GENETIC LANDSCAPE OF THE PEOPLE OF BANGLADESH DEPICTED WITH 17 Y-CHROMOSOME-SPECIFIC MICROSATELLITES

Bioresearch Communications
Volume 10, Issue 2, July 2024

DOI:
doi.org/10.3329/brc.v10i2.74500

Md. Mahamud Hasan[1], Kanchan Chakma[2], Saikat Bhattacharjee[3], Mohammad Abul Hasnat[4] and Sharif Akhteruzzaman[5]*

[1]National Forensic DNA Profiling Laboratory, Dhaka Medical College Campus, Dhaka-1000, Bangladesh
[2]Department of Biochemistry and Molecular Biology, University of Chittagong, Chittagong -4331, Bangladesh
[3]Divisional DNA Screening Lab, National Forensic DNA Profiling Laboratory, MAG Osmani Medical College Hospital, Sylhet, Bangladesh
[4]Department of Biochemistry and Molecular Biology, Shahjalal University of Science and Technology, Sylhet, Bangladesh
[*5]Department of Genetic Engineering and Biotechnology, University of Dhaka, Dhaka-1000, Bangladesh

ABSTRACT

Seventeen microsatellite loci from the non-recombining region of the human Y chromosome were typed using AmpF/STR® Yfiler® PCR amplification systems in 404 male subjects belonging to the three largest ethnic groups in Bangladesh. A total of 150 haplotypes from the Chakma, 144 from the Tripura, and 110 from the Khasia were detected with a corresponding discrimination capacity of 73.885%, 65.563%, and 81.250%, respectively. The highest allele frequency of 0.828 was detected in DYS391 locus in the Tripuras, while the lowest allele frequency of 0.009 was detected at the same locus for the Chakma population. The highest gene diversity (0.964) was observed at DYS385a/b locus in the Khasias, while the lowest gene diversity (0.301) was detected at DYS391 locus in Tripuras population. The overall haplotype diversity for the studied populations was 0.986141. Both the Neighbour-Joining tree and pairwise genetic distances showed that Chakma lies closer to a clade consisting of Tripuras (Khagrachari, Bangladesh) and Tripuri (Tripura, India). In contrast, the Khasias demonstrated a close affinity with the Oraon (Chhattisgarh, India), followed by the Santals. The Y-STR haplotype matching probabilities within and between populations demonstrated that the Chakma, Tripura, and Khasia were 100% genetically distinct. The studied ethnic populations exhibited higher frequency for haplogroups L and Q as opposed to haplogroups R1a, H, and L found in the mainstream Bengali population. The Median-joining networking showed haplogroups L and R1a have the most compact clustering within populations, followed by haplogroups Q and H. The presence of haplogroup R1a suggests that Bengali may have originated through west-to-east migration, whereas haplogroups L and Q distribution in the studied tribes reveal a very significant affinity with the South-East Asian populations and may have shared a common ancestral origin with the Mongoloid stock populations.

KEYWORDS: Y-STR, Discrimination, Haplotype, Diversity, Haplogroup, Network.

RECEIVED: 09 March 2024, ACCEPTED: 18 May 2024

TYPE: Original Research

***CORRESPONDING AUTHOR: Dr. Sharif Akhteruzzaman, Department of Genetic Engineering and Biotechnology, University of Dhaka, Dhaka-1000, Bangladesh.**
Email: sazaman@du.ac.bd

Introduction

Bangladesh is a country in South Asia that shares borders with India to the east, north, and west, and with Myanmar to the southeast for a short distance. Situated at the apex of the Bay of Bengal is its southern region. Bangladesh is the 8th largest and one of the most densely populated country in the world. Together with the majority Bengali population, Bangladesh is home to nearly 40 indigenous communities, totaling 169.4 million people. Being ethnically homogeneous, the Bengalis comprise 98% of the total population of Bangladesh and belong to the Bengali ethnolinguistic group. The remaining population, which makes up 2% of the total, is primarily made up of indigenous ethnic minorities who reside in various areas of the country's mountainous zones in the northeast, center, and southeast. In this study, the Chakmas, Tripuras, Marmas, Rakhines, Garos, Khasias, and Manipuris belong to the Mongoloid stock and Sino-Tibetan in origin. The majority of the indigenous people live in Chattogram Hill Tracts, Sylhet, Mymensingh, Dinajpur, and Rajshahi regions. Based on anthropological studies, it is believed that these ethnic minorities migrated to Bangladesh from the neighboring countries over centuries. This study is intended to evaluate the genetic diversity of the people of Bangladesh with reference to Y-chromosome specific microsatellites and their utility in

population genetic research and as well as its application in forensic genetics and genetic genealogy studies.

Materials and Methods

Sample collection

Buccal swab samples were collected from randomly selected unrelated male individuals from the Chakma (n = 157), Tripura (n = 151), and Khasia (n = 112) populations with the written informed consent of the contributors and following procedures that are in accordance with the Helsinki Declaration of 1975, revised in 1983 **(**Carlson *et. al.*, 2004). This study was approved by the Ethical Review Committee, Faculty of Biological Sciences, University of Dhaka.

Extraction and Quantitation of Genomic DNA

Genomic DNA was extracted using the Chelex®-100 methods (Walsh *et. al.*,1991). The extracted DNA samples were quantified using NanoDrop™ 1000 Spectrophotometer (NanoDrop Technologies Inc., USA). Due to highly concentrated DNA obtained from saliva, about 1 µL of genomic DNA was diluted with ddH2O (1:99 µL ddH2O) prior to the spectrphotmoetric measurement.

PCR amplification of Y-STR loci and genotyping

About 1 to 2 ng of genomic DNA was amplified by polymerase chain reaction using AmpF*l*STR® Yfiler™ PCR Amplification Kit (Applied Biosystems, USA). PCR amplification was carried out using a Veriti® Thermal Cycler (Applied Biosystems, USA) following manufacturer's instructions (Mulero *et. al.*, 2006). The amplified PCR products were then separated and typed using 3500 Genetic Analyzer with POP-4 polymer and Data Collection Software v1.0 (Applied Biosystems, USA). GeneScan-600 LIZ™ was used as an internal size standard. A peak detection threshold of 100 RFU was used for allele designation using GeneMapper ID-X software v1.2, and each allele call at all Y-STR loci was determined by comparing with an allelic ladder. The forensic Y-STR alleles were designated based on the number of repeat motifs according to the recommendations for the DNA commission of the International Society of Forensic Genetics (ISFG) guidelines (Gusmão *et. al.*, 2006). Our DNA laboratory participated in the YHRD (Y-Chromosome Haplotype Reference Database) quality assurance verification test by typing blind samples for 17 Y-STR loci included in the AmpF*l*STR® Y-Filer™ Kit (certification date: September 11, 2008). The population data included in this study has been submitted to YHRD and assigned the following accession numbers: Chakma (YA004320), Tripura (YA004321), and Khasia (YA004172). The populations designated herein can be searched at www.yhrd.org by population name, contributor name, and accession number.

Statistical analysis

The number of haplotypes were counted for each population. Allele frequencies and power of discrimination at each Y-STR locus were calculated using PowerStat Microsoft Excel Workbook (Tereba *et. al.*, 1999). Haplotype discrimination capacity (DC) was calculated as $D=N_{diff}/N$, where N_{diff} is the total number of different haplotypes (not unique haplotypes) and N is the total number of haplotypes obtained in a population. The haplotype diversity (HD) was calculated as $HD = 1-\sum (x_i /N_x)^2$. Gene diversity (GD), power of matching (PM), power of discrimination (PD), and power of exclusion paternity (PE) were calculated as $GD = N (1-\sum pi^2) / (N-1)$; $PM = \sum pi^2$; $PD = 1-PM$; and $PE = 1-\sum pi^2$, respectively. Here N is the total number of haplotypes and pi is the frequency of the ith haplotype (Nei *et. al*, 1987).

Diversity parameters, including the minimum diversity within the population (dw_{min}), the maximum diversity between two populations (db_{max}), the maximum matching probability within the population (mw_{max}), the minimum matching probability within the population (mw_{min}), and the minimum matching probability between two populations (mb_{min}) were calculated for the Bangladeshi populations as described previously (Brinkmann *et. al*., 1999, Pfeiffer *et. al*., 1999, Zupanic *et. al*., 2004, Lacau *et. al*., 2011, Gayden et. al., 2012). The ratio of mw_{max}/mb_{min} gives an estimate of the upper limits of the probability to find a match within a population rather than between two populations. Conversely, the ratio of mw_{min}/mb_{min} then yields a lower estimate of how much more probable it is to obtain a match within a population than between two populations. Regular haplotype diversity dw_{min} within a population, i.e. the probability of obtaining different haplotypes when sampling two individuals, was calculated as-

$$dw_{min} = 1-\sum_{i=1}^{h} (x_i / N_x)^2$$

where x_i is the absolute frequency of the i-th haplotype and h is the total number of different haplotypes in the Nx samples. While comparing between populations, the formula for the probability db_{max} of obtaining different haplotypes when sampling two individuals from two different populations simplifies to-

$$db_{max} = 1-\sum_{i=1}^{m} (x_i y_i) / (N_x N_y)$$

where only the m sequence matches between populations X and Y need to be considered. Both formulas assume that the absolute frequency of each haplotype in a sample reflects its relative frequency in the population, which may be an overestimate in small, highly diverse samples. Thus, the ratio $(1–dw_{min})/(1– db_{max})$ or its equivalent expression mw_{max} / mb_{min} is an overestimate of the probability ratio of obtaining a match when sampling a pair of individuals from the same population group versus from two different populations. To obtain a lower bound for this interpopulation comparison, we assume that at the extreme, all unique haplotypes in the sample reflect close to zero relative frequency. Then the within-population matching pair probability mw_{min} equals.

$$mw_{min} = \sum_{i=1;\, x_i>1}^{h} [x_i!/2(x_i-2)!] / [N_x!/2(N_x-2)!]$$

To examine and portray the genetic affinities between neighboring populations, pair-wise genetic distances (R_{st}) and associated probability values (p-values, 10,000 permutations) were calculated using the AMOVA tools of YHRD. A total of 8,677 haplotypes from 32 reference populations were included in the analysis. To portray the genetic relationship between

populations, N-J tree was constructed using MEGA version 11.0.13. To demonstrate the relationship between populations, a Principal Coordinate Analysis (PCoA) plot was generated using GenAlEx v6.51b2. Besides, a total of 1,890 samples with 1,424 unique haplotypes data were used for haplogroup distribution in the studied populations, including Bengali, Rakhine, Marma, Hajong, Manipuri, Garo, and Santal (Hasan *et. al*., 2015, Hasan *et. al*., 2018). The haplogroups were predicted by Whit Athey's Haplogroup Predictor v5 (Athey *et. al*., 2005, Athey *et. al*., 2006). The median-joining network were constructed with the assistance of the NETWORK v5.0.0.3 and NETWORK Publisher v1.2.0.0 program using the reduced median algorithm (Bandelt *et. al*., 1999). The selected haplotype had a definite certainty of which haplogroups they belonged to, which was determined by this predictor method. Moreover, haplotypes that had a bi-allelic marker (DYS385a/b), microvariants, or null alleles were excluded in this network analysis. The original population dataset contained, on average, 150 unique haplotypes per population. To obtain a transparent network, about 10 most frequent haplotypes from each population were selected to be analyzed. In this study, major haplogroups were selected for haplotype network analysis, while minors were excluded. Simultaneously, mutations between haplotypes were estimated by network analysis of a cluster in a haplogroup.

Results and Discussion

A total of 150 haplotypes were identified in the Chakma males, of which 94 were unique (62.666%). In this population, 14, 5, 2, and 1 haplotype were shared in 2, 3, 4, and 5 individuals, respectively. In total, there were 116 different haplotypes among 157 individuals, which corresponded to a discrimination capacity of 73.885%. In the Tripuras, 144 haplotypes were observed, of which 77 were unique (53.472%), where 12, 4, 3, 2, and 1 haplotype were shared in 2, 3, 4, 6, and 7 individuals, respectively. In total, there were 99 different haplotypes among 151 individuals, which corresponded to a discrimination capacity of 65.563%. In contrast, a total of 110 haplotypes were examined in the Khasias, of which 78 were unique (70.909%), where 10, 2, and 1 haplotype were shared in 2, 3, and 6 individuals, respectively. In general, there were 91 different haplotypes among 112 individuals in the Khasias, which correspond to a discrimination capacity of 81.250% (**Tables S1-S3**). Haplotype frequency, allele frequency, genotype frequency, and gene diversity are presented in **Table S4**. The highest allele frequency (0.828) was monitored in the Tripuras whereas the lowest allele frequency (0.009) was observed in the Chakmas. The most informative locus DYS385a/b showed the highest degree of diversity (above 88%), whereas the lowest diversity values were found in DYS391 (below 40%), the least informative locus, for all three groups. The maximum gene diversity (0.964) was observed at DYS385a/b in the Khasias whereas the minimum gene diversity (0.301) was calculated at DYS391 in the Tripuras. In contrast, less genetic diversity has been observed in the Manipuri population as compared to other populations. The forensic statistical parameters of the Bengali, Rakhine, Marma, Hajong, and Manipuri populations were reported in the previous study (Hasan *et. al*., 2018).

We also searched 1,587 haplotypes in the studied populations and the YHRD database, which currently includes 205,059 haplotypes from over 932 populations (release version R59; November 01, 2018). The most frequent 2 haplotypes (Bn98: 16-13-24-30-17-15-12/12-14-10-11-19-11-12-14-11-19 and Bn316: 15-14-25-32-16-15-11/14-13-10-10-23-11-13-14-11-20) shared by 3 individuals in the Bengalis (0.899%). Interestingly, 5 haplotypes from the Bengali individuals were shared with 6 individuals in four ethnic groups, i.e. 2 haplotypes with the Rakhine (Bn138 was identical to Rk22, and Bn485 was identical to Rk106); 2 haplotypes with the Marma (Bn138 was identical to Mr72, and Bn237 was identical to Mr88). Furthermore, 1 haplotype shared with the Santals (Bn265 was identical to Sn87), and 1 haplotype shared with the Garos (Bn365 was identical to Gr83). Moreover, out of the 667 haplotypes, haplotype Bn11 was identical to the Tripuris (Ghosh *et. al*., 2011); 3 haplotypes (Bn307, Bn371, and Bn476) were identical to the Pakistani samples (Ghosh *et. al*., 2014); 2 haplotypes (Bn86 and Bn610) were identical to the Tamils (Tamil Nadu, India) (Balamurugan *et. al*., 2010); 1 haplotype (Bn96) was identical to the Pathans (Pakistan) (Lee *et. al*., 2014); 1 haplotype (Bn371) was identical to the Iranian samples (Tabrizi *et. al*., 2014); and 2 haplotypes (Bn314 and Bn476) were identical to the Croatian samples (Pokupčić *et. al*., 2008). The most frequent haplotype (Ck3: 16-13-24-29-16-17-15/20-14-10-12-22-14-13-14-10-18) was shared in 5 individuals in the Chakmas (3.333%). Amusingly, among the studied population, 4 haplotypes from Chakma individuals were shared with 5 individuals from two tribal groups i.e. 3 Y-chromosomes with the Rakhine (Ck14 was identical to Rk101; Ck16 was identical to Rk66; and Ck54 was identical to Rk12); 2 with the Marmas (Ck16 was identical to Mr10; and Ck115 was identical to Mr70) (Hasan *et. al*., 2018). Also, haplotype Ck22 remained found in ten matches out of 205,059 haplotypes in the global populations. In the Tripuras, 1 haplotype (Tr37: 15-12-23-28-17-14-13/18-12-10-12-20-14-12-14-11-20) was shared by 7 individuals within this population (4.861%). In addition, another two most frequent haplotypes (Tr18: 14-14-23-30-17-14-14/19-13-10-12-23-11-11-14-9-21 and Tr22: 15-12-24-27-18-14-13/18-12-10-12-21-14-11-15-11-20) were shared in 6 individuals in the Tripuras (8.333%), respectively. Interestingly, only one haplotype from Tripuras (Tr57) was shared with one male in the Rakhines (Rk96). Moreover, haplotype Tr34 remained found 41 matches in the global populations. Besides, 5 haplotypes from the Tripuras (Tr1, Tr22, Tr29, Tr41, and Tr50) were identical to the Tripuris (Tripura, India)[17]. In the Khasias, 1 haplotype (Kh16: 15-13-25-29-17-15-15/19-14-10-11-21-13-10-14-10-18) was shared by 6 individuals (5.455%). Interestingly, 1 haplotype from Khasia (Kh11) was shared with 1 male in the Marmas (Mr87). In addition, haplotype Kh11 was shared between 8 males in global populations. Furthermore, haplotype Kh11 was found to match one individual of the Pakistani samples (Perveen *et. al*., 2014).

The Y-chromosome STR matching probabilities within and between Bangladeshi populations are represented in **Table S5**. Consistent with the haplotype diversity values, the highest maximum probability of finding a match within the population (mw_{max}) was monitored in the Manipuris (0.018239), followed by Tripuras (0.015914) and Marmas (0.014808). In contrast, the lowest maximum probability within the population was monitored in the Bengalis (0.001589). When compared among the studied populations, the maximum diversity (db_{max}) of obtaining two different haplotypes when sampling a pair of

individuals. In this study, there were observed twenty-eight pairs of individuals with the maximum diversity values i.e. from Bengali and Chakma (100%), Bengali and Tripura (100%), Bengali and Rakhine (99.991%), Bengali and Marma (99.996%), Bengali and Hajong (100%), Bengali and Manipuri (100%), Bengali and Khasia (100%), Chakma and Tripura (100%), Chakma and Rakhine (99.981%), Chakma and Marma (99.980%), Chakma and Hajong (100%), Chakma and Manipuri (100%), Chakma and Khasia (100%), Tripura and Rakhine (99.995%), Tripura and Marma (100%), Tripura and Hajong (100%), Tripura and Manipuri (100%), Tripura and Khasia (100%), Rakhine and Marma (99.890%), Rakhine and Hajong (99.987%), Rakhine and Manipuri (100%), Rakhine and Khasia (100%), Marma and Hajong (100%), Marma and Manipuri (100%), Marma and Khasia (99.993%), Hajong and Manipuri (100%), Hajong and Khasia (100%), and Manipuri and Khasia (100%). These values are supported by the observed haplotype sharing between populations in the above-mentioned pairs. The elevated db_{max} values for each population pair indicate genetic uniqueness not only within the Bengali population but also among them. The highest haplotype sharing occurs between Rakhine and Marma, which have seven different 17 Y-STRs haplotypes in common. On the contrary, Tripura, Hajong, and Khasia are more unique, as these populations share only one distinct haplotype with Rakhine and Marma, respectively. Interestingly, the Manipuris does not share any haplotypes with the other seven studied populations. As suggested by the ratio mw_{max}/mb_{min}, it is more probable to find a match within each studied population than between any two of these populations. Particularly, it is about 454 times more likely to find a match within Marma, followed by 332 times within Tripura and 283 times within Rakhine, than between either of these populations and the rest of the studied populations. These particular ratios are also prominent at the lower estimate of the parameter (mw_{min}/mb_{min}) and these values are about 234, 189, and 144 times for Marma, Tripura, and Rakhine, respectively. These ratios designate a greater degree of genetic heterogeneity and distinctiveness in Bengali, Hajong, and Manipuri than in Chakma, Tripura, Rakhine, Marma, or Khasia. The high power of discrimination emphasizes the genetic heterogeneity and uniqueness of the studied populations and the need for independent databases for forensic applications and studies in population genetics.

The pair-wise R_{st} genetic distances and associated *p*-values for comparisons of the studied populations and 29 reference populations are presented in **Table S6**. The comparative analysis indicated that the Bengali population was statistically significantly different from Khasia (R_{st}=0.1064, p=0.0000), Chakma (R_{st}=0.1994, p=0.0000), and Tripura (R_{st}=0.2485, p=0.0000), respectively. Phylogenetic and molecular evolutionary analysis of Nei's genetic distances was performed using a neighbor-joining tree for the three studied populations under research with those of the seven Bangladeshi ethnicities and 29 reference populations available in the YHRD database to understand genetic affinities, if any (**Fig. 1**). The results exhibited that the Chakmas lies closer to a clade consisting of Tripura, Hajong, and Marma from Bangladesh; Tripuri from India. On the other hand, the Khasias are closely related to the Santal from Bangladesh, followed by the Oraon and Munda populations from India. The Principal Coordinate Analysis (PCoA) of pairwise genetic distance values revealed groups of genetically related populations (**Fig. S1**), same as the phylogenetic tree mentioned above. A previous study of Y-chromosomal STRs also revealed a difference in distribution between Bengali and the other major four tribal populations (Rakhine, Marma, Hajong, and Manipuri) in Bangladesh (Hasan *et. al*., 2018).

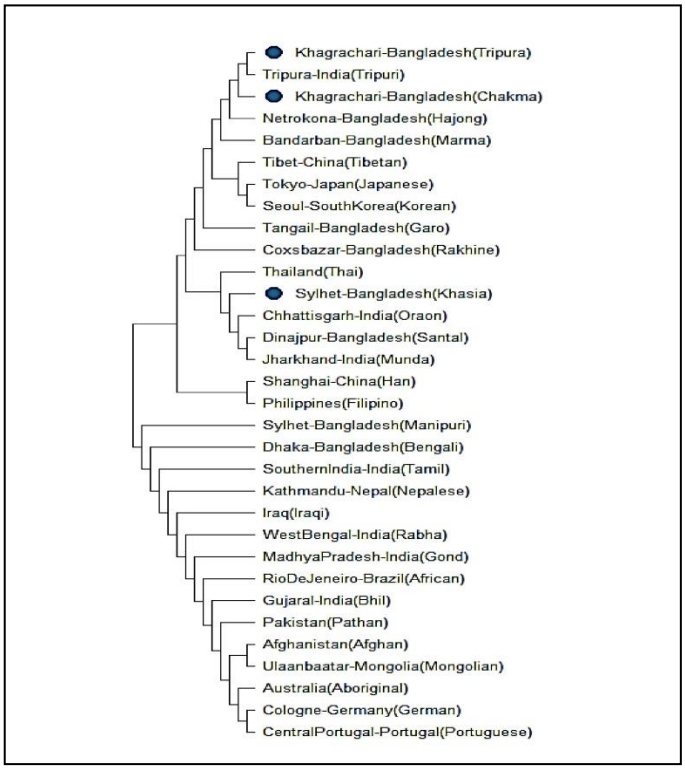

Fig 1. The Neighbor-Joining tree constructed using the Nei's genetic distance on a same set of 17 Y-STR loci in the Bangladeshi population and other 29 populations.

The Y-chromosome haplogroup frequency distributions in the studied populations, including Bangladeshi populations as well as Asian populations are shown in **Fig. 2** and **Table S7**. According to the Whit Athey's Predictor algorithm, Bangladeshis consisted of twenty haplogroups; four of them are accounted for as major haplogroups, namely R1a, L, H, and Q. The least abundant haplogroups are assigned as minor haplogroups, namely J1, J2a1xJ2a1-bh, T, J2b, J2a1b, E1b1b, G2a, I2b1, E1b1a, I2a1, I2a(xI2a1), J2a1h, N, R1b, I1, and I2b(xI2b1). The percentage distribution of haplogroups in Chakmas showed a total of two most frequent (percentage >10%) haplogroups: L (52.586%) and Q (13.793%), contributing in total to 66.379%. Haplogroups J1, H, R1a, G2a, I1, I2a(xI2a1), N, and T were monitored in 11, 5, 5, 4, 3, 3, 2, and 2 individuals, respectively. Besides, haplogroups E1b1a, E1b1b, I2a1, and J2a1h were represented by only one individual. Interestingly, haplogroups I2b(xI2b1), I2b1, J2a1b, J2a1xJ2a1-bh, J2b, and R1b were not detected in the Chakmas. The Tripuras also showed two most frequent (percentage >10%) haplogroups; the most prominent haplogroup appears to be L, which accounts for a total of 56.566%, while haplogroup J1 represents the second most abundant haplogroups, accounting for 14.141%. Haplogroups E1b1b, J2a1xJ2a1-bh, Q, H, and I2a1(xI2a1) were observed in 8, 8, 4, 3, and 3 individuals, respectively. Besides, haplogroups G2a, J2a1b, and T were represented by only one individual. Furthermore, haplogroups E1b1a, I1, I2a1, I2b(xI2b1), I2b1, J2a1h, J2b, N, R1a, and R1b were not detected in the Tripura population. In the Khasias, the analyzed male samples were categorized into twelve different haplogroups and the most abundant haplogroups were L (25.275%), T (17.582%), and Q (15.385%). Furthermore, haplogroups L, T, and Q were found on 23, 16, and 14 of the total 91 Y-chromosomes. The remaining nine haplogroups seemed to exist to a smaller extent in the following method: R1a (10.989%), I2b1 (9.890%), G2a (5.495%), J2a1b (4.496%), J2a1 x J2a1-bh (3.297%), H (2.198%), J1 (2.198%), J2a1h (2.198%), and N (1.099%). Haplogroups R1a, I2b1, G2a, J2a1b, J2a1 xJ2a1-bh, H, J1, and J2a1h were detected in 10, 9, 5, 3, 3, 2, 2, and 2 individuals, respectively. Also, haplogroup N was represented by only one individual. Interestingly, haplogroups E1b1a, E1b1b, I1, I2a (xI2a1), I2a1, I2b (xI2b1), J2b, and R1b were not detected. The distribution of haplogroups among the tribal populations is relatively uniform except for the Santals. The Chakma haplogroup percentages also correspond with Tripura, Rakhine, Marma, Hajong, Manipuri, Khasia, and Garo haplogroup structures. The Y-chromosomes of Chakma, Tripura, and Khasias mainly belong to the L, Q, J1, and T haplogroups, respectively, while the Santal, a proto-Australoid stock, belongs to the Q, T, and H haplogroups.

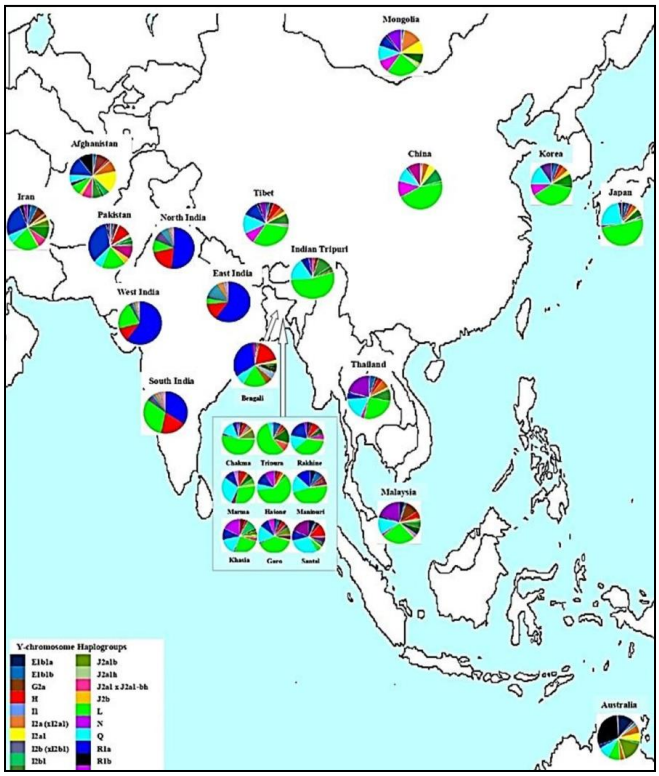

Fig 2. The frequency of Y-chromosomal haplogroups in Asian and Australian populations. The Y-STR haplotype data were used from the current published studies.

We also investigated the possible origins of the major haplogroups R1a, L, Q, and H that were obtained from the Bangladeshi populations. R1a, H, and L were the major haplogroups with a frequency of 51.5, 16.2, and 15.8% present throughout India, respectively, and accounted for more than three-fourths of the Y lineages (Singh *et. al.*, 2018). R1a, L, H, and J2a1xJ2a1-bh were the most common haplogroups found in Pakistan and accounted for 34.307, 17.518, 9.854, and

9.489% of Pakistani Y-chromosomes, respectively (Perveen *et. al*., 2014). I2a1, R1a, R1b, I2a(xI2a1), and L were the most frequent haplogroups observed in Afghanistan and accounted for 15.924, 14.013, 10.191, 8.280, and 7.643% of Afghani Y-chromosomes (Älgenäs *et. al*., 2014). In addition, R1a, L, and J2a1b were the most common haplogroups detected in the Iranian population and accounted for 24.710, 21.236, and 9.653% of Iranian Y-chromosomes (Tabrizi *et. al.*, 2015). R1a, H, and L were the most prominent Bengali male haplogroups, and these were shared with Indian, Pakistani, Afghan, and Iranian populations. Also, these haplogroups had predominately West Asian ancestry. It is suggesting that R1a, including H and L haplogroups, presence in Bengali may originate from demic diffusion by way of west-to-east migration. Interestingly, haplogroup I2b(xI2b1) belongs to one individual in Bengalis, whereas it was not found in the studied ethnic populations. Haplogroup T was also found at low frequency in Bengalis, whereas it was found in Khasia, Santal, Tibetan, Mongolian, Chinese, Thai, and Malaysian populations at moderate frequencies. Besides, we found the moderate frequency of J2a1b to be characteristic of Bengali, Indian, Iranian, and Australian populations, but its distribution elsewhere in the Northeastern and Southeastern Asian populations was very insignificant.

On the whole, we found the studied ethnic population groups of Bangladesh to be characterized by male haplogroup homogeneity, showing mostly expansions of haplogroups L, Q, and R1a. L and Q were the most frequent Y-chromosome haplogroups found in the studied ethnic populations, and these were shared primarily with populations in eastern Asia, including Indian Tripuri, Tibetan, Chinese, Mongolian, Thai, Malaysian, Korean, and Japanese (Chang *et. al*., 2007, Kang *et. al*., 2007, Mizuno *et. al*., 2008, Kim *et. al*., 2010, Taylor *et. al*., 2012, Bing *et. al*., 2013, Fu *et. al*., 2016). The Northeastern and Southeastern Asian networks and diversity analyses of haplogroups L and Q may support an eastern origin that migrated into Bangladesh via the eastern route. Haplogroups L and Q were found to be widely distributed across both Northeastern and Southeastern Asia. In addition, most of the South-East Asian population groups seemed to have L, Q, T, and J1 chromosomes, whereas haplogroups R1a and H were not present. Haplogroups T, J1, and H were present in the studied ethnic populations at moderate frequencies, suggesting that these populations may have shared a common origin with Mongoloid stock populations in the Southeast Asian regions. Haplogroups I2a(xI2a1) and I2a1 were found at extremely low frequencies in the studied ethnic populations, whereas these haplogroups were present at high frequencies in the Mongolian males.

A number of haplotypes of the studied populations, including reference populations in Bangladesh, originated from Whit Athey's Predictor algorithm, and the most common haplotypes were assigned for network analysis in haplogroups: 83 haplotypes for R1a, 97 haplotypes for L, 89 for Q, and 54 for H. The clusters of haplogroups L and Q showed the most proper organization. The L predicted haplotypes have the highest compactness in their clusters, which is to be expected as this is the most predicted haplogroup within the Bangladeshi populations, and there are three minor branches that indicate a certain degree of differences between these haplotypes. The majority of Y-chromosome haplotypes are connected by one- or two-step mutation events (**Fig. S2**). The haplogroup Q network exhibited the second highest compactness and also exhibited two distinct clusters of haplotypes with considerable haplotype sharing among the populations of Bangladesh. On the other hand, R1a and H haplogroups in the network analysis tree do not have as compact clusters as observed in the cases of L and Q, respectively. The tribal populations from Bangladesh, except Bengali, branched out of the network with more than three mutation steps.

Based on the result of the Y-chromosomal haplogroup prediction analysis, the Bengali population contains paternal lineages from India, Pakistan, Afghanistan, and Iran in West Asia. In contrast, the studied ethnic populations contain paternal lineages from the northern and southern areas of East Asia.

Acknowledgements

This research was supported with by the Multi-Sectoral Program on Violence Against Women (MSPVAW), Ministry of Women and Children Affairs, Government of the People's Republic of Bangladesh. We would also like to extend our thanks to all the contributors who donated their buccal swab sample for this study.

Conflict of interest

The authors declare that they have no conflict of interest.

Compliance with ethical standards

Buccal swab samples were collected from randomly selected unrelated male individuals from Chakma, Tripura, and Khasia populations with written informed consent following procedures that are in accordance with the Helsinki Declaration of 1975, revised in 1983. This study was approved by the Ethical Review Committee, Faculty of Biological Sciences, University of Dhaka, Bangladesh.

References

1. Älgenäs C, Tillmar AO. Population genetics of 29 autosomal STRs and 17 Y-chromosomal STRs in a population sample from Afghanistan. *International Journal of Legal Medicine*., 2014; 128(2):279-280.
2. Athey TW. Haplogroup prediction from Y-STR values using an allele-frequency approach. *Journal of Genetic Genealogy*, 2005; 1:1-7.
3. Athey TW. Haplogroup prediction from Y-STR values using a Bayesian-allele frequency approach. *J. Genetic Genealogy*, 2006; 2:34-39.
4. Balamurugan, K., Suhasini, G., Vijaya, M., Kanthimathi, S., Mullins, N., Tracey, M. & Duncan, G. 2010. Y chromosome STR allelic and haplotype diversity in five ethnic Tamil populations from Tamil Nadu, India. *Legal Medicine*. 12, 265-269.
5. Bandelt, H.J., Forster, P. & Röhl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology Evolution* 16(1), 37-48.
6. Bing, L., Liang, W., Pi, J., Zhang, D., Yong, D., Luo, H., Zhang, L. & Lin, Z. 2013. Population genetics for 17 Y-STR loci (AmpFISTR® Y-filer™) in Luzhou Han ethnic

group. *Forensic Science. International: Genetics*. 7(2), 23-26.

7. Brinkmann, C., Forster, P., Schürenkamp, M., Horst, J., Rolf, B, & Brinkmann, B. 1999. Human Y-chromosomal STR haplotypes in a Kurdish population sample. *International Journal of Legal Medicine*. 112, 181-183.
8. Carlson, R.V., Boyd, K.M., & Webb, D.J. 2004. The revision of the Declaration of Helsinki: past, present and future. *British Journal of Clinical Pharmacology*. 57, 695-713.
9. Chang, Y.M., Perumal, R., Ke at, P.Y. & Kuehn, D.L. 2007. Haplotype diversity of 16 Y-chromosomal STRs in three main ethnic populations (Malays, Chinese and Indians) in Malaysia. *Forensic Science International*. 167(1), 70-76.
10. Fu, X., Fu, Y., Liu, Y., Guo, J., Liu, Y., Guo, Y., Yan, J., Cai, J., Liu, J. & Zha, L. 2016. Genetic polymorphisms of 26 Y-STR loci in the Mongolian minority from Horqin district, China. *International Journal of Legal Medicine* 130 (4), 941-946.
11. Gayden, T., Bukhari, A., Chennakrishnaiah, S., Stojkovic, O. & Herrera, R. J. 2012. Y-chromosomal microsatellite diversity in three culturally defined regions of historical Tibet. *Forensic Science International: Genetics*. 6, 437-446.
12. Ghosh, T., Kalpana, D., Mukerjee, S., Mukherjee, M., Sharma, A.K., Nath, S., Rathod, V. R., Thakar, M.K. & Jha, G. N. 2011. Genetic diversity of 17 Y-short tandem repeats in Indian population. *Forensic Science International Genetics*. 5(4), 363-367.
13. Gusmão, L., Butler, J.M., Carracedo, A., Gill, A.P., Kayser, M., Mayr, W.R., Morling, N., Prinz, M., Roewer, L., Tyler-Smith. C. & Schneider, P. M. 2006. DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendation on the use of Y-STRs in forensic analysis. *Forensic Science International*. 157, 187-197.
14. Hasan, M.M., Momtaz, P., Hosen, I., Das, S.A. & Akhteruzzaman, S. 2015. Population genetics of 17 Y-chromosomal STRs loci in Garo and Santal tribal populations in Bangladesh. *International Journal of Legal Medicine*. 129, 251-252.
15. Hasan, M.M., Sufian, A., Momtaz, P., Mazumder, A.K., Khondaker, J.A., Bhattacharjee, S., Chakma, K. & Akhteruzzaman, S. 2018. Phylogenetic analysis and forensic evaluation among Rakhine, Marma, Hajong, and Manipuri tribes from four culturally defined regions of Bangladesh using 17 Y-chromosomal STRs. *International Journal of Legal Medicine*. 132(6), 1641-1644.
16. Kang, LL., Liu, K. & Ma, Y. 2007. Y chromosome STR haplotypes of Tibetan Living Tibet Lassa. *Forensic Science International*. 172(1), 79-83.
17. Kim, S.H., Han, M.S., Kim, W. & Kim, W. 2010. Y-chromosome homogeneity in the Korean population. *International Journal of Legal Medicine*. 124(6), 653-657.
18. Lacau, H., Bukhari, A., Gayden, T., Salvia, J, L., Regueiro, M., Stojkovic, O. & Herrera, R.J. 2011. Y-STR profiling in two Afghanistan populations. *Legal Medicine*. 13, 103-108.
19. Lee, E.Y., Shin, K.J., Rakha, A., Sim, J.E., Park, M.J., Kim, N.Y., Yang, W.I. & Lee, H.Y. 2014. Analysis of 22 Y chromosomal STR haplotypes and Y haplogroup distribution in Pathans of Pakistan. *Forensic Science International: Genetics*. 11, 111-116.
20. Li, L., Xu, Y., Luis, J.R., Alfonso-Sanchez, M.A., Zeng, Z., Garcia-Bertrand, R. & Herrera, R.J. 2019. Cebú, Thailand and Taiwanese aboriginal populations according to Y-STR loci. *Gene*. 721, (Supplementary issue): 100001.
21. Mizuno, N., Nakahara, H., Sekiguchi, K., Yoshida, K., Nakano, M. & Kasai, K. 2008. 16 Y-chromosomal STR haplotypes in Japanese. *Forensic Science International*. 174(1), 71-76.
22. Mulero, J.J., Chang, C.W., Calandro, M., Green, R.L., Li, Y., Johnson, C.L. & Hennessy, L.K. 2006. Development and validation of the AmpF*l*STR®Yfiler PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system. *Journal of Forensic Science*. 51, 64-75.
23. Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press. New York, pp, 176-179.
24. Perveen, R., Rahman, Z., Shahzad, M.S., Israr, M., Shafique, M., Shana, M.A., Zar, M.S., Iqbal, M. & Husnain, 2014. T. Y-STR haplotype diversity in Punjabi population of Pakistan. *Forensic Science International: Genetics*. 9, 20-21.
25. Pfeiffer, H., Brinkmann, B., Hühne, J., Rolf, B., Morris, A.A., Steighner, R., Holland, M.M. & Forster, P. 1999. Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. *International Journal of Legal Medicine*. 112, 291-298.
26. Pokupčić, K., Cukrov, S., Klaric´, I.M., Salihovic´, M.P., Lauc, L.B., Blazˇanovic´, A. & Janic´ijevic´, B. 2008. Y-STR genetic diversity of Croatian (Bayash) Roma. *Forensic Science International: Genetics*. 2,11-13.
27. Singh, M., Sarkar, A. & Nandineni, M.R. 2018. A comprehensive portrait of Y-STR diversity of Indian populations and comparison with 129 worldwide populations. *Scientific Reports*. 8, 15421.
28. Tabrizi, A.A., Hedjazi, A., Hedjazi, A., Kerachian, M.A., Honarvar, Z., Dadgarmoghaddam, M. & Raoofian, R. 2015. Genetic profile of 17 Y-chromosome STR haplotypes in East of Iran. *Forensic Science International: Genetics*. 14, 6-7.
29. Taylor, D.A. & Henry, J.M. 2012. Haplotype data for 16 Y-chromosome STR loci in Aboriginal and Caucasian populations in South Australia. *Forensic Science International: Genetics*. 6(6), 187-188.
30. Tereba, A. Tools for analysis of population statistics, Profiles in DNA. Promega Corporation. http://www.promega.com/geneticidtools/powerstats.
31. Walsh, P.S., Metzger, D.A. & Higuchi, R. 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques*. 10, 506-518.
32. Zupanic, P.I., Balazic, J. & Komel, R. 2004. Sequence polymorphism of the mitochondrial DNA control region in the Slovenian population. *International Journal of Legal Medicine*. 118(1), 1-4.