Data Article

Sequence polymorphism data of the hypervariable regions of mitochondrial DNA in the Yadav population of Haryana

Kapil Verma, Sapna Sharma, Arun Sharma, Jyoti Dalal, Tapeshwar Bhardwaj

Department of Genetics, Maharshi Dayanand University, Rohtak, Haryana 124001, India
Govt. of Himachal Pradesh, Junga, Himachal Pradesh 173216, India

Abstract

Genetic variations among humans occur both within and among populations and range from single nucleotide changes to multiple-nucleotide variants. These multiple-nucleotide variants are useful for studying the relationships among individuals or various population groups. The study of human genetic variations can help scientists understand how different population groups are biologically related to one another. Sequence analysis of hypervariable regions of human mitochondrial DNA (mtDNA) has been successfully used for the genetic characterization of different population groups for forensic purposes. It is well established that different ethnic or population groups differ significantly in their mtDNA distributions. In the last decade, very little research has been conducted on mtDNA variations in the Indian population, although such data would be useful for elucidating the history of human population expansion across the world. Moreover, forensic studies on mtDNA variations in the Indian subcontinent are also scarce, particularly in the northern part of India. In this report, variations in the hypervariable regions of mtDNA were analyzed in the Yadav population of Haryana. Different molecular diversity indices were computed. Further, the obtained haplotypes were classified into

Keywords:
Genetic variation
Hypervariable regions
mtDNA
Forensic
Yadav

Article info

Article history:
Received 23 January 2018
Received in revised form 9 February 2018
Accepted 1 March 2018
Available online 8 March 2018

Keywords:
Genetic variation
Hypervariable regions
mtDNA
Forensic
Yadav

A B S T R A C T

Genetic variations among humans occur both within and among populations and range from single nucleotide changes to multiple-nucleotide variants. These multiple-nucleotide variants are useful for studying the relationships among individuals or various population groups. The study of human genetic variations can help scientists understand how different population groups are biologically related to one another. Sequence analysis of hypervariable regions of human mitochondrial DNA (mtDNA) has been successfully used for the genetic characterization of different population groups for forensic purposes. It is well established that different ethnic or population groups differ significantly in their mtDNA distributions. In the last decade, very little research has been conducted on mtDNA variations in the Indian population, although such data would be useful for elucidating the history of human population expansion across the world. Moreover, forensic studies on mtDNA variations in the Indian subcontinent are also scarce, particularly in the northern part of India. In this report, variations in the hypervariable regions of mtDNA were analyzed in the Yadav population of Haryana. Different molecular diversity indices were computed. Further, the obtained haplotypes were classified into

https://doi.org/10.1016/j.dib.2018.03.004
2352-3409 © 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
different haplogroups and the phylogenetic relationship between different haplogroups was inferred.

© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Specifications Table

| Subject area          | Forensic Science             |
|-----------------------|------------------------------|
| More specific subject area | Forensic Genetics            |
| Type of data          | Tables and Figure            |
| How data was acquired | Data was acquired by extracting, amplifying, sequencing and analysing the target region of mtDNA from the blood samples by using SureCycler 8800 (Agilent Technologies, USA), Gel Documentation System (Alpha Innotech, USA) DNA sequencer (Applied Biosystems by Life Technologies, CA, USA) Arlequin software version 3.5 (Computational and Molecular Population Genetics Lab, Zoological Institute, Switzerland), HaploGrep 2 software (Medical University of Innsbruck, Austria) |
| Data format           | Analysed                     |
| Experimental factors  | Blood sample collection, DNA Extraction, PCR Amplification, Sequencing and Interpretation of Data |
| Experimental features | During the experiments of extraction and amplification, the contamination is eliminated by using the filtered pipette tips, gloves, masks, lab coats, autoclaving of stock chemicals/tubes and separation of pre and post amplification areas in the laboratory |
| Data source location  | Haryana (A northern state of India) |
|                       | Latitude: 29.0588°N          |
|                       | Longitude: 76.0856°E         |
| Data accessibility    | The data is available with this article |

Value of the data

- The present data is highly useful for the identification of individuals hypervariable involved in mass disasters, missing person cases and criminal cases in the Yadav population of Haryana.
- This data will help assess matches in mtDNA sequences in forensic casework in Haryana, and will be useful for population analyses based on specific sequence polymorphisms in the Yadav population of Haryana.
- The data report will provide baseline information for genetic studies based on the control region of mtDNA for tracking families related to the Yadav population of Haryana.
- This report is important for anthropological and evolutionary research, as well as for phylogenetic studies on the Yadav population of Haryana.
- This report could also be used by evolutionary biologists to study genetic variations in order to understand the possible relationships of the Yadav population with other populations.
- The data presented here can be used as reference material for future genetic studies on the Yadav population of Haryana.
- The mtDNA haplogroups generated in this data report can be used for tracing the migration and ancestry of the Yadav population of Haryana.
- The present data will contribute to the DNA database for the Yadav population of Haryana, which can be used for calculating the probability of matches based on mtDNA.
1. Data

- Table 1: Primers used for amplification of the hypervariable regions (HVI, HVII and HVIII). Nucleotide position, primer name, sequence, length and melting temperature of the primers are mentioned.
- Table 2: Constituents of the PCR reaction mixture for each reaction volume of 25 µl per sample for amplification of the HVI, HVII and HVIII regions.
- Table 3: PCR cycling conditions for amplification of three hypervariable regions of mitochondrial DNA.
- Table 4: Molecular diversity indices for the HVI region, HVII+HVIII region & HVI+HVII+HVIII region.
- Table 5: Frequency distribution of the mtDNA haplotypes in the Yadav population.
- Table 6: Sequence polymorphism of the three hypervariable regions and their respective haplogroups in the Yadav population.
- Table S1: GenBank accession numbers for the mtDNA polymorphisms identified in the Yadav population.
- Figure 1: Phylogenetic tree of haplogroups including all related polymorphisms relative to the rCRS for Yadav population

2. Experimental design, materials and methods

2.1. Blood samples and DNA extraction

Blood samples were collected from 66 maternally unrelated individuals belonging to the Yadav population of Haryana from nearly all districts of Haryana. A sample of 2–5 ml of venous blood was drawn into 5 ml EDTA vacutainer tubes (Greiner Bio-One, USA). A consent form was signed by each participating individual at the site of collection. DNA was extracted using the phenol-chloroform-isoamyl (PCI) method [1].

2.2. PCR amplification

The three hypervariable regions were amplified by using two sets of PCR reactions. The primers F15900 and R00159 were used to amplify the HVI region. The primers F00015 and R00599 were used to amplify the HVII and HVIII regions [2] (Table 1). They were synthesized at Integrated DNA Technologies (IDT, USA). The PCR reaction was carried out in a final volume of 25 µl (Table 2). PCR was performed on SureCycler 8800 (Agilent Technologies, USA) (Table 3). Positive and negative controls were also used to ensure that no contamination was present at any stage during the experiments. The amplified PCR product was visualized using 1.6% agarose gel in the Gel Documentation System (Alpha Innotech). The amplified PCR product was cleaned with a GeneJET PCR Purification Kit (Thermo Fisher, USA) according to the manufacturer’s guidelines to remove any impurities present in the template.

Table 1

| MtDNA region | Nucleotide position | Primers | Primer sequence (5′-3′) | Length (Bases) | Tm Value | PCR product Size (bp) |
|--------------|---------------------|---------|-------------------------|----------------|----------|----------------------|
| HVI          | 16024–16365         | L15900  (F) | TACACCAGTCTTGTAACCATGTAAC  | 19             | 49.1 °C  | 828                  |
|              |                     | H00159  (R) | AAATAATAGGATGAGGAATC      | 24             | 52.5 °C  |                      |
| HVII & HVIII | 73–576              | L00015  (F) | CACCCATTACCACCACCATCACC  | 20             | 52.7 °C  | 585                  |
|              |                     | H00599  (R) | TTGAGGAGGTAAGCTACATAA     | 21             | 50.3 °C  |                      |
Table 2
Constituents of the PCR reaction mixture for each reaction volume of 25 µl per sample for amplification of the HVI, HVII and HVIII regions.

| S.No. | Chemical                            | Quantity     |
|-------|-------------------------------------|--------------|
| 1     | 10 × PCR buffer                     | 2.5 µl       |
| 2     | 0.2 mM of each dNTPs                | 2.5 µl       |
| 3     | 0.6 µM each of forward primer (F)   | 1.5 µl       |
| 4     | 0.6 µM each of reverse primer (R)   | 1.5 µl       |
| 5     | 5 Units/µl of Taq DNA polymerase    | 0.5 µl       |
| 6     | D/DH2O                              | 15.5 µl      |
| 7     | 50 ng of template DNA               | 1 µl         |

Table 3
PCR cycling conditions for amplification of three hypervariable regions of mitochondrial DNA.

| Cycle Step                  | Temperature | Time duration |
|-----------------------------|-------------|---------------|
| Hot start                   | 95 °C       | 11 min        |
| Denaturation                | 95 °C       | 15 s          |
| Annealing                   | 62 °C       | 30 s          |
| Elongation                  | 72 °C       | 1 min         |
| End cycle Elongation        | 72 °C       | 10 s          |
| Hold                        | 4 °C        | ∞             |

Table 4
Molecular diversity indices for the HVI region, HVII + HVIII region & HVI + HVII + HVIII region.

| Diversity indices                              | HVI region | HVII + HVIII region | HVI + HVII + HVIII region |
|------------------------------------------------|------------|---------------------|---------------------------|
| No. of polymorphic sites                       | 50         | 55                  | 105                       |
| No. of observed transitions                    | 48         | 41                  | 89                        |
| No. of observed transversions                  | 4          | 10                  | 14                        |
| No. of observed substitutions                  | 52         | 51                  | 103                       |
| No. of observed indels                         | –          | 5                   | 5                         |
| Nucleotide composition (%) C                   | 33.31      | 34.59               | 34.07                     |
| T                                              | 22.41      | 22.76               | 22.62                     |
| A                                              | 33.01      | 30.45               | 31.49                     |
| G                                              | 11.27      | 12.20               | 11.82                     |
| Mean number of pairwise differences            | 4.27 ± 2.14| 5.603 ± 2.7054      | 9.8331 ± 4.5583           |
| Heterozygosity/sample                          | 0.086 ± 0.09| 0.1013 ± 0.12     | 0.09433 ± 0.11           |
| No of haplotypes                              | 50         | 55                  | 66                        |
| Gene diversity                                | 0.9869 ± 0.0065 | 0.9925 ± 0.0046 | 1.000 ± 0.0026           |
| Nucleotide diversity                          | 0.01238 ± 0.0068 | 0.01092 ± 0.0058 | 0.01151 ± 0.005918       |
| Ss2 of haplotype frequencies (RMP)             | 0.028      | 0.0225              | 0.0152                    |
| Alleles frequency (Mean ± SD)                  | 1.151 ± 0.374| 1.11 ± 0.31        | 1.156 ± 0.388            |
| Sum of square deviation                        | 0.0045     | 0.0028              | –                         |
| Harpending's raggedness index                  | 0.0159     | 0.0103              | –                         |
| Mismatch distribution observed mean            | 4.273      | 5.560               | –                         |
| Mismatch observed variance                     | 4.62       | 6.031               | –                         |
| Tajima's D test                                | −1.9763    | −1.8781             | −1.9896                   |
| Fu's FS test                                   | −25.8062   | −25.3770            | −24.6038                  |

a Hypervariable region I.
b Hypervariable region II.
c Hypervariable region III.
Table 5
Frequency distribution of the mtDNA haplotypes in the Yadav population.

| Number of times a haplotype repeated | Numbers of Haplotypes |
|--------------------------------------|-----------------------|
|                                      | HVI region            | HVII + HVIII region |
| 1                                    | 40                    | 48                   |
| 2                                    | 7                     | 4                    |
| 3                                    | 2                     | 2                    |
| 4                                    | –                     | 1                    |
| 6                                    | 1                     | –                    |
| Total                                | 50                    | 55                   |
| Random Match Probability             | 0.028                 | 0.022                |

Table 6
Sequence polymorphism of the three hypervariable regions and their respective haplogroups in the Yadav population.

| Sample ID | HVI region | HVII region | HVIII region | Haplogroup |
|-----------|------------|-------------|--------------|------------|
| YA1       | 16129A     | 16242T     | 16356C       | H1b        |
| YA2       | 16189C     | 16223T     | 16278T       | L3b1a+@16124 |
| YA3       | 16095T     | 16223T     | 16335G       | M          |
| YA4       | 16292T     |             |              | W5a2       |
| YA5       | 16111A     | 16223T     |              |            |
| YA6       | 16036R     | 16189C     | 16223T       |            |
| YA7       | 16309G     | 16318T     |              |            |
| YA8       | 16192T     | 16223T     | 16278T       |            |
| YA9       | 16309G     | 16318T     |              |            |
| YA10      | 16126C     | 16181G     | 16209C       |            |
| YA11      | 16223T     | 16181G     | 16325C       |            |
| YA12      | 16048A     | 16129A     | 16223T       |            |
| YA13      | 16126C     | 16223T     | 16311C       |            |
| YA14      | 16309G     | 16318C     |              |            |
| YA15      | 16174T     | 16354T     |              |            |
| YA16      | 16126C     | 16223T     | 16311C       |            |
| YA17      | 16111T     | 16184T     | 16189C       |            |
| YA18      | 16036R     | 16048A     | 16129A       |            |
| YA19      | 16150T     | 16223T     | 16298C       |            |
| YA20      | 16036R     | 16309G     | 16318T       |            |
| YA21      | 16126C     | 16129A     | 16183C       |            |
| YA22      | 16093W     | 16172C     | 16293G       |            |
| YA23      | 16095T     | 16184T     | 16223T       |            |
| YA24      | 16111T     | 16183C     | 16189C       |            |
| YA25      | 16036R     | 16355T     |              |            |
| Sample ID | HVI region | HVII region | HVIII region | Haplogroup |
|-----------|------------|-------------|--------------|------------|
| YA26      | 16111Y 16129R 16189Y 16218Y 16223T 16293Y 16311Y | 73G 151T 152C 263G 309.1C 315.1C | 544d | N5 |
| YA27      | 16226T 16304C | 253T 263G 309.1C 315.1C | 487T 544d | H2a2 |
| YA28      | 16183C 16193.1C | 73G 263G 309.1C 315.1C | 448M 482C | M3a1+204 |
| YA29      | 16036R | 253T 263G 309.1C 315.1C | 443M 489C | M1a3b1 |
| YA30      | 16095T 16184T 16223T 16249C 16359C | 253T 263G 309.1C 315.1C | 489C | U8c |
| YA31      | 16126C 16186T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M3a1 |
| YA32      | 16189C 16223T | 523d 524d | N9b |
| YA33      | 16309G 16318T | 73G 263G 309.1C 315.1C | 489C | M5a2a1a |
| YA34      | 16174Y 16223Y 16243Y 16304Y | 16129A 16223T | 489C | M3a1 |
| YA35      | 16218T 16223T | 150T 263G 309.1C 315.1C | 489C | M3a1 |
| YA36      | 16129A 16223T | 253T 263G 309.1C 315.1C | 489C | M3a1 |
| YA37      | 16174Y 16223Y | 16129A 16223T | 489C | M3a1 |
| YA38      | 16189C 16223T | 253T 263G 309.1C 315.1C | 489C | M3a1 |
| YA39      | 16292T | 523d 524d | N9b |
| YA40      | 16126C 16186T 16223T | 150T 263G 309.1C 315.1C | 489C | M3a1 |
| YA41      | 16218T 16223T | 150T 263G 309.1C 315.1C | 489C | M3a1 |
| YA42      | 16189C 16223T | 253T 263G 309.1C 315.1C | 489C | M3a1 |
| YA43      | 16189C 16223T | 253T 263G 309.1C 315.1C | 489C | M3a1 |
| YA44      | 16218T 16223T | 150T 263G 309.1C 315.1C | 489C | M3a1 |
| YA45      | 16126C 16186T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA46      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA47      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA48      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA49      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA50      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA51      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA52      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA53      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA54      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA55      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA56      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA57      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA58      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA59      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA60      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA61      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA62      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA63      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA64      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA65      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
| YA66      | 16218T 16223T | 150T 263G 309.1C 315.1C | 523d 524d | M5a2a1a |
2.3. Sequencing

The cleaned PCR product was sequenced by commercial DNA sequencing service (Xcelris Labs Limited, Ahmedabad, India). Both the strands were sequenced with the ABI BigDye Terminator Cycle Sequencing Kit on the ABI 3700 Genetic Analyzer (Applied Biosystems). All the samples were sequenced with the same primers used in PCR amplification of the HVI, HVII and HVIII regions. An additional primer (16410R-GAGGATGGTGGTGGTCAA) was used in cases where there was slippage due to ‘C’ stretch in hyper variable region 1.

2.4. Statistical analysis

A total of 66 mtDNA sequences was obtained, which included all the base pairs from nucleotide positions 16021−16365, 69−576bp. Sequence data files were edited and aligned with the revised Cambridge reference sequences (rCRS) [3] by using the Mega 7 software [4]. Manual alignment was also done to cross check the results by creating data analysis sheets. The interpretation was done as per the guidelines [5–7]. The coding for heteroplasmic sites was done according to the IUPAC codes in the interpretation guidelines to interpret the mtDNA data analysis results [8]. Any observed C-stretch length heteroplasmy in the HVI, HVII and HVIII region sequences was excluded from statistical analysis. Statistical analysis was first performed for each hypervariable segment separately, and then combined for the HVI+HVII+HVIII regions. Gene diversity was calculated according to Tajima [9]. Population pairwise differences were determined based on genetic distances [10]. Haplotype diversity, mean pairwise differences, nucleotide diversity, Harpending’s raggedness index, mismatch distributions, Fu’s Fs and Tajima’s D test statistics were calculated using the Arlequin software, version 3.5.1.2 [11] (Table 4). A Random match probability (RMP) was calculated according to Stoneking et al. [12] (Table 5). Haplogroup classification was performed using the HaploGrep 2 software [13] (Table 6). A phylogenetic tree of all the haplogroups was constructed using HaploGrep 2, while all the classified samples were combined to produce a resulting (rooted) tree, which included all the related polymorphisms relative to the rCRS (Fig. 1). GenBank accession numbers for the mtDNA polymorphisms identified in the Yadav population are provided in Table S1 (Supplementary table).

Acknowledgements

This work was supported by the University Grant Commission (UGC), New Delhi, Govt. of India under Major Research Project (MRP) scheme (Grant no. F.No. 42-45/2013 (SR)).

Transparency document. Supplementary material

Transparency document associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2018.03.004.
Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2018.03.004.

References

[1] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, New York, 1989.
[2] A. Brandstätter, C.T. Peterson, J.A. Irwin, S. Mpoke, D.K. Koech, W. Parson, T.J. Parsons, Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database, Int. J. Leg. Med. 118 (5) (2004) 294–306.
[3] R.M. Andrews, I. Kubacka, P.F. Chinnery, D.M. Turnbull, R.N. Lightowlers, N. Howell, Reanalysis and revision of the Cambridge Reference Sequence, Nat. Genet. 23 (2) (1999) 147.
[4] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol. Biol. Evol. 33 (2016) 1870–1874.
[5] M.M. Holland, T.J. Parsons, Mitochondrial DNA sequence analysis- validation and use for forensic casework, Forensic Sci. Rev. 11 (1999) 21–50.
[6] A. Carracedo, W. Bär, P.J. Lincoln, W. Mayr, N. Morling, B. Olaisen, DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing, Forensic Sci. Int. 110 (2000) 79–85.
[7] SWGDAM. Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories, 2013. Available at http://swgdam.org/SWGDAMS20mtDNA_Interpretation_Guidelines_APPROVED_073013.pdf.
[8] SWGDAM. Guidelines for mitochondrial DNA (mtDNA) nucleotide sequence interpretation, Forensic Science Communications, vol. 5(2), 2003. Available at (http://www2.fbi.gov/hq/lab/fsc/backissu/april2003/swgdammitodna.htm).
[9] F. Tajima, Evolutionary relationship of DNA sequences in finite populations, Genetics. 105 (1983) 437–460.
[10] J.B. Reynolds, B.S. Weir, C.C. Cockerham, Estimation of the coancestry coefficient: basis for a short-term genetic distance, Genetics 105 (1983) 767–779.
[11] L. Excoffier, H.E. Lischer, Arlequin Suite Ver. 3.5: a new series of programs to perform population genetic analyses under Linux and Windows, Mol. Ecol. Resour. 10 (3) (2010) 564–567.
[12] M. Stoneking, D. Hedgecock, R.G. Higuchi, L. Vigilant, H.A. Erlich, Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes, Am. J. Hum. Genet. 48 (1991) 370–382.
[13] H. Weissensteiner, D. Pacher, A. Brandstätter, L. Forer, G. Specht, H.J. Bandelt, et al., HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing, Nucleic Acids Res. 44 (2016) W58–W63.