Meta Gene 10 (2016) 108–117

Contents lists available at ScienceDirect

Meta Gene

journal homepage: www.elsevier.com/locate/mgene

Population genetic analysis of five northwest Punjabi endogamous groups using microsatellite markers

Manpreet Kaur Ph.D.Research Scientist [a,*], Badaruddoza M.Phil, Ph.D.Director [b]

[a] Viral Research and Diagnostic Laboratory, Government Medical College, Amritsar, Punjab, India
[b] Aligarh Muslim University Murshidabad Centre, Murshidabad, West Bengal, India

ARTICLE INFO

Article history:
Received 14 July 2016
Revised 10 August 2016
Accepted 11 August 2016
Available online 26 August 2016

Keywords:
Population genetic analysis
Microsatellite polymorphism
Punjab
North-West India

ABSTRACT

Previous episodes of foreign migration and invasion led to massive gene flow into the population of Punjab, especially, Northwest Punjab, as invaders came through this route. Furthermore, regional conversions from different castes occurred during the period of Islamic dominance. The present study is motivated by the fact that negligible study is available regarding genetic structure of Punjabi population. In purview of this the present study was conducted to analyze the genetic structure, relationship and gene flow in five endogamous population groups, namely Jat Sikh, Mazhbi Sikh, Brahmin, Ramdasia and Muslim, from the Northwest districts of Punjab, based on microsatellite loci. A total of 751 subjects, including 154 Jat Sikh, 148 Mazhbi Sikh, 151 Brahmin, 148 Ramdasia and 150 Muslim samples were analyzed for six autosomal microsatellite markers (THO1, TPOX, CSF1PO, vWA, D7S820 and FGA). Statistical analyses were performed to interpret allele frequencies, heterozygosities, gene diversity, analysis of molecular variance (AMOVA), phylogenetic analysis (neighbor joining (NJ) method) and multidimensional scaling (MDS). The average heterozygosity for each locus in each population as a whole was substantial. Significant deviations from HWE were observed in almost all the populations and all the loci. In the entire analyses the Brahmin and the Mazhbi Sikh formed same cluster, while Jat Sikh and Ramdasia were placed close to their cluster, whereas, Muslim population group maintained the genetic distance from all the other population groups. Jat Sikh, Mazhbi Sikh, Brahmin and Ramdasia groups formed a clear and distinct separate cluster when compared with other Indian populations. The present study has also examined the affinities of Northwest Punjabi population with international populations, and some elements of central Asia, Middle East, Caucasoid and European populations were evident among the present studied groups. In conclusion, overall, a low level of genetic differentiation was observed in the studied population groups, especially, Jat Sikh, Mazhbi Sikh, Brahmin and Ramdasia indicating that genetic drift might have been small or negligible in shaping the genetic structure of the Northwest Punjabi populations.

1. Introduction

The inter-population differences have a definite pattern in Punjab which leads to having more genetic similarity within the same groups and genetically being more distant from population of other groups (Kaur and Badaruddoza, 2014). However, the previous episodes of foreign migration and invasion to this region have led to massive gene flow into the Indian subcontinent, particularly Punjab (Sekhon, 2000; Badaruddoza et al., 2008). It is assumed that a considerable genetic heterogeneity would be found in Punjabi population, especially, in North West Punjab as invaders came through this route (Tripathi et al., 2008). Therefore, Punjabi population probably experienced very high gene flow during the period of co-existence with foreign populations since 15th century or even earlier. It is also noteworthy that Punjabi population exhibited higher affinity with central Asia and European populations (Kashyap et al., 2003). Furthermore, regional conversions from different caste occurred during the period of Islamic dominance. However, negligible studies are available regarding genetic structure of Punjabi population (Badaruddoza and Brar, 2006; Badaruddoza et al., 2008; Kaur and Badaruddoza, 2014).

There has been a degree of controversy in the analysis of genetic differentiation among groups. It has been argued that race is appropriate to determine ancestry while genetic variation should determine

Abbreviations: AMOVA, Analysis of molecular variance; df, degree of freedom; D_{ST}, average gene diversity between the subpopulations; F_{CT}, the fraction of variation between the three groups; F_{SC}, the fraction of genetic variation within groups between the sub-populations; F_{ST}, the fraction of total genetic variation between sub-populations; G_{ST}, the coefficient of gene differentiation; H_{exp}, Expected Heterozygosity; H_{obs}, Observed Heterozygosity, H_S, average gene diversity within the subpopulation; H_T, gene diversity of the total population; HWE, Hardy-Weinberg equilibrium; MDS, Multidimesional Scaling; NJ, neighbor joining; PCR, Polymerase Chain Reaction; SD, Standard Deviation.

* Corresponding author at: Viral Research and Diagnostic Laboratory, Government Medical College, Amritsar, India.
E-mail address: manpreet024@gmail.com (M. Kaur).

http://dx.doi.org/10.1016/j.mgene.2016.08.007
2214-5400/

population structure. There is a need to comprehend the migration and demographic history of humans, ancestry and the relationship between races. Markers that are informative for inferring regional ancestry are also suggestive of inferences among populations in the same regions (Bamshad et al., 2004). Genetic data can be used to differentiate groups and allocate individuals into groups.

Microsatellite loci, due to its uniformity in distribution throughout the genome and high level of polymorphism, gained popularity for diverse genetic applications such as intro-group phylogenetic reconstruction, disease analysis, gene mapping and identification purposes (Jeffreys et al., 1994; Ferdous et al., 2010; Balanovsky et al., 2011). Microsatellites have been important from human evolutionary point of view since their discovery as an important form of human genetic variation. The very first microsatellite study of global human variation was done by Bowcock et al. (1991) based on thirty markers among 148 individuals from 14 indigenous populations. They discovered that population clusters on a neighbor joining tree according to its geography and the highest diversity of microsatellites was observed among the Africans (Pemberton et al., 2013). In the present time, microsatellite loci are considered as one of the most important criteria for phylogenetic analysis of the human population.

In the present study, the genetic structure, relationship and gene flow in five endogamous population groups, namely Jat Sikh, Mazhbi Sikh, Brahmin, Ramdasia and Muslim from the North West districts of Punjab have been analyzed based on six microsatellite markers (THO1, TPOX, CSF1PO, vWA, D7S820 and FGA) which were selected bearing in mind the feasibility of genotypic analysis and availability of databases for comparisons. The genetic similarity and phylogenetic position of Punjabi population with respect to admixture of other Indian and global populations in the past history has also been studied.

2. Materials and methodology

2.1. Sample collection

The present study is an attempt to explore the genetic differentiation, diversity and population structure of Punjab. The caste groups selected for the present study were Jat Sikh (higher caste of Sikh religion), Mazhbi Sikh (lower caste of Sikh religion), Brahmins (higher caste of Hindu religion) and Ramdasia/Valmiki (lower caste of Hindu religion) with their informed consents. All selected groups practice a high degree of endogamy (Sekhon, 2000). Samples were collected from healthy individuals and special care was taken to avoid sampling from related individuals. Individuals whose families had been from a same particular area for at least three generations were selected and this was recorded with the questionnaires filled by the donors.

The sample size was determined by the population frequency of the least representative allele/s. The appropriate sample size for the present study was determined by following assumptions and formulae given by Yan and Zhang (2004):

$$P_0 = (1-q)^{2n.}$$

(where, n is the sample size, q is the frequency of allele in question, P_0 is the frequency of allele being not represented in the sample)

Rare alleles (frequency < 0.05) are a common feature of microsatellite loci, whereas, very less alleles (frequency < 0.01) provide no useful information for most of the population based analysis (Hale et al., 2012). Therefore, in order to detect the allele frequency of allele with frequency < 0.01 (q) with 95% probability for the allele being present in the population ($P_0 = 0.05$), the sample size calculated was, $n = 149$. Therefore, 149 samples for each population were sufficient to get a statistical power of 90% with 95% confidence interval.

A total of 751 subjects, including 154 Jat Sikh, 148 Mazhbi Sikh, 151 Brahmin, 148 Ramdasia and 150 Muslim samples were recruited from the selected six districts of Northwest Punjab, namely, Amritsar, Fazilka, Ferozepur, Gurdaspur, Pathankot and Tarn Taran, and the Malerkotla city. 3 ml of intravenous blood sample was collected by a certified technician from each subject after obtaining written consent.

2.2. Ethical considerations

The study was ethically approved for collecting blood samples from the human subjects for carrying out the present study by Ethical Committee of Guru Nanak Dev University, Amritsar, Punjab (India).

2.3. Genotype analysis

Extraction of DNA was the first step in the molecular analysis of the collected blood samples. The extraction was carried out using phenol-chloroform (organic) method (Gill et al., 1987). After the extraction of genomic DNA all the samples were quantified using nanodrop (Thermo Scientific, Germany).

The PCR reactions were performed with initial denaturation (95 °C for 5 min) followed by 30 or 35 cycles of denaturation (95 °C for 30 s), annealing (30 s) and extension (72 °C for 30 s-1 min), followed by final extension (72 °C for 10 min). Annealing temperature was generally set two degrees below the T_m ($T_m = [2 \times n\,(A + T)] + [4 \times n\,(G + C)]$) of the primer with the lowest melting temperature and was adjusted according to specific standardization condition of each primer. The concentration of PCR reagents like dNTPs, primers and Taq polymerase were standardized for each target sequence.

The analysis here was done for six microsatellite loci: THO1, TPOX, CSF1PO, vWA, D7S820 and FGA. The specific oligonucleotide primer sequences and their respective annealing temperature used for each microsatellite locus are given in Table 1.

Table 1
Primer pairs and annealing temperature used for the amplification of selected microsatellite markers.

Locus	Primer pair	Annealing temperature	Reference
THO1	Forward: 5′-ATT CAA AGG GTA TCT GGG CTC TGG-3′ Reverse: 5′-GTG GGC TGA AAA GCT CCC GAT TAT-3′	58 °C	Brinkmann et al., 1996
TPOX	Forward: 5′-ACT GGC ACA GAA CAG GCA CTT AGG-3′ Reverse: 5′-GGA GGA ACT GGG AAC CAC ACA GGT TA-3′	58 °C	Huang et al., 1995
CSF1PO	Forward: 5′-AAC CTG AGT CTG CCA AGG ACT AGC-3′ Reverse: 5′-TTC CAC ACA CCA CTG GCC ATC TTC-3′	58 °C	Hammond et al., 1994
vWA	Forward: 5′-CCC TAG TGG ATG ATA AGA ATA ATC-3′ Reverse: 5′-GGA CAG ATG ATA AAT ACA TAG GAT GGA TGG-3′	53 °C	Moller et al., 1994
D7S820	Forward: 5′-TGT CAT AGT TTA GAA CGA ACT AAC G-3′ Reverse: 5′-CTG AGG TAT CAA AAA CTC AGA GG-3′	52 °C	Jin et al., 1997
FGA	Forward: 5′-GCC CCA TAG GTT TTG AAC TCA-3′ Reverse: 5′-TGA TTT GTC TGT AAT TGC CAG C-3′	51 °C	Urquhart et al., 1995