Journal of Bioinformatics and Sequence Analysis Vol. 3(1), pp. 1-5, January 2011
Available online at http://www.academicjournals.org/JBSA
ISSN 2141-2464 ©2011 Academic Journals

Full length Research Paper

Evolutionary analysis of gorilla, chimps and humans using sequence divergence

Vibhu Ranjan Prasad, Soumya Chaurasia and Rao Sethumadhavan*

School of Bioscience and Technology, Vellore Institute of Technology, Vellore, Tamil Nadu – 632014, India.

Accepted 30 September, 2010

We have analyzed the sequence divergence amongst the three species that is, gorilla, chimpanzee and human varying from Hominidae and Pongidae. Apart from the genomic phylogeny, we compared the protein and rRNA phylogenies to increase the importance of phylogenetic analysis. The proteins selected are from the mitochondrial origin as mtDNA which codes for mitochondrial proteins, mutate at a higher rate compared to nuclear DNA, so as to give a more useful, magnified view of the diversity present in a population, and its history. The phylogenetic analysis of the mitochondrial genome, rRNA and proteins are done using parsimony, BIONJ and PHYML methods which $resulted into variable results. The proteins that were able to infer the already stated phylogeny between these members were ATP synthase subunit 6 and 8, cytochrome b, cytochrome oxidase subunit 1 and 3 and NADH dehydrogenase subunit 2, 3 and 5. Although BIONJ and PHYML methods predicted similar results most often parsimony was found to be predicting contradictory phylogeny with respect to above two methods. To verify the results as obtained from various methods and to further analyze the evolutionary relationship between these members, we applied the method of calculating sequence divergence using bioinformatics tools. Taking into account that local point mutations generate a considerable genetic variation and this variation being very high in mitochondrial DNA, we analyzed the sequence divergence between these three species using 0, ±1 and ±2 error in identical cut sites by different restriction enzymes. No identical site between chimpanzee-human clade at different errors indicated gorilla to be common ancestor of these two members. At ±2 error, an identical sequence divergence was obtained between the chimpanzee-gorilla and human-gorilla clades but overall results suggest chimpanzee to be a closer relative of gorilla than human.

Key words: Sequence divergence, in silico restriction digestion, phylogeny.

INTRODUCTION

The endless number of study related to the evolutionary analysis of the three species varying from Hominidae to Pongidae confirms the importance which these organisms hold in evolutionary analysis. One of the species among these, the *Homo sapiens* has also been studied widely in evolutionary and phylogenetic terms. Traditionally chimps are classified with other great apes, gorillas and orangutans in the family *Pongidae,* separated from the human family Hominidae, but at the level of DNA humans are so closely related to chimps that chimps should not be part of the same taxonomic family, but also the same genus. Even the humans are also being called "the third chimpanzee" (Jeff, 2003). However, the close similarity in the nucleotides of two species does not mean that proteins coded by them are also the same. A recent study has proposed that 80% of the proteins between chimps and humans are different that leads to a considerable morphological difference between the two species (Galina et al., 2005). The molecular phylogeny using the complete sequences of mitochondrial DNAs for the evolution of prosimians is already being done (Atsushi et al., 2009). Moreover, taxonomic and phylogenetic analysis between great apes and humans has identified phylogentic relationship between these

*Corresponding author. E-mail: rsethumadhavan@vit.ac.in.

species as a never ending conflict (Galina et al., 2005). We have analyzed the phylogenetic relationship between the great apes, that is chimps and gorillas with humans using sequence divergence in their mitochondrial genome.

The paper deals with the use of novel bioinformatics tool in analyzing the phylogeny and stating some new points in the evolutionary relationship of chimpanzee, gorilla and human. The approach involves the use of parsimony, BIONJ and PHYML methods to analyze the phylogenetic relations between the mitochondrial DNA, 12s and 16s rRNA and mitochondrial proteins on which the life of animals are crucially dependent. The new method that we are using here is the analysis of sequence divergence between the mitochondrial genome of these organisms. Since mitochondria is found to be evolving much more rapidly than single-copy nuclear DNA in higher animals (Brown et al., 1979), the number of cut sites obtained after digesting the DNA with a variety of restriction enzymes were highly unlikely to be similar. We used a new approach considering the local point mutations to be one of the main causes of genetic variation among these sequences and obtained a result that can be used to state evolutionary relationship as well as the degree of separation between gorilla, chimpanzee and human. The results presented here may be further used in the evolutionary studies of Hominidae and great apes.

MATERIALS AND METHODS

Dataset

The mitochondrial genome, mitochondrial rRNA and mitochondrial proteins of *Pan troglodytes* (Chimpanzee), *Gorilla gorilla* (Gorilla) and *Homo sapiens* (humans) were taken from from Entrez Genome (available at http://www.ncbi.nlm.nih.gov/entrez) using the NCBI Sequence Viewer version 2. The details of the sequences taken and their accession numbers are depicted in Tables.

Multiple sequence alignment

The multiple sequence alignment of mitochondrial genome, mitochondrial rRNA and mitochondrial proteins was done using Clustal X 1.83 (Thompson et al., 1997). The program can align the sequences in both multiple alignment mode as well as Profile alignment mode. For the present work, we have used the multiple alignment mode to align the mitochondrial genome, mitochondrial rRNA and mitochondrial proteins of chimpanzee, gorilla and human.

Phylogenetic tree using parsimony

Phylogenetic tree using parsimony was calculated using PHYLIP's dnapars/protpars algorithm. The tool used for finding the tree is a standalone tool known as SeaView 4.0 (Galtier et al., 1996).

BioNJ tree

The distance based phylogenetic analysis of the mitochondrial genome, mitochondrial rRNA and mitochondrial proteins of chimpanzee, gorilla and human was done using the BioNJ algorithm which is an improved version of NJ algorithm (Gascuel, 1997). The algorithm is well suited for distance estimation from DNA and protein sequences. BIONJ has better topological accuracy than NJ in all evolutionary conditions; its superiority becomes important when the substitution rates are high and varying among lineages.

PhyML tree

PhyML 3.0 program was used to estimate the phylogeny by maximum likelihood method (Thompson et al., 1997). The PhyML algorithm is a simple, fast and accurate algorithm to estimate the large phylogenies by maximum likelihood method. The tool used for calculating the PhyML tree is a standalone tool known as SeaView 4.0 (Galtier et al., 1996). The algorithm behind this program dramatically reduces time and can be used with much larger and more complex data sets.

In silico restriction digestion

The *in silico* restriction digestion of the mitochondrial genome, mitochondrial rRNA and mitochondrial protein was done using Webcutter 2.0 tool available at http://rna.lundberg.gu.se/cutter2. Webcutter 2.0 accesses the restriction enzymes from the Restriction Enzyme Database or REBASE available at http://rebase.neb.com (Gascuel, 1997). The mitochondrial genome was provided as a RAW input to the tool for restriction digestion analysis and the restriction digestion for chimpanzee, gorilla and human genomes are done separately. The selection of restriction enzymes was based upon the earlier works of Ferris et al. (1981). The 19 restriction enzymes used for the present work are AvaI, BglII, BstEII, EcoRI, EcoRV, FaunDI, HincII, HindIII, HpaI, KpnI, PstI, PvuII, SacI, SalI, ScaI, SmaI, Xba I and XhoII with the exception of FaunDII. Since FaunDII is not available in the Webcutter 2.0, it was replaced by FaunDI for the present work.

RESULTS AND DISCUSSION

Nucleic acid phylogeny

Phylogenetic analysis of the mitochondrial genome, 12s rRNA and 16s rRNA of chimpanzee, gorilla and human was done and the trees are drawn in Table 2. BioNJ and PhyML phylogeny supported the already stated results of chimpanzee-human clade. However, the parsimony phylogeny using mitochondrial genome supported human-gorilla clade while using the 12s and 16s rRNA, phylogeny supported the chimpanzee-gorilla clade.

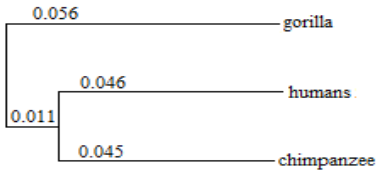
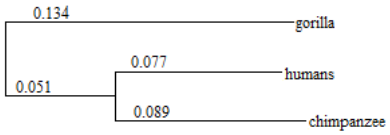
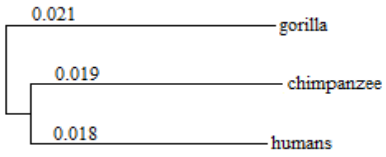
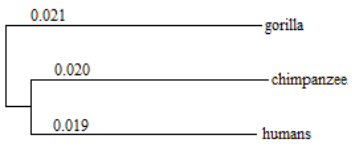
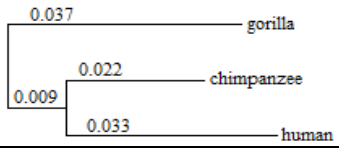
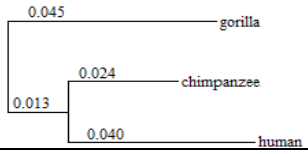
Protein phylogeny

Phylogenetic analysis of the mitochondrial proteins listed in Table 1 also gave variable results. Among the thirteen mitochondrial proteins selected for the analysis, cytochrome oxidase subunit 2 and NADH dehydrogenase subunit 1 and 4 were supporting the chimpanzee-gorilla clade and hence were considered faulty. ATP synthase

Table 1. Dataset of mitochondrial genome, rRNA and proteins.

mtDNA or proteins coded by mitochondria	Chimpanzee	Human	Gorilla
mt DNA	NC 001643	AC_000021	NC_011120
ATP synthase S6	NP_008191	AP_000644	YP_002120664
ATP synthase S8	NP_008190	AP_000643	YP_002120663
Cytochrome B	NP_008198	AP_000651	YP_002120670
Cytochrome ox S1	NP_008188	AP_000641	YP_002120661
Cytochrome ox S2	NP_008189	AP_000642	YP_002120662
Cytochrome ox S3	NP_008192	AP_000645	YP_002120665
NADH dehy sub 1	NP_008186	AP_000639	YP_002120659
NADH dehy sub 2	NP_008187	AP_000640	YP_002120660
NADH dehy sub 3	NP_008193	AP_000646	YP_002120666
NADH dehy sub 4	NP_008195	AP_000648	YP_002120667
NADH dehy sub 4L	NP_008194	AP_000647	YP_002117971
NADH dehy sub 5	NP_008196	AP_000649	YP_002120668
NADH dehy sub 6	NP_008197	AP_000650	YP_002120669

Table 2. BioNJ and PHYML trees for mitochondrial genome and 12s and 16s rRNA.

	BioNJ	PhyML
Genome		
12 s rRNA		
16s rRNA		

subunit 6 and 8, cytochrome b, cytochrome oxidase subunit 1 and 3, NADH dehydrogenase subunit 2, 3, 4 and 5 supported the chimpanzee-human clade. Mitochondiral protein phylogenetic analysis in these three organisms also indicated that the phylogenies predicted by parsimony method was contradictory to that obtained from BIONJ and PHYML methods.

In silico restriction digestion

As we know, most of the variation among genes in a gene family is caused by point mutations and positive selection (Xinli et al., 2006). *In silico* restriction digestion using Webcutter 2.0 and analysis of the number of sites shared between mitochondrial genome of chimpanzee, gorilla and human showed that there were no sites shared between the genome of chimpanzee-human and gorilla-human and the results are tabulated in Table 3. This could be due to the intense evolution that has resulted in a change in the position of restriction sites of the restriction enzymes. However, a total of 2 sites (PvuII454 and Xbal643) were found to be exactly common between chimpanzee-gorilla. This directly confers the strong similarity between these two and hence less sequence divergence during the course of evolution.

Table 3. Cut sites alongwith the restriction enzymes at increased level of error for gorilla, chimps and human mitochondrial genomic DNA.

Error	Chimpanzee	Gorilla	Human
0	PvuII 454	PvuII 454	-
	XbaI 643	XbaI 643	-
±1	EcoRI 3571	EcoRI 3570	-
	EcoRI 4724	EcoRI 4723	-
	HincII 1874	HincII 1873	-
	HincII 5143	HincII 5142	-
	HpaI 1874	HpaI 1873	-
	HpaI 5143	HpaI 5142	-
	PvuII 4416	PvuII 4415	-
	KpnI 2028	KpnI 2027	-
±2	-	AvaI 19	AvaI 17
	-	PstI 30	PstI 28
	-	SmaI 21	SmaI 19
	-	XbaI 18	XbaI 16
	-	XhoI 21	XhoI 19
	AvaI 21	AvaI 19	-
	PstI 32	PstI 30	-
	SmaI 23	SmaI 21	-
	XbaI 20	XbaI 18	-
	XhoI 23	XhoI 21	-

Table 4. Fraction of common sites between chimpanzee-gorilla, chimpanzee-human and gorilla-human clades.

Error	Chimpanzee-Gorilla	Chimpanzee-Human	Gorilla-Human
0	0.0163 (0.9837)	0	0
±1	0.068 (0.932)	0	0
±2	0.042 (0.958)	0	0.043 (0.957)

Percentage of sequence divergence is given in parenthesis.

As it is reported earlier that local point mutations continually generate considerable genetic variation that is capable of altering gene expression (Jonathon and Gregory, 2001) that we can consider as evolution, the error of ±1 between the sites digested by restriction enzymes in different organisms can be taken closer to the exact sites digested by the same restriction enzymes. Using this approach, I have tried to analyze the number of sites shared between chimpanzee, gorilla and human with ±1 error. Extending the same approach to ±2 error, number of sited shared are calculated and sequence divergence is calculated. The percentage of sequence divergence at increased levels of error is tabulated in Table 4.

Conclusion

We have analyzed the phylogeny between three members of the Hominidae family that is, human, gorilla and chimpanzee using parsimony, BIONJ and PHYML methods and calculated the sequence divergence between them using *in silico* restriction digestion. For our analysis, we have taken the genome, 12s and 16s rRNA and a set of 13 proteins coded by the mitochondrial DNA. Phylogeny based upon mitochondrial genome, 12s and 16srRNA supported the chimpanzee-human clade. Mitochondrial genome phylogeny was not found to predict the correct phylogeny by any of the three methods. However, some mitochondrial proteins were also found to be supporting the chimpanzee-gorilla clade. Moreover, not a single protein was found for which the phylogeny predicted by parsimony, BIONJ and PhyML methods were same. PhyML and BIONJ methods were found to predict correct phylogeny often. However, neither mitochondrial genome, rRNA nor any of the mitochondrial proteins was found for which the three methods predicted similar results. In an attempt to

analyze the phylogenies more clearly, we calculated the sequence divergence between the genome of these three organisms. The results suggest the human mitochondrial genome to have diverged more from gorilla in comparison to chimpanzee. Even considering the evolutionary process to be influenced more by point mutation, we analyzed the digested sites of these three organisms and achieved the similar results. At ±2 error, human mitochondrial genome was found to have diverged 95.7% from gorilla genome compared to the 95.8% divergence of chimpanzee genome. No similar cut sites were found between human and chimpanzee at 0, ±1 and ±2 errors. This indicate gorilla to be having similarity with both human and chimpanzee at increased levels of error and hence to be the common ancestor. The results from this study might be helpful in further evolutionary studies of these species.

REFERENCES

Atsushi M, Felix R, Isao M, Hasegawa M, Horai S (2009). Molecular phylogeny and evolution of prosimians based on complete sequences of mitochondrial DNAs. Gene, 441: 53–66.

Brown WM, George M, Wilson AC (1979). Rapid evolution of animal mitochondrial dna (primates/restriction endonuclease cleavage maps/gel electrophoresis/dna melting) Proc. Natl. Acad. Sci., 76(4): 1967-1971.

Ferris SD, Wilson AC, Brown WM (1981). Evolutionary tree for apes and humans based on cleavage maps of mitochondrial DNA. Proc. Nat. Acad. Sci., 78: 2432-2436.

Galina G, Veeramachaneni V, Nei M, Makayowski W (2005), Eighty percent of proteins are different between humans and chimpanzees. Gene, 346: 215–219.

Galtier N, Gouy M, Gautier C (1996). SeaView and Phylo_win, two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci., 12: 543-548.

Gascuel O (1997). Molecular Biology and Evolution. 14: 685-695.

Guindon S and Gascuel O (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol., 52(5): 696-704.

Jeff Hecht (2003). Chimps are human, gene study implies. New Scientist.

Jonathon RS, Gregory AW (2001). Rapid Evolution of cis-Regulatory Sequences via Local Point Mutations. Mol. Biol. Evol., 18: 1764-1770.

Lockwood CA, Kimbel WH, Lynch JM (2004), Morphometrics and hominoid phylogeny: Support for a chimpanzee–human clade and differentiation among great ape subspecies. PNAS, 101(13): 4356-4360.

Maarek Y and I Shaul (1997). Webcutter: A System for dynamic and tailorable site mapping. Comput. Netw., ISDN Syst., 29: 1269-1279.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res., 25: 4876-4882.

Xinli S, Yinglong C, Shiping W, (2006). Point Mutations with Positive Selection Were a Major Force during the Evolution of a Receptor-Kinase Resistance Gene Family of Rice. National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan 430070, China. 140(3): 998-1008.