Readers Insight
Publisher of open access books and journals

PJMLS

Print ISSN: 2707-4471. Online ISSN: 2707-448X

Mini Review	Pak-Euro Journal of Medical and Life Sciences
DOI: 10.31580/pjmls.v6i2.2169	Copyright © All rights are reserved by Corresponding Author
Vol. 6 No. 2, 2023: pp. 85-90	
www.readersinsight.net/pjmls	**Revised:** April 01, 2023 **Accepted:** April 03, 2023
Submission: December 04, 2022	**Published Online:** June 30, 2023

GENETIC LINEAGE OF HAZARA SUB-TRIBES THROUGH SHORT TANDEM REPEATS (STRS)

Masooma[1], Irfan Shehzad Sheikh[1]*, Niamatullah Kakar[1], Heera Batool[1]

[1]**Center for Advanced Studies in Vaccine and Biotechnology**, University of Balochistan, Quetta, Pakistan

[2]**Department of Natural and Basic Sciences,** University of Turbat, Kech, Pakistan

***Corresponding Author:** Dr. Niamatullah Kakar. Email: niamatullah.kakar@gmail.com

Abstract

The Hazaras counting over 7 million, are primarily found in central Afghanistan and southwest Pakistan's Quetta contributing about 0.09 percent of Pakistan's overall population, Hazaras are one of the unique but minor ethnicities. Studies have been done on Hazara people through Short Tandem Repeats (STRs) but there is no any study on the genetic lineage of Hazara sub-tribes. Short Tandem Repeats (STRs) have a key role in the forensic study these days. These short tandem repeats (STRs) contain short nucleotide repeats which vary from individual to individual and can be used in fingerprinting in modern days. There is a wide usage of STRs such as genetic lineage, historical background, paternity testing, finding culprits in harassment and theft cases because small amount of DNA is required to process the study. The Hazara population, with their distinctive genetic profile, offers a unique opportunity to investigate the genetic lineage of sub-tribes within this minority group. Understanding the genetic diversity and ancestry of Hazara sub-tribes can shed light on their migration patterns and historical origins. By employing Short Tandem Repeats (STRs), which exhibit high variability and can be easily analyzed, this study aims to unravel the intricate genetic landscape of Hazara sub-tribes in Quetta, Pakistan. The findings from this research will not only contribute to the scientific knowledge of Hazara genetics but also have implications for forensic investigations, population studies, and paternity testing. Furthermore, the utilization of STRs in this study highlights the importance of this molecular tool in unraveling the complex genetic identities of human populations. Through a comprehensive analysis of genetic lineage using STRs, this study will provide valuable insights into the genetic diversity and sub-tribal structure of the Hazara population in Quetta.

Keywords: *Genetic lineage, Hazaras, Nucleotide repeats, STRs, Sub-tribes*

INTRODUCTION

Pakistan, the sixth most populous and the 33rd largest country in the world, is a diverse nation with a rich cultural and ethnic tapestry (1). It is home to more than 150 million people, comprising at least 18 different ethnic groups and speaking over 60 distinct languages (2). Among the various ethnic groups, the Hazaras hold a significant place in the country's demographic landscape. The Hazaras, numbering over 7 million, are primarily concentrated in central Afghanistan and Quetta, a city in the southwest province of Balochistan, Pakistan (3). Despite their relatively small population, the Hazaras are an important and unique ethnic group with a distinct cultural and genetic heritage.

The Hazara people trace their origins to Genghis Khan's army during the early 13th century AD, and their distinctive Mongolian facial features set them apart from other ethnic groups (4). Hazaragi, their spoken language, is a fusion of Mongolian and Persian elements. While they have primarily settled in Quetta, Balochistan, a significant number of Hazaras also reside in Parachinar, a town in the northwestern province of Khyber Pakhtunkhwa (5-8). The name "Hazara" itself derives from the Persian term "hazar," meaning "thousand," reflecting their historical connection to the Mongol Empire.

Genetic variation and polymorphism play a crucial role in shaping the characteristics and diversity of different populations. Deoxyribonucleic acid (DNA) serves as the hereditary material in all cells, containing the genetic information that distinguishes individuals and determines their unique traits (9). While human DNA is nearly 99.9% identical among individuals, the remaining 0.1% accounts for the genetic

variations responsible for the differences observed between individuals (10). These genetic variations, or polymorphisms, can be examined using various molecular techniques, including mitochondrial DNA analysis, polymerase chain reaction (PCR)-based methods for single nucleotide polymorphism (SNP) typing, and short tandem repeat (STR) analysis (11).

Among these techniques, STR analysis has emerged as a powerful tool for studying genetic variation in human populations. Short tandem repeats are repetitive DNA sequences composed of 2 to 6 base pair motifs that exhibit varying numbers of repetitions at specific loci in the genome (12, 13). These repetitive sequences, also known as microsatellites or simple tandem repeats, are abundant in the human genome and display high levels of polymorphism (14, 15). The ability to analyze STRs using PCR amplification has made them invaluable in diverse fields, including forensic genetics, population genetics, genetic mapping, and paternity testing (16).

In the context of the Hazara population, previous studies have utilized STR analysis to explore their genetic characteristics and establish population-specific databases (17, 18). However, there is a notable research gap regarding the genetic lineage of Hazara sub-tribes. Investigating the genetic lineage of sub-tribes within the Hazara population can provide insights into their migration patterns, ancestral connections, and population structure.

This mini-review aims to evaluate the genetic lineage of Hazara sub-tribes from Quetta, Balochistan, using STR analysis. By examining the genetic diversity and patterns of variation within and between sub-tribes, we can gain a deeper understanding of the Hazara population's genetic heritage. Furthermore, such insights can contribute to broader discussions on population genetics, migration history, and genetic relationships in this region.

In summary, the evaluation of the genetic lineage of Hazara sub-tribes through STR analysis offers an opportunity to explore the genetic diversity and unique characteristics of this distinctive ethnic group. By filling the research gap in the understanding of Hazara sub-tribes' genetic lineage, this study will contribute to the broader knowledge of population genetics and shed light on the historical and ancestral connections of the Hazara population in Quetta, Balochistan.

GENETIC VARIATION OR POLYMORPHISM

Except for erythrocytes, each cell has hereditary material called DNA. People can be recognized and their genetic make-up may be ascertained, thanks to DNA, which serves as the genetic blueprint. Saliva, blood, bones, vaginal discharge, sperm, teeth and hair all contain traces of DNA (19). Human DNA is 99.9% same and only 0.1 percent unique. Polymorphism, which distinguishes each individual, reveals the difference. Polymorphisms can be identified using mitochondrial DNA (mtDNA), Polymerase chain reaction (PCR)-based methods for single nucleotide polymorphism (SNPs) and (STR) short tandem repeat loci, as well as typing of variable number tandem repeats (VNTR) loci and restriction fragment length polymorphism (RFLP). The degree and kind of genetic diversity within a population affects its structure, background (migration and recent expansion) and the distribution of diseased genes (20).

DNA AS A HEREDITARY MATERIAL

There are approximately three billion bases of genetic material in humans, among which 2% are used to code for different proteins. The other DNA can be referred to either "junk" or as "non-coding" DNA since it does not include any protein-coding regions (21). This non-coding DNA, often known as junk DNA, is composed of repeating nucleotides of various lengths. Individuals' vastly repeated DNA differs from that of other people (22). Moreover, individuals' DNA nucleotide sequences at the same chromosomal location vary (23). Thus, genetically speaking, any two randomly chosen individuals are 99.9% genetically similar, while the other three millions of the three billion genetic bases give each person their own distinctiveness (24). Due to the variability in repetitive DNA, a method termed "DNA fingerprinting" (also termed "DNA profiling " or "DNA typing") has been developed. Using a person's DNA, this technique generates a distinctive design that may be employed to recognize people (25).

SHORT TANDEM REPEATS (STRs)

A new name was created for shorter repeat motifs (often 10 to 60 bp) called "minisatellite" (26). These repetitive DNA sequences are widespread and exhibit less repetition at a particular location in the genome. In the human genome, there is another class of repetitive DNA sequences that are composed of relatively short DNA sequences, typically 1–10 bps. These repetitive sequences were termed as “Simple Sequence Repeats” (SSRs) (27, 28). After that, these sequences were called as “Microsatellite” or “Simple Tandem Repeats” or “Short Tandem Repeats” (STRs) (29, 30). Microsatellites are widely spread in eukaryotic genomes, according to increasing research (31-34). Different types of repetitive DNA sequences have different core repeat unit lengths, neighboring repeat unit counts, and total repeat region lengths. A few hundred to several thousand bases may be included in the core repetition of long repeat units. These areas, which are frequently referred to as satellite DNA, are located close to the chromosomal centromere. The word "satellite" was coined since early tests that also used equilibrium density gradient centrifugation frequently showed the presence of one or more minor satellite bands (35).

Simple sequence or short tandem repeats (STRs)are tandemly repeated DNA slices with repetition lengths comprising up to 6 bp with maximum lengths of 60 bp. More number of STRs can be found in mammalian genomes. As the number of repetitions inside individual STRs vary extremely so these short tandem repeat polymorphisms (STRPs) can be easily studied using PCR. STR distribution within genetic material is not equal, they are found rarely in sub telomeric areas (being the site for VNTRs) (36).

The human genome contains short tandem repeats (STR), which make up 3% of the entire genome and are small DNA fragments with 2–6 base pair repetitions that may be found everywhere. One of the few DNA profiling techniques now available for identifying the molecular makeup of people and animals is called "STR fingerprinting," which is also referred to as "DNA fingerprinting" (37).

STRs are plenteous, codominant, highly polymorphic and practically neutral in terms of selection. Furthermore, The DNA fragments found in STRs are so small that they might be detached on high-definition medium like polyacrylamide and can be amplified by PCR. Since high-throughput capillary sequencers and mass spectrographs are now readily available, the size of alleles is no longer taken into account when evaluating STRs. STRs are commonly used in both scientific and practical research as an essential result. STRs can be used to create genetic maps, localize genes, conduct genetic linkage studies, identify people, test for paternity, and diagnose illnesses. Population genetics has also employed STR analysis (38). By definition, STRs are brief, three or more can be evaluated at once. DNA tests can be multiplexed or run with several STRs. Multiplex STRs are helpful since they may produce high discriminatory results and effectively evaluate sample mixtures and biological materials including damaged DNA molecules. As the need for DNA testing rises, it is also advantageous that multiplex STRs can be detected automatically (39).

USES OF STRs

When analyzing DNA using the sex and autosomal chromosomes, Short Tandem Repeats (STRs) are widely utilized. They are commonly used in forensic cases, population research and medico-legal analysis (40). Short tandem repeat loci are necessary for forensic DNA analysis in the Combined DNA Index System (CODIS). By using CODIS, these core loci have been increased to 20 STRs (41, 42). FGA, TPOX, CSF1PO, TH01, vWA, D3S1358, D5S818, D7S820, D13S317, D21S11, D8S1179, and D18S51 are the 13 core loci that CODIS prescribed for forensic purposes up until 2016. Since the beginning of 2017 seven additional markers including D10S1248, D1S1656, D2S1338, D19S433, D2S441 (43). The human genome initiative and other programs demanding a significant number of biological and genetic markers can use STRs due to their great diversity and ease of usage as polymorphic markers. Sequence-based alleles have the capability to be more precise than continuous-allele methods, which are already used in forensic medicine and science for genetic typing (e.g., VNTRs) (44). Additionally, it demonstrated that trimeric and tetrameric STRs are common in the human genome, as compared to single nucleotides, their mutation rate is substantially higher, surpassing 10^{-3} per generation and are appropriate for use in DNA typing and genetic mapping (45-48). There are certain benefits to adopting STR analysis for forensic purposes. The loci are 200 base pair in

length, making it possible to identify them even in highly damaged DNA, allowing STRs to be successfully analyzed from bones that are 70 years old (39). With automated photoluminescence, alleles may still be clearly recognized even though there are fewer alleles than in large molecular mass (2 kB) hypervariable DNA because each electrophoretic path has an internal standard size marker that enables accurate measurement of a fragment of DNA to around 1 base pair (40, 41). Micro satellites are a collection of molecular markers used for a variety of tasks, such as relatedness analysis and forensic detection and tracking (42-44). Little amounts of template DNA are needed (20-100 ng) (44). Each STR is passed down as a single unit, referred to as a haplotype, and serves as a single allele for every individual because of the lack of crossover. (44-46). Additionally, STRs tend to be helpful in historical study, various parentage testing circumstances, biological ancestry, and inquiries into missing individuals (44, 47, 48).

CONCLUSION

The study highlights the unique genetic lineage of Hazara sub-tribes in Quetta through the use of Short Tandem Repeats (STRs). While previous studies have explored the Hazara population using STRs, this research specifically focuses on the genetic lineage of Hazara sub-tribes, which has not been studied before. STRs have proven to be valuable tools in forensic analysis, and their usage extends to genetic lineage studies, historical background analysis, paternity testing, and criminal investigations. The prevalence of Hazaras, a minor ethnic group, in central Afghanistan and Quetta, Pakistan, contributes to their distinct genetic profile. Understanding the genetic variation and polymorphism within the Hazara sub-tribes provides insights into their ancestry and population structure. The study emphasizes the importance of STR analysis in uncovering the unique genetic characteristics of sub-tribes and sheds light on the diversity within the Hazara population.

References:

1. Chen P, Adnan A, Rakha A, Wang M, Zou X, Mo X, He G. Population background exploration and genetic distribution analysis of Pakistan Hazara via 23 autosomal STRs. Annals of Human Biology. 2019;46(6):514-8.
2. Grimes BF. Ethnologue: Languages of the World, Summer Institute of Linguistics, Dallas, TX, 2006.
3. Mehdi SQ, Qamar R, Ayub Q, Khaliq S, Mansoor A, Ismail M, Hammer MF, Underhill PA, Cavalli-Sforza LL. The origins of Pakistani populations: evidence from Y chromosome markers. InGenomic diversity: applications in human population genetics. Boston, MA: Springer US. 1999:83-90.
4. Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, Qamar R, Ayub Q, Mohyuddin A, Fu S, Li P. The genetic legacy of the Mongols. The American Journal of Human Genetics. 2003;72(3):717-21.
5. Mousavi SA. The Hazaras of Afghanistan. Routledge; 2018.
6. Perveen R, Shahid AA, Ahmad J. Forensic and phylogenetic characterization of 15 autosomal STRs in Hazara population of Pakistan. Legal Medicine. 2020;47:101786.
7. Adnan A, Rakha A, Kasim K, Noor A, Nazir S, Hadi S, Pang H. Genetic characterization of Y-chromosomal STRs in Hazara ethnic group of Pakistan and confirmation of DYS448 null allele. International Journal of Legal Medicine. 2019;133:789-93.
8. Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C, Mehdi SQ. Y-chromosomal DNA variation in Pakistan. The American Journal of Human Genetics. 2002;70(5):1107-24.
9. Nahar Sultana GN. Medicolegal aspects and examination techniques in exhumation cases with alleged human rights violation. J forensic sci crim investig. 2018;9(1).
10. Budowle B, Allard MW, Wilson MR, Chakraborty R. Forensics and mitochondrial DNA: applications, debates, and foundations. Annual review of genomics and human genetics. 2003 Sep;4(1):119-41.
11. Akhtar M, Watson JL, Nazli A, McKay DM. Bacterial DNA evokes epithelial IL-8 production by a MAPK-dependent, NFκB-independent pathway. The FASEB Journal. 2003;17(10):1319-21.
12. Jeffreys AJ, Wilson V, Thein SL. Individual-specific 'fingerprints' of human DNA. Nature. 1985;316(6023):76-9.
13. Lewis SE, Aitken RJ. DNA damage to spermatozoa has impacts on fertilization and pregnancy. Cell and tissue research. 2005;322:33-41.
14. Yamamoto T, Uchihi R, Nozawa H, Huang XL, Leong YK, Tanaka M, Mizutani M, Tamaki K, Katsumata Y. Allele distribution at nine STR loci—D3S1358, vWA, FGA, TH01, TPOX, CSF1PO,

D5S818, D13S317 and D7S820—in the Japanese population by multiplex PCR and capillary electrophoresis. Journal of forensic sciences. 1999;44(1):167-70.
15. Brar JS, Bains TS, Shanmugasundaram S, Singh S. Developing short duration mungbean genotypes suitable for rice-wheat cropping system. Improving Income and Nutrition by Incorporating Mungbean in Cereal Fallows in the Indo-Gangetic Plains of South Asia DFID Mungbean Project for. 2002;2004:61.
16. Tautz D, Renz M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic acids research. 1984;12(10):4127-38.
17. Jacob HJ, Lindpaintner K, Lincoln SE, Kusumi K, Bunker RK, Mao YP, Ganten D, Dzau VJ, Lander ES. Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. Cell. 1991;67(1):213-24.
18. Litt M, Luty JA. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. American journal of human genetics. 1989;44(3):397.
19. Craig J, Fowler S, Skinner JD, Burgoyne LA, McInnes JL. Improved separation of multi-locus hypervariable DNA restriction fragments by field inversion gel electrophoresis and fragment detection using a biotinylated probe. Applied and Theoretical Electrophoresis: the Official Journal of the International Electrophoresis Society. 1988;1(1):23-8.
20. Hamada H, Petrino MG, Kakunaga T. A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proceedings of the National Academy of Sciences. 1982;79(21):6465-9.
21. Condit R, Hubbell SP. Abundance and DNA sequence of two-base repeat regions in tropical tree genomes. Genome. 1991;34(1):66-71.
22. Morgante M, Olivieri AM. PCR-amplified microsatellites as markers in plant genetics. The plant journal. 1993;3(1):175-82.
23. Stallings RL, Ford AF, Nelson D, Torney DC, Hildebrand CE, Moyzis RK. Evolution and distribution of (GT) n repetitive sequences in mammalian genomes. Genomics. 1991;10(3):807-15.
24. Ashley MV, Dow BD. The use of microsatellite analysis in population biology: background, methods and potential applications. Molecular ecology and evolution: approaches and applications. 1994:185-201.
25. Britten RJ, Kohne DE. Repeated sequences in DNA: hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. Science. 1968;161(3841):529-40.
26. Koreth J, O'LEARY JJ, O'D. McGEE JA. Microsatellites and PCR genomic analysis. The Journal of pathology. 1996;178(3):239-48.
27. Fu J, Cheng J, Liu X, Li J, Wei C, Zheng X, He T, Fu J. Evaluation genotypes of cancer cell lines HCC1954 and SiHa by short tandem repeat (STR) analysis and DNA sequencing. Molecular biology reports. 2018;45:2689-95.
28. Fan H, Chu JY. A brief review of short tandem repeat mutation. Genomics, proteomics & bioinformatics. 2007;5(1):7-14.
29. Butler JM. Forensic DNA typing: biology & technology behind STR markers. Academic Press; 2001.
30. Srivastava A, Kumawat R, Dixit S, Kaitholia K, Shrivastava D, Yadav VK, Nigam K, Sharma H, Trivedi VB, Chaubey G, Shrivastava P. Genetic data for PowerPlex 21™ autosomal and PowerPlex 23 Y-STR™ loci from population of the state of Uttar Pradesh, India. International journal of legal medicine. 2019;133:1381-3.
31. Moretti TR, Moreno LI, Smerick JB, Pignone ML, Hizon R, Buckleton JS, Bright JA, Onorato AJ. Population data on the expanded CODIS core STR loci for eleven populations of significance for forensic DNA analyses in the United States. Forensic Science International: Genetics. 2016;25:175-81.
32. Hares DR. Selection and implementation of expanded CODIS core loci in the United States. Forensic Science International: Genetics. 2015;17:33-4.
33. Kumar A, Kumar R, Kumawat RK, Tilawat A, Shrivastava P, Chaubey G. Genetic variation (population database) at 20 autosomal STR loci in the population of Rajasthan (north-western India). International Journal of Legal Medicine. 2020;134(5):1667-9.
34. Devlin B, Risch N, Roeder K. No excess of homozygosity at loci used for DNA fingerprinting. Science. 1990;249(4975):1416-20.
35. Edwards A, Civitello A, Hammond HA, Caskey CT. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. American journal of human genetics. 1991;49(4):746.
36. White TL, Adams WT, Neale DB. Genetic markers-morphological, biochemical and molecular markers. InForest genetics. Wallingford UK: CABI.2007:53-74).

37. Zhang DX, Hewitt GM. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. Molecular ecology. 2003;12(3):563-84.
38. Bentayebi K, Abada F, Ihzmad H, Amzazi S. Genetic ancestry of a Moroccan population as inferred from autosomal STRs. Meta gene. 2014;2:427-38.
39. Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E, Sullivan K. Identification of the remains of the Romanov family by DNA analysis. Nature genetics. 1994;6(2):130-5.
40. Kimpton C, Fisher D, Watson S, Adams M, Urquhart A, Lygo J, Gill P. Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. International journal of legal medicine. 1994;106:302-11.
41. Gill P, Evett I. Population genetics of short tandem repeat (STR) loci. Human Identification: The Use of DNA Markers. 1995:69-87.
42. Nakamura Y. DNA variations in human and medical genetics: 25 years of my experience. Journal of human genetics. 2009;54(1):1-8.
43. Yamamoto T, Uchihi R, Nozawa H, Huang XL, Leong YK, Tanaka M, Mizutani M, Tamaki K, Katsumata Y. Allele distribution at nine STR loci—D3S1358, vWA, FGA, TH01, TPOX, CSF1PO, D5S818, D13S317 and D7S820—in the Japanese population by multiplex PCR and capillary electrophoresis. Journal of forensic sciences. 1999;44(1):167-70.
44. Hameed IH, Jebor MA, Ommer AJ, Yoke C, Zaidian HK, Al-Saadi AH, Abdulazeez MA. Genetic variation and DNA markers in forensic analysis. African Journal of Biotechnology. 2014;13(31).
45. Parson W, Niederstätter H, Brandstätter A, Berger B. Improved specificity of Y-STR typing in DNA mixture samples. International journal of legal medicine. 2003;117:109-14.
46. Kwak KD, Jin HJ, Shin DJ, Kim JM, Roewer L, Krawczak M, Tyler-Smith C, Kim W. Y-chromosomal STR haplotypes and their applications to forensic and population studies in east Asia. International journal of legal medicine. 2005;119:195-201.
47. Verzeletti A, Cerri N, Gasparini F, Poglio A, Mazzeo E, De Ferrari F. Population data for 15 autosomal STRs loci and 12 Y chromosome STRs loci in a population sample from the Sardinia island (Italy). Legal Medicine. 2009;11(1):37-40.
48. Park MJ, Lee HY, Chung U, Kang SC, Shin KJ. Y-STR analysis of degraded DNA using reduced-size amplicons. International journal of legal medicine. 2007;121:152-7.