PENSOFT.

MUSEUM FÜR
NATURKUNDE
BERLIN

Deciphering conserved identical sequences of mature miRNAs among six members of great apes

Aftab Ali Shah[1], Mushtaq Ahmad[1], Taqweem-Ul-Haq[1]

1   Department of Biotechnology, University of Malakand, Chakdara, Pakistan

http://zoobank.org/84AC151C-9946-42BB-A41F-F30BAAB92867

Corresponding author: *Aftab Ali Shah* (aftabuom@gmail.com)

Abstract

MicroRNAs (miRNAs) are a group of small RNA molecules which act as negative regulators of gene expression by controlling post-transcriptional regulation through binding to their corresponding mRNAs. Due to their small size, their nucleotide compositions are expected to be similar, but until now, the extent of similarity has not been reported in humans and their six phylogenetically closely related members of hominids. The present study allows direct comparison among six members of hominid species (*Homo sapiens*, *Gorilla gorilla*, *Pan paniscus*, *Pongo pygmaeus*, *Pan troglodytes* and *Symphalangus syndactylus*) in terms of their miRNA repertoire, their evolutionary distance to human, as well as, the categorization of identical species-specific miRNAs. For this purpose, a total of 2694, 370, 157, 673, 590 and 10 mature miRNA sequences of *Homo sapiens*, *Gorilla gorilla*, *Pan paniscus*, *Pongo pygmaeus*, *Pan troglodytes* and *Symphalangus syndactylus* respectively were retrieved from miRbase 22. A total of 12, 4, 4 and 3 conserved clusters with identical miRNA sequences that belong to the same gene families were found in *Homo sapiens*, *Gorilla gorilla*, *Pongo pygmaeus*, *Pan troglodytes* respectively by neighbor-joining method using MEGA7 software. Interestingly, cross-species comparison has also shown a set of conserved identical miRNA sequences. Homologs of human mature miRNAs with 100% sequence identity are expected to have similar functions in the studied primates. Further *in-vitro* study is required to investigate common targets for identical miRNAs in the studied primates.

Introduction

MiRNAs are small (19–23nt) RNA molecules that regulate messenger RNA through binding to their 3'-UTR, mediated by the RNA induced silencing (RISC) complex in all living organisms (Zhang et al. 2018). Binding of miRNA to their corresponding target mRNA leads to translational repression and/ or mRNA degradation (Unterbruner et al. 2018). To date, considerable number of mature miRNAs have been identified in *Homo sapiens*, *Gorilla gorilla*, *Pan paniscus*, *Pongo pygmaeus* and *Pan troglodytes*, *Symphalangus syndactylus* (no=2694, 370, 157, 673, 590, 10) respectively, as shown in miRbase database (http://www.mirbase.org/) (Griffiths-Jones et al. 2006; Kozomara and Griffiths-Jones 2013).

It is already known that multiple miRNAs are produced from the same primary transcript and majority of miRNA clusters are transcribed as a single unit (Marco et al. 2013). The evolutionary importance of miRNA clusters has been the subject of much speculation (Wang et al. 2016). Many clusters contain members of the same family, suggesting an important role of gene duplication in their evolution (Berezikov 2011). On the other hand, some miRNA clusters also contain members of different miRNA families, particularly in animal kingdom (McCreight et al. 2017). Like other gene families, miRNAs are also prone to forming paralogs, with the result that many miRNAs appear as members of families as homologs (Hertel et al. 2006). However, the origin and evolution of these miRNA clusters has not been investigated in detail (Altuvia et al. 2005; Tanzer and Stadler 2004). Phylogenetic studies have shown that miRNAs are present throughout the evolution of metazoans. Comparison of pre-miRNA sequences demonstrate that they are less conserved and
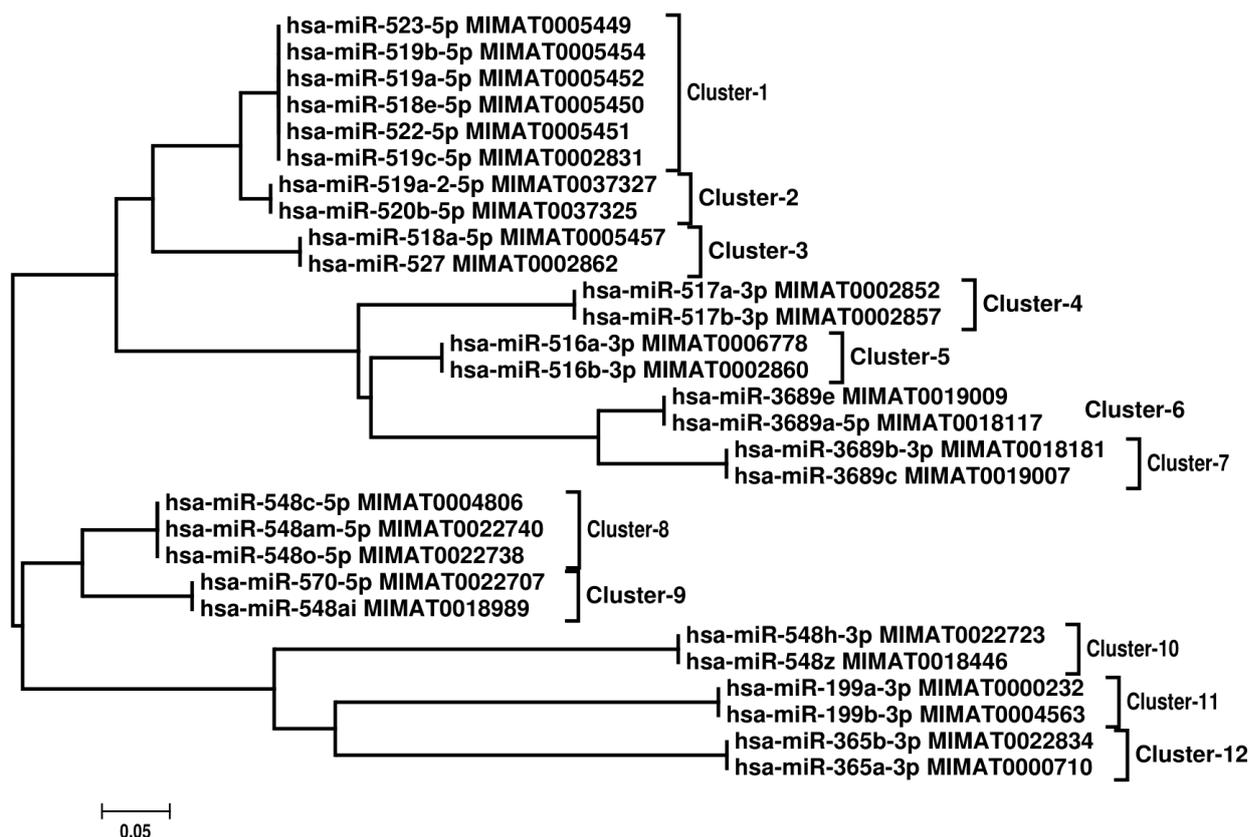
Figure 1. Shows evolutionary relationships of taxa for all mature miRNAs in *Homo sapiens*. The analysis involved 29 nucleotide sequences.

therefore are more prone to phylogenetically preserved than the mature sequences alone. High degree of identity across different species was observed for mature miRNAs (Li et al. 2010). It is also noted that many matured miRNAs are prevailing in several species and are highly conserved and are confined to specific lineages. There are several polycistronic transcripts that suggest a potential mode of evolution for polycistronic miRNAs (Truscott et al. 2016). It is already known that the miRNA repertoire has continuously increased during evolution of metazoan. However, the advent ratio of these molecules is diverse over evolutionary time (Bartel 2018). The expansions of miRNA have been linked with evolutionary innovations that lead to the diversification of bilaterians. Till now, identification of orthologous miRNAs in different species has been investigated in primates.

Among the six members of great apes, *Homo sapiens* are the deepest explored group with 2694 mature miRNAs miRNAs described. In the present study, we took advantage of a recently available set of mature miRNA from six members of the great ape population to systematically detect identical miRNA by comparing patterns of intra- and inter-species sequence similarity and their evolutionary distance. Interestingly, it was found that intra- and inter-species sequence set of identical mature miRNA exists in great apes including humans. Further *in-vitro* study is required to investigate common targets for identical miRNAs in the studied primates.

Materials and methods

Alignment of sequences and phylogenetic analysis

For miRNA, a very limited open and free data is available. The miRBase is one of the highly referred databases, easily accessible and in its latest release 10883 pre-miRNAs are available. Dataset of mature miRNAs sequences of *Homo sapiens* (no=2694), *Gorilla gorilla* (no=370), *Pan paniscus* (no=157), *Pongo pygmaeus* (no=673 mature), *Pan troglodytes* (no=590 mature) and *Symphalangus syndactylus* (no=10 mature) were retrieved from miRBase sequence database (a data repository of published miRNA sequences and its annotation) (release 22.0) at http://microrna.sanger.ac.uk. ClustalW was used to generate multiple alignments of nucleic acid sequences (Chenna et al. 2003) and MEGA7 was used to generate phylogenetic analyses using Neighbor-Joining method (Kumar et al. 2016).

Results

Identification of Homologous mature miRNA sequences in intra-species in hominides

Homologous sequences in *Homo sapiens* were clustered based on their phylogenetic relationship and sequence identity using ClustalW. Multiple alignment of mature miRNAs revealed a conserved consensus. Neighbor-Join-

Table 1. List of miRNAs grouped into clusters, their genomic coordinates, gene family names and their matured miRNA sequences in *Homo sapiens*.

S. No	Members	Gene Family Name	Genomic coordinate	Mature miRNA sequence
Cluster 1	hsa-miR-523-5p		chr19: 53698385-53698471 [+]	CUCUAGAGGGAAGCGCUUUCUG
	hsa-miR-519b-5p		chr19: 53695213-53695293 [+]	
	hsa-miR-519a-5p		chr19: 53752397-53752481 [+]	
	hsa-miR-518e-5p		chr19: 53729838-53729925 [+]	
	hsa-miR-522-5p		chr19: 53751211-53751297 [+]	
	hsa-miR-519c-5p		chr19: 53686469-53686555 [+]	
Cluster 2	hsa-miR-519a-2-5p	MIPF0000020; mir-515	chr19: 53762344-53762430 [+]	CUGCAAAGGGAAGCCCUUUC
	hsa-miR-519b-2-5p		chr19: 53754018-53754102 [+]	
Cluster 3	hsa-miR-518a-5p		chr19: 53731006-53731090 [+]	
	hsa-miR-527		chr19: 53754018-53754102 [+]	
Cluster 4	hsa-miR-517a-3p		chr19: 53712268-53712354 [+]	AUCGUGCAUCCCUUUAGAGUGU
	hsa-miR-517b-3p		chr19: 53721076-53721142 [+]	
Cluster 5	hsa-miR-516a-3p		chr19: 53756741-53756830 [+]	UGCUUCCUUUCAGAGGGU
	hsa-miR-516b-3p		chr19: 53736845-53736934 [+]	
Cluster 6	hsa-miR-3689e	MIPF0001144; mir-3689	chr9: 134850570-134850641 [-]	UGUGAUAUCAUGGUUCCUGGGA
	hsa-miR-3689a-5p		chr9: 134849487-134849564 [-]	
Cluster 7	hsa-miR-3689b-3p		chr9: 134850125-134850272 [-]	CUGGGAGGUGUGAUAUUGUGGU
	hsa-miR-3689c		chr9: 134849298-134849369 [-]	
Cluster 8	hsa-miR-548c-5p	MIPF0000317; mir-548	chr12: 64622509-64622605 [+]	AAAAGUAAUUGCGGUUUUUGCC
	hsa-miR-548am-5p		chrX: 16627012-16627085 [-]	
Cluster 9	hsa-miR-570-5p		chr3: 195699401-195699497 [+]	AAAGGUAAUUGCAGUUUUUCCC
	hsa-miR-548ai		chr6: 99124609-99124696 [+]	
Cluster 10	hsa-miR-548h-3p		chr8: 27048853-27048963 [-]	CAAAAACCGCAAUUACUUUUGCA
	hsa-miR-548z		chr12: 64622509-64622605 [-]	
Cluster 11	hsa-miR-199a-3p	MIPF0000040; mir-199	chr19: 10817426-10817496 [-]	ACAGUAGUCUGCACAUUGGUUA
	hsa-miR-199b-3p		chr9: 128244721-128244830 [-]	
Cluster 12	hsa-miR-365b-3p	MIPF0000061; mir-365	chr17: 31575411-31575521 [+]	UAAUGCCCCUAAAAAUCCUUAU
	hsa-miR-365a-3p		chr16: 14309285-14309371 [+]	

ing method was used for inferring the evolutionary history. The optimal tree with the sum of branch length = 2.32453654 is shown in Figure 1. The p-distance method was used for computing the evolutionary distances. The analysis involved 29 nucleotide sequences. All uncertain positions were deleted for each sequence pair. In the final dataset, there was a total of 36 positions. Evolutionary analyses were conducted in MEGA7. The conserved miRNA sequences were grouped into 12 clusters. Number of miRNA members in each cluster, showing 100% identity in their sequences and their corresponding genomic coordinates, are shown in Table 1. It is interesting to note that the genomic location of all the identical miRNAs are in clustered form. This indicates that the genes for these identical miRNAs are originated through gene duplication during evolution and speciation. Identical miRNA sequences were also noted in *Gorilla gorilla* using the same phylogenetic model. The optimal tree with the sum of branch length = 0.40606061 is shown in Figure 2. The analysis involved 12 nucleotide sequences. There were a total of 35 positions in the final dataset. However, only 12 miRNAs were found to be 100% identical. These 12 miRNAs are grouped into four clusters. Number of miRNA members in each cluster showing 100% identity in their sequences and their corresponding genomic coordinates are shown in Table 2. Interestingly, all the identified conserved miRNAs

belong to a single gene family (MIPF0000020; mir-515). Similarly, identical miRNA sequences were also noted in *Pongo pygmaeus* using the same phylogenetic model as shown in Figure 3. The total number of identical miRNAs were 8 that were grouped into 4 clusters as shown in Table 3. The optimal tree with the sum of branch length = 1.37798461 is shown in Figure 3. The analysis involved 8 nucleotide sequences. There were a total of 32 positions in the final dataset. Likewise, identical miRNA sequences were also noted in *Pan troglodytes* using the same phylogenetic model. Seven conserved miRNAs were also found in *Pan troglodytes* that were grouped into three clusters as shown in Table 4. The optimal tree with the sum of branch length = 0.71717172 is shown in Figure 4. The analysis involved 7 nucleotide sequences. There were a total of 26 positions in the final dataset.

Identification of Homologous mature miRNA Sequences in inter-species in Hominides

To assess whether any cross-species conserved miRNA in hominids exists, all known matured miRNAs were aligned to generate multiple alignments of nucleic acid sequences using ClustalW, and MEGA7 was used to generate phylogenetic analyses. The optimal tree with the sum of branch length = 3.16412289 is shown in Figure 4. There were a total of 31 positions in the final dataset.
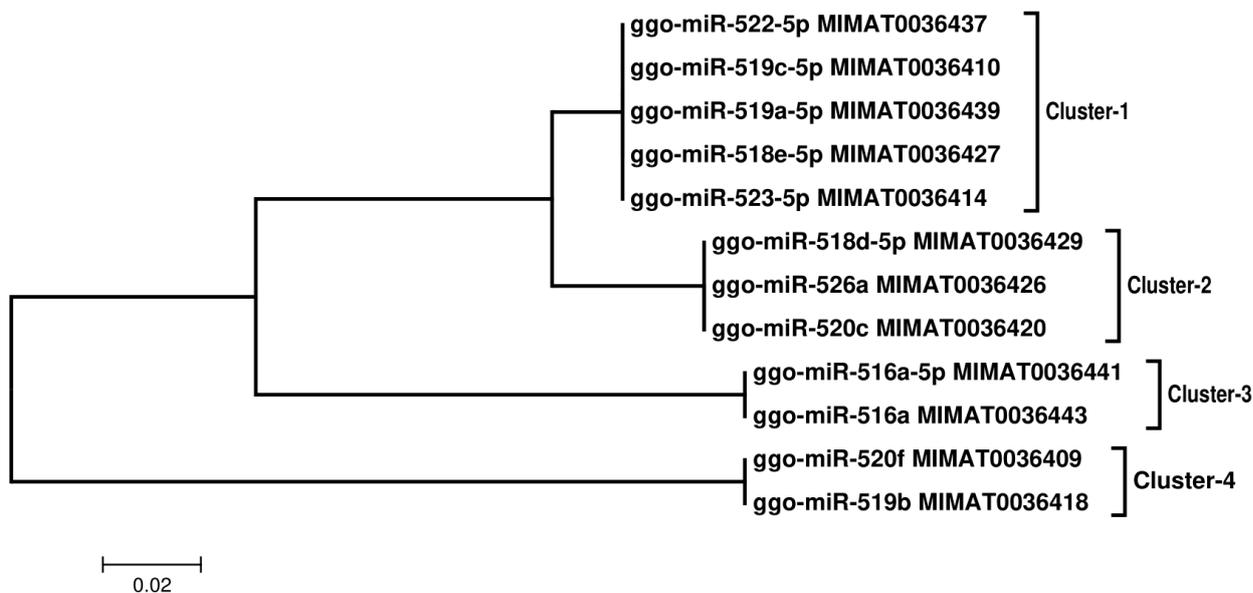
Figure 2. Represents evolutionary relationships of taxa for *Gorilla gorilla*. The optimal tree with the sum of branch length = 0.40606061 is shown. The analysis involved 12 nucleotide sequences.
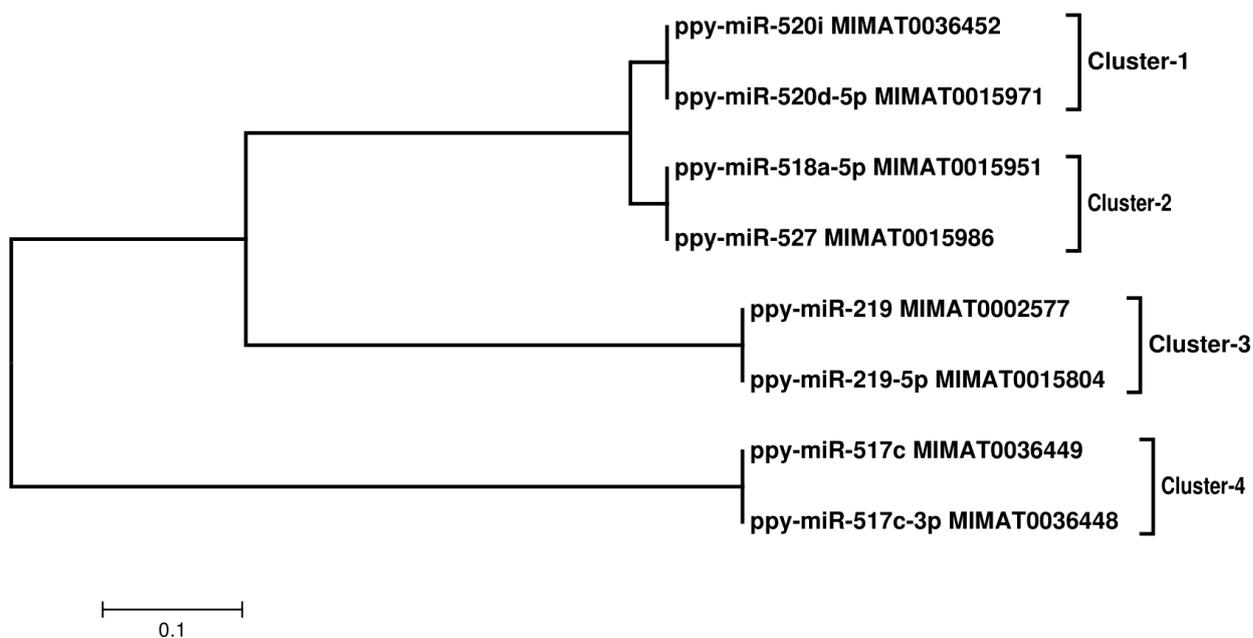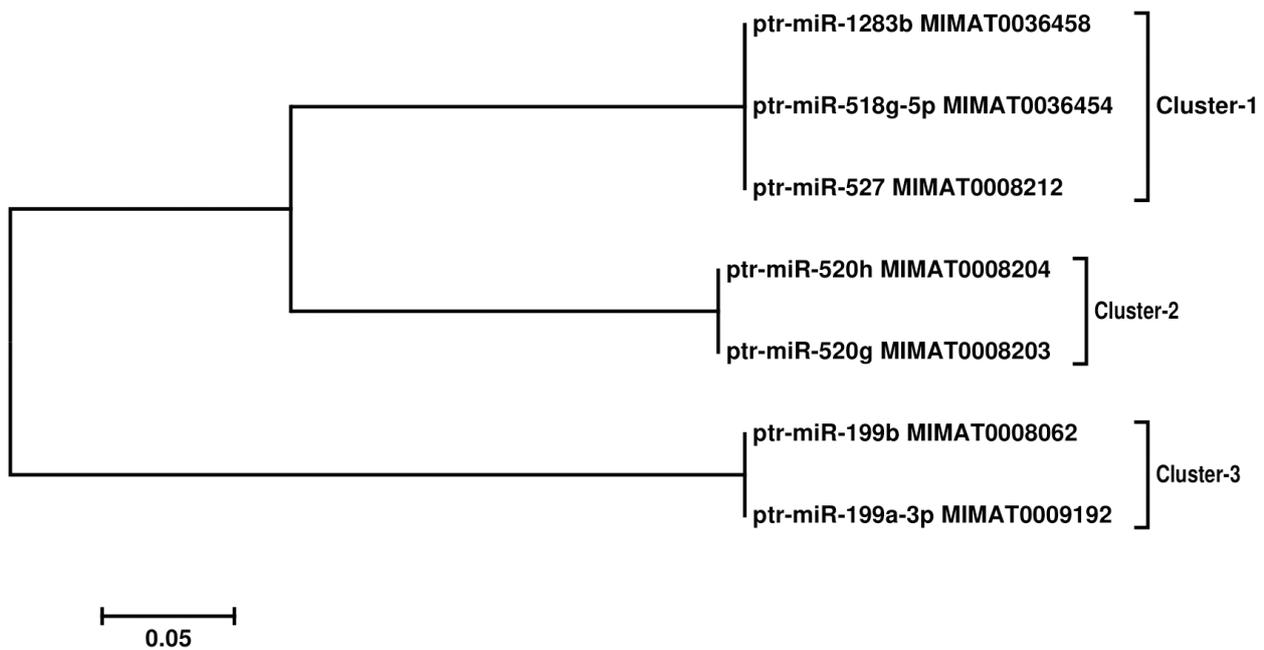

Figure 3. Illustrate evolutionary relationships of taxa for *Pongo pygmaeus*. The analysis involved 8 nucleotide sequences.

Using online tool jvenn (Bardou et al. 2014), it was interesting to note that one miRNA (CUCUAGAGGGAAGC-GCUUUCUG) was conserved and overlapping in *Homo sapiens* and *Gorilla gorilla*. Similarly, another conserved sequence (CUGCAAAGGGAAGCCCUUUC) was overlapping in all the three members of hominde species (*Homo sapiens*, *Pan troglodytes* and *Pongo pygmaeus*) as shown in Figures 5, 6. It was also noted that one miRNA (ACAGUAGUCUGCACAUUGGUUA) was overlapping in *Homo sapiens* and *Pan troglodytes*.

Discussion

MiRNA-mediated gene regulation is novel mechanism among all lineages in animal kingdom (Zhang et al. 2004). Due to their smaller size, many known miRNA genes in animal genomes are found as clusters. MiRNA clusters are a group of related miRNAs closely localized in the genome with an evolution that remains poorly understood (Chen et al. 2015). Therefore, most of the clusters are transcribed as a single polycistronic transcripts (Lagos-Quintana et
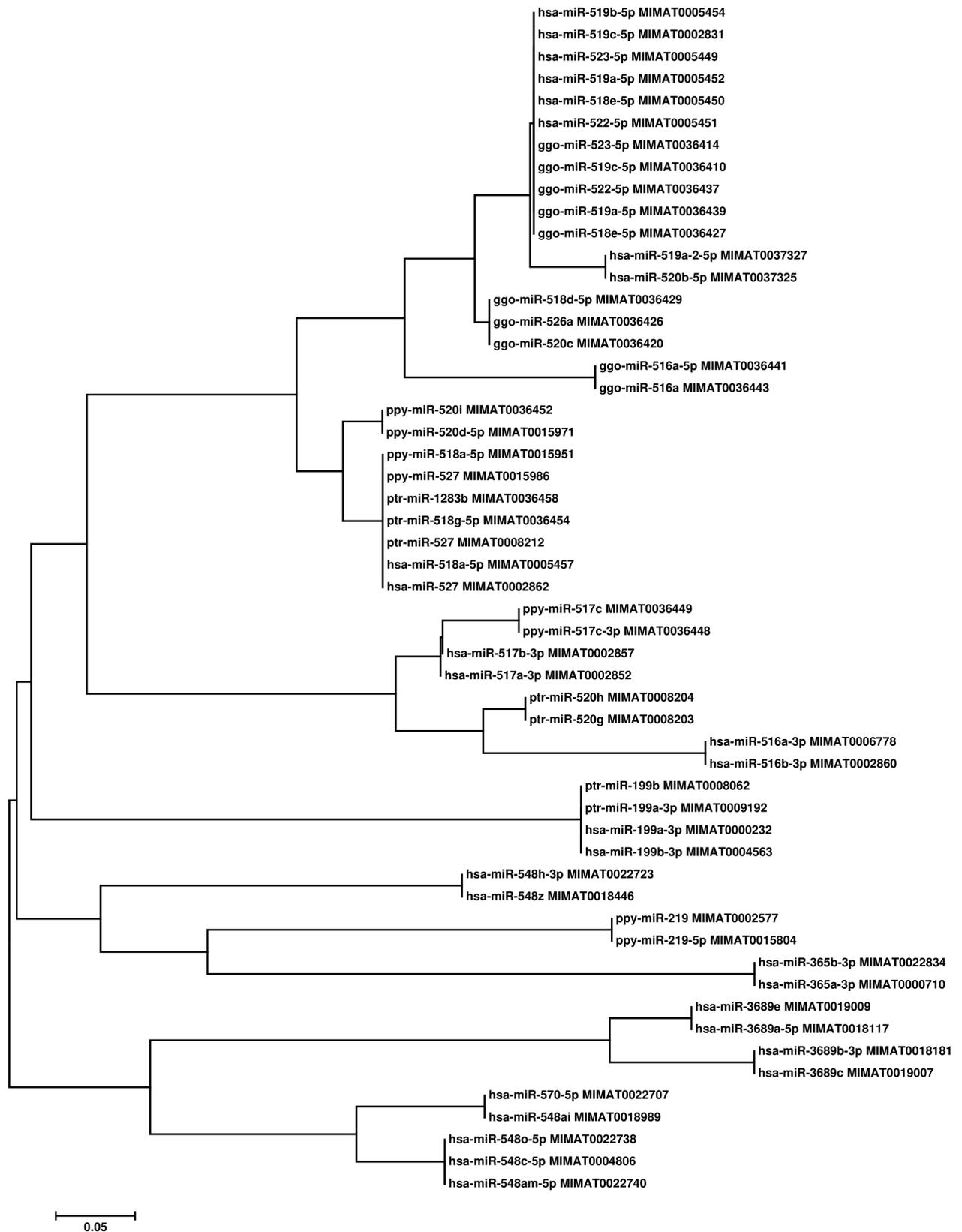
Figure 4. Represents evolutionary relationships of taxa for *Pan troglodytes*. The optimal tree with the sum of branch length = 0.71717172 is shown. The analysis involved 7 nucleotide sequences.

Table 2. List of miRNAs grouped into clusters, their genomic coordinates, gene family names and their matured miRNA sequences in *Gorilla gorilla*.

S. No	No. of members	Gene family	Genomic coordinate	Mature miRNA sequence
Cluster 1	ggo-miR-522-5p	MIPF0000020; mir-515	chr19: 53095952-53096058 [+]	CUCUAGAGGGAAGCGCUUUCUG
	ggo-miR-519c-5p		chr19: 53038851-53038969 [+]	
	ggo-miR-519a-5p		chr19: 53097153-53097260 [+]	
	ggo-miR-518e-5p		chr19: 53073319-53073416 [+]	
	ggo-miR-523-5p		chr19: 53042042-53042160 [+]	
Cluster 2	ggo-miR-518d-5p		chr19: 53078317-53078435 [+]	
	ggo-miR-526a		chr19: 53069990-53070073 [+]	
	ggo-miR-520c		chr19: 53049906-53050015 [+]	
Cluster 3	ggo-miR-516a-5p		chr19: 53101171-53101280 [+]	UUCUCGAGGAAAGAAGCACUUUC
	ggo-miR-516a		chr19: 53105582-53105691 [+]	
Cluster 4	ggo-miR-520f		chr19: 53025391-53025489 [+]	AAGUGCUUCCUUUUAGAGGGUU
	ggo-miR-519b		chr19: 53044871-53044969 [+]	

Table 3. List of miRNAs grouped into clusters, their genomic coordinates, gene family names and their matured miRNA sequences in *Pongo pygmaeus*.

S. No	No. of members	Gene Family	Genomic coordinate	Mature miRNA sequence
Cluster 1	ppy-miR-520i	MIPF0000020; mir-515	chr19: 55513927-55514025 [+]	CUACAAAGGGAAGCCCUUUC
	ppy-miR-520d-5p		chr19: 55491990-55492076 [+]	
Cluster 2	ppy-miR-518a-5p		chr19: 55505640-55505726 [+]	CUGCAAAGGGAAGCCCUUUC
	ppy-miR-527		chr19: 55533164-55533250 [+]	
Cluster 3	ppy-miR-219	MIPF0000044; mir-219	NW_002874576.1: 1044940-1045049 [+]	UGAUUGUCCAAACGCAAUUCU
	ppy-miR-219-5p		chr9: 125361031-125361127 [-]	
Cluster 4	ppy-miR-517c	MIPF0000020; mir-515	chr19: 55485875-55485961 [+]	AUCGUGCAUCCCUUUAGAGUGU
	ppy-miR-517c-3p			

Figure 5. A comparative evolutionary relationships of taxa for *Homo sapiens*, *Gorilla gorilla*, *Pongo pygmaeus* and *Pan troglodytes.* The analysis involved 54 nucleotide sequences.

Table 4. List of miRNAs grouped into clusters, their genomic coordinates, gene family names and their matured miRNA sequences in *Pan troglodytes*.

S. No	No. of members	Gene family	Genomic coordinate	Mature miRNA sequence
Cluster 1	ptr-miR-1283b	MIPF0000020; mir-515	chr19: 56105638-56105756 [+]	CUGCAAAGGGAAGCCCUUUC
	ptr-miR-518g-5p		chr19: 56078436-56078532 [+]	
	ptr-miR-527		chr19: 56101441-56101526 [+]	
Cluster 2	ptr-miR-520h		chr19: 56089945-56090033 [+]	ACAAAGUGCUUCCCUUUAGAGUGU
	ptr-miR-520g		chr19: 56069121-56069209 [+]	
Cluster 3	ptr-miR-199b	MIPF0000040; mir-199	chr9: 106385739-106385847 [-]	ACAGUAGUCUGCACAUUGGUUA
	ptr-miR-199a-3p		chr19: 11396978-11397047 [-]	
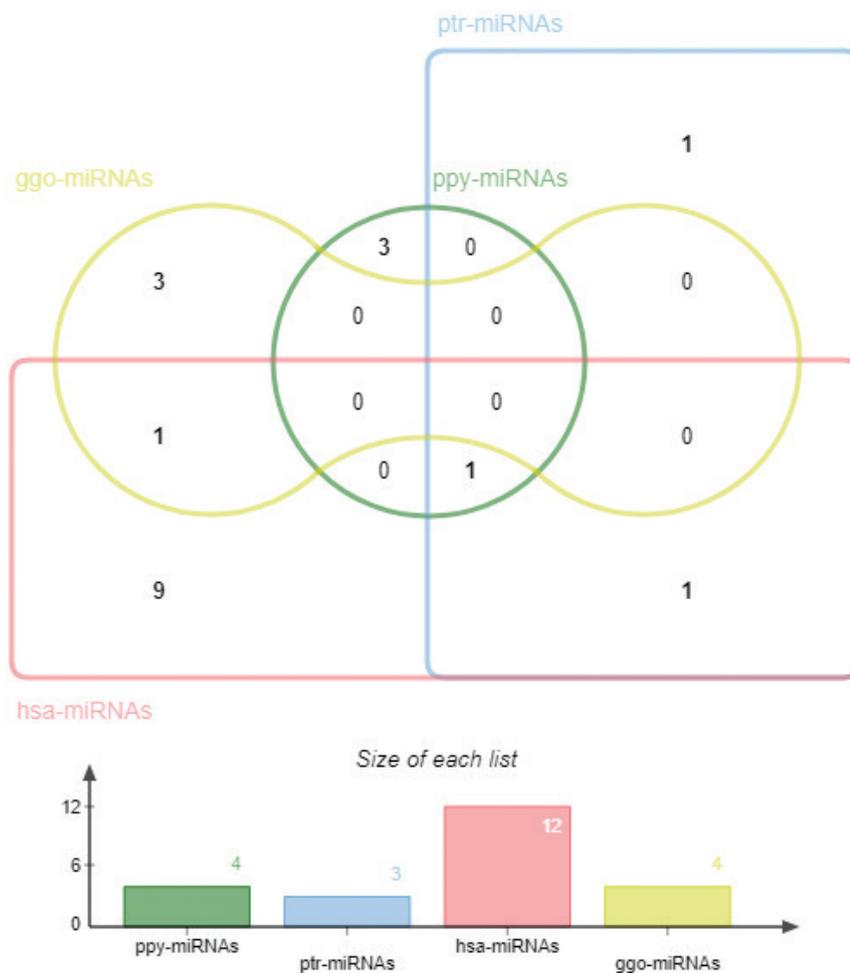
Figure 6. Venn diagram representing number of overlapping miRNA in *Homo sapiens*, *Gorilla gorilla*, *Pongo pygmaeus* and *Pan troglodytes*.

al. 2003; Mourelatos et al. 2002). It was found that these clusters are highly conserved in most mammals. Insertions of new miRNAs, deletions of individual miRNAs, and a cluster duplication observed in different species suggest an actively evolving cluster. In the present study, intra-species and inter-species conserved identical miRNAs were identified in the six species of hominoids. Interestingly, there were few miRNAs that were conserved across all studied species, indicating their evolutionary distance to humans, as well as, the categorization of identical species-specific miRNAs were identified in six members of great apes. It

was investigated that most conserved miRNA clusters in all the studied members of hominoids belong to the two families i.e., mir-515 and mir-199, suggesting that the ancestral clusters may be originated by tandem duplication. It has been demonstrated that some miRNA genes exhibited the phenomena of clustering (Kurkewich et al. 2018). Interestingly, several studies have shown that miRNA clusters comprise of two or more miRNA genes that display high level of identity in sequences. Moreover, they are situated contiguously with each other in the genome (Gonzalez-Vallinas et al. 2018). Through experimental

and computational identification, the miRBase database is one of the primary repository resources for collecting miRNA genes (Griffiths-Jones et al. 2007; Kozomara and Griffiths-Jones 2013). No futuristic data related to miR-NA clusters is available in miRBase. Similarly, no further information about miRNA gene clusters was rendered to explore the evolutionary conservation between miRNA clusters across several species. The present data highlighted intra and inter phylogenetic relationship of matured miRNA in six species of great apes.

Conclusion

In this comparative study, conserved identical miRNA sequences were found among four hominid species. The applied prediction algorithm (mentioned in the materials and method section) proves several criteria based on similarity to identified conserved sequences of miRNAs to detect both more distantly-related and closely-related homologs. Further study is required to identify potential targets for identical miRNAs in the studied primates.

Compliance with ethical standards

Conflict of interest: No conflict of interest exists.

Ethical approval: This article does not contain any study with human participants or animals performed by any of the authors.

References

Altuvia Y et al. (2005) Clustering and conservation patterns of human microRNAs Nucleic acids research 33: 2697–2706. https://doi.org/10.1093/nar/gki567

Bardou P, Mariette J, Escudie F, Djemiel C, Klopp C (2014) jvenn: an interactive Venn diagram viewer BMC Bioinformatics 15: 293. https://doi.org/10.1186/1471-2105-15-293

Bartel DP (2018) Metazoan MicroRNAs Cell 173:20–51. https://doi.org/10.1016/j.cell.2018.03.006

Berezikov E (2011) Evolution of microRNA diversity and regulation in animals Nature Reviews Genetics 12: 846. https://doi.org/10.1038/nrg3079

Chen L, Heikkinen L, Emily Knott K, Liang Y, Wong G (2015) Evolutionary conservation and function of the human embryonic stem cell specific miR-302/367 cluster Comp Biochem Physiol Part D Genomics Proteomics 16: 83–98. https://doi.org/10.1016/j.cbd.2015.08.002

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs Nucleic acids research 31: 3497–3500. https://doi.org/10.1093/nar/gkg500

Gonzalez-Vallinas M, et al. (2018) Epigenetically Regulated Chromosome 14q32 miRNA Cluster Induces Metastasis and Predicts Poor Prognosis in Lung Adenocarcinoma Patients Mol Cancer Res 16: 390–402. https://doi.org/10.1158/1541-7786.MCR-17-0334

Griffiths-Jones S, Grocock RJ, Van Dongen S, Bateman A, Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature Nucleic acids research 34: D140–D144. https://doi.org/10.1093/nar/gkj112

Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2007) miR-Base: tools for microRNA genomics Nucleic acids research 36: D154–D158. https://doi.org/10.1093/nar/gkm952

Hertel J et al. (2006) The expansion of the metazoan microRNA repertoire BMC genomics 7: 25. https://doi.org/10.1186/1471-2164-7-25

Kozomara A, Griffiths-Jones S (2013) miRBase: annotating high confidence microRNAs using deep sequencing data Nucleic acids research 42: D68–D73. https://doi.org/10.1093/nar/gkt1181

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets Molecular biology and evolution 33: 1870–1874. https://doi.org/10.1093/molbev/msw054

Kurkewich JL, Boucher A, Klopfenstein N, Baskar R, Kapur R, Dahl R (2018) The mirn23a and mirn23b microrna clusters are necessary for proper hematopoietic progenitor cell production and differentiation Exp Hematol 59: 14–29. https://doi.org/10.1016/j.exphem.2017.12.007

Lagos-Quintana M, Rauhut R, Meyer J, Borkhardt A, Tuschl T (2003) New microRNAs from mouse and human RNA 9: 175–179. https://doi.org/10.1261/rna.2146903

Li SC, Chan WC, Hu LY, Lai CH, Hsu CN, Lin WC (2010) Identification of homologous microRNAs in 56 animal genomes Genomics 96: 1–9. https://doi.org/10.1016/j.ygeno.2010.03.009

Marco A, Ninova M, Ronshaugen M, Griffiths-Jones S (2013) Clusters of microRNAs emerge by new hairpins in existing transcripts Nucleic acids research 41: 7745–7752. https://doi.org/10.1093/nar/gkt534

McCreight JC, Schneider SE, Wilburn DB, Swanson WJ (2017) Evolution of microRNA in primates. PloS ONE 12: e0176596. https://doi.org/10.1371/journal.pone.0176596

Mourelatos Z et al. (2002) miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs Genes Dev 16: 720–728. https://doi.org/10.1101/gad.974702

Tanzer A, Stadler PF (2004) Molecular evolution of a microRNA cluster Journal of molecular biology 339: 327–335. https://doi.org/10.1016/j.jmb.2004.03.065

Truscott M, Islam AB, Frolov MV (2016) Novel regulation and functional interaction of polycistronic miRNAs RNA 22: 129–138. https://doi.org/10.1261/rna.053264.115

Unterbruner K, Matthes F, Schilling J, Nalavade R, Weber S, Winter J, Krauss S (2018) MicroRNAs miR-19, miR-340, miR-374 and miR-542 regulate MID1 protein expression PLoS One 13:e0190437. https://doi.org/10.1371/journal.pone.0190437

Wang Y, Luo J, Zhang H, Lu J (2016) microRNAs in the same clusters evolve to coordinately regulate functionally related genes Molecular biology and evolution 33: 2232–2247. https://doi.org/10.1093/molbev/msw089

Wheeler BM, Heimberg AM, Moy VN, Sperling EA, Holstein TW, Heber S, Peterson KJ (2009) The deep evolution of metazoan microRNAs Evolutionary Development 11, 50–68. https://doi.org/10.1111/j.1525-142X.2008.00302.x

Zhang J, Yang J, Zhang X, Xu J, Sun Y, Zhang P (2018) MicroRNA-10b expression in breast cancer and its clinical association PLoS One 13: e0192509. https://doi.org/10.1371/journal.pone.0192509

Zhang P, Min W, Li WH (2004) Different age distribution patterns of human, nematode, and Arabidopsis duplicate genes Gene 342: 263–268. https://doi.org/10.1016/j.gene.2004.08.001