The evolution of gene regulation by transcription factors and microRNAs

Kevin Chen and Nikolaus Rajewsky*‡*

Abstract | Changes in the patterns of gene expression are widely believed to underlie many of the phenotypic differences within and between species. Although much emphasis has been placed on changes in transcriptional regulation, gene expression is regulated at many levels, all of which must ultimately be studied together to obtain a complete picture of the evolution of gene expression. Here we compare the evolution of transcriptional regulation and post-transcriptional regulation that is mediated by microRNAs, a large class of small, non-coding RNAs in plants and animals, focusing on the evolution of the individual regulators and their binding sites. As an initial step towards integrating these mechanisms into a unified framework, we propose a simple model that describes the transcriptional regulation of new microRNA genes.

The emergence of complex, multicellular organisms was accompanied, and perhaps facilitated, by dramatic increases in the complexity of gene regulatory mechanisms[1,2]. At the level of transcriptional regulation, this can be clearly seen in the massive expansions of transcription-factor families and the pervasive combinatorial control of genes by multiple transcription factors in higher organisms[1,3] (BOX 1). At the level of post-transcriptional control, entirely new mechanisms of gene regulation arose, typified by a large and growing class of ~22-nucleotide-long non-coding RNAs, known as microRNAs (miRNAs), which function as repressors in all known animal and plant genomes[4,5] (BOX 1). Although transcription factors and miRNAs are two of the best-studied gene regulatory mechanisms, there are many other layers of gene regulation, including: cell signalling; mRNA splicing, polyadenylation and localization; chromatin modifications; and mechanisms of protein localization, modification and degradation (FIG. 1).

But why have higher plants and animals evolved such complex, multilayered gene regulatory systems? Is combinatorial transcriptional regulation alone insufficient to specify a developmental programme? What are the relative contributions of the various mechanisms of gene regulation to changes at the phenotypic level? Do all the different modes of gene regulation evolve in the same manner and at the same rate? Despite half a century of research on gene regulation, such questions have yet to be tackled seriously, because most of the effort in the field so far has been devoted to studying the evolution of transcriptional regulation. Although the primacy of transcription as the necessary first step in gene expression is undeniable, this does not imply that transcriptional regulation has the largest effect on the final concentration of the active gene product, which is the most relevant quantity to the phenotype.

An important goal for future research is to elucidate how complex gene regulatory networks evolve and how their evolution results in phenotypic change and speciation. However, a necessary first step towards this goal is to understand the basic principles that underlie the evolution of the individual regulators and their regulatory interactions with their target genes. Recent computational and experimental work has made it possible to begin to study the evolution of transcription factors, miRNAs and their binding sites, and to compare the rate and manner in which these two important regulatory mechanisms evolve. Here we mainly focus on animal evolution because most of the work on gene regulatory evolution has been carried out in animal systems, although we refer to plant evolution whenever possible.

Ultimately, a complete picture of the evolution of gene regulation will require a synthesis of information about all the diverse components of gene regulatory networks. As an initial step, we propose a simple model in which the evolution of transcriptional control of animal miRNA genes themselves is an important step in the successful acquisition of a novel miRNA. A corollary of this model is that there are miRNAs that are transcribed at low levels and in specific cell types, which might have little biological function in regulating target genes in *trans*.

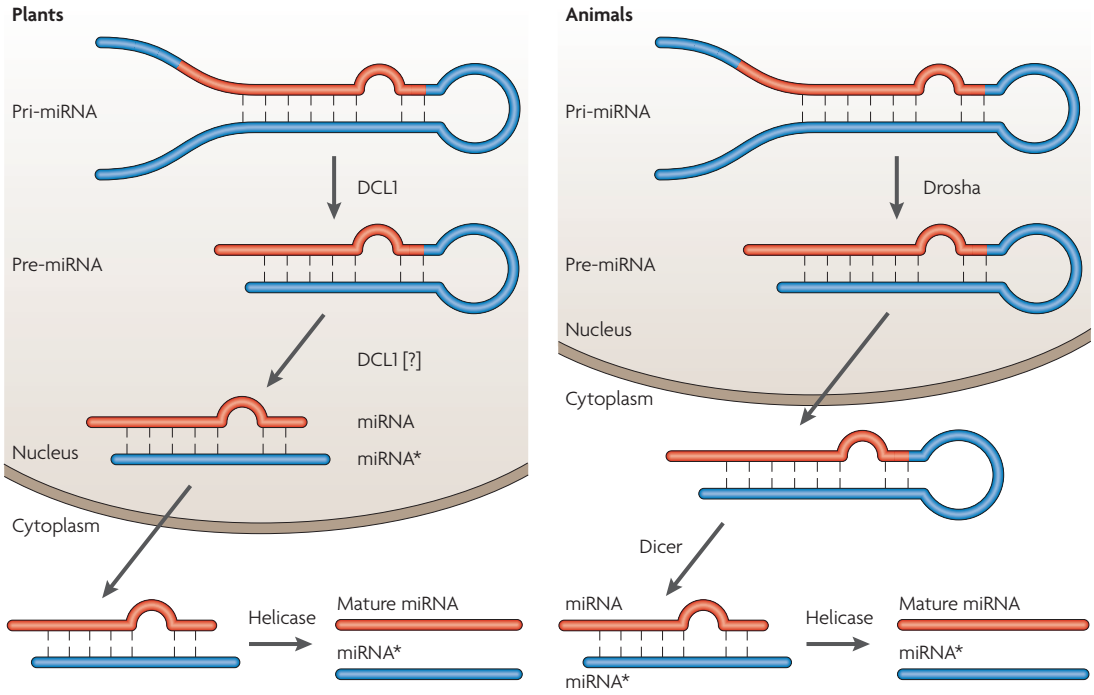
**Center for Comparative Functional Genomics, Department of Biology, New York University, New York, New York 10003, USA.*
‡Max Delbrück Centrum für Molekulare Medizin, Robert-Rössle-Strasse 10, Berlin-Buch 13125, Germany.
Correspondence to N.R.
e-mail:
rajewsky@mdc-berlin.de
doi:10.1038/nrg1990

Transcription factors and miRNAs in development

No discussion of the evolution of gene regulation is complete without a consideration of the developmental roles of the regulators themselves, because these roles imply certain constraints on the evolvability of the regulatory relationships[6,7]. Furthermore, if we accept that development is the consequence of the unfolding of precise and robust spatio-temporal patterns of gene expression, then it is in the context of development that the evolution of gene regulation is most closely related to the evolution of organismal form[3,7–9].

In complex multicellular organisms, transcription factors generally do not work in isolation, but instead, together with co-regulators, they form large networks of cooperating and interacting transcription factors[3,8]. It is widely believed that the rate of evolution of regulatory relationships is not homogeneous over the entire network. For example, in one model that was proposed by Davidson and colleagues, animal developmental networks can be decomposed into subnetworks, including: highly conserved 'kernels' that specify the spatial domain in which a particular body part will develop; 'plug-in'

Box 1 | **Transcription factors and microRNA genes — an overview**



Transcription factors are proteins that either activate or repress transcription of other genes by binding to short *cis*-regulatory elements called transcription-factor binding sites that lie in the vicinity of the target genes. Transcription-factor binding sites, especially in developmental regulatory networks, are often organized into clusters called *cis*-regulatory modules (CRMs)[3], which typically span a few hundred nucleotides and can contain dozens of binding sites for ~3–10 transcription factors. CRMs produce the initial spatio-temporal expression pattern of the target gene by 'reading out' the concentrations of multiple transcription factors that are present in a particular cell at a particular time. So, dependence on cellular context and combinatorial control are common themes in transcriptional regulation. Transcription factors are usually grouped into families on the basis of shared DNA-binding domains, which are an important determinant of transcription-factor binding specificity.

Mature microRNAs (miRNAs) are short (~22-nucleotide), non-coding ssRNAs that repress mRNAs post-transcriptionally by binding to partially complementary sites, called miRNA binding sites, in their target mRNAs. In animals, miRNA-mediated repression is often relatively weak, whereas transcription-factor-mediated repression can be much stronger. Mature miRNAs are cleaved from ~70-nucleotide hairpin structures, called precursor miRNAs (pre-mRNAs), by the enzyme dicer. Pre-miRNAs are in turn excised from a primary miRNA (pri-mRNA) transcript by the enzyme drosha. Pri-miRNAs are typically transcribed by RNA polymerase II[104] and seem to possess promoter and enhancer elements that are similar to those of protein-coding genes (for example, REFS 105,106). They can be thousands of nucleotides long and contain multiple pre-miRNAs. In some cases, however, pre-miRNAs are contained in introns of protein-coding genes and are excised by the splicing machinery. In metazoans, pre-mRNAs are exported into the cytoplasm where they are processed into mature miRNAs, whereas in plants, miRNA maturation occurs within the nucleus. miRNAs are grouped into families on the basis of their target recognition motifs (BOX 3).

The predominant regulatory effect of miRNAs is to repress their target mRNAs; mechanisms for this include translational repression, mRNA cleavage, mRNA deadenylation or alteration of mRNA stability (reviewed in REF. 107,126). miRNA-mediated cleavage of mRNAs seems to be the exception in animals. By contrast, it is believed to be the dominant regulatory mode in plants. DCL1, dicer-like 1; miRNA*, the rapidly decaying strand that is complementary to the mature miRNA.

components, such as signal transduction cassettes, which are re-used in multiple developmental contexts; terminal differentiation gene batteries that consist of all the genes that define a particular cell type; and 'input–output switches', such as Hox genes, which allow or disallow the action of particular developmental processes in specific spatio-temporal contexts[3,7]. The authors proposed that each of these types of subnetwork has its own set of evolutionary constraints, and that changes in different subnetworks result in qualitatively different types of change at the phenotypic level. For example, kernels are by definition highly conserved, and their consistency over long periods of evolutionary time might provide an explanation for the high degree of conservation of body plans within animal phyla. By contrast, differentiation gene sets, being the least pleiotropic of all the regulatory relationships, are easiest to alter, and such changes might result in species-level phenotypic differences (for details, see REFS 3,7).

Various studies have demonstrated that miRNAs have important roles in animal and plant development (see REFS 10–14 for a consideration of the developmental roles of miRNAs). Some well known examples include miRNAs with switch-like roles, such as *lin-4* and *let-7* in *Caenorhabditis elegans* developmental timing[15,16] or the miRNAs that are involved in plant leaf or flower development (reviewed in REF. 13), and miRNAs that confer more general tissue or temporal identity, such as *miR-1* in *Drosophila melanogaster* muscle development[17] and *miR-430* in the zebrafish maternal–zygotic transition[18].

Much of the current evidence for an early developmental role for miRNAs is conflicting or difficult to interpret[11]. First, zebrafish embryos that lack maternal dicer, a protein that is required for miRNA biogenesis (BOX 2), progress through axis formation and regionalization[19], a fact that strongly argues against a role for miRNAs in early zebrafish development. But dicer knockout mice[20] and *Arabidopsis thaliana* that carry hypomorphic alleles of the dicer homolog *DCL-1* die in early embryogenesis[21]. Second, although miRNAs have not been detected in early zebrafish and medaka embryos[22,23], mature miRNAs have been detected in mice[24] and *D. melanogaster* embryos[25]. Primary miRNA transcripts are spatially regulated in early *D. melanogaster* embryogenesis[26], with the caveat that the processing of primary miRNAs into mature miRNAs (BOX 2) can be regulated[24]. Third, on the basis of miRNA knockdowns using 2′-*O*-methyl antisense oligoribonucleotides, it has been reported that miRNAs are involved in patterning the *D. melanogaster* embryo[27], although a number of these results disagree with experimental data from genetic knockouts[28].

The prevailing opinion (for example, REFS 3,29 and the references therein) seems to be that miRNAs as a class tend to function as lock-down mechanisms for already-differentiated states, or confer an additional layer of robustness or 'noise' reduction[30,31] on the developmental processes, rather than having fundamental roles in body-plan patterning. However, because the functions of only a few miRNAs have been dissected in detail, we believe that this point has not yet been proven unambiguously, at least not as a general principle.

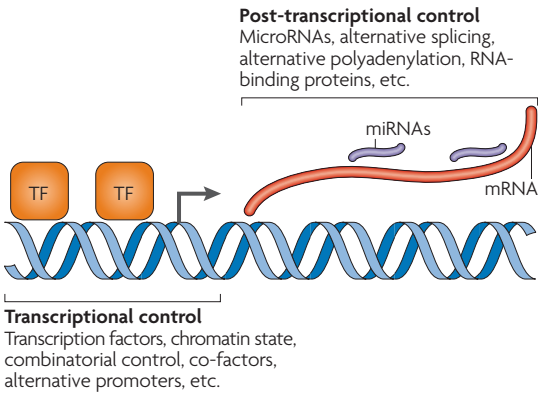


Figure 1 | **Gene regulation by transcription factors and microRNAs.** One or more transcription factors activate transcription by binding to *cis*-regulatory sites, which are often, although not always, situated upstream of protein-coding genes. After transcription, one or more microRNAs bind to *cis*-regulatory sites, usually in the 3′ UTR of the mRNA, and repress protein translation. In addition to these two mechanisms, many more gene regulatory mechanisms work at either the transcriptional or post-transcriptional level, including: cell signalling; mRNA splicing, polyadenylation and localization; chromatin modifications; and mechanisms of protein localization, modification and degradation (not shown). TF, transcription factor.

trans-factor evolution

Independent evolution of transcription factors and miRNAs in the plant and animal kingdoms. It is widely believed that the last common ancestor of plants and animals was unicellular, and therefore that animal and plant development evolved independently (for example, REF. 32). Both animal and plant development depends on 'master' transcription-factor regulators, perhaps most famously the MADS-box proteins in plants and homeodomain proteins in animals[32]. Both of these transcription-factor families, and most of the other transcription-factor families that are found in plants or animals, predate the divergence of the two kingdoms. However, at the sequence level, these ancient transcription-factor families are typically poorly conserved beyond the DNA-binding domain that defines the family. More importantly, their developmental functions seem to be different. For example, MADS-box proteins do not seem to have the same fundamental roles in animals as they do in plants; the opposite is true for homeodomain proteins. Furthermore, a substantial proportion of transcription-factor families are in fact kingdom-specific (for example, >45% in *C. elegans* and *A. thaliana*)[33]. So, despite the fact that many ancient transcription-factor families predate the divergence of plants and animals, the overall picture of plant and animal transcription-factor evolution involves the acquisition of novel transcription-factor families and the diversification of existing families.

Similarly, it is generally accepted that miRNAs have evolved independently in the animal and plant kingdoms, because there are no known homologous miRNAs between plants and animals (but see REF. 127 for some

Box 2 | MicroRNA gene discovery: bioinformatics and experimental methods

The two general approaches to microRNA (miRNA) gene discovery are bioinformatics and experimental methods (reviewed in REFS 108,109). Generally, bioinformatics methods use RNA-folding algorithms (for example, REFS 110,111) to search for approximate hairpin structures in non-coding and non-repetitive regions of the genome, and filter them using patterns of evolutionary conservation. Known examples of miRNA precursors are used as training examples for machine learning algorithms to discriminate between true predictions and false positives (reviewed in REFS 108,109). Predictions are generally verified by northern blots, PCR or microarray analysis. Naturally, bioinformatic predictions have false-positive rates and can miss species-specific miRNAs, as many, but not all[112], of the current methods use evolutionary conservation as an indicator of biological function.

The traditional experimental approach to miRNA discovery is cloning and sequencing (for example, REF. 113). This approach successfully detects species-specific miRNAs, but tends to miss miRNAs that are expressed at low levels, in a small number of cells or only under particular cellular conditions. More recently, high-throughput sequencing methods, especially 454 sequencing[114], have become popular for surveying small RNA populations (for example, REF. 115).

Currently, 328 miRNAs have been annotated in the human genome and 199 in *Arabidopsis thaliana*, the metazoan and plant species in which their small RNA complements have been well surveyed[116]. Several groups have shown that metazoan miRNAs probably regulate thousands of genes in mammals[41,117,118], flies[62,119], and nematodes[120]. For example, in humans, known miRNAs make up >1% of the gene repertoire and are thought to regulate >30% of all protein-coding genes.

Both computational and sequencing approaches indicate that there are likely to be many more miRNAs, many of which are lineage-specific (for example, REFS 86,92).

evidence to the contrary) and miRNA biogenesis and the mechanism of miRNA-mediated gene repression are significantly different between plants and animals (BOX 1). Further evidence for this comes from the phylogenetic distribution of miRNAs, specifically the apparent lack of miRNAs in sponges and fungi, as discussed below. To a first approximation then, miRNA genes and most transcription-factor families have evolved independently in the animal and plant kingdoms.

Deep conservation of transcription factors and miRNAs. It is well known that many transcription factors are highly conserved over large evolutionary distances, and some have similar developmental roles in diverse species. Hox genes, which regulate development along the anterior–posterior axis in most animals, are the textbook example of this phenomenon. Other examples in vertebrates and many invertebrates include the paired box 6 (Pax6) genes, which direct eye and anterior nervous system development, and Csx/Nkx2-5/Tinman genes, which direct visceral mesoderm and heart development (reviewed in REF. 8). Although such broad similarities in function are intriguing, it is worth noting that they do not necessarily imply complete functional redundancy between distant homologues, or even evolutionarily conserved developmental roles[34].

Likewise, many miRNAs seem to be extremely well conserved. The best known example is *let-7*, which is phylogenetically conserved in all bilaterians that have been tested so far, with the single exception of acoel flatworms[35,36]. Furthermore, the exact sequence of the mature form of *let-7*, its temporal expression pattern and in some cases its syntenic position in the genome are conserved[35,36]. Some other well known examples are muscle-specific *miR-1*, which is conserved in nematodes, mammals and flies (REF. 17 and the references therein) and *miR-7*, the mature form of which is perfectly conserved between mammals and flies, and which lies in an intron of the same host gene in both clades.

Several groups have used bioinformatic sequence comparisons or northern blots to study the conservation of miRNAs across many animal species[37–39]. They found that 18–30 miRNA families seem to be conserved in all bilaterians that have been studied, depending on the stringency of the bioinformatic methods used. Three of these miRNA families were also found in cnidarians, but none was found in sponges[38,39]. As previously discussed, this supports the idea of independent evolution of miRNAs in plants and animals. One of the cnidarian miRNAs, *miR-10*, has been detected by northern blots, and is particularly interesting because it is present in a Hox cluster and regulates Hox genes in *Drosophila* and vertebrates[38,40]. Overall, the level of sequence conservation of many miRNAs is generally high, although sequence conservation need not imply functional conservation[23].

Further evidence for deep conservation of animal miRNA genes at a genome-wide level can be found in cross-clade comparisons of highly conserved motifs (HCMs) in 3′ UTRs. Xie *et al.* showed that HCMs in vertebrates are highly enriched in miRNA-binding sites[41]. We and others have extended these results to nematodes and flies[42,43], and have shown that HCMs are highly conserved between vertebrates, nematodes and flies (Supplementary information S1 (figure)). Because many HCMs are expected to represent miRNA recognition motifs, these analyses provide indirect evidence for deep conservation of miRNA genes.

The sampling of miRNAs in plants is less extensive than in animals, with most of the plant miRNAs that are known so far having been discovered in *A. thaliana* (for example, REF. 44). Nevertheless, several examples of deeply conserved plant miRNAs are known from sequencing of miRNAs in moss[45–47]. Some of these examples are particularly impressive because their regulatory relationships with their targets also seem to be conserved, a point we return to below.

Lineage-specific expansions of transcription factors and miRNAs. Lineage-specific expansions of transcription-factor families are common and are widely believed to have an important role in both plant and animal diversification and complexity[1,3,9]. Interestingly, there is evidence that the expansion of transcription-factor families is greater in plants than in animals[48–50]. In principle, this could reflect a fundamental difference between plant and animal biology. However, an alternative explanation is that *A. thaliana* has undergone a recent whole-genome duplication, whereas the putative whole-genome duplications in animal lineages are more ancient: immediately after a large-scale duplication event, a transcription-factor loss might be deleterious if it changes the relative concentrations of the set of transcription factors that are expressed in a cell[51], leading to increased rates of transcription-factor retention in plants compared with animals.

Bilaterians
Members of the animal kingdom that have bilateral symmetry — the property of having two similar sides, with definite upper and lower surfaces, and anterior and posterior ends.

Acoel flatworms
A basal bilaterian clade that diverged from the rest of bilaterians before the split between protostomes and deuterostomes.

Synteny
Collinearity in the order of genes or other DNA sequences in chromosomal regions of two species or in the same species.

Clade
A group of organisms that includes a common ancestor and all of its descendants, representing a distinct branch on a phylogenetic tree.

Cnidarians
Radially symmetrical animals that have sac-like bodies with only one opening. They include jellyfish, corals, hydra and anemones.

Several detailed studies in plants[52] and animals[53,54] have shown that miRNA families can expand by the same processes of tandem, segmental and whole-genome duplication as protein-coding genes. Plant miRNA families tend to be larger and their members more similar to each other than animal miRNA families (reviewed in REF. 55). This indicates that the expansion of plant families is more recent, and that the main effect of having multiple paralogous copies of the same miRNA in plants is to increase dosage. By contrast, family members of animal miRNAs might tend to have synergistic but functionally distinct roles, as shown for the *let-7* family in *C. elegans*[56,57].

cis-element evolution

Large-scale rewiring of miRNA-mediated regulatory relationships over large evolutionary distances in animals, but deep conservation in plants. Although miRNAs themselves seem to be highly conserved, there are only a few miRNA-target regulatory relationships that are known to be conserved over large evolutionary distances in animals (for example, between vertebrates and *Drosophila* species). There are two cases, namely *let-7*:*lin-41* and *let-7*:*let-60*–RAS, in which the target relationship has been experimentally dissected in *C. elegans,* and for which there is evidence for conservation in mammals that was obtained from computational target predictions and reciprocal expression patterns of the miRNA and mRNA in cell lines[58,59]. Another miRNA-target relationship, *lin-4*:*lin-28*, has been established experimentally in *C. elegans*, and there is computational and gene-expression evidence that it is conserved in a number of other animal species, including mammals[60,61]. However, even in these cases, caution in interpretation is warranted, as the existence of the same regulatory relationship in highly diverged species does not always imply that it is evolutionarily conserved as opposed to independently evolved. For example, the well known Pax6 transcription factor is involved in mammalian and *Drosophila* eye development, and ectopic expression of the mouse homologue in *Drosophila* can induce the development of ectopic *Drosophila* eyes, but it is not clear whether even this regulatory relationship represents conservation or convergent evolution[3].

These isolated examples aside, the conservation of computationally predicted miRNA targets between vertebrates, *Drosophila* and nematodes seems to be close to what is expected by chance[28,42,43,62]. These results seem to be robust, even when accounting for the false-positive and false-negative rates of the target prediction algorithms and possible errors in assignments of homologous genes and miRNAs[42]. Together with the high level of conservation that is observed for the miRNA genes themselves, this indicates that miRNA regulatory networks have undergone extensive rewiring during animal evolution.

In plants, a number of miRNA regulatory relationships are known to be conserved between *A. thaliana* and moss[46,47], which diverged over 400 million years ago. These relationships are supported by strong experimental evidence, with verification by 5′-RACE (rapid amplification of cloned ends) of the cleavage products of the target mRNA (BOX 2). Because many of these regulatory relationships are involved in crucial biological processes (for example, auxin signalling) this indicates that at least some miRNAs have held central positions within plant developmental regulatory networks for a long time.

Because transcription-factor binding sites are difficult to predict computationally, it is harder to decipher the global dynamics of transcription-factor binding site turnover over large evolutionary distances using computational methods. However, the recent emergence of large-scale experimentally defined transcription-factor binding-site data (reviewed in REF. 63), particularly from ChIP-chip analysis, might make this problem tractable in the near future. Towards this end, an interesting study that compared the targets of the RNA-binding protein Pumilio in *D. melanogaster* and yeast[64] found that, although the binding affinity of the regulator had remained virtually the same in these two species, its targets had diverged almost completely. Therefore, the high conservation of *trans*-acting factors and low overall conservation of *cis*-regulatory sites might be common to many regulatory mechanisms.

High turnover of binding sites even over short evolutionary distances. A number of groups have studied the turnover of experimentally verified transcription-factor binding sites between humans and mice[65], and between various *Drosophila* species[66–69]. The general conclusion is that sequence conservation for known binding sites is surprisingly low (for example, ~50% for *D. melanogaster* and *Drosophila pseudoobscura*[67]), although this picture is complicated by the fact that selection is more likely to work at the level of an entire *cis*-regulatory module than on an individual binding site (BOX 1). In particular, *cis*-regulatory modules often contain redundant binding sites and, as elegantly demonstrated by Ludwig *et al.*[70,71], there can be compensatory mutations that maintain the function of the enhancer despite the loss of individual binding sites.

Similar studies have not been carried out for miRNAs due to the paucity of experimentally verified miRNA binding sites. However, several recent microarray-based studies have indicated that the rate of binding-site conservation is also around 50% (between humans and mice or between zebrafish, *Tetraodon* and *Fugu*)[18,72,73]. Furthermore, computational miRNA target predictions indicate that many lineage-specific miRNA binding sites exist in *Drosophila* and vertebrates (N.R., unpublished observations). Finally, we have used human SNP data and population-genetics techniques to show that 30–50% of non-conserved miRNA binding sites in the human genome might be functional when the mRNA and miRNA are expressed in the same tissue[74]. It should be noted that terms such as 'non-conserved binding sites' generally refer to cases in which a binding site cannot be aligned to its homologous sequence. Because these alignments often suffer from technical problems (for example, almost all alignment algorithms are unable to deal with genomic rearrangements on various scales), 'non-conserved' sites might actually turn out to be conserved in a strict evolutionary sense. Nevertheless, these studies together indicate that both miRNA and transcription-factor binding sites are gained and lost quickly over short evolutionary distances.

ChIP-chip analysis
A method that combines chromatin immunoprecipitation with microarray technology to identify *in vivo* targets of a transcription factor.

Taking these observations to the species level, it has been shown that human promoter regions harbour a surprisingly high amount of variation that significantly affects expression levels, presumably by interfering with transcriptional control (reviewed in REF. 75). Examples of functional polymorphism have also been identified recently in miRNA binding sites in humans[76] and sheep[77], and sets of SNPs in predicted miRNA binding sites in the human and mouse genomes have been collected by several groups[74,77]. By contrast, resequencing of miRNA genes in humans showed almost no polymorphism in the sequences of mature miRNA genes[78], consistent with the higher levels of constraint on *trans*-acting regulators compared with *cis*-regulatory sites.

Rates of evolution

It has been suggested that repressors should evolve faster than activators, as there are many ways to repress a gene but relatively few ways to activate it[79]. As transcription factors can function as activators or repressors, but all known miRNAs work as repressors, one might expect miRNA binding sites to evolve faster than transcription-factor binding sites.

However, a more fundamental difference between transcription factors and miRNAs is that transcription-factor binding sites are typically 'fuzzy' (that is, the same transcription factor can bind to many similar DNA sequences, possibly with different binding affinities), whereas many miRNA binding sites exhibit almost exact Watson–Crick complementarity, either to the first 6–8 bases from the 5′ end of the mature miRNA in animals, or to the entire mature miRNA in plants (BOX 3). Therefore, under neutral evolution, one would expect that it is more difficult to destroy a functional transcription-factor binding site than to create a new one, whereas the converse would be expected for miRNA binding sites.

Plant and animal miRNA binding sites might also evolve at different rates. Plant miRNA binding sites are typically found in coding regions, and if we assume that non-synonymous sites are highly constrained and synonymous sites are neutrally evolving, then approximately one-third of the ~22 bases in a plant miRNA binding site can accommodate a substitution without highly deleterious consequences for the organism. Therefore, the sizes of the mutational target for plant and animal miRNA binding sites are comparable, implying that the probabilities of losing a plant or animal miRNA binding site are similar under simple neutral evolutionary models. On the other hand, the length of a plant miRNA binding site means that in theory it would be virtually impossible for a plant gene to gain a new miRNA binding site by point mutation, whereas the same is not true for animals.

Box 3 | **MicroRNA target prediction: bioinformatics and experimental methods**

Most microRNA (miRNA) targets have been identified using bioinformatics methods (reviewed in REFS 88,109,121). In fact, miRNAs are one of the few classes of *trans*-acting regulatory factors for which computational approaches can, with reasonable confidence, successfully predict a large number of *cis*-regulatory binding sites. This is primarily owing to the fact that miRNAs recognize their targets at least partly on the basis of simple sequence complementarity between the miRNA and its binding site. In other words, pure knowledge of the sequence of a miRNA is sufficient to predict many targets. This is typically not yet possible for transcription factors, for which large training data sets or other experimental information are needed to accurately identify targets computationally[122,123].

In plants, miRNA binding sites are usually contained in coding regions and have extensive complementarity to the mature miRNA. Therefore, many plant miRNA binding sites have been successfully discovered using relatively straightforward bioinformatics screens[89,124]. Because the usual mechanism of repression in plants is mRNA cleavage, predicted binding sites can be verified with confidence using 5′-RACE (rapid amplification of cloned ends) to identify the cleavage products. In addition, smaller off-target effects are also possible[89].

Animal miRNA binding sites usually lie in 3′ UTRs of target mRNAs and exhibit imperfect complementarity to the mature miRNA (reviewed in REFS 88,109,121). Many target prediction methods are based on a model in which miRNA–mRNA binding is nucleated by an exact Watson–Crick complementary match to the first 6–8 bases from the 5′ end of the mature miRNA. Estimates of false-positive rates on the basis of comparative genomics (see REF. 88 and references therein), population genetics[74] and experimental assays[95] all indicate that the accuracy of these algorithms is high. For example, the estimated accuracy for targets that are conserved in humans, chimpanzees, mice, rats and dogs is 50–85% (REFS 74,88). However, other classes of miRNA sites have also been predicted, such as imperfect miRNA sites without an exact Watson–Crick match to the first 6–8 bases of the miRNA (reviewed in REFS 88,109,121).

Various assays have been developed to verify predicted animal miRNA targets, for example, by expressing the miRNA in a localized domain *in vivo* while simultaneously expressing and monitoring the target mRNA in a broader domain (for example, REF. 119). However, the number of target sites that have been validated *in vivo* under endogenous conditions and by mutagenesis of the predicted sites is extremely small. An entirely different approach from computational methods is to assay the expression levels of mRNAs directly, using the observation that miRNAs can not only downregulate protein levels, but also downregulate mRNA levels of their targets (for example, REF. 125). Therefore, a direct approach to target discovery is to knockout or overexpress the miRNA and use microarrays to identify the genes that show expression changes (for example, REFS 18,72,73), and to correlate these changes with 3′ UTR sequence motifs, for example, by a linear regression model[73]. Because of the additional information, evolutionary conservation filters on the target-site predictions can be relaxed. Therefore, this method is expected to have a higher sensitivity than approaches that are purely sequence-based, although it suffers from the possibility of indirect effects.

Several databases and web resources have been developed to store and maintain predicted target sites and sites with experimental support (reviewed in REF. 88,121).

Quantitative models of transcription-factor and miRNA binding-site evolution. One promising way of making the qualitative arguments of the previous section more quantitative involves developing models of binding-site evolution on the basis of point substitutions. Note that this approach does not accommodate the acquisition of new binding sites by large-scale rearrangements such as transpositions (for example, REF. 80), or small-scale rearrangements such as tandem duplications[81].

Stone and Wray[82] calculated the expected time for the appearance of a new transcription-factor binding site by neutral point substitution and concluded that this time was short (for example, 55,000 years for two 6-bp sites to evolve within a 200-bp region in *D. melanogaster*[82]), assuming that all members of the population evolve independently. The problem was revisited under the more realistic assumption that the members of the population are related by descent[83–85]. Although the methods and models differed, the three studies arrived at similar conclusions. First, neutral mutation is too slow to efficiently evolve new binding sites by point substitutions. Second, positive selection on partial binding sites can effectively speed up the rate of evolution, making it feasible to evolve new binding sites. Third, the time taken to create new sites scales linearly with the length of the regulatory region (for example, the promoter or enhancer regions for a transcription-factor binding site, or the 3′ UTR for a miRNA binding site) but exponentially with the length of the binding site. Fourth, the base composition of the region is important, particularly the presence of 'pre-sites' that are a single point mutation away from being a functional binding site.

As a specific example, consider a 1-kb region of non-coding DNA with equal base composition in the human genome. Assume that the binding sites for a miRNA and a transcription factor are each 8 bp long, but a miRNA requires exactly eight matches, whereas a transcription factor requires any seven matches out of eight. Durrett and Schmidt calculated that, given neutral point substitutions, it would take ~650 million years for the miRNA binding site to appear in the absence of a pre-site, and ~375,000 years in the presence of a pre-site. By contrast, the transcription-factor binding site would take ~60,000 years to appear[85].

The rate of acquisition of miRNA genes versus transcription factors. As more genome sequences are completed, it emerges that few novel transcription-factor families have arisen since the divergence between animals and plants[3,79] (although the number of transcription factors in each family can be different in individual genomes). The situation seems to be different for miRNAs. As a result of a combination of bioinformatics and sequencing efforts, it is now apparent that the process of miRNA creation is both active and ongoing (for example, REFS 37,38,86,87) (BOX 2). For example, it seems that the human genome alone might contain more than 1,000 miRNAs[86], of which many have been proposed to be primate-specific or even human-specific[87]. It should be noted that this global picture of transcription-factor-gene versus miRNA-gene acquisition ignores important issues such as combinatorial transcriptional control, co-factors and mutations outside of the DNA-binding domain that can affect transcription-factor binding specificity, as well as difficulties in computational identification of homologous miRNAs in different genomes. Nevertheless, even taking these points into account, the difference in the rates of creation of transcription-factor families and miRNA families remains striking, and it seems reasonable to propose that the speed of creation of new miRNA families has been faster in animal evolution than that of new transcription-factor families.

Do animals need so many miRNAs, and if so, why? Also, given that an animal miRNA can apparently target easily hundreds of genes (BOX 3) (reviewed in REF. 88), how can new miRNAs be acquired with such apparent ease without seriously disrupting the existing regulatory network of the organism? Even in plants in which the specificity of miRNA–mRNA binding is higher than in animals, significant off-target effects can occur and so similar issues can arise[89]. To address these questions, we next consider models of how new transcription factors and miRNAs evolve.

Creating new *trans*-factors

Transcription factors typically contain multiple functional domains, which mediate binding to DNA, interactions with other proteins and the subcellular localization of the transcription factor. A transcription factor with a new binding specificity can be created by duplication of an existing transcription factor followed by mutations, often, although not always, in the DNA-binding domain. Transcription factors can also evolve by the acquisition or loss of one of its other functional domains. For example, the loss of a transcriptional activation domain could turn an activator into a repressor, whereas the acquisition of a new protein–protein interaction domain that facilitates heterodimerization with a novel binding partner could significantly alter the targets of the transcription factor. It is often assumed that mutations in the coding sequence of the transcription factor itself are likely to be highly deleterious because they potentially affect the expression of many downstream target genes. However, recent examples of transcription factors that are important in *Drosophila* development that have significantly diverged in sequence and function indicate that this assumption might be worth revisiting (reviewed in REF. 34).

A duplication-mutation model accounts also for the evolution of at least some miRNA genes. For example, human *miR-10a* and *miR-100* are homologues but differ by a single nucleotide insertion-deletion in the predicted target recognition region of the respective mature miRNAs. If this were the predominant mode of miRNA-family creation, one might expect that miRNA target recognition motifs would not be randomly distributed in sequence space, but would show a propensity to cluster in groups of similar sequences. Although this problem has not been fully studied, preliminary bioinformatic analyses indicate that the distribution of these sequences seems to be close to random (K.C. and N.R., unpublished observations). If true, this would indicate that a duplication-mutation process accounts for a

Deep sequencing
Sequencing to high coverage, where coverage (or depth) corresponds to the average number of times that a nucleotide is sequenced.

relatively small fraction of miRNAs, and that instead most new miRNA families arise *de novo*.

One model of *de novo* miRNA acquisition, the inverted duplication model, suggests that new miRNAs evolve by inverted duplication of a stretch of coding sequence followed by subsequent erosion of the sequence into an imperfect hairpin structure[90]. In some cases, part of the coding sequence itself might be duplicated before the creation of the miRNA, so the miRNA shows homology to two regions of the target gene[128]. This model is attractive for plant miRNAs because it accounts for the long stretches of sequence similarity that are required between the miRNA and its target.

The inverted duplication model seems less appropriate for animal miRNAs, because the length of complementary sequence in miRNA binding sites is much smaller in animals than in plants, and no examples of newly formed miRNAs arising in this manner are known in animals. A second model, the random creation model, proposes that new miRNAs simply arise randomly from existing hairpin structures in the genome[54,91]. Hairpin structures are generally abundant in eukaryotic genomes — for instance, Bentwich *et al.*[92] identified 11 million hairpins in the human genome in a bioinformatic screen. Therefore, the problem of creating a new miRNA in animals might be less involved with creating a new hairpin structure, but rather with appropriately transcribing an existing hairpin structure in the genome and providing the prerequisite signals for biogenesis of the new miRNA (for example, signals for processing by the RNase III Drosha) (BOX 1).

A model of transcriptional control of new miRNAs. Because the minimal binding site of an animal miRNA is short, a new miRNA should be able to target many mRNAs simply by chance, and many of these interactions are likely to be selectively deleterious, as is the case for all types of mutations. Indeed, the existence of many genes for which the presence of a miRNA binding site would be deleterious ('anti-targets') was proposed[93] and later demonstrated by several groups in various species using microarray data and computational studies[72,73,89,94,95].

These observations raise the question of how a new miRNA could ever be acquired without seriously impairing the fitness of the organism. We propose that one way this could happen is if the miRNA were initially transcribed only weakly, and in a specific tissue or at a specific developmental stage. Current studies indicate that multiple sites for the same miRNA in the same target mRNA are needed to generate a strong regulatory effect of the miRNA on this target (REF. 88 and references therein). If this is true, natural selection can eliminate slightly deleterious miRNA sites over time — recall that it is easier for an mRNA to lose a miRNA site than to gain one — while also maintaining or creating beneficial binding sites for the new miRNA. Once this elimination process is complete, the expression level of the miRNA can be increased and its tissue-specificity can be relaxed (FIG. 2).

Although this model is clearly a simplified picture of a complex process, it makes the testable prediction that more-recently evolved miRNAs should be expressed weakly and in specific spatio-temporal domains. Indeed, this prediction is generally supported by the available miRNA expression data (for example, REFS 87,96), which indicate that more-recently acquired human miRNAs are more likely to be weakly expressed than ancient conserved miRNAs. A recent study by Berezikov *et al.*[87], in which miRNA expression in the human and chimpanzee brain was determined using deep sequencing (BOX 2), is also consistent with this prediction.

The model puts the enormous number of putative small RNA transcripts that are being revealed by deep sequencing efforts in perspective: it suggests that a number of them are randomly transcribed hairpins that might not have a significant biological role as *trans*-acting regulators, although it does not preclude the possibility that they could acquire such functionality in the future. The model is consistent with the hypothesis proposed by Sempere *et al.*[38] and Prochnik *et al.*[39] that the acquisition of new miRNAs contributed to the acquisition of novel tissue types and organs in animal development. It is also consistent with an intriguing idea put forward by Davidson[7] that it is relatively easy to evolve a new miRNA gene that targets a specific sequence motif, but the same is not true of transcription factors, because the sequence specificity of a protein is a complex function of its amino-acid sequence. So, if a transcription factor were to acquire a new domain of expression (for example, in a new tissue), it would be expected to regulate genes that it already regulates in its original domain of expression, probably leading to deleterious effects. However, assuming that a miRNA with an arbitrary target sequence can evolve easily, such complications can be avoided, and consequently it might be easier to insert a miRNA into a developmental network than a transcription factor.

The expression and conservation of miRNAs are not as well understood in plants compared with animals. The correlation between the age of a miRNA and its expression level is expected to be weaker in plants than in animals for two reasons. First, a newly arising plant miRNA is expected to have relatively few targets compared with

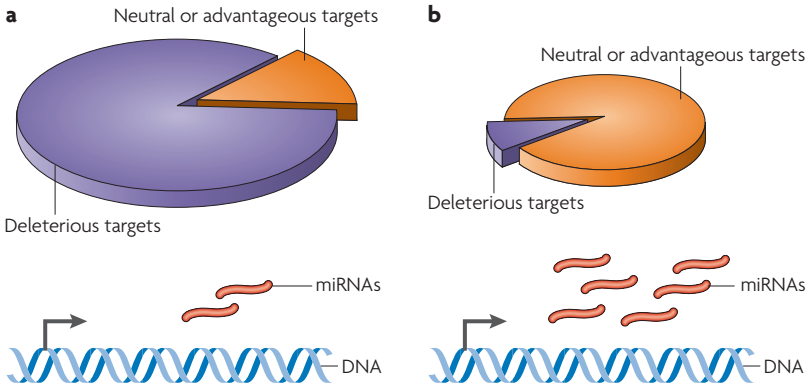


Figure 2 | **A model of the acquisition of a new microRNA. a** | Initially, a new microRNA (miRNA) is expressed at low levels and in a specific spatio-temporal domain. It has many targets that appear at random in the genome, many of which are selectively slightly deleterious and a few of which are selectively neutral or advantageous. **b** | At a later point in time, natural selection has purged many deleterious targets from the genome while preserving many of the neutral or advantageous targets, and the expression of the miRNA can increase without being highly deleterious to the organism.

an animal miRNA, and if these are either highly beneficial or highly deleterious, then selection can drive up or down the expression level of the miRNA more quickly. Second, a single plant miRNA molecule can target and cleave many mRNAs, whereas it seems that animal miRNAs must be bound to their targets to confer repression, a process that is reversible under specific conditions[97,98]. Nonetheless, it is likely that some plant miRNAs that are weakly expressed or tissue specific have little or no biological function as *trans*-acting regulators[128].

Conclusion

With our rapidly advancing knowledge of the different mechanisms of gene regulation in higher eukaryotes, we can begin to consider the evolutionary implications of these different mechanisms within a unified framework. In the past, much work focused on a synthesis between transcriptional regulation and cell signalling mechanisms[3,99]; here we have concentrated on the evolution of transcription factors and miRNAs. Ultimately, all other mechanisms of gene regulation should be brought into the discussion in order to form a holistic picture of the evolution of gene regulation.

Many open questions and directions for future research remain. This Review gives a local view of the evolution of individual regulators and binding sites as a necessary first step to understanding the evolution of gene regulation as a whole. In the future, it will be necessary to move towards the broader view of the evolution of developmental regulatory networks, and from there, towards the even bigger picture of changes in organismal form. The model of Davidson and colleagues that is discussed above is one promising way of thinking about the evolution of network structure and body plans on a global scale. Our current knowledge of how transcription factors, miRNAs, signalling pathways and other regulators are wired together into developmental networks is much too rudimentary to make any sensible statements about the effect of different regulatory mechanisms on global network evolution. However, as our knowledge of the developmental roles of these regulatory mechanisms increases, it should be possible to extend the model to account for these different components. For example, if it indeed turns out that miRNAs tend to work at the periphery of developmental networks to confer additional layers of robustness, and not as the primary agents of developmental patterning, then this might lead us to postulate a certain amount of evolutionary pliability for miRNA-mediated regulation. It will also be interesting to investigate whether different eukaryotic lineages, particularly plants and animals, use different regulatory mechanisms in similar ways or not.

Such global trends in network evolution have a natural counterpart in local subnetwork motifs that exist within the overall network. One particularly interesting example of such a motif is a feedback loop that involves multiple transcription factors and miRNAs. Examples of this motif have been identified in recent work on neuron cell-fate determination[100] and vulval development[101] in *C. elegans*, and granulocytic differentiation in humans[102]. Whether multicomponent feedback loops and other complex subnetwork motifs are common features of developmental networks, and whether they have any significance for organismal traits and evolution, are intriguing questions for future research. Within the Davidson model, signalling cassettes function as plug-in components that are re-used repeatedly in regulatory networks. Likewise, certain subnetwork motifs could also potentially form re-usable plug-in components.

Since the work of Mary-Claire King and Allan Wilson three decades ago[103], scientists have asked whether changes in gene regulation or protein sequence have made bigger contributions to phenotypic differences between species. Today, we are well positioned to broaden the question to ask about the relative contributions of the evolution of different mechanisms of gene regulation to the evolution of phenotypic diversity in animals and plants. The long journey towards a comprehensive understanding of the evolution of gene regulation is only beginning.

1. Levine, M. & Tjian, R. Transcription regulation and animal diversity. *Nature* **424**, 147–151 (2003).
2. Moore, M. J. From birth to death: the complex lives of eukaryotic mRNAs. *Science* **309**, 1514–1518 (2005).
3. Davidson, E. H. *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution* (Academic, New York, 2006).
This book synthesizes several decades of work on gene regulation, animal development and evolution, with an emphasis on transcriptional regulation.
4. Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350–355 (2004).
5. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism and function. *Cell* **116**, 281–297 (2004).
6. Wittkopp, P. J. Genomic sources of regulatory variation in *cis* and in *trans*. *Cell. Mol. Life Sci.* **62**, 1779–1783 (2005).
7. Davidson, E. H. & Erwin, D. H. Gene regulatory networks and the evolution of animal body plans. *Science* **311**, 796–800 (2006).
8. Carroll, S., Grenier, J. & Weatherbee, S. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Blackwell Scientific, Malden, 2005).
9. Doebley, J. & Lukens, L. Transcriptional regulators and the evolution of plant form. *Plant Cell* **10**, 1075–1082 (1998).
10. Kloosterman, W. P. & Plasterk, R. H. The diverse functions of microRNAs in animal development and disease. *Dev. Cell* **11**, 441–450 (2006).
11. Kosik, K. S. The neuronal microRNA system. *Nature Rev. Neurosci.* **7**, 911–920 (2006).
12. Carrington, J. C. & Ambros, V. Role of microRNAs in plant and animal development. *Science* **301**, 336–338 (2003).
13. Jones-Rhoades, M. W., Bartel, D. P. & Bartel, B. MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* **57**, 19–53 (2006).
14. Alvarez-Garcia, I. & Miska, E. A. MicroRNA functions in animal development and human disease. *Development* **132**, 4653–4662 (2005).
15. Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843–854 (1993).
16. Reinhart, B. J. *et al.* The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901–906 (2000).
17. Sokol, N. S. & Ambros, V. Mesodermally expressed *Drosophila* microRNA-1 is regulated by Twist and is required in muscles during larval growth. *Genes Dev.* **19**, 2343–2354 (2005).
18. Giraldez, A. J. *et al.* Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**, 75–79 (2006).
19. Giraldez, A. J., Cinalli, R. M., Glasner, M. E., Thomson, J. M. & Baskerville, S. MicroRNAs regulate brain morphogenesis in zebrafish. *Science* **308**, 833–838 (2005).
20. Bernstein, E. *et al.* Dicer is essential for mouse development. *Nature Genet.* **35**, 215–217 (2003).
21. Schauer, S. E., Jacobsen, S. E., Meinke, D. W. & Ray, A. DICER-LIKE1: blind men and elephants in *Arabidopsis* development. *Trends Plant Sci.* **7**, 487–491 (2002).
22. Wienholds, E. *et al.* MicroRNA expression in zebrafish embryonic development. *Science* **309**, 310–311 (2005).
23. Ason, B. *et al.* Differences in vertebrate microRNA expression. *Proc. Natl Acad. Sci. USA* **103**, 14385–14389 (2006).
24. Thomson, J. M. *et al.* Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev.* **20**, 2202–2207 (2006).
25. Aravin, A. A. *et al.* The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell* **5**, 337–350 (2003).
26. Aboobaker, A. A., Tomancak, P., Patel, N., Rubin, G. M. & Lai, E. C. *Drosophila* microRNAs exhibit diverse spatial expression patterns during embryonic development. *Proc. Natl Acad. Sci. USA* **102**, 18017–18022 (2005).
27. Leaman, D. *et al.* Antisense-mediated depletion reveals essential and specific functions of microRNAs in *Drosophila* development. *Cell* **121**, 1097–1108 (2005).
28. Brennecke, J., Stark, A. & Cohen, S. M. Not miR-ly muscular: microRNAs and muscle development. *Genes Dev.* **19**, 2261–2264 (2005).

29. Hornstein, E. & Shomron, S. Canalization of development by microRNAs. *Nature Genet.* **38**, S20–S24 (2006).
30. Li, Y., Wang, Y., Lee, J. A. & Gao, F. B. MicroRNA-9a ensures the precise specification of sensory organ precursors in *Drosophila*. *Genes Dev.* **20**, 2793–2805 (2006).
31. Cohen S. M., Brennecke, J. & Stark A. Denoising feedback loops by thresholding — a new role for microRNAs. *Genes Dev.* **20**, 2769–2772 (2006).
32. Meyerowitz, E. M. Plants compared to animals: the broadest comparative study of development. *Science* **295**, 1482–1485 (2002).
33. Riechmann, J. L. *et al. Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* **290**, 2105–2110 (2000).
A detailed study of the evolution of transcription-factor families in plants and animals, which showed that they have diversified greatly during eukaryotic evolution.
34. Hsia, C. C. & McGinnis, W. Evolution of transcription factor function. *Curr. Opin. Genet. Dev.* **13**, 199–206 (2003).
35. Pasquinelli, A. E. *et al.* Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. *Nature* **408**, 86–89 (2000).
This paper demonstrated that *let-7*, one of the two founding miRNAs, is highly conserved in animals.
36. Pasquinelli, A. E. *et al.* Expression of the 22 nucleotide *let-7* heterochronic RNA throughout the metazoa: a role in life history evolution? *Evol. Dev.* **5**, 372–378 (2003).
37. Hertel, J. *et al.* The expansion of the metazoan miRNA repertoire. *BMC Genomics* **7**, 25 (2006).
38. Sempere, L. F., Cole, C. N., McPeek, M. A. & Peterson, K. J. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *J. Exp. Zool. B. Mol. Dev. Evol.* **306**, 575–588 (2006).
The authors studied the evolution of miRNAs in bilaterians, cnidarians and sponges, and proposed that the acquisition of new miRNAs has had an important role in the development of new animal organs.
39. Prochnik, S. E., Rokhsar, D. S. & Aboobaker, A. A. Evidence for a microRNA expansion in the bilaterian ancestor. *Dev. Genes Evol.* 14 Nov 2006 (doi: 10.1007/s00427-006-0116-1).
40. Lemons, D. & McGinnis, W. Genomic evolution of Hox gene clusters. *Science* **313**, 1918–1922 (2006).
41. Xie, X. *et al.* Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* **434**, 338–345 (2005).
42. Chen, K. & Rajewsky, N. Deep conservation of miRNA-target relationships and 3′ UTR motifs in vertebrates, flies and nematodes. *Cold Spring Harb. Symp. Quant. Biol.* 12 Dec 2006 (doi:10.1101/sqb.2006.71.039).
43. Chan, C. S., Elemento, O. & Tavazoie, S. Revealing posttranscriptional regulatory elements through network-level conservation. *PLoS Comput. Biol.* **1**, e69 (2006).
44. Lu, C. *et al.* MicroRNAs and other small RNAs enriched in the *Arabidopsis* RNA-dependent RNA polymerase-2 mutant. *Genome Res.* **16**, 1276–1288 (2006).
45. Arazi, T. *et al.* Cloning and characterization of micro-RNAs from moss. *Plant J.* **43**, 837–848 (2005).
46. Floyd, S. K. & Bowman, J. L. Gene regulation: ancient microRNA target sequences in plants. *Nature* **428**, 485–486 (2004).
47. Axtell, M. J. & Bartel, D. P. Antiquity of microRNAs and their targets in land plants. *Plant Cell* **17**, 1658–1673 (2005).
These authors showed that many miRNAs and miRNA-target relationships are well conserved in plants.
48. Shiu, S. H., Shih, M. C. & Li, W. H. Transcription factor families have much higher expansion rates in plants than in animals. *Plant Physiol.* **139**, 18–26 (2005).
49. Blanc, G. & Wolfe, K. H. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* **16**, 1679–1691 (2004).
50. Seoighe, C. & Gehring, C. Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* promoterome. *Trends Genet.* **20**, 461–464 (2004).
51. Davis, J. C. & Petrov, D. A. Do disparate mechanisms of duplication add similar genes to the genome? *Trends Genet.* **21**, 548–551 (2005).
52. Maher, C., Stein, L. & Ware, D. Evolution of *Arabidopsis* microRNA familes through duplication events. *Genome Res.* **16**, 510–519 (2006).
53. Tanzer, A., Amemiya, C. T., Kim, C. B. & Stadler, P. F. Evolution of microRNAs located within Hox gene clusters. *J. Exp. Zool. B. Mol. Dev. Evol.* **304**, 75–85 (2005).
54. Tanzer, A. & Stadler, P. F. Molecular evolution of a microRNA cluster. *J. Mol. Biol.* **339**, 327–335 (2004).
55. Li, A. & Mao, L. Evolution of plant microRNA gene families. *Cell Res.* 28 Nov 2006 (doi:10.1038/sj.cr.7310113).
56. Abbott, A. L. *et al.* The *let-7* microRNA family members *mir-48*, *mir-84*, and *mir-241* function together to regulate developmental timing in *Caenorhabditis elegans*. *Dev. Cell* **9**, 403–414 (2005).
57. Hayes, G. D., Frand, A. R. & Ruvkun, G. The *mir-84* and *let-7* paralogous microRNA genes of *Caenorhabditis elegans* direct the cessation of molting via the conserved nuclear hormone receptors NHR-23 and NHR-25. *Development* **133**, 4631–4641 (2006).
58. Johnson, R. M. *et al.* RAS is regulated by the *let-7* microRNA family. *Cell* **120**, 635–647 (2005).
59. Schulman, B. R. M., Esquela-Kerscher, A. & Slack, F. J. Reciprocal expression of *Lin-41* and the microRNAs *Let-7* and *Mir-125* during mouse embryogenesis. *Dev. Dyn.* **234**, 1046–1054 (2005).
60. Moss, E. G. & Tang, L. Conservation of the heterochronic regulator *Lin-28*, its developmental expression and microRNA complementary sites. *Dev. Biol.* **258**, 432–442 (2003).
61. Yang, D.-H. & Moss, E. G. Temporally regulated expression of *Lin-28* in diverse tissues of the developing mouse. *Gene Expr. Patterns* **3**, 719–726 (2003).
62. Grun, D., Wang, Y., Langenberger, D., Gunsalus, K. C. & Rajewsky, N. microRNA target predictions across seven *Drosophila* species and comparison to mammalian targets. *PLoS Comput. Biol.* **1**, e13 (2005).
63. Elnitski, L., Jin, V. X., Farnham, P. J. & Jones, S. J. M. Locating mammalian transcription factor binding sites: A survey of computational and experimental techniques *Genome Res.* **16**, 1455–1464 (2006).
64. Gerber, A. P., Luschnig, S., Krasnow, M. A., Brown, P. O. & Herschlag, D. Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **103**, 4487–4492 (2006).
65. Dermitzakis, E. T. & Clark, A. G. Evolution of transcription factor binding sites in mammalian gene regulatory regions: conservation and turnover. *Mol. Biol. Evol.* **19**, 1114–1121 (2002).
66. Dermitzakis, E. T., Bergman, C. M. & Clark, A. G. Tracing the evolutionary history of *Drosophila* regulatory regions with models that identify transcription factor binding sites. *Mol. Biol. Evol.* **20**, 703–714 (2003).
67. Emberly, E., Rajewsky, N. & Siggia, E. D. Conservation of regulatory elements between two regions of *Drosophila*. *BMC Bioinformatics* **4**, 57 (2003).
68. Richards, S. *et al.* Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene and *cis*-element conservation. *Genome Res.* **15**, 1–18 (2005).
69. Moses, A. M. *et al.* Large-scale turnover of functional transcription factor binding sites in *Drosophila*. *PLoS Comput. Biol.* **2**, e130 (2006).
70. Ludwig, M. Z., Bergman, C. M., Patel, N. H. & Kreitman, M. Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* **403**, 564–567 (2000).
An elegant demonstration of the complex evolution of the well studied *D. melanogaster* even-skipped stripe 2 enhancer.
71. Ludwig, M. Z. *et al.* Functional evolution of a *cis*-regulatory module. *PLoS Biol.* **3**, e93 (2005).
72. Krutzfeldt, J. *et al.* Silencing of microRNAs *in vivo* with 'antagomirs'. *Nature* **438**, 685–689 (2005).
73. Sood, P., Krek, A., Zavolan, M., Macino, G. & Rajewsky, N. Cell-type-specific signatures of microRNAs on target mRNA expression. *Proc. Natl Acad. Sci. USA* **103**, 2746–2751 (2006).
74. Chen, K. & Rajewsky, N. Natural selection on human miRNA binding sites inferred from SNP data. *Nature Genet.* **38**, 1452–1456 (2006).
This paper examined polymorphism in human miRNA binding sites and used population genetic techniques to estimate levels of selective constraint on these binding sites.
75. Rockman, M. V. & Wray, G. A. Abundant raw material for *cis*-regulatory evolution in humans. *Mol. Biol. Evol.* **19**, 1991–2004 (2002).
A systematic survey of the literature on naturally occurring polymorphisms in human promoter regions that significantly affect gene expression levels.
76. Abelson, J. F. *et al.* Sequence variants in *SLITRK1* are associated with Tourette's syndrome. *Science* **310**, 317–320 (2005).
77. Clop, A. *et al.* A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nature Genet.* **38**, 813–818 (2006).
The clearest example to date of a naturally occurring SNP in a miRNA binding site that affects a gross phenotypic trait.
78. Iwai, N. & Naraba, H. Polymorphisms in human pre-miRNAs. *Biochem. Biophys. Res. Commun.* **331**, 1439–1444 (2005).
79. Wray, G. A. *et al.* The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **20**, 1377–1419 (2003).
A comprehensive review of the evolution of transcription factors and their binding sites.
80. Balhoff, J. P. & Wray, G. A. Evolutionary analysis of the well characterized endo16 promoter reveals substantial variation within functional sites. *Proc. Natl Acad. Sci. USA* **102**, 8591–8596 (2005).
81. Sinha, S. & Siggia, E. D. Sequence turnover and tandem repeats in *cis*-regulatory modules in *Drosophila*. *Mol. Biol. Evol.* **22**, 874–885 (2005).
82. Stone, J. & Wray, G. Rapid evolution of *cis*-regulatory sequences via local point mutations. *Mol. Biol. Evol.* **18**, 1764–1770 (2001).
83. Berg, J., Willmann, S. & Laessig, M. Adaptive evolution of transcription factor binding sites. *BMC Evol. Biol.* **4**, 42 (2004).
84. MacArthur, S. & Brookfield, J. F. Y. Expected rates and modes of evolution of enhancer sequences. *Mol. Biol. Evol.* **21**, 1064–1073 (2004).
85. Durrett, R. & Schmidt, D. Waiting for regulatory sequences to appear. *Ann. Appl. Probab.* (in the press).
86. Berezikov, E. *et al.* Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis. *Genome Res.* **16**, 1289–1298 (2006).
87. Berezikov, E. *et al.* Diversity of microRNAs in human and chimpanzee brain. *Nature Genet.* **38**, 1375–1377 (2006).
These authors used 454 sequencing to show that many miRNAs that are expressed in the human brain are not conserved in the chimpanzee brain.
88. Rajewsky, N. microRNA target predictions in animals. *Nature Genet.* **38**, S8–S13 (2006).
89. Schwab, R. *et al.* Specific effects on microRNAs on the plant transcriptome. *Dev. Cell* **8**, 517–527 (2005).
90. Allen, E. *et al.* Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nature Genet.* **36**, 1282–1290 (2004).
91. Svoboda, P. & Di Cara, A. Hairpin RNA: a secondary structure of primary importance. *Cell. Mol. Life Sci.* **63**, 901–918 (2006).
92. Bentwich, I. *et al.* Identification of hundreds of conserved and nonconserved human microRNAs. *Nature Genet.* **37**, 766–770 (2005).
93. Bartel, D. P. & Chen, C. Z. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nature Rev. Genet.* **5**, 396–400 (2004).
94. Farh, K. K. *et al.* The widespread impact of mammalian microRNAs on mRNA repression and evolution. *Science* **310**, 1817–1821 (2005).
95. Stark, A., Brennecke, J., Bushati, N., Russell, R. B. & Cohen, S. M. Animal microRNAs confer robustness to gene expression and have a significant impact on 3′ UTR evolution. *Cell* **123**, 1133–1146 (2005).
96. Baskerville, S. & Bartel, D. P. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* **11**, 241–247 (2005).
97. Bhattacharyya, S. N., Habermacher, R., Martine, U., Closs, E. I. & Filipowicz, W. Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* **125**, 1111–1124 (2006).
98. Ashraf, S. I., McLoon, A. L., Sclarsic, S. M. & Kunes, S. Synaptic protein synthesis associated with memory is regulated by the RISC pathway in *Drosophila*. *Cell* **124**, 191–205 (2006).

99. Davidson, E. H. *Genomic Regulatory Systems. Development and Evolution* (Academic, San Diego, 2001).
100. Johnston, R. J. J., Chang, S., Etchberger, J. F., Ortiz, C. O. & Hobert, O. MicroRNAs acting in a double-negative feedback loop to control a neuronal cell fate decision. *Proc. Natl Acad. Sci.* **102**, 12449–12454 (2005).
101. Yoo, A. S. & Greenwald, I. *LIN-12*/Notch activation leads to microRNA-mediated down-regulation of *Vav* in *C. elegans*. *Science* **310**, 1330–1333 (2005).
102. Fazi, F. *et al.* A minicircuitry comprised of microRNA-223 and transcription factors NFI-A and C/EBPα regulates human granulopoiesis. *Cell* **123**, 819–831 (2005).
103. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
104. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23**, 4051–4060 (2004).
105. Johnson, S. M., Lin, S. Y. & Slack, F. J. The time of appearance of the *C. elegans let-7* microRNA is transcriptionally controlled utilizing a temporal regulatory element in its promoter. *Dev. Biol.* **259**, 364–379 (2003).
106. Biemar, F. *et al.* Spatial regulation of microRNA gene expression in the *Drosophila* embryo. *Proc. Natl Acad. Sci.* **102**, 15907–15911 (2005).
107. Engels, B. M. & Hutvagner, G. Principles and effects of microRNA-mediated post-transcriptional gene regulation. *Oncogene* **25**, 6163–6169 (2006).
108. Berezikov, E., Cuppen, E. & Plasterk, R. H. Approaches to microRNA discovery. *Nature Genet.* **38**, S2–S7 (2006).
109. Bentwich, I. Prediction and validation of microRNAs and their targets. *FEBS Lett.* **579**, 5904–5910 (2005).
110. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
111. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429–3431 (2003).
112. Pfeffer, S. *et al.* Identification of virus-encoded microRNAs. *Science* **304**, 734–736 (2004).
113. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. Identification of novel genes coding for small expressed RNAs. *Science* **294**, 853–858 (2001).
114. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
115. Lu, C. *et al.* Elucidation of the small RNA component of the transcriptome. *Science* **309**, 1567–1569 (2005).
116. Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A. & Eddy, S. R. Rfam: an RNA family database. *Nucleic Acids Res.* **31**, 439–441 (2003).
117. Krek, A. *et al.* Combinatorial microRNA target predictions. *Nature Genet.* **37**, 495–500 (2005).
118. Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).
119. Brennecke, J., Stark, A., Russell, R. B. & Cohen, S. M. Principles of microRNA-target recognition. *PLoS Biol.* **3**, e85 (2005).
120. Lall, S. *et al.* A genome-wide map of conserved microRNA targets in *C. elegans*. *Curr. Biol.* **16**, 460–471 (2006).
121. Sethupathy, P., Megraw, M. & Hatzigeorgiou, A. G. A guide through present computational approaches for the identification of mammalian microRNA targets. *Nature Methods* **3**, 881–886 (2006).
122. Tompa, M. *et al.* Assessing computational tools for the discovery of transcription factor binding sites. *Nature Biotechnol.* **23**, 137–144 (2005).
123. Siggia, E. D. Computational methods for transcriptional regulation. *Curr. Opin. Genet. Dev.* **15**, 214–221 (2005).
124. Rhoades, M. W. *et al.* Prediction of plant microRNA targets. *Cell* **110**, 513–520 (2002).
125. Lim, L. P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–773 (2005).
126. Pillai, R. S., Bhattacharyya, S. N. & Filipowicz, W. Repression of protein synthesis by miRNAs: how many mechanisms? *Trends Cell Biol.* 2 Jan 2007 (doi:10.1016/j.tcb.2006.12.007).
127. Arteaga-Vazquez, M., Caballero-Perez, J. & Vielle-Calzada, J. P. A family of microRNAs present in plants and animals. *Plant Cell* 22 Dec 2006 (doi:10.1105/tpc.106.044420).
128. Rajagopalan, R., Vaucheret, H., Trejo, J., Bartel, D. P. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev.* **20**, 3407–3425.

Acknowledgements

We regret that due to space constraints, we are unable to cite the work of many colleagues who have made key contributions to the field. We thank K. Birnbaum for insightful discussions about gene regulation in plants, and V. Ambros, M. Hammell, S. Small and N. Sokol for helpful comments on a preliminary version of the manuscript.

Competing interests statement

The authors declare no competing financial interests.

DATABASES

The following terms in this article are linked online to:
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
let-7 | *miR-1* | *miR-7* | *miR-10* | *miR-100* | Pax6 | Nkx2-5 |
UniProtKB: http://ca.expasy.org/sprot
dicer | Pumilio

FURTHER INFORMATION

The Rajewsky laboratory web site:
http://www.mdc-berlin.de/rajewsky
NYU Department of Biology and Center for Comparative Functional Genomics: http://www.nyu.edu/fas/dept/biology

SUPPLEMENTARY INFORMATION

See online article: S1 (figure)
Access to this links box is available online