



Genomic portrait of Odisha, India drawn by using 21 autosomal STR markers

Subhasish Sahoo¹ · Rashmita Samal¹ · Suchismeeta Behera¹ · Aruna Kumar Swain¹ · Sumit Biswas² · Pankaj Shrivastava³ · R. K. Kumawat⁴ · Gyaneshwer Chaubey⁵ · Hirak Ranjan Dash³

Received: 6 January 2020 / Accepted: 13 March 2020 / Published online: 24 March 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

In order to find out the genetic diversity in the eastern Indian population of Odisha consisting of various linguistic and ethnic groups, we undertook a study on 508 unrelated healthy individuals belonging to Odisha, India. We assessed genetic variation and compared the data with published literature of Indian population consisting of different ethnic groups from different geographical areas using 21 autosomal STR markers. The most polymorphic and discriminatory STR locus in the studied population was found to be SE33 with the calculated values of 0.94 and 0.991 respectively for both the parameters. The combined power of discrimination (CPD) and combined power of exclusion (CPE) were found to be 1 and 0.999999999704865 respectively. The combined probability of match (CP_m) and combined paternity index (CPI) for all 21 autosomal STR loci were found to be 8.01×10^{-26} and 3.45×10^9 respectively. Though, the studied eastern Indian population of Odisha shared its closest genetic affinity with nearest Indian geographical regions, i.e., the population of Jharkhand which is geographically located in eastern India as well as the Central Indian population.

Keywords STRs · Polymorphic · Genetic diversity · Odisha · Discrimination

Introduction

Odisha is a coastal state in the eastern India with 485 km of coastline along the Bay of Bengal which stands out as the 9th

Subhasish Sahoo, Pankaj Shrivastava, and R. K. Kumawat have equal credentials.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00414-020-02281-5>) contains supplementary material, which is available to authorized users.

✉ Hirak Ranjan Dash
hirakdash@gmail.com

- ¹ DNA Profiling Unit, State Forensic Science Laboratory, Rasulgarh, Bhubaneswar, Odisha 751010, India
- ² Department of Biological Sciences, Birla Institute of Technology & Sciences, Pilani K.K. Birla Goa Campus, NH 17B, Bypass Road, Zuarinagar, Sancoale, Goa 403726, India
- ³ DNA Fingerprinting Unit, State Forensic Science Laboratory, Department of Home (Police), Govt. of MP, Sagar 470001, India
- ⁴ DNA Division, State Forensic Science Laboratory, Jaipur, Rajasthan 302016, India
- ⁵ Cytogenetics Laboratory, Department of Zoology, Banaras Hindu University, Varanasi, India

largest state, spreading over an area of 155,707 km² (Fig. S1). As per the 2011 national census, Odisha has a population of around 42 million and is ranked 11th among the Indian states on the basis of its population [1]. Its population constitutes about 3.47% of the total population of India. The extant Odia population of Indo-Aryan ethnic group varies in ethnicity, hierarchical caste system, and native language. Odisha's relative isolation and lack of external invasions have conserved the social, religious, and genetic structure of the Odia population. The state is inhabited by the largest section of diversified tribal population including Santhal, Munda, Kandha, Bonda, Mahali, and Oraon. High-density population genetic data are needed to establish reference databases that can be used to statistically strengthen the results of forensic analyses. This is the first report on Odia population on 21 autosomal STR loci using a GlobalFiler™ PCR amplification kit (Thermo Scientific, US). The purpose of this study was to evaluate the extent of genetic diversity in the ethnic population of Odisha and to investigate the genetic diversity and allele frequency of the Odia population.

In the present study, 508 healthy unrelated individuals were considered, and the population samples were taken from various geographical regions of Odisha. All the samples were taken from the routine casework of the authors of DNA

Profiling Unit, State Forensic Science Laboratory, Rasulgarh, Bhubaneswar, Odisha, India, during July 2015 to November 2018 with written informed consent following the declaration of Helsinki. DNA was extracted from the samples by an automated DNA extraction system, EZ1 Advanced XL (Qiagen, Germany) using EZ1 DNA Investigator Kit (Qiagen) as per the protocol supplied by the manufacturer. Quantification of the isolated DNA samples was performed by 7500 real-time PCR (Thermo Scientific, US) using Quantifiler Duo DNA Quantification Kit (Thermo Scientific, US) following the manufacturer's protocol. The isolated DNA samples were further amplified at 21 STR loci which included the 20 CODIS loci (D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, D10S1248, D1S1656, D12S391, D2S1338) and another additional autosomal loci SE33 besides 3 gender-determining loci (Amelogenin, Y indel, and DYS391) using the GlobalFiler™ PCR amplification kit (Thermo Scientific, US) as per the protocol supplied by the manufacturer. Veriti™ 96-Well Fast Thermal Cycler (Thermo Scientific, US) was used for amplification of STR markers. The amplified products along with the positive and negative control samples, supplied by the kit manufacturer, were run on the Genetic Analyzer Model 3500xL (Thermo Scientific, US) using Gene Scan™ 600 Liz size standard (Thermo Scientific, US). The data was analyzed using the GeneMapper® IDX Software v1.4 (Thermo Scientific, US). The obtained data were statistically analyzed by the GenAlix 6.5 software [2], PowerStats v1.2 spreadsheet program [3], Arlequin v3.5 software [4], POPTREE2 program [5], and PAST 3.02a software [6].

The observed allele frequencies ranged from 0.001 to 0.409 (Table S2), out of which CSF1PO locus was found to have a maximum allele frequency and allele 12 (0.409) being the most frequent allele in the studied population. The power of discrimination (PD) ranged from 0.863 (D2S441) to 0.991 (SE33), and polymorphic information content (PIC) ranged from 0.65 (TPOX) to 0.94 (SE33). The most polymorphic and discriminatory STR locus in the studied population was found to be SE33 with values of 0.94 and 0.991 respectively. The study revealed a combined power of discrimination (CPD) and combined power of exclusion (CPE) for all the studied loci with the value of 1 and 0.99999999704865 respectively. The combined probability of match (CPM) and combined paternity index (CPI) for all the studied STR loci were found to be 8.01×10^{-26} and 3.45×10^9 respectively. The observed heterozygosity (H_{obs}) value varied from 0.679 (TPOX) to 0.925 (SE33).

Locus-wise allele frequencies of the studied population were compared at 15 common STR loci of previously published data on Indian populations [7–15] (Table S1), using pairwise F_{st} distance (Table S3). The population of Odisha

showed significant variations at 12 loci with Kora (Bengal), 11 loci with Yerukula (Andhra Pradesh), 8 loci with Konkanastha Brahmin (Maharashtra) and Bhil (Gujrat), 7 loci with Kurmans (Tamilnadu) and Central Indian population (Madhya Pradesh), 6 loci with Chenchu (Andhra Pradesh), Santhal (Chota Nagpur) and Jat Sikh (Punjab), 2 loci with Oraon (Chhattisgarh), and the population of Jharkhand out of 15 common STR loci (Table S3). In order to verify the results obtained from F_{st} and distance analysis, a neighbor-joining (NJ) tree was drawn (Fig. S2). The NJ tree revealed two major clusters, and Jat Sikh (Punjab), Yerukula (Andhra Pradesh), and Bhil (Gujrat) were found to be the outlier population. The studied population clustered along with the population of Jharkhand belonging to the eastern geographical region of India and with the Oraon population of Chhattisgarh. This shows a close genetic affinity of the studied population with Central Indian population as well as geographically closely located populations with higher bootstrap values which ranged from 41 to 77. Second cluster consists of Konkanastha Brahmin (Maharashtra) and Kurmans (Tamil Nadu) with the bootstrap value 57. Interestingly, the studied population showed significant genetic distance from distinct geographic regions, and the possible cause for the same may be a long term geographical isolation (Fig. S2). Similarly, the principal component analysis (Fig. S3) also showed the same clustering pattern in the compared populations. Thus, the NJ tree and PC analyses produced a consistent result with the previous observations.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical statement The study was approved by the ethical committee of Banaras Hindu University, Varanasi, India (Ref. No. - I.Sc./ECM-XII/2018-19/06).

References

1. Census of India (2011) <http://censusindia.gov.in/>. accessed 31.12.2018
2. Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295
3. Tereba A (1999) Tools for analysis of population statistics. Profiles DNA 2:14–16
4. Excoffier L, Laval G, Schneider S (2005) Arlequinver 3.0: an integrated software package for population genetics data analysis. Evol Bioinformatics Online 1:47–50
5. Takezaki N, Nei M, Tamura K (2009) POPTREE2: software for constructing population trees from allele frequency data and computing other population statistics with windows-interface. Mol Biol Evol 27:747–752

6. Hammer O, Harper DAT, Ryan PD (2001) PAST: palaeontological statistics software package for education and data analysis. *Palaeontol Electron* 4:9
7. Bindu HG, Trivedi R, Kashyap VK (2005) Genotypic polymorphisms at fifteen tetranucleotides and two pentanucleotide repeat locus in four tribal populations of Andhra Pradesh, southern India. *J Forensic Sci* 50(4):978–983
8. Banerjee J, Trivedi R, Kashyap VK (2005) Polymorphism at 15 short tandem repeat AmpFISTR Identifiler locus in three aboriginal populations of India: an assessment in human identification India. *J Forensic Sci* 50(5):1229–1234
9. Singh A, Trivedi R, Kashyap VK (2006) Genetic polymorphism at 15 tetrameric short tandem repeat locus in four aboriginal tribal populations of Bengal. *J Forensic Sci* 51(1):183–187
10. Chaudhari RR, Dahiya MS (2014) Genetic diversity of 15 autosomal short tandem repeats locus using the AmpFLSTR® Identifiler™ kit in a Bhil tribe population from Gujarat state, India. *Indian J Hum Genet* 20(2):148–152
11. Ghosh T, Kalpana D, Mukerjee S, Mukherjee M, Sharma AK, Nath S, Rathod VR, Thakar MK, Jha GN (2011) Genetic diversity of autosomal STRs in eleven populations of India. *Forensic Sci Int Genet* 5(3):259–261
12. Shrivastava P, Jain T, Trivedi VB (2015) Genetic polymorphism study at 15 autosomal locus in central Indian population. *SpringerPlus* 4:566
13. Imam J, Reyaz R, Singh RS, Bapuly AK, Shrivastava P (2017) Genomic portrait of population of Jharkhand, India, drawn with 15 autosomal STRs and 17 Y-STRs. *Int J Legal Med* 130(5):1185–1186
14. Giroti R, Talwar, (2013) Diversity and differentiation in Khatri, Baniyas, and Jat sikhs of Punjab; a study with forensic microsatellites. *Ind J Phys Anthropol Hum Genet* 32(2):309–328
15. Shrivastava P, Jain T, Trivedi VB (2016) Genetic portrait of Oraon Indian tribe drawn with 15 autosomal and 17 Y chromosomal STR markers. *Int J Legal Med* 130(5):1185–1186

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.