Origin, Evolution, and Biological Role of miRNA Cluster in *DLK-DIO3* Genomic Region in Placental Mammals

Evgeny A. Glazov, Sean McWilliam, Wesley C. Barris, and Brian P. Dalrymple

Commonwealth Scientific and Industrial Research Organisation CSIRO Livestock Industries, Queensland Bioscience Precinct QBP, St Lucia, Brisbane, Queensland, Australia

MicroRNAs (miRNAs) are a rapidly growing family of small regulatory RNAs modulating gene expression in plants and animals. In animals, most of the miRNAs discovered in early studies were found to be evolutionarily conserved across the whole kingdom. More recent studies, however, have identified many miRNAs that are specific to a particular group of organisms or even a single species. These present a question about evolution of the individual miRNAs and their role in establishing and maintaining lineage-specific functions and characteristics.

In this study, we describe a detailed analysis of the miRNA cluster (hereafter mir-379/mir-656 cluster) located within the imprinted *DLK-DIO3* region on human chromosome 14. We show that orthologous miRNA clusters are present in all sequenced genomes of the placental (eutherian) mammals but not in the marsupial (metatherian), monotreme (prototherian), or any other vertebrate genomes. We provide evidence that the locus encompassing this cluster emerged in an early eutherian ancestor prior to the radiation of modern placental mammals by tandem duplication of the ancient precursor sequence. The original amplified cluster may have contained in excess of 250 miRNA precursor sequences, most of which now appear to be inactive. Examination of the eutherian genomes showed that the cluster has been maintained in evolution for approximately 100 Myr.

Analysis of genes that contain predicted evolutionarily conserved targets for miRNAs from this cluster revealed significant overrepresentation of the Gene Ontology terms associated with biological processes such as neurogenesis, embryonic development, transcriptional regulation, and RNA metabolism. Consistent with these findings, a survey of the miRNA expression data within the cluster demonstrates a strong bias toward brain and placenta samples from adult organisms and some embryonic tissues.

Our results suggest that emergence of the mir-379/mir-656 miRNA cluster was one of the factors that facilitated evolution of the placental mammals. Overrepresentation of genes involved in regulation of neurogenesis among predicted miRNAs targets indicates an important role of the mir-379/mir-656 cluster in this biological process in the placental mammals.

Introduction

MicroRNAs (miRNAs) are small 21–25 nt regulatory RNAs modulating gene expression in animals and plants. In animals, regulation of gene expression by miRNAs is achieved by sequence-specific targeting of the 3′ untranslated regions (UTRs) of messenger RNAs (mRNAs) by the RNA-induced silencing complex that results in translational repression of the protein synthesis (He and Hannon 2004). In the past few years, the number of discovered miRNAs has increased from tens to thousands and is likely to grow further (Griffiths-Jones et al. 2006). Although most of the miRNAs discovered early were found to be highly conserved in evolution, more and more of the newly identified miRNAs are present in only a small group of organisms and in some cases in a single species (Bentwich et al. 2005; Berezikov, Thuemmler, et al. 2006; Berezikov, van Tetering, et al. 2006; Ruby et al. 2007). The functional significance of these evolutionarily divergent miRNAs has not been established experimentally. However, it is hypothesized that these miRNAs might play a role in establishing and maintaining phenotypic diversity between different groups of organisms (Plasterk 2006; Sempere et al. 2006). A few comparative studies have established connections between some miRNAs and evolutionary changes in animal body plan (Tanzer and Stadler 2004; Sempere et al. 2006; Prochnik et al. 2007).

The miRNA cluster mir-379/mir-656 was originally described as 2 families of related repeats adjacent to a small nucleolar RNA cluster located within the imprinted *DLK-DIO3* region on human chromosome 14 (Cavaille et al. 2002). Since then, mature miRNAs derived from most of these repeats have been experimentally identified in mouse, rat, chimpanzee, human, and cow (Houbaviy et al. 2003; Lagos-Quintana et al. 2003; Kim et al. 2004; Seitz et al. 2004; Suh et al. 2004; Bentwich et al. 2005; Berezikov, Thuemmler, et al. 2006; Berezikov, van Tetering, et al. 2006; Coutinho et al. 2007). Currently, the miRBase miRNA database contains 38 human and 35 mouse miRNAs that originate from this cluster, which makes it the largest known miRNA cluster in vertebrates (Griffiths-Jones et al. 2005). In mouse, some of these miRNAs were shown to be expressed as products of a large noncoding transcript named *Mirg* (Seitz et al. 2003).

We used a combination of comparative genomics and bioinformatics approaches to examine the evolutionary history of the mir-379/mir-656 cluster and its function in vertebrate biology and evolution.

Materials and Methods

Sources of Sequences and Assemblies

Draft genome assemblies of armadillo, elephant, and tenrec were produced by the Broad Institute at Massachusetts Institute of Technology and Harvard (http://www.broad.mit.edu/). Draft genome assembly of platypus genome was produced by the Genome Sequencing Center at Washington University School of Medicine in St Louis (http://genome.wustl.edu/genome_group_index.cgi). Draft genome assembly of cow genome was produced by Baylor College of Medicine Sequencing Center (http://www.hgsc.

Table 1
Synteny of the mir-379/mir-656 Cluster

Organism	Genome Size (Gb)	Chromosome, Scaffold, Contig	*DLK1-DIO3* Locus Size (kb)	mir-379/mir-656 Size (kb)
Human	2.8	chr14	826.2	44.7
Chimp	2.8	chr14	846.3	46.5
Mouse	2.5	chr12	818.5	35.4
Rat	2.7	chr6	ND	37.9
Dog	2.4	chr8	726.7	41.4
Cow	2.4	chr21	859.5	42.9
Armadillo	~3.0	Scaffold 5303, scaffold 19965	ND	~38
Elephant	~3.0	Scaffold 4770	ND	ND
Tenrec	~3.0	ND	ND	ND
Possum	3.4	chr1	1603.5	A
Platypus	~3.0	Ultracontig 378	~700	A
Chicken	1.0	chr5	346.3	A
Puffer fish	0.33	chrUn	ND	A

NOTE.—The table summarizes sizes and genomic locations of *DLK1-DIO3* syntenic loci and the miRNAs clusters in different vertebrate genomes. Where whole-genome assemblies are available, the chromosomes bearing *DLK1-DIO3* loci and the miRNA cluster are listed. Genomic scaffold or contig numbers are provided for partially assembled genomes. ND, not determined; A, absent; the miRNA cluster is completely absent in nonplacental vertebrates.

bcm.tmc.edu/projects/bovine/). Human, chimpanzee, mouse, rat, dog, possum, chicken, and puffer fish genomes were produced by their respective genome sequencing consortiums (Lander et al. 2001; Aparicio et al. 2002; Waterston et al. 2002; Consortium 2004, 2005; Gibbs et al. 2004; Lindblad-Toh et al. 2005; Mikkelsen et al. 2007). Unless specified otherwise, sequences of miRNA precursors and mature miRNAs were obtained from the latest release of miRNA registry (miRBase 10.0, August 2007, http://microrna.sanger.ac.uk/sequences/) (Griffiths-Jones et al. 2005, 2006).

Sequence Searches and Analysis

Pairwise whole-genome sequence alignments and "RefSeq" gene annotation data were obtained from UCSC genome browser (http://genome.ucsc.edu/) (Kent et al. 2002, 2003; Karolchik et al. 2003). *DLK-DIO3* syntenic regions were identified using UCSC whole-genome chained sequence alignments as described by (Kent et al. 2003). The regions from different species were considered syntenic if the gene order was preserved. BLAT was used for sequence similarity cross-searches between different genomes (Kent 2002). BLAT parameters were determined empirically by searching for known human and mouse miRNAs reciprocally in the genomes of these 2 organisms. Maximum search sensitivity was achieved with a tile size set to 6 and overall minimum sequence identity set to 65%. These parameters were used to query vertebrate genomes using known human and mouse pre-miRNA sequences. Only the alignments covering at least 90% of the query pre-miRNA sequence were considered as orthologous pre-miRNA candidate sequences. The following genome assemblies were used in this study: human, hg18; chimpanzee, panTro2; mouse, mm8; rat, rn3; dog, canFam2; cow, bTau2; chicken, galGal3; elephant, loxAfr1; armadillo, dasNov1; opossum, monDom4; tenrec, echTel1; and fugu, fr2.

Detection and Analysis of Sequence Motifs

Sequence motifs were identified using MEME algorithm at the San Diego Supercomputing Center Web site (http://meme.sdsc.edu/meme/intro.html) (Bailey and Elkan 1994). The following parameters were applied: model = tcm, minimum width = 6, maximum width = 100, minimum sites = 2, maximum sites = 300. Sequence logos were generated using WebLogo 2.8.2 at http://weblogo.berkeley.edu/logo.cgi (Crooks et al. 2004).

The miRNA Target Genes, Gene Ontology Enrichment, and *P* Values

Predicted miRNA target genes were obtained from the TargetScan 4.0 Web site (http://www.targetscan.org/) (Lewis et al. 2003). To reduce false-positive rate of miRNA target prediction in our analyses, we considered a gene to be a true miRNA target if it contained at least 2 evolutionarily conserved miRNA target sites within its 3′ UTR. Gene Ontology (GO) annotations were downloaded from the GO consortium Web site (April 2007, http://www.geneontology.org/) (Camon et al. 2004; Harris et al. 2004). "Known Isoforms" identifiers for UCSC human (hg18) and mouse (mm8) "Known Genes" were used to make sure that a gene was only counted once where there were multiple isoforms. A Perl script and Structured Query Language code were created to calculate enrichment of terms and "Fisher's exact" *P* values against a background of all GO-annotated genes in the UCSC Known Genes database. For significance, we required at least 2-fold enrichment, $P < 1 \times 10^{-5}$, and at least 10 associated Known Genes in the target genes sample.

Results

Evolution of the mir-379/mir-656 Cluster

To address the evolutionary origin of the mir-379/mir-656 cluster, we examined the sequenced genomes of 11 mammals, chicken, and puffer fish. Using whole-genome sequence alignments from the UCSC genome browser, we searched for the *DLK1-DIO3* syntenic regions in the assembled vertebrate genomes. We found that *DLK1-DIO3* synteny is maintained in all mammalian and bird genomes but is absent in the puffer fish genome (table 1).

Table 2.
Evolution of the mir-379/mir-656 miRNA Cluster

	Human	Chimp	Mouse	Rat	Dog	Cow	Armadillo	Elephant	Tenrec	Opossum	Platypus	Chicken	Xenopus	Fugu
mir-379	✓		✓	✓										
mir-411	✓		✓	✓										
mir-299	✓		✓	✓										
mir-A1.2														
mir-380	✓		✓	✓		✓								
mir-A1.5														
mir-A1.6														
mir-323	✓		✓	✓										
mir-758	✓		✓	✓										
mir-329-1	✓		✓	✓										
mir-329-2	✓													
mir-A4.2														
mir-494	✓		✓	✓										
mir-G														
mir-679			✓											
mir-666			✓											
mir-543	✓		✓	✓										
mir-495	✓		✓	✓										
mir-667			✓											
mir-A7														
mir-A8														
mir-B1														
mir-B2														
mir-376c	✓		✓	✓										
mir-376a-2	✓													
mir-654	✓		✓											
mir-376b	✓		✓	✓										
mir-376a-1	✓		✓	✓										
mir-300	✓		✓	✓										
Hs212														
Hs213														
mir-381	✓		✓	✓										
mir-487b	✓		✓	✓		✓								
mir-539	✓		✓	✓										
mir-889	✓													
mir-544	✓		✓											
mir-655	✓													
cand.15232														
mir-487a	✓					✓								
mir-382	✓		✓	✓										
mir-134	✓	✓	✓	✓										
mir-668	✓		✓											
mir-485	✓		✓	✓										
mir-453	✓		✓											
mir-154	✓	✓	✓	✓										
mir-496	✓		✓											
mir-377	✓		✓	✓										
mir-541	✓		✓	✓										
mir-409	✓		✓	✓										
mir-412	✓		✓	✓										
mir-369	✓		✓	✓		✓								
mir-410	✓		✓	✓										
mir-A22.2														
mir-656	✓													

Distribution of the individual miRNAs from mir-379/mir-656 cluster in different vertebrate genomes. The miRNAs are listed in order of their position on the genomic DNA strand in 5′ to 3′ orientation. White cells marked with a tick indicate known experimentally confirmed miRNAs; light gray indicates orthologous miRNAs identified by sequence similarity; dashed gray cells show divergent orthologous miRNA sequences that are not likely to code for functional miRNAs; dark gray indicates that no identifiable orthologous sequences were found. Note that in some cases the absence of an orthologous sequence might be due to incomplete genomic assembly.

To identify miRNAs orthologous to human miRNAs from the mir-379/mir-656 cluster within the syntenic *DLK1-DIO3* regions of vertebrate genomes, we performed sequence similarity searches using BLAT (Kent 2002). The orthologous miRNAs clusters were easily identifiable in all examined genomes of placental mammals but were not detectable in the marsupial (*Monodelphis domestica*), monotreme (*Ornithorhynchus anatinus*), or any other non-mammalian vertebrate genome (tables 1 and 2 and Supplementary Material online). These data are consistent with the results of the similar analysis performed by Seitz et al. (2004) in worm (*Caenorhabditis elegans*), fruit fly (*Drosophila melanogaster*), and puffer fish (*Fugu rubripes*) genomes. Together, these results demonstrate that the mir-379/mir-656 cluster is an evolutionary innovation that is uniquely present in the placental mammals.

To examine the evolution of this cluster in detail, we looked at the evolutionarily conservation of the individual miRNA sequences within the cluster between different vertebrate genomes. The results of this analysis are summarized in the table 2. It is evident that most of the sequences of the known experimentally validated miRNAs are present in all examined genomes of the placental mammals. Although the assemblies of elephant (*Loxodonta africana*), lesser hedgehog (tenrec, *Echinops telfairi*), and armadillo (*Dasypus novemcinctus*) genomes are incomplete and the exact evolutionary fate of some miRNAs could not be resolved at present, it is important to note that most of the miRNA sequences from the mir-379/mir-656 cluster are present in these genomes. These 3 species are descendants of the lineages that diverged from the common placental ancestor at the early stages of mammalian evolution (fig. 1). The estimated divergence time between elephant, tenrec, armadillo, and human lineages is approximately 100 Myr, whereas separation of the eutherian lineage from the common mammalian ancestor are thought to have occurred between 180 and 140 MYA (Hedges et al. 2006). This leads us to conclude that the mir-379/mir-656 cluster emerged early in the eutherian lineage prior to radiation of modern placental mammals. The fact that the cluster has been maintained in different groups of placental mammals for approximately 100 Myr without any major structural rearrangements indicates that the whole cluster may function as a coordinated unit with an important biological role in this group of organisms.

Origin of the mir-379/mir-656 Cluster, Novel miRNA Candidates, and Regulatory Sequence Motifs

Sequence similarity observed between the individual miRNA precursors within the mir-379/mir-656 cluster led previous studies to conclude that these miRNAs originated from a common ancestral sequence by a process of tandem duplication (Seitz et al. 2004; Hertel et al. 2006). To identify the unit of amplification and to determine whether miRNAs from the cluster may share some regulatory elements, we examined human genomic sequences located between miRNA precursor sequences and 1 kb adjacent to the miRNA cluster on either side. We searched for overrepresented sequence motifs of a variable length using a motif discovery algorithm MEME (Bailey and Elkan
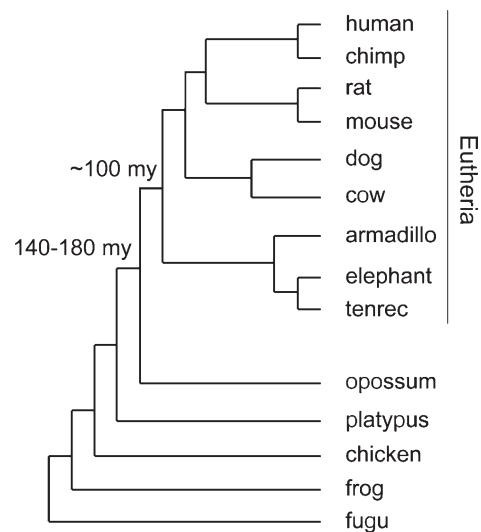

FIG. 1.—Summary of phylogenetic relationships of vertebrate species addressed in this study. The tree structure and estimated divergence times are used with modifications from Hedges et al. (2006) and Murphy et al. (2001).

1994). This analysis identified 2 motifs that were very significantly overrepresented within the cluster as compared with a random set of genomic sequences of a similar total length. Motif 1 was 21 nt long and was present 147 times within the mir-379/mir-656 cluster, which corresponds to the MEME-calculated e value of 1.2×10^{-209} (fig. 2*A*). Motif 2 was 23 nt long and was present at 115 sites within the cluster, which corresponds to the MEME-calculated e value of 2.4×10^{-70} (fig. 2*A*). Further inspection of the distribution of the motifs within the miRNA cluster revealed that both motifs are often present adjacent to known experimentally validated miRNA precursor sequences suggesting a regulatory function in expression and/or processing of the primary miRNA transcripts. We also noticed that copies of motif 2 frequently followed a copy of motif 1 and that both motifs have a regular periodic distribution across the ~45-kb genomic region encompassing the mir-379/mir-656 cluster (supplementary fig. 2, Supplementary Material online). To analyze this further, we calculated distances between neighboring pairs of motifs 1 and 2. After plotting the resulting distribution of the distances, we found that it had 1 major peak at ~160 bases (fig. 3). Interestingly, the distribution of the distances constituting this peak strongly resembled length distributions of a subset of the 38 known human miRNA precursor sequences with the adjacent motifs 1 and 2 (fig. 2*B*). This result suggests that we are observing the vestiges of an original amplified array consisting of an ~160 bases long repeat unit comprised of a single copy each of motif 1, motif 2, and miRNA precursor sequence (fig. 2*B*). Consistent with this, we were able to identify several additional sequences located within the ~45-kb locus that share similarity with known active miRNA precursors. Although some of these are only partially similar to the known miRNAs and are likely to be remnants of the ancestral repeat sequences that gave rise to the mir-379/mir-656 cluster, others are highly similar to the known
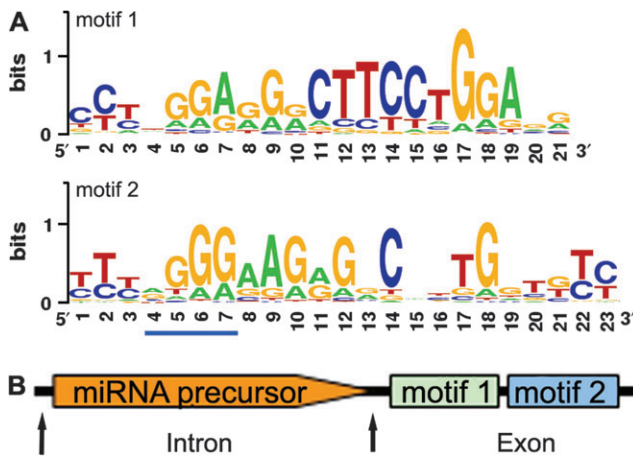
Fig. 2.—Amplified repeat units and sequence motifs within the mir-379/mir-656 miRNAs cluster. (*A*) Sequence logos of the 2 over-represented sequence motifs within human mir-379/mir-656 miRNAs cluster. The *y* axis shows informational content at each base position within the motif. Sequence logos for motifs 1 and 2 were created based on 147 and 115 individual sequences present within the mir-379/mir-656 miRNAs cluster, respectively. A blue bar indicates a possible hnRNP A1–binding site. (*B*) Structure of the proposed amplified repeat unit. The figure shows relative positions of the miRNA precursor sequence and of the 2 motifs. Vertical arrows indicate suggested positions of the splice sites boundaries in the ancestral amplified array.
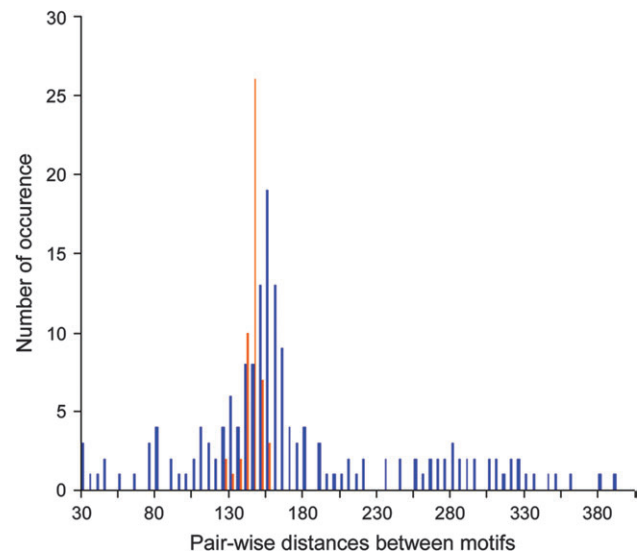
Fig. 3.—Size of the amplified repeat unit. Blue bars represent the distribution of pairwise distances between the adjacent motifs within human mir-379/mir-656 miRNAs cluster. Orange bars represent distribution of lengths of the 38 known human miRNA precursors from within the cluster measured with the 2 adjacent motifs. Each vertical bar represents a number of occurrences within a bin size 5. There were 21 individual single-point motif distances beyond cutoff of the *x* axis at 400 nt.

miRNAs and show evolutionary sequence conservation in at least 2 of the examined mammalian genomes. Table 2 shows 13 of these sequences. Importantly, 2 of them were experimentally validated by Berezikov, Thuemmler, et al. (2006a) during the course of this study (supplementary table 2, Supplementary Material online). Another 6 were previously identified as putative miRNA candidates (Seitz et al. 2004). The remaining 5 sequences are reported here as potential miRNA candidates for the first time (table 2, Additional file 2 [Supplementary Material online]). Although there is no experimental evidence to identify these sequences as functional miRNAs, the pattern of their evolutionary conservation suggests that they are expressed at least in some of the eutherian mammals.
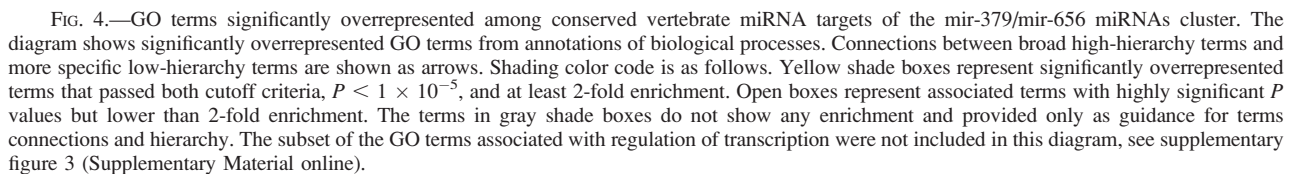
Evolving miRNAs

Despite the overall similarity in structure and sequence conservation between the orthologous mir-379/mir-656 miRNA clusters in placental mammals, we found several examples demonstrating an ongoing evolutionary selection of the individual miRNAs within the cluster. The loss and gain of the individual miRNAs is best illustrated by the rodent lineage. For example, we were able to identify mouse and rat sequences orthologous to human miRNAs mir-329-2, mir-655, mir-487a, and mir-656. However, the detailed analysis of multiple sequence alignments between rodents and other mammals showed that mouse and rat sequences have accumulated nucleotide substitutions, small deletions, and/or insertions, which are likely to affect secondary structure necessary for correct processing of the mature miRNAs. To assess this, we compared MFOLD 3.2–predicted RNA

secondary structures of the rodent sequences and experimentally validated orthologous miRNA sequences from human (Zuker 2003). We found that rodent sequences failed to produce characteristic hairpin-like miRNA precursor structures (supplementary fig. 1, Supplementary Material online). Therefore, we conclude that these sequences do not code for functional miRNAs in either mouse or rat and are likely to be remnants of the ancestral miRNA sequences (table 2, supplementary fig. 1 [Supplementary Material online]). The opposite process of the evolutionarily fixation of the individual lineage-specific miRNAs is also evident in rodents. Using similar analysis, we found that whereas degenerate orthologous sequences for rodent mir-679, mir-666, and mir-667 are present in all genomes of the placental mammals, the functional miRNAs seem to be present only in mouse and rat. Examples of the lineage-specific evolutionary selection of miRNAs have been also reported for primate and other lineages (table 2) (Berezikov, Thuemmler, et al. 2006; Hertel et al. 2006).

The miRNA Target Genes Point to Eutherian-Specific Biological Processes

In the publication describing the mir-379/mir-656 cluster, Seitz et al. (2003, 2004) hypothesized that the common origin of the miRNAs within the cluster and their co-expression from a large polycistronic transcript *Mirg* may also result in a common set of target genes. To investigate this possibility, we used TargetScan 4.0 data for the predicted evolutionary conserved vertebrate target sites for miRNAs from mir-379/mir-656 cluster to examine GO annotations of biological process associated with miRNA

Fig. 4.—GO terms significantly overrepresented among conserved vertebrate miRNA targets of the mir-379/mir-656 miRNAs cluster. The diagram shows significantly overrepresented GO terms from annotations of biological processes. Connections between broad high-hierarchy terms and more specific low-hierarchy terms are shown as arrows. Shading color code is as follows. Yellow shade boxes represent significantly overrepresented terms that passed both cutoff criteria, $P < 1 \times 10^{-5}$, and at least 2-fold enrichment. Open boxes represent associated terms with highly significant P values but lower than 2-fold enrichment. The terms in gray shade boxes do not show any enrichment and provided only as guidance for terms connections and hierarchy. The subset of the GO terms associated with regulation of transcription were not included in this diagram, see supplementary figure 3 (Supplementary Material online).

target genes in the human and mouse genomes (Lewis et al. 2003). The results of this analysis show that 5 functional categories of GO terms were significantly overrepresented among predicted miRNA target genes (fig. 4, supplementary fig. 3 [Supplementary Material online]). These categories can be broadly defined as regulation of transcription, RNA metabolism, cell motility, neurogenesis, and embryonic development. While genes involved in regulation of transcription and RNA metabolism appear to be common targets for many miRNAs and have been reported in several studies (Lewis et al. 2003; John et al. 2004; Grun et al. 2005), overrepresentation of target genes involved in neurogenesis, cell motility, and embryonic development is highly specific for the mir-379/mir-656 cluster. Consistent with this result, these 3 groups of genes did not show any significant overrepresentation when the same analysis was repeated with a random set of miRNAs of the same size, although as expected regulation of transcription genes were overrepresented in this set (data not shown).

Because the miRNA cluster has emerged after the divergence of the bird lineage and prior to the radiation of the eutherian mammals, it is likely that most, if not all, of the targets would be present in the eutherian mammals but not in the birds. To test this hypothesis, we repeated the GO analysis with only those genes that contained predicted miRNA-binding sites that are conserved in the eutherian genomes but not in chicken genome. We found that GO terms associated with neurogenesis and cell motility were still significantly enriched in this set of target genes, but GO terms associated with embryonic development were no longer significantly overrepresented.

Interestingly, GO terms related to different aspects of nervous system development were most common in the whole set of the overrepresented terms. These terms showed higher overall enrichment and lower P values compared with terms related to other biological processes. The biological process term that showed the highest enrichment was axon guidance (fig. 4). Logically, this term unites 2 other significantly overrepresented biological process terms: cell migration and axonogenesis. Examination of the individual target genes within this class demonstrates that some of them, like brain-derived neurotrophic factor, contain up to 7 evolutionarily conserved miRNA target sites within their 3′ UTR for different miRNAs from the mir-379/mir-656 cluster. As can be expected from the GO annotations, the predicted miRNA target genes would be highly expressed in the tissues related to GO annotations—namely, embryonic tissues and various parts of the developing and the adult brain. Indeed, out of 18 miRNA target genes associated with the GO term of axon

guidance, 14, including human homolog of Robo1, 2 ephrin receptors, and neurogenin 2, show high expression levels in various parts of the brain and the remaining 4 genes show moderate expression in at least one brain region (see GNF Gene Expression Atlas at http://symatlas.gnf.org/SymAtlas/ and Allen Brain Atlas at http://www.brain-map.org data [Su et al. 2004; Lein et al. 2007]). More importantly, our survey of the miRNA expression data from within the cluster also shows that most of these miRNAs were frequently detected in or cloned from the various adult brain–derived samples and some embryonic tissue samples (Seitz et al. 2004; Bentwich et al. 2005; Berezikov, Thuemmler, et al. 2006; Cummins et al. 2006) (for detailed summary and additional references, see supplementary table 2, Supplementary Material online).

These results demonstrate a significant overlap between the expression profile of the miRNAs from the mir-379/mir-656 cluster and their predicted target genes. Such overlap in the expression patterns between miRNAs and their predicted target genes strongly suggest that these results reflect biologically relevant miRNA–target interactions rather than unexpected biases in miRNA target predictions or GO annotations.

Discussion

Origin and Evolution of the mir-379/mir-656 miRNA Cluster

Consistent with the earlier studies, our results show that the mir-379/mir-656 cluster is an evolutionary innovation that appears first in the eutherian mammals (Seitz et al. 2004; Hertel et al. 2006). Acquisition of novel miRNA genes in evolution is a common trend in different groups of the metazoa that has been well documented recently (Hertel et al. 2006). However, unlike other novel miRNA genes that mostly originated from individual duplications of existing miRNA genes or exaptation of different genomic sequences (Smalheiser and Torvik 2005), mir-379/mir-656 cluster has a different origin. We have shown that not only the individual members of mir-379/mir-656 cluster but also the entire ~45-kb genomic region encompassing these miRNAs originated from an ancestral repeat unit that was amplified over 250 times (fig. 2B). Although initially the amplified copies of the repeat unit might have been functionally identical, their subsequent evolutionary fate followed 1 of the 3 alternatives described by the duplication–degeneration–complementation model (Force et al. 1999). This model suggests that duplication of a gene results in complete loss of function of the redundant copy (degeneration) without necessarily loss of the sequence itself (generation of a pseudogene), evolutionary preservation of both copies if they evolve to perform complimentary functions (subfunctionalization), or one of the copies evolves to perform an entirely new function (neofunctionalization) (Force et al. 1999). It is evident that in the case of mir-379/mir-656 cluster, all these possibilities have been realized, resulting in the array of all known miRNAs within the cluster. Events of functional degeneration are readily identifiable within the mir-379/mir-656 cluster. Although the remains of many of the ancestral repeat units can still be readily recognized within the ~45-kb region, most appear

to have degenerated over the last 100 Myr and have lost one or more components required for activity. Interestingly, in humans, motif 1 and motif 2 appear to be preserved better than some of the ancestral sequences that gave rise to miRNA precursors. Significant overrepresentation of the detected sequence motifs within the mir-379/mir-656 cluster, but not elsewhere in the genome, clearly indicates a functional relationship between these motifs and miRNAs within the cluster. The fact that motifs are still detectable even in places where miRNA precursor sequences have degenerated beyond recognition, as well as the overall uniform distribution of the motifs across the 45-kb region of the mir-379/mir-656 cluster, suggests that their role may be in the regulation of the entire cluster as well as individual miRNAs within it. In this context, it is noteworthy that events of sub- and neofunctionalization also apply to regulatory elements controlling function of the duplicated gene. In practice, this could mean that some miRNAs with identical sequences may still perform different functions due to sub- and/or neofunctionalization of their regulatory elements. Consistent with this is the fact that the 45-kb region is relatively depleted in transposable elements indicating that the sequences between active pre-miRNA sequences may also be functional despite the lack of evolutionary conservation at the primary sequence level (Simons et al. 2006).

One possible explanation of the biological role of these motifs could be in the regulation of processing of the primary miRNA (pri-miRNA) transcript. In fact, transcription and expression data available for the ~45-kb locus encompassing mir-379/mir-656 cluster suggest that the entire region may be transcribed into a single noncoding RNA precursor called *Mirg*, which is then processed to give rise to the individual miRNA precursors (Seitz et al. 2003, 2004; Mineno et al. 2006). Recent studies have demonstrated that processing of pri-miRNA transcripts can be complex and may include alternative pathways such as the mirtron pathway in Drosophila and the splicing repressor protein heterogeneous nuclear ribonuclear protein (hnRNP) A1–dependent miR-18a processing in humans (Guil and Caceres 2007; Okamura et al. 2007). In this context, it is worth noting that of the 11 known intron–exon boundaries from the miRNA cluster supported by expressed sequence tag and mRNA data (e.g., GenBank accessions AK021542 and AA861571 in humans, AJ517767 and AW244689 in mouse, and AW916103 in the rat), 5 are located between positions 5 and 9 of copies of motif 1. In contrast, no association between exon–intron boundaries and either motif is observed. However, one such splice site is in the vicinity of the probable 5′ end of the mir-369 precursor in mouse (in GenBank accession AJ517767). We speculate that the ancient amplified repeat unit may have contained a 3′ splice site within the motif 1 and perhaps also a 5′ splice site at the 5′ end of miRNA precursor sequence (fig. 2A). Over time many of these sites may have lost their function in splicing and other new sites may have evolved. We also note that motif 2 contains a conserved sequence that is similar to hnRNP A1–binding sites. Although we favor a role in processing for the motifs within the cluster, it is important to point out that several potentially overlapping molecular processes have been reported to take place within this cluster: maternal imprinting, RNA

editing, and tissue-specific expression (Seitz et al. 2003, 2004; Kawahara et al. 2007). Each of these processes requires regulation at different levels, which can result in various sequence and structural constraints present in this genomic region. Our analyses suggest several hypotheses to be tested experimentally in the future.

Role of mir-379/mir-656 Cluster in Placental Mammals

To discuss possible biological roles of the members of the mir-379/mir-656 cluster, it is important to summarize the results of this and other studies that have shown that the mir-379/mir-656 cluster is uniquely present in the placental mammals, it originated from a common ancestral precursor sequence, and it is imprinted and expressed from the maternally derived chromosome predominantly in embryonic brain and placental tissues (Cavaille et al. 2002; Seitz et al. 2003, 2004; Hertel et al. 2006). Together, these findings consistently indicate involvement of the mir-379/mir-656 cluster in biological functions specific to eutherian mammals.

Our results showed that genes associated with the biological process of axon guidance are among the most likely candidates targeted by miRNAs from the mir-379/mir-656 cluster. Although neither axon guidance nor associated processes of neurogenesis and cell migration are exclusive to eutherian mammals, closer investigation reveals that the nervous system underwent a significant upgrade and rewiring in this group of organisms as compared with nonplacental mammals. For example, one of the most significant evolutionarily innovations in the eutherian brain is the emergence of a large intrahemispheric connective structure called "corpus callosum" (reviewed by Mihrshahi [2006]). Like the mir-379/mir-656 cluster, the corpus callosum is exclusively present in placental mammals and has not been found in any of the nonplacental species. Formation of the corpus callosum relies on the correct specification of the commissural neurons and precise axon guidance across the midline to their final destination in the opposite hemisphere (Mihrshahi 2006; Lindwall et al. 2007).

Although we do not have strong evidence to suggest that any of the miRNAs from the mir-379/mir-656 cluster are is directly involved in the regulation of axon guidance in developing corpus callosum, we find that a few genes with known functions in the development of corpus callosum, including Robo1 and SLIT-like proteins (SLITRK1, SLITRK2, SLITRK3, and SLITRK6), are present among predicted targets of miRNAs from the mir-379/mir-656 cluster (Lindwall et al. 2007). Other genes implicated in biological processes that involve regulation of axon guidance, such as thalamocortical patterning and motoneuron projections, were also predicted to be targeted by several miRNAs from the mir-379/mir-656 cluster.

Our survey of miRNA expression data also revealed that miRNAs from the mir-379/mir-656 cluster are often detectable in the placenta. However, analysis of miRNA target genes and associated GO biological processes failed to show any significant overrepresentation of terms related to placental development or function. There also appears to be relatively limited knowledge about many biological processes in the placenta and consequently a lack of explicit GO annotations relating to the placenta.

Conclusions

It is clear that the mir-379/mir-656 cluster of miRNAs was generated by a large amplification event between the branching of the marsupial lineage and the radiation of the eutherian mammals. This appears to have been followed by a fairly rapid divergence of the miRNA sequences some of which evolved into new specificities and have become fixed in evolution. The remnants of the original event can be seen today, but most of the sequence in the region appears to be nonfunctional. Consistency between the results of bioinformatics analyses of miRNA target genes, their function and expression pattern, as well as analyses of miRNA expression pattern strongly suggest that the miRNAs in the cluster are likely to act cooperatively to influence novel regulatory pathways that emerged in the eutherian mammals.

Supplementary Material

Supplementary tables and figures are available as additional files 1–3 at *Molecular Biology and Evolution* online (http://mbe.oxfordjournals.org/).

Acknowledgments

Literature Cited

Aparicio S, Chapman J, Stupka E, et al. (41 co-authors). 2002. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science. 297:1301–1310.

Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology; Menlo Park (CA). Menlo Park (CA): AAAI Press. p. 28–36.

Bentwich I, Avniel A, Karov Y, et al. (13 co-authors). 2005. Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet. 37:766–770.

Berezikov E, Thuemmler F, van Laake LW, Kondova I, Bontrop R, Cuppen E, Plasterk RHA. 2006. Diversity of

microRNAs in human and chimpanzee brain. Nat Genet. 38:1375–1377.

Berezikov E, van Tetering G, Verheul M, et al. (14 co-authors). 2006. Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis. Genome Res. 16:1289–1298.

Camon E, Magrane M, Barrell D, Lee V, Dimmer E, Maslen J, Binns D, Harte N, Lopez R, Apweiler R. 2004. The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. Nucleic Acids Res. 32:D262–D266.

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

Consortium C. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature. 437:69–87.

Consortium CG. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature. 432:695–716.

Coutinho LL, Matukumalli LK, Sonstegard TS, Van Tassell CP, Gasbarre LC, Capuco AV, Smith TP. 2007. Discovery and profiling of bovine microRNAs from immune-related and embryonic tissues. Physiol Genomics. 29:35–43.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res. 14:1188–1190.

Cummins JM, He YP, Leary RJ, et al. (16 co-authors). 2006. The colorectal microRNAome. Proc Natl Acad Sci USA. 103:3687–3692.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 151:1531–1545.

Gibbs RA, Weinstock GM, Metzker ML, et al. (203 co-authors). 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature. 428:493–521.

Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. 2006. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 34:D140–D144.

Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. 2005. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. 33:D121–D124.

Grun D, Wang YL, Langenberger D, Gunsalus KC, Rajewsky N. 2005. microRNA target predictions across seven Drosophila species and comparison to mammalian targets. PLoS Comput Biol. 1:e13.

Guil S, Caceres JF. 2007. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. Nat Struct Mol Biol. 14:591–596.

Harris MA, Clark J, Ireland A, et al. (59 co-authors). 2004. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 32:D258–D261.

He L, Hannon GJ. 2004. MicroRNAs: small RNAs with a big role in gene regulation. Nat Rev Genet. 5:522–531.

Hedges SB, Dudley J, Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics. 22:2971–2972.

Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, Flamm C, Hofacker IL, Stadler PF. 2006. The expansion of the metazoan microRNA repertoire. BMC Genomics. 7:25.

Houbaviy HB, Murray MF, Sharp PA. 2003. Embryonic stem cell-specific MicroRNAs. Dev Cell. 5:351–358.

John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. 2004. Human microRNA targets. PLoS Biol. 2:e363.

Karolchik D, Baertsch R, Diekhans M, et al. (13 co-authors). 2003. The UCSC genome browser database. Nucleic Acids Res. 31:51–54.

Kawahara Y, Zinshteyn B, Sethupathy P, Iizasa H, Hatzigeorgiou AG, Nishikura K. 2007. Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science. 315:1137–1140.

Kent WJ. 2002. BLAT–the BLAST-like alignment tool. Genome Res. 12:656–664.

Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. 2003. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. Proc Natl Acad Sci USA. 100:11484–11489.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. Genome Res. 12:996–1006.

Kim J, Krichevsky A, Grad Y, Hayes GD, Kosik KS, Church GM, Ruvkun G. 2004. Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. Proc Natl Acad Sci USA. 101:360–365.

Lagos-Quintana M, Rauhut R, Meyer J, Borkhardt A, Tuschl T. 2003. New microRNAs from mouse and human. RNA. 9:175–179.

Lander ES, Linton LM, Birren B, et al. (255 co-authors). 2001. Initial sequencing and analysis of the human genome. Nature. 409:860–921.

Lein ES, Hawrylycz MJ, Ao N, et al. (108 co-authors). 2007. Genome-wide atlas of gene expression in the adult mouse brain. Nature. 445:168–176.

Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. 2003. Prediction of mammalian microRNA targets. Cell. 115:787–798.

Lindblad-Toh K, Wade CM, Mikkelsen TS, et al. (236 co-authors). 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 438:803–819.

Lindwall C, Fothergill T, Richards LJ. 2007. Commissure formation in the mammalian forebrain. Curr Opin Neurobiol. 17:3–14.

Mihrshahi R. 2006. The corpus callosum as an evolutionary innovation. J Exp Zool B Mol Dev Evol. 306:8–17.

Mikkelsen TS, Wakefield MJ, Aken B, et al. (235 co-authors). 2007. Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. Nature. 447:167–177.

Mineno J, Okamoto S, Ando T, et al. (11 co-authors). 2006. The expression profile of microRNAs in mouse embryos. Nucleic Acids Res. 34:1765–1771.

Murphy WJ, Eizirik E, O'Brien SJ, et al. (11 co-authors). 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science. 294:2348–2351.

Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC. 2007. The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell. 130:89–100.

Plasterk RHA. 2006. Micro RNAs in animal development. Cell. 124:877–881.

Prochnik SE, Rokhsar DS, Aboobaker AA. 2007. Evidence for a microRNA expansion in the bilaterian ancestor. Dev Genes Evol. 217:73–77.

Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, Lai EC. 2007. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. Genome Res. 17:1850–1864.

Seitz H, Royo H, Bortolin ML, Lin SP, Ferguson-Smith AC, Cavaille J. 2004. A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. Genome Res. 14:1741–1748.

Seitz H, Youngson N, Lin SP, Dalbert S, Paulsen M, Bachellerie JP, Ferguson-Smith AC, Cavaille J. 2003. Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. Nat Genet. 34:261–262.

Sempere LF, Cole CN, McPeek MA, Peterson KJ. 2006. The phylogenetic distribution of metazoan microRNAs: insights

into evolutionary complexity and constraint. J Exp Zoolog B Mol Dev Evol. 306B:575–588.

Simons C, Pheasant M, Makunin IV, Mattick JS. 2006. Transposon-free regions in mammalian genomes. Genome Res. 16:164–172.

Smalheiser NR, Torvik VI. 2005. Mammalian microRNAs derived from genomic repeats. Trends Genet. 21:322–326.

Su AI, Wiltshire T, Batalov S, et al. (13 co-authors). 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci USA. 101:6062–6067.

Suh MR, Lee Y, Kim JY, et al. (12 co-authors). 2004. Human embryonic stem cells express a unique set of microRNAs. Dev Biol. 270:488–498.

Tanzer A, Stadler PF. 2004. Molecular evolution of a microRNA cluster. J Mol Biol. 339:327–335.

Waterston RH, Lindblad-Toh K, Birney E, et al. (222 co-authors). 2002. Initial sequencing and comparative analysis of the mouse genome. Nature. 420:520–562.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31: 3406–3415.

Andrew Roger, Associate Editor