Int J Hum Genet, 7(2): 123-131 (2007)
PRINT: ISSN 0972-3757 ONLINE: 2456-6360
DOI: 10.31901/24566330.2007/07.02.01

Phylogenetic Applications of HLA Class II Loci

Uddalak Bharadwaj, Faisal Khan, Sanjeev Srivastava, Himanshu Goel and Suraksha Agrawal

Department of Medical Genetics, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Raebareli Road, Lucknow 226 014, Uttar Pradesh, India

KEYWORDS HLA; Phylogenetic reconstruction; maximum likelihood; heterozygosity; gene diversity; bootstrapping

ABSTRACT Human Leukocyte Antigen (HLA) loci widely known for their role in generation of immune responses are often considered to be effective in reconstructing human phylogenies due to high degree of polymorphism and rarity of recombination observed at HLA loci. In the present study, we have made an attempt to substantiate the phylogenetic potential of HLA class II loci by analyzing DRB1, DQA1 and DQB1 in 202 North Indians. High allelic diversity was found at all the three loci total 69 alleles were observed (39 at DRB1, 10 at DQA1 and 19 at DQB1), along with high-observed heterozygosity (Avg. Heter. 0.73). Data generated from the study was then assessed for phylogenetic reconstruction based on maximum likelihood criterion along with statistical bootstrapping procedure involving 1000 replicates. The ensuing tree is further compared with those obtained in earlier phylogenetic reports. The compiled database of 20 populations got segregated and finely resolved in three basal clusters with very high bootstrap values corresponding to four geo-ethnic groups of African, Orientals, American and Caucasians. ML (maximum likelihood) phylogram has placed North Indian Hindus alongside other Caucasian populations, strengthening the findings of other markers. This indicates that if an appropriate analysis is carried out on a set of populations which represents different geographic followed by proper interpretation based on more logistic statistical model, then there is s high possibility that HLA class II loci can infer exact and accurate phylogenetic assessments as revealed by mt-DNA and Y-chromosome markers.

INTRODUCTION

Findings of 'Human Genome sequencing' project bestowed an imposing challenge to genetists all over the world. The challenge involves development of a detailed understanding of the heritable variation in the human genome. This understanding of human genetic variation encompasses the endeavor of reconstructing the human phylogenies and determining the amount, pattern, distribution and structuring of genetic diversity across different geo-ethnic, socio-cultural and linguistic human groups (Cavalli-Sforza 2005; Bamshad et al. 2004).

The past decade of advances in molecular genetic technology and bionforamtics tools have opened a new era for the science of human evolution (Cavalli-Sforza and Feldman 2003; Jorde and Wooding 2004). Human DNA has become the blueprint of all the genetic diversity studies carried out in last decade. Still, the choice of population, polymorphic marker, genotyping procedures and statistical methods formulates the main strategies for human genetic diversity analysis. Choice of population has been concentrated on the populations of widely separated geographical zones like sub-Saharan Africans, Han Chinese and Europeans. However, the admixed populations of geogra-phically intermediate zones like Central Asia, Middle East or Southwest Asia are generally unexplored or rarely studied (Agrawal et al. 2001; Agrawal and Khan 2005, 2006).

The choice of marker has a long journey starting from blood grouping to lineage based NRY and mt-DNA and LD based SNP haplotype blocks. However, the preferred regions are those that exhibit high degree of polymorphism; have significantly different frequency distribution in different groups and carries signature alleles/ haplotypes in segregating populations (Cavalli-Sforza 2005). Several polymorphic regions of human genome have been used in recent years to infer human phylogenies. These polymorphic regions include Short tandem repeats (STR) (Rosenberg et al. 2002; Rowold and Herrera 2003; Agrawal and Khan 2005), mt-DNA (Pallanichamy et al. 2004; Ingman et al. 2000), Y-chromosome (Underhill et al. 2001; Semino et al. 2000; Agrawal et al. 2005), Polymorphic Alu insertion (PAIs) (Antunez-de-Mayolo et al. 2002) and Human Leukocyte antigens (HLA) (Tsuneto et al. 2003)

Corresponding author: Suraksha Agrawal, Department of Medical Genetics, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Raebareli Road, Lucknow, 226014 Uttar Pradesh, India
Phone: 091-522 -2668004-8 Ext 2338, 2346, 2347, 2339
Fax: 091-522 -26680973/6680017
Email: suraksha@sgpgi.ac.in

etc. Each of these markers carries various features, which makes them suitable and successful in inferring human phylogenetic relationships.

STR loci exhibit substantial allelic variability due to high rate of germline mutation and hence preferred to infer recent evolutionary patterns (Khan and Agrawal 2005). On the contrary PAIs represents identity by descent and a known ancestral state defined by the absence of the *Alu* element, having more chances to carry signature alleles (or genotypes) for different populations (Antunez-de-Mayolo et al. 2002). These two regions have an added advantage of simple and robust PCR based allelic detection. However, the UEPs of non recombining region of Y-chromosome (NRY) and control region (HVRI and II) of mt-DNA have paved the way for reconstructing the genealogies of both extinct and extant haplotypes in the form of bifurcating trees remains favorite due to their uni-parental segregation that can be used to delineate paternal and maternal lineages respectively (Jobling and Tyler Smith 2003). Further, negligible recombination, uni-parental segregation and enblock segregation of different SNPs on mt-DNA and Y-chromosome have been used to delineate paternal and maternal lineages respectively (Underhill et al. 2001; Ingman et al. 2000). At the same time, various reports based on the analysis of HLA class I and class II loci continue to appear, however, most of them are focused on genetic structuring of the populations with little or no emphasis on biological relationship among different populations (Tsuneto et al. 2003).

Human Leukocyte Antigen (HLA) system, a genetic set of more than 100 genes is the most important component of human Major Histocaompatibility Complex (MHC). Genes of HLA system resides on short arm of chromosome 6 (6p21.31) and spans a region of approximately four million base pair, equivalent to 0.1% of total human genome (Perez-Miranda et al. 2003). HLA genes display an extreme degree of genetic polymorphisms due to their role in generation of immune responses by encoding cell-surface heterodimers which are involved in antigen presentation, tolerance and self and non self recognition, The HLA loci comprises of three main regions based on the molecular structure they codify. They include the Class I and Class III (telomeric); and class II (centromeric) loci. Out of these three regions, HLA class II are least polymorphic, while HLA class I loci are the most polymorphic with 382,554 and 167 alleles reported respectively for A, B and C class I loci. HLA class II loci are moderately polymorphic with 459, 27 and 59 alleles reported for DRB1, DQA1 and DQB1 class II loci respectively (http://www.allele-frequencies.net). The high rate of polymorphism of HLA loci offers possibility to infer phylogenetic assessment of different populations.

There are several reasons to believe that HLA class II loci are at par with mt-DNA and Y-chromosome markers in phylogenetic assessment. They are both vastly polymorphic and highly linked loci that results into unique linkage disequilibria pattern, signature alleles and haplotype segregation among different populations (Begovitch et al. 1992; Grubic et al. 2000). High preference of lineage based markers (NRY and mtDNA) is mainly because of their non-recombining property, but only maternal or paternal lineages do not represent the entire genetic makeup of an individual. Therefore, a biparentally segregated HLA class II loci can be an ideal candidate for inferring phylogenetic relationships between different populations.

Present study has been carried out to explore the utility of HLA class II loci in inferring phylogenetic relationships. For this purpose, we have analyzed three HLA class II loci in 202 random North Indian Hindus from the State of Uttar Pradesh. Indian subcontinent is one of the most heterogeneous populations in the world, having enormous genetic, cultural and linguistic diversity (Roychoudhury et al. 2000). The important fact about the present day North Indian Hindu population (north-west India in particular) is the impending role of extensive gene flow through a series of migrations and invasions that have shaped and distributed the contemporary genetic variation (Balakrishnan et al. 1978). The complex history and the action of evolutionary processes operating on the numerous subdivisions of the Indian population are reflect-ed in the heterogeneous genetic constitution of this sub-continent.

To approach our goal of inferring phylogenetic relationships between different population groups by using HLA class II loci, we have compiled a database of geographically targeted and ethnically diverse set of 19 popu-lation in which all the three HLA class II loci – DRB1, DQA1 and DQB1 have been studied. Same set of loci has been studied in the North Indian Hindu

population. Both analysis of intra population genetic variation and phylogenetic relationship with other global populations have been carried out. The ensuing phylogenetic patterns were then compared with those obtained in earlier phylogenetic investigations to ensure the success of HLA class II loci in human genetic variation studies.

MATERIALS AND METHODS

Populations: A total of 202 unrelated North Indian Hindus were randomly selected. Before sample collection, regional addresses and detailed computerized lists were prepared. Random numbers were generated with the help of computer and samples were collected from the different collection sites of Uttar Pradesh-Lucknow, Kanpur and Agra. Whole blood obtained by venipuncture was collected in EDTA vacutainer tubes. Three-generation pedigree charts were prepared to assure un-relatedness in all the samples. The ethical committee of the institute approved the study and blood samples were taken after obtaining informed consent from the subjects.

DNA Extraction and HLA Class II Loci Genotyping: DNA was extracted by phenol chloroform method as described by Comey et al. (1994) and purified by ethanol precipitation. HLA typing was done using primers and sequence specific oligonucleotide probes (SSOP) according to the 13th International Histocompatibility Workshop (http://www.ihwg.org/protocols/hla_c.pdf).

Population Database: We have selected 19 populations; the criterion of selection was to cover the major geographical regions i.e. Africa, Asia, Europe and America and ethnic groups such as African, Caucasoid, Amerindians and Orientals. All the populations selected have allele frequency data for three HLA class II loci – DRB1, DQA1, and DQB1. In Table 1 all the populations and their sources compiled from allele frequency database of 13th Histocompatibility Workshop has been shown.

Statistical Analysis: Allele frequencies were calculated by a simple gene count method. Three different parameters average gene diversity, average observed heterozygosity, and mean numbers of pair wise differences were calculated using ARLEQUIN V2 software to quantify intra population diversity. Phylogenetic reconstruction was done based on maximum likelihood (ML) and the HLA class II loci- DRB1, DQA1 and DQB1 allele frequency distribution using CONTML in PHYLIP v3.5c (Felsenstein 1993). A statistical bootstrap involving 1000 replicates

Table 1: Details of the 19 different world populations included in the analysis obtained from 13th Histocompatibility Workshop

S.No.	*Populations*	*No. of samples*	*References*
Africans			
1	Cameroon	126	Pimtanothai et al. 2001
2	Algerians	47	Djoulah et al. 1994
3	Moroccan Arabs	96	Gomez-Casado et al. 2000
4	Ethiopian Amharas	98	Fort et al. 1998
5	Ethiopian Oromo	83	Fort et al. 1998
Europeans			
6	Germans	174	13th Histocompatibility workshop
7	Italy	53	Lulli et al. 1998
8	Spanish Pesiegos	88	Sanchez Yelasco et al. 2003
9	Spanish Basque	83	Sanchez Yelasco et al. 2003
10	Russian	140	Evseeva et al. 2002
11	French CEPH	248	Bugawan et al. 2000
12	Czech Republic	106	13th Histocompatibility workshop
13	Turkey	250	Sarukan-Direskineli et al. 2000
East Asians			
14	Chinese	76	13th Histocompatibility workshop
15	Thailand	142	13th Histocompatibility workshop
Native Americans			
16	Brazilian Kaingangs	235	Tsuneto et al. 2003
17	Brazilian Nandevas	87	Tsuneto et al. 2003
18	Argentinians	157	13th Histocompatibility workshop
19	Canadian Athabaskans	62	Monsalxe et al. 1998

was carried out using SEQBOOT option of PHYLIP v3.5c. Finally, a consensus of 1000 trees was drawn using CONSENSE option of PHYLIP v 3.5.

RESULTS

1. Allele Frequency Distribution

Analysis of three HLA class II loci- DRB1, DQA1 and DQB1 has revealed high allelic variability among North Indian Hindus. Total 68 alleles have been observed at all the three HLA class II loci, 39 at DRB1, 10 at DQA1 and 19 at DQB1 loci. Multilocus genotype frequencies were utilized for testing the conformity of the assumption of HWE. After applying Bonferroni correction to the 'p' value, no significant departures from HWE were observed at any of the three loci. Allele frequency distribution at all the three HLA class II loci among North Indian Hindus is shown in Table 2.

1.1 Allele Frequency Distribution at DRB1 Loci: Total 39 alleles have been observed at DRB1 loci. Most frequent among these is 0701 (18.8%) followed by 1001, 1101, 1501 and 1502 (9% each). SSOP probes used in the study do not distinguish DRB*080x subtypes, similarly DRB1*090x has not been subtyped due to lack of specific probes. Allele 0301 and 1301 were observed in a frequency 7%. Allele distribution at DRB1 suggested presence of strong Caucasoid element among North Indian Hindus as most of the Mediterranean and European population have similar allelic profile. 1101 has been found in a similar frequency among Turkey (10%), Czechs (8.7%) and Germans (6.5%), although 1101 is found in similar frequency among oriental populations also (9% in Chinese), but relatively low among Americans and African groups (Pimtanothai et al. 2001; Fort et al. 1998). Similarly 1301 is observed in nearly equal frequency in Germans (9%) and French Caucasians (8.5%), but occur in very low frequencies among other ethnic groups (Bugawan et al. 2000). Moreover, some of the alleles frequent among Orientals – 0803 (6.7% among Chinese and 7% among Thais) and 0901 (15% among Chinese and 12% among Thais) are absent or found in a very low frequency among North Indian Hindus. At the same time, alleles highly frequent among African populations- 0103 (13% in Moroccans and Cameroon and 15 % in

Table 2: Allele frequency distribution at DRB1, DQA1 and DQB1 loci among North Indian Hindus (n= 202)

DRB1 alleles	*n*	*Frequency*
*01	9	0.022
*0101	4	0.01
*0103	5	0.01
*02	86	0.212
*1501	36	0.09
*1502	35	0.087
*1503	6	0.015
*1504	5	0.012
*150x	2	0.005
*160x	2	0.005
*0301	31	0.077
*04	30	0.074
*0401	1	0.002
*0402	2	0.005
*0403	13	0.032
*0404	6	0.015
*0405	3	0.007
*0406	2	0.005
*0407	2	0.005
*040x	1	0.003
*0701	76	0.188
*080x	1	0.003
*090x	4	0.01
*1001	36	0.09
*11	44	0.108
*1101	36	0.09
*1102	2	0.005
*1103	3	0.007
*1104	2	0.005
*1108	1	0.003
*12	11	0.027
*1201	1	0.003
*1202	8	0.02
*1203	1	0.003
*120x	1	0.003
*13	52	0.128
*1301	31	0.077
*1302	13	0.032
*1303	5	0.012
*1308	1	0.003
*1313	2	0.005
*14	24	0.059
*1401	8	0.02
*1402	3	0.007
*1403	4	0.01
*140x	9	0.022

DRB1 alleles	*n*	*Frequency*
*0101	13	0.032
*0102	62	0.153
*0103	56	0.138
*0104	57	0.141
*0201	74	0.183
*0301	43	0.106
*0302	4	0.01
*0401	7	0.017
*0501	77	0.19
*0601	11	0.027

Table 2: Contd. .....

DRB1 alleles	*n*	*Frequency*
*0201	87	0.215
*0301	59	0.146
*0302	20	0.05
*0303	27	0.067
*0304	1	0.002
*0305	2	0.005
*030x	1	0.002
*0401	1	0.002
*0402	7	0.017
*0501	48	0.119
*0502	6	0.015
*0503	26	0.064
*0504	2	0.005
*0601	61	0.151
*0602	13	0.032
*0603	30	0.076
*0604	10	0.025
*0605	2	0.005
*060x	1	0.002

Algerians) (Djoulah et al. 1994; Gomez-Casado et al. 2000) were also not found among North Indian Hindus.

1.2 Allele Frequency Distribution at DQA1 Loci: Total 10 alleles have been observed at DQA1 loci. 0501 (19%) and 0201 (18%) were the most frequent alleles. 0102 (13%), 0103 (14%) and 0104 (15%) were other frequent alleles. The frequent alleles ie 0201 and 0501 is seen at a high frequency among different Caucasian groups 0501 is found in a frequency of 20% among Germans and 39% among Turks, while 0201 is found with a frequency of 12% among Turks and 16% among French Caucasians (Sarukan-Direskineli et al. 2000; Bugawan et al. 2000). However, both these alleles are frequently found in other ethnic groups therefore it could be inferred that these are not specific to any particular population group.

1.3 Allele Frequency Distribution at DQB1 Loci: 19 segregating alleles have been observed at DQB1 loci with 0201 being the most frequent (21%) of all, followed by 0601 (16%), 0301 (14%) and 0501(11%). Allele distribution at DQB1 is not like DRB1 and DQA1 as most of the frequent alleles are found in more or less equal frequencies among all the ethnic groups. It is interesting to note that DQB1* 0601, is highly frequent among North Indian Hindus (15%), is quite rare among other Caucasians (0.5-3%) (http://www.allelefrequencies.net).

2. Intra Population Genetic Variation Analysis

HLA class II loci exhibits high degree of polymorphism and hence are preferred region of the human genome to evaluate intra population diversity. In the present study, three different parameters were used to quantify intra population diversity namely, average gene diversity, average observed heterozygosity and mean number of pair wise differences (Table 3). High allelic variability was further depicted by high-observed heterozygosity, men value being 73% (0.69 at DQA1, 0.72 at DQB1 and 0.78 at DRB1). Average gene diversity at diploid loci is indicative of average expected heterozygosity and high value of gene diversity corresponds to high intra population diversity. North Indians Hindus have depicted a high value of gene diversity, average being 74.5% (0.71 at DQA1, 0.74 at DQB1 and 0.79 at DRB1). The mean number of pairwise differences indicates the mean number of loci that vary between any pair of haplotypes drawn from the same population. This mean number of pairwise difference was found to be 2.64 ± 1.41. All the three estimates indicated high genetic diversity among north Indian Hindus.

Table 3: Different parameters inferring in tra-population genetic variation

Parameters depicting intra population diversity	
Average gene diversity	0.745 ± 0.52
Average observed Heterozygosity	0.732 ± 0.07
Mean number of pair wise differences	2.641 ± 1.41

3. Phylogenetic Assessment

Based on the allele frequency distribution of the three HLA class II loci, phylogenetic analysis was carried out for 20 populations (present study and compiled database of 19 other populations). A Model based approach (Maximum Likelihood – ML) was used to determine the phylogenetic relationship among different global populations. The ML-phylogenetic tree generated from present data set was drawn from CONTML algorithm that works upon the conjecture that random action of genetic drift is the solitary basis of the differences between allele frequencies in different populations (Felenstein 1993). An enrooted radial phylogram (Maximum Likelihood) is shown in Figure 1**.** The edge lengths displayed in the phylogram indicated that the amount of evolutionary change occurred along each branch. The scores next to the nodes represent the number of bootstrap replicates (out of 1000) exhibiting these specific bifurcations.

The ML phylogenetic tree depicts three

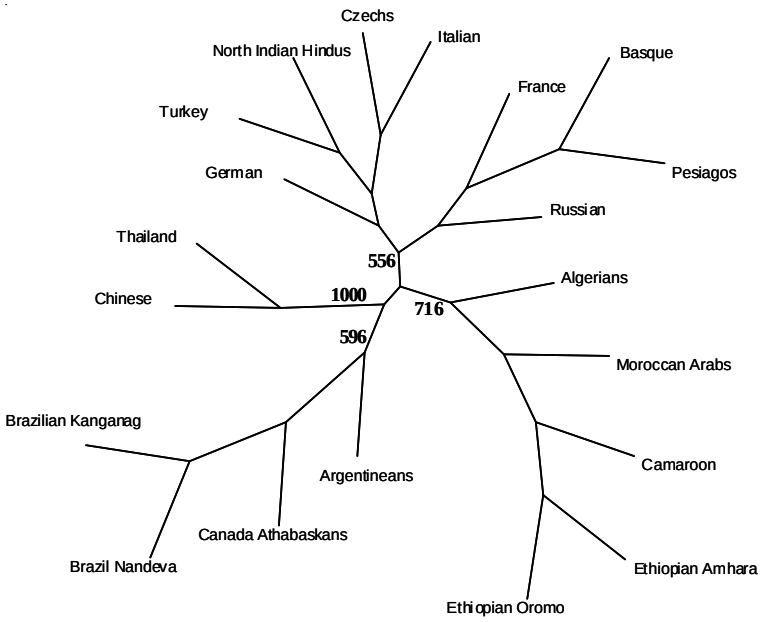

Fig. 1. ML Phylogram based on allele frequency distribution in 20 different populations at three HLA class II loci with statistical bootstrap involving 1000 replicates

different basal groups corresponding to three ethnic groups- African, Caucasian and Orientals, which further branched into East Asians and Americans. The Caucasian cluster has two monophyletic units- Turkish/ German/ North Indian and other Spanish Pesiagos/ Basque/ French Caucasoids. Total 13 out of the 18 nodes have bootstrap values more than 50%, most important among them is (i) division between Africans and other three groups of Caucasians, Americans and Oriental-716, (ii) the division between Caucasian and other three groups –556, (iii) the division between Orientals and other two groups-1000 and the node separating Orientals from American – 596. Overall, ML phylogenetic tree exhibits a clearer picture showing distinct and parallel bio-geographical associations.

DISCUSSION

Present study apart from quantifying the amount of genetic diversity in north Indian Hindus, also deciphered substantial and meaningful inferences about the phylogenies of diverse human populations solely based on three HLA class II loci. The Maximum Likelihood (ML) phylograms generated from allele frequencies of different HLA class II loci in 20 geographically diverse and ethnically varied populations have yielded finer resolution of recent human evolutionary history.

Indian sub-continent had experienced a massive gene flow from at least two Neolithic episodes of migrations. Firstly, ~ 10-15 KYA, when agriculture developed in the Fertile Crescent region, a part of an eastward wave of human migration (Cavalli-Sforza et al. 1994) entered India. This wave brought Dravidian languages into India (Renfrew 1987) mainly, Elamo-Dravidian languages (Ruhlen 1991). They are now confined to southeastern India and to some isolated groups in Pakistan and northern India. A later episode, the arrival of pastoral nomads from the central Asian steppes to the Iranian plateau, ~4,000 YBP, brought with it the Indo-European language family, which eventually replaced Dravidian languages from most of Pakistan and northern India, perhaps by an elite dominance process (Renfrew 1987). Therefore, the present north Indians are mainly considered to be the Indo-Aryan speakers with a possible origin in Eurasia.

All the three different analysis, allele frequency distribution, parameters quantifying the amount of genetic variation and maximum likelihood based phylogenetic assessment revealed that north Indians Hindus have strong Caucasian element in the genetic structure. Allele distribution at all the three loci, DRB1, DQA1 and DQB1 suggested that North Indian Hindus have similar allelic profile with that of the Mediterranean and European population mainly Turks, Czechs and Germans (Sarukan-Direskineli et al. 2000; Bugawan et al. 2000). Intra population variation analysis have depicted that the north Indians Hindus reveal a high range of intra population diversity deduced from average gene diversity (0.745), average observed heterozygosity (0.73), and mean number of pair wise differences (2.48 ± 6.78). A comparison of the average heterozygosity values at 3 HLA class II loci across population groups inhabiting varied geographical locations clearly establishes that heterozygosity estimates of North Indian Hindus are nearly similar to that of other Caucasian group, higher than that of Orientals but lower than that of the Africans (Bugawan et al. 2000; Sarukan-Direskineli et al. 2000; Pimtanothai et al. 2001; Tsuneto et al. 2003). Similar observations have been reported by several reports based on the regions of human genome like mtDNA (Roychudhary et al. 2000) and autosomal STR loci (Khan et al. 2004; Agrawal and Khan 2005).

Researchers have used various statistical models, mathematical algorithms and computational software to decipher the phylogenetic relationships between closely related and distant population groups. Most popular among them is Maximum likelihood criterion (Rowold and Herrera 2003). The ML algorithm works upon the assumption that random action of genetic drift is the only basis of the differences between allele frequencies between different population groups. Random genetic drift occurs due to a finite number of individuals participating in the formation of the next generation (Cavalli-Sforza 1994). This process is responsible behind the genetic differences between African and non-African populations and has typically shaped the present day genetic variation (Cavalli-Sforza and Feldman 2003). It affects the geographic distribution of genetic variation in two ways. First the 'founder population' like the east Africans in standard model of human evolution carries with them a sub set of genetic variation found in the ancestral population (Africans). Secondly, these *'founders'* becomes more widely separated by again confining into groups corresponding to different geographical locations (like East Asians or Europeans), where mating becomes more common between same groups. Therefore, it is the process of random genetic drift, which is mainly responsible for the expansion of humans in the world (Bamshad et al. 2004). There are several reports available that have suggested that ML criterion is a better option in assessing phylogenies (Perez-Lezaun et al. 1997) Recently, Rowald and Herrera (2003) and Agrawal and Khan (2005) have categorically stated that the genetic drift component is pivotal in finding out differences or relatedness among population groups.

In the present study, we have used ML based approach along with the statistical bootstrapping method of 1000 replicates to ensure the reliabilities of branching order in the consensus tree. The ML phylogram has depicted a finely resolved basal cluster pattern with three different clades specific to Caucasians, Africans and Asian-American clad which later on bifurcated into Oriental and Native American branches. North Indian Hindus clustered with Caucasians especially with Turks, Italians and Germans in the ML tree.

Apart from revealing the phylogenetic relationship of north Indian Hindus, ML based phylogram has also revealed some interesting aspects about distribution of genetic variation among global populations. African populations depict longest branch length revealing the higher diversity. This is because Africans being the most ancestral population have maintained a larger effective population size and have had more time for mutation and recombination (Bamshad et al. 2004). The Caucasian branch showed a bifurcation into two sub clusters roughly corresponding to the western and eastern Eurasia – one carrying Turks, Germans and North Indian Hindus, while other carrying Spaniard Basques, Pasiegos and French CEPH. Semino et al. (2000) exhibited similar bifurcation between east and west European populations on the basis of differential presence of R1a and R1b NRY haplogroups. Another scintillating demons-tration was of the Asian-American branch that got bifurcated into Oriental and Native American branches which strengthen the Grenburg theory that Native Americans came from Africa through a Bering Land Bridge, 30,000- 60,000 years ago in three

waves- Amerindians, Nan-Den speakers like Athabaskans and Eskimos (Gomez-Casado et al. 2003).

Overall, the analysis of the three HLA class II loci have depicted a strong geo-ethnic phylogeny indicating presence of numerous alleles (against bi-allelic SNPs of mt-DNA and Y-chromosome) does not interfere with phylogenetic information content of the loci, provided that frequency distribution of the populations is significantly different. Instead it increases the chance of presence of signature alleles and specific haplotypes in case of closely linked loci like HLA. However, there are potential problems associated with the use of HLA loci in inferring human phylogenies. Most important of them is non-uniformity of HLA genotyping methodologies as some studies are based on low resolution typing while others on high resolution. Even the technique used varies between serology, SSP, SSOP and SBT. Another discrepancy is in the choice of loci as only few studies involve analysis of all HLA class I and class II loci.

Preferably, if all the HLA loci (both class I and class II, if not, at least all class II loci) are analyzed with high resolution of all the alleles and proper statistical interpretation based on more logistic approaches is carried out, then HLA loci can be an ideal marker to infer genetic differences between inter and intra geo-ethnic groups. Moreover, if the property of rarity of recombination at HLA loci is fully explored to assign extended haplotypes, then, there is a high possibility that HLA loci can reconstruct the human phylogenies as exactly and accurately as deciphered by mt-DNA and Y-chromosome markers.

ACKNOWLEDGEMENTS

This work was supported by Indian Council of Medical Research (ICMR) New Delhi. Authors are thankful to Sanjay Gandhi Post Graduate Institute of Medical Sciences Lucknow for providing various lab facilities and other assistance.

REFERENCES

Agrawal S, Bhatnagar S, Bhardwaj U, Khan F 2001. Distribution of HLA Class II Antigens in Three North Indian Populations. *Int J Hum Genet,* **1:** 283-29.

Agrawal S, Khan F 2005. Reconstructing Recent Human Phylogenies with forensic STR loci: A Statistical Approach. *BMC Genet,* **6:** 47.

Agrawal S, Khan F, Pandey A, Tripathi M, Herrrera RJ 2005. YAP, signature of an African-Middle Eastern migration into northern India. *Current Sci,* **88:** 174-79.

Antunez-de-Mayolo G, Antunez-de-Mayolo A, Antunez-de-Mayolo P, Papiha SS, Hammer M, Yunis JJ, Yunis EJ, Damodaran C, Martinez de Pancorbo M, Caeiro JL, Puzyrev VP, Herrera RJ 2002. Phylogenetics of worldwide human populations as determined by polymorphic Alu insertions. *Electrophoresis*, **23:** 3346-56.

Balakrishnan V 1978. A preliminary study of genetic distances among some populations of the Indian sub-continent. *J Hum Evol*, **7:** 67-75

Bamshad M, Wooding S, Salisbury BA, Stephens JC 2004. Deconstructing the relationship between genetics and race. *Nat Rev Genet,* **5:** 598-609.

Begovich AB, McClure GR, Suraj VC, Helmuth RC, Fildes N, Bugawan TL, Erlich HA, Klitz W 1992 Polymorphism, recombination, and linkage disequilibrium within the HLA class II region. *J Immunol*, **148:** 249-258.

Bugawan TL, Klitz W, Blair A, Erlich HA 2000. High-resolution HLA class I typing in the CEPH families: analysis of linkage disequilibrium among HLA loci. *Tissue Antigens,* **56:** 392-404.

Cavalli-Sforza LL, Menozzi P, Piazza A 1994. *The History and Geography of Human Genes*. Princeton: Princeton University Press.

Cavalli-Sforza LL 2005. The human genome diversity project: past, present and future. *Nat Rev Genet,* **6:** 333-40.

Cavalli-Sforza LL, Feldman MW 2003. The application of molecular genetic approaches to the study of human evolution. *Nat Genet*, **33S:** 266-75.

Comey CT, Koons BW, Presley KW 1994. DNA extraction strategies for amplified fragment length polymorphism analysis. *J Forensic Sci*, **39:** 12541

Djoulah S, Sanchez-Mazas A, Khalil I, et al. (1994) HLA DRBI, DQA1and DQB1 DNA polymorphisms in healthy Algerians and genetic relationships withy other populations. *Tissue Antigens,* **43:** 102-109.

Evseeva I, Spurkland A, Thorsby E, et al. (2002). HLA profile of three ethnic groups living in the northwestern region of Russia. *Tissue Antigens,* **59:** 38-43.

Felsenstein J 1993. *Phylogeny Inference Package (PHYLIP) Version 3.5c*/ Distributed by The author, Department of Genetics, University of Washington, Seattle.

Fort M, de Stefano GF, Cambon-Thomsen A, Giraldo-Alvarez P, Dugoujon JM, Ohayon E, Scano G, Abbal M 1998. HLA class II allele and haplotype frequencies in Ethiopian Amhara and Oromo populations. *Tissue Antigens*, **51:** 327-36.

Gomez-Casado E, del Moral P, Martinez-Laso J, Garcia-Gomez A, Allende L, Silvera-Redondo C, Longas J, Gonzalez-Hevilla M, Kandil M, Zamora J, Arnaiz-Villena A 2000. HLA genes in Arabic-speaking Moroccans: close relatedness to Berbers and Iberians. *Tissue Antigens*, **55:** 239-49.

Gomez-Casado E, Martinez-Laso J, Moscoso J, Zamora J, Martin-Villa M, Perez-Blas M, Lopez-Santalla M, Lucas Gramajo P, Silvera C, Lowy E, Arnaiz-Villena A 2003. Origin of Mayans according to HLA genes and the uniqueness of Amerindians. *Tissue Antigens*, **61:** 425-36

Grubic Z, Zunec R, Cecuk-Jelicic E, Kerhin-Brkljacic V, Kastelan A 2000. Polymorphism of HLA-A, -B, -DRB1, -DQA1 and -DQB1 haplotypes in a Croatian population. *Eur J Immunogenet*, **27:** 47.

Ingman M, Kaessmann H, Paabo S, Gyllensten U 2000. Mitochondrial genome variation and the origin of modern humans. *Nature*, **408:** 708-13.

Jobling MM, Tyler-Smith C 2003. The human Y-chromosome: an evolutionary marker comes of age. *Nat Rev Genet*, **4:** 598-612.

Jorde LB, Wooding SP 2004. Genetic variation, classification and 'race'. *Nat Genet*, **36:** S28-33.

Khan F, Phadke S, Nityanand S, Talwar S, Agrawal S 2004. Use of ApoB' hyper variable region in studying mixed chimerism and maternal contamination in North Indian populations. *J Clin Forensic Med*, **11:** 183-8.

Lulli P, Grammatico P, Brioli G, Catricala C, Morellini M, Roccella M, Mariani B, Pennesi G, Roccella F, Cappellacci S, Trabace S 1998. HLA-DR and -DQ alleles in Italian patients with melanoma. *Tissue Antigens*, **51:** 276-80.

Monsalve MV, Edin G, Devine DV 1998. Analysis of HLA class I and class II in Na-Dene and Amerindian populations from British Columbia, Canada. *Hum Immunol*, **59:** 48-55.

Palanichamy MG, Sun C, Agrawal S, Bandelt HJ, Kong QP, Khan F, Wang CY, Chaudhuri TK, Palla V, Zhang YP 2004. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. *Am J Hum Genet*, **75:** 966-78.

Perez-Lezaun A, Calafell F, Mateu E, Comas D, Bosch E, Bertranpetit J 1997. Allele frequencies for 20 microsatellites in a worldwide population survey. *Hum Hered*, **47:** 189-96.

Perez-Miranda AM, Alfonso-Sanchez MA, Pena JA, Calderon R 2003. HLA-DQA1 polymorphism in autochthonous Basques from Navarre (Spain): genetic position within European and Mediterranean scopes. *Tissue Antigens*, **61:** 465-74.

Pimtanothai N, Hurley CK, Leke R, Klitz W, Johnson AH 2001. HLA-DR and -DQ polymorphism in Cameroon. *Tissue Antigens*, **58:** 1-8.

Renfrew C 1987. *Archeology and Language: The puzzle of Indo-European Origins*. London: Jonathan Cape.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW 2002. Genetic structure of human populations. *Science*, **298:** 2381-5.

Rowold Diane J, Herrera Rene J 2003. Inferring recent human phylogenies using forensic STR technology. *Forensic Science International*, **133:** 260–265

Roychoudhury S, Roy S, Dey B, Chakraborty M, Roy M, Roy B 2000. Fundamental genomic unity of ethnic India is revealed by analysis of mitochondrial DNA. *Curr Sci*, **79:** 1182–92.

Ruhlen M 1991. TITLE In: Edward Arnold, London Sherratt A (Eds.). 1994. *A Guide to the World's Languages*. *The Cambridge Encyclopedia of Archaeology*. Stanford, California: Stanford University Press.

Sanchez-Velasco P, Gomez-Casado E, Martinez-Laso J, Moscoso J, Zamora J, Lowy E, Silvera C, Cemborain A, Leyva-Cobian F, Arnaiz-Villena A 2003. HLA alleles in isolated populations from North Spain: origin of the Basques and the ancient Iberians. *Tissue Antigens*, **61:** 384-92.

Saruhan-Direskeneli G, Uyar FA, Bakar S, Eraksoy M 2000. Molecular analysis of HLA-DRB1, -DQA1 and -DQB1 polymorphism in Turkey. *Tissue Antigens*, **55:** 171-4.

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA 2000. The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: A Y chromosome perspective. *Science*, **290:** 1155-9.

Tsuneto LT, Probst CM, Hutz MH, Salzano FM, Rodriguez-Delfin LA, Zago MA, Hill K, Hurtado AM, Ribeiro-dos-Santos AK, Petzl-Erler ML 2003. HLA class II diversity in seven Amerindian populations. Clues about the origins of the Ache. *Tissue Antigens*, **62:** 512-26.

Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet*, **65:** 43-62.