Rapid evolution of an X-linked microRNA cluster in primates

Rui Zhang, Yi Peng, Wen Wang and Bing Su

*Genome Res.* published online Apr 6, 2007; Access the most recent version at doi:10.1101/gr.6146507

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to *Genome Research* go to: http://www.genome.org/subscriptions/

© 2007 Cold Spring Harbor Laboratory Press
Rapid evolution of an X-linked microRNA cluster in primates

Rui Zhang,1,2,3 Yi Peng,1,2 Wen Wang,1 and Bing Su1,2,4

1Key Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; 2Kunming Primate Research Center, Chinese Academy of Sciences, Kunming 650223, China; 3Graduate School of Chinese Academy Sciences, Beijing 100039, China

MicroRNAs (miRNAs) are a growing class of small RNAs (about 22 nt) that play crucial regulatory roles in the genome by targeting mRNAs for cleavage or translational repression. Most of the identified miRNAs are highly conserved among species, indicating strong functional constraint on miRNA evolution. However, nonconserved miRNAs may contribute to functional novelties during evolution. Recently, an X-linked miRNA cluster was reported with multiple copies in primates but not in rodents or dog. Here we sequenced and compared this miRNA cluster in major primate lineages including human, great ape, lesser ape, Old World monkey, and New World monkey. Our data indicate rapid evolution of this cluster in primates including frequent tandem duplications and nucleotide substitutions. In addition, lineage-specific substitutions were observed in human and chimpanzee, leading to the emergence of potential novel mature miRNAs. The expression analysis in rhesus monkeys revealed a strong correlation between miRNA expression changes and male sexual maturation, suggesting regulatory roles of this miRNA cluster in testis development and spermatogenesis. We propose that, like protein-coding genes, miRNA genes involved in male reproduction are subject to rapid adaptive changes that may contribute to functional novelties during evolution.

[Supplemental material is available online at www.genome.org. The sequence data from this study have been submitted to GenBank under accession nos. EF466135–EF466137.]
Figure 1. (Legend on next page)
(Bentwich et al. 2005), we identified two new miRNAs in dog (designated as cfa-mir-513 and cfa-mir-X) and one in rodents (designated as mmu-(rno)-miR-X). The cfa-mir-513 is orthologous to mir-513 in primates (bootstrap value, 99%) and the cfa-mir-X and mmu-(rno)-miR-X are phylogenetically close to the lineage clustering mir-509, mir-510, mir-510L, mil-514 in primates (bootstrap value, 85%) (Fig. 1C). Analysis of amphibian, bird, and fish species failed to detect any homologs, implying that either this cluster emerged relatively recently or the miRNA sequences have diverged too far to recognize orthology.

Figure 1C shows the phylogenetic relationships among all the precursor miRNAs in primates, rodents, and dog. Again, as observed in the human miRNAs (Fig. 1A), the physical distances between miRNAs correlate to the phylogenetic relationships, supporting the proposed tandem duplications in primates.

To rule out the possibility that some of the putative miRNAs in primates might be pseudogenes, we tested the expression of three miRNAs with copy number variations (mir-509, mir-513, and mir-514) in rhesus monkey testis. We detected the expression of all copies of the precursors (data not shown), suggesting that all the copies are probably functional. Hence, the difference of copy numbers among primate species may have dosage effects in regulating tests development and/or function. However, we cannot rule out the possibility that, despite their transcription, some miRNA copies may represent processed pseudogenes.

It is well known that copy number variations (CNVs) are common in human populations (Redon et al. 2006). To test whether the miRNAs with CNVs between species could also be polymorphic within human populations, we tested mir-509 in three major continental populations (10 Chinese, 10 Caucasians, and 10 Africans) and all the three mir-509 copies were detected in all the human samples, an implication of fixation of the duplicated miRNAs in human.

We also observed rapid sequence evolution of the miRNAs. In five miRNA members with confirmed primate orthologs (mir-506, mir-507, mir-508, mir-510 in all six species, and mir-514 in four species), we observed a total of 77 sites with sequence substitutions in the precursors. The average substitution rate is 15.4 per miRNA, which is much higher than the rate reported by Berezikov et al. (2005) (2.27 per miRNA in 122 miRNAs based on 2 single substitutions, No. expected No. observed

<table>
<thead>
<tr>
<th>Type of substitution</th>
<th>No. expected</th>
<th>No. observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base pairing to base pairing</td>
<td>7.1</td>
<td>22</td>
</tr>
<tr>
<td>Base pairing to non-base pairing</td>
<td>49.9</td>
<td>35</td>
</tr>
<tr>
<td>Double</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base pairing to base pairing</td>
<td>1.3</td>
<td>5</td>
</tr>
<tr>
<td>Base pairing to non-base pairing</td>
<td>3.7</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1. Substitutions observed in precursor miRNAs

Single substitutions were calculated for mir-506, mir-507, mir-508, mir-510 of six primate species, and miR-514 of siamang, Yunnan snub-nosed monkey, rhesus monkey, and black-handed spider monkey with confirmed orthologs (see Methods). Single substitutions, P = 0.002, double substitutions, P = 0.05 (two-tailed Fisher’s exact test).

with confirmed orthology with that of 102 known intergenic miRNAs (data from http://miRnamap.mbc.nctu.edu.tw/). We calculated the between-species substitution rates (human vs. rhesus monkey) for the miRNA precursor (Kp) and the flanking genomic sequence (Kf) (presumably nonfunctional). The average Kf values are similar between the X-linked (0.064) and the intergenic miRNAs (0.053), an indication of nearly equal mutation rates. However, the X-linked miRNA cluster has a much larger average Kp (0.047) compared to the intergenic miRNAs (0.013) (P < 0.001, two-tailed Student’s t-test), again supporting the proposed rapid evolution of the X-linked miRNA cluster. We also compared Kp with Kf between human and chimpanzee. Kf is the substitution rate of the transposable elements (30 in total) within the X-linked cluster shown to be neutral (Lunter et al. 2006). Three out of the eight Kp values are out of the 99% confidence interval of the Kf distribution, an indication of excess of substitutions over neutrality (Supplemental Fig. 3).

It should be noted that, despite the rapid sequence substitutions, the general secondary structures (hairpin and loop structures) are highly conserved in all the primate miRNAs, as reflected by the excess of compensatory substitutions over random expectation (Table 1) (single substitutions, P = 0.002, double substitutions, P = 0.05, two-tailed Fisher’s exact test), consistent with strong functional constraint on miRNA secondary structures. This substitution pattern suggests that compensatory mutations could be the mechanism of miRNA evolution, as shown in other functional RNA genes (Hancock et al. 1998; Higgs 2000). In addition, the few observed deletions in the precursors did not affect the secondary structures (Berezikov et al. 2005).

Sequence substitutions may lead to the emergence of novel miRNAs. In mir-513 and mir-509, we observed sequence substitutions in the mature miRNAs both between species and between copies within species. Target gene prediction for mir-513 (two human copies, one chimpanzee copy, and one rhesus copy) by miRanda (Enright et al. 2003) showed that the duplicated miRNAs target fewer genes than do the ancestral copy, and that
many of the targets are novel (not targeted by the ancestral copy). This observation suggests that the new miRNA copies might be functionally more specialized (data not shown). Moreover, for miR-508 and miR-510, we observed lineage-specific substitutions important for target recognition (Bartel 2004; Brennecke et al. 2005) (human miR-508 at site 16 of the mature miRNA; siamang miR-508 at site 12; human and chimpanzee miR-510 at site 6; chimpanzee miR-510 at site 4). These sequence changes of miRNAs are possibly driven by natural selection on testis-expressed miRNAs, as previously found in protein-coding genes involved in male reproduction (Wyckoff et al. 2000; Swanson and Vacquier 2002; Dorus et al. 2004; Clark and Swanson 2005). It should be noted that current target prediction methods may suffer from a high false positive rate, and the miRanda algorithm used reported a success rate of \(~64\%\) in Drosophila (Stark et al. 2005). Therefore, the suggested functional specialization of the miRNAs is yet to be verified with experimental data.

To understand the functional roles of the X-linked miRNA cluster, we tested the expression of seven miRNA seeds (miR-510L was not tested) in testis of infant (1–2 yr, sexually immature) and adult rhesus monkeys (8–10 yr, sexually matured), using real-time quantitative PCR. All seven showed a significant expression reduction of the mature miRNAs in adults compared with infants (Fig. 2A). In particular, miR-514 is only expressed in infants (Fig. 2B). Recent studies have shown that animal miRNAs can not only repress translation of target gene miRNAs, but also induce their degradation by nonperfect sequence recognition (Bagga et al. 2005; Lim et al. 2005). Hence, the observed reduced expression of the pri-miRNAs in adults would allow development-stage-related functioning of the target genes during male sexual maturation. Thus the X-linked miRNA cluster likely plays regulatory role in primate testis development and spermatogenesis, and experimental data are needed to confirm the suggested regulatory role of this cluster.

Previous genome-wide analysis identified a series of fast-evolving genomic regions or lineage-specific structure changes (deletions, insertions, and inversions) in human (Newman et al. 2005; Pollard et al. 2006; Prabhakar et al. 2006). Our data add another example by showing rapid evolution of a microRNA cluster in human and nonhuman primates. However, the previously identified fast-evolving regions do not include the X-linked microRNA cluster, which is probably due to the different strategies used since the previous studies were aimed to identify human-specific changes (Pollard et al. 2006; Prabhakar et al. 2006).

In conclusion, we report rapid evolution of the X-linked miRNA cluster in primates, with frequent copy number changes and sequence substitutions. This observation is consistent with the finding on a protein-coding gene cluster, i.e., PRAME, which is testis-expressed and has expansion of copy numbers in primates (Birle et al. 2005). Hence, like protein-coding genes (Wyckoff et al. 2000; Swanson and Vacquier 2002; Dorus et al. 2004; Clark and Swanson 2005; Zhang et al. 2007), miRNA genes related with male reproduction are subject to adaptive changes due to selection, which may contribute to the emergence of functional novelties during evolution.

**Methods**

**BAC library screening, sequence assembly, and sequence analysis**

The pooled PCR-based method was used in screening the primate BAC libraries and identifying positive BAC clones. Primer sequences are shown in Supplemental Table 1. A total of seven positive BAC clones were detected, three of which (one for each species, 150 kb, 160 kb, 210 kb) were selected by end-sequencing, then full-length sequenced using shotgun sequencing method with \(6 \times\) coverage at Beijing Genomics Institute, CAS. The sequences were aligned and assembled by phred/phrap/consed package (Ewing and Green 1998; Ewing et al. 1998; Gordon et al. 1998). The sequence gaps were filled by PCR-based sequencing on an ABI-3130 sequencer. The sequence quality of the putative miRNAs was checked by looking at the raw electro-morph data.

We reanalyzed the human X-linked cluster using BLASTN (E-value cutoff \(10^{-4}\)) with previously identified 10 miRNAs as query sequences. Paralogous relationship of the newly identified members of miR-513 and miR-509 in human was determined by phylogenetic analysis (neighbor-joining method, sequence similarity, 97%–99%; bootstrap values, 98%) (Nei 1987). We searched the homologs of the miRNAs in nonhuman primates with BLASTN (E-value cutoff \(10^{-10}\)), and all the 15 human miRNA precursors were used as query sequences. The orthologous miRNAs were identified based on reciprocal best hits. The orthologous miRNAs of rodent (mouse and rat) and dog were identified using multiple alignment data in the UCSC genome browser (the Vertebrate Multiz Alignment & Conservation track). The identified orthologous sequences were also supported by synteny check. We used the UCSC genome database (Hirnrichs et al. 2006) in our sequence analysis, including the Human Mar. 2006 (hg18) assembly, the Chimp Mar. 2006 (partTro2) assembly, the Rhesus Jan. 2006 (rheMac2) assembly, the Dog May 2005 (canFam2) assembly, the Mouse Feb. 2006 (mm8) assembly, and the Rat Nov. 2004 (rm4) assembly. Gene sequences were aligned by ClustalX (Jeanmougin et al. 1998) with manual adjustment. Phylogenetic trees were reconstructed by using the NJ method in MEGA3 (Kumar et al. 2004) and evaluated by 1000 bootstrap replications. For calculation of between-species sequence substitutions, the orthologous precursor miRNA sequences were aligned by ClustalX (Jeanmougin et al. 1998). For miR-506, 507,
508, 510, the orthologs were identified in all the six primate species. Due to multiple duplications, for mir-514, there are only four species with unambiguous orthologs, and no orthologs from more than three species could be unambiguously identified for mir-509, 513 (Berezikov et al. 2005).

For target prediction, the ancestral copy of mir-513 was inferred using the core region of the within-species duplication units with dog mir-513 as outgroup. Dotter (Sonnhammer and Durbin 1995) was used to find the duplication units of mir-513, mir-509, and mir-514. Repeatmasker (http://repeatmasker.org) was used to find repeat elements.

**Calculation of compensatory substitutions**

The secondary structures of the pre-miRNAs were predicted by mfold (Mathews et al. 1999; Zuker 2003). The miRNA sequences and secondary structures were provided in Supplemental Figure 4. The pre-miRNA sequences were divided into four parts as described before (Han et al. 2006). We considered two classes of changes in the stem region: substitutions that change one pair of complementary bases to another pair of complementary bases (e.g., C-G to U-G) and substitutions that change one pair of complementary bases to a pair of noncomplementary bases, or vice versa (e.g., C-G to C-C). For a particular site in the stem region, a single substitution refers to change of either of the two corresponding bases, and a double substitution refers to changes of both bases. The probability of a single substitution or a double substitution converting one pair of complementary bases to another pair of complementary bases was provided before (Dixon and Hillis 1993). The expected values of substitutions were calculated following the method of Dixon and Hillis (1993). We aligned the orthologs or paralogs of the X-linked miRNAs from the six primate species by ClustalX. For single substitution analysis, only the confirmed orthologs were used to avoid multiple counts of shared mutations between species in paralogs. The single substitution pattern in paralogs is listed in Supplemental Table 2, which is similar to the result of the orthologs. Both the orthologs and paralogs were used in calculating double substitutions (multiple counts were removed) to test if compensatory substitutions are overrepresented.

**Calculation of the substitution rates of precursor miRNAs (Kp) and flanking genomic sequences (Kf)**

For calculation of Kf, the 5’ and 3’ flanking sequences (500 bp each side, 1000 bp in total) of the pre-miRNAs were used. Due to the close proximity of mir-506 and mir-507 in the genome (142 bp), only 642 bp were used. A total of 108 known intergenic human miRNAs (physical distance >500 bp from each other) (http://mirnamap.mbc.ntcu.edu.tw/) were used to blast the rhesus monkey genome (rhemac2), and 102 rhesus orthologs were identified. The sequences were then aligned by ClustalX. Kimura 2-parameter model was used to calculate substitution rates of the precursor miRNAs (Kp) and flanking genomic sequences (Kf) (Kimura 1980).

**Calculation of K of transposable elements (TE)**

Due to the variable and complex nature of this region, it is difficult to collect enough orthologous TEs in distantly related species (such as between human and rhesus monkey). With the use of siamang as outgroup, we identified 30 TE sequences within the sequenced X-linked cluster in human and chimpanzee. The orthology was determined by three criteria: TE classification, reciprocal best hits, and synteny. Kimura 2-parameter model was used to calculate substitution rates (Kimura 1980). By One-Sample Kolmogorov-Smirnov Test, the normal distribution of Kf was confirmed.

**Target gene prediction**

Human, chimpanzee, and rhesus monkey 3’UTRs are from Ensembl BioMart. For each gene, we only choose one transcript. If a gene has more than one transcript, the longest one is chosen for analyses. miRanda (Enright et al. 2003) was used to identify miRNA binding sequences in the 3’UTR sequences. Due to the rapid substitution nature of this cluster, we did not use the conservation criterion because it was shown that the nonconserved target sites also mediated repression (Farh et al. 2005). The parameters used for scanning are S:90, ΔG: -17 kcal/mol, and -shuffle.

**Real-time quantitative PCR analysis of miRNA expression**

The method used to quantify the expression level of mature miRNAs was described previously (Raymond et al. 2005). The LNA primers were purchased from Prologio LCC. The potential dimers of GS-primer and L-primer were evaluated by Autodimer (Vallone and Butler 2004). All primers are listed in Supplemental Table 3. RNA was extracted in duplicates from three infant and three adult rhesus testis samples using TRIzol (Invitrogen). After DNase treatment, a mixture of GS-primer and GAPDH specific primer was used for reverse-transcription by Omniscript Reverse Transcriptase (QIAGEN). The sequences of each miRNA amplicon were determined by TA-cloning (Takara). Quantitative RT-PCR by SYBR green (Bio-Rad) for each cDNA was run in triplicates in MJ Opticon. We calculated the relative amounts of miRNAs by the comparative Ct (2–ΔΔCt) method (Livak and Schmittgen 2001). The target miRNA values were normalized based on the internal control of GAPDH. Each Ct value used for the calculation was the mean obtained from each CDNA in triplicates. Statistical analysis was done with Student’s t-test or t-test for unequal variances.

**Precursor miRNAs detection**

The list of primers is provided in Supplemental Table 1. Total RNA was isolated by TRIzol. After DNase treatment, we used GS-primer for reverse transcription. The PCR amplicons were validated using gel electrophoresis, purified using MinElute PCR Purification Kits (QIAGEN), and subcloned with TA-cloning vectors. Thirty randomly chosen positive clones were sequenced for each pre-miRNA.

Primer for detection of mir-509 copy numbers are listed in Supplemental Table 1.

**Acknowledgments**

We thank Scott W. Roy for extensive discussion and critical reading of the manuscript, Ayelet Chajut for kindly providing sequence data of mouse and dog, and Chris Raymond for technical suggestion on miRNA quantification. We also thank Xiao-jing Yu, Yin-juan Wang, Ao-lei Niu, Cheng-hong Liao, and Ruo-lin Yang for their help in this study. We thank Xiao-na Fan and Hui Zhang for their technical assistance. This study was supported by grants from the Chinese Academy of Sciences (KSCX1-YW-R-34), the National Natural Science Foundation of China (30630013, 30525028), the Natural Science Foundation of Yunnan Province of China, and the National 973 project of China (2006CB701506).

**References**


