

Evolution of MicroRNAs and the Diversification of Species

Yong-Hwee E. Loh, Soojin V. Yi, and J. Todd Streebman*

School of Biology and Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology

*Corresponding author: E-mail: todd.streebman@biology.gatech.edu.

Accepted: 9 December 2010

Abstract

MicroRNAs (miRNAs) are ancient, short noncoding RNA molecules that regulate the transcriptome through post-transcriptional mechanisms. miRNA riboregulation is involved in a diverse range of biological processes, and misregulation is implicated in disease. It is generally thought that miRNAs function to canalize cellular outputs, for instance as “fail-safe” repressors of gene misexpression. Genomic surveys in humans have revealed reduced genetic polymorphism and the signature of negative selection for both miRNAs themselves and the target sequences to which they are predicted to bind. We investigated the evolution of miRNAs and their binding sites across cichlid fishes from Lake Malawi (East Africa), where hundreds of diverse species have evolved in the last million years. Using low-coverage genome sequence data, we identified 100 cichlid miRNA genes with mature regions that are highly conserved in other animal species. We computationally predicted target sites on the 3′-untranslated regions (3′-UTRs) of cichlid genes to which miRNAs may bind and found that these sites possessed elevated single nucleotide polymorphism (SNP) densities. Furthermore, polymorphic sites in predicted miRNA targets showed higher minor allele frequencies on average and greater genetic differentiation between Malawi lineages when compared with a neutral expectation and nontarget 3′-UTR SNPs. Our data suggest that divergent selection on miRNA riboregulation may have contributed to the diversification of cichlid species and may similarly play a role in rapid phenotypic evolution of other natural systems.

Key words: cichlid, single nucleotide polymorphism, riboregulation, evolutionary radiation, minor allele frequency.

Introduction

Ever since King and Wilson compared protein sequence between chimpanzee and human and concluded that there was insufficient coding divergence to explain phenotypic differences (King and Wilson 1975), biologists have highlighted regulatory change in gene expression as a source for adaptive evolution (Wray 2007; Carroll 2008). There is now ample direct evidence that *cis*-acting mutations cause phenotypic variation among closely related organisms by modulating gene expression (Sucena et al. 2003; Miller et al. 2007). These data, coupled with the signature of divergent and positive selection at putative gene regulatory elements (Haygood et al. 2007; Sethupathy et al. 2008), have established the general consensus that 5′ promoters act as evolutionary engines of transcriptional change (e.g., tinker where the tinkering's good; Rockman and Stern 2008).

Plausible scenarios for the evolution of animal diversity hinge on the ever-growing complexity of 5′ promoters and the modification of transcriptional regulatory networks (Levine and Tjian 2003). Notably, evolutionary “tinkering” with transcription at 5′ promoters may have evolved in con-

cert with post-transcriptional safeguards encoded at the 3′ end of cistrons. Reports suggest that microRNAs (miRNAs), potent agents of riboregulation, are as old as metazoan 5′ *cis*-regulatory logic (Grimson et al. 2008; Wheeler et al. 2009). miRNAs are short (~22 nucleotides), endogenous noncoding RNA molecules that regulate gene expression after transcription. Generally, animal miRNA targeting is achieved by complementary base pairing between the miRNA and the specific sequences in the 3′-untranslated region (3′-UTR) of messenger RNAs (mRNAs). Target recognition is thought to be determined by perfect Watson–Crick base pairing at a miRNA “seed” region (base positions 2–7 counting from the 5′ end; Lewis et al. 2005), although this is not a necessary condition, and targeting may include other determinants (Grimson et al. 2007; Barbato et al. 2009). Transcript silencing then occurs through inhibition of translation or via mRNA degradation (Bartel 2004). Individual miRNAs may regulate hundreds of loci, and it has been estimated that a majority of human genes are potential miRNA targets (Lewis et al. 2005; Friedman et al. 2009).

© The Author(s) 2011. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.5>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

miRNAs generally act as “fail-safe” buffers against gene misexpression in time and/or space, in effect canalizing the transcriptome (Carrington and Ambrose 2003; Stark et al. 2005). Consistent with this notion, miRNA misexpression and/or genetic polymorphism in target sequences can cause abnormality and disease (Clop et al. 2006; Eberhart et al. 2008; Sethupathy and Collins 2008; Mencia et al. 2009). Likewise, and in contrast to predicted transcription factor binding sites in 5′ promoters, human miRNAs and their 3′-UTR target sequences evolve under purifying selection (Chen and Rajewsky 2006; Saunders et al. 2007; Sethupathy et al. 2008).

As humans and chimps diverged from a common ancestor during the last 5–7 My, the East African Rift lakes Tanganyika, Malawi, and Victoria spawned three of the most spectacular evolutionary radiations known to biology (Kornfield and Smith 2000; Salzburger et al. 2005). In Lake Malawi alone, hundreds of cichlid fish species have evolved from a common ancestor over the last million years (Won et al. 2005). These species are remarkably diverse in size, shape, color, and behavior (Streelman et al. 2003; Albertson et al. 2005; Carleton et al. 2008; Fraser et al. 2008; Sylvester et al. 2010), yet their genomes are highly similar and share ancestral polymorphism (Moran and Kornfield 1993; Loh et al. 2008). We have shown recently that most of the genome is not genetically differentiated among Malawi species and major lineages; only 2–4% of single nucleotide polymorphism (SNP) loci exhibit the statistical signature of strong evolutionary divergence (Loh et al. 2008). Cichlids are models of the mapping of phenotype to genotype; the problem of so many biological species in so little time (Kocher 2004) is equally matched by the problem of rapid diversification and evolutionary novelty (Streelman et al. 2007).

We hypothesized that divergence of miRNAs or their target sequences might be one of the genomic mechanisms contributing to the rapid phenotypic evolution observed in Lake Malawi cichlids. To this end, we analyzed available low-coverage genome sequence and SNP data (Loh et al. 2008) and computationally identified 1) putative cichlid miRNAs and 2) the target sequences in 3′-UTRs to which miRNAs may bind. Most studies of miRNA focus on evolutionary conservation of the molecules and their target sites (Bartel 2004; Alexiou et al. 2009; Barbato et al. 2009). Our goal of evaluating the link(s) between miRNAs, polymorphism in putative miRNA targets and diversity among Lake Malawi cichlid species predicates that we not only consider target sequences conserved for hundreds of millions of years but also those that may have evolved more recently. Such “nonconserved” targets are known to be functional and may be generated by single mutations to standing sequence (Farh et al. 2005; Clop et al. 2006).

We observed that predicted cichlid mature miRNAs are strongly conserved in sequence. On the other hand, miRNA targets exhibited greater SNP densities than flanking

sequences and the overall 3′-UTR average. Moreover, polymorphic sites in target sequences showed higher minor allele frequencies (MAFs) and divergence among Malawi evolutionary lineages when compared against a neutral expectation and nontarget SNPs in the same set of 3′-UTRs. Our data reveal a signature of divergent selection on cichlid miRNA binding sites and suggest an evolutionary role for miRNA riboregulation in the diversification of species.

Materials and Methods

Lake Malawi Genomes

We obtained Lake Malawi cichlid genomic data, consisting of 304,310 sequences from five species, 25,458 multispecies alignments, and 32,417 SNPs, from a previous study (Loh et al. 2008), which applied various criteria to ensure that alignments are allelic and not products of paralogous loci. Sequence data were generated by the Sanger method, allowing the detection of variable sites with an even distribution across the data set and with high confidence (Loh et al. 2008). Examination of these data and subsequent genotyping revealed very low genetic variation, and the persistence of ancestral polymorphism across the Malawi cichlid flock. Molecular genetic analyses across multiple cichlid species are thus highly analogous to within-species polymorphism studies conducted in other organisms (e.g., humans, Chen and Rajewsky 2006; Saunders et al. 2007). Our use of the term “SNP” in this context therefore extends to include variable sites across multiple cichlid species (For more details, see Loh et al. 2008).

miRNA Gene Detection

A database of 623 known teleost precursor miRNA (pre-miRNA) sequences was downloaded from miRBase release 14.0 (Griffiths-Jones et al. 2008). To detect miRNA genes in cichlids, we conducted a BlastN similarity search of these pre-miRNAs against the cichlid genomic sequences as described above, with an *E* value cutoff of 0.001. The BlastN hits were then manually inspected and compared with their query sequences in order to extract adjacent nucleotides that might form part of the pre-miRNA. RNA secondary structure of the cichlid putative miRNA sequences was predicted using Mfold (Zuker 2003) to ensure proper stem-loop folding, and excess bases were trimmed. A reciprocal BlastN of the putative cichlid miRNAs against known teleost miRNAs was performed to identify the cichlid miRNA and to assign orthology. Multiple sequence alignments of the putative cichlid miRNAs and their orthologs were then generated using ClustalW (Larkin et al. 2007). Mutations in the alignments were marked and counted based on the region (mature miRNA, stem, or loop) where they reside.

3'-UTR Annotation

Cichlid genomes have yet to be fully sequenced and annotated; therefore, we first annotated cichlid 3'-UTRs from partial genomic sequence. We chose to work with genomic and not transcript sequences because our ultimate goal was to map SNPs to putative miRNA targets found within 3'-UTRs (below); SNP data exist for genome survey sequences (Loh et al. 2008) but not for the small number of publicly available cichlid expressed sequence tags (ESTs). Sequences used to annotate cichlid 3'-UTRs include *Fugu rubripes*, *Tetraodon nigroviridis*, *Oryzias latipes*, *Gasterosteus aculeatus*, and *Danio rerio* proteins (98,037 entries) downloaded from Ensembl version 56, all "Actinopterygii" proteins (41,746 entries) from Refseq release 39, and all "Eukaryota" proteins (158,696 entries) from UniProtKB/Swiss-Prot release 2010_02 databases.

We applied the TBlastN algorithm with an *E* value cutoff of 1×10^{-10} to identify similarity between the protein sequences above and the cichlid multispecies alignments (Loh et al. 2008). High-scoring segment pairs (HSPs) of the TBlastN output with lengths of at least 30 amino acids were parsed and retained, and in cases where the end position of a HSP query was found to be within three amino acids from the known 3' end of the full-length query protein, it was deemed that a corresponding cichlid coding region might also have ended in this region. We further looked within the ± 9 nucleotide region of the HSP subject (i.e., cichlid) end to confirm that a stop codon was indeed present and in frame with codon phase of the HSP. Cichlid 3'-UTRs were thus annotated to begin at the next nucleotide beyond the stop codon and presumed to continue for 500 nucleotides in length. This approximation of 3'-UTR length was based on a calculation of the mean 3'-UTR length in zebra fish (513 nucleotides), as annotated by Ensembl. During our work on this project, an additional $\sim 56,000$ unique ESTs were released for the tilapia cichlid, roughly 10–15 My divergent from the Malawi assemblage (Lee et al. 2010). Comparing our annotations with these data, we observed that 66% of our predicted 3'-UTRs showed significant similarity (*E* value $< 1 \times 10^{-5}$) to ESTs.

miRNA Target Prediction

A total of 249 unique mature miRNA sequences, consolidated from the 623 known pre-miRNAs from *Fugu*, *Tetraodon*, and *Danio* (miRBase) and the 100 derived from miRNA loci in cichlids (this study), was used for the prediction of target sites on annotated cichlid 3'-UTRs. The target prediction algorithm (hereby termed the SeedMatch algorithm) was written in Perl programming language, implementing the seed-matching requirements similar to that of Target ScanS (Lewis et al. 2005): namely, 1) a six nucleotide Watson–Crick complementary match between miRNA and mRNA at positions 2–7 of the miRNA plus 2) an anchor

of either an adenosine at the mRNA target aligned to miRNA position 1 and/or a Watson–Crick match at position 8 of the miRNA.

Conservation of predicted cichlid miRNA target sites in other fish species was determined by 1) generating multiple sequence alignments (MLAGAN; Brudno et al. 2003) of cichlid 3'-UTRs and their orthologs (when determined) in puffer fishes, medaka, stickleback, and zebra fish, 2) applying the SeedMatch algorithm separately to each sequence in the multiple alignment to identify target sites, and 3) calling a cichlid target site conserved, when an identical target site was found in at least one other fish at a location within 50 nucleotide positions along the alignment. We defined conservation as such, in contrast to other target prediction strategies requiring strict conservation across multiple species (Alexiou et al. 2009; Barbato et al. 2009) for two reasons. First, the fishes with complete genome sequences noted above are all at least 100 My divergent from Malawi cichlids. Second, fish genomes are generally more divergent with greater neutral nucleotide substitution rates compared with mammals (Brunet et al. 2006). The latter consideration influences the degree of target conservation observed between species and also our initial task of generating robust multiple sequence alignments.

Target SNP Density Calculations

Subsequent to predicting miRNA target sites on 3'-UTRs, we mapped SNPs to these same data (Loh et al. 2008). For statistical analysis of observed SNP densities in predicted miRNA targets, we obtained a distribution of randomized target SNP densities by running 1,000 simulations that permute the occurrence of SNPs along the 3'-UTRs. In each simulated run, every empirical SNP in the 3'-UTRs was shuffled to a random position maintaining the same trinucleotide sequence (i.e., the SNP position itself and the nucleotides immediately before and after). For example, a G[AT]C trinucleotide where [A/T] represents the SNP would be shifted to a random GAC or GTC position. The "randomized" target SNP density was then calculated for each run. This simulation strategy controls for neighbor-dependent mutation rates and has been used previously to investigate SNP densities in miRNA target sites (Hiard et al. 2010).

3'-UTR Resequencing, Alignment, and Target Prediction

The analyses described above using data from Loh et al. (2008) allow us to identify cichlid miRNAs, their putative targets, and to calculate SNP densities in target sequence. However, because those data do not represent full genomes from the five species sequenced, alignments of orthologous sequence rarely contain more than three species (Loh et al. 2008). To better understand, evolutionary processes acting on putative cichlid miRNA target sequences, we

resequenced annotated 3'-UTRs in a diverse and standardized collection of species. Polymerase chain reaction primers were designed (supplementary file 3, Supplementary Material online) and used for amplification and sequencing of a subset of annotated 3'-UTRs from the genomic DNA of eight individuals: *Labeotropheus fuelleborni* (LF), *Melanochromis auratus* (MA), and *Maylandia zebra* (MZ) are members of the rock-dwelling mbuna lineage; *Tyrannochromis maculiceps* (TM), *Docimodus evelynae* (DE), *Nimbochromis polystigma* (NP), and *Mchenga conophoros* (MC) belong to a sister lineage of pelagic and sand-dwelling species (henceforth termed non-mbuna); and *Rhamphochromis esox* (RE) represents an early diverging, deepwater group within the radiation (pictures at <http://www.malawicichlids.com>). The individuals of LF, MA, MZ, MC, and RE were those surveyed and sequenced by the JGI (Loh et al. 2008). Sequences were aligned using ClustalW (Larkin et al. 2007), from which polymorphic positions were identified at locations exhibiting at least seven species depth of coverage (supplementary file 5, Supplementary Material online). We applied the target site prediction algorithms and SNP density calculations to these data as described above. We also carried out additional analyses, described below, with these resequenced data.

MAFs of SNPs in Resequenced 3'-UTRs

We calculated the MAF of each SNP (in and out of putative miRNA targets) identified in the resequenced data set. We then compared these MAF distributions with a neutral expectation. From a set of 70 nongenic SNPs typed across a diverse mix (183 individuals, 62 species) of Lake Malawi cichlids (Cichlid Genome Consortium, Broad Institute), we randomly sampled eight individuals to match our resequenced 3'-UTR data set (three mbuna, four non-mbuna, and one deepwater species) and calculated the allele frequency distribution of the sample. This process was repeated 1,000 times to approximate a neutral distribution of allele frequencies and the 95% confidence intervals at each allele frequency. Because we sequenced and resampled eight individuals or 16 total alleles, the empirical and simulated allele frequency data are largely discrete, with the majority of observations falling around multiples of 1/8 (0.125). Therefore, bins were set around multiples of 0.125, and bin edges fall at the midpoint of consecutive bins; for example, the first bin edge (0.1675) is the midpoint between 0.125 and 0.25. Z-tests were implemented within each allele frequency bin to detect significant shifts in the proportion of SNPs exhibiting that particular range of MAFs between empirical and resampled neutral distributions.

Genetic Differentiation of High-MAF Target SNPs in Resequenced 3'-UTRs

We observed that SNPs in predicted targets exhibited higher MAFs than expected under neutrality. To test, whether these

high-MAF ($31.25 < \text{MAF} < 50\%$) miRNA target SNPs exhibited greater genetic differentiation among Malawi lineages than expected under neutrality, we generated 1,000 sets of matching "neutral" genotype data using the same nongenic SNP data set and sampling strategy as described above. For each set of genotypic data, we calculated for each SNP the 1) overall population, 2) mbuna, and 3) non-mbuna allele frequencies, where each allele frequency value lies between 0 and 1. We defined an SNP as displaying clear lineage-specific differentiation when the difference in mbuna and non-mbuna allele frequencies was equal or greater than 0.75, and hence calculated the proportion of high-MAF SNPs that were well differentiated between lineages. Values were aggregated for the 1,000 data sets to obtain a distribution from which a Z-test was used to determine the statistical significance of our observed data.

Results

miRNA Prediction

We used a reference set of 623 known teleost pre-miRNA sequences from *Fugu*, *Tetraodon*, and *Danio*, obtained from miRBase release 14.0 (Griffiths-Jones et al. 2008), in a similarity search (see Materials and Methods) against a database of 304,310 cichlid genomic sequences (Loh et al. 2008). We manually curated the similarity hits to extract putative cichlid pre-miRNAs and confirmed that they were able to fold into the secondary stem-loop structure necessary for miRNA biogenesis (Bartel 2004). This resulted in the identification of 100 distinct cichlid pre-miRNA genes (supplementary file 2, Supplementary Material online) that produce 87 unique mature miRNAs.

We compared cichlid pre-miRNA loci with their orthologues in other fish species and found a total of 1002 of 6422 nucleotide positions where substitutions had occurred. This results in an overall nucleotide divergence of 0.156 (variable sites/nucleotide positions). When the pre-miRNAs were divided into mature miRNA, stem, and loop regions (fig. 1A), we observed nucleotide divergences of 0.015, 0.172, and 0.485, respectively (fig. 1B), with no mutations found in the miRNA seeds. A similar trend of region-specific variation holds for the subset of substitutions, where cichlids exhibit a different nucleotide than all other species; a divergence of 0.008, 0.060, and 0.185 at the mature miRNA, stem, and loop regions, respectively (fig. 1B).

Polymorphism in Cichlid miRNA Targets

To study genetic variation in putative cichlid miRNA targets, we mapped SNPs (Loh et al. 2008) to target sequences predicted to fall within 3'-UTRs. We first annotated 731 cichlid 3'-UTRs (supplementary file 4, Supplementary Material online) that contained 367 SNPs (0.28% SNP density). To direct our computational prediction of targets, we used 249

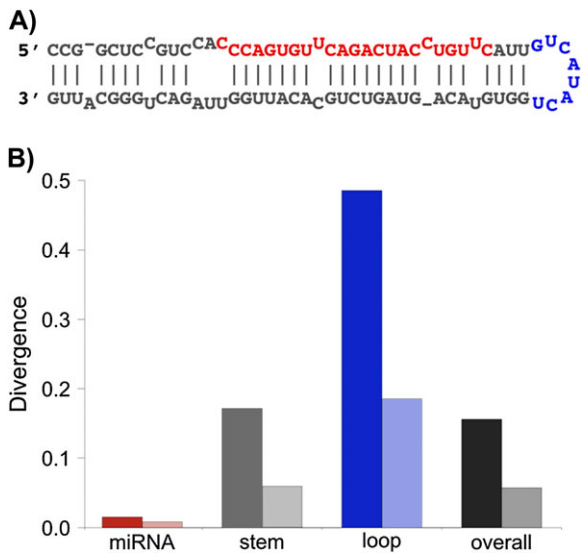


FIG. 1.—Evolutionary divergence in pre-miRNA sequences. (A) An example of predicted stem–loop secondary structure for a cichlid miRNA (lfu-mir-199-1 shown here), classified into separate regions for analysis. Nucleotide symbols are colored red for the mature miRNA region, blue for the loop region, and gray for the stem region excluding the mature miRNA. Vertical bars represent Watson–Crick or G:U base-pairing matches. (B) Distribution of divergence across different regions of the pre-miRNA. Bar colors correspond to the regions defined in A, with black representing the divergence over the entire molecule. Solid-colored bars are calculated from all observed variable sites. Shaded bars are calculated from variable sites where cichlids displayed a different nucleotide than all other species.

unique mature miRNAs, derived from miRNA loci in cichlids (above) as well as known miRNAs from other fish species *Fugu*, *Tetraodon*, and *Danio*. These miRNAs are highly conserved among vertebrates; 86% are in miRNA families that extend outside of fishes. Note that the 100 cichlid miRNAs we identified here (above) possess identical seed sequences to their fish orthologues; this justifies our use of additional

fish miRNAs, conserved among vertebrates but not yet identified in cichlids (see below), to facilitate target prediction.

Putative miRNA binding sites in 3′-UTR sequences were predicted using a Perl script written to implement a “Seed-Match” algorithm incorporating rules similar to those of TargetScanS (Lewis et al. 2005). Briefly, 7- and 8-mer target sites were identified that had exact Watson–Crick base pair matches at seed sequences (positions 2–7 counting from the miRNA 5′ end) plus a corresponding base anchor at position 1 and/or 8 (see Materials and Methods).

Considering all putative 3′-UTRs identified from the Loh et al. (2008) data, we detected 6,299 miRNA target sites on 719 of 731 3′-UTR sequences (an average of 8.62 miRNA target sites per 3′-UTR; table 1). As expected, we observed overlaps among predicted target sites for multiple miRNAs; 13.0% of the total 3′-UTR length (39,660 nucleotides) were predicted to be bound by one or more miRNA(s), similar to results reported in human and mouse (Hiard et al. 2010). Seventy-eight SNPs mapped within 17,607 informative bases of miRNA target sites. Thus, the SNP density for miRNA target sites is 0.44%, higher than 1) the average 3′-UTR SNP density (0.28%), 2) the SNP densities of target-flanking sequence (0.21–0.28%), and 3) the average randomized target SNP density of 0.28% (Z-test, $P = 2.41 \times 10^{-6}$; fig. 2A). For reference, the SNP densities of synonymous and replacement coding sites in the same set of data are 0.42% and 0.20%, respectively (Loh et al. 2008).

Enforcing a criterion of target site conservation reduced the size of our data set considerably (see Materials and Methods and below; table 1). We assigned orthology to single genes in other fish genomes for 481 of 731 predicted cichlid 3′-UTRs. Other predicted 3′-UTRs showed similarity to members of gene families or to specific pairs of duplicated loci, but we could not specify reciprocal orthology with confidence. Conserved sites accounted for 21% of cichlid miRNA targets (875 of 4,182), similar to previous study (Friedman et al. 2009; Hiard et al. 2010), and covered

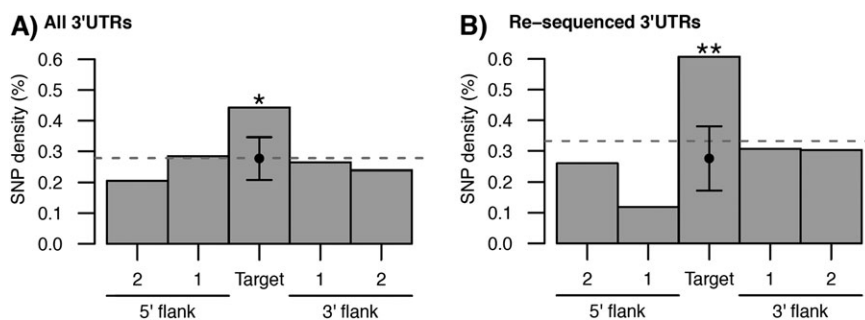


FIG. 2.—SNP densities within computationally predicted miRNA target sites and their flanking regions on data from (A) all predicted 3′-UTRs and (B) select resequenced 3′-UTRs. Flanking regions 1–2 on both 5′ and 3′ ends of target represent successive, nonoverlapping windows of sizes equal to that of the target sites. Dotted lines show the average 3′-UTR SNP density. Filled circle with error bars represents the mean and 95% confidence intervals of SNP densities calculated from 1,000 simulated replicates of randomized SNP shuffling. Asterisk symbols indicate significant deviation from simulated distributions (Z-test, $*P < 10^{-5}$, $**P < 10^{-9}$).

Table 1.

miRNA Target Prediction Results on All Putative and Select Resequenced 3'-UTRs

	All 731 Putative 3'-UTRs			130 Resequenced 3'-UTRs		
	All (731 UTRs)	Conserved Targets (481 UTRs)	Nonconserved Targets (481 UTRs)	All (130 UTRs)	Conserved Targets (124 UTRs)	Nonconserved Targets (124 UTRs)
Number of targets predicted	6,299	875	3,307	1,296	360	639
Number of targets (per 3'-UTR)	8.62	1.82	6.88	9.97	2.90	5.15
Total coverage of targets (nt)	39,660	5,505	21,157	6,602	2,159	4,089
3'UTR coverage by targets (%)	13.0	2.7	10.5	13.7	4.76	9.01
Informative sites within targets ^a (nt)	17,607	2,761	9,355	6,602	2,159	4,089
Number of SNPs in targets	78	8	40	40	7	29
SNP density in targets (%)	0.443	0.290	0.428	0.606	0.324	0.709

^a Only a subset of 3'-UTR positions had multispecies sequence data to determine polymorphism.

only 2.7% of nucleotides in these 481 3'-UTRs. The SNP density in conserved target sites was 0.29%, similar to the average SNP density for flanking and overall 3'-UTRs and within the 95% confidence interval of randomized target SNP densities (supplementary fig. S1, Supplementary Material online).

MAFs and Genetic Differentiation of "Target" SNPs in Resequenced 3'-UTRs

We resequenced a set of 130 3'-UTRs in eight individuals of Malawi cichlid species spanning a range of morphologies and behaviors, representing the three major evolutionary lineages in the lake (Won et al. 2006; Loh et al. 2008). Our rationale here was 2-fold. First, we reasoned that 3'-UTR sequence variation across samples, in and out of putative miRNA target sites, could be examined for the evolutionary signature of natural selection (Chen and Rajewsky 2006; Saunders et al. 2007). Second, in order to better validate predicted miRNA–mRNA interactions against previous literature, we chose certain gene subsets whose molecular functions have been well characterized for interactions with miRNAs (e.g., development, Plasterk 2006; immunity, Xiao and Rajewsky 2009).

From 48,114 base positions of multiple sequence alignments (supplementary file 5, Supplementary Material online), we identified 160 SNPs, an overall SNP density of 0.33%. We then applied the SeedMatch algorithm to these data. SeedMatch targets covered 6,602 total bases, within which we mapped 40 SNPs (table 1). This resulted in an SNP density in predicted targets of 0.606%, higher than the overall average in resequenced data (0.33%), target-flanking sequence (0.12–0.31%), and randomized target SNP densities (0.28%; Z-test, $P = 4.88 \times 10^{-10}$; fig. 2B). Similar to the analysis of all putative 3'-UTRs (above), enforcing a strong conservation criterion for target sites reduced the size of the data set (only 4.8% of 3'-UTR bases are covered by conserved target sites). Conserved sites

accounted for 36% of all targets on 124 cichlid 3'-UTRs; the empirical SNP density in conserved targets was 0.32%, elevated from flanking sequence but similar to the overall 3'-UTR and randomized target SNP densities (supplementary fig. S1, Supplementary Material online).

Next, we examined the allele frequency distribution of SNPs in predicted miRNA target sites in relation to 3'-UTR nontarget sites compared against a neutral expectation. We approximated a neutral distribution by subsampling from a data set of 70 randomly chosen, nongenic SNPs typed in a diverse mix of Lake Malawi cichlids. Significant departure from a neutral distribution of allele frequencies might be indicative of natural selection (Chen and Rajewsky 2006; Drake et al. 2006; Sethupathy et al. 2008). Notably, allele frequencies at nontarget 3'-UTR SNPs did not depart from the neutral distribution (nearly 80% of polymorphisms exhibit minor alleles that are relatively rare) but predicted target SNPs differed significantly, with a bias toward high MAFs (fig. 3; supplementary fig. S2, Supplementary Material online).

We asked if high-MAF SNPs in predicted miRNA targets were differentiated among lineages (i.e., mbuna vs. non-mbuna) to a degree beyond expectation under neutrality. We found that a significantly elevated proportion (86%) of high-MAF ($31.25 < \text{MAF} < 50\%$) target SNPs exhibit genetic differentiation between Malawi evolutionary groups (Z-test, $P = 9.32 \times 10^{-7}$). Predicted miRNA–gene interactions, highlighting evolutionarily differentiated SNPs, are shown in figure 4 and as discussed below.

Discussion

Lake Malawi cichlids have evolved in a brief evolutionary window. Their genomes are highly similar and segregate ancestral polymorphism. For comparison, nucleotide diversity across the flock (0.26%, Loh et al. 2008) is less than that observed among laboratory strains of the zebra fish (0.48%, Guryev et al. 2006), comparable to that of chimpanzees (0.24%, Fischer et al. 2004) and humans (0.11%,

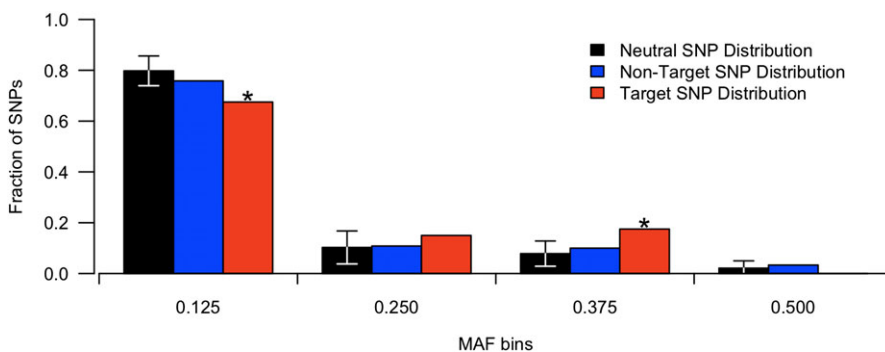


Fig. 3.—Comparison of MAF distributions. 3'-UTR miRNA target SNPs are colored in red, nontarget SNPs in blue, and nongenic (neutral) SNPs in black. Error bars represent the 95% confidence interval of the neutral expectation. Asterisk symbols indicate significant deviation from neutral expectation within each bin (Z-test, * $P < 10^{-4}$).

International Hapmap Consortium 2007), and contrasts against the ~1.2% divergence between chimps and humans (King and Wilson 1975; Chen and Li 2001). It is notable, then, that the range of variation across Malawi species for many phenotypes (body size, tooth, and taste bud number) spans an order of magnitude and that the diversity of other traits (color pattern, feeding and breeding biology, brain organization) is comparable to that observed in other vertebrate taxonomic orders. The cichlid system is thus a model of the genotype to phenotype mapping function (Streelman et al. 2007), with speculation revolving around the rapid evolution of novelty. Here, we test the hypothesis that evolutionary divergence of miRNAs and/or

their binding sites may have contributed to the diversification of species (Plasterk 2006).

Cichlid miRNA Target Sites Exhibit Elevated SNP Densities

We identified 100 distinct miRNA loci in the genomes of cichlid fishes. The mature miRNAs encoded by these loci are highly conserved among fishes (fig. 1B). The trend of higher divergence in stems and loops (vs. the mature miRNA) has been observed in other species (Hertel et al. 2006) and may be indicative of purifying selection against change to the functional component of the miRNA molecule (and/or a relaxation of constraint at stems and loops). The number

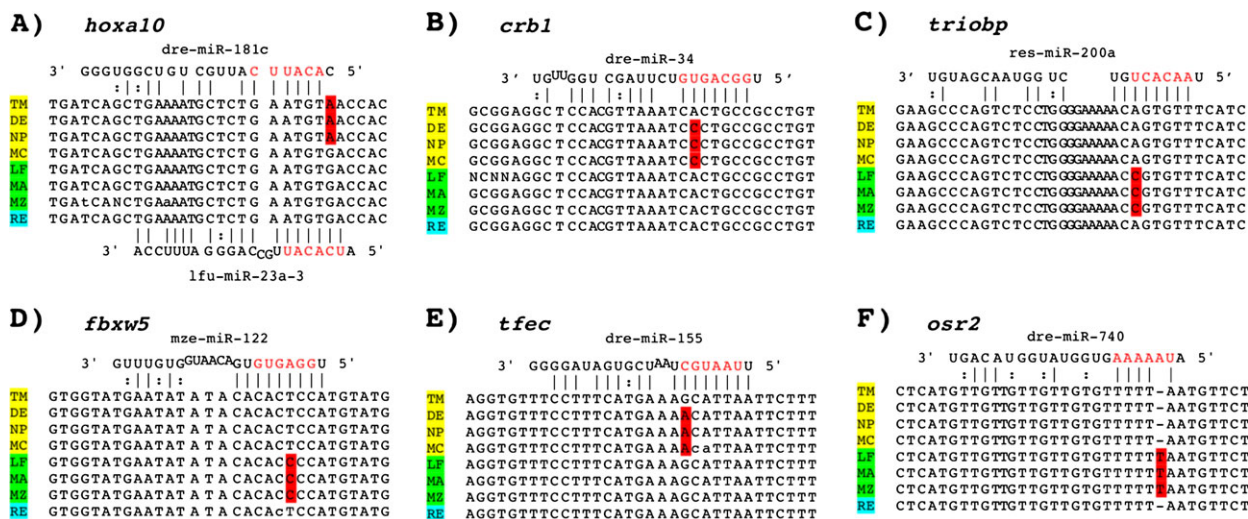


Fig. 4.—Multiple sequence alignments of several miRNA targets containing differentiated SNPs, discussed in the text. Red blocks indicate SNP minor alleles. Dashes represent gaps in sequence (indel in *osr2*). miRNAs predicted to bind to the target are shown, with the seed region in red font. Vertical bars represent Watson-Crick base pairing and colors represent G:U base pairing. Raised and lowered nucleotides illustrate bulges in the predicted miRNA binding. TM, *Tyrannochromis maculiceps*; DE, *Docimodus evelynae*; NP, *Nimbochromis polystigma*; MC, *Mchenga conophoros*; LF, *Labeotropheus fuelleborni*; MA, *Melanochromis auratus*; MZ, *Maylandia zebra*; and RE, *Rhamphochromis esox*. Yellow, green, and blue boxes over abbreviated species names represent non-mbuna, mbuna, and deepwater lineages, respectively.

of miRNAs we identified is likely to be an incomplete count, as the available cichlid genomic resources used here comprise only about 32% coverage of the cichlid genome (Loh et al. 2008). As a reference, there are 360 zebra fish (characterized from an assembled genome and by deep RNA sequencing; Wienholds et al. 2005; Soares et al. 2009) and 132 puffer fish miRNAs in miRBase.

Predicted miRNA target sites, located in the 3'-UTRs of cichlid genes, showed elevated SNP densities when compared with flanking regions, the overall 3'-UTR average and randomized simulations that account for nucleotide composition (table 1 and fig. 2). For a more restricted set of evolutionarily conserved targets, SNP densities were not distinguishable from those in flanks, the overall 3'-UTR average and simulation values. This trend held in both the genome-wide 3'-UTR data set and the directed set of resequenced 3'-UTRs. Our observation of elevated or equivalent SNP densities in both conserved and nonconserved miRNA targets runs counter to results from previous study within humans, where average SNP density in predicted target sites (both conserved and nonconserved) was reduced compared with flanking regions (Chen and Rajewsky 2006; Saunders et al. 2007).

miRNA Target Sites Show the Signature of Divergent Natural Selection

The observation of increased SNP density at predicted miRNA target sites does not provide conclusive information about the evolutionary forces shaping this pattern; for instance, even though the SNP density of predicted targets is high within the context of 3'-UTR sequence, minor alleles at variable sites could be rare. We therefore resequenced a collection of 3'-UTRs in a standard set of species and designed a test to evaluate the allele frequency distribution of 1) SNPs predicted in miRNA binding sites and 2) other 3'-UTR nontarget SNPs, against a neutral expectation. This test is conceptually similar to the derived allele frequency approach (Chen and Rajewsky 2006; Drake et al. 2006; Sethupathy et al. 2008). However, because Lake Malawi cichlid fishes retain ancestral polymorphism that may predate the species flock (Loh et al. 2008) we have not attempted to designate ancestral versus derived alleles.

We found that whereas the allele frequency distribution of nontarget SNPs in 3'-UTRs was not different than the neutral expectation, the distribution of predicted miRNA target SNPs was biased toward high MAFs (fig. 3). In addition, we observed that 86% of putative miRNA target SNPs with high MAFs showed a clear pattern of evolutionary divergence between major Malawi lineages (fig. 4 and below). To put this in greater context, we have previously observed that <5% of haphazardly chosen SNPs are outliers for genetic differentiation in a large sample of mbuna versus non-mbuna (Loh et al. 2008). The alternative that the differentiated polymorphisms we highlight in figure 4 are

not in fact in miRNA targets but are each physically linked to other, as yet unidentified nucleotide sites, is unlikely because it would require that we happened upon these unidentified sites in six independent loci through the sole discovery operation of searching for miRNA targets.

Taken together, our observations of 1) elevated SNP densities, 2) a bias toward high MAFs, and 3) the pattern of genetic differentiation among lineages for high-MAF SNPs suggest that select miRNA binding sites have experienced divergent selection during the evolution of the Lake Malawi species flock.

Differentiated SNPs in miRNA Targets are Biologically Relevant

A secondary goal of our resequencing project was to investigate putative miRNA binding site polymorphism in gene sets whose molecular functions have been well-studied vis-à-vis miRNAs. We reasoned that such data would add biological plausibility to our computational predictions and population genetic analyses. Figure 4 displays examples of high-MAF SNPs, genetically differentiated among Malawi cichlid lineages, mapped to miRNA target sites in 3'-UTRs. These examples represent miRNA–gene pairs supported by previous research in humans and other model organisms.

The interplay between miRNAs and Hox gene riboregulation is well known (Yekta et al. 2008). We predict an association between two miRNAs, miR-181c and miR-23a, which share a target site SNP in the cichlid *hoxa10* 3'-UTR (fig. 4A); this target site in *hoxa10* is conserved between cichlid and stickleback. The SNP differentiates non-mbuna predators (TM, DE, NP) from other species. miR-181 is known to target mouse *Hoxa11* (a Hox cluster family member of *hoxa10*) during muscle differentiation (Naguibneva et al. 2006); fish *hoxa10* genes are expressed in paired fins and associated musculature (Ahn and Ho 2008). Recently, it has been shown that miR-181 is upregulated whereas miR-23 is downregulated in mouse leg muscle during endurance exercise (Safdar et al. 2009). These data raise the possibility that a single SNP modulates the miRNA riboregulation of Hox-mediated fin muscle development and regeneration in Lake Malawi predators.

We highlight two miRNA–gene pairs that may modify sensory modalities among Lake Malawi cichlids. We predict differential binding of miR-34 to cichlid *crb1* (fig. 4B), a member of the Crumbs protein complex. *crb1* contributes to photoreceptor morphogenesis and sensitivity, mutations cause retinal degeneration in humans, mice, and flies (Bulgakova and Knust 2009). miR-34 is expressed in neural tissue (including the optic tectum) of larval and adult zebra fish (Kapsimali et al. 2007), also in the retina of embryonic and adult mice (Arora et al. 2010). This association is of particular interest given the vast literature implicating the role of vision in Malawi cichlid ecology, mate choice, and evolution (Carleton et al. 2008). Next, we predict that the

TRIO and F-actin binding protein (*triobp*) is differentially bound by miR-200a (fig. 4C). *triobp* functions in the hair cell cilia of the inner ear (Kitajiri et al. 2010), mutations result in nonsyndromic hearing loss (Shahin et al. 2006). miR-200a is expressed in sensory epithelia, including those of the inner ear of zebra fish, chicken, and mouse (Soukup 2009). Recent reports have linked hearing to mate choice and communication in East African cichlids (Simões et al. 2008; Verzijden et al. 2010).

Two SNPs are predicted to affect binding of miRNAs to genes involved in immune response (Xiao and Rajewsky 2009). *fbxw5* (fig. 4D) is a F-box protein with a role in interleukin signaling (Minoda et al. 2009); a T ↔ C SNP differentiated among Malawi cichlids is predicted to modulate binding of miR-122, a liver-specific miRNA (Sarasin-Filipowicz et al. 2009; Soares et al. 2009). The miR-122 binding site in *fbxw5* is conserved between cichlid and medaka. Second, *tfec* (fig. 4E) is a macrophage-restricted basic helix-loop-helix transcription factor, also involved in interleukin signaling (Rehli et al. 2005). We predict that a differentiated G ↔ A SNP modifies binding of miR-155, a well-known regulator of immune function (O'Connell et al. 2009).

Finally, our data may be useful to identify new interactions between miRNAs and genes of interest. For example, we predict that an indel in the 3'-UTR of cichlid *osr2* should differentially regulate binding of miR-740 in mbuna cichlids (LF, MA, MZ) versus others (fig. 4F). *Osr2* restricts the teeth of mice to a single row (Zhang et al. 2009), among other functions in the craniofacial skeleton. Tooth row number is highly variable among cichlid species (Fraser et al. 2008). miR-740 is poorly understood (Kloosterman et al. 2006); our data suggest it may play a role in craniofacial development.

Conclusion

Biologists recognize that 5' *cis*-acting mutations regulate gene expression and contribute to phenotypic evolution (King and Wilson 1975; Wray 2007; Carroll 2008). Correspondingly, studies have reported the signature of diversifying selection on population genetic variants in computationally predicted 5' promoter elements (Haygood et al. 2007; Sethupathy et al. 2008). The situation is different for 3'-UTRs. miRNAs and their binding sites collaborate as post-transcriptional capacitors to canalize the transcriptome (Carrington and Ambrose 2003; Stark et al. 2005). Evidence suggests that both miRNAs and their target sequences in 3'-UTRs evolve under purifying selection (Chen and Rajewsky 2006; Saunders et al. 2007). Metazoan cis-trons may therefore have evolved for transcriptional exploration at 5' promoters, with post-transcriptional safeguards encoded at the back.

We provide evidence that the evolution of miRNA binding sites may play a role in evolutionary diversification. We dem-

onstrate that 1) computationally predicted miRNA targets in cichlid 3'-UTRs harbor elevated SNP densities, 2) a greater frequency of polymorphic sites in predicted targets have high MAFs compared with a neutral expectation, and 3) these sites are often genetically differentiated among Malawi lineages.

It has been argued that polymorphisms in miRNA target sites are deleterious within species because even single base mismatches (especially to the seed) can abrogate binding and disrupt riboregulation (Clop et al. 2006; Sethupathy et al. 2008; Mencía et al. 2009). We suggest that mutations in 3'-UTRs where miRNAs may bind, whether breaking transcriptome canalization or introducing new regulation, may contribute to phenotypic differentiation among rapidly evolving lineages. Further analyses, with fully annotated and assembled cichlid genomes (<http://www.genome.gov/10002154>), deeper genotyping, next-generation miRNA, and miRNA target prediction algorithms (Chaudhuri and Chatterjee 2007; Barbato et al. 2009), and experimental validation of predicted miRNAs and their interactions with target genes (Kuhn et al. 2008; Sethupathy and Collins 2008) will reveal additional intricacies of miRNA riboregulation and evolution.

Supplementary Material

Supplementary files 1–5 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

We thank Craig Albertson, Greg Gibson and two anonymous reviewers for comments on previous drafts of this manuscript and Michael Norsworthy for experimental assistance. This work was supported by the National Science Foundation (IOS 0546423) and the National Institutes of Health (R01 DE019637).

Literature Cited

- Ahn D, Ho RK. 2008. Tri-phasic expression of posterior Hox genes during development of pectoral fins in zebrafish: implications for the evolution of vertebrate paired appendages. *Dev Biol.* 322(1): 220–233.
- Albertson RC, Strelman JT, Kocher TD, Yelick PC. 2005. Integration and evolution of the cichlid mandible: the molecular basis of alternative feeding strategies. *Proc Natl Acad Sci U S A.* 102(45):16287–16292.
- Alexiou P, Maragkakis M, Papadopoulos GL, Reczko M, Hatzigeorgiou AG. 2009. Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics.* 25(23):3049–3055.
- Arora A, et al. 2010. Prediction of microRNAs affecting mRNA expression during retinal development. *BMC Dev Biol.* 10:1.
- Barbato C, et al. 2009. Computational challenges in miRNA target predictions: to be or not to be a true target? *J Biomed Biotechnol.* 2009(803069):1–9.
- Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism and function. *Cell.* 116(2):281–297.

- Brudno M, et al. 2003. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* 13(4): 721–731.
- Brunet FG, et al. 2006. Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. *Mol Biol Evol.* 23(9): 1808–1816.
- Bulgakova NA, Knust E. 2009. The Crumbs complex: from epithelial-cell polarity to retinal degeneration. *J Cell Sci.* 122(Pt 15):2587–2596.
- Carleton KL, et al. 2008. Visual sensitivities tuned by heterochronic shifts in opsin gene expression. *BMC Biol.* 6:22.
- Carrington JC, Ambrose V. 2003. Role of microRNAs in plant and animal development. *Science.* 301(5631):336–338.
- Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell.* 134(1):25–36.
- Chaudhuri K, Chatterjee R. 2007. MicroRNA detection and target prediction: integration of computational and experimental approaches. *DNA Cell Biol.* 26(5):321–337.
- Chen FC, Li WH. 2001. Genomic divergences between humans and homonoids and the effective population size of the common ancestor of humans and chimpanzees. *Am J Hum Genet.* 68(2): 444–456.
- Chen K, Rajewsky N. 2006. Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet.* 38(12):1452–1456.
- Clop A, et al. 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat Genet.* 38(7):813–818.
- Drake JA, et al. 2006. Conserved noncoding sequences are selectively constrained and not mutation cold spots. *Nat Genet.* 38(2): 223–227.
- Eberhart JK, et al. 2008. MicroRNA mirn140 modulates Pdgf signalling during palatogenesis. *Nat Genet.* 40(3):290–298.
- Farh KK, et al. 2005. The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science.* 310(5755):1817–1821.
- Fischer A, Wiebe V, Paabo S, Przeworski M. 2004. Evidence for a complex demographic history of chimpanzees. *Mol Biol Evol.* 21(5):799–808.
- Fraser GJ, Bloomquist RF, Streelman JT. 2008. A periodic pattern generator for dental diversity. *BMC Biol.* 6:32.
- Friedman RC, Farh KK, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19(1): 92–105.
- Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* 36:D154–D158 (Database issue).
- Grimson A, et al. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell.* 27(1):91–105.
- Grimson A, et al. 2008. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature.* 455(7217):1193–1197.
- Guryev V, et al. 2006. Genetic variation in the zebrafish. *Genome Res.* 16(4):491–497.
- Haygood R, Fedrigo O, Hanson B, Yokoyama KD, Wray GA. 2007. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. *Nat Genet.* 39(9):1140–1144.
- Hertel J, et al. 2006. The expansion of the metazoan microRNA repertoire. *BMC Genomics.* 7:25.
- Hiard S, Charlier C, Coppieters W, Georges M, Baurain D. 2010. Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates. *Nucleic Acids Res.* 38:D640–D651.
- International HapMap Consortium. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 449(7164):851–861.
- Kapsimali M, et al. 2007. MicroRNAs show a wide diversity of expression profiles in the developing and mature central nervous system. *Genome Biol.* 8(8):R173.
- King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science.* 188(4184):107–116.
- Kitajiri S, et al. 2010. Actin-bundling protein TRIOBP forms resilient rootlets of hair cell stereocilia essential for hearing. *Cell.* 141(5): 786–798.
- Kloosterman WP, et al. 2006. Cloning and expression of new microRNAs from zebrafish. *Nucleic Acids Res.* 34(9):2558–2569.
- Kocher TD. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. *Nat Rev Genet.* 5(4):288–298.
- Kornfield I, Smith PF. 2000. African cichlid fishes: model systems for evolutionary biology. *Ann Rev Ecol Evol Syst.* 31:163–196.
- Kuhn DE, et al. 2008. Experimental validation of miRNA targets. *Methods.* 44(1):47–54.
- Larkin MA, et al. 2007. ClustalW and ClustalX version 2. *Bioinformatics.* 23(21):2947–2948.
- Lee BY, et al. 2010. An EST resource for tilapia based on 17 normalized libraries and assembly of 116,899 sequence tags. *BMC Genomics.* 11:278.
- Levine M, Tjian R. 2003. Transcription regulation and animal diversity. *Nature.* 424(6945):147–151.
- Lewis BP, Burge CB, Bartel DP. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell.* 120(1):15–20.
- Loh YH, et al. 2008. Comparative analysis reveals signatures of differentiation amid genomic polymorphism in Lake Malawi cichlids. *Genome Biol.* 9(7):R113.
- Mencia A, et al. 2009. Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nat Genet.* 41(5):609–613.
- Miller CT, et al. 2007. cis-Regulatory changes in Kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell.* 131(6):1179–1189.
- Minoda Y, Sakurai H, Kobayashi T, Yoshimura A, Takaesu G. 2009. An F-box protein, FBXW5, negatively regulates TAK1 MAP3K in the IL-1beta signaling pathway. *Biochem Biophys Res Commun.* 381(3):412–417.
- Moran P, Kornfield I. 1993. Retention of ancestral polymorphism in the Mbuna species flock of Lake Malawi. *Mol Biol Evol.* 10:1015–1029.
- Naguibneva I, et al. 2006. The microRNA miR-181 targets the homeobox protein Hox-A11 during mammalian myoblast differentiation. *Nat Cell Biol.* 8(3):278–284.
- O'Connell RM, Chaudhuri AA, Rao DS, Baltimore D. 2009. Inositol phosphatase SHIP1 is a primary target of miR-155. *Proc Natl Acad Sci U S A.* 106(17):7113–7118.
- Plasterk RH. 2006. Micro RNAs in animal development. *Cell.* 124(5): 877–881.
- Rehli M, et al. 2005. Transcription factor Tfec contributes to the IL-4-inducible expression of a small group of genes in mouse macrophages including the granulocyte colony-stimulating factor receptor. *J Immunol.* 174(11):7711–7722.
- Rockman MV, Stern DL. 2008. Tinker where the tinkering's good. *Trends Genet.* 24(7):317–319.
- Safdar A, Abadi A, Akhtar M, Hetingga BP, Tarnopolsky MA. 2009. miRNA in the regulation of skeletal muscle adaptation to acute endurance exercise in C57Bl/6J male mice. *PLoS One.* 4(5):e5610.
- Salzburger W, Mack T, Verheyen E, Meyer A. 2005. Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. *BMC Evol Biol.* 5(1):17.

- Sarasin-Filipowicz M, Krol J, Markiewicz I, Heim MH, Filipowicz W. 2009. Decreased levels of microRNA miR-122 in individuals with hepatitis C responding poorly to interferon therapy. *Nat Med.* 15(1):31–33.
- Saunders MA, Liang H, Li WH. 2007. Human polymorphism at microRNAs and microRNA target sites. *Proc Natl Acad Sci U S A.* 104(9):3300–3305.
- Sethupathy P, Collins FS. 2008. MicroRNA target site polymorphisms and human disease. *Trends Genet.* 24(10):489–497.
- Sethupathy P, Giang H, Plotkin JB, Hannenhalli S. 2008. Genome-wide analysis of natural selection on human cis-elements. *PLoS One.* 3(9):e3137.
- Shahin H, et al. 2006. Mutations in a novel isoform of *TRIOBP* that encodes a filamentous-actin binding protein are responsible for DFNB28 recessive nonsyndromic hearing loss. *Am J Hum Genet.* 78(1):144–152.
- Simões JM, Duarte IS, Fonseca PJ, Turner GF, Amorim MC. 2008. Courtship and agonistic sounds by the cichlid fish *Pseudotropheus zebra*. *J Acoust Soc Am.* 124(2):1332–1338.
- Soares AR, et al. 2009. Parallel DNA pyrosequencing unveils new zebrafish microRNAs. *BMC Genomics.* 10:195.
- Soukup GA. 2009. Little but loud: small RNAs have a resounding affect on ear development. *Brain Res.* 1277:104–114.
- Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM. 2005. Animal microRNA confer robustness to gene expression and have a significant impact on 3' UTR evolution. *Cell.* 123(6):1133–1146.
- Streelman JT, Albertson RC, Kocher TD. 2003. Genome mapping of the orange blotch colour pattern in cichlid fishes. *Mol Ecol.* 12(9):2465–2471.
- Streelman JT, Peichel CL, Parichy DM. 2007. Developmental genetics of adaptation in fishes: the case for novelty. *Ann Rev Ecol Evol Syst.* 38:655–681.
- Sucena E, Delon I, Jones I, Payre F, Stern DL. 2003. Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. *Nature.* 424(6951):935–938.
- Sylvester JB, et al. 2010. Brain diversity evolves via differences in patterning. *Proc Natl Acad Sci U S A.* 107(21):9718–9723.
- Verzijden MN, et al. 2010. Sounds of male Lake Victoria cichlids vary within and between species and affect female mate preferences. *Behav Ecol.* 21(3):548–555.
- Wheeler BM, et al. 2009. The deep evolution of metazoan microRNAs. *Evol Dev.* 11(1):50–68.
- Wienholds E, et al. 2005. MicroRNA expression in zebrafish embryonic development. *Science.* 309(5732):310–311.
- Won YJ, Sivasundar A, Wang Y, Hey J. 2005. On the origin of Lake Malawi cichlid species: a population genetic analysis of divergence. *Proc Natl Acad Sci U S A.* 102(1 Suppl):6581–6586.
- Won YJ, Wang Y, Sivasundar A, Raincrow J, Hey J. 2006. Nuclear gene variation and molecular dating of the cichlid species flock of Lake Malawi. *Mol Biol Evol.* 23(4):828–837.
- Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet.* 8(3):206–216.
- Xiao C, Rajewsky K. 2009. MicroRNA control in the immune system: basic principles. *Cell.* 136(1):26–36.
- Yekta S, Tabin CJ, Bartel DP. 2008. MicroRNAs in the Hox network: an apparent link to posterior prevalence. *Nat Rev Genet.* 9:789–796.
- Zhang Z, Lan Y, Chai Y, Jiang R. 2009. Antagonistic actions of *Msx1* and *Osr2* pattern mammalian teeth into a single row. *Science.* 323(5918):1232–1234.
- Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31(13):3406–3415.

Associate editor: Wen-Hsiung Li