

Mitochondrial DNA Control Region Variation from Bangladesh: Sequence Analysis for the Establishment of a Forensic Database

Gazi Nurun Nahar Sultana^{1*}, Jannatul Ferdoush Tuli², Rokeya Begum¹, Rakesh Tamang³

¹Centre for Advanced Research in Sciences, University of Dhaka, Dhaka, Bangladesh

²Department of Genetic Engineering and Biotechnology, University of Dhaka, Dhaka, Bangladesh

³University of Kolkata, Kolkata, India

Email: *nngazi@gmail.com

Received 28 August 2014; revised 30 September 2014; accepted 20 October 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Mitochondrial DNA sequences of the entire control region and three coding regions were analyzed in 108 unrelated individuals from three regions of Bangladesh. Sequence evaluation was performed with validated primers and combined sequence comparison led to the identification of 14 different haplotypes characterized by 37 variable polymorphic sites. The Bangladeshi sequences exhibited high variations and low random match probability, indicating for forensic application. The mean pairwise difference between individual was 9.698 ± 1.8658 nucleotides (95% CI 9.67 - 9.69), compared to a mean pairwise difference of 9.890 ± 4.189 nucleotides reported from North-east Asia and suggested significant differences in the mtDNA composition of the various populations. The sequence diversity of 108 Bangladeshi Bengali samples ($n = 216$ chromosomes) was estimated to be 0.8475 ± 0.13406 . This study first time reports that the comparison of closely related mtDNA sequences can be very useful for improving mtDNA database quality, as well as provide haplotype information for forensic study in mainstream population of Bangladesh.

Keywords

Forensic Database, Mainstream Population, mtDNA Control Region, Coding Region, Haplotype, Bangladesh

*Corresponding author.

1. Introduction

Sequence analysis of human mitochondrial DNA (mtDNA) has been demonstrated to be a valid and reliable tool for genetic characterization of an individual [1]. Typically, mtDNA typing is attempted on samples for which nuclear DNA is not likely to be successful, such as hair, shaft, bones, teeth and other samples that are severely decomposed or have been exposed to substantial environmental abuse. Mitochondrial DNA is less prone to degradation than nuclear DNA because of the circular nature of mtDNA genome and its sub cellular sequestration [2] [3]. Although mtDNA comprises less than 1% of total cellular DNA, the high copy number of mtDNA per cell enables successful typing of samples containing only trace amount of amplifiable nuclear DNA [4]. In the forensic community, the sequence variations used for identification are localized to the D-loop region because this region appears to have a higher rate of mutation.

The noncoding, displacement (D-loop) loop is approximately 1100 bp segment situated between the mitochondrial tRNA^{pro} and the tRNA^{phe} genes and contains two hypervariable regions, HVR-I and HVR-II. mtDNA mutations also evolve five to ten times faster than chromosomal DNA, and this relatively higher mutation rate gives rise to more polymorphic sites [5]. The hypervariable regions can be sequenced to provide a high degree of information for discriminating between unrelated individuals of Bangladesh. Simultaneous construction of haplogroups is the most common approach in the forensic field and coding region information is indispensable for phylogenetic analysis of mtDNA [6]. We applied direct sequencing in this study samples. The most commonly observed haplogroup was M5, M3, M7, R, A4 and U2b in main stream population of Bangladesh.

2. Materials and Methods

2.1. Population

About 3 - 5 mL of blood samples of 108 healthy unrelated anonymous individuals belonging to three districts of Bangladesh was collected. Genealogical information of sampled individual was recorded for at least 2 previous generations. The research aim of the project was explained in Bangla (local language) to each individual and informed consent forms in both Bangla (local language) and English were signed. The project was approved by the ethical committee of University of Dhaka, Bangladesh. Genomic DNA was extracted by the standard phenol-chloroform method [7].

2.2. Amplification of mtDNA

The hyper-variable regions (HVR-1) and (HVR-II) and 3 coding regions of mtDNA were amplified from 10 ng of template DNA using 10 pM of each primer, 100 μM dNTPs, 1.5 mM MgCl₂ and 1 U of *Taq* polymerase. Generally, 35 cycles of reaction was performed with optimized protocol of PCR in addition with positive and negative control. Annealing temperature and the time were slightly modified for few sets of primers. The reactions were carried out in Bio-Rad C1000 thermal cycler.

2.3. Sequencing of the PCR Products

PCR products were checked by 1.8% agarose gel electrophoresis and directly sequenced using BigDye™ Terminator cycle sequence kit v3.1 (Applied Biosystems) in ABI Prism 3130 DNA Genetic Analyzer following manufacture's protocol. The mtDNA sequences were aligned and compared with the revised Cambridge Reference Sequence (rCRS) [8] using the Autoassembler software v 1.4.0. Mitochondrial haplogroups were assigned to all samples according to [9] [10].

2.4. Data Analysis

All sequences were compared to the reference standard described by Anderson *et al.* [11]. Sequence and nucleotide diversity were calculated using the Arlequin package [12]. All sequences were aligned and trimmed to a greatest common range of nt 16024-16565 and nt 73-340 as well as 3 diagnostic coding regions.

3. Results

In this study, presentation of complete sequences of HVR-I and HVR-II regions and diagnostic coding regions

of all 108 individuals are most informative display of data for the first time in mainstream population of Bangladesh. The low value of sequence diversity (0.8475 ± 0.0134) and mean number of pairwise differences (9.698 ± 1.8658) were calculated from the data. Consequently the random match probability is (8.09%).

A total of 31 polymorphic sites were observed in these study population. We report a total of 14 haplotypes in HVR-I, 8 in HVR-II regions. Over 79% of the individuals analyzed exhibited between 2 to 8 polymorphisms in the control region. Nucleotide substitutions (42%) were the most common, compared to insertion (10.41%) or deletion (0%). The most common transition scored was G → A and T → C as well as C → T respectively. While transversion of C → G substitution scored was (6.24%).

The probability of two randomly selected individuals having identical mtDNA types in a population is 0.42%. The most frequent haplotype showed differences with respect to the rCRS (T16083C, T16093C, G16129C, C16223T, 16319.1A, 73G, 263G). The haplotype accumulating more nucleotide substitutions relative to the rCRS was characterized by polymorphisms C16223T, C16278T, A16289T, 16296, 16200, 16325, T16319C, T16362C, G16390A, T16519C, A73G, A93G, T146C, C150T, T152C, C182T, T195C, C198T. Only one sample has C-stretch at (nt 16184-16192). Most of the common polymorphisms (56%) identified in Bangladeshi population were detected in Indian and Himalayas population [13] [14].

Haplogroup Structure and Most Common Haplotypes

The proportion of the most common macrohaplogroups M, N, R, A4 and U2b were 47.5%, 36%, and 12.5%, 3.0% and 1.0% respectively. The most frequent haplotype was M5 (14.7%; with sequence motif; 73-750-16129-16290-16304) and M3 (5.9%; 16223-16311), followed by A4 (3.4%; 16129-16223-16290-16362) and R21 (3.2%; 152-16223-16362, 16519) [15]-[18] most common haplogroup frequencies shown in **Table 1**.

Forensic parameters and diversity is calculated in order to compare mainstream population of Bangladesh with tribal population. mtDNA control regions sequence of this population showed lower sequence diversity 0.8475 ± 0.0134 , than the tribal Indian-specific diversity 0.989 (95% CI: 0.975, 0.969) indicating gene flow shares with East and Southeast Asian specific diversity than Indian diversity [19]. The random match probability is 8.09% with 43 substitutions. Only one sample has “C” stretch and five samples have 319A-ins which is high in tribal population of Bangladesh **Table 2**.

Haplogroups C and U each occurred at frequencies lower than 3%. We found no remarkable new lineages in the mainstream population of Bangladesh and haplotypes are less complex rather than tribal population of Chittagong Hill Tracts of Bangladesh. A detail list of diversity comparison is given in **Table 3**.

Table 1. mtDNA haplogroup frequencies among the Bengali population of Bangladesh.

mtDNA							
Population	Location	n	A4	F1a	H2	M	M3
Bengali	Dhaka	31	0.24	0.19	0.25	0.45	0.40
	Chittagong	48	0.18	0.27	0.05	0.27	0.08
	Bogra	29	-	0.06	0.10	0.19	0.16
Population	Location	n	M5	M2a1a	M7	M21	M45
Bengali	Dhaka	31	0.88	0.01	0.27	0.08	0.02
	Chittagong	48	0.11	-	0.45	-	1
	Bogra	29	0.11	-	0.18	0.08	0.02
Population	Location	n	N9a	R21	U2b	C1	
Bengali	Dhaka	31	0.02	0.15	0.03	0.01	
	Chittagong	48	-	0.05	0.03	-	
	Bogra	29	-	0.05	0.03	-	

Table 2. Diversity and forensic parameters computed from entire mtDNA control region.

Parameters	HVS-I	HVS-II
Sequence diversity	0.8475 ± 0.0134	0.8291 ± 0.0149
Pairwise difference	9.698 ± 1.8658	9.8890 ± 1.5213
Random match probability	8.09%	8.69%
Haplotypes	14	8
Substitutions	43	0
Insertion in poly C-stretch	1	0
Insertion at 319A-ins	5	0

Table 3. Comparison of diversity measures from Northeast India.

Population statistics	Bangladesh	Assam	Meghalaya	Mizoram
N	108	752	56	45
Reference	This study	[16]	[17]	[18]
No. of haplotypes	14	316	31	124
Number of unique	3	153	12	36
Haplotypes diversity	0.847	0.995	0.986	0.991
Mean pairwise difference	9.890 ± 4.189	9.790 ± 4.14	9.519 ± 4.43	10.631 ± 4.07

4. Discussion

Mitochondrial (mtDNA) haplogroup is very useful tool not only for phylogenetic and clinical study, but it is also useful for forensic study. The direct sequencing of samples in this study has provided a premature database for the initiation of mass population screening. This study reports the first time analysis of the mtDNA control region and three coding regions sequences of mainstream population of Bangladesh and was generated according to standard operating procedure (SOP) of laboratory. Nevertheless, our analysis report of the entire mtDNA control region of 108 individuals proved to be useful to determine the ancestry of the samples examined, by contributing to the confirmation, and on occasion, even to the refinement of the haplogroup assignment. In fact, a precise haplogroup classification is an issue of increasing importance when mtDNA analysis is applied to forensic casework.

The distributions of mtDNA haplogroups in mainstream population of Bangladesh are less complex than those of the tribal populations of Chittagong Hill-Tracts [19] and West Bengal of India [20]. The low sequence diversity of mainstream Bangladeshi population indicated that this population are distinct than those of Thailand, Myanmar, and Northeast India. As the land between NE India and Bangladesh is of lower elevation it is not clear why the Bangladeshi population appears to be bottlenecked (low genetic diversity and limited distribution of haplotypes across the haplotype network). It could be due to the low sample size ($n = 108$) considered in this study. The diversity estimates for the mtDNA-control region, and the sequence diversity values indicate that the use of mitochondrial DNA is very informative for forensic case work, since the probability of discriminating two maternal lineages is high.

The random match probability of the mtDNA sub-haplotypes was <9%. Thus, the panel of 37 variable polymorphic sites identified in this study represents a useful forensic tool to further resolve the identity of individual from mitochondrial DNA (mtDNA).

The frequencies of the M5 and M3 haplogroups were the most common found in this group of population. Uttar Pradesh of India and individuals from Rajputs of Bihar and Muslims of Karnataka were also found to harbor this haplogroups [21]-[23]. Further M haplogroup is very common in Bangladesh; M2a is the oldest haplo-

group found in Indian Sub-continent and M2a is the third most common in Bangladesh. Other haplogroup; M7, M21, and M45, A4, N9, U2b, R21 show genetic affinity between Bangladesh and East/Southeast Asia, which was also concluded for Tibeto-Burman population of Bangladesh.

Haplogroup U2 is most common in South Asia and sub clades U2b was found in Bangladeshi population. While U2 is typically found in India, it is also present in the Nogais people descendants of various Mongolic and Turkic tribes, who formed the Nogai Horde indicating the [24] admixture of Indian specific branches of haplogroup in Bangladesh. Haplogroup M7 is found in East Asia, especially in Japan, Southern China, Vietnam and Laos (explains the gene flow from East-Asia). It also suggest that China, near Tibet, as a possible origin, from there they entered Myanmar around the 6th-7th century AD, where they still live in large number. Some group might be then migrated to Bangladesh and Thailand [25].

Obviously, knowing the ancestry (geographic origin and/or ethnicity) of a sample may be the first important step in forensic cases for which only a very limited amount of DNA is available. In cases where direct information from coding regions is limited, one can at least anticipate the haplogroup according to mutations of HVR-I and HVR-II regions. In these later cases, implementing the analysis of the entire mtDNA control region as a routine procedure in forensic investigation would be suggested [26].

5. Conclusion

In all, we can conclude that in isolated human groups, where mitochondrial genome diversities may be substantially low, although the improvement of the power of discrimination between sequences by analysis of the entire mtDNA control region seems to be somewhat limited, this procedure might constitute a very useful tool to carry out reliable, unambiguous haplogroup identification. Because of the feasibility of this system and the need for only small amounts of DNA, we can anticipate its use in studies of mtDNA-associated disorders, individual/species identification, and maternity testing in forensic genetics for general population of Bangladesh.

Acknowledgements

The authors would like to thank to all blood donors. Special thanks to Professor Altaf Hussain, Director, Centre for Advanced Research in Sciences, University of Dhaka for financial support. Our special thanks to Dr. K. Thangaraj, Scientist Centre for Cellular and Molecular Biology, Hyderabad 50007, India, for supporting us with his suggestion on mtDNA haplogroup identification.

References

- [1] Stoneking, M., Hedgecock, R.G., Higuchi, L. and Vigilant, H.A. (1991) Population Variation of Human mtDNA Control Region Sequences Detected by Enzymatic Amplification and Sequence-Specific Oligonucleotide Probes. *American Journal of Human Genetics*, **48**, 370-382.
- [2] Tajima, F. (1989) Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics*, **123**, 585-595.
- [3] Karafet, T., Xu, R.L., Du, R.F., Wang, W., Feng, S., Wells, R.S., Redd, A.J., Zegura, S.L. and Hammer, M.F. (2001) Paternal Population History of East Asia: Sources, Patterns, and Micro Evolutionary Processes. *American Journal of Human Genetics*, **69**, 615-628. <http://dx.doi.org/10.1086/323299>
- [4] Cobl, M.D., Just, R.S., O'Callaghan, J.E., Letmanyi, I.H., Peterson, C.T., Irwin, J.A. and Parsons, T.J. (2003) Single Nucleotide Polymorphisms over the Entire mtDNA Genome That Increase the Forensic Power of mtDNA Testing in Caucasians. *International Journal of Legal Medicine*, **118**, 137-146. <http://dx.doi.org/10.1007/s00414-004-0427-6>
- [5] Imaizumi, K., Parsons, T.J., Yoshino, M. and Holland, M.M. (2002) A New Database of Mitochondrial DNA Hyper-variable Regions I and II Sequences from 162 Japanese Individuals. *International Journal of Legal Medicine*, **116**, 68-73. <http://dx.doi.org/10.1007/s004140100211>
- [6] Anita, B., Christine, T.P., Jodi, A.I., Solomon, M., Davy, K.K., Walther, P. and Thomas, J.P. (2004) Mitochondrial DNA Control Region Sequences from Nairobi (Kenya): Inferring Phylogenetic Parameters for the Establishment of a Forensic Database. *International Journal of Legal Medicine*, **118**, 294-306. <http://dx.doi.org/10.1007/s00414-004-0466-z>
- [7] Sultana, G.N.N., Rahman, A., Karim, M.M., Shahinuzzaman, A.D.A., Begum, R. and Begum, R.A. (2011) Breast Cancer Risk Associated Mitochondrial NADH-Dehydrogenase Subunit-3 (ND3) Polymorphisms (G10398A and T10400C) in Bangladeshi Women. *Journal of Medical Genetics Genomics*, **3**, 131-135.

- [8] Kivisild, T., Bamshad, M.J., Kaldma, K., Metspalu, M., Reidla, M., Laos, S., Parik, J., Watkins, W.S., *et al.* (1999) Deep Common Ancestry of Indian and Western-Eurasian Mitochondrial DNA Lineages. *Current Biology*, **9**, 1331-1334. [http://dx.doi.org/10.1016/S0960-9822\(00\)80057-3](http://dx.doi.org/10.1016/S0960-9822(00)80057-3)
- [9] Kivisild, T., Tolk, H.V., Parik, J., Wang, Y., *et al.* (2002) The Emerging Limbs and Twigs of the East Asian mtDNA Tree. *Molecular Biology & Evolution*, **19**, 1737-1751. <http://dx.doi.org/10.1093/oxfordjournals.molbev.a003996>
- [10] Bandelt, H.J., Forster, P., Sykes, B.C. and Richards, M.B. (1995) Mitochondrial Portraits of Human Populations Using Median Networks. *Genetics*, **141**, 743-753.
- [11] Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., *et al.* (1981) Sequence and Organization of Human Mitochondrial Genome. *Nature*, **290**, 457-465. <http://dx.doi.org/10.1038/290457a0>
- [12] Schneider, S., Guillaume, L. and Excoffier, L. (2005) Arlequin (Version 3.0): An Integrated Software Package for Population Genetics Data Analysis. *Evolutionary Bioinformatics*, **1**, 47-50.
- [13] Kong, Q.P., Yao, Y.G., Sun, C., Bandelt, H.J., Zhu, C.L. and Zhang, Y.P. (2003) Phylogeny of East Asian Mitochondrial DNA Lineages Inferred from Complete Sequence. *American Journal of Human Genetics*, **73**, 671-676. <http://dx.doi.org/10.1086/377718>
- [14] Thangaraj, K., Chaubey, G., Kivisild, T., Reddy, A.G., Singh, V.K., Rasalkar, A.A. and Singh, V.K. (2005) Reconstructing the Origin of Andaman Islanders. *Science*, **308**, 996. <http://dx.doi.org/10.1126/science.1109987>
- [15] Bergen, A.W., Wang, C.Y., Tsai, J., Jefferson, K., Dey, C., Smith, K.D., Park, S.C., Tsai, S.J. and Goldman, D. (1999) An Asian-Native American Paternal Lineage Identified by RPS4Y Resequencing and by Microsatellite Haplotyping. *Annals of Human Genetics*, **63**, 63-80. <http://dx.doi.org/10.1046/j.1469-1809.1999.6310063.x>
- [16] Bertorelle, G. and Excoffier, L. (1998) Inferring Admixture Proportions from Molecular Data. *Molecular Biology & Evolution*, **15**, 1298-1311.
- [17] Rajkumar, R., Banerjee, J., Gunturi, H.B., Trivedi, R. and Kashyap, V.K. (2005) Phylogeny and Antiquity of M Macrohalogroups Inferred from Complete mtDNA Sequence of Indian Specific Lineages. *BMC Evolutionary Biology*, **5**, 26. <http://dx.doi.org/10.1186/1471-2148-5-26>
- [18] Miller, K.W. and Budowle, B. (2001) A Compendium of Human Mitochondrial DNA Control Region: Development of an International Standard Forensic Database. *Croatian Medical Journal*, **42**, 315-327.
- [19] Gazi, N.N.S., Tamang, R., Singh, V.P., Pathak, A.K., Singh, M., Chaubey, G., Thangaraj, K., *et al.* (2013) Genetic Structure of Tibeto-Burman Population of Bangladesh: Evaluating the Gene Flow along the Sides of Bay-of-Bengal. *PLoS ONE*, **8**, e75064. <http://dx.doi.org/10.1371/journal.pone.0075064>
- [20] Basu, A., Mukherjee, N., Roy, S., Sengupta, S., Banerjee, S., Chakraborty, M., *et al.* (2003) Ethnic India: A Genomic View, with Special Reference to Peopling and Structure. *Genome Research*, **13**, 2277-2290. <http://dx.doi.org/10.1101/gr.1413403>
- [21] Ballinger, S.W., Schurr, T.G., Torroni, A., Gan, Y.Y., Hodge, J.A., Hassan, K., Chen, K.H. and Wallace, D.C. (1992) Southeast Asian Mitochondrial DNA Analysis Reveals Genetic Continuity of Ancient Mongoloid Migrations. *Genetics*, **130**, 139-152.
- [22] Cordaux, R., Saha, N., Bentley, G.R., Auger, R., Sirajuddin, S.M. and Stoneking, M. (2003) Mitochondrial DNA Analysis Reveals Diverse Histories of Tribal Populations from India. *European Journal of Human Genetics*, **11**, 253-264. <http://dx.doi.org/10.1038/sj.ejhg.5200949>
- [23] Sun, C., Kong, Q.P., Palnichamy, M.G., Agrawal, S., Bandelt, H.J., Yao, Y.G., *et al.* (2006) The Dazzling Array of Basal Branches in the mtDNA Macrohalogroups M from India as Referred from Complete Genomes. *Molecular Biology & Evolution*, **23**, 683-690. <http://dx.doi.org/10.1093/molbev/msj078>
- [24] Bermisheva, M.A., Kutuev, I.A., Korshunova, T., Dubova, N.A., Villems, R. and Khusnutdinova, E.K. (2004) Phylogeographic Analysis of Mitochondrial DNA Nogays: The High Level of Mixture of Maternal Lineages from Eastern and Western Eurasia. *Molekuliarnaia Biologiya*, **38**, 617-624.
- [25] Bodner, M., Zimmermann, B., Röck, A., Kloss-Brandstätter, A., Horst, D., Horst, B., *et al.* (2011) Southeast Asian Diversity: First Insights into the Complex mtDNA Structure of Laos. *BMC Evolutionary Biology*, **11**, 49. <http://www.biomedcentral.com/1471-2148/11/49>
<http://dx.doi.org/10.1186/1471-2148-11-49>
- [26] Yang, Y., Zhang, P., He, Q., Zhu, Y.L., Yang, X.H., Lv, R. and Chen, J.H. (2011) A New Strategy for the Discrimination of Mitochondrial DNA Haplogroups in Han Population. *Journal of Forensic Science*, **56**, 586-590. <http://dx.doi.org/10.1111/j.1556-4029.2011.01711.x>