Asian Journal of Biological and Life Sciences | Original Research

# Genome wide analysis of Single Nucleotide Polymorphism of Human mitochondrial Genome in Indian Population

**Vijaylaxmi Saxena[1], Supriya Dixit[2]*, Krati Shukla[3]**

1 Co-ordinator: Bioinformatics Infrastructure Facility, Centre of DBT (Gov. of India), D.G. (P.G.) College, Kanpur- 208001 (U.P), India.
2,3 Bioinformatics Infrastructure Facility, Centre of DBT (Govt. of India), D.G. (P.G.) College, Kanpur- 208001 (U.P), India.

E-mail : supriya.dixit.28@gmail.com Contact No. : 0512-2530657

Submitted : 20.09.2015 Accepted : 28.10.2015 Published : 31.12.2015

## Abstract

The human mitochondrial genome is extremely small compared with the nuclear genome and also have high rate of mutations which will lead to a disease like cancer, ageing and neurodegenerative disease. A major challenge for linkage disequilibrium in mitochondrial genome studies is the high cost of genotyping large number of single nucleotide polymorphisms and to predict correlations between them to find out informative SNP which is able to cause variations among population. The main aim of this research to find out an efficient method for finding SNPs, and linkage within, thus analyzing polymorphism in human mitochondrial genome in Indian population. A total of 1284 SNPs were identified in the dataset of mitochondrial genome of Indian population and LD analysis was done to know the potential SNP which will be the cause of disease in the particular population but no significance results were found. This set of SNP markers found in dataset represents a useful tool for genetic studies in mitochondrial genome and will be useful in identifying candidate genes that affect diverse ranges of phenotypes.

Key words : Mitochondrial genome, linkage disequilibrium, single nucleotide polymorphisms, Mutations.

## INTRODUCTION

The term mutation was introduced by Hugo devries in 1901 for marked inheritable variation that is not inherited, these changes may occur in amount or structure of genetic material of a cell of an organism or nucleotide sequence which may result from errors during DNA replication, or covalent changes in structure because of reaction with chemical or physical agents in the environment or transpositions[1]. The mutation may be chromosomal (in chromosome) or genetic (genes) and when these natural variations in a gene, DNA sequences or chromosomes have no adverse effects on individuals it is called polymorphism[2]. It may be the presence of two or more allelic forms in a species.

The polymorphismat phenotypic or genotypic level may be due to mutation or insertion or deletion of section of DNA, its simplest form is single base mutation, it is just a single base change with a usual alternative of two possible nucleotides[3]. Single nucleotide polymorphism (SNP) represent the most frequent type of genetic polymorphism and thus provides a high density marker near the locus of interest[4].When the variation accruing in population have their allele frequency greater than 1% is termed as single nucleotide polymorphism and when less than 1% it is mutation[2]. The genomic distribution of SNPs is not homogenous and are found throughout the genome e.g. in introns, exons, intergenic regions in promoter or enhancers. The SNP may be Synonymous that is which on substitution causes no change in amino acid or protein it produces, also called silent mutation, whereas when the substitution affects the amino acid as well as the product it is called Non-synonymous mutation.

SNP's can act as genetic markers, and can be utilized in gene identification, genetic mapping, in finding predisposition to diseases, in drug discovery, in drug response, as well as for investigation purpose. SNP's are most frequent form of DNA variation and show slow mutation rate also easy to score[5].

SNP's are also required for tagging mitochondrial genetic variation, and thus the extent of the correction for multiple testing in mitochondrial association studies, depends on the amount of linkage disequilibrium in the mitochondrial genome[6]. The linkage analysis is done to identify regions in which gene expression variation segregates in pedigrees, and then linkage disequilibrium (LD) mapping is done in regions of larger sample of unrelated individuals with additional markers[7].

Linkage disequilibrium is the amount of statistical association of alleles (nucleotides) at different loci (sites). A number of evolutionary factors selections, genetic drift, gene flow, Mutation can generate linkage disequilibrium. However for tightly linked, neutral loci, a balance between the generation of linkage disequilibrium and its reduction by recombination is maintained[8].Most cancer and cardiovascular, respiratory and infectious disease are influenced by variation at multiple loci, to test the involvement of SNP in a specific condition, allele frequency are compared in affected and unaffected individual , in this way it would be possible to identify those variants that influence both the common disease and variable drug reactions.

SNP's can be found in nuclear genome as well as mitochondrial genome. The genome is the genetic material of an organism. It is encoded either in DNA or for many types of viruses in RNA. The genome includes both the genes and the non-coding sequences of the DNA/RNA. The Nuclear genome is linearly arranged diploid genome having biparental inheritance. The mitochondrial genome is circular in structure with haploid genome having maternal inheritance; the mitochondrial DNA (mtDNA) is highly polymorphic, providing an aspect of the human genome that has been thus far widely excluded from studies looking for genetic factors involved in human variation[5].

The main aim of establishment of this study is as follows:

1. Find out list of SNP's in Indian populations of mitochondrial genome.
2. Pairwise LD analysis of founded SNP's.

## MATERIALS & METHODS

### Data Source:

MtDB: Human Mitochondrial Genome Database is used in this project to find out the SNP sequences. The MtDB is a resource for population genetics and medical sciences; this has provided a comprehensive database of complete human mitochondrial genome since 2000[9].

### Tool:

Clustal Omegais used asmultiple sequence alignments (MSAs) are essential in most bioinformatics analyses that involve comparing homologous sequences[10].

SNP Finderis a web-based program developed to facilitate the SNPs discovery process. SNPs discovery is important for detection, forensic analysis and for understanding the genetic variations among closely related microorganisms[11].

SNP Analyzer 2 is a web based analysis tool embracing data quality check, linkage disequilibrium analysis and genetic association analysis in an integrated user Interface[12].

## METHODOLOGY

### 1. Sequence and SNP Analysis:

Multiple sequence alignment was done with the help of Clustal Omega to find out star sequence among 100 sequences. Then these results were analyzed with help of SNP Finder tool to find useful markers for analyzing the mitochondrial genomic structure of Indian population.

### 2. LD calculation:

LD analysis between SNP's was carried out with the help of SNP ANALYSER tool. Previous studies of LD in human's shows greater variability and pattern of LD are due to mutation, recombination, selection, population demography and genetic drift (Figure 1).

## RESULTS

For the analysis of mitochondrial genome, 100 Mitochondrial DNA sequences were taken as a dataset from Indian population namely: INDIA AF382013 (1 SEQUENCE), INDIA AF 346966 (1 SEQUENCE), PALANICHAMY INDIANS (AY71 3976 to AY71 4050) (75 SEQUENCES), RAJKUMAR INDIANS (DQ 246811 to DQ 246833) (23 SEQUENCES).

A high quality anchor sequence is used to find out SNPs onto annotated DNA regions to felicitate further evaluation of predicted SNPs and for that multiple alignments with the help of Clustal omega was done to know star sequence i.e. anchor sequence among 100 DNA sequences.

A total of 100 mitochondrial DNA sequences of Indian population were analyzed and total of 1284 SNP's were detected in our dataset and cut-off used for this analysis is 95% (Supplementary data).

The entire pattern of linkage disequilibrium and tagging SNP's are displayed in a reverse triangle shown in figure 2 & 3.

## DISCUSSION

The project was aimed to find out various SNP's from mitochondrial genome because they are highly polymorphic and cataloguing that will help to identify mutations causing mitochondrial dysfunction[9] and after that LD analysis was carried out in indian population of human mitochondrial genome is to facilitate identification and characterization of genetic variants related to common complex diseases[13].

SNP's Finder program allows to finding SNP from anchor sequence. The anchor sequence found was from Locus No.AY714041 with 16575 bp, Circular typed and the sequence

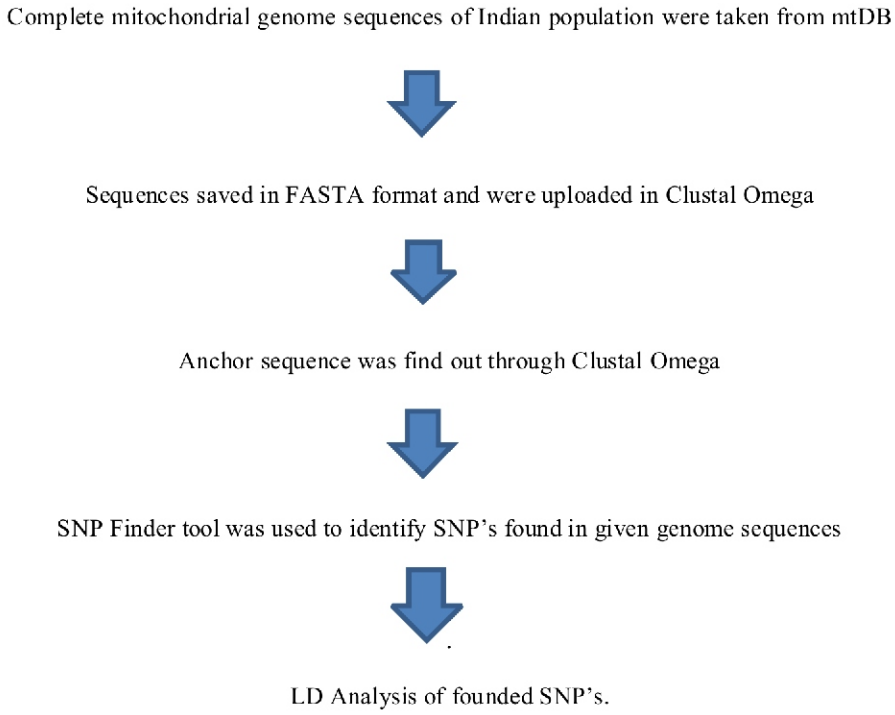


**Figure 1:** Flowchart of Methodology.

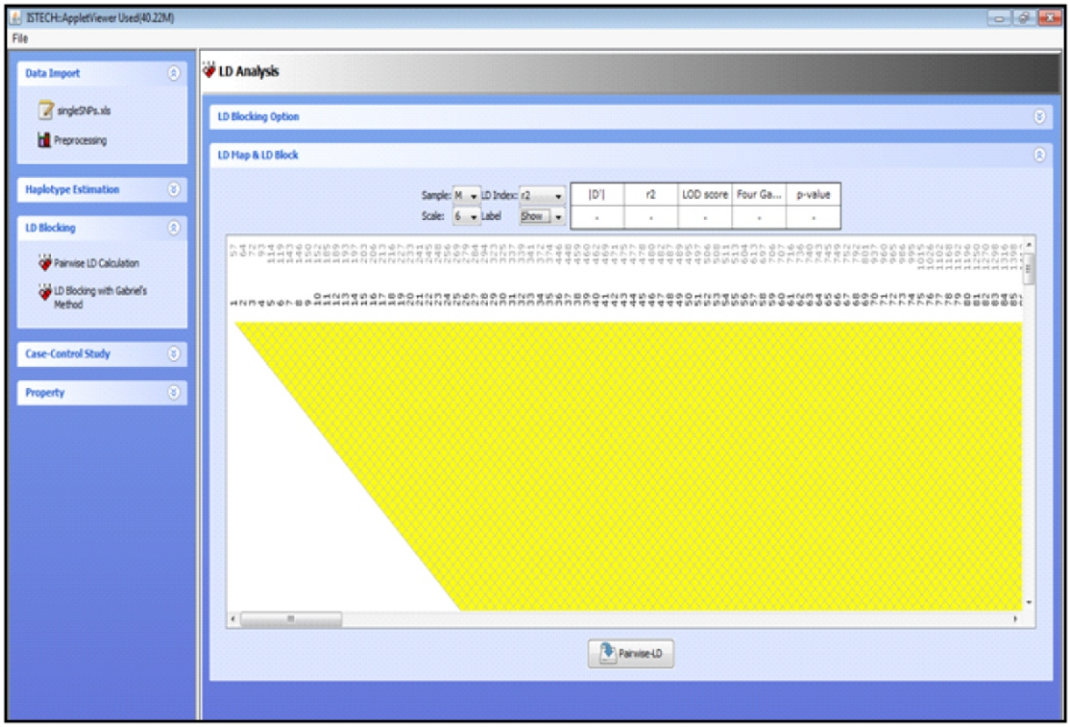


**Figure 2:** Result of SNPAnalyzer for LD blocking by pairwise LD calculation.

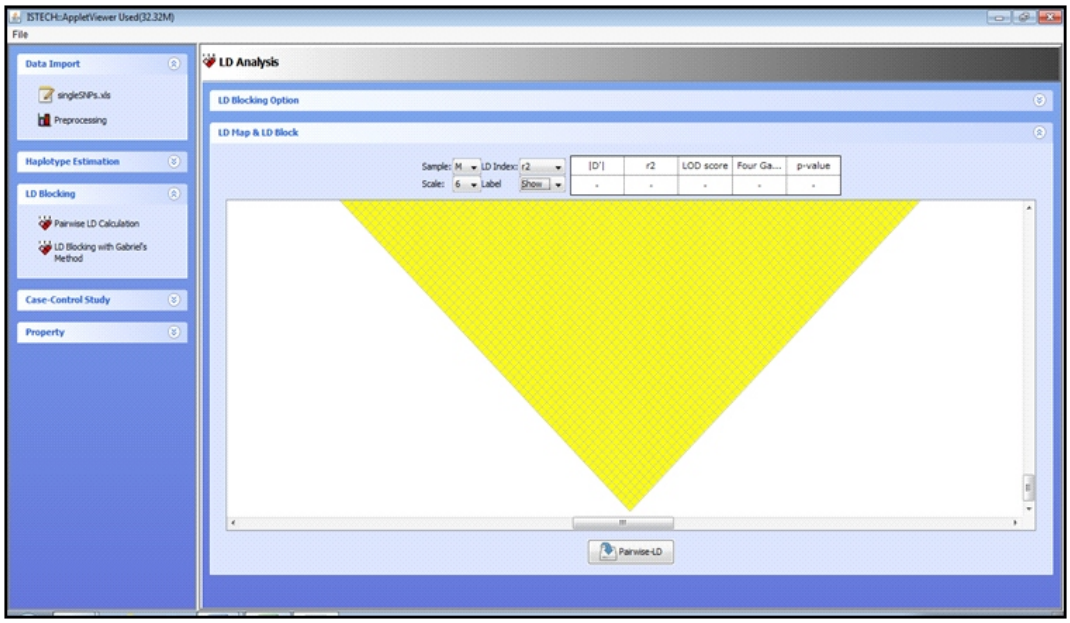


**Figure 3:** The point of triangle in result shows the least appropriate sequence by SNP finder

with least quality was from Locus No.AY713987 with 16569 bp, circular type. Both the sequences were from PALANICHAMY INDIANS population.

Single nucleotide polymorphisms (SNPs) are the simplest and most frequent type of DNA sequence variation among individuals and they represent one of the most powerful tools for the analysis of genomes. Owing to their widespread distribution, SNPs are particularly valuable as genetic markers in the search for disease susceptibility genes, drug response-determining genes[14]. A total of 1284 SNP's were found in mitochondrial DNA which can be used as a marker and also these changes can leads towards a disease.

The use of functional SNPs could be an important factor for increasing significantly the sensitivity of association tests. In fact, several complex genetic disorders have been associated with functional SNPs. With the idea of extracting as much information as possible form SNPs with phenotypic effect, SNP Finder has been developed. This tool retrieves all the SNPs present in a set of genes of interest that potentially affect the functionality of the gene product[15].

Typically, SNPs have been used as markers; that is, the real

determinant of the disease was not the SNP itself but some other mutation in linkage disequilibrium with it. So, further analysis was carried out by calculating pairwise LD of SNP's. The primary motivation of LD analysis is to facilitate identification characterization of genetic variants related to common complex diseases[13].

To investigate the pattern of LD found SNP's, SNP Analyzer tool was used and there are certain parameters like D', r2, LOD score, or by four gamete testwere used to calculate LD.

### Pairwise LD Calculation

Figure 2 and 3 represents the first line shows samples number by SNP finder, and second line shows the sequence number by SNPAnalyzer. The figure drawn by tool represents the Mt genome sequences without LD or potential SNP found which cause of variations in Indian population since there are no LD found thus the diagram is represented only yellow colour.

## CONCLUSION

Single nucleotide polymorphisms are single base variations between genomes within a species. SNP's are useful markers for diseases in haplotype-based association studies and in linkage disequilibrium analysis[24]. As there are many researches going on this field and it will be very helpful to explain the differential drug-response behaviour of individuals and also in population. Like the main of this study is to find out mutations found in mitochondrial genome of Indian population and to construct LD map.

Our results shows the list of mutations occurring in mt genome of Indian population since no LD is found in those SNP's but this study suggest *in-silico* method of finding SNP and also constructing LD map also it will provide distinctive Indian mitochondrial genomic features and further association studies can be done.

## ACKNOWLEDGMENTS

I like to put my sincere acknowledgements to DBT for providing us such platform and financial assistance. And my sincere thanks to BIF Centre at D.G.P.G. College, Kanpur and all staff members there.

## REFERENCE

1. Friedberg E C, Walker G C, and Siede W. DNA repair and mutagenesis. ASM Press. 1995, Washington, D.C.

2. Barnes M R and Gray C I. Human Genetic Variation: Databases and Concepts. *Bioinformatics for Geneticists. 2003: 40-70.*

3. Kato T. The other, forgotten genome: mitochondrial DNA and mental disorders. Mol Psychiatry. 2001: 6: 625-33.

4. Komar A A. Brief Introduction of Single Nucleotide Polymorphism: Basic Concept and Research Methods. Humana Press. (2009).

5. Kuntal H & Sharma V. SNP analysis in core mitochondrial genes of Chlorophytes 5, International Journal of Computational Bioinformatics and In Silico Modeling. 2013: 234-238.

6. McRae F A, Byrne M E, Zhao Z Z, Montgomery W G, Visscher M P. Power and SNP tagging in whole mitochondrial genome association studies. Genome Research, 18 2008: 911917. ()

7. Reich E D, Cargill M, Bolk S, Ireland J, Sabeti C P et al. Linkage disequilibrium in the human genome. Nature. 2001: 411: 199-204.

8. Hedrick P and Kumar S. Mutation and linkage disequilibrium in human mtDNA. European Journal of Human Genetics, 9 (2001), 969-972. (7)

9. Ingman M &Gyllensten U. mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. 2006: 34: D749D751.

10. Sievers F,Wilm A, Dineen D, Gibson T J, Karplus K et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular System Biology. 2011: 7: 539.

11. Conde L, Vaquerizas J M, Santoyo J, Al-Shahrour F, Ruiz-Llorente S et al. PupaSNP Finder: a web tool for finding SNPs with putative effect at transcriptional level. Nucleic Acids Res. 2004: 32: W242-8.

12. Yoo J, Lee Y, Kim Y, Rha S Y, Kim Y. SNPAnalyzer 2.0: a web-based integrated workbench for linkage disequilibrium analysis and association analysis. BMC Bioinformatics.2005: 21:20832084.

13. Kim K J, Lee H J, Park M H, Cha S H, Kim K S et al. SNP identification, linkage disequilibrium, and haplotype analysis for a 200-kb genomic region in a Korean population. Genomics. 2006: 88:535-40.

14. Schork N J, Fallin D, Lanchbury S. Single nucleotide polymorphisms and the future of genetic epidemiology. Clin Genet. 2000: 58:250264.

15. Balasubramanian S, Harrison P, Hegyi H,Bertone P, *Luscombe N et al.* SNPs on human chromosomes 21 and 22 analysis in terms of protein features and pseudogenes. Pharmacogenomics. 2002: 3:393-402.

16. Bai K R, Leal M S, Covarrubias D, et al. Mitochondrial Genetic Background Modifies Breast Cancer Risk. *Cancer Res.* 2007: 67: 4687-4694.

17. Grosjean H & Rother K. Web Links to Databases about RNA and DNA Modifications and Related Topics. 2009: 640-644.

18. Jehan T, Lakhanpaul S. SNP method and application in plants genetics, Indian Journal of biotechnology. 2006: 5: 1420-1421.

19. Johnson D A. Single-Nucleotide Polymorphism Bioinformatics: A Comprehensive Review of Resources. *CircCardiovasc Genet.*2009: 2: 530-536.

20. Karchin R. Next generation tools for the annotation of human SNPs 10. Oxford University Press. 2008: 35-52.

21. Parsons J T, Mitochondrial DNA Genome Sequencing and SNP Assay Development for Increased Power of Discrimination, NIJ Grant (2006).

22. Punt L C, *Smalley E.* Identification of Single Nucleotide Polymorphisms in the Coding Region of Human Mitochondrial DNA. JUR. (2008).

23. Sequeira A, Martin V M, Rollins B, Moon A E, Bunney E W et al. Mitochondrial mutations and polymorphisms in psychiatric disorders. Front. Genet. 2012: 1-18.

24. Smith K. Genetic Polymorphism and SNPs, Genotyping, Haplotype Assembly Problem, Haplotype Map, Functional Genomics and Proteomics. 2002. available at http://www.cs.mcgill.ca/~kaleigh/compbio/snp/snp_summary.html.

25. Taylor W R & Turnbull M D .Mitochondrial DNA Mutations in Human Disease. Nat Rev Genet. 2005: 6: 389-402.

26. Vignala A, Milana D, Cristobala S M, Eggen A. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol. 2002: 34: 275-305.

27. Voisey J & Morris P C. SNP Technologies for Drug Discovery: A Current Review, Current Drug Discovery Technologies. 2008: 5: 230-235.