MicroRNA Evolution in the Human Genome

Bing Su, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China
Rui Zhang, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China

Micro ribonucleic acids (miRNAs) are a growing class of small RNA [~22 nucleotide (nt)] that function as negative regulators in the genome by targeting messenger RNA (mRNA) for translational repression, cleavage and destabilization. Hundreds of miRNAs have been identified in the human genome. miRNAs can be conserved or nonconserved according to their cross-species sequence conservation, and these two classes of miRNAs in the human genome are subject to distinct evolutionary patterns.

Introduction

Micro ribonucleic acids (miRNAs) are a class of small single-stranded noncoding RNAs (20–24 nt) in the genome. In the past few years, studies have shown that miRNAs are an important composition of the transcriptional output of the genome of plants and animals. They usually show tissue and developmental stage-specific expression and play pivotal roles in regulating diverse biological processes by targeting messenger RNAs (mRNAs) for translational repression, cleavage or destabilization. Understanding the origin and evolution of miRNAs is still an on-going process for research scientists. So far, what we know about this subject is limited. In plants, we know that some miRNAs can be generated by inverted duplication of target gene sequences, whereas in animals, little is known about the origin of miRNAs. Only a few animal miRNAs were reported derived from transposable elements (TEs) (Piriyapongsa et al., 2007). To date, hundreds of miRNAs have been identified in the human genome, and the number is still increasing. Thousands of 3’ untranslated regions (3’-UTR) of mRNAs are under selective constraint to maintain pairing to miRNA seed region (nucleotide 2–8 of the miRNA), enabling prediction of miRNA targets by searching the conserved 7-nt region among species matching to the miRNA seed region.

The Discovery of miRNAs

In 1993, through a genetic screen in Caenorhabditis elegans for defects in the temporal control of post-embryonic development, Lee and colleagues discovered the gene lin-4. It is interesting that lin-4 does not code a protein rather produces a 22-nt noncoding RNA, lin-4 is partially complementary to multiple 7-nt sites located in the 3’-UTR of a nuclear protein lin-14 and the base pairing between lin-4 and the lin-14 3’-UTR was essential for the ability of lin-4 to repress lin-14 expression through regulation of protein synthesis. The discovery of lin-4 and its target-specific translational inhibition suggested a new mechanism of gene regulation. In 2000, almost 7 years after the initial identification of lin-4, the second miRNA let-7 was discovered also in C. elegans. let-7 encodes a 21-nt small RNA. It also performs its function by binding to the 3’-UTR of lin-41, and inhibiting its translation. Unlike lin-4, both let-7 and its target gene lin-41 are evolutionarily conserved throughout metazoans. This observation suggests that there might be considerable number of miRNAs among metazoans (for review, refer to Kim, 2005).

miRNA Biogenesis and Functional Significance

Biogenesis of miRNAs

Thousands of miRNAs have now been identified in various organisms. In both animals and plants, miRNA genes are transcribed by RNA polymerase II as capped and polyadenylated primary miRNA transcripts (pri-miRNA). In animal, the RNase III enzyme Drosha initiates the nuclear processing of the pri-miRNA into a ~70-nt hairpin precursor miRNA (pre-miRNA). Drosha is a large protein of ~160 kDa, and it is conserved in animals. During processing, Drosha interacts with the double-stranded RNA-binding protein DiGeorge syndrome critical region gene 8 (DGCR8) to form the microprocessor complex. Although it is not fully understood, it seemed that the tertiary structure of pri-miRNAs is the primary determinant for substrate specificity of Drosha complex (Figure 1).

Following nuclear processing by Drosha, pre-miRNAs are exported to the cytoplasm. Export of pre-miRNA is
mediated by one of the nuclear transport receptors, exportin-5, which is a member of the karyopherin family of nucleocytoplasmic factors. Then the RNase III Dicer processes the pre-miRNA into an imperfect RNA duplex about 22 nt, with 2 nt 3′ overhangs. Dicer is a highly conserved protein that is found in almost all eukaryotic organisms. It associates with several other proteins, such as R2D2, FMR1 in Drosophila melanogaster, and Argonaute family proteins in various organisms. These Dicer-interacting proteins do not seem to be required for the cleavage reaction because purified Dicer itself can catalyse the cleavage reaction. Instead, Dicer-interacting proteins have various roles in miRNA stability and effector complex formation and action. The strand of the miRNA duplex with the weakest base pairing at its 5′ terminus is preferably loaded into the effector complexes that are known as ‘miRNP’ (miRNA-containing ribonucleoprotein complex) (for review, refer to Kim, 2005). Finally, miRNP performs miRNA function through interacting with mRNAs (Figure 1).

In plants, DCL1, one of the four Dicer-like proteins in Arabidopsis thaliana, might be responsible for the processing of pre-miRNA. Moreover, the export of pre-miRNA is mediated by HASTY (HST), which is a plant homologue of exportin-5 (for review, refer to Kim, 2005.)

Recently, in the unicellular green alga Chlamydomonas reinhardtii, miRNAs are also identified. The Chlamydomonas genome encodes both Dicer and Argonaute nuclease (AGO) proteins, indicating the processing of miRNAs is similar to plant and animal. However, the length of hairpin in Chlamydomonas is more variable compared to multicellular organisms, suggesting a relatively unstable and early stage in miRNA evolution (Molnar et al., 2007).

Functional significance of animal miRNAs
Animal miRNAs have been implicated in early development, tissue morphogenesis, cellular processes such as apoptosis and major signalling pathways. They are also involved in metabolic regulation and diseases such as cancer. The widespread and important role of miRNAs in animals is highlighted by recent estimates that ~30% of all genes are miRNA targets. In general, the miRNA targets can be categorized into two groups according to their sequence conservation among species. It is suggested that the conserved targets are often highly expressed at developmental stages before miRNA expression and that their levels tend to fall as the miRNAs that target them begin to accumulate. In addition, the genes containing nonconserved sites have a propensity of not expressing in the same tissue as the targeting miRNAs (Farh et al., 2005). A large set of genes involved in basic cellular processes avoid miRNA regulation due to short 3′-UTRs that are specifically depleted of miRNA binding sites (Stark et al., 2005). Consequently, miRNAs regulate the expression and influence the evolution of most animal mRNAs.

Genomics of Human miRNAs
Increasing number of identified miRNAs in the human genome
There is an increasing number of identified miRNAs in the human genome registered in miRBase after the first human miRNA was identified (Figure 2). As mentioned earlier, let-7
Genomic organization of miRNAs

miRNAs are located universally in the human genome. A total of 53% of miRNAs are located in the intergenic region, followed by 31% in the intron region and 13% in the UTR (Megraw et al., 2007). The 44% miRNAs are located in the intron or UTR of known genes, suggesting the share of regulation elements between protein-coding genes and miRNAs.

Previous study suggests that miRNAs in the human genome are apt to clustering. By the criterion that the maximum distance smaller than 1 kb between any two miRNAs is considered to be in the same cluster, there are 30.5% miRNAs in the human genome forming clusters (Megraw et al., 2007). Considering the small size of miRNA genes, it is evident that the miRNA clustering in the human genome is significantly higher than the expected of random distribution, suggesting the co-regulation of clustered miRNAs by shared cis-acting regulatory regions or higher-order, chromatin-mediated regulatory coordination. Sometimes, miRNAs in one cluster show pre-miRNA similarity between each other (i.e. they form a miRNA family), suggesting formation of this cluster by tandem duplication. In other cases, members within a family show no similarity. In addition, many miRNA paralogues with one or more nucleotides substitutions also existed among different miRNA clusters, suggesting the ancestral duplications respond to the expansion of miRNA cluster during evolution.

miRNA families in the human genome

The miRNA families are usually defined as a group of pre-miRNAs that show similarity between each other. And the within-family pre-miRNAs produce similar, if not identical, mature miRNAs. In the human genome, the precursors within family can scatter among genome or group together by forming cluster. These patterns suggest the expansion of miRNA family via tandem duplication or segmental duplication in the human genome. Compared to plants, the number of human miRNA families is larger though the average number of precursor/family in human (1.37 per family) is smaller than that of plants (ranging from 2.51 ~ 6.45 per family). The largest miRNA family identified so far in human is a cluster located in chromosome 19. It contains 43 precursors and expressed preferentially in placenta. The precursors are fully conserved only in primates; and many of the miRNAs are embedded in long (400 ~ 700 nt) sequences that are repeated along the cluster, suggesting they evolved through local duplication and mutation events unique to primates (Bentwich et al., 2005).

Evolution of Conserved miRNAs in the Human Genome

As mentioned earlier, miRNAs can be classified into two groups, the conserved and nonconserved miRNAs, according to their between-species sequence conservation. In the human genome, most conserved mature miRNAs, especially their seed regions, are totally conserved compared to other primates, even to rodents and dog. It seemed that the conserved miRNAs are subject to more stringent evolutionary constraint compared to protein coding genes, suggesting the functional conservation and importance of them in the human genome. A good example of conserved miRNA evolution in the human genome is presented by Berezikov and colleagues (Berezikov et al., 2005). They sequenced 122 human miRNAs in 10 diverse primate species, most of which were identified relying on sequence conservation. These species cover the major lineages of

Figure 2  The rapid increase of identified miRNAs in the human genome registered in miRBase. From Version 3 (2004 January) to Version 9 (2007 February).
primates, including ape, Old World monkey, New World monkey and prosimian, reflecting 63 million years of primate evolutionary history. **Figure 3** shows the secondary structure of the pre-miRNA. Strict conservation is observed in terms of miRNA stem regions and increased variation occurs in loop sequences. Sixty-eight of 122 miRNAs have variations in the miRNA precursor regions, but only three are variable in the mature miRNA, suggesting strong evolutionary constraint in mature miRNA regions. In total, they observed 154 variations that do not affect the secondary structure of a pre-miRNA sequence and 118 variations that lead to disruption of base pairing in a hairpin. Most of the observed variations occur in terminal loop sequences and at the ends of a precursor, suggesting that there is less functional constraint on the primary sequence of these regions of miRNAs. Interestingly, a striking drop in conservation was found for sequences immediately flanking the miRNA hairpins (**Figure 4a**, using miR-194 as an example), suggesting that no common *cis*-acting elements can be immediately recognized in human miRNAs. Since the flanking sequences are presumably nonfunctional and evolved neutrally, this pattern also suggests that the conserved miRNAs are subject to strong negative selection, which leads to stringent function constraint.

**Evolution of Nonconserved miRNAs in the Human Genome**

Nonconserved miRNAs may contribute to the functional novelty during evolution. The nonconserved miRNAs in the human genome show a different evolutionary pattern compared to conserved ones. For example, we studied the evolution of a nonconserved X-linked human miRNA cluster in primates (**Zhang et al., 2007**). When compared with conserved miRNAs, this cluster showed both varied sequence substitution rate and locations. Firstly, they show frequent tandem duplications in primates. **Figure 5** shows the genomic structure of the miRNA cluster in the six primate species compared. Three miRNAs (miR-513, miR-509 and miR-514) have different copy numbers among species. For example, miR-514 has three copies in human, four copies in chimpanzee, but only one copy in the other primate species, indicating common miRNA duplications in primates. Secondly, the variation rate is significantly higher in closely related species in the X-linked cluster (**Table 1**). In addition, the conserved miRNAs show evolutionary constraint in mature miRNAs compared to the other regions of the precursors, whereas the nonconserved miRNAs have nearly equal substitution rate between mature miRNAs and the other regions of precursors. Thirdly, compared to the flanking regions or the local ancestral repeat, some miRNAs of the X-linked cluster have significantly higher substitution rate, an indication of excess of substitutions over neutrality (**Zhang et al., 2007**). For example, **Figure 4** demonstrates the difference in conservation between miR-194 (conserved miRNA) and miR-510 (nonconserved). The conservation of miR-194 in primates (including human, ape, Old World monkey, New World monkey) is higher than that of miR-510. In
addition, there is evident drop of conservation in the flanking sequences of miR-194, whereas, no such pattern was found when comparing miR-510 and its flanking region. This suggests that recently emerged miRNAs can evolve rapidly.

Rapid evolution of the nonconserved miRNAs would allow the emergence of lineage-specific miRNAs, which opens up the question of whether there are human-specific miRNAs. Experimental support for nonconserved miRNAs in human has recently been reported. Using massively parallel sequencing method, Berezikov and colleagues identified hundreds of miRNAs not conserved beyond primates, and some miRNAs are likely human-specific (Berezikov et al., 2006). They proposed that the different miRNA repertoire between human and chimpanzee in the brain could have a role in establishing or maintaining cellular diversity and could thereby contribute to the differences in human and chimpanzee brain evolution and function. Since only brain sample was analysed, we should expect that more nonconserved miRNAs, or human-specific miRNAs could be identified when more tissues are analysed.

Perspective

Although the functional importance of miRNAs in diverse biological processes has been widely understood, the origin and evolution of human miRNAs, especially nonconserved miRNAs, is not well studied. Recently developed small RNA library deep sequencing method has made it possible to identify many more low-expressed nonconserved miRNAs in human. With the recently finished genome sequences in diverse primate species, the comparison of nonconserved miRNAs between close-related species could shed light on the evolutionary details of human miRNAs.

Another important research subject will involve understanding the function of nonconserved miRNAs in the human genome. The identification of relationship between nonconserved miRNAs and the emerging target genes could tell more about the co-evolution of miRNA and mRNAs compared to conserved miRNAs. In addition, large-scale genetic screening at cellular level may reveal function of nonconserved, or human-specific miRNAs, and contribute to a better understanding of human physiology and pathology.

References


Further Reading


Web Links

miRBase http://microrna.sanger.ac.uk/
mirGen http://www.diana.pcbi.upenn.edu/mirGen.html