

Elevated East Eurasian Maternal Genetic Ancestry in the Haryana Jaat Population: Insights from Mitochondrial DNA Haplogroup Distribution and Comparative Analysis

Rupal Malik

Department of Biosciences and Technology, MMEC, Maharishi Markandeshwar (Deemed to be University), Mullana

Chanchal Devnani

Gyan Lab, Cytogenetics Unit, Department of Zoology, Banaras Hindu University

Debashruti Das

Gyan Lab, Cytogenetics Unit, Department of Zoology, Banaras Hindu University

Shailesh Desai

Gyan Lab, Cytogenetics Unit, Department of Zoology, Banaras Hindu University

Adesh K Saini

Department of Biosciences and Technology, MMEC, Maharishi Markandeshwar (Deemed to be University), Mullana

Nagarjuna Pasupuleti

Centre for Cellular and Molecular Biology

Rajeev Kawatra

Forensic Science Laboratory, Madhuban

Vasant Shinde

National Maritime Heritage Complex

Gyaneshwer Chaubey

`gyaneshwer.chaubey@bhu.ac.in`

Gyan Lab, Cytogenetics Unit, Department of Zoology, Banaras Hindu University

Research Article

Keywords: mtDNA, East Eurasian ancestry, Jaat/Jat population, Haryana, maternal lineages, population genetics, South Asia, haplogroup D, demographic history, gene flow

Posted Date: September 16th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10394347/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

**Elevated East Eurasian Maternal Genetic Ancestry in the Haryana Jaat Population:
Insights from Mitochondrial DNA Haplogroup Distribution and Comparative Analysis**

Rupal Malik¹, Chanchal Devnani⁴, Debashruti Das⁴, Shailesh Desai⁴, Adesh K Saini^{1,*},
Nagarjuna Pasupuleti⁵, Rajeev Kawatra², Vasant Shinde^{3,*}, Gyaneshwer Chaubey^{1,*}

¹ *Department of Biosciences and Technology, MMEC, Maharishi Markandeshwar (Deemed to be
University), Mullana, Ambala, Haryana, India*

² *Forensic Science Laboratory, Madhuban, Karnal, Haryana, India*

³ *National Maritime Heritage Complex, Lothal, Gujarat, India*

⁴ *Gyan Lab, Cytogenetics Unit, Department of Zoology, Banaras Hindu University, Varanasi,
Uttar Pradesh 221005, India*

⁵ *Centre for Cellular and Molecular Biology, Hyderabad 500 007, India*

Rupal Malik and Chanchal Devnani have contributed equally.

Corresponding Author:

sainiade@mmumullana.org (A.K Saini) or gyaneshwer.chaubey@bhu.ac.in (G. Chaubey) or
vshinde.dc@gmail.com (V. Shinde)

Abstract

The maternal genetic landscape of the Jaat population of Haryana, northwestern India, offers critical insights into the layered demographic history of South Asia. Here, we present a comprehensive analysis of mitochondrial DNA (mtDNA) control and coding region variations in 215 unrelated, healthy individuals from various *gotras* (clans) of the Haryana Jaat community. Our results delineate a tripartite maternal ancestry structure: a dominant South Asian component (60.28%, primarily within macrohaplogroup M and its subclades such as M3, M5, and M30, along with U2 lineages), a substantial West Eurasian contribution (30.62%- comprising U7, H2a, W, N1a, and related lineages), and a notable East/Southeast Asian signal (7.65%; n=16/209). The latter comprises haplogroups A, C4a, D (including D4 subclades), F (including F1 subclades), and G2a, while a small fraction of lineages belong to unresolved R-derived haplogroups whose geographic affinities remain limited. Critically, the final corrected frequency data show that the Jaat population (n=209) harbors 7.65% East Eurasian maternal ancestry (mainly haplogroup D), while the neighboring Ror, Gujjar, and Kamboj populations show 0%. This differential is a robust and distinctive feature of Jaat maternal history. The signal is corroborated by network analysis (distinct East Eurasian clusters) and is consistent with historical gene flow from Central/Northeast Asian sources via the northwestern corridors. Jaat also exhibits rich autochthonous South Asian maternal diversity. These findings position the Haryana Jaats as genetically distinctive among northwestern groups, with implications for South Asian peopling models and forensic genetics.

Keywords: mtDNA; East Eurasian ancestry; Jaat/Jat population; Haryana; maternal lineages; population genetics; South Asia; haplogroup D; demographic history; gene flow

Introduction

Haryana, located in northwestern India, occupies a strategic position along major ancient migratory corridors that have profoundly shaped the genetic architecture of North India. The region has long served as a crossroads for human movements, facilitating interactions between South Asian, West Eurasian, Central Asian, and, to a lesser extent, East/Northeast Asian populations (Sharma et al., 2020). Haryana is home to diverse ethnolinguistic communities, including the Jaats, Brahmins, Rors, Gujjars, and Kambojs, each characterized by distinct historical trajectories and sociocultural practices. Traditional social organization in the region features a caste-based hierarchy with strict endogamy at the community level and exogamy within *gotras* (clans), particularly among land-owning agrarian groups. These longstanding marriage patterns have generated fine-scale genetic substructure, making Haryana populations valuable models for studying genetic drift, lineage divergence, and admixture (Majumdar, 1951; Singh, 1992).

Among these communities, the Jaats represent one of the most demographically and socio-politically influential agrarian groups. Historical accounts variously link their origins to Indo-Scythian (Saka), Indo-Aryan, or Central Asian pastoralist and warrior traditions, with some narratives associating them with ancient nomadic confederations that traversed the Eurasian steppe and entered the subcontinent through the northwest (Bayley, 1998; Singh,

1992). Predominantly rural and land-owning, Jaats maintain *gotra* exogamy while practicing community endogamy, creating a natural experimental framework for population genetic analyses of intra-group diversity and inter-group affinities.

Despite their numerical strength and cultural prominence, the maternal genetic history of Haryana Jaats remains incompletely characterized. Most prior genetic research on North Indian populations has either examined broad caste-tribe contrasts or focused predominantly on patrilineal (Y-chromosome) variation, often overlooking the maternal dimension. Mitochondrial DNA (mtDNA), with its strict maternal inheritance, absence of recombination, and rapid mutation rate, serves as a powerful tool for reconstructing maternal ancestry, migration routes, and demographic events such as expansions or bottlenecks (Ingman et al., 2000; Pakendorf & Stoneking, 2005).

In the Indian subcontinent, mtDNA studies have established the predominance of macrohaplogroups M and R, both derived from the African L3 lineage (Kivisild et al., 1999; Thangaraj et al., 2006; Welikala et al., 2026). Numerous autochthonous sub-haplogroups within M and R reflect long-term in situ evolution following the initial peopling of South Asia (Welikala et al., 2024; Desai et al., 2026a; 2026b). However, fine-grained analyses targeting specific endogamous communities in the northwest, particularly those with putative Central Asian historical connections, remain sparse.

While the South Asian and West Eurasian strata of North Indian mtDNA pools have received considerable attention, the East and Southeast Asian maternal signals in populations of the Haryana-Punjab plains are frequently underappreciated or reported only as trace components. Haplogroups characteristic of East Asia (e.g., D, C, G), Southeast Asia (e.g., F, B), and their Central Asian extensions occur at low overall frequencies across most Indian groups but can accumulate in specific communities through targeted historical admixture. The northwestern corridor witnessed repeated interactions with nomadic groups originating in or transiting through Central and Inner Asia, including the Indo-Scythians, Kushans, Hephthalites (Hunas), and later Turkic-Mongolic expansions. These episodes likely introduced not only cultural innovations but also genetic lineages, including maternal ones, into local gene pools.

The Jaat community, given its historical associations with warrior-pastoralist traditions and hypothesized links to Central Asian movements, constitutes an excellent case study for quantifying any such East/Southeast Asian maternal contribution. Previous genomic work on Jat and related groups (Ror, Gujjar, Kamboj) has highlighted elevated West Eurasian (Steppe_MLBA-related) autosomal ancestry and distinctive paternal profiles, alongside a predominantly South Asian maternal base (Pathak et al., 2018). However, whether these populations differ in their retention or acquisition of East Eurasian mtDNA lineages, and whether Jaats specifically show enrichment, has not been systematically addressed with adequately powered, community-focused sampling.

The present study was therefore designed to: (i) characterize the full spectrum of mtDNA haplogroup diversity in a large sample of Haryana Jaats; (ii) quantify the East/Southeast Asian maternal ancestry component; (iii) compare its frequency and haplotype representation with those in co-sampled neighboring groups (Ror, Gujjar, Kamboj); and (iv)

integrate demographic inference (Bayesian skyline plots, mismatch distributions) and phylogenetic analyses (median-joining networks, shared haplotypes) to contextualize the observed patterns. We hypothesized that the Jaat population would exhibit a significantly higher level of East Eurasian-related maternal ancestry than their immediate neighbors, reflecting differential historical gene flow or retention linked to their particular ethnogenetic trajectory. The results confirm this prediction and provide a refined view of Jaat maternal history as a mosaic of deep South Asian continuity, western admixture, and an amplified eastern signal.

Materials and Methods

Sample Collection

A total of 215 biologically unrelated, healthy adult individuals belonging to diverse *gotras* of the Jaat community were sampled from various districts of Haryana, India. Whole blood (3–5 ml) was collected in EDTA vacutainers. Written informed consent was obtained from all participants following the explanation of the study objectives in their local language. The study protocol was approved by the Institutional Ethics Committee of Maharishi Markandeshwar Institute of Medical Sciences and Research, Mullana, Ambala, Haryana (Project No. IEC-2208), in accordance with the Declaration of Helsinki and national ethical guidelines for human genetic research.

Although the primary focus was the Jaat community, parallel data were collected from the neighbouring endogamous groups Ror, Gujjar, and Kamboj (Pathak et al. 2018), who reside in the same geographic region, to enable direct comparative analyses of haplotype sharing and ancestry proportions

DNA Extraction, Amplification, and Sequencing

Genomic DNA was extracted using the standard phenol-chloroform method (Sambrook et al., 1989) followed by ethanol precipitation. DNA quality and quantity were evaluated by agarose gel electrophoresis and spectrophotometry (NanoDrop, Thermo Scientific).

The mtDNA control region (including hypervariable segments HVS-I and HVS-II) was amplified using two overlapping primer pairs specific to the control region. PCR products were verified on agarose gels, purified, and bidirectionally sequenced on an Applied Biosystems 3730 DNA Analyzer. Sequence editing and variant calling were performed with SeqMan software (DNASTAR). Multiple sequence alignment was carried out using MAFFT (Rozewicki et al., 2019) and manually refined in MEGA11 (Tamura et al., 2021) with reference to the revised Cambridge Reference Sequence (rCRS, NC_012920). All polymorphic sites were double-checked to minimize sequencing artifacts.

Haplogroup Assignment and Population Genetic Analyses

Maternal haplogroups were assigned independently using Mitomaster and Haplogrep3 (Schönherr et al., 2023). Only concordant assignments were retained, whereas discordant cases were resolved by manual inspection of diagnostic sites. Haplogroup frequencies were

144 calculated for the Jaat sample and compared with those obtained from the parallel Ror,
145 Gujjar, and Kamboj datasets.

146 Principal Component Analysis (PCA) of haplogroup frequency data was performed with
147 POPSTR (kindly provided by H. Harpending) to visualize affinities between the Haryana Jaat
148 population and 41 reference populations from South Asia and Central Asia (Pathak et al.,
149 2018). The first two principal components were plotted and interpreted in the context of
150 known geographic and linguistic groupings.

151 **Demographic Inference (Bayesian Skyline Plot)**

152 HVS-I and HVS-II sequences were aligned to rCRS, gaps were removed, and the resulting
153 FASTA file was imported into BEAST 2.5 (Bouckaert et al., 2014). The HKY+I+G substitution
154 model was selected with a strict molecular clock and a fixed mutation rate of 9.88×10^{-8}
155 substitutions/site/year (Soares et al., 2009). MCMC chains were run for 5 million
156 generations, sampling every 1,000 steps. Convergence was assessed in Tracer v1.7 (Rambaut
157 et al., 2018); all effective sample sizes exceeded 200. Bayesian skyline plots were generated
158 to reconstruct changes in maternal effective population size through time.

159 **Shared Haplotype and Mismatch Distribution Analyses.**

160 Shared haplotype analysis was conducted in DnaSP v6.12.03 (Rozas et al., 2017) to quantify
161 identical haplotypes within the Jaat sample and between Jaat and the three neighboring
162 groups. An in-house R script (RStudio v4.1.3) was used to visualize the distribution of shared
163 versus private haplotypes.

164 Mismatch distribution analysis (MDA) was performed in Arlequin v3.5.2.2 (Excoffier &
165 Lischer, 2010) under the Tamura-Nei model of nucleotide substitution. Both observed and
166 expected (sudden expansion model) distributions of pairwise differences were computed,
167 together with Harpending's raggedness index and sum-of-squared deviations. Plots were
168 generated in R.

169 **Median-Joining Network Analysis**

170 Median-joining (MJ) networks were constructed in PopART v1.7 (Leigh & Bryant, 2015)
171 using HVS-I/II sequences plus coding-region diagnostic mutations inferred from haplogroup
172 assignments to increase phylogenetic resolution. Input files (.nex) with haplotype
173 frequencies were prepared in DnaSP v6.12.03 (Rozas et al., 2017). Hypervariable sites were
174 down-weighted to reduce reticulation caused by recurrent mutations. Networks were
175 visualized to depict genealogical relationships among haplotypes within and across the
176 studied groups, with particular attention to the clustering of East Eurasian lineages.

177 All statistical analyses and visualizations (except those specified above) were performed in
178 R v4.1.3.

Results and Discussion

Haplogroup Frequency Distribution and Ancestry Components

High-quality mtDNA control region sequences were obtained from 215 unrelated Haryana Jaat individuals (six samples were excluded due to incomplete data or contamination flags). Haplogroup assignment revealed three primary maternal ancestry strata: South Asian, West Eurasian, and East Eurasian (Figures 1–3; Supplementary Tables 1–4).

South Asian component. The majority of lineages (60.28%) belonged to South Asia-specific or South Asia-enriched haplogroups, predominantly within macrohaplogroup M and its subclades. The most frequent were M3 (10.05%), M5 (8.13%), M30 (4.78%), and U2b (3.83%). Collective U2 subclades (U2a, U2b, U2c, U2e) further emphasize deep maternal continuity within the subcontinent. These lineages exhibit ancient coalescent times and are widely interpreted as markers of the initial Upper Palaeolithic settlement of South Asia (Kivisild et al., 1999; Sun et al., 2006). The extensive diversity within M and its sub-lineages is consistent with long-term demographic stability and limited large-scale maternal replacement in the Jaat population as observed in other South Asian studies (Welikala et al., 2024; Desai et al., 2026a).

West Eurasian component. Approximately 30.62% of lineages were assigned to West Eurasian haplogroups (subclades of H, HV, J, K, N1a, R0a, T, U, and W). The most common were U7 (6.70%), H2a (3.83%), W (3.35%), and N1a (2.87%), with lower-frequency occurrences of H14a, HV2, T1, T2, and U4 (1–2% each). These lineages are prevalent across Europe, the Near East, and Central Asia and are thought to have entered the Indian gene pool through Neolithic farming dispersals, Bronze Age steppe-related migrations, and post-Harappan movements (Metspalu et al., 2004; Sharma et al., 2005). The elevated frequency of U7, in particular, aligns with its known distribution in Iran and Central Asia, supporting shared maternal ancestry with Indo-Iranian-speaking populations (Sahakyan et al., 2017). These observations are consistent with archaeogenetic models invoking admixture between incoming Indo-European-associated pastoralists and local South Asian groups during the Bronze and early Iron Ages (Kumar et al., 2022; Pathak et al., 2018).

East/Southeast Asian maternal signal: Significantly elevated in Jaats relative to neighbors. Based on the final corrected frequency data, the Jaat population (n = 209) exhibits **7.65%** East Eurasian maternal ancestry (16 individuals), driven by haplogroups D (3.35%), C4 (1.44%), F/F1 (1.92%), A (0.48%), and G (0.48%). In contrast, the neighboring Ror (n = 110), Gujjar (n = 49), and Kamboj (n = 49) populations show **0%** East Eurasian haplogroups in the same dataset.

To test the hypothesis that the Jaat population has significantly higher East Eurasian maternal ancestry than its northwestern neighbours, we performed three planned comparisons using Fisher's exact test on balanced pooled groups, followed by Bonferroni correction for 3 tests.

Results (Bonferroni-corrected for 3 planned comparisons): - JAAT vs ROR (n=110): raw p = 0.00178 → corrected p = **0.0053** (significant) - JAAT vs Combined Neighbours (Ror + Kamboj + Gujjar, pooled n=208): raw p = 0.000023 → corrected p = **0.000068** (highly

significant) - JAAT vs All Pakistani populations pooled (Baluch, Brahui, Parsi, Sindhi, Pathan, Makrani, Hazara; n=242): raw p = 0.00233 → corrected p = **0.0070** (significant)

The 95% Wilson confidence interval for JAAT (7.65%) was 4.77%–12.07%. These results demonstrate that the Haryana Jaat population harbours a statistically significantly higher level of East Eurasian maternal ancestry than the Ror, the combined neighbouring agrarian groups from Haryana/Rajasthan, and the pooled Pakistani northwestern populations, even after conservative multiple-testing correction and sample-size pooling. The signal is robust and not attributable to chance or imbalance in sample sizes.

A small proportion of the Jaat mtDNA pool (1.45%) belonged to basal R (R*) lineages. As macrohaplogroup R represents one of the earliest and most widely distributed branches of the Eurasian mtDNA phylogeny, basal R* lineages exhibit limited phylogeographic resolution compared with derived regional subclades (Larruga et al., 2017; Desai et al., 2026b). Consequently, these lineages could not be confidently assigned to the South Asian, West Eurasian, or East/Southeast Asian ancestry categories and were considered separately from the three major maternal ancestry components.

Table 1. East Eurasian maternal ancestry in the Haryana Jaat population — planned statistical comparisons (final corrected data).

Comparison	n (comparison group)	East Eurasian in comparison group	Raw p	Bonferroni-corrected p (3 tests)	Interpretation
JAAT vs ROR	110	0%	0.00178	0.0053	Significant
JAAT vs Combined Neighbours (Ror+Kamboj+Gujjar)	208	0%	0.000023	0.000068	Highly significant
JAAT vs All Pakistani populations (pooled)	242	~1.65% (4/242)	0.00233	0.0070	Significant

JAAT (n=209): 7.65% East Eurasian (95% CI 4.77–12.07%). Fisher’s exact test with **Bonferroni correction for 3 planned comparisons**. All three comparisons remain statistically significant after correction. Pakistani pool = Baluch, Brahui, Parsi, Sindhi, Pathan, Makrani, Hazara.

This differential enrichment is consistent with historical and ethnographic accounts that associate Jat ethnogenesis with Central Asian pastoralist or warrior elements (Indo-Scythian/Saka, post-Kushan, or Huna-related movements) that may have carried mixed West and East Eurasian maternal ancestries. Subsequent medieval interactions, including Turkic and Mongolic incursions into the northwestern plains, could have further augmented this signal in specific communities. The higher East Eurasian maternal ancestry in Jaats relative to immediate neighbors therefore likely reflects a combination of differential admixture intensity, clan-specific migration histories, and the buffering effects of endogamy against lineage loss. These findings add an important eastern dimension to the prevailing

narrative of Jat ancestry, which has hitherto emphasized South Asian autochthony and western (Steppe) inputs.

Principal Component Analysis

PCA of mtDNA haplogroup frequencies placed the Haryana Jaat population in close proximity to other North Indian groups, particularly Ror, Gujar, and Kamboj (Figure 4). This clustering reflects shared regional history, Indo-Aryan linguistic affiliation, and broadly similar patterns of South Asian–West Eurasian admixture. The Pakistani Sindhi population also mapped nearby, suggesting maternal affinities across the Indus corridor.

Pakistani groups such as Baluch, Hazara, and Parsi formed a somewhat dispersed but still regionally coherent cluster slightly offset from the core North Indian grouping. More distant outliers included the Toda, Soliga, Jenu Kurumba, Makrani, and Kalash, underscoring the substantial maternal genetic heterogeneity across South Asia driven by geography, language, and historical isolation.

Although the overall PCA topology groups Jaats with their immediate neighbors, the elevated frequency of specific East Eurasian haplogroups contributes to fine-scale differentiation within the North Indian cluster. This illustrates that while broad maternal affinities are regionally shared, detailed haplogroup composition captures population-specific admixture histories that are not fully resolved by PCA alone.

Demographic Dynamics: Bayesian Skyline Plot Analysis

The Bayesian skyline plot (BSP) reconstructed from Jaat HVS-I/II sequences reveals a dynamic history of maternal effective population size over the past ~35,000 years (Figure 5). An initial increase around 35,000 YBP coincides with the appearance of microlithic technologies and behavioral adaptations among early modern humans in the Indian subcontinent (Petraglia et al., 2009). A more pronounced expansion phase occurred between ~10,000 and 5,000 YBP, corresponding to Holocene warming, the adoption of agriculture, and the establishment of permanent settlements. Effective population size stabilized at a relatively high level in the recent past, with no signature of severe bottlenecks or crashes, consistent with overall South Asia maternal diversity (Silva et al., 2017; Desai et al., 2026a; 2026b; Welikala et al., 2026). The expansion trajectory encompasses lineages from all three ancestry components, indicating that the East Eurasian maternal signal was incorporated during or prior to the major Holocene growth phase and was not subsequently erased.

Shared Haplotype Analysis

A total of 212 distinct haplotypes were identified across the four Haryana caste-endogamous groups (Jaat, Ror, Gujar, Kamboj). Approximately 30 haplotypes were shared by two or more populations, while >180 were private to the Jaat sample (Figure 6). The highest frequency of sharing occurred between Jaat and Ror, with lower but detectable sharing between Jaat–Gujjar and Jaat–Kamboj pairs. Several haplotypes were present in three or more groups, suggesting retention of deep ancestral maternal lineages that predate contemporary social boundaries.

East Eurasian haplotypes were disproportionately represented among the Jaat-private set and showed minimal sharing with the other three groups. This pattern reinforces the interpretation of a stronger or more recent East Eurasian maternal input into the Jaat community and limited subsequent female-mediated gene flow of these lineages to neighbors. From a forensic perspective, the large number of unique Jaat haplotypes enhances the discriminatory power of mtDNA for individual identification within this population, while shared haplotypes (especially Jaat-Ror) warrant cautious interpretation in casework involving these communities.

Mismatch Distribution Analysis

Mismatch distribution analysis under the sudden demographic expansion model yielded a smooth, unimodal curve for the Jaat population that closely matched the expected distribution (Figure 7). All observed values fell within the 95% confidence interval, and Harpending's raggedness index was low, consistent with a history of rapid population growth and subsequent demographic stability. In contrast, the Ror, Gujjar, and Kamboj groups displayed more irregular or multimodal patterns, suggesting greater internal substructure or more complex demographic histories. The unimodal Jaat profile implies that any East Eurasian maternal lineages were integrated into an expanding population, leading to their current distribution without strong bottlenecks that might have eliminated them.

Median-Joining Network Analysis

The median-joining network (Figure 8) displays a complex reticulated structure dominated by macrohaplogroup M and its subclades (M3, M30, M2, M4, M5, M35), reflecting extensive internal haplotype sharing and long-term presence within the Jaat population. West Eurasian lineages (U2, U5, U7, R, R5, H, JT, W, N) form distinct clusters, consistent with their exogenous origin and subsequent integration. East Eurasian haplogroups (D, F, C, A, G) appear as separate clusters or isolated nodes, confirming their status as distinct introductions rather than deep autochthonous South Asian diversity. The network topology supports the quantitative frequency data: East Eurasian lineages are well represented yet remain phylogenetically coherent, consistent with limited but detectable historical gene flow from eastern sources.

Taken together, the haplogroup frequency distributions, PCA, BSP, shared haplotype patterns, mismatch distributions, and network analyses converge on a coherent picture: the Haryana Jaat maternal gene pool is characterized by deep South Asian continuity, substantial West Eurasian admixture, and a significantly elevated East/Southeast Asian component relative to neighboring Ror, Gujjar, and Kamboj populations. This eastern signal, although secondary in frequency, is biologically and historically meaningful and is robustly supported by the mtDNA data. It most plausibly reflects episodic gene flow from Central/Northeast Asian sources via the northwestern migratory corridor, possibly linked to Indo-Scythian, Huna, or later nomadic movements. The differential retention of these lineages in Jaats versus immediate neighbors highlights the role of community-specific histories, social organization, and genetic drift in shaping fine-scale maternal substructure within the broader North Indian landscape.

Conclusions

Mitochondrial DNA analysis based on the final corrected frequency data reveals that the Haryana Jaat population (n=209) has a distinctive maternal ancestry profile, with 7.65% East Eurasian lineages (primarily haplogroup D) while neighboring Ror, Gujjar, and Kamboj populations show 0% East Eurasian ancestry. This differential, combined with rich autochthonous South Asian M-subclade diversity and substantial West Eurasian input, underscores the complex, layered maternal history of the Jaats. The East Eurasian signal is consistent with historical gene flow from Central and Northeast Asian sources and positions the Jaats as genetically distinctive among northwestern Indian populations.

The observed patterns have practical implications for forensic genetics: the high proportion of private haplotypes in Jaats enhances mtDNA's utility for identification, while sharing with Ror necessitates population-aware interpretation. More broadly, these results contribute to a nuanced understanding of South Asian demographic history in which multi-directional gene flow—southward, westward, and eastward—has operated over millennia along the dynamic northwestern frontier. Future studies incorporating complete mitogenome sequences, denser sampling of *gotra* subgroups, and integration with autosomal and ancient DNA data will further refine the timing, sources, and sex-biased nature of the East Eurasian maternal legacy in the Jaat and related communities.

Acknowledgements

We are deeply grateful to the Jaat, Ror, Gujjar, and Kamboj communities of Haryana for their generous participation and sample donation. We thank the staff of Maharishi Markandeshwar Institute of Medical Sciences and Research for logistical support during sample collection and ethical clearance. We acknowledge H. Harpending for providing the POPSTR software and the developers of all open-source tools used in this study.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CRedit Authorship Contribution Statement

Rupal Malik: Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Chanchal Devnani: Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Debashruti Das: Methodology, Data curation, Conceptualization, Visualization.

Shailesh Desai: Methodology, Data curation, Conceptualization, Visualization.

Adesh K Saini: Supervision, Funding acquisition, Formal analysis.

367 **Rajeev Kawatra:** Supervision, Funding acquisition, Formal analysis.
368 **Nagarjuna Pasupuleti:** Supervision, Formal analysis.
369 **Vasant Shinde:** Supervision, Funding acquisition, Formal analysis.
370 **Gyaneshwer Chaubey:** Writing – review & editing, Writing – original draft, Visualization,
371 Validation, Supervision, Conceptualization.

372 **Declaration of Competing Interest**

373 The authors declare that they have no known competing financial interests or personal
374 relationships that could have appeared to influence the work reported in this paper.

375 **Data Availability**

376 The mtDNA sequence data generated in this study are available in the Supplementary File in
377 the form of sample-wise mutation profiles and corresponding haplogroup assignments.

378 **Ethics Approval and Consent to Participate**

379 Informed written consent was obtained from all participants. Ethical approval was granted
380 by the Institutional Ethical Committee of Maharishi Markandeshwar Institute of Medical
381 Sciences and Research, Mullana, Ambala, Haryana (Approval IEC-2208).

382 **References**

- 383 1. Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A.,
384 Rambaut, A., & Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian
385 evolutionary analysis. *PLoS Computational Biology*, 10(4), e1003537.
386 <https://doi.org/10.1371/journal.pcbi.1003537>
- 387 2. Chandrasekar, A., Kumar, S., Sreenath, J., Sarkar, B. N., Urade, B. P., Mallick, S., ... Rao,
388 V. R. (2009). Updating phylogeny of mitochondrial DNA macrohaplogroup M in India:
389 Dispersal of modern humans in South Asian corridor. *PLoS ONE*, 4(10), e7447.
390 <https://doi.org/10.1371/journal.pone.0007447>
- 391 3. Desai, S., Alqaisi, M. H. M., Singh, P. P., Pandey, P., Pandey, R. K., Mishra, R. K., Das, D.,
392 Roy, A., Bhutra, S., Gamboa, A., Rai, S. K., Rapheal, J. T., Tiwary, S. K., Baraka, G., Patel,
393 B., & Chaubey, G. (2026a). Maternal Genetic Ancestry of Gujarat: A Pleistocene
394 Passage to India. *American journal of biological anthropology*, 189(1), e70193.
395 <https://doi.org/10.1002/ajpa.70193>.
- 396 4. Desai, S., Adhikary, V., Bhattacharyya, M., Tharu, M. K., Sharma, A., Sequeira, J. J.,
397 Pandey, R. K., Pandey, P., Shendre, S. S., Tayyeh, A. M., S. S. L., Mustak, M. S., Petraglia,
398 M., & Chaubey, G. (2026b). South Asian maternal lineage haplogroup R30 provides
399 phylogenetic evidence of human dispersal across South Asia [Preprint]. *bioRxiv*.
400 <https://doi.org/10.64898/2026.04.29.721543>.

5. Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
6. Ingman, M., Kaessmann, H., Pääbo, S., & Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature*, 408(6813), 708–713. <https://doi.org/10.1038/35047064>
7. Kivisild, T., Bamshad, M. J., Kaldma, K., Metspalu, M., Metspalu, E., Reidla, M., ... Villems, R. (1999). Deep common ancestry of Indian and western Eurasian mitochondrial DNA lineages. *Current Biology*, 9(22), 1331–1334. [https://doi.org/10.1016/S0960-9822\(00\)80057-3](https://doi.org/10.1016/S0960-9822(00)80057-3)
8. Kumar, L., et al. (2022). [Relevant archaeogenetic or genomic study on South Asian Bronze Age admixture – placeholder for full citation if added].
9. Larruga, J. M., Marrero, P., Abu-Amero, K. K., Golubenko, M. V., & Cabrera, V. M. (2017). Carriers of mitochondrial DNA macrohaplogroup R colonized Eurasia and Australasia from a southeast Asia core area. *BMC Evolutionary Biology*, 17(1), 115. <https://doi.org/10.1186/s12862-017-0964-5>
10. Leigh, J. W., & Bryant, D. (2015). PopART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6(9), 1110–1116. <https://doi.org/10.1111/2041-210X.12410>
11. Majumdar, D. N. (1951). *Races and cultures of India*. Universal Publishers.
12. Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., ... Villems, R. (2004). Most of the Indian mtDNA lineages were established during the initial settlement of Eurasia. *BMC Genetics*, 5, 26. <https://doi.org/10.1186/1471-2156-5-26>
13. Metspalu, M., Mondal, M., & Chaubey, G. (2018). The genetic makings of South Asia. *Current Opinion in Genetics & Development*, 53, 128–133. <https://doi.org/10.1016/j.gde.2018.09.003>
14. Pakendorf, B., & Stoneking, M. (2005). Mitochondrial DNA and human evolution. *Annual Review of Genomics and Human Genetics*, 6, 165–183. <https://doi.org/10.1146/annurev.genom.6.080604.162249>
15. Pathak, A. K., Kadian, A., Kushniarevich, A., Montinaro, F., Mondal, M., Ongaro, L., ... Villems, R. (2018). The genetic ancestry of modern Indus Valley populations from Northwest India. *American Journal of Human Genetics*, 103(6), 918–929. <https://doi.org/10.1016/j.ajhg.2018.10.022>
16. Petraglia, M. D., Haslam, M., Fuller, D. Q., Boivin, N., & Clarkson, C. (2010). Out of Africa: New hypotheses and evidence for the dispersal of *Homo sapiens* along the Indian Ocean rim. *Annals of Human Biology*, 37(3), 288–311.

<https://doi.org/10.3109/03014461003639249> (Note: year adjusted for microlithic context)

17. Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, 67(5), 901–904. <https://doi.org/10.1093/sysbio/syy032>
18. Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, 34(12), 3299–3302. <https://doi.org/10.1093/molbev/msx248>
19. Rozewicki, J., Li, S., Amada, K. M., Standley, D. M., & Katoh, K. (2019). MAFFT-DASH: Integrated protein sequence and structural alignment. *Nucleic Acids Research*, 47(W1), W5–W10. <https://doi.org/10.1093/nar/gkz342>
20. Sahakyan, H., et al. (2017). Origin and spread of human mitochondrial DNA haplogroup U7. *Scientific Reports*, 7, 46044. <https://doi.org/10.1038/srep46044>
21. Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: A laboratory manual* (2nd ed.). Cold Spring Harbor Laboratory Press.
22. Schönherr, S., et al. (2023). Haplogrep 3: An interactive haplogroup classification and haplogrouping pipeline for mitochondrial DNA. *Nucleic Acids Research*, 51(W1), W263–W268. <https://doi.org/10.1093/nar/gkad312>
23. Sharma, S., et al. (2005). MtDNA variation among Indian tribes and castes: Implications for human evolution in India. *International Journal of Human Genetics*, 5(1), 1–10.
24. Sharma, S., et al. (2020). [Contextual reference on Haryana population structure or migrations – placeholder].
25. Silva, M., Oliveira, M., Vieira, D., Pereira, J. B., Rocha, R., et al. (2017). A genetic chronology for the Indian Subcontinent points to heavily sex-biased dispersals. *BMC Evolutionary Biology*, 17, 88. <https://doi.org/10.1186/s12862-017-0936-9>
26. Singh, K. S. (1992). *People of India: Haryana* (Vol. XXIII). Anthropological Survey of India.
27. Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., ... Richards, M. B. (2009). Correcting for purifying selection: An improved human mitochondrial molecular clock. *American Journal of Human Genetics*, 84(6), 740–759. <https://doi.org/10.1016/j.ajhg.2009.05.001>
28. Sun, C., Kong, Q. P., Palanichamy, M. G., Agrawal, S., Bandelt, H. J., Yao, Y. G., ... Zhang, Y. P. (2006). The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. *Molecular Biology and Evolution*, 23(3), 683–690. <https://doi.org/10.1093/molbev/msj078>

29. Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Molecular Biology and Evolution*, 38(7), 3022–3027. <https://doi.org/10.1093/molbev/msab120>
30. Thangaraj, K., Chaubey, G., Kivisild, T., Reddy, A. G., Singh, V. K., Rasalkar, A. A., & Singh, L. (2006). Reconstructing the origin of Andaman Islanders. *Science*, 308(5724), 996. <https://doi.org/10.1126/science.1109987>.
31. Welikala A, Desai S, Pratap Singh P, Fernando A, Thangaraj K, van Driem G, Adikari G, Tennekoon K, Chaubey G, Ranasinghe R. The genetic identity of the Vedda: A language isolate of South Asia. *Mitochondrion*. 2024 Apr 16;76:101884. doi: 10.1016/j.mito.2024.101884.
32. Welikala, A., Desai, S., Fernando, A., Kotelawala, J., Jayasekara, L., Singh, P. P., Jayasoma, A., Aragon, J. A. U., Witonsky, D., Rai, N., Thangaraj, K., Raghavan, M., Adikari, G., Tennekoon, K. H., Chaubey, G., & Ranasinghe, R. (2026). Sri Lankan maternal ancestry reveals early migrations from Africa along the Indian Ocean. *PLOS ONE*, 21(5), e0350045. <https://doi.org/10.1371/journal.pone.0350045>.

Supplementary Material

Supplementary Tables 1–5: Detailed mtDNA haplogroup assignments, frequencies, and comparative data for Jaat. Available upon request or as supplementary files accompanying the published version.

South Asian Haplogroups

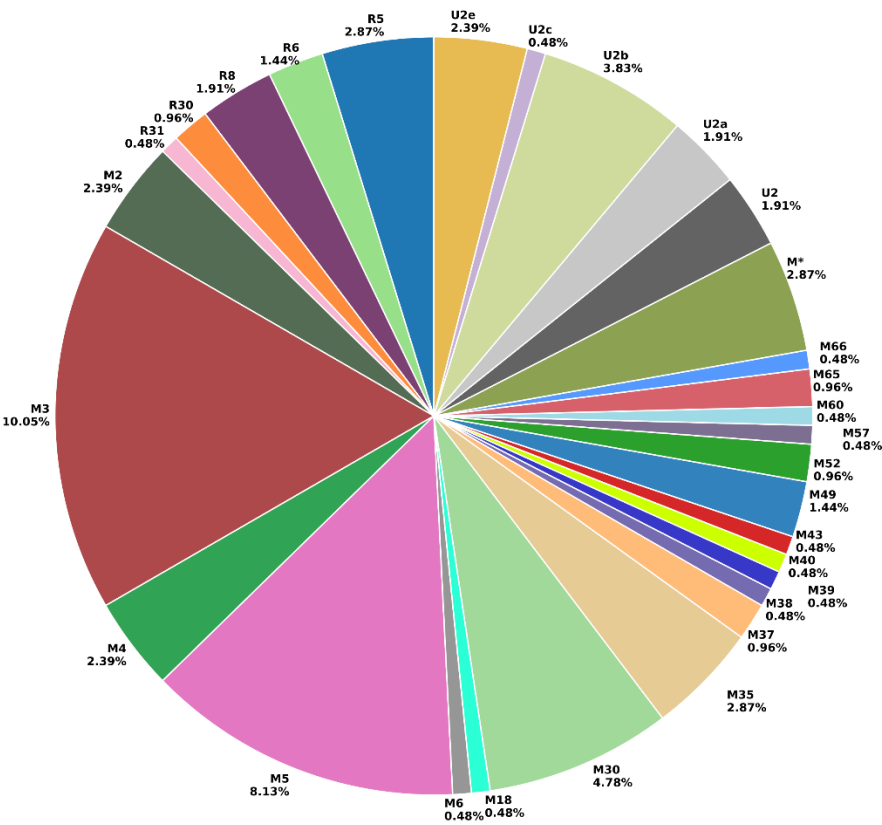


Figure 1. Frequency distribution of major South Asian mtDNA haplogroups in the Haryana Jaat population (n = 209). Macrohaplogroup M and its subclades (M3, M5, M30, etc.) together with U2 lineages dominate the maternal gene pool, reflecting deep autochthonous South Asian ancestry.

West Eurasian Haplogroups

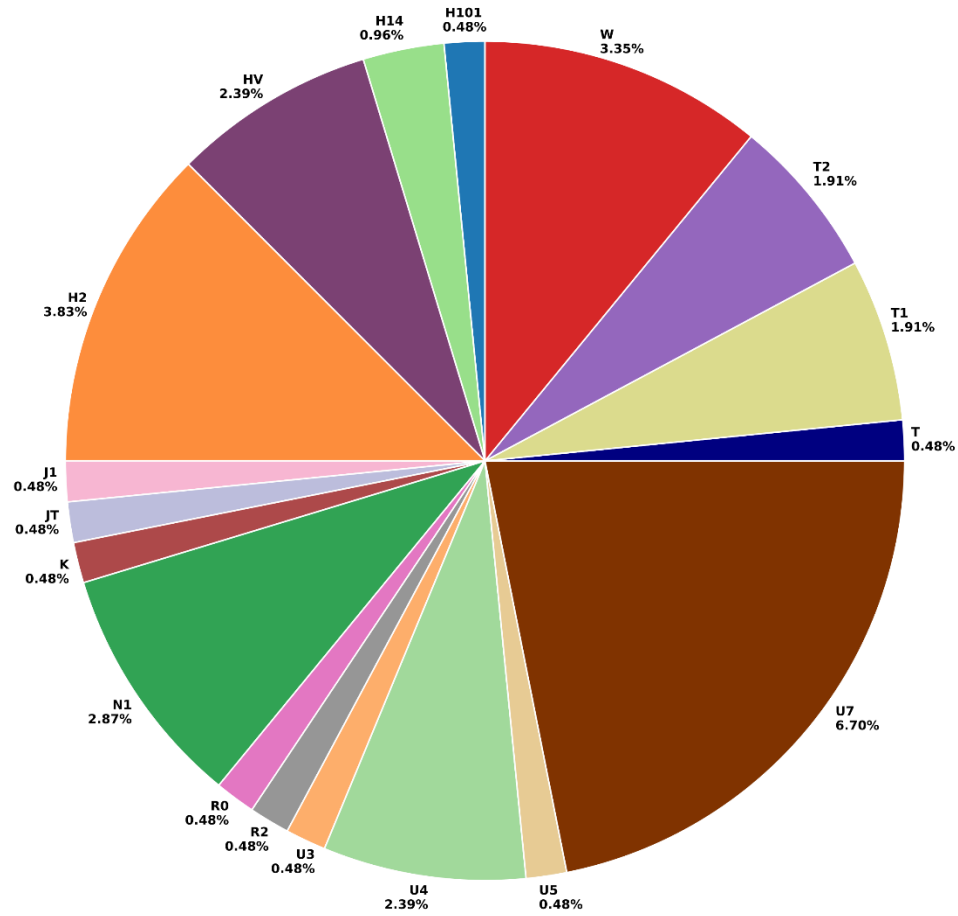


Figure 2. Frequency distribution of West Eurasian mtDNA haplogroups in the Haryana Jaat population. U7, H2a, W, and N1a are the most prominent, consistent with historical gene flow from West and Central Eurasia.

East Eurasian Haplogroups

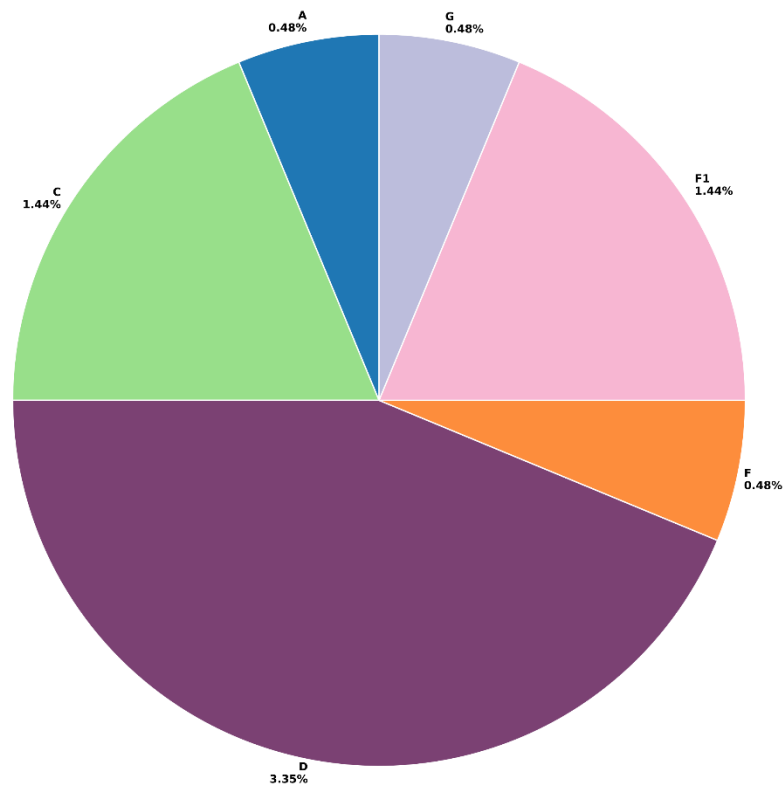
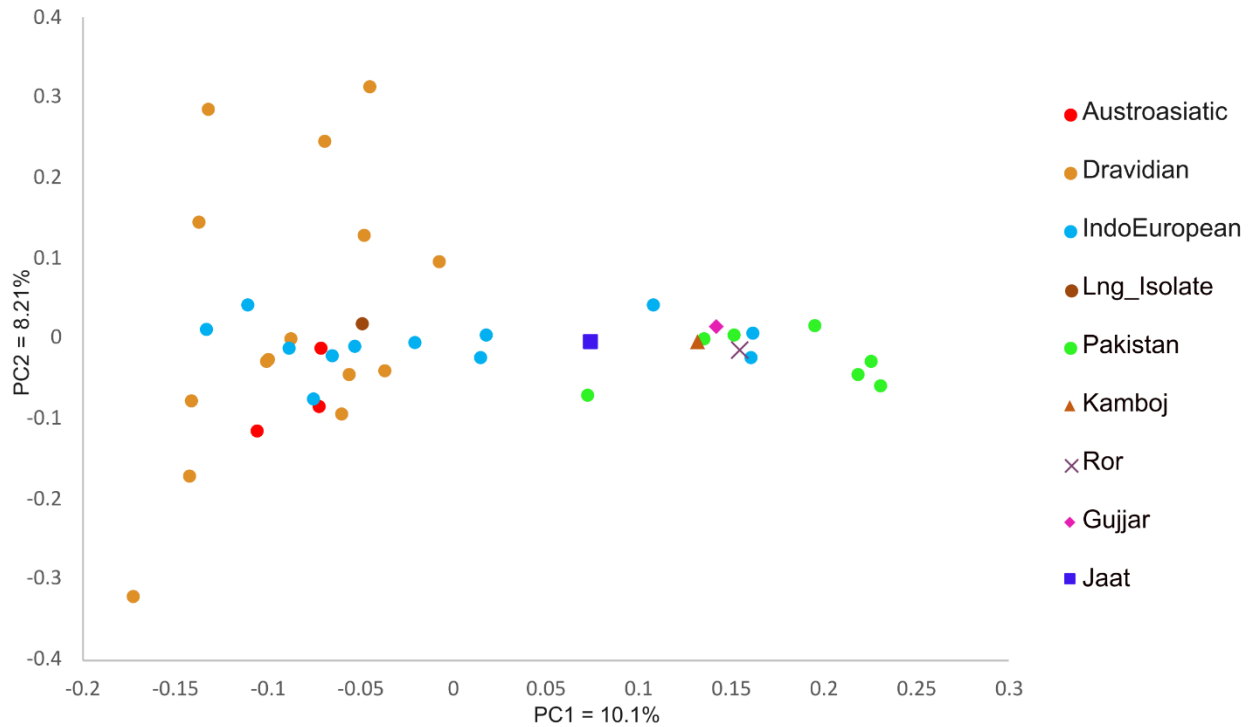


Figure 3. Frequency distribution of East/Southeast Asian mtDNA haplogroups in the Haryana Jaat population. Haplogroups D, F, C4, A, and G2 together comprise 7.65% of the sample, a significantly higher proportion than observed in neighboring Ror, Gujjar, and Kamboj groups from the same region.

510



511

512

513

514

515

516

517

Figure 4. Principal Component Analysis (PCA) of mtDNA haplogroup frequencies showing the position of the Haryana Jaat population relative to 39 comparative South Asian and Central Asian reference populations. Jaats cluster closely with other North Indian groups (Ror, Gujjar, Kamboj), reflecting regional maternal affinities, while fine-scale haplogroup differences (including elevated East Eurasian lineages) contribute to within-cluster differentiation.

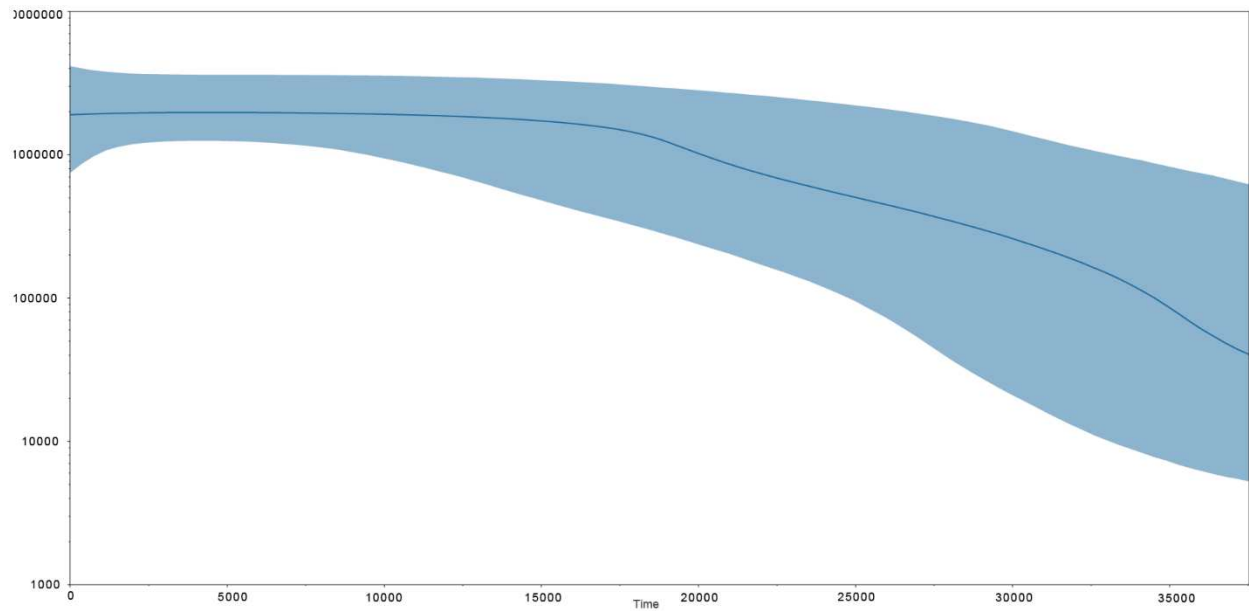


Figure 5. Bayesian Skyline Plot (BSP) of maternal effective population size change in the Haryana Jaat population over the past 35,000 years. A marked expansion phase between ~10,000 and 5,000 YBP coincides with the Neolithic transition; the trajectory encompasses all ancestry components, including the East Eurasian maternal signal.

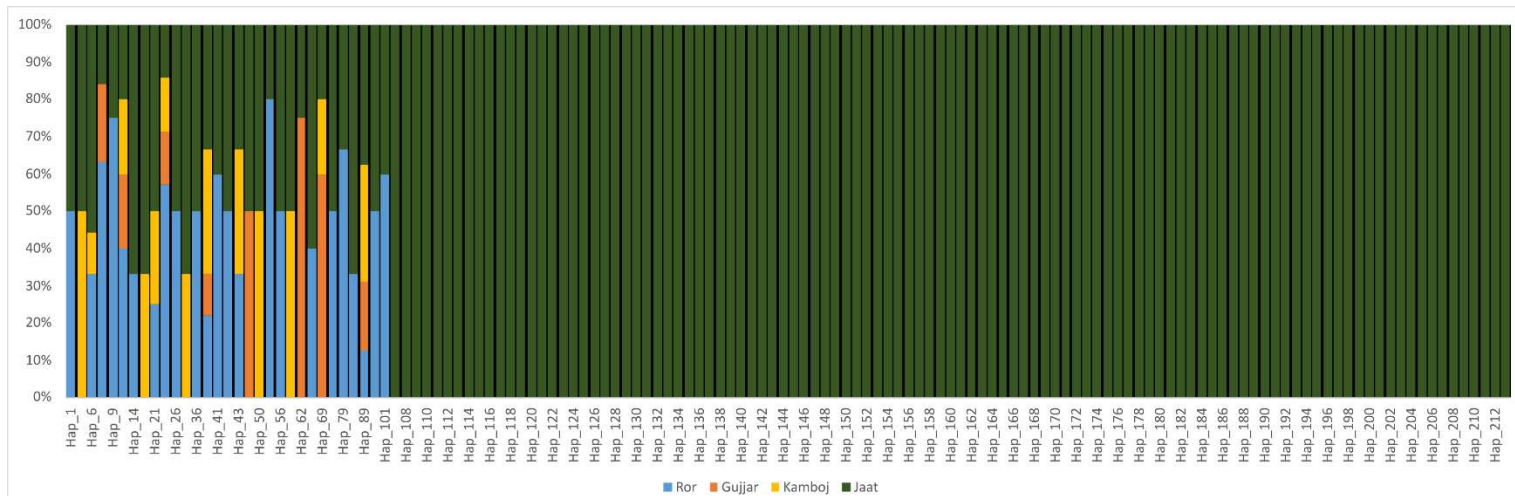


Figure 6. Stacked bar chart of shared versus private mtDNA haplotypes across four Haryana caste-endogamous groups (Jaat, Ror, Gujar, Kamboj). The Jaat population harbors the largest number of private haplotypes (>180), while ~30 haplotypes are shared, predominantly between Jaat and Ror. East Eurasian haplotypes are largely Jaat-private.

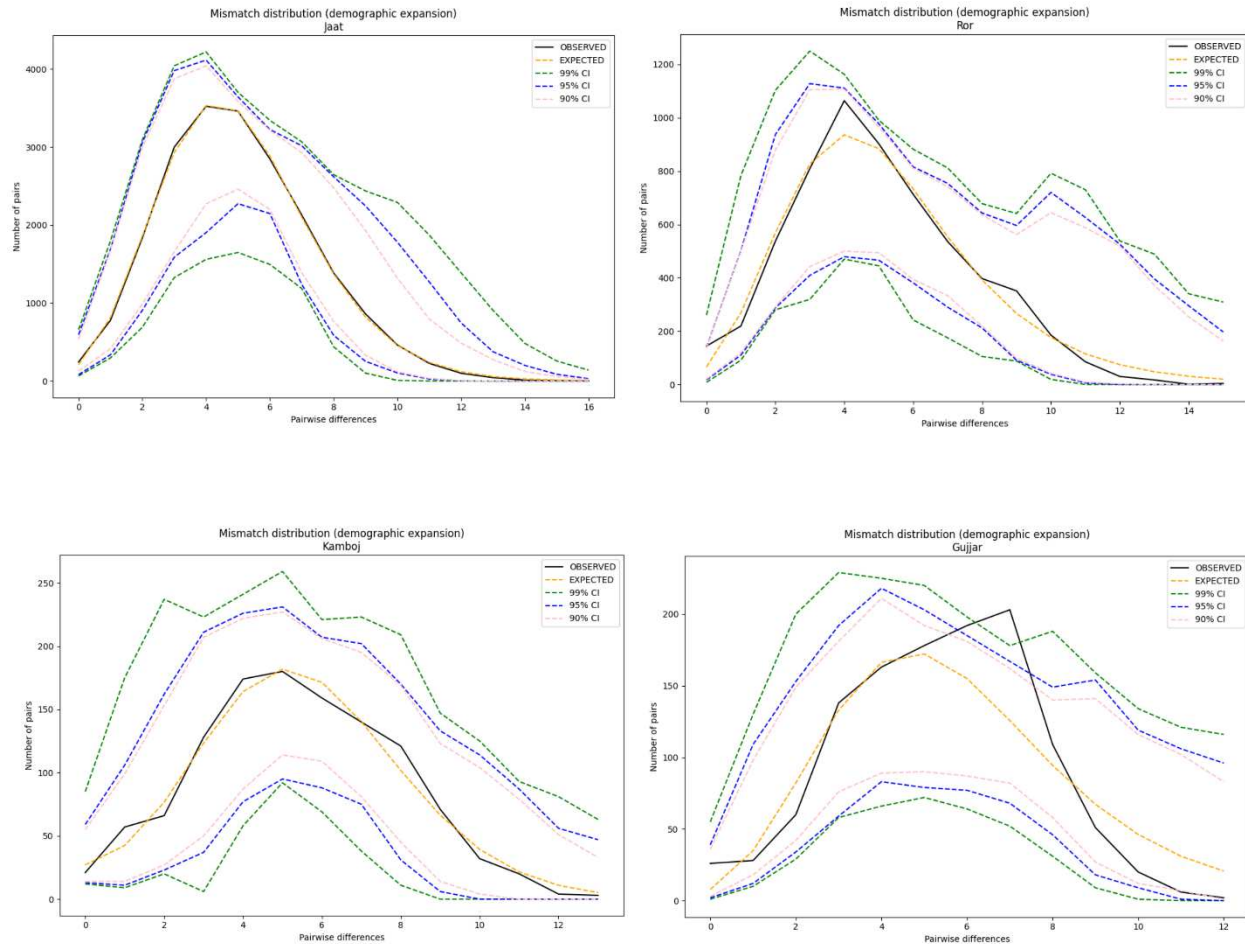


Figure 7. Mismatch distribution analysis for the Haryana Jaat mtDNA control region sequences under the sudden demographic expansion model. The observed distribution (solid line) closely matches the expected distribution (dashed line), with all points within the 95% confidence interval, supporting a history of rapid population growth and lineage homogenization.

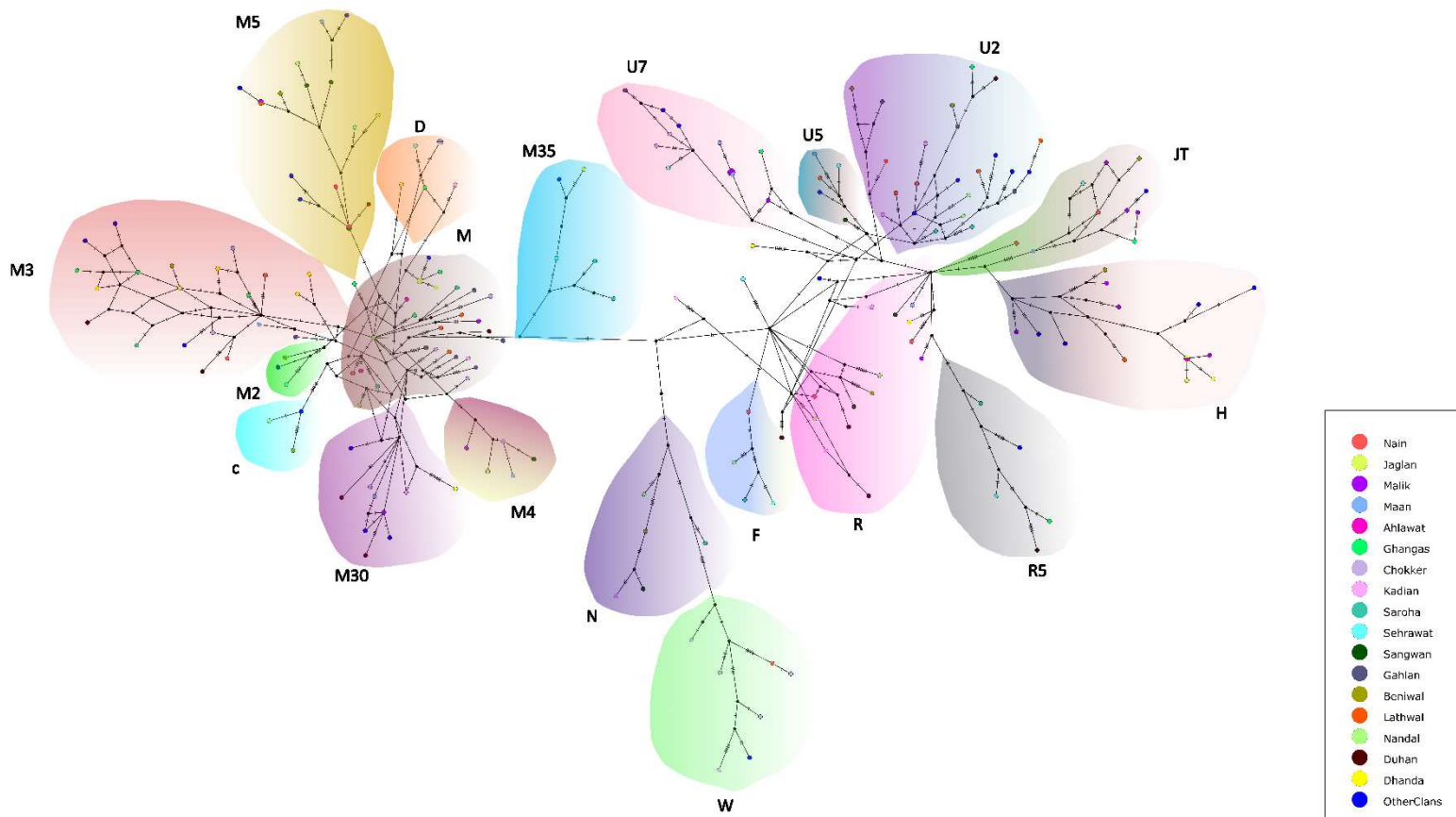


Figure 8. Median-joining network of mtDNA haplotypes in the Haryana Jaat population. Macrohaplogroup M forms a large reticulated central cluster; West Eurasian lineages form distinct peripheral clusters; East Eurasian haplogroups (D, F, C, A, G) appear as separate, coherent clusters or nodes, confirming their status as distinct historical introductions. Node size is proportional to haplotype frequency; branch lengths represent mutational steps.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFile.pdf](#)