



Characterization of Simple Sequence Repeats: Evolutionary Implications from Ancient Human Mitochondrial Genome

Poulami Ghosh^(✉) and Anasua Sarkar

Department of Computer Science and Engineering, Jadavpur University, Kolkata, India
poulamighosh738@gmail.com

Abstract. Evolution is an extraordinarily long and complex process that all lives on earth undergo. Therefore, studying genetic origins and mutation scans reveal much information about the evolution of life. In this work, we consider the Simple Sequence Repeats (SSRs) of four subspecies of Hominini forms of the taxonomic tribes and derive biological features of the extinct species and search for repeat markers concerning modern humans. We also draw a phylogenetic relationship among the four sub-species from their mitochondrial genomes that had originated during the Pleistocene period and determine the alteration of microsatellite repeats resulting from changes in the formation of protein and, in the long run, leading to evolution.

Keywords: Evolution · Mutation · SSRs · Hominini · Taxonomic tribe · Phylogenetic relationship · Pleistocene period · Microsatellite

1 Introduction

If we travel back in time, Archaic *Homo sapiens* originated and spread through the Afro-Asian subcontinent during the Pleistocene period. The different sub-species of Archaic humans are the Neanderthals, Denisovans and the Heidelbergensis. The modern *Homo sapiens* emerged close to 20,000 to 45,000 years ago, assumingly in Africa. We have also implied that the *Homo neanderthalensis* or Neanderthals emerged in Africa, and *Homo sapiens sapiens* emerged in Europe and West Asia around the same period.

It has been discovered that Neanderthal-derived DNA can be found in the DNA of possibly all contemporary populations, which vary regionally. The Denisovan ancestry is also called *Homo sapiens Altai*, which emerged during the lower and middle Paleolithic periods. The first identification of Denisovan individuals occurred in 2010 based on the mitochondrial DNA extracted from a juvenile female finger bone from the Siberian Desinovan cave in the Altai Mountains. The DNA indicates a close affinity with the Neanderthal DNA sequence [7].

Another sub-species of Archaic *Homo sapiens* is the *Homo heidelbergensis*. It is considered a dynamic species that had evolved from an African form of *Homo erectus*. It is the most recent common ancestor between modern humans and Neanderthals [8].

This study will consider the mitochondrial DNA genomes of the sub-species mentioned above of the tribe Homo and will discuss their distributions as well as in Simple Sequence Repeats characterization, which may imply their evolutionary tree and for a human to escape from extinction.

1.1 The Mitochondrial DNA

Mitochondria is an organelle found in all organisms with nuclei releasing energy from food and storing it as ATP (Adenosine triphosphate) molecules. Mitochondria have their DNA. It was once a free-living bacteria that ancestors of eukaryotes engulfed but could not digest. Bacteria enlarged cells established a symbiotic relationship in which the larger cell protected while bacteria produced food. As a result, the mitochondrial chromosomes store many features like bacteria. They are circular molecules. The human mitochondrial chromosome is around 16,546 base pairs long which is approximate to the size of any bacterial plasmid. Unlike nuclear chromosomes, the mitochondrial chromosome is packed tightly with genes, with a significant region for non-coding DNA between genes. Most mitochondrial genes lack introns which are non-coding information of nuclear genes. Human mtDNA contains 37 genes. 13 are involved in oxidative phosphorylation processes. 22 genes code for tRNA for specific amino acids, whereas 2 regulate the subunit of ribosomes.

The human mitochondrial chromosome has been vastly reduced. Over time, the functionality of genes that the host could provide was lost. Also, some genes needed for respiration were transferred to the nucleus. As a result, millions of years of evolution have resulted in small mitochondrial chromosomes. When an egg and sperm unite, the sperm mitochondria are discarded during fertilisation. MtDNA is, therefore, only inherited from the mother within a family. Unlike the DNA of the nucleus, the mtDNA does not recombine with other chromosomes. Consequently, we can use them to understand the ancestry and evolution of a species [1, 2].

2 Materials and Method

2.1 Genome Sequences

We have considered the whole mitochondrial genome sequences of the species: Neanderthals, *Homo sapiens*, Denisovans and Heidelbergensis. We have downloaded the FASTA files of their reference sequences. We have used sequences with NC 011137.1, NC 012920.1, NC 013993.1, and NC023100.1, respectively, from www.ncbi.nlm.nih.gov.

The lengths of those mitochondrial genomes are 16.9 kb, 16.9 kb, 16.9 kb, and 16.9 kb, respectively.

2.2 Microsatellite Identification and Investigation

Simple microsatellites are extracted with the help of a sliding window algorithm. We have computed the program using Python 3.0 programming language with the assistance

of Regular expression package and Biopython modules in Conda environment in Jupyter notebook (version: 6.0.3). For finding the GC content, we have used Bio.SeqUtils package from Biopython module. We have also used the standard Sequence Input/Output interface (Bio.SeqIO) for accessing the FASTA files. We have considered repeat type to be Perfect and motif sizes to vary from 6 to 1, i.e., Hexamer, Pentamer, Quadmer, Trimer, Dimer and monomers. The Considered repeat number is 3,3,3,3,3,6 in the same order, and other parameters are set as default. We have also determined the phylogenetic analysis of the given hominini sub-species with the help of the Molecular Evolutionary Genetics Analysis (MEGA) tool. Version 11.0.10. In the analytical process, we have not found any hexamer and pentamer repeats therefore have considered the repeats that were found.

2.3 Statistical Analysis

We have done the statistical analysis and graphical representations with the help of NumPy and matplotlib libraries in Python 3.

3 Observations and Results

In Table 1, we see that the number of SSRs in Neanderthals mtDNA is 31, which counts to be the highest among the four subspecies. Therefore they have a higher mutation rate than the other species.

Table 1. Overview of the computational results of the whole genome of the four Hominini sub-species.

Sub species	Homo sapiens mitochondrion	Homo sapiens neanderthalensis mitochondrion	Homo sapiens Desinova mitochondrion	Homo Sapiens Heidelbergensis mitochondrion
Length of genome	16568	16565	16570	16568
No. of SSR	26	31	25	28
No. of Tri-mer	9	16	9	9
No. of Di-mer	13	11	12	14
No. of Mono-mer	4	4	4	4
GC Content	44.35	44.39	44.3	43.42

Relative Abundance (RA) in Fig. 1 estimates the biological varieties of a species. RA can be defined as the number of microsatellite repeats divided by genome size in kb. It measures how common or rare a species is comparable to other species in a particular location. According to Fig. 1, the Neanderthals had the highest Relative Abundance of the other species in one specific area. We conclude from this context that Neanderthals were the most diverse, and Denisovans were the least varied in a given location among the given species. Relative Density (RD), shown in Fig. 2, measures the total length contributed by each microsatellite upon the length of the genome in kb. It tells us about the number of a given species expressed as a percentage of all species present.

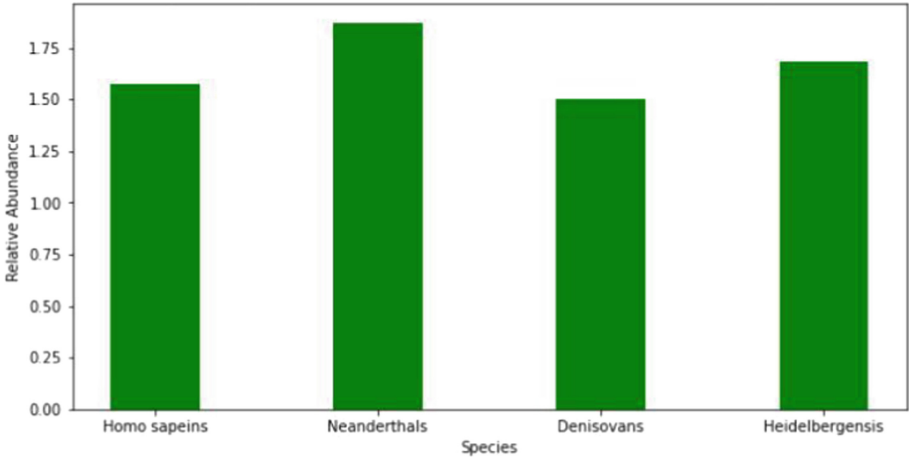


Fig. 1. Comparison of Relative Abundance (RA) of the Hominini sub-species.

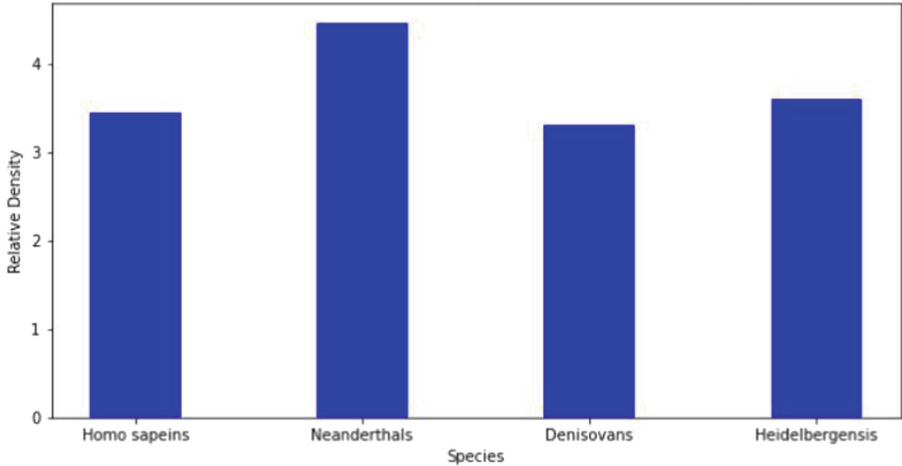


Fig. 2. Comparison of Relative Density (RD) of the Hominini sub-species.

Therefore, according to our observation in Fig. 2, the relative density of Neanderthals is higher than all the given species, and Denisovans have the lowest relative density among the four species. From this angle, we can conclude that Relative Abundance and Relative Density are directly proportional [3, 10].

Moreover, the percentage of GC content showcases the lifespan of a particular species.

Therefore, according to Table 1, we can conclude that Neanderthals had the longest life span and Heidelbergensis the shortest [4, 6]. Figure 3 displays the various simple sequence repeats that show a significant rate of length polymorphism due to mutations of one or more repeat types. The comparison of the common simple sequence repeats in

the four species shows that there has been missing and gaining of microsatellites among the four species, which has led to the emergence of one and the extinction of the other. For example, $(ACC)_3$ and $(GGA)_3$ are gained by all species other than *homo sapiens*. According to Fig. 4, we can see that the *Homo sapiens* and the Neanderthals emerged during the same period, and *Homo Heidelbergensis* is the most recent common ancestor between the two. In contrast, the *Homo sapiens Altai* (Denisovans) emerged a long time after their emergence. While performing the phylogenetic analysis, we had to perform sequence alignment of the mentioned sequences. In the process, we observed mutations in the amino acid sequence that had specifically occurred in the mtDNA sequence of the discussed species, showing some beautiful examples of evolution [4].

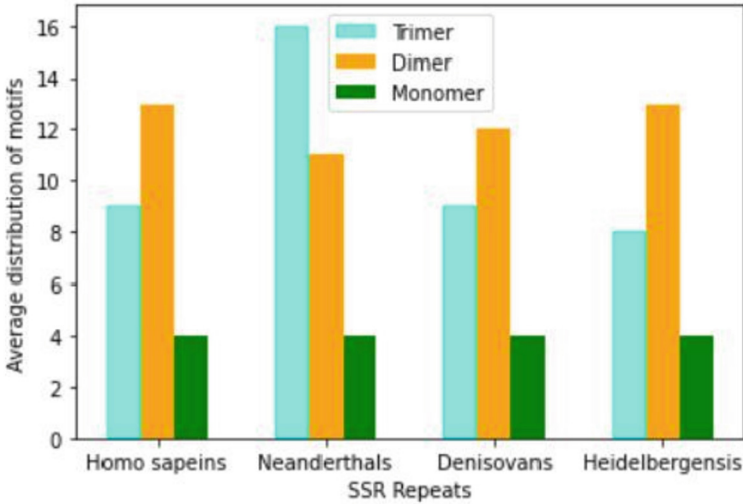


Fig. 3. Observation of SSR distribution in the given Hominini sub-species.

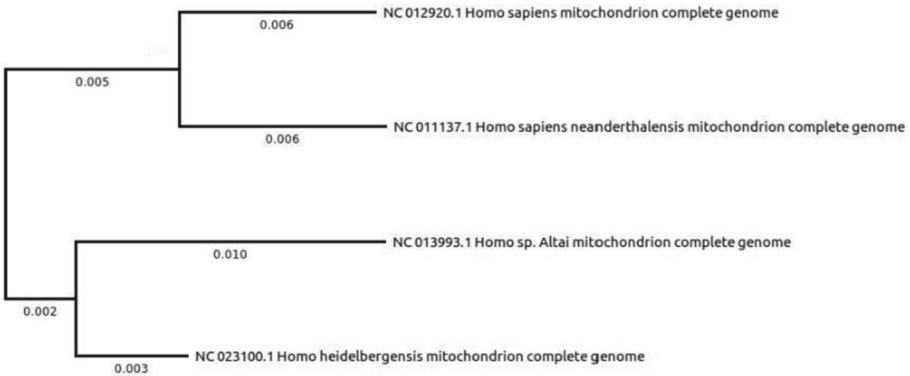


Fig. 4. Observation of the phylogenetic analysis of the Hominini sub-species.

Table 2 displays some of those positions and changes in the amino acid. In Fig. 5, we observe the trimer repeat $(ATC)_3$ is gained by *Homo sapiens* and Denisovans, whereas Neanderthals and Heidelbergensis missed it. We also see that the trimer repeat $(ATA)_3$ is found in *Homo sapiens*, Neanderthals and Heidelbergensis, but it is missing in the mtDNA of Denisovans. The repeats $(CAG)_3$ and $(GCA)_3$ are found in all the subspecies except that of Denisovans. $(CCT)_3$ is found in *Homo sapiens* and Denisovans, whereas it is missing in Neanderthal and Heidelbergensis. $(ACT)_3$ is found in all the subspecies but Heidelbergensis. Trimer motifs $(GGA)_3$ and $(ACC)_3$ are gained by Neanderthals and Heidelbergensis, whereas missed by the *Homo sapiens* and Denisovans. $(TTA)_3$ is another repeat found in all the subspecies other than the *Homo sapiens. $(CAA)_3$, $(GCC)_3$ in Neanderthals and $(CCG)_3$, $(GAG)_3$, $(AAC)_3$ in Denisovans are the repeats exclusively found in their respective mtDNA sequences. Figure 6 represents the changes that have occurred in the SSRs in the subspecies. We observe that most dimers are present in all the mtDNA sequences except $(GA)_2$, which is missing in Neanderthals and Denisovans. Another repeat $(AG)_2$ is present in all the subspecies except the Neanderthals. We have compared our observation with that of the MicroSATellite Identification tool (MISA), which has led us to the conclusion that the repeats $(ATA)_3$, $(ACT)_3$ and $(CCT)_3$ are the SSR markers for *Homo sapiens*. The presence of $(CCG)_3$, $(CCT)_3$ and $(TTC)_3$ confirms that the mtDNA is of the Denisovan subspecies. The presence of $(CAA)_3$ and $(ACC)_3$ in the mtDNA sequence assures that the mtDNA is of the Neanderthal subspecies. Lastly, the presence of $(GGA)_3$, $(ACC)_3$, and $(AG)_2$ confirms that the mtDNA is of Heidelbergensis.*

Table 2. Observed changes in protein sequence after sequence alignment.

Homo sapiens mitochondrion	A	G	T	T	C	T	T	C	G	T	A	C	A	C
Homo sapiens Neanderthalensis mitochondrion	A	A	T	T	T	T	T	A	C	A	T	A	T	
Homo sapiens Denisova mitochondrion	G	A	T	T	T	T	C	T	A	T	G	C	A	T
Homo Sapiens Heidelbergensis mitochondrion	A	A	C	C	T	C	T	T	A	T	A	C	G	T
Position:	929	1019	1041	1244	4905	4908	4938	5388	5472	5971	6047	6207	6360	6459

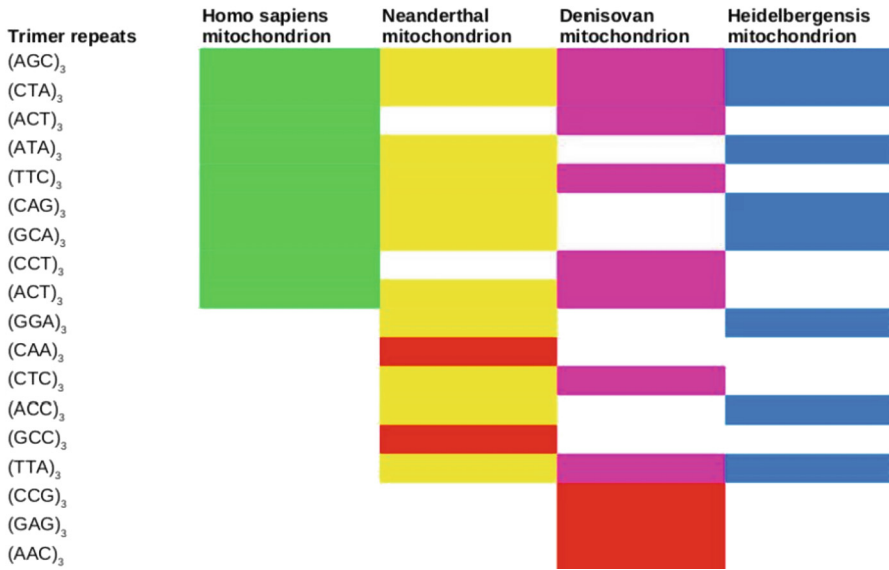


Fig. 5. Representation of the Trimer distribution among the four Hominini sub-species.

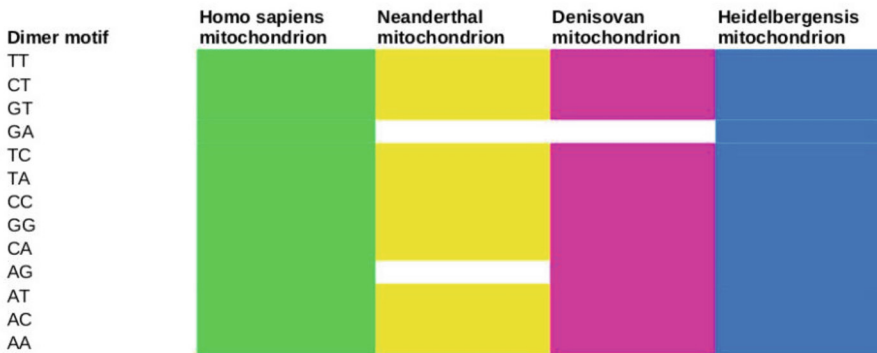


Fig. 6. Representation of the Dimer distribution among the four Hominini sub-species.

4 Conclusion

We can identify the different aspects of evolution in the given subspecies of the Hominini tribe by performing the SSR extraction. From Table 1, we observe that the SSR count in the mtDNA of the Neanderthals is a maximum of 31, and that of the Desinovans is a minimum. The observation concludes that the Neanderthals have maximum relative density and relative abundance, giving us an idea about the vast diversity of the species across a particular location. The high GC content of the Neanderthals also reveals that the species had the longest life span than the rest. Moreover, by extracting the microsatellite

sequences, we have tried to understand the alteration of functionalities in the mtDNA sequences. Working more on these data using advanced computational tools will reveal more information about the prehistoric ancestors of humanity to reframe the history of evolution better.

References

1. Khan, S.M., Smigrodzki, R.M., Swerdlow, R.H.: Cell and animal models of mtDNA biology: progress and prospects. *Am. J. Physiol. Cell Physiol.* **292**(2), C658–C669 (2007). <https://doi.org/10.1152/ajpcell.00224.2006>
2. Nesheva, D.: Aspects of ancient mitochondrial DNA analysis in different populations for understanding human evolution. *Balkan J. Med. Genet.* **20**(1), 5–14 (2014)
3. Chaudhary, A., Singh, A., Choudhary, S., Safdar, A.: In-silico analysis of simple and imperfect microsatellites in diverse tobamovirus genomes. *Gene* **530**(2), 193–200 (2013)
4. Johnathan, A., Robert, P., Corey, C., Aaron, C., David, D.: Finding and extending ancient simple sequence repeat-derived regions in the human genome. *Mob. DNA* **530**(2), 193–200 (2013)
5. Lehmann, L., Budovsky, A., Khachik, K., Vadimr, E.: Mitochondrial genome anatomy and species-specific lifespan. *Rejuven. Res.* **9**(2), 223–226 (2006)
6. Lehmann, G., Segal, E., Muradian, K., Fraifeld, V.: Do mitochondrial DNA and metabolic rate complement each other in determination of the mammalian maximum longevity? *Rejuven. Res.* **11**(2), 409–417 (2008)
7. Gomez, A., Robles, A.: Dental evolutionary rates and its implications for the Neanderthal-modern human divergence. *Sci. Adv.* **5**(5) (2019)
8. Richter, D.: The age of Hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age. *Nature* **546**, 293–296 (2017)
9. Barshad, G., Marom, S., Cohen, T., Mishmar, D.: Mitochondrial DNA transcription and its regulation: an evolutionary perspective. *Trends Genet.* **94**(9), 682–692 (2018)
10. Falkenberg, M., Larsson, N.: DNA replication and transcription in mammalian mitochondria. *Annu. Rev.* **76**, 679–699 (2007)