Segmental Duplications: A Possible Mechanism of Hominid Uplift through MicroRNA Diversification

Marla A. Endriga[1], Aldrich Ivan Lois D. Burog[1], Denise Lauren V. Dalmacion[1] and Custer C. Deocaris[2]*

[1] Department of Biology, College of Arts and Sciences, University of the Philippines Manila, 1000 Ermita, Manila, Philippines
[2] Laboratory of Exercise Biochemistry, Institute of Health and Sports Sciences, University of Tsukuba, Japan

Abstract

MicroRNAs (miRNA) are ~21 nucleotide-long gene silencers. Segmental duplications (SD) are among the driving forces in acquiring new genes. Both miRNA and SD are believed to have played a significant role in evolution, particularly in the divergence of humans (*Homo sapiens*) from the chimpanzee (*Pan troglodytes*). This study determines the distribution of miRNAs in humans and in chimpanzees, and presents a hypothesis on its significance in the occurrence of segmental duplications. MiRNA sequences from miRBASE were subjected to BLAT and BLAST to determine if miRNAs are located in SD regions or not. Homology between miRNAs was determined with ClustalW. BLAST was then used to determine whether the non-homologous human miRNA are homologous to any other part of the chimpanzee genome. We found that all 695 human miRNAs are found exclusively in SD regions, and that 67 are *de novo* miRNAs. Thirteen are homologues of chimpanzee miRNA, and 11 were possibly derived from non-miRNA regions in chimp. Of these, 6 were located in SD regions of the chimpanzee genome. Results indicate that miRNA evolution occurs within regions of segmental duplication and suggest that the presence of miRNA duplicates allows more exposure to mutations that could necessitate diversification, and possibly evolution, through sub- and neofunctionalization.

Keywords:

Introduction

Genetic events such as gene translocations and mutations have been thought to be the sole driving force for evolution. However, recent studies have suggested that there exist other elements that contribute to the adaptive selection of a population through time. MicroRNAs and segmental duplication are genetic events that have gained much attention due to the new discoveries of their roles in disease manifestation, gene regulation and evolution (1-5).

The prevalence of segmental duplications or low copy repeats in the human genome characterizes and distinguishes humans from lower mammals. This provides an important insight on the potential role of segmental duplications in gene and genomic evolution (6). But the exact mechanism of evolution that allowed the human genome to diverge from that of the lower mammals to its present state remains uncertain.

Gene duplication results in the creation of novel gene function that could possibly drive the process of evolution (7). These duplicated genes are either lost or degraded into pseudogenes or retained through the course of evolution, depending on the selective advantage they confer on the organism (8,9). The new gene duplicates acquire novel functions, though rarely, as they are driven by the force of adaptive evolution. They are subjected to a set of selective constraints that differs from that of the parental genes (10,11). This is the mechanism of selection by which many gene functions come about from novel duplicated segments (12). This is true for a lot of primate-specific gene families (13-17). The ability to retain gene duplicates (or gene duplicability) increases the complexity of an organism derived from high adaptability (18). However, this duplicability varies among the eukaryotic gene families. Under the influence of gene regulation, it has been hypothesized that the primary factor that can explain the vast phenotypic divergence of humans and great apes are changes in regulation rather than having altered protein-coding gene sequences. Through comparative analysis, it has been shown that a

Corresponding address:
Custer C. Deocaris, PhD,
Laboratory of Exercise Biochemistry,
Institute of Health and Sports Sciences,
University of Tsukuba, Japan
Email : cdeocaris@gmail.com

high frequency of evolutionary conserved non-coding sequences exists in the vertebrate genome. These non-coding sequences might provide insight and contain the decisive alternatives for mode of regulation that may have motivated the course of human evolution (19).

This study determines the miRNA distribution in chimpanzees and humans as well as their occurrence within regions of segmental duplication, and proposes a hypothesis on their role in human evolution. This study adds new insights on the mode of miRNA diversification within regions of segmental duplication.

Experimental Procedures

Sources of miRNA sequences- The miRNA sequences of both human (*Homo sapiens*) and chimpanzee (*Pan troglodytes*) were acquired from the miRBase database (20) at http://microrna.sanger.ac.uk/sequences/index.shtml. Stem-loop (precursor) sequences in FASTA format were utilized. As of September 2008, there were 595 chimp miRNA sequences, and as of March 2009, 695 human miRNA sequences were available.

Determination of miRNA distribution within segementally duplicated regions- A BLAT (21) search was done to determine if the human (695) and chimpanzee (595) miRNA sequences are located in SD regions or not. A miRNA sequence was considered if it met the following criteria: sequence identity of 100% with that of the query sequence derived from miRBase; and similar score and span size, with a span of at least 20kb. This size represents the minimum length of a mature or functional miRNA. The BLAT search for human miRNA sequences was applied to the March 2006 Assembly at the Centre for Applied Genomics Human Segmental Duplication Database (http://projects.tcag.ca/humandup/), while the BLAT search for nonhuman organisms, such as the chimpanzee, was done using the November 2003 genome assembly at the Centre for Applied Genomics Non-Human Segmental Duplication Database (http://projects.tcag.ca/xenodup/). Both are essentially databases of SD regions. The location of each miRNA -- chromosome number, strand, start and end bases – was noted.

Determination of miRNA in non-SD regions- Both human and chimpanzee miRNA sequences were also subjected to a BLAST (22) search at http://blast.ncbi.nlm.nih.gov/Blast.cgi to determine possible locations of miRNAs in non-SD regions. The locations of the resulting miRNA sequences were compared with those in the BLAT search. This was done by comparing the coordinates of the miRNAs found using the two algorithms. This was done for both human and chimpanzee miRNA. Locations in BLAT that coincided with those in BLAST indicate that the corresponding miRNAs are restricted to SD regions. Locations in BLAT that are dissimilar with those in BLAST indicate that these miRNAs may be found in SD and in non-SD regions.

MicroRNA Homology Search- Distinguishing non-homologous from homologous miRNA sequences between human and chimpanzee was first achieved by comparing the miRNA IDs between chimp and human (http://microrna.sanger.ac.uk/sequences/). Based on the miRBASE nomenclature, similar numbers and letter suffixes of the ID between miRNA of different species indicate highly conserved or related (homologous) sequences. Human miRNAs with IDs not similar with any of the chimpanzee miRNA IDs were considered non-homologous (or *de novo*) miRNA.

Homologous human and chimpanzee miRNAs were subjected to multiple sequence alignment via ClustalW (available at http://www.ebi.ac.uk/Tools/clustalw/). Pairwise scores (in percent) indicate the degree of similarity between sequences. This verifies if the sequences in the pair are identical and can be considered as homologous human-chimp miRNAs.

Each human miRNA that did not have a chimp miRNA counterpart was subjected to BLAST against the chimpanzee genome. This essentially aligns each human miRNA sequence against the entire chimpanzee genome and allows identification of sequences in the chimp genome (miRNA or non-miRNA) that are homologous to human miRNA. Human miRNAs with 100% identity, maximum of 4 mismatches, and zero gaps were considered homologous to the chimpanzee genome. Searches that generated no results were noted and considered as possible human miRNAs without chimp ancestry. To determine if those human miRNAs homologous to chimpanzee gene sequences are also chimpanzee miRNA, comparisons of BLAST chimpanzee miRNA-chimpanzee genome alignments with BLAST human miRNA-chimpanzee genome alignments were done. These were further aligned with chimpanzee miRNAs using ClustalW to ensure

Table 1. *De novo* human miRNAs resulting from chimpanzee BLAST search

Accession No.	ID no.	Accession No.	ID no.	Accession No.	ID no.
MI0000274	hsa-mir-187	MI0003653	hsa-mir-638	MI0006391	hsa-mir-1257
MI0000299	hsa-mir-222	MI0003662	hsa-mir-647	MI0006395	hsa-mir-548g
MI0000342	hsa-mir-200b	MI0003668	hsa-mir-548d-1	MI0006396	hsa-mir-1261
MI0000449	hsa-mir-132	MI0003669	hsa-mir-661	MI0006400	hsa-mir-548m
MI0000476	hsa-mir-138-1	MI0003670	hsa-mir-662	MI0006402	hsa-mir-548o
MI0000486	hsa-mir-190	MI0003671	hsa-mir-548d-2	MI0006405	hsa-mir-1268
MI0000764	hsa-mir-363	MI0003686	hsa-mir-542	MI0006406	hsa-mir-1269
MI0000824	hsa-mir-325	MI0003763	hsa-mir-767	MI0006411	hsa-mir-548h-1
MI0001641	hsa-mir-429	MI0003815	hsa-mir-1301	MI0006412	hsa-mir-548h-2
MI0002467	hsa-mir-483	MI0003833	hsa-mir-675b	MI0006413	hsa-mir-548h-3
MI0003132	hsa-mir-493	MI0003834	hsa-mir-769	MI0006414	hsa-mir-548h-4
MI0003560	hsa-mir-92b	MI0005416	hsa-mir-675	MI0006419	hsa-mir-1277
MI0003577	hsa-mir-570	MI0005537	hsa-mir-888	MI0006430	hsa-mir-1283-2
MI0003581	hsa-mir-574	MI0006318	hsa-mir-1228	MI0006435	hsa-mir-1255b-1
MI0003592	hsa-mir-585	MI0006321	hsa-mir-1231	MI0006436	hsa-mir-1255b-2
MI0003612	hsa-mir-548a-3	MI0006332	hsa-mir-1200	MI0006441	hsa-mir-1308
MI0003615	hsa-mir-602	MI0006344	hsa-mir-548e	MI0006652	hsa-mir-1321
MI0003616	hsa-mir-603	MI0006347	hsa-mir-1285-2	MI0007075	hsa-mir-1470
MI0003617	hsa-mir-604	MI0006371	hsa-mir-1304	MI0007259	hsa-mir-1538
MI0003634	hsa-mir-620	MI0006376	hsa-mir-548f-3	MI0007260	hsa-mir-1539
MI0003639	hsa-mir-625	MI0006377	hsa-mir-548f-4	MI0008330	hsa-mir-1909
MI0003643	hsa-mir-629	MI0006378	hsa-mir-548f-5	MI0008335	hsa-mir-1914
		MI0006389	hsa-mir-1255a		

that sequences from the chimpanzee genome that will appear homologous to human miRNAs are also miRNAs (i.e. not coding regions of the genome). Searches yielding no similarity indicate that the human miRNA sequences are non-homologous with the chimpanzee genome.

Human miRNAs found to be in the chimp genome but represent non-miRNAs in chimp were subjected to a BLAT search to identify possible locations of these human miRNA sequences in SD regions. Those query sequences (human miRNAs) with 100% identity and a similar query score and span were the criteria used to determine the presence of the query sequence in SD regions.

The methodology used in this study is summarized in Figure S.5.1 and Figure S.5.2.

Results

MicroRNA Distribution. The BLAT search revealed that out of the 595 of the chimpanzee miRNA, 564 are located in least one SD site. Twenty-three did not meet the criteria to be considered a miRNA sequence. No matches were found for two of the chimpanzee miRNA (ptr-mir-1281; ptr-mir-489).

On the other hand, of the 695 human miRNAs, all were found to be in SD regions. Locations for all 695 human miRNAs retrieved from BLAT coincided with those from BLAST.

Table 2. Human miRNAs with similarity to the chimp genome (BLAST score = 100%)

Accession number	ID no.
MI0000650	hsa-mir-200c
MI0000468	hsa-mir-9-3
MI0000093	hsa-mir-92a-1
MI0000112	hsa-mir-105-2
MI0000444	hsa-mir-124-2
MI0000455	hsa-mir-138-2
MI0000464	hsa-mir-153-2
MI0000488	hsa-mir-194-1
MI0003165	hsa-mir-517b
MI0003182	hsa-mir-519a-2
MI0006346	hsa-mir-1285-1
MI0006429	hsa-mir-1282
MI0006657	hsa-mir-1324
MI0003677	hsa-mir-655
MI0006273	hsa-mir-1180
MI0006373	hsa-mir-1243
MI0007074	hsa-mir-1469
MI0006358	hsa-mir-1297
MI0007261	hsa-mir-103-1-as
MI0007262	hsa-mir-103-2-as
MI0003683	hsa-mir-659
MI0006379	hsa-mir-1244
MI0008333	hsa-mir-1912
MI0008336	hsa-mir-1915

MicroRNA Homology Search. Out of the 695 human miRNAs, 162 were not similar with any of the chimpanzee miRNA. BLAST alignment

of these human miRNAs with the chimpanzee genome showed that 24 are conserved (0-4 mismatches, no gaps), while the other 67 had no similarities and hence, were considered human miRNAs without chimp ancestry (or *de novo* miRNAs). The latter are shown in Table 1 and the former are in Table 2.

BLAST results indicate that 13 out of the 24 human miRNAs have 100% identity with known chimpanzee miRNAs and pairwise scores of 100%, whereas the remaining 11 miRNAs showed 100% identity, but less than 100% pairwise scores in ClustalW. This means that 13 human miRNAs are identical to chimp miRNA, while 11 human miRNAs are present in chimpanzees, but not as miRNA. Table 3 shows the 13 human miRNAs with their respective chimpanzee miRNA counterpart identified with ClustalW alignments.

The chimpanzee BLAT search revealed that out of the 11 human miRNAs present in chimpanzee as non-miRNAs, 6 were in SD regions. All the hits from this search met the criteria for *de novo* miRNAs. Table 4 shows the summary of the query size, chromosome locations and scores that resulted from the search. Results of this study are summarized in Figure S.5.3.

Table 3. Human-chimpanzee miRNA pairs with pairwise score of 100%

Human miRNA		Chimpanzee miRNA	
Accession No.	ID no.	Accession No.	ID no.
MI0000093	hsa-mir-92a-1	MI0003000	ptr-mir-92-1
MI0000112	hsa-mir-105-2	MI0002749	ptr-mir-105
MI0000444	hsa-mir-124-2	MI0002765	ptr-mir-124a
MI0000455	hsa-mir-138-2	MI0008539	ptr-mir-138
MI0000464	hsa-mir-153-2	MI0008553	ptr-mir-153
MI0000488	hsa-mir-194-1	MI0003026	ptr-mir-194
MI0003165	hsa-mir-517b	MI0008711	ptr-mir-517b-2
MI0003182	hsa-mir-519a-2	MI0008718	ptr-mir-519a
MI0006346	hsa-mir-1285-1	MI0008486	ptr-mir-1285
MI0006429	hsa-mir-1282	MI0008481	ptr-mir-1282-1
MI0006657	hsa-mir-1324	MI0008531	ptr-mir-1324-1
		MI0008532	ptr-mir-1324-2
MI0006358	hsa-mir-1297	MI0008500	ptr-mir-1297-1
MI0006379	hsa-mir-1244	MI0008441	ptr-mir-1244-2
		MI0008442	ptr-mir-1244-3
		MI0008444	ptr-mir-1244-5
		MI0008445	ptr-mir-1244-6
		MI0008447	ptr-mir-1244-8
		MI0008450	ptr-mir-1244-11

Table 4. BLAT Human miRNA with non-miRNA counterpart located in SD regions in chimpanzee genome.

Query	Score	Start	End	QSize	Identity	Chromosome	Strand	Start	End	Span
hsa-mir-103-1-as	62	1	62	62	100.00%	4	+	175165060	175165121	62
hsa-mir-103-2-as	62	1	62	62	100.00%	21	-	3821998	3822059	62
hsa-mir-1243	93	1	93	93	100.00%	3	+	131042991	131043083	93
hsa-mir-1469	47	1	47	47	100.00%	16	+	95697059	95697105	47
hsa-mir-1912	80	1	80	80	100.00%	X	+	117570186	117570265	80
hsa-mir-1915	80	1	80	80	100.00%	8	-	22032997	22033076	80

Discussion

It was discovered that the human miRNA sequences lie only in SD regions. With some exceptions, miRNA locations derived from the BLAT search of the Human Segmental Duplication Database were similar to those using BLAST and miRBase. For chimpanzees, on the other hand, BLAST-determined locations did not match those with BLAT. Not one result from BLAT showed similar coordinates with BLAST and miRBASE. This indicates that chimpanzee miRNAs are not found exclusively in SD regions.

What then is the significance of the preferential distribution of the human miRNAs in SD regions?

The exclusivity of human miRNAs in SD regions provides a major insight on the role of segmental duplications in the diversification of human miRNA. The duplicative transposition characteristic of segmental duplications probably provides a mechanism by which miRNAs can be derived from these transposable elements. The ubiquity, abundance and high evolutionary rate of transposable elements, such as those in segmental duplications, provide an impetus for miRNAs to emerge from such sequences. That the transposable elements are lineage-specific and mostly non-conserved suggests that miRNAs derived from such elements could also acquire the non-conserved trait that could provide diversifying regulatory effects on multiple genes.

Segmental duplication events may have led to the production of *de novo* miRNAs, which in turn creates new functions (e.g. biochemical pathways, proteins, etc.). The presence of miRNAs in SD suggests that these miRNAs, through the course of evolution, may have been subjected to different modes of mutation and selection, which resulted in specialization or novel functions. It is possible that the presence of human miRNAs solely in SD regions conferred the selective advantages in terms of miRNA diversification. MiRNAs could have possibly acquired the opportunity in SD regions to diversify and be stably maintained in the genome, possibly through sub- or neofunctionalization. The interplay between conserved miRNAs and the miRNAs that are not of chimp ancestry might have contributed to increasing the functional complexity, allowing that certain organism to acquire new traits. And if these *de novo* miRNAs are species-specific, it would entail species-specific traits. Further, if these non-chimp human miRNAs have brain-specific targets, this could necessitate human-specific traits, which could explain the differences between human and chimpanzee brains.

De novo miRNAs that are not of chimp ancestry were identified in this study. These might have played a role in the advancement of humans from the chimpanzees in molecular, anatomical and social aspects. Human miRNAs that were not found in any part of the chimpanzee genome might have particular gene targets in the human genome that corresponds to brain-specific functions. The presence of these miRNAs and their corresponding targets could provide a possible mechanism by which humans developed complex learning skills, language, enhanced memory and recall, and heightened capacity for interpretations, which led to the ability for more advanced cognitive tasks and develop social interactions and relationships with one another (23). Human miRNAs that were found in the chimp genome, but as non-miRNA, supports the idea that most miRNAs originate from non-miRNA regions, and that these may have been selected to become miRNAs depending on the environment (positive selection). It should be noted that a number of these human miRNA with counterparts in the chimpanzee genome are located in SD regions. This indicates that segmental duplication could be one of the mechanisms miRNAs evolve and possibly acquire new functions.

Conclusion and Future Directions

This study shows that human miRNAs are found exclusively in SD regions, while chimpanzee miRNA are not. Human miRNAs possibly derived from non-miRNA chimp gene sequences are present in humans, and some of these are found in SD regions in chimps. *De novo* miRNAs without chimp ancestry are also present in humans.

The presence of *de novo* miRNAs exclusively in human SD regions suggests that, over evolutionary time, these genes could have been subjected to numerous duplications. Some of these duplications could have provided sites for mutations to occur, particularly in miRNAs. Consequently, changes in miRNA could have resulted in the creation of functions specific to humans, for example, brain-specific ones. The fact that a number of human miRNAs are found in SD regions also indicates that segmental duplication could be one of the primary modes for miRNA evolution.

The preferential distribution of human miRNAs in segmentally duplicated regions supports the hypothesis that new miRNAs can be derived from transposable elements, particularly that of SD regions. However, we recommend that identification and annotation of the human miRNA gene targets be performed to determine brain-specific targets of both the homologous and *de novo* human miRNAs. Possible promoters and regulatory regions that might be adjacent to the novel miRNAs should be identified to determine their role in gene regulation and transcription. Network on the gene targets relevant to brain function should be performed to derive significant connections between brain-specific regulated genes.

References

[1] Eichler EE. *Trends Genet.* 2001; **17**(11),661-9

[2] Pillai R S. *RNA* 2005; **11**: 1753-1761

[3] Maher C, Stein L and Ware D. *Genome Res* 2006; **16**:510-519

[4] Zhao Y and Srivastava D. *Trends Biochem Sci* 2007; **32**(4):189-97

[5] Liu N, Okamura K, Tyler D, Phillips M, Chung W and Lai E. *Cell Res* 2008; 985-996

[6] Bailey J A, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW and Eichler EE. *Science* 2002; **297**:1003–1007

[7] Zhang J. *Trends in Ecology and Evolution* 2003; **18**:292-298

[8] Nadeau JH**,** Sankoff D. *Genetics* 1997; **147:** 1259-1266

[9] Li WH, Gu Z, Wang H, Nekrutenko A. *Nature* 2001; **409:** 847-849

[10] Walsh JB.*Genetics* 1995; **139:** 421-428

[11] Lynch M, Conery JS.*Science* 2000; **290:** 1151-1155

[12] Prince VE, Pickett FB.*Nat Rev Genet* 2002; **3:** 827-837

[13] Courseaux A**,**Nahon JL.*Science* 2001; **291:** 1293-1297

[14] Johnson ME, Viggiano L, Bailey JA, Abdul-Rauf M, Goodwin G, Rocchi M, Eichler E.E.*Nature* 2001; **413:** 514-519

[15] Maston GA, Ruvolo M. *Mol Biol Evol* 2002; **19:** 320-335

[16] Zhang J, Zhang YP, Rosenberg HF. *Nat Genet* 2002; **30:** 411-415

[17] Paulding CA, Ruvolo M, Haber DA. *Proc Natl Acad Sci* 2003; **100:**2507-2511

[18] Yang J, Lusk R, Li WH. *Proc Natl Acad Sci* 2003; **100:**15661-15665

[19] Bird CP, Stranger BE, Liu M, Thomas DJ, Ingle CE, Beazley C, Miller W, Hurles ME, Dermitzakis ET. *Genome Biol 2007;* **8:**R118

[20] Griffiths-Jones S. *Nucleic Acids Res* 2004; **32**: D109-D111 (http://microrna.sanger.ac.uk/sequences/index.shtml.)

[21] Kent W J. (2002) *Genome Res* **12**, 656-664

[22] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. *J Mol Biol 1990; **215** (3): 403–410.*

[23] Bradbury J. *PLoS Biol* 2005; **3**(3):e50