *J. Exp. Zool.: Mol. Dev. Evol.*, 304B:75–85, 2005.

173. A. Tanzer and P. F. Stadler. Molecular evolution of a microRNA cluster. *J. Mol. Biol.*, 339:327–335, 2004.

174. A. Tanzer and P. F. Stadler. Evolution of microRNAs. In S. Y. Ying, editor, *MicroRNA Protocols*, volume 342 of *Methods in Molecular Biology*, pages 335–350. Humana Press, Totowa, NJ, 2006.

175. The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447:799–816, 2007.

176. Y. Tomari, T. Du, and P. D. Zamore. Sorting of *Drosophila* small silencing RNAs. *Cell*, 130:299–308, 2007.

177. Y. Tomari, C. Matranga, B. Haley, N. Martinez, and P. D. Zamore. A protein sensor for siRNA asymmetry. *Science*, 306:1377–1380, 2004.

178. R. Triboulet, B. Mari, Y. L. Lin, C. Chable-Bessia, Y. Bennasser, K. Lebrigand, B. Cardinaud, T. Maurin, P. Barbry, V. Baillat, J. Reynes, P. Corbeau, K. T. Jeang, and M. Benkirane. Suppression of microRNA-silencing pathway by HIV-1 during virus replication. *Science*, 315:1579–1582, 2007.

179. E. Ullu, C. Tschudi, and T. Chakraborty. RNA interference in protozoan parasites. *Cell Microbiol.*, 6:509–519, 2004.

180. V. Vagin, A. Sigova, C. Li, H. Seitz, V. Gvozdev, and P. D. Zamore. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*, 313:320–324, 2006.

181. A. R. van der Krol, L. A. Mur, M. Beld, J. N. Mol, and A. R. Stuitje. Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell*, 2:291–299, 1990.

182. H. Vaucheret, F. Vazquez, P. Crété, and D. Bartel. The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes Dev.*, 18:1187–1197, 2004.

183. F. Vazquez. *Arabidopsis* endogenous small RNAs: highways and byways. *Trends Plant Sci*, 11:460–468, 2006.

184. M. C. Vella, K. Reinert, and F. J. Slack. Architecture of a validated MicroRNA::Target interaction. *Chem Biol*, 11:1619–1623, 2004.

185. X. Wang, J. Gu, M. Q. Zhang, and Y. Li. Identification of phylogenetically conserved microRNA cis-regulatory elements across 12 *Drosophila* species. *Bioinformatics*, 24:165–171, 2008.

186. X. Wang, J. Zhang, F. Li, J. Gu, X. He, T. Zhang, and Y. Li. MicroRNA identification based on sequence and structure alignment. *Bioinformatics*, 21:3610–3614, 2005.

187. Y. Wang, T. Hindemitt, and K. F. Mayer. Significant sequence similarities in promoters and precursors of *Arabidopsis thaliana* non-conserved microRNAs. *Bioinformatics*, 22:2585–2589, 2006.

188. N. Warthmann, S. Das, C. Lanz, and D. Weigel. Comparative analysis of the MIR319a microRNA locus in *Arabidopsis* and related Brassicaceae. *Mol. Biol. Evol.*, 25:892–902, 2008.

189. T. Watanabe, Y. Totoki, A. Toyoda, M. Kaneda, S. Kuramochi-Miyagawa, Y. Obata, H. Chiba, Y. Kohara, T. Kono, T. Nakano, M. A. Surani, Y. Sakaki, and H. Sasaki. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*, 453:539–543, 2008.

190. M. J. Weber. New human and mouse microRNA genes found by homology search. *FEBS J.*, 272:59–73, 2005.

191. M. R. Willmann and R. S. Poethig. Conservation and evolution of miRNA regulatory programs in plant development. *Curr Opin Plant Biol.*, 10:503–511, 2007.

192. X. Xie, J. Lu, E. J. Kulbokas, T. R. Golub, V. Mootha, K. Lindblad-Toh, E. S. Lander, and M. Kellis. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature*, 434:338–345, 2005.

193. Z. Xie, E. Allen, N. Fahlgren, A. Calamar, S. Givan, and J. Carrington. Expression of *Arabidopsis* MIRNA genes. *Plant Physiol.*, 138:2145–2154, 2005.

194. Z. Xie, K. Kasschau, and J. Carrington. Negative feedback regulation of Dicer-Like1 in *Arabidopsis* by microRNA-guided mRNA degradation. *Curr. Biol.*, 13:784–789, 2003.

195. Z. Xie and X. Qi. Diverse small RNA-directed silencing pathways in plants. *Biochim Biophys Acta*, 2008. doi:10.1016/j.bbagrm.2008.02.009.

196. N. Xu, B. Segerman, X. Zhou, and G. Akusjärvi. Adenovirus virus-associated RNAII-derived small RNAs are efficiently incorporated into the rna-induced silencing complex and associate with polyribosomes. *J Virol*, 81:10540–10549, 2007.

197. Y. Xu, X. Zhou, and W. Zhang. MicroRNA prediction with a novel ranking algorithm based on random walks. *Bioinformatics*, 24:i50–i58, 2008.

198. C. Xue, F. Li, T. He, G. Liu, Y. Li, and X. Zhang. Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. *BMC Bioinformatics*, 6:310 [epub], 2005.

199. H. YY, Y. XM, and L. WJ. Computational approaches to microRNA discovery. *Yi Chuan*, 30(6):687–96, 2008. Article in Chinese. PubMed - in process.

200. B. Zhang, X. Pan, C. H. Cannon, G. P. Cobb, and T. A. Anderson. Conservation and divergence of plant microRNA genes. *Plant J.*, 46:243–259, 2006.

201. Y. Zhang. miRU: an automated plant miRNA target prediction server. *Nucleic Acids Res.*, 33:W701–704, 2005.

202. T. Zhao, G. Li, S. Mi, S. Li, G. J. Hannon, X. J. Wang, and Y. Qi. A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. *Genes Dev.*, 21:1190–1203, 2007.

203. H. Zhou and K. Lin. Excess of microRNAs in large and very 5' biased introns. *Biochem Biophys Res Commun*, 368:709–715, 2008.

204. X. Zhou, J. Ruan, G. Wang, and W. Zhang. Characterization and identification of microRNA core promoters in four model species. *PLoS Comput. Biol.*, 3:e37, 2007.

205. Y. Zhou, J. Ferguson, J. T. Chang, and Y. Kluger. Inter- and intra-combinatorial regulation by transcription factors and microRNAs. *BMC Genomics*, 8:396, 2007.

206. Q. H. Zhu, A. Spriggs, L. Matthew, L. Fan, G. Kennedy, F. Gubler, and C. Helliwell. A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. *Genome Res*, 2008. epub ahead of print.