# *Allele frequencies of SE33 locus and evaluation of its reliability for molecular forensics in the Pakhtun population of Pakistan*

**Muhammad Jamil[1], Umer Rashid*[1], Nadia Zeeshan[1]**

[1] Department of Biochemistry and Biotechnology, University of Gujrat, Gujrat, Pakistan

## Abstract

Forensic DNA investigation is considered the pinnacle of various forensic sciences, as it can extend probabilistic predictions beyond the universe of observations, underpinned by a robust theoretical foundation. Forensic DNA investigations typically involves the development of Short Tandem Repeat (STR) DNA profiles from samples and interpretation of the evidentiary weight of these profiles. The SE33 STR DNA locus, renowned for its global forensic identification applications, is the focus of this study. We assess SE33 polymorphism frequencies, utilizing a sample of 494 unrelated Pakhtun individuals' genomic DNA which underwent amplification using the GlobalFiler PCR (AmpFlSTR) amplification kit. A total of 44 alleles were observed including few least frequent alleles with a 9.1% homozygote genotype rate and a 90.9% heterozygote genotype rate. Statistical calculations revealed a high power of discrimination (PD) of 0.9921, Power of Exclusion (PE) of 0.814 and a low matching probability (PM) of 0.0078. We further broadened the scope by juxtaposing our allele frequencies with published datasets from neighboring populations. This comparative analysis, not only highlights the unique genetic architecture of the Pakhtun population but also underscores their partial genetic relatedness and distinction from other South Asian populations. The genetic affiliations of Pakhtun population with South Asian communities, enriching our understanding of historical migratory patterns and genetic exchange.

The SE33 locus can be confidently applied in Pakhtun and other population of Pakistan and neighboring countries with shared ancestry across regions.

**Keywords:** Pakhtun population; Forensic STR; SE33 Locus.

## 1. Introduction

Forensic DNA analysis is the gold standard for different forensically applied sciences due to its ability to make probabilistic predictions beyond the universe of observations based on a solid theoretical background [1]. A distinct collection of alleles from several STR loci is assembled to provide a DNA profile. STRs are resilient, display high levels of variation among human groups, and are simple to multiplex for Polymerase Chain Reaction (PCR) amplification [2]. Due to their small fragment size, they are less likely to degrade, making them more suitable for forensic casework [3]. In forensic sciences, to properly use a genetic marker, one must be aware of the polymorphism characteristics of a population, including allele frequency, homozygote, and heterozygote. Several autosomal loci are used in the process of forensic identification. Among these STR loci, one of the most unique loci is SE33.

SE33 was discovered as a STR marker on chromosome 6 next to a beta actin related pseudogene (ACTBP2) [4]. Due to its high polymorphism, discriminating capacity, and the fact that more than 140 alleles have been identified to far, SE33 has been included in a number of STR kits [5]. In forensic, SE33 is valuable for identification and paternity testing because it has been shown to have great individualization power and hyper variability in specific Pakistani groups [6]. SE33 locus is part of the extended 20 CODIS core STR loci and has also been added to the newly formed European standard STR loci set for forensic analysis [7].

In Pakistan, no systematic research has been conducted regarding the distribution of allele frequencies in local populations; hence, DNA profile probability in forensic is calculated using the databases of other countries, which results in lower discrimination power. Significant differences in allele frequencies in different ethnic groups or subpopulations may be possible and these differences may affect the statistical calculations. It is crucial to determine the allele frequencies of each subpopulation or ethnic group within a larger population in order to assess these potential disparities [8]. The current study was therefore conducted to analyze the highly polymorphic SE33 locus in Pakhtun population who are basically residents of Afghanistan and Pakistan and are also extensively scattered over the subcontinent due to their traits of being dominating, independent, and courageous.

This study demonstrates that among the STR marker, the SE33 locus has the potential to enhance the forensic assessment in Pakistan as well as in Asian populations.

## 2. Materials and Methods

With the explicit informed consent of 494 unrelated, healthy Pakhtun residents of Pakistan, buccal swab samples were obtained. The Chelex-100 extraction procedure was used to extract DNA [9]. The Qubit™ dsDNA High Sensitivity (HS) Assay Kit, the Fluorometer Qubit® 3.0 (Invitrogen, Carlsbad, California), and the Human Dou DNA Quantifiler™ (Applied Biosystems, Foster City, CA, USA) kit were used to quantify DNA in accordance with the manufacturer's instructions [10].

Samples underwent genotyping for 21 STR Genetic Markers through the use of the AmpFLSTR® Global Filer PCR amplification kit (Thermo Fisher Scientific, MA, USA). This kit encompasses the SE33 locus within the set of 21 autosomal STR Loci and includes three gender-specific markers. Following PCR amplification, 1 µl of the amplified DNA was combined with 9.5 µl Hi-Di Formamide. Cappilary electrophoresis was carried out to separate and detect the fluorescently labeled PCR fragments using a 3500 Genetic Analyzer (Thermo Fisher Scientific, MA, USA) equipped with J6 Matrix Standards. Data analysis was carried out using GeneMapper IDX software V1.4 (Thermo Fisher Scientific, MA, USA). As controls, male DNA of 9948 from Global Filer PCR amplification kit and molecular biology grade water were employed as positive and negative controls, respectively. The Gene Mapper software v1.4 was used to analyze and compile the electrophoretic data. The reliability of genotyping findings was tested twice, and the results were compared.

Analyzing 494 unrelated individuals, Power Stats v 1.2 software (Promega, Madison, WI, USA) was used to carry out statistical analyses, including evaluations of allelic and genotype frequencies, Genetic Distance (GD), power of discrimination (PD), matching probability (MP), polymorphism information content (PIC), observed heterozygosity (Ho), power of exclusion (PE) and typical paternity index (TPI).

## 3. Results and Discussion

The SE33 locus found on chromosome number 6, with a reported mutation rate of 0.64% displayed the maximum number of alleles which are depicted in Figure 1. On the SE33 locus,

allele 18 was the most frequent with a frequency of about 9.1% in the Pakhtun population. In a study on SE33 locus in Turkish population, they have also reported high frequency for this allele [11]. Few alleles 6.3, 7, 15.2, 15.3, 16.2, 18.2, 23, 28 and 32 has a lower frequency of just above 0.1%. A total of 44 alleles are visible with their respective frequency percentages, indicating that SE33 is highly polymorphic in the Pakhtun population of Pakistan and other populations studied, such as U.S. populations [12] and the Pakistani population [6].

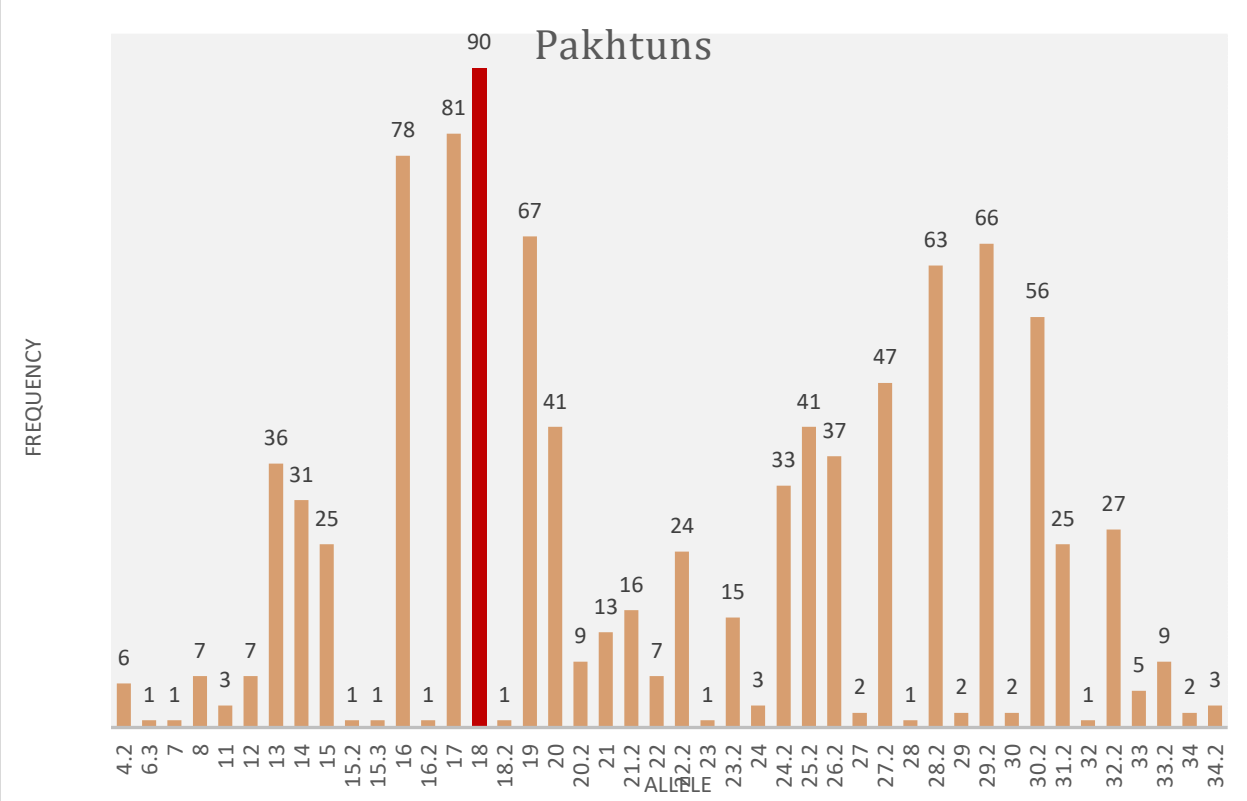


Figure 1. Distribution of detected alleles (Allele 18 detected 90 times).

There were a total of 494 genotypes identified, consisting of 449 heterozygous (90.9%) and 45 homozygous (9.1%) genotypes. This study establishes that SE33 stands out as an excellent STR marker for forensic DNA based identification, as evident by the values of forensic statistical parameters given in table 01.

Table 01. Forensic Parameters of SE33 Locus in Pakhtun population

| Locus | No | No. of alleles | GD | PIC | PM | PD | Hobs | PE | TPI |
|---|---|---|---|---|---|---|---|---|---|
| SE33 | 494 | 44 | 0.94911583 | 0.945707 | 0.0078595 | 0.9921405 | 0.90890688 | 0.81365054 | 5.48888889 |

Outcomes of different studies conducted on SE33 locus around the globe, few of which are given in table 2, reported this marker as highly polymorphic and valuable for human identification. The forensic parameters reported for SE33 locus in Turkish population as PM: 0.008, observed heterozygosity: 0.946, PD: 0.992, PE: 0.95, TPI: 9.26 and PIC: 0.95 [11]. In a study conducted on German population, forensic parameters like the power of discrimination and observed heterozygosity were 0.987 and 0.951 respectively for SE33 locus [13]. PD of SE33 STR loci in individuals from Bahrain was 0.9937 [14], individuals from İndia was 0.990 [15] and individuals from Jordan was 0.983 [16] reveals that this marker is the most polymorphic and informative. Forensic parameters of the SE33 locus for Japanese population were reported as observed Homozygosity: 0.0743, PIC: 0.9331, PD: 0.9872 and TPI: 6.7308 [17]. In a study conducted on a large number of Korean individuals, forensic parameters of SE33 locus were determined as Probability of Matching (PM): 0.0080, observed heterozygosity: 0.9475 and polymorphic information content (PIC): 0.8930 [18]. In Portuguese population, forensic parameters of SE33 locus were determined as PD: 0.995, PE: 0.903 and heterozygosity: 0.947 [19]. As mentioned earlier, our findings align with population studies and results observed globally for SE33 locus. Even upon examining forensic parameters like PD, PM, PIC, TPI, and heterozygous rates, higher values have been identified.

**Table 2**. Overview of forensic parameters for the SE33 locus of Pakhtun population and other populations.

| **Forensic Parameters** | Pakhtun (current study | Pakhtun (Bhinder) | Pakistani (Bhinder) | Indian | German | Polish | Portuguese | Turkish | Japanese | Hungarian | Brazilian |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Heterozygosity** | 0.909 | 0.906 | 0.887 | 0.938 | 0.951 | 0.933 | 0.947 | 0.946 | 0.9458 | 0.8652 | 0.916 |
| **Power of discrimination** | 0.992 | 0.978 | 0.991 | 0.990 | 0.987 | 0.994 | 0.995 | 0.992 | 0.9872 | 0.9851 | 0.991 |
| **Power of exclusion** | 0.814 | 0.807 | 0.769 | 0.770 | 0.903 | 0.8223 | 0.903 | 0.89 | 0.8892 | 0.8832 | 0.815 |

In a study in the Pakistani population by Bhinder et al., 2018, Forensic genetics and statistical parameters for the SE33 STR locus were; TPI: 4.43, PD: 0.991, PE: 0.769, PIC: 0.95, with 88.7% heterozygotes detected [6]. This study contains a population sample of 53 Pakhtun individuals as well. In another study by Shan et al, 2020, for a sample of 529 Pakistani population including 107 Pakhtun individuals, different forensic and genetic parameters were calculated for 21 STR loci containing SE33 locus which showed the greatest power of discrimination [20].

As noted above, our study stands out in forensic parameters when juxtaposed with the aforementioned studies, particularly those conducted on the Pakistani population. We fortified our research with a significantly larger sample size, underscoring the robustness and reliability of our findings. Bhinder et al., 2018 and the current study have almost the same values for PD and PIC parameters. Conversely, higher values were obtained for the power of exclusion, TPI and heterozygosity parameters in our study with lower matching probability. Furthermore, the heightened discernment in the meticulous selection of unrelated individuals, coupled with the meticulous confirmation of their affiliation to the well-defined sub-caste within the original native Pakhtun tribes, fortifies the robustness of our study. Hence, the meticulously calculated forensic parameters and observed alleles possess the capability to accurately and comprehensively depict the homogeneity and diversity within the Pakhtun population.

## 4. Conclusion

The extensive genetic variability inherent in the SE33 STR marker underscores its universal application as a dependable tool in forensic casework worldwide. Our dataset elucidates that the SE33 STR locus emerges as a stalwart tool, furnishing essential forensic parameters crucial for human identification. Grounded in our findings, we assert with confidence that SE33 can be reliably employed in the identification procedures of individuals belongs to sub-continent populations as well as across diverse global populations.

## Conflict of interest

The authors declares that there is no conflict of interest.

## Ethical Approval

The Ethical Board of the University of Gujrat granted the approval (UOG/ORIC/2021/70) to collect the sample and conduct the study.

## References

1. Antonio, A., & Bruce, B. (2016). Handbook of Forensic Genetics: Biodiversity and Heredity in Civil and Criminal Investigation (Vol. 2). World Scientific.
2. Butler, J. M. (2006). Genetics and genomics of core short tandem repeat loci used in human identity testing. *Journal of forensic sciences*, *51*(2), 253-265.

3. Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, *5*(6), 435-445.
4. Degerli I,Yukseloglu E,Korkmaz M,Evaluation of the reliability of SE33 locus in forensic DNA identification in Turkish Population.Turk J Vasc Surg 2022;11(3):1044-8
5. A. Gaviria MV, G.Morejon, J Galarza, et al. Analysis of the most efficient autosomal STRs and genetic data for the locus SE33 in ecuadorian population. Forensic Sci Int: Genetics Supplement Series. 2015;5:93-5.
6. Bhinder, Dr Munir Ahmad; Zahoor, Muhammad Yasir; Sadia, Haleema; Qasim, Muhammad; Perveen, Rukhsana; Anjum, Ghulam Murtaza; Iqbal, Muhammad; Ullah, Najeeb; Shehzad, Waseem; Tariq, Muhammad; And Waryah, Ali Muhammad (2018) "SE33 locus as a reliable genetic marker for forensic DNA analysis systems," *Turkish Journal of Medical Sciences*: Vol. 48: No. 3, Article 23. https://doi.org/10.3906/sag-1801-21
7. Green RL, Lagacé RE, Oldroyd NJ, Hennessy LK, Mulero JJ. Developmental validation of the AmpFℓSTR® NGM SElect™ PCR Amplification Kit: a next-generation STR multiplex with the SE33 locus. Forensic Sci Int-Genet 2013; 7: 41-51.
8. Akram, N., & Farooqi, S. (2012). DNASF: A Statistical Package to Analyze the Distribution and Polymorphism of CODIS STR Loci in a Heterogeneous Population. *J Forensic Res*, *3*, 170.
9. Walsh, P. S., Metzger, D. A., & Higuchi, R. (2013). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques, 54(3), 134-139.
10. Barbisin, M., Fang, R., O'Shea, C. E., Calandro, L. M., Furtado, M. R., & Shewale, J. G. (2009). Developmental validation of the Quantifiler® Duo DNA quantification kit for simultaneous quantification of total human and human male DNA and detection of PCR inhibitors in biological samples. Journal of Forensic Sciences, 54(2), 305-319.
11. Ismail D., Emel H., Mehmet K. (2022). Evaluation of the reliability of SE33 locus in forensic DNA identification in Turkish Population. Medicine Science | International Medical Journal 11(3):1044.

12. Butler JM, Hill CR, Kline MC, Duewer DL, Sprecher CJ, McLaren RS, Rabbach DR, Krenke BE, Storts DR. The single most polymorphic STR locus: SE33 performance in U.S. populations. Forensic Sci Int-Genet 2009; 2: 23-24.
13. Schmid D, Anslinger K, Rolf B. Allele frequencies of the ACTBP2 (=SE33), D18S51, D8S1132, D12S391, D2S1360, D3S1744, D5S2500, D7S1517, D10S2325 and D21S2055 loci in a German population sample. Forensic Sci Int. 2005;151:303-5
14. Al-Snan NR, Messaoudi S, S RB, Bakhiet M. Population genetic data of the 21 autosomal STRs included in GlobalFiler kit of a population sample from the Kingdom of Bahrain. PLoS One. 2019;14:e0220620.
15. Dixit S, Shrivastava P, Kumawat RK, et al. Forensic genetic analysis of population of Madhya Pradesh with PowerPlex Fusion 6C() Multiplex System. Int J Legal Med. 2019;133:803-5.
16. Al-Eitan LN, Darwish NN, Hakooz NM, Dajani RB. Investigation of the forensic GlobalFiler loci in the genetically isolated Circassian subpopulation in Jordan. Gene. 2020;733:144269.
17. Ozeki M, Tamaki K. Allele frequencies of 37 short tandem repeat loci in a Japanese population. Leg Med (Tokyo). 2013;15:342-6.
18. Jung JY, Kim EH, Oh YL, et al. Evaluation of forensic genetic parameters of 12 STR loci in the Korean population using the Investigator HDplex kit. Int J Legal Med. 2017;131:1247-9.
19. Cruz C V-SC, Ribeiro T, Espinheira R. Genetic data for the locus SE33 in a south Portuguese population with PowerplexR ES System. Int Congress Series-elsevier. 2006;1288:427-9
20. Shan, M. A., Børsting, C., & Morling, N. (2021). Forensic application and genetic diversity of 21 autosomal STR loci in five major population groups of Pakistan. International journal of legal medicine, 135, 775-777.