Original Article

Analysis of single-nucleotide polymorphisms of PEO1 gene in 55 ethnic groups of India

Abstract

Background: Progressive External Opthalmoplegia (PEO1) or Chromosome 10 open reading frame 2 gene (OMIM ID 606075) encodes Twinkle protein, a phage T7 gene 4-like hexameric helicase, and is associated with mitochondrial DNA (mtDNA) deletions and neuromuscular disease called autosomal dominant PEO (adPEO). Twinkle has also been known to play an important role in the stability and maintenance of the mtDNA. **Aims:** In this study as an effort of Indian Genome Variation Consortium, we screened the SNPs of PEO1 gene such as rs7184, rs1535349, rs2863095, rs3740484, rs3740488, rs3740489, rs4919511, rs17113613, rs3824783, rs3740485, rs3740486, and rs3740487 in discovery panel (a population set of 40 DNA samples), and four synonymous SNPs, namely rs3824783 (ancestral allele=A), rs3740485 (ancestral allele=T), rs3740486 (ancestral allele=C), and rs3740487 (ancestral allele=A) in a large validation panel composed of 55 Indian subpopulations. **Materials and Methods:** In present study, a total of 55 Indian subpopulations were identified and collected for validation panel to check the frequencies of SNPs in PEO1 gene. **Results and Conclusion:** The allelic and genotype frequencies are found to be variable among different ethnic groups of India.

Key words:
Indian populations, progressive external opthalmoplegia, single-nucleotide polymorphisms

Introduction

Progressive External Ophthalmoplegia (PEO) is said to be associated with mutations in nuclear gene, PEO1, or Chromosome 10 open reading frame 2 (C10orf2). The gene is located on chromosome 10 and encodes mitochondrial DNA maintenance protein Twinkle which co-localize with mitochondrial nucleoid (meaning nucleus like).[1] A number of mutations have so far been identified in the PEO patients which are located in PEO1 and are involved in subunit interactions of the hexameric helicase.[1] Furthermore, mutations of the 22 mitochondrial tRNAs in human (10% of mtDNA) and large mtDNA deletions have been implicated in the disease.[2,3] Besides PEO1, PEO disease is also associated with POLG and ANT1 genes.[4,5] PEO disease, like any other mitochondrial disease, is rare in occurrence, maternally inherited, and life-threatening or chronically debilitating, resulting in considerable morbidity due to absence of any therapy. In India its occurrence is also rare and studies dealing with this disease is almost absent in the literature.

In this study, as an effort of Indian genome variation Consortium, we screened the SNPs of or PEO1 gene such as rs7184, rs1535349, rs2863095, rs3740484, rs3740488, rs3740489, rs4919511, rs17113613, rs3824783, rs3740485, rs3740486, and rs3740487 in discovery panel (a population set of 40 samples and each sample representing a population for initial screen) and four synonymous SNPs, namely rs3824783 (ancestral allele=A), rs3740485 (ancestral allele=T), rs3740486 (ancestral allele=C), and rs3740487 (ancestral allele=A) in a large validation panel composed of 55 Indian subpopulations. The allelic and genotype frequencies are found to be different among different ethnic subpopulations of India.

Access this article online	
Website: http://www.cysonline.org	Quick Response Code
DOI: 10.4103/2229-5186.103100	

Ashok Singh, Amit Kumar Mitra, Indian Genome Variation Consortium[1], Srikanta Kumar Rath

Toxicology Division, Central Drug Research Institute, Lucknow, Uttar Pradesh, [1]Nodal Laboratory, Institute of Genomics and Integrated Biology (A unit of CSIR), New Delhi, India

Address for correspondence:
Dr. Srikanta Kumar Rath,
Scientist, Genotoxicity Laboratory, Division of Toxicology,
Central Drug Research Institute, Lucknow – 226 001, Uttar Pradesh, India.
E-mail: rathsk_2000@yahoo.com

Materials and Methods

Study population and sample collection

This study was conducted on DNA samples collected from 55 different ethnic populations of India. A complete detail composition of DNA validation panel and the Indian Genome Variation Consortium is given elsewhere (http://www.igvdb.res.in/discovery.php). Each SNP was screened in a panel of 1871 unrelated DNA samples from different region, linguistic group, and morphological types of India. Among these populations, genetic heterogeneity was significantly greater than zero showing population heterogeneity among different populations.[6] Prior to sample collection, ethical clearance was obtained from the Institutional Ethics Committee (IEC) following the guidelines of Indian Council of Medical Research (ICMR) (http://icmr.nic.in). Several field trips were made to collect samples, and voluntary participants were informed in the beginning about the purpose of the study and data handling.[7]

Briefly, a total of 55 subpopulations (1871 DNA samples) were identified and collected for validation panel to check the frequencies of SNPs in PEO1. Validation panel comprises 31 Indo-European, 4 Tibeto-Burman, 12 Dravidian, and 8 Austro-Asiatic linguistic subpopulations of Indian origin.[6] For each subpopulation of validation panel, a maximum 46 and minimum 23 samples were collected.[7]

Blood samples (5–10 ml) were collected by venipuncture and transferred to the tube with acid citrate dextrose as anticoagulant agent (0.9 ml for each 5 ml of blood) (Vacutainers, Medigene Co. Ltd). The vacutainers were kept on ice or at 4°C until process for DNA isolation. Genomic DNA was isolated using GenElute™ blood DNA isolation kit (Sigma) and quantified using DNA Quant (GE HealthCare).

Primer design, polymerase chain reaction and electrophoresis

Primers were designed for all the five exons of PEO1 with the help of Primer Select module of Lasergene v6.0, (DNA STAR™) (See supplementary [Table 1]) and procured from Sigma. Gene-specific amplification of PEO1 was performed in thermal cycler (MJ Research) using standard polymerase chain reaction reagents (Sigma) and the amplified product was resolved on 1.2% agarose (Sigma) gel.

Validation/genotyping of SNPs using MALDI-TOF

Following electrophoresis, the polymerase chain reaction product was eluted using MinElute™ Gel Extraction Kit (Qiagen) and sequenced (ABI 3100) bidirectionally using exon specific primers at the Centre for Genomic Application (TCGA), Okhla, New Delhi. All the SNPs before validation were screened through the discovery panel (http://www.igvdb.res.in/discovery.php). All the validation of SNPs in different Indian subpopulations was done using SEQUENOMEplatform in TCGA using the homogenous MassExtend (hME) assay based on allele-specific primer extension followed by MALDI-TOF technology.

Sequence analysis and statistics

DNA sequences were analyzed with the help of SeqMan module of Lasergene v6.0 (DNA STAR™) and compared with the reference sequence (NT_030059.12) for PEO1 (http://www.ncbi.nlm.nih.gov/) for initial SNP identification and comparison to reference sequence. For genotype data analysis, softwares like GENCOUNT and ALLHET (written by SujitMaiti, Center for Population Genetics, Indian Statistical Institute, Kolkata) were used.

Results and Discussion

SNPs and allele frequencies in PEO1 gene in discovery panel

Four SNPs (s3740485, rs3740486, rs3740487, and rs3824783) reported in NCBI database in non-coding regions of PEO1 gene were found in different population samples. The SNPs rs3740485 (intron 3, position 3300 in gene, allele "C" replaced by "T"), rs3740486 (intron 3, position 3302 in gene, allele "T" replaced by "C"), rs3740487 (intron 4, position 3462 in gene, allele "C" replaced by "A"), and rs3824783 (intron 4, position 3530 in gene, allele "G" replaced by "A") were present in populations of discovery panel (for location of these SNPs, see [Figure 1]). All these four SNPs (rs3740485; rs3740486, rs3740487, and rs3824783) were present in three caste (dSNP9, -18, and -22, all IE) and three tribal (dSNP 29, 31, 32, all AA) subpopulations of India. In addition, the "C" allele of rs3740486 was also present in dSNP6 (IE, Tribe) and dSNP24 (DR, Tribe) of Indian populations. Similarly, allele "A" of SNP rs3740487 and rs3824783 was present in dSNP1 (caste, IE) [Table 1]. Other eight validated SNPs reported at NCBI database (rs7184, rs1535349, rs2863095, rs3740484, rs3740488, rs3740489, rs4919511, and rs17113613) were absent in the discovery samples for PEO1. So we did not choose them for further validation.

The allelic frequencies were calculated for all four SNPs within the PEO1 [Table 2]. Nine populations for SNP

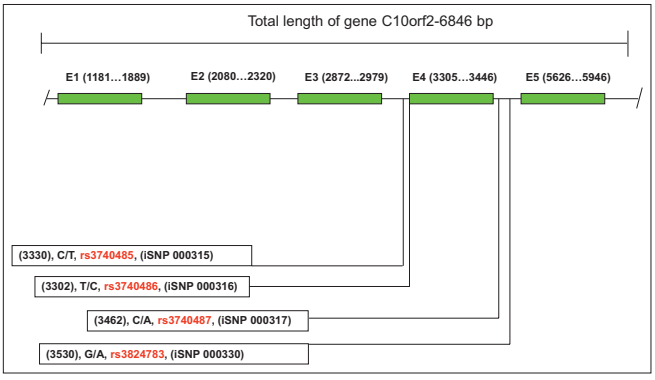

Figure 1: Location of the four SNPs of PEO1 gene in present study. Note: iSNP = Indian SNP ID

rs3740485 (dSNP1, 5, 6, 10-12, 23, 24, and 28), 7 for rs3740786, (dSNP1, 5, 10, 12, 17, 28, and 30), 4 for rs3740787 (dSNP5, 10, 11, and 30), and 10 for rs3824783, (dSNP4, 5, 6, 11, 12, 23, 24, 25, 26, and 30) in gene C10orf2 were found to be heterozygous out of 32 populations [Table 2].

SNPs and allele frequencies in PEO1 gene in validation panel

Allele frequencies for the allele "C" of the SNP rs3740485 were found to be <0.5 in 36 populations; =0.5 in 3 populations, between 0.50 and 0.90 in 9 populations, >0.90 in 1 population, 1 in 3 populations (IE-NE-LP_NSD, IE-E-LP_ORB, and DR-S-IP_KRM), and absent in 3 populations (IE-N-LP_SPB, IE-S-IP_HPK, and DR-S-IP_PNY). As a result, 20 subpopulations were found to be heterozygous [Figure 2]. A list of validation population details is given in Supplementary [Table 2].

Similarly, allele frequency for the allele "C" of the SNP rs3740486: 7 populations were with frequency <0.50,

Table 1: Genotypes of four SNPs of PEO1, rs3740485, rs3740486, rs3740487 and rs3824783 in representative discovery panel of Indian populations

Sample (Geography)	Linguistics	Discovery panel ID	rs3740485	rs3740486	rs3740487	rs3824783
West Bengal	IE-E-LP	dSNP1	CT	CT	AA	AA
Uttar Pradesh	IE-N-LP	dSNP2	CC	TT	CC	GG
Himachal Pradesh	IE-N-IP_T	dSNP3	CC	TT	CC	GG
West Bengal	IE-E-LP	dSNP4	CC	TT	CC	AG
Bihar	IE-E-LP	dSNP5	CT	CT	AC	AG
Rajasthan	IE-W-IP_T	dSNP6	CT	CC	CC	AG
Rajasthan	IE-W-LP	dSNP7	CC	TT	CC	GG
Maharashtra	IE-W-LP	dSNP8	CC	TT	CC	GG
Rajasthan	IE-W-SP	dSNP9	TT	CC	AA	AA
Maharashtra	IE-W-SP	dSNP10	CT	CT	AC	GG
Maharashtra	IE-W-SP	dSNP11	CT	TT	AC	AG
Maharashtra	IE-N-LP_T	dSNP12	CT	CT	CC	AG
Himachal Pradesh	IE-N-LP	dSNP13	CC	TT	CC	GG
Himachal Pradesh	IE-N-LP	dSNP14	CC	TT	CC	GG
West Bengal	IE-N-LP	dSNP15	CC	TT	CC	GG
Uttar Pradesh	IE-N-SP	dSNP16	CC	TT	CC	GG
Haryana	IE-N-SP	dSNP17	CC	CT	CC	GG
Punjab	IE-N-SP	dSNP18	TT	CC	AA	AA
Punjab	IE-N-LP	dSNP19	CC	TT	CC	GG
Himachal Pradesh	IE-N-LP	dSNP20	CC	TT	CC	GG
Karnataka	IE-S-IP_T_T	dSNP21	CC	TT	CC	GG
Maharashtra	IE-C-LP	dSNP22	TT	CC	AA	AA
Meghalaya	TB-NE-IP_T	dSNP23	CT	TT	CC	AG
Andhra Pradesh	DR-S-IP_T	dSNP24	CT	CC	CC	AG
Andhra Pradesh	DR-S-LP	dSNP25	CC	TT	CC	AG
Karnataka	DR-S-LP	dSNP26	CC	TT	CC	AG
Chattisgarh	DR-C-LP	dSNP27	CC	TT	CC	GG
West Bengal	AA-E-IP_T	dSNP28	CT	CT	CC	GG
Jharkhand	AA-E-IP_T	dSNP29	TT	CC	AA	AA
Andaman & Nicobar	AA-S-IP_T	dSNP30	CC	CT	AC	AG
Andaman & Nicobar	AA-S-IP_T	dSNP31	TT	CC	AA	AA
Maharashtra	AA-C-IP_T	dSNP32	TT	CC	AA	AA

IE – Indo-European; TB – Tibeto-Burman; DR – Dravidian; AA – Austro-Asiatic; LP – Large population; SP – Small population; IP – Isolated population; T – Tribes; and all others are caste populations

Table 2: Genotypes and allelic frequencies of different SNPs in discovery panel for PEO1

Gene name	SNP ID	p Allele (no in bracket is population samples)	q Allele	Heterozygous	Total no. of genotype	Allele freq. (p)	Allele freq. (q)
PEO1	rs3740485	CC (17)	TT (6)	CT (9)	32	0.672	0.328
	rs3740486	CC (8)	TT (17)	CT (7)	32	0.359	0.641
	rs3740487	CC (21)	AA (7)	AC (4)	32	0.719	0.281
	rs3824783	GG (15)	AA (7)	AG (10)	32	0.625	0.375

PEO1 - Progressive external opthalmoplegia

1 population with 0.50 (DR-S-LP_PDC), 13 populations between 0.50 and 0.90, 22 populationd between 0.90 and 1.00, 11 populations with frequency 1.00, and absent in 1 population (IE-N-SP_SYD). Total 44 subpopulations were heterozygous for rs3740486 [Figure 3].

Allele frequencies for the allele "C" of the SNP rs3740487 were found to be 0.5 in 1 population (IE-NE-IP_HJG), between 0.50 and 0.90 in 51 populations, between 0.90 and 1.00 in 2 populations (IE-N-LP_KKB and AA-E-IP_STL), and 1 population with frequencies 1.00 (IE-N-LP_RJU). Total 51 subpopulations were heterozygous in nature for rs3740487 [Figure 4].

Allele frequencies for the allele "G" of the SNP rs3824783 were found to be less than 0.5 in 2 subpopulations (IE-NE-IP_HJG, DR-S-IP_PNY), =0.5 in 1 subpopulation (IE-W-IP_BHL), and between 0.50 and 0.90 in 52 subpopulations. All 55 subpopulations were heterozygous for rs3824783 [Figure 5].

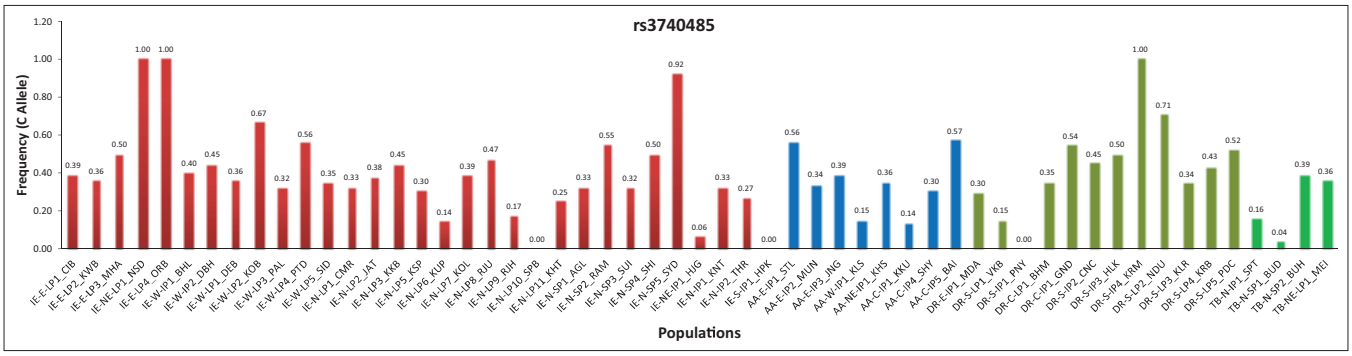

Figure 2: Frequencies of "C" allele of SNP rs3740485 in different subpopulations of India

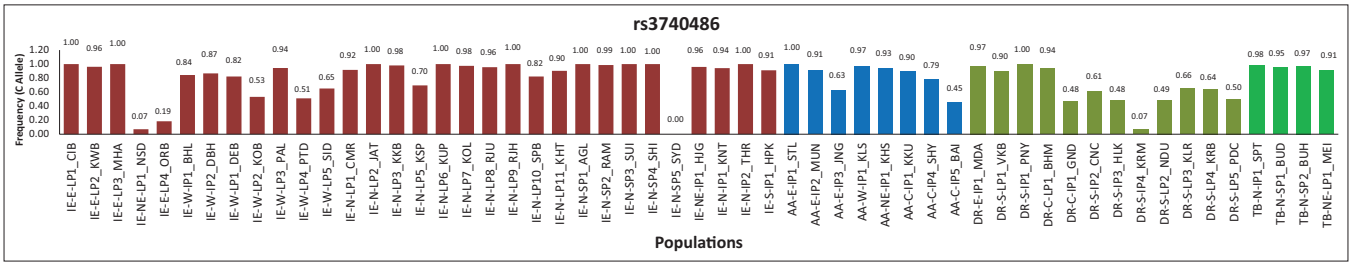

Figure 3: Frequencies of "C" allele of SNP rs3740486 in different subpopulations of India

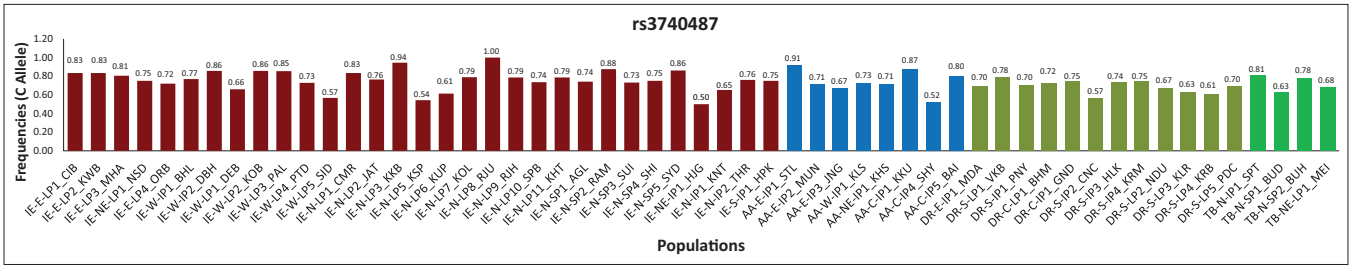

Figure 4: Frequencies of 'C' allele of SNP rs3740487 in different sub populations of India

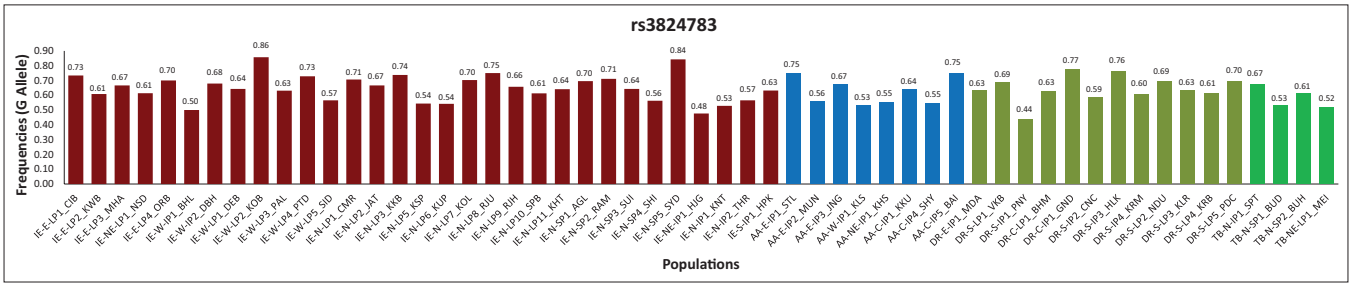

Figure 5: Frequencies of 'G' allele of SNP rs3824783 in different sub populations of India

The following populations were found to be monomorphic (allele frequency = 1) for the "C" allele of rs3740485, 3 subpopulations (IE-NE-LP_NSD, IE-E-LP_ORB and DR-S-IP_KRM), for "C" allele of rs3740486, 11 subpopulations, for "C" allele of rs3740487, 1 subpopulation (IE-N-LP_RJU), and for "G" allele of rs3824783 (none of the population). The data show the maximum number of 11 subpopulations were monomorphic for "C" allele of rs3740486. Additionally, in rs3740485 there were three subpopulations, where allele "C" was replaced with allele "T". Similarly, in rs3740486, only one subpopulation was with 100% frequency of allele "T". In case of rs3740487 and rs3824783, no subpopulations were with alternative alleles "A" and "G", respectively.

Interestingly, the Indian populations found to be polymorphic in discovery panel such as dSNP6, -24, -49, and -53 were also polymorphic in large population samples of validation panel.

Frequency and genetic status of each SNP was not uniform in all the populations, indicating that all of these SNPs were in different phase of their life cycle (homozygous allele1/heterozygous condition, and homozygous allele 2) in different Indian subpopulations. It may be recalled that each SNP has to undergo a continuous cycle of birth and death within the long time of population history.[8] Following analysis of four SNPs of PEO1 in validation panel, the frequencies of the particular alleles were analyzed extensively. There were very few subpopulations, three for rs3740485, and one for rs3740486, rs3740487, and rs3824783, where the frequencies of both the alleles were equal. These frequencies represent an equal distribution of both the allele within Indian subpopulations (i.e., p = q). The following populations were found to be monomorphic 3 subpopulations for the "C" allele of rs3740485; 11 subpopulations for "C" allele of rs3740486; 1 subpopulation for "C" allele of rs3740487, and 1 subpopulation for "C" allele of the SNP rs3176388, indicating that the frequencies of the allele "C" became fixed and the frequency of alternative allele was zero.

For allele "C" in rs3740485, a total of 36 subpopulations are with frequencies <0.50, indicating that the allele "C" is a recent allele in these populations and now in the third phase of its life cycle. This is the potentially the lengthy phase of a SNP, where risk of loss is reduced, and the allele will increase the frequencies after survival in early stages.[8] Other SNPs such as rs3740486 (for allele C), rs3740487 (for allele C), rs3824783 (for allele G), and rs3176388 (for allele C) in maximum number of subpopulations are with frequencies >0.50 [Figures 4-9]. Additionally, for allele "C" in rs3740486, a maximum of 22 subpopulations are with frequency category >0.90 and <1.0 [Figure 3]. The result demonstrates that these alleles are increasing their frequencies in the different subpopulations toward fixation.

Additionally, in rs3740485 there were three subpopulations, where allele "C" is replaced with allele "T". Similarly, in rs3740486, only one subpopulation is with 100% frequency of allele "T" (fixation). This indicates an increase in the frequencies of the allele "T" in those populations in contrast to other subpopulations. In case of rs3740487 and rs3824783, no subpopulations were with alternative alleles "A" and "G", respectively [Figures 4,5,8,9]. These allelic frequency variations may be due to within and between population stratification, ancestral geographical migration, marriage practices, reproductive expansions and bottlenecks, and stochastic variation.[9,10]

The extent of genetic diversity among different Indian population is well observed in the entire globe with the exception of African population.[11] However, the Indian

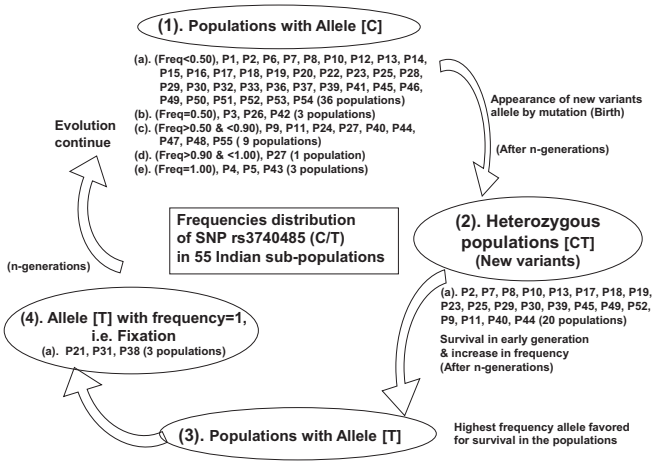

Figure 6: Sketch diagram showing distribution of frequencies of SNP rs3740485 (C to T change after n-generation time) among 55 Indian sub populations along with four life cycle stages of the SNP within the same sets of Indian sub populations

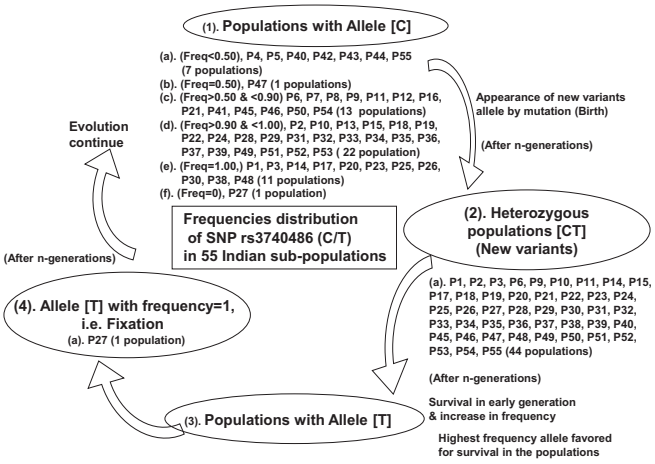

Figure 7: Sketch diagram showing distribution of frequencies of SNP rs3740486 (C to T change after n-generation time) among 55 Indian sub populations along with four life cycle stages of the SNP within the similar sets of Indian sub populations

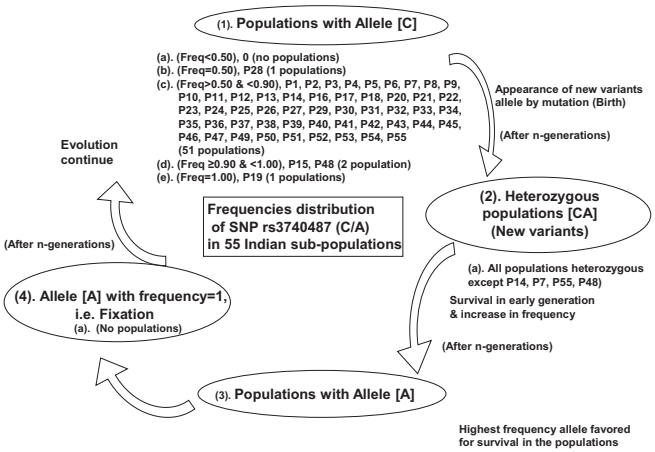

Figure 8: Sketch diagram showing distribution of frequencies of SNP rs3740487 (C to A change after n-generation time) among 55 Indian sub populations along with four life cycle stages of the SNP within the similar sets of Indian sub populations

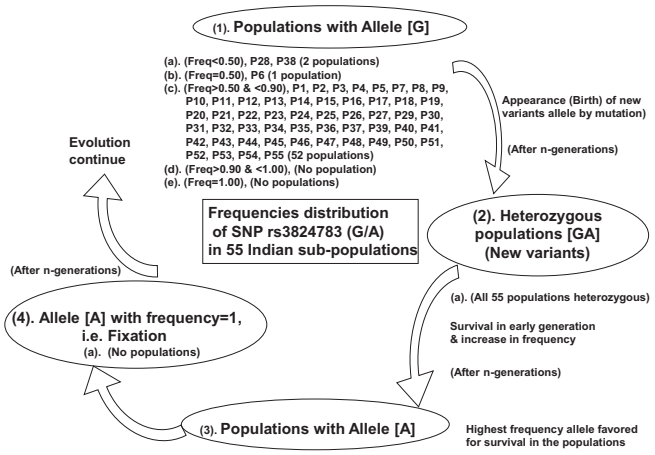

Figure 9: Sketch diagram showing distribution of frequencies of SNP rs3824783 (G to A change after n-generation time) among 55 Indian sub populations along with four life cycle stages of the SNP within the similar sets of Indian sub populations

population comprises more than a billion people, 4693 communities, and several thousands of endogamous groups (in-marrying),[12] which makes Indian population more diverse at genetic level. Moreover, the genetic variability may be contributed due to inbreeding,[13] different migration ways, admixture, and population stratification.[14,15]

Results from this study within a limited number of SNPs indicate extensive diversity in SNP and their frequency distribution among various Indian subpopulations. Collectively, these frequencies indicate that Indian subpopulation reveals an enormous variability at genetic level and such population can be used as a canvas for disease association study such as whole genome association, new gene discovery, and future pharmacogenomics studies.

References

1. Spelbrink JN, Li FY, Tiranti V, Nikali K, Yuan QP, Tariq M,*et al.* Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria. Nat Genet 2001;28:223-31.
2. Ruiz-Pesini E, Lott MT, Procaccio V, Poole JC, Brandon MC, Mishmar D, *et al.* An enhanced MITOMAP with a global mtDNA mutational phylogeny. Nucleic Acids Res 2007;35:D823-8.
3. Zeviani M, Servidei S, Gellera C, Bertini E, DiMauro S, DiDonato S. An autosomal dominant disorder with multiple deletions of mitochondrial DNA starting at the D-loop region. Nature 1989;339:309-11.
4. Van Goethem G, Dermaut B, Löfgren A, Martin JJ, Van Broeckhoven C. Mutation of POLG is associated with progressive external ophthalmoplegia characterized by mtDNA deletions. Nat Genet 2001;28:211-2.
5. Kaukonen J, Juselius JK, Tiranti V, Kyttälä A, Zeviani M, Comi GP,*et al.* Role of adenine nucleotide translocator 1 in mtDNA maintenance. Science 2000;289:782-5.
6. Genetic landscape of the people of India: A canvas for disease gene exploration. J Genet 2008;87:3-20.
7. The Indian Genome Variation database (IGVdb): A project overview. Hum Genet 2005;118:1-11.
8. Miller RD, Kwok PY. The birth and death of human single-nucleotide polymorphisms: New experimental evidence and implications for human history and medicine. Hum Mol Genet 2001;10:2195-8.
9. Cavalli-Sforza LL, Piazza A. Human genomic diversity in Europe: A summary of recent research and prospects for the future. Eur J Hum Genet 1993;1:3-18.
10. Slatkin M. Inbreeding coefficients and coalescence times. Genet Res 1991;58:167-75.
11. Majumder PP. People of India: Biological diversity, affinities. Evol Anthropol 1998;6:100-10.
12. Singh K. People of India: Introduction National Series, Delhi, India: Oxford University Press; 2002.
13. Malhotra K, Vasulu TS (Editors.) Structure of human populations in India. , New York:Plenum Press; 1993.
14. Gadgil M, Joshi N, Manoharan S, Patil S, PrasadUV, (Editors.) Peopling of India; in The human heritage (eds) Hyderabad, India: Universities Press; 1998.
15. Bhasin M, Walter H, Danker-Hopfe. An Investigation of Biological Variability in Ecological Ethnoeconomic and Linguistic Groups. in, People of India, Delhi, India: Kamla-Raj Enterprises; 1994.

How to cite this article: Singh A, Mitra AK, I, Rath SK. Analysis of single-nucleotide polymorphisms of PEO1 gene in 55 ethnic groups of India. Chron Young Sci 2012;3:304-9.

Source of Support: Financial support from Council of Scientific and Industrial Research (CSIR) Government of India, Task Force Project on 'Predictive Medicine using repeat and single nucleotide polymorphisms (CMM0016), **Conflict of Interest:** None declared