Naif Arab University for Security Sciences
Arab Journal of Forensic Sciences and Forensic Medicine
المجلة العربية لعلوم الأدلة الجنائية والطب الشرعي
https://journals.nauss.edu.sa/index.php/AJFSFM

Revealing Y-STR Diversity of Koli Populations (Gujarat) by Studying 23 Y-STR Loci

الكشف عن التنوع في التتابعات القصيرة المتكررة (STRs) في كروموسوم واي (Y) لسكان كولي في منطقة غوجارات من خلال دراسة هذه التتابعات في 23 موقع وراثي على ذلك الكروموسوم

Sarthak Misra[1], Ulhas Gondhali[2], Deepesh Saini[3], Aditi Mishra[4]*
[1]Indira Gandhi Institute of Medical Sciences, Patna, India.
[2]John Jay College of Criminal Justice, City University of New York; Jindal Institute of Behavioural Sciences, O.P. Jindal Global University, Sonipat, India.
[3]Wildlife Forensic and Conservation Genetics Cell, Wildlife Institute of Forensic Science, India.
[4]Kristu Jayanti College, Bangalore, India.

Received 18 Aug. 2023; Accepted 13 Dec. 2023; Available Online 27 Dec. 2023.

Abstract

Genetic analysis of Y-STR loci is pivotal for forensic libraries and genetic analysis. The Koli population in Gujarat, India, however, lacks such genetic characterization. This study aims to develop an allele frequency database for 23 Y-STR loci in the Koli population, examining forensic parameters and assessing genetic connections with neighboring tribes.

A total of 153 unrelated Koli males were genotyped using the PowerPlex®Y23 multiplex commercial kit. We identified 117 distinct haplotypes. The Haplotype Diversity (HD) and Discrimination Capacity (DC) for the 23 Y-STR loci were 0.993 and 0.8034, respectively. DYS385b locus exhibited the highest allele variability (10 alleles), whereas DYS391, DYS389I, and DYS437 showed the least (4 alleles each). The highest Polymorphic Information Content (PIC) was observed in DYS385b (0.775), with the lowest in DYS391 (0.386). The dominant haplogroup R1a accounted for 45% of the

Keywords: Forensic science; Y-STR; Koli population; Haplotype diversity; Forensic parameters.

المستخلص

يُشكل التحليل الوراثي لمواقع التتابعات القصيرة المتكررة (STRs) في كروموسوم واي (Y) أهمية محورية في تكوين المحتوى الوراثي الجنائي والتحليل الوراثي لهذا الكروموسوم. ومع ذلك، يفتقر سكان كولي في منطقة غوجارات الهندية إلى مثل هذا التصنيف الوراثي. تهدف هذه الدراسة إلى تطوير قاعدة بيانات تكرارات الأليل في 23 موقع وراثي في كروموسوم واي (Y) لسكان كولي، مع فحص المعايير الجنائية وتقييم العلاقات الوراثية مع القبائل المجاورة.

تم عمل تحليل وراثي لـ 153 شخص من الذكور البالغين غير الأقارب من سكان كولي باستخدام مجموعة تكثير الحمض النووي (Power-Plex®Y23) التجارية. أظهرت النتائج 117 نمطاً وراثياً أحاديا فريداً. كان تنوع النمط الوراثي (HD) وقدرة التمييز (DC) لـ 23 موقع وراثي على كروموسوم واي (Y-STR) يساوي 0.993 و 0.8034 على التوالي. أظهر موقع DYS385b أعلى تباين للأليلات (10 أليلات)، بينما أظهر كل من المواقع التالية؛ DYS391 و DYS389I و DYS437 أقل تباين للأليلات (4 أليلات لكل منها)، كما لوحظ أعلى محتوى معلومات متعدد الأشكال (PIC) في الموقع الوراثي DYS385b بقيمة بلغت (0.775)، وأقلها في الموقع الوراثي DYS391 بقيمة بلغت (0.386). شكلت الفصيلة السائدة (R1a)

الكلمات المفتاحية: علوم الأدلة الجنائية، التتابعات القصيرة المتكررة في كروموسوم واي، سكان كولي، تنوع النمط الوراثي الأحادي، المعايير الجنائية.

* Corresponding Author: Aditi Mishra
Email: aditi@kristujayanti.com

doi: 10.26735/RQGQ9805

population. Comparative analysis with other Indian populations from YHRD revealed two distinct clusters, placing the Koli population in cluster 2, indicating significant genetic similarity within this group.

This inaugural study of Y-STRs in the Koli population demonstrates the utility of the Y23 kit in male identification, highlighted by substantial haplotype diversity and discrimination capacity.

النسبة الأعلى (45 %) من السكان. كشف التحليل المقارن لسكان غوجارات مع المجموعات السكانية الهندية الأخرى في خلال موقع بيانات (YHRD) عن مجموعتين منفصلتين، حيث وضِعت قبيلة كولي في المجموعة الثانية، مما يشير إلى تشابه جيني كبير داخل هذه المجموعة.

تُظهر هذه الدراسة التي تُعد الأولى من نوعها لسكان كولي للمواقع الوراثية (Y-STR) فائدة مجموعة تكثير الحمض النووي PowerPlex®Y23 في التعرف على الذكور، كما تظهر التنوع الكبير للأنماط الوراثية والقدرة على التمييز.

1. Introduction

The Indian subcontinent, known for its rich cultural and genetic diversity, reflects a complex history shaped by prolonged human occupation, cultural customs, and genetic drift. This diversity is manifested in its people, broadly categorized into castes, tribes, linguistic groups, and religious affiliations [1, 2]. Among the diverse tribes, the Koli people of western India, primarily found in Gujarat, Maharashtra, and parts of Goa, are of particular interest. With significant populations also in Madhya Pradesh, Rajasthan, and Orissa, they form one of the largest tribal communities in western India, accounting for about 24% of Gujarat's population, especially concentrated in the Saurashtra and Kutch regions [3]. The Koli's ancestry, debated between theories linking them to the White Huns from Sindh or to east-central Indian tribes like the Kol and Munda, remains a subject of ongoing research [4].

Despite their demographic significance, the Koli people have been underrepresented in genetic studies. This gap underscores the necessity for more focused research to elucidate their genetic diversity and historical background. Notably, a comprehensive study using Y-STR (Y-chromosomal short tandem repeat) markers on the Koli population of Gujarat is absent, indicating a gap in the genetic understanding of this tribe [5,6].

Gujarat, bordering the west coast of India, boasts the nation's longest coastline, mostly along the Kathiawar peninsula. As the fifth-largest state in India, covering an area of 196,024 km² and housing over 60.4 million people, it presents a significant region for genetic study due to its diverse population and strategic location Fig. 1.

Y-STR markers, known for their high polymorphism, are invaluable in forensic science, aiding in individual identification, disaster victim identification, parental kinship testing, and crucially, in rape investigations where autosomal DNA profiling may be insufficient [7,8]. This study aims to assess the efficacy of the commercial PowerPlex®Y23 multiplex kit for forensic applications within the Koli population. Additionally, it seeks to understand their demographic composition and explore genetic relationships both within the Koli group and with other populations. Our approach involves extensive Y-STR marker analysis to examine the Koli population's genetic associations and their inter- and intra-population affinities, complemented by comparisons with data from regional and global human populations.

Materials and Methods

Study Population

The focus of our study, the Koli people, are an indigenous group primarily found in Western India, notably in Maharashtra and Gujarat. For this research, the Ahmedabad district in Northern Gujarat was selected as the study area (Figure 1). This region is known for its significant Koli population, providing a representative sample for our genetic analysis. Ethical clearance for this study was obtained from the Ethical Committee of Raksha Shakti

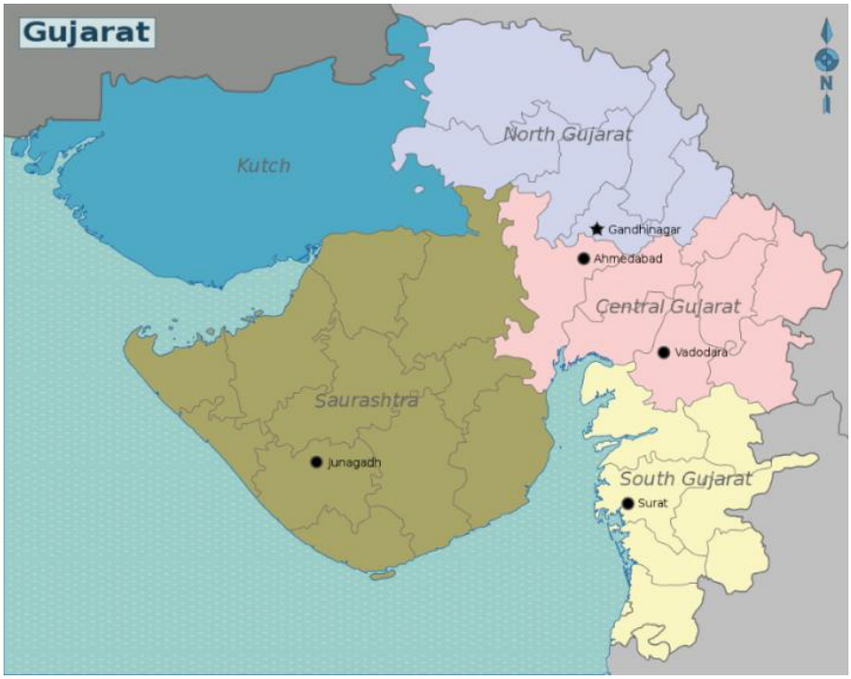

Figure 1- *Geographical location of studied population (Koli, Gujarat)*

University. In total, peripheral blood samples from 153 unrelated male individuals of the Koli tribe were collected. The use of EDTA tubes ensured the proper preservation of these samples. Recognizing the importance of community engagement, we collaborated with the local panchayat head in the rural areas of the Gujarat region. Their assistance was instrumental in reaching out to potential participants and facilitating the sample collection process. Individuals within the age range of 20-35 years were approached, and each participant was thoroughly briefed about the study objectives. Adhering to the ethical standards of the Declaration of Helsinki (2013), informed consent was obtained from all participants prior to sample collection. Additionally, participants provided information regarding their ethnicity and place of birth through a questionnaire, designed in compliance with the University Ethical Committee's guidelines [9]. This data collection process not only ensured ethical compliance but also enriched the demographic information of our study sample.

Data collection

Genomic DNA was isolated from whole blood specimens (n=153) using the Flinders Technology Associates (FTA) card technique. Each blood sample was applied to an FTA card, which was then air-dried at room temperature for 2 hours. The dried FTA cards were subsequently cleaned to remove inhibitors; each card was punched with a 1.2 mm Harris Micro Punch. The punched FTA card samples were then placed directly into the PCR reaction for amplification. In parallel, blood samples were also processed for DNA extraction using the standard salting-out method. This provided an additional source of DNA for comparison with the FTA card samples. DNA quantification was performed using the Quantifiler® Duo DNA Quantification Kit (Thermo Fisher Scientific, USA). The concentration of each DNA sample, both from FTA cards and blood extractions, was adjusted to 1 ng/µl. Amplification of samples was carried out using the PowerPlex®Y23 multiplex

commercial kit (Promega Corporation, Madison, WI, USA), following the manufacturer's protocol. The amplified products were separated on an ABI 3130 Genetic Analyzer (Thermo Fisher Scientific) using capillary electrophoresis. For DNA fragment sizing, the CC5 Internal Lane Standard 500 Y23 was employed as the internal lane standard. The polymer used for electrophoresis was POP-4 (Performance Optimized Polymer 4), and the samples were injected at 1.2 kV for 5 seconds. DNA profiles were generated using Gene Mapper ID v3.2 software (Thermo Fisher Scientific). To ensure quality control, each PCR batch included a negative control (sterile distilled H2O) and a positive control (DNA9948).

Data management and statistical analysis

For the analysis of Y-STR haplotype data, we employed the Software STR Analysis for Forensics (STRAF) [10]. This software facilitated the calculation of allelic frequencies, Haplotype Diversity (HD), Gene Diversity (GD), Matching Probability (MP), Polymorphic Information Content (PIC), and Power of Discrimination (PD) [10]. The Discrimination Capacity (DC) of our data set was computed using the formula DC = h/n, where 'h' represents the total number of unique observed haplotypes, and 'n' is the total number of haplotypes analyzed. Haplotype Analysis was conducted using Haplotype Analysis v1.04 software [11], which aided in the detailed examination of the haplotype data and its forensic relevance. To assess the genetic affinity and distances between the Koli population and other populations, the Y-Chromosome Haplotype Reference Database (YHRD, available at http://www.yhrd.org) was utilized. We calculated the genetic distance (Rst) between the Koli population and other published populations using YHRD. Additionally, Multidimensional Scaling (MDS) plots and Analysis of Molecular Variance (AMOVA) were developed using the same database to visualize genetic relationships [12]. For phylogenetic interpretation, a Neighbor-Joining (NJ) tree was constructed using MEGA 6.0 software, in-

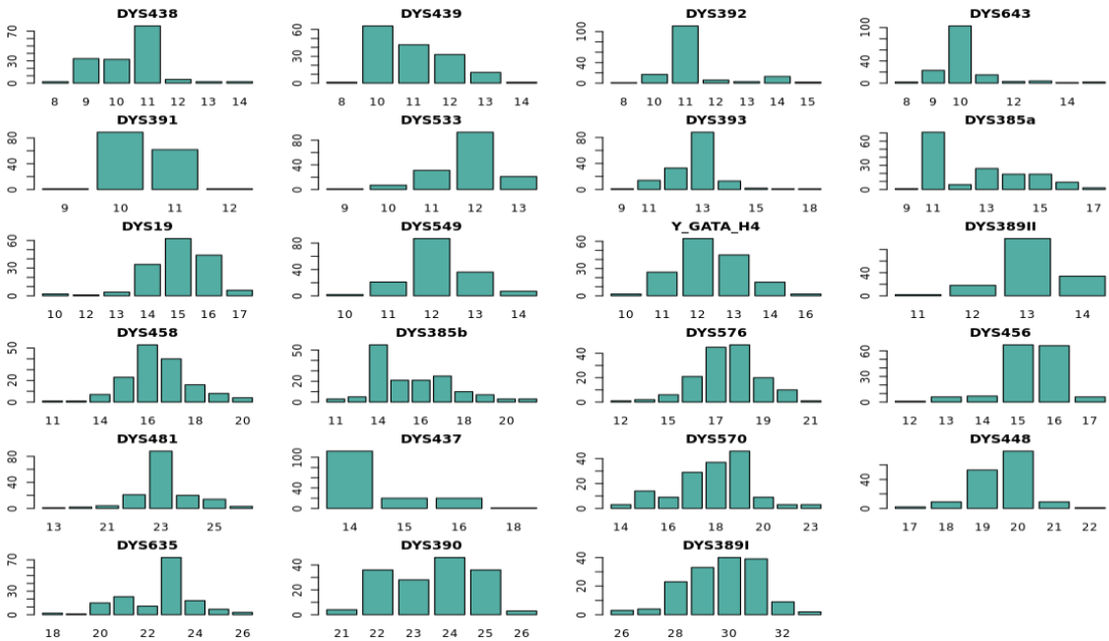

Figure 2- *Allelic frequency distribution at 23 Y-STR loci in the Koli community.*

corporating 1000 bootstrap replicates to assess the tree's reliability [13]. Haplogroups were assigned using Whit Athey's Y-DNA Haplogroup Predictor [14].

Results

Allele Frequencies, Allele Numbers, and GD Values

117 complete haplotypes were developed. The Supplementary Tables 1 and 2 present the allele frequencies, and GD values of 23 Y-STR loci in the Koli populations. Across all loci, 159 unique alleles were documented. Figure 2 displays the study population's allele frequencies for each locus.

The allelic frequencies of these alleles ranged from 0.007 to 0.732. The most alleles are found in DYS385b (n=10), DYS635, DYS570, DYS576, DYS635, DYS458 (n=9), and DYS389II, DYS393, DYS643, DYS481, DYS385a, DYS389II (n=8). A greater degree of polymorphism was noted in the remaining loci, whereas DYS391, DYS389I, and DYS437(n=4) locus exhibited a lower level of polymorphism, indicating the least amount of variation. Table 1 displays that locus DYS570 had the highest Gene Diversity (GD) value of 0.804, while DYS437 had the lowest value of 0.432. Seventeen Y-STR loci exhibiting GD values exceeding 0.5 demonstrated an elevated level of gene diversity, while six loci with GD values below 0.5 exhibited a diminished degree of gene variety.

Haplotypes and Forensic Parameters

The supplementary table 4 provide the information on the frequencies of the haplotypes for each of the 23 Y-STRs in the Koli population. Among the 153 Koli male individuals, 117 distinct haplotypes were identified, 94 (80%) of which were unique. The remaining haplotypes were shared by the population under study. Twenty haplotypes were found to occur twice, one haplotype three times, and only two haplotypes were found to be shared by eight individuals (Supplementary Table 4). For 23 Y-STR loci, the values for haplotype diversity (HD) and discrimination capacity (DC) were 0.8034 and 0.993, respectively. Table 1 and Figure 3 provide summaries of the values for Polymorphic Information Content (PIC), Matching Probability (PM), and Discrimination Power (DC). The DYS385b locus had the highest PIC values and the DYS391 locus had the lowest PIC values. Based on the calculated data, it can be concluded that the DYS385b locus outperforms the

Table 1- *Forensic parameters such as Genetic Diversity (GD), Polymorphic Information Content (PIC), Matching Probability (PM), and Discrimination Power (DC) studied in the Koli, Gujarat, population (n=153)*

Locus	GD	PIC	PM	PD
DYS19	0.706	0.648	0.299	0.701
DYS385a	0.725	0.689	0.280	0.720
DYS385b	0.803	0.775	0.202	0.798
DYS389I	0.798	0.762	0.207	0.793
DYS389II	0.521	0.463	0.482	0.518
DYS390	0.769	0.724	0.236	0.764
DYS391	0.501	0.386	0.503	0.497
DYS392	0.455	0.429	0.548	0.452
DYS393	0.611	0.564	0.393	0.607
DYS437	0.433	0.393	0.570	0.430
DYS438	0.659	0.604	0.345	0.655
DYS439	0.701	0.643	0.304	0.696
DYS448	0.610	0.537	0.394	0.606
DYS456	0.621	0.542	0.383	0.617
DYS458	0.778	0.741	0.228	0.773
DYS481	0.628	0.592	0.376	0.624
DYS533	0.572	0.521	0.432	0.569
DYS549	0.604	0.548	0.400	0.600
DYS570	0.804	0.771	0.201	0.799
DYS576	0.782	0.745	0.223	0.777
DYS635	0.723	0.692	0.282	0.718
DYS643	0.517	0.482	0.487	0.513
Y_GATA_H4	0.710	0.655	0.295	0.705

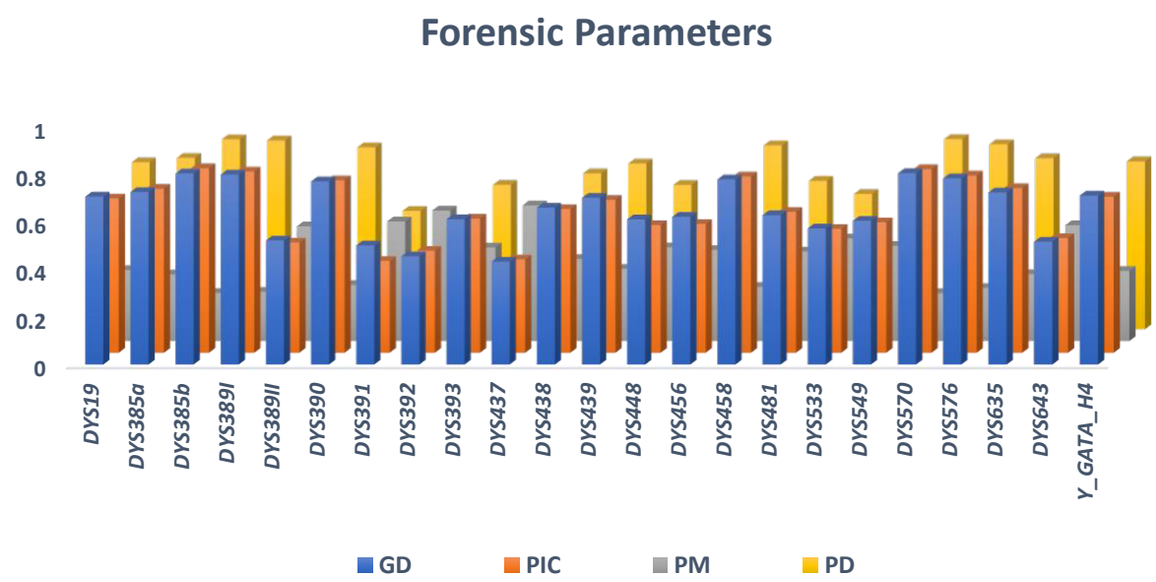

Figure 3- *Analysis of Genetic Diversity (GD), Polymorphic Information Content (PIC), Matching Probability (PM), and Discrimination Power (DC) within the Koli, Gujarat, Population (n=153).*

DYS391 locus in identifying polymorphism within the population under study. Among all the Y-STR loci examined, the DYS385b locus emerged as the most effective marker due to its superior gene diversity, polymorphic information content, and power of discrimination, with respective values of 0.803, 0.775, and 0.798.The values of the population under study's genetic diversity and discriminatory capacity were very similar to those of previously published information on the Gujarat population [15–18,18,19].

Genetic relationship between the Koli population and previously studied populations in India

Dimensionality reduction analyses (MDS) were carried out by comparing the data with previously studied Indian populations such as Odisha, India [Bhotra], Haryana, India [Brahmin], Rajasthan, India [Brahmin], Chhattisgarh, India [Dorla], Jammu and Kashmir, India [Gujjar], Uttar Pradesh, India [Gujjar], Andhra Pradesh, India [Indian], Himachal Pradesh, India [Indian], Nayagarh-Odisha, India [Indian], Uttar Pradesh, India [Indian], Uttar Pradesh, India [Kahar], Ladakh, India [Ladakhi], Chhattisgarh, India [Majhi], Chhattisgarh, India [Muria], Rajasthan, India [Indian], Maharashtra, India [Indian], Madhya Pradesh, India [Indian], Madhya Plradesh, India [Bhil] and Uttarakhand, India [Indian] [20–34], available at YHRD to analyse the genetic topographies of dispersed human populations, particularly for Koli populations. Each population was depicted as a small point within a multi-dimensional space, and the distances between these points reflected the genetic relationships among populations from different geographic regions (Figure 4). The scatterplot encompassed seven populations (Uttar Pradesh, India [Gujjar], Chhattisgarh, India [Dorla], Odisha, India [Bhotra], Chhattisgarh, India [Majhi], Ladakh, India [Ladakhi], Nayagarh-Odisha, India [Indian], and Jammu and Kashmir, India [Gujjar]) within its periphery, as indicated by the MDS diagram.

Two clusters of related Indian groups were obtained when considering reference populations and the Koli population of Gujarat. The current population, along with Rajasthan, India [Indian] and Maha-

rashtra, India [Indian], is part of Cluster 2. Cluster 1 includes Madhya Pradesh, India [Indian], Madhya Pradesh, India [Bhil], and Uttarakhand, India [Indian]. In Figure 4, our investigated population is clustered together in the first quadrant with Maharashtra, India [Indian], Rajasthan, India [Indian], Haryana, India [Brahmin], and Rajasthan, India [Brahmin]. The investigated Koli population is notably distant from Chhattisgarh, India [Dorla], which is mainly distributed in the Second quadrant. Uttar Pradesh, India [Kahar], Odisha, India [Bhotra], Chhattisgarh, India [Muria], Andhra Pradesh, India [Indian], and Chhattisgarh, India [Majhi] are localized in the fourth quadrant, far from the Koli population in the MDS plot. Uttar Pradesh, India [Gujjar] and Ladakh, India [Ladakhi] form separate clades in the upper center quadrant and the bottom center quadrant of the MDS plot, respectively. Other reference populations were found to be genetically dissimilar to the current population. Rajasthan, Haryana, and Maharashtra are geographically close. Figure 5 depicts the population structures of the 19 reference populations through the reconstructed N-J phylogenetic tree derived from the Rst analysis. The clustering pattern seen in the neighbor-joining tree (Figure5) supports MDS observation.

RST and Phylogenetic Analyses

In population genetics, the Rst (or RST) value—also called the Genetic Differentiation Index—is a metric used to evaluate the genetic affinity or differentiation between various populations. By comparing genetic variation within populations and genetic differences between populations, it calculates the genetic distance between populations. The Rst values presented in the provided table demonstrate

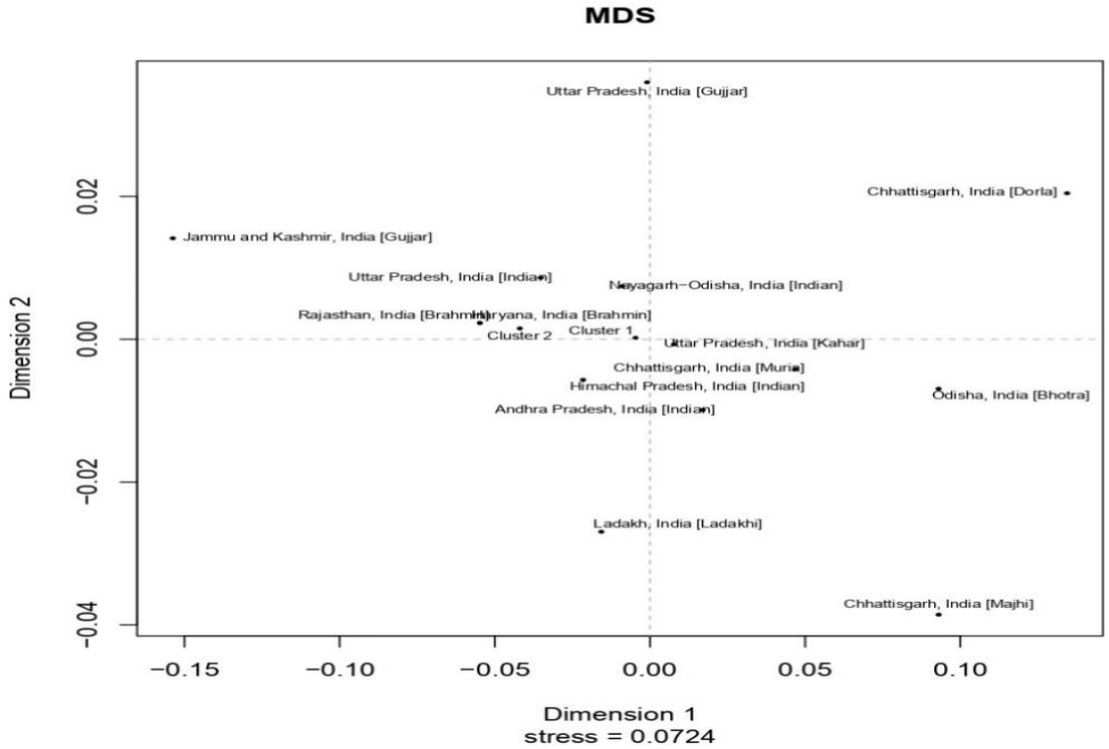

Figure 4- *The genetic relationships between the Koli population and 19 other Indian populations were depicted through a multidimensional scaling plot (MDS) utilizing 23 Y-STRs.*

Figure 5- *A Neighbourhood Joining (NJ) phylogenetic tree was constructed using 23 Y-STRs to analyse the genetic relationship between the Koli population and 19 other Indian populations*

the genetic differentiation between the Gujarati Koli population and numerous other Indian populations. Greater genetic divergence or dissimilarity is indicated by a higher Rst value, whereas a lower Rst value points to a closer genetic affinity or resemblance. The Supplementary Table 3 contains the results of the pairwise RST calculations and the corresponding p values ($p < 0.05/253 \approx 0.00019763$ after Bonferroni correction) between Koli populations and 19 other Indian populations from various regions in Odisha, India [Bhotra], Haryana, India [Brahmin], Rajasthan, India [Brahmin], Chhattisgarh, India [Dorla], Jammu and Kashmir, India [Gujjar], Uttar Pradesh, India [Gujjar], Andhra Pradesh, India [Indian], Himachal Pradesh, India [Indian], Nayagarh-Odisha, India [Indian], Uttar Pradesh, India [Indian], Uttar Pradesh, India [Kahar], Ladakh, India [Ladakhi], Chhattisgarh, India [Majhi], Chhattisgarh, India [Muria], Rajasthan, India [Indian], Maharashtra, India [Indian], Madhya Pradesh, India [Indian], Madhya Pradesh, India [Bhil], and Uttarakhand, India [Indian], based on 23 Y-STR haplotypes. The genetic affinity between the Gujarati Koli community and other Indian populations exhibited variations. The Koli community displayed the highest genetic distance with Chhattisgarh India [Dorla] (Rst = 0.1886), followed by Chhattisgarh India [Majhi] (Rst = 0.1777). Conversely, the smallest genetic distance was observed

with Maharashtra India [Indian] (Rst = -0.0028). There was some evidence of moderate genetic resemblance with Brahmin groups in Rajasthan and Himachal Pradesh, Haryana Brahmins, and Bhil in Madhya Pradesh. Populations such as the Gujjar community in Jammu and Kashmir, the Dorla population in Chhattisgarh, and the Bhotra people in Odisha have shown significant genetic difference. Maharashtra's Indian population showed negative Rst scores, a sign of genetic closeness. Some communities showed a moderate degree of genetic kinship with the Gujarati Koli population, such as those found in Uttar Pradesh, Andhra Pradesh, and Nayagarh, Odisha. The results show how these Indian communities have interacted historically and through intricate genetic linkages, which has led to the observed genetic divergence and variety.

Prediction of Y-Haplogroups

A total of 139 individuals (90.8% of the total sample size of 159) were assigned to haplogroups based on 23 Y-STRs using Whit-Athey's algorithm [35]. In this study, eleven haplogroups (A1, B, E1b1b, G2a, H1, J2a, J2b, L, R1a, R1b, R2) were observed, while 7 other haplogroups (G2c, I1, I2a1, I2a (xI2a1), I2b1, J2a1h, and N) mentioned in the haplogroup assigning tool were not found in these samples (Table 2). A total of 14 individuals (9.1%) did not have a haplogroup prediction, and one individual (0.6%) was categorized as having a poor haplogroup. The major haplogroups observed throughout the country were R1a (45.1%), H1 (13.7%), and L (11.8%), which accounted for more than three-fourths of the Y lineages. Figure 6 illustrates the distribution of haplogroups in Koli populations. It is important to exercise caution when interpreting these results, as the haplogroups were derived through prediction algorithms.

However, in this study, a substantial number of STRs (more than 17 loci) were utilized, which can offer a reasonably accurate estimate for haplogroup prediction [14]. This approach is consistent with previous research that has employed Y-chromosomal STRs to support haplogroup predictions (36). The Koli community in Gujarat exhibits a wide array of Y-chromosomal haplogroups, indicating a diverse genetic ancestry. This assortment is a testament to the region's intricate genetic composition, influenced by historical migrations and interactions.

The prevalence of R1a in South Asian populations, commonly associated with Indo-European migrations, should be considered. Additionally, the widespread presence of Haplogroup H1 in South Asia, often linked to Dravidian-speaking populations, should be considered. Smaller proportions of other haplogroups, such as E1b1b, L, and R2, also exist, ranging from 0.04% to 1.35%, representing genetic contributions from various historical and regional sources. Furthermore, there are haplogroups like A1, B, G2a, and Bad Haplotype, which have very low frequencies, each comprising less than 0.05% of the population. These may indicate less common or relatively recent genetic lineages within the Koli population. It is worth noting that approximately 0.82% of individuals in the Koli population lack any specific haplogroup prediction, which could be attributed to unique or rare genetic lineages, or limitations in the prediction tool's accuracy in assigning haplogroups in these cases.

The genetic diversity and presence of multiple haplogroups in the Koli population suggest a complex genetic ancestry influenced by historical migrations, regional interactions, and potentially a combination of Indo-European and Dravidian genetic elements. These findings have significant implica-

Table 2- *The predicted Y-Chromosomal Haplogroup Assignments within the Koli Population of Gujarat, employing Whit Athey's Y-DNA Haplogroup Predictor*

Haplogroup	Number of individuals	Frequencies of in silico Y-chromosomal haplogroup
A1	2	1.3%
B	1	0.7%
E1b1b	3	2.0%
G2a	1	0.7%
H1	21	13.7%
J2a	4	2.6%
J2b	4	2.6%
L	18	11.8%
R1a	69	45.1%
R1b	3	2.0%
R2	12	7.8%
Bad Haplotype	1	0.7%
Devoid from any haplogroup prediction	14	9.2%

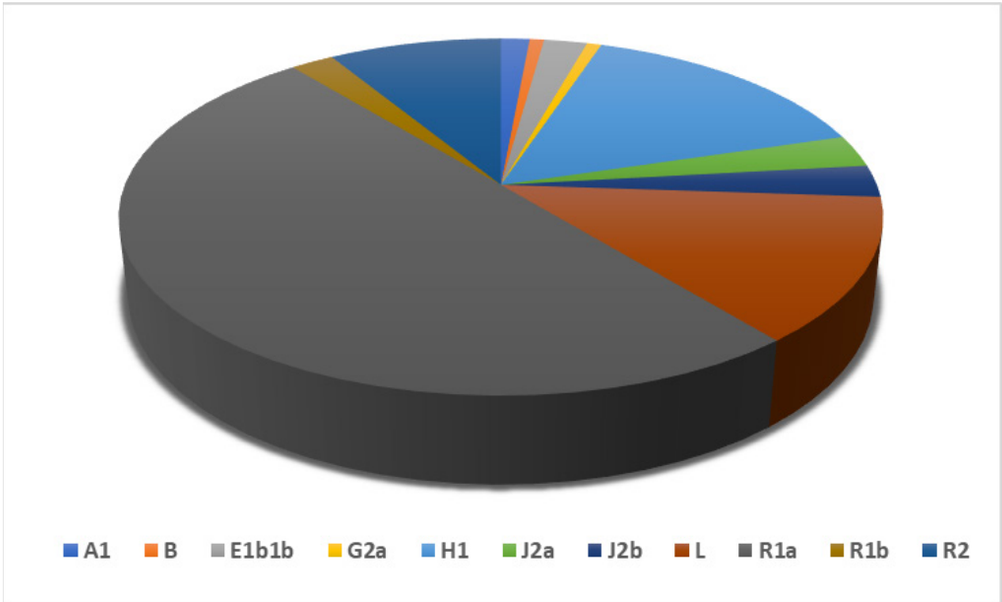

Figure 6- *Y-Chromosomal haplogroups frequency in Koli population estimated through Whit-Athey's algorithm.*

tions for genetic research, anthropology, and population genetics in the Indian subcontinent, emphasizing the necessity for further studies to explore the historical and demographic factors that have shaped the genetic landscape of the Koli population in Gujarat. In conclusion, the predicted distribution of haplogroups in the Koli population in Gujarat highlights the diverse genetic heritage of this population, reflecting the intricate history and varied origins of the people in this region.

Discussion

This study introduces the first Y-STR genetic database for the Koli population in Gujarat, significantly enriching the existing Y-STR population database. By employing the PowerPlex®Y23 multiplex commercial kit set, we observed a notable degree of genetic diversity and informativeness in the Koli population. The calculated Haplotype Diversity (HD) of 0.8034 and Discrimination Capacity (DC) of 0.993 highlight the effectiveness of these genetic markers for individual differentiation within the Koli community, underscoring their value in both forensic and population genetic studies.

Our analysis extended to comparing the Koli population's genetic affinities with 19 other geographic regions in India, building upon existing literature. The multidimensional scaling (MDS) plot (Figure 4) and its detailed examination (Supplementary Table S3) revealed a close genetic relationship between the Koli population and other Indian populations, such as those in Madhya Pradesh [Bhil] [15] and Maharashtra [Indian] [28]. Additionally, similarities with populations from Rajasthan [Brahmin], Himachal Pradesh [Indian], and Haryana [Brahmin] [21,24,29] were observed. The genetic distances among these populations, particularly those from Northern, Central, and Southern India, underscore the regional genetic diversity within the Koli gene pool, shaped by historical migrations, founder effects, endogamy, and regional isolation.

These findings illuminate the complex genetic landscape of India, molded by centuries of interaction and separation. The genetic linkage between the Koli population and the Maharashtra region corroborates earlier studies [5,6,37,38], while also supporting the notion of a common origin and close clustering with the Bhil populations of Gujarat and Madhya Pradesh. This suggests a shared linguistic and cultural heritage, providing insights into the historical and cultural interconnections of these groups. The convergence of genetic and linguistic data in these studies underscores the importance of examining both aspects to decipher the intricate relationships among various communities.

This research contributes essential genetic information about the Koli population of Gujarat, illustrating their unique genetic markers and diversity. Comparing this data with other Indian populations enhances our understanding of the broader genetic landscape and historical connections among different tribes in India. The study highlights the importance of genetic research across diverse groups to unravel the complex web of human ancestry and variation. For future research, we recommend incorporating additional molecular genetic markers and expanding the sample size. Such an approach would provide a more exhaustive and nuanced analysis of the Koli community's genetic diversity, further enriching our understanding of their genetic heritage.

Conclusion

This pilot study embarked on an exploration of the genetic traits of the Koli populations in Gujarat, India, demonstrating the efficacy of a specific panel of markers in forensic contexts. Through extensive haplogroup analysis with a representative sample, we uncovered multiple lines of descent within the

Koli community, offering insights into their diverse geographical origins. A notable outcome of this study is the highlighted potential of the PowerPlex®Y23 multiplex technology for forensic applications specific to this demographic.

Ethics approval and consent to participate

The study was approved by the ethical committee of the Raksha Shakti University, Gujarat, India (RSU/IRD/RSUIEC/5 2019/72/2019). The written consent was obtained from all the participants.

Conflict of interest

The authors declare no conflicts of interest.

Source of funding

The authors received no financial support for the research, authorship or publication of this paper.

References

1. Cavalli-Sforza LL, Menozzi P, Piazza A. The history and geography of human genes. Princeton university press; 1994.
2. Driem G van. Languages of the Himalayas: an ethnolinguistic handbook of the greater Himalayan Region: containing an introduction to the symbiotic theory of language. (No Title). 2001;
3. Singh KS. People of India: India's communities. Vol. 5. Oxford University Press; 1998.
4. Pancholi D. Particular Vulnerable Tribes of Gujarat. International Journal of Research and Analytical Reviews. 2018;5(04):422–31.
5. Ghosh T, Kalpana D, Mukerjee S, Mukherjee M, Sharma AK, Nath S, et al. Genetic diversity of autosomal STRs in eleven populations of India. Forensic Science International: Genetics. 2011;5(3):259–61.
6. Gaikwad S, Kashyap V. Genetic diversity in four tribal groups of western India: a survey of polymorphism in 15 STR loci and their application in human identification. Forensic Science International. 2003;134(2–3):225–31.
7. Jobling MA, Pandya A, Tyler-Smith C. The Y chromosome in forensic analysis and paternity testing. International journal of legal medicine. 1997;110(3):118–24.
8. Calafell F, Larmuseau MH. The Y chromosome as the most popular marker in genetic genealogy benefits interdisciplinary research. Human genetics. 2017;136(5):559–73.
9. PP R. Human experimentation. Code of ethics of the world medical association. Declaration of Helsinki. British medical journal. 1964;2(5402):177–177.
10. Gouy A, Zieger M. STRAF—a convenient online tool for STR data evaluation in forensic genetics. Forensic Science International: Genetics. 2017;30:148–51.
11. Eliades N, Eliades D. HAPLOTYPE ANALYSIS: software for analysis of haplotype data. Forest Goettingen (Germany): Genetics and Forest Tree Breeding, Georg-August University Goettingen. 2009;
12. Roewer L, Kayser M, Dieltjes P, Nagy M, Bakker E, Krawczak M, et al. Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations. Human Molecular Genetics. 1996;5(7):1029–33.
13. Caspermeyer J. MEGA software celebrates silver anniversary. Molecular biology and evolution. 2018;35(6):1558–60.
14. Athey W. Comments on the article,"Software for Y haplogroup predictions, a word of caution." International journal of legal medicine. 2011;125(6):901–3.
15. Chaudhari RR, Dahiya M. Genetic diversity of 15 autosomal short tandem repeats loci using the AmpFLSTR® IdentifilerTM kit in a Bhil Tribe Population from Gujarat state, India. Indian Journal of Human Genetics. 2014;20(2):148.
16. Mishra A, Dixit S, Choudhary SK, Sharma H, Shrivastava P. Forensic genetic analysis of the population of Gujarat with PowerPlex 21 multiplex system. Forensic Science International: Genetics Supplement Series. 2019;7(1):167–8.
17. Mishra A, Kumari A, Choudhary S, Gondhali U. Evaluation and comparison of population genetics software in Rabari Tribe of Gujarat population. Egyptian Journal of Forensic Sciences. 2021;11(1):1–12.

18. Mishra A, Misra S, Gondhali U. Genetic polymorphisms and phylogenetic analysis of the Gujarat population via 17 Y chromosomal STR systems. Rechtsmedizin. 2022;32(3):184–90.
19. Mishra A, Gondhali U, Choudhary S. Estimating the genetic polymorphisms of gujarat population using 17 Y chromosomal STR Loci. OnLine Journal of Biological Sciences. 2021;21(2):388–94.
20. Sharma AN, Kumawat R, Shrivastava P. Revealing genetic structure in the paternal lineages and forensic profiling of Bhotra tribe by the analysis of Y-STR multiplex platform. Human Gene. 2022;33:201038.
21. Sharma S, Yadav R, Sahajpal V, Zieger M, Ranga S, Ahuja P, et al. Genetic portrait of 23 Y-STR loci in the Brahmin population of Haryana, India. Rechtsmedizin. 2023;33(5):410–20.
22. Singh M, Sarkar A, Kumar D, Nandineni MR. The genetic affinities of Gujjar and Ladakhi populations of India. Scientific Reports. 2020;10(1):2055.
23. Singh M, Sarkar A, Madhusudan R. 'A comprehensive portrait of Y-STR diversity of Indian populations and comparison with 129 worldwide populations. Scientific Reports. 2018;8(1).
24. Mohapatra B, Chauhan K, Shrivastava P, Sharma A, Dagar S, Kaitholia K. Haplotype data for 17 Y-STR loci in the population of Himachal Pradesh, India. International journal of legal medicine. 2019;133(5):1401–2.
25. Imam J, Reyaz R, Singh RS, Bapuly AK, Shrivastava P. Genomic portrait of population of Jharkhand, India, drawn with 15 autosomal STRs and 17 Y-STRs. International journal of legal medicine. 2018;132(1):139–40.
26. Shrivastava P, Jain T, Trivedi VB. Haplotype data for 17 Y-STR loci in the population of Madhya Pradesh, India. Forensic Science International: Genetics. 2017;26:e31–2.
27. Shrivastava P, Gupta U, Jain T, Trivedi VB. Y STR haplotype diversity in central Indian population. Annals of human biology. 2016;43(5):480–7.
28. Panda M, Kumawat R, Dixit S, Sharma AN, Shankar H, Chaubey G, et al. Forensic features and phylogenetic analyses of the population of Nayagarh (Odisha), India using 23 Y-STRs. Annals of Human Biology. 2022;49(1):54–68.
29. Kumawat R, Shrivastava P, Shrivastava D, Mathur G. Molecular diversity of 23 Y-STR genetic markers in the population of Rajasthan, India. Meta Gene. 2020;24:100694.
30. Kumar A, Kumar R, Kumawat R, Mathur B, Shrivastava P, Chaubey G, et al. Genetic portrait study for 23 Y-STR loci in the population of Rajasthan, India. International Journal of Legal Medicine. 2020;134(5):1691–3.
31. Henke J, Henke L, Chatthopadhyay P, Kayser M, Dulmer M, Cleef S, et al. Application of Y-chromosomal STR haplotypes to forensic genetics. Croatian Medical Journal. 2001;42(3):292–7.
32. Shrivastava P, Kushwaha K, Kumawat R, Chauhan T, Saiz M, Lorente JA, et al. Untangled the genetic structure of Kahar and Tharu, using 23 Y chromosomal paternal lineage markers. Meta Gene. 2020;26:100797.
33. Perez-Benedico D, Chennakrishnaiah S, Gayden T, Rowold DJ, Garcia-Bertrand R, Herrera RJ. Y-STR markers from Ladakh in the Himalayas. Legal Medicine. 2016;21:29–32.
34. Jain T, Shrivastava P, Trivedi VB. Genetic portrait of Majhi tribe of Chhattisgarh, India based on 15 autosomal STRs and 23 Y-STRs. International journal of legal medicine. 2017;131(1):113–4.
35. Athey TW. Haplogroup prediction from Y-STR values using an allele-frequency approach. J Genet Geneal. 2005;1:1–7.
36. Toscanini U, Brisighelli F, Llull C, Berardi G, Gómez A, Andreatta F, et al. Charting the Y-chromosome ancestry of present-day Argentinean Mennonites. Journal of human genetics. 2016;61(6):507–13.
37. Badiye A, Kapoor N, Kumawat R, Dixit S, Mishra A, Dixit A, et al. A study of genomic diversity in populations of Maharashtra, India, inferred from 20 autosomal STR markers. BMC Research Notes. 2021;14(1):1–8.
38. Thangaraj K, Ramana GV, Singh L. Y-chromosome and mitochondrial DNA polymorphisms in Indian populations. ELECTROPHORESIS: An International Journal. 1999;20(8):1743–7.