Genetic characterization of Y-chromosomal STR in Hazara ethnic group of Pakistan and confirmation of DYS448 null allele

Atif Adnan[1], Allah Rakha[2], Anam Noor[2], Shahid Nazir[2], Jinfeng Xuan[1], *Jiaxin Xing[1]*, Sibte Hadi[3] and Hao Pang[1*]

1 Department of Forensic Genetics and Biology, School of Forensic Medicine, China Medical University, Shenyang 110122, P.R. China

2 Department of Forensic Sciences, University of Health Sciences Lahore, 54000, Pakistan

3 University of Central Lancashire, School of Forensic and Investigative Sciences, Preston, UK

[*]Corresponding author:

Hao Pang (hpang@cmu.edu.cn)

ABSTRACT:

Pakistan harbors 18 major ethnic groups and Hazara is one of the distinct but smaller groups comprising 0.090% of the total population of Pakistan. Hazara individuals have typical Mongolian facial features and they claim to be descendants of Genghis Khan's army in the first quarter of the thirteenth century AD. Previous study based on limited number of Y STRs with limited number of samples showed that Hazaras are descendants of Mongols. In this study, we genotyped 198 unrelated males (153 Hazara, 15 Baluchi and 30 Mongols) living in Quetta, Baluchistan, Pakistan and Inner Mongolia, China respectively, for a total of 26 (n=153) to 29 (n=92) Y-chromosomal STR loci. 140 unique haplotypes were developed for Hazara population using the PowerPlex Y23 loci. The Y-STR locus showed a genetic diversity ranging from 0.2384 to 0.7918, and an overall discrimination capacity (DC) of 91.5%. The Hazara population samples were profiled for three additional Y-STRs (DYS388, DYS449 and DYS460), which increased the number of unique haplotypes to 144 while the DC increased to 94.11% in Hazara Population of Pakistan. Interestingly null alleles were observed at DYS448 with a specific mutation patterns in 25 individuals of Hazara population. The Hazaras showed significant differences from other local populations of Pakistan as well as neighboring populations, but had considerable genetic affinities to Kazakhs and Mongols. There was a lack of data for Hazara population of Pakistan and our results thus contribute to understanding the potential forensic usefulness of the 26 Y-STRs studied and also shed light on the population history of Pakistani Hazara population.

INTRODUCTION

Pakistan harbors more than 18 major ethnic groups, over 210 million people (http://www.pbscensus.gov.pk). The genetic legacy of many of the ethnic groups in Pakistan has not been studied in depth and data for these ethnic groups is scarce[1]. The origin of Hazara population is disputed. The Hazara could be of Turko-Mongol ancestry, and theorized to be the descendants of an occupying army left in Afghanistan by Genghis Khan[2]. The Hazara population speaks Persian with some Mongolian words. An earlier study done on a limited number of samples (n = 33) had shown them to be closer to populations in Mongolia1 and the star shape phylogenetic tree of this population using Y-STR haplotypes was suggestive of their Mongolian roots[3].

The human Y chromosome markers are remarkably variable, and significantly used for population lineage and human migration studies; this primarily has been due to their ability to differentiate different paternal lineages[4–6]. Molecular population genetics studies have shown that many structural variants exist within human Y chromosome, includes deletions[7–9], duplications[10–12], and inversions[11–15]. Null alleles are known to occur with any PCR-based STR typing system. These could be either due to deletions within the target region or primer binding site mutations[16,17]. DYS448 is located adjacent to the azoospermia factor c (AZFc) region, which is considered important in spermatogenesis and made up of ‘‘ampliconic’’ repeats that act as substrates for nonallelic homologous recombination (NAHR). NAHR might delete longer chunks of the Y chromosome which included DYS448[18].

Short tandem repeats (STRs) can reveal more recent events in population history than single nucleotide polymorphisms (SNPs), because of their higher mutability and allelic polymorphism.

Y-STRs Plays vital role in forensic genetics, particularly for sexual offences analysis; paternity testing and investigations related missing persons [19–22]. Specifically Y STR analysis can detect low quantity of male DNA within high background female DNA as typically challenged within cases of sexual assault[23].

In this study we have studied Pakistani Hazara population using extended sets of Y STRs to determine genetic diversity and to explore their history. 25 null alleles were observed at DYS448 in 153 Hazara male samples. This data has defined the Hazara population better and will be an addition to the Y STR haplotype reference database (YHRD).

MATERIALS AND METHODS

Samples used in the study:

~~Blood samples were collected from, unrelated~~ One hundred and fifty three unrelated ~~153~~ Hazara & 15 Baluchi males residing in Quetta, Baluchistan, Pakistan and 30 Mongols living in Inner Mongolia, P.R. China volunteered to provide blood samples. The samples were collected as blood stains on TIANamp blood Spots DNA Kit (TIANGEN BIOTECH BEIJING CO., LTD). Informed consent was obtained from all participants before obtaining the sample.~~All participants gave their informed consent either orally (in case they could not write) or in writing, after the study aims and procedures were explained them carefully. The study was approved by the~~ The ethical review committee of China Medical University, Shenyang, Liaoning Province, P.R. China ~~and in accordance with the standards of the Declaration of Helsinki~~approved the project and obtaining of samples.

DNA extraction:

Genomic DNAs were extracted from blood stains by xxxxxxx ~~using TIANamp blood Spots DNA Kit (TIANGEN BIOTECH BEIJING CO., LTD)~~ according to the manufacturer's instructions .(TIANGEN BIOTECH BEIJING CO., LTD).

Commented [SH1]: Pl say how this was done e.g. a 1 mm punch was taken from the stain and was heated at 56 C for 30 mins and so on.

PCR amplification and Y-STR typing:

26 Y STR loci (which contains PowerPlex Y23 loci and three additional Y-STRs DYS388, DYS449 and DYS460) were amplified using the Goldeneye® 26Y system (PEOPLESPOT R&D, China) as described elsewhere[24]. The 2800M (Promega Corporation, USA) was used as positive control in all batches. All PCR batches included a PCR negative control. The PCR products were separated and detected by capillary electrophoresis on an ABI 3500 Genetic Analyzer (Applied Biosystems, USA). The genotyping results were analyzed using GeneMapper ID v3.2 (Applied Biosystems, USA).

Identification of DYS448 deletions

Yfiler™ Plus PCR Amplification Kit (Applied Biosystems, USA) and Microreader™ 29Y Direct ID System (DYS393, DYS570, DYS19, DYS392, DYS549, Y-GATAH4, DYF387S1, DYS460, DYS458, DYS481, DYS635, DYS448, DYS533, DYS627, DYS456, DYS389I, DYS390, DYS389II, DYS438, DYS576, DYS449, DYS391, DYS439, DYS437, DYS385a/b, DYS643, DYS518) kits were used to confirm the null alleles at DYS448 according to the manufacturer's protocol. 47 samples were used for this exercise which generated haplotypes comprising of Yfiler™ Plus kit and 30 Y STRs.

DYS 448 Sequencing

25 Hazara samples showed DYS448 null alleles (detected as allelic dropouts) in three Y STR kits (GoldenEye Y26, Microreader™ Y29 and Yfiler™ Plus). The region was amplified using a

nested PCR to reduce unspecific PCR products. First-round PCR was performed with the forward primers 5′-AGAAGAATGGATGATGAGAGATCTG-3′ and the reverse primer 5′-CATCTCCTGGAGTCAGACAGTAATC-3′, resulting in an 1193bp PCR product. Second-round of PCR was performed using forward primer

5′-GCATGCCAACAACAAGGATAT-3′ and the reverse primer

5′-GTTTCCCTGACTCCATTAGCTG-3′ resulting in approximately 500bp product. Both-rounds of PCR reaction was carried out in a 20µl volume containing 2µl of 10 × LA PCR™ Buffer II (Mg^{2+} Plus) (TaKaRa Bio Inc., Dalian, China.), 2µl dNTP Mixture (2.5 mM each) (TaKaRa), 0.25 µl TaKaRa LA Taq™ (5 units/µl) (TaKaRa), 14.75µl ddH_2O and 1µl of genomic DNA. In first round 1µl of DNA was used and in second round we 1µl of 20× to 100× PCR products of first round, were used. Samples were amplified in a GeneAmp 9700 PCR machine (Applied Biosystems, Foster City, CA, USA) under following conditions for both rounds of amplification: one denaturation cycle of 94 ℃ for 1 min; 35 amplification cycles of 94 ℃ for 30 s, 65 ℃ for 30 s and 72℃ for 1 min; and one full extension cycle of 72 ℃ for 5 min. The PCR products around 500bp were purified by centrifugation by using a Suprec-02 tube (Takara) and then were sequenced directly with second round primers. The Sequencing reaction was carried out in 20µl reaction volume, which contains 1µl of PCR product, 8 µl of BigDye (2.5×), each primer (F and R) (3.2 mM) 1µl and dH_2O 10µl under following conditions 96 ℃ for 1 min; 25 amplification cycles of 96 ℃ for 10 sec, 50 ℃ for 5 sec; and one full extension of 60 ℃ for 4 min. Then purified PCR products were used for sequencing by capillary electrophoresis using 3500 Genetic Analyzer (Thermo Fisher Scientific) according to manufacturer's manual.

Commented [SH2]: These can be tables

Statistical analyses:

Allelic frequencies were developed through simple count method. ~~computed using the method of gene counting.~~ ~~Gene diversity (GD) or haplotype diversity (HD) was calculated by~~ Nei's formula was used to calculate Gene diversity (GD) which is equivalent to haplotype diversity (HD) for Y STRs [25]. Y STR data generated using Goldeneye® 26Y kit and the Yfiler™ Plus kit allowed the development of ~~H~~haplotype diversities~~y~~ (HD) were~~as~~ calculated~~estimated~~ for the 26/27 Y STR loci included in these kits and also the haplotype diversities for the ~~9 Y STR loci of~~ minimal (9 loci) ~~haplotype, 11 loci of;~~ extended (11 loci); ~~haplotype,~~ Powerplex Y (12 loci); ~~,~~ Yfiler™ (17 loci) and ~~,~~ Powerplex Y 23 (23 loci)~~.~~ ~~Goldeneye® 26Y kit, Yfiler™ Plus (27 loci) and a~~ An extended haplotype comprising of Yfiler™ Plus with three additional loci (30 loci) was also developed and haplotype diversity calculated. Discrimination capacity (DC) was calculated as the ratio of unique haplotypes in the sample. Match probabilities (MP) were calculated as ΣP_i^2, where P*i* is the frequency of the *i*-th haplotype. Genetic distances between Hazara population and ~~reference population analysis of~~ other world populations were calculated using molecular variance (AMOVA) and multidimensional scaling (MDS). Both these analyses were ~~that exploits variation among populations were~~ performed using YHRD online tools (http://www.yhrd.org) based on pairwise R*st* values.

Phylogenetic analysis:

A neighbor-joining phylogenetic tree was constructed for the Hazara and the reference populations based on a distance matrix of R*st* using the Mega7 software[26]. We also predicted Y-SNP haplogroups in the samples from Y STR haplotypes using the Y-DNA Haplogroup Predictor NEVGEN (http://www.nevgen.org).

The median joining network

Using the program Network 4.1.1.2.(http://www.fluxus-engineering.com/sharenet.htm), median joining network was constructed from data of Hazara population for 23 Y STRs (DYS19, DYS389II-I, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS456, DYS458, DYS635, Y_GATA_H4, DYS576, DYS570, DYS481, DYS533, DYS549, DYS643, DYS460, DYS449 and DYS388).

Linear discriminant analysis and Correlation

Linear discriminant analysis (LDA) was performed on Hazara, Central Asia, East Asia; Middle East and South West Asian samples using R program[27] and correlation between STRs were calculated using XLSTAT (http://www.xlstat.com/en/). The multi-copy marker like (DYS385ab) and markers that have null alleles or duplication variants in the Hazara population or any of the reference populations were excluded from the analysis. For DYS389I and DYS389II, we have subtracted DYS389I from DYS389II and used DYS389II-I for analysis.

RESULTS AND DISCUSSION

~~We successfully obtained~~ Y-STR haplotypes from 153 male individuals in Hazara population were developed. The data showed~~with~~ 25 null alleles for DYS448. 144 (out of 153) different haplotypes were observed for 26 Y STRs while 47 (n =47) haplotypes observed for 30 Y STRs (Supplementary material Table S1A). Allelic frequencies and GD values of various sets of Y-STR loci in Hazara individuals were calculated (Table 1). 199 alleles were observed for all 30 Y STRs. The allelic frequencies of single-copy STR loci ranged from 0.0065 to 0.8627, while the frequencies of genotypes of multi-copy STR loci ranged from 0.0065 to 0.4183. DYS385 (0.960), DYS387S1 (0.813) and DYS627 (0.792) revealed highest GD values, revealing a high level of genetic polymorphism in the Hazara population of Pakistan. Hazara population of

Pakistan showed almost same allelic frequencies on minimum haplotype (MH) STRs with Mongols of Inner Mongolia and Kazak population of Kazakhstan, while have differences on GD values. However, since the focus of this study was on Hazara population, we did not aim to discuss Baluchi and Mongol population in detail due to small data set but the genotype data for Baluchi and Mongol populations was tabulated (Supplementary material Table S1B and Supplementary material Table S1C). Allelic frequencies and GD values of Baluchi and Mongols were also calculated (Table 2 and Table 3).

HAPLOTYPE DIVERSITY

The increase in the number of Y STR loci enhanced haplotype diversity for the set of 153 male samples. This ranged from 0.93 to almost 1 as the number of loci increased from 9 minimum haplotype (MH) STRs loci to 30 extended STRs loci (Table 4). Haplotype data already made accessible via YHRD under accession number YC000340. The increase of haplotype diversity was small between the Powerplex Y 23 and Goldeneye® 26Y kit (0.9981 & 0.9985 respectively). This increase in haplotype diversity was parallel to the increase in the number of unique haplotypes which ranged from 75-97% (Table 4). In this set of 153 male samples the Goldeneye® 26Y kit generated 144 haplotypes among which 140 were unique haplotypes whereas if the PowerPlex Y23 kit loci were considered there were 140 haplotypes with only 132 being unique. This substantial increase in the proportion of unique haplotypes revealed the greater power of discrimination of Goldeneye 26 Y STR loci in the Hazaras. This trend was also noted for the smaller set of 45 samples for the Baluchi and Mongol populations which were profiled using Yfiler™ plus and 2 additional Y STR loci as well (Table 5 and 6). Thus both, Goldeneye® 26Y and Yfiler™ systems can potentially be used for forensic casework and population genetic studies for smaller ethnic groups like the Hazaras in Pakistan.

GENETIC RELATIONSHIP OF THE HAZARAS WITH MONGOL POPULATION

Previous studies based on uni-parental markers suggesting that Hazaras are the male descendants of Genghis Khan, or other Mongol populations[1,28]. In order to study the genetic relationship of the Hazara population, the genotypes of the sample set were used to calculate the R*st* values for the Harazas (Supplementary Table S1A). Then the multidimensional scaling analysis (MDS) analysis was performed among the Hazara population of Pakistan and 27 other reference populations based on R*st* values (Table 7 and Figure 1). Hazara population was located closest to previously studied Hazara population (R*st* -0.004), Kohistani population (R*st* -0.0049) and Kazakh population (R*st* -0.0048) of Kazakhstan. The Khalkh population (R*st* 0.0117) of Central Mongolia and Mongolian population (R*st* 0.0158) of Ulaanbaatar, Mongolia also appeared to be clustered with the Hazara population. It was established earlier that at least third to half of Hazara chromosomes were of East Asian origin[29]. In current study linear discrimination analysis (LDA) analysis was performed (Figure 2), which showed that Hazara not even have East Asian origin there is also influence of Central Asian populations. LDA was also performed between Hazara population of Pakistan and five other populations including Hazara population of Afghanistan, Kazakh and Uyghur population of Xinjiang, Mongol population from Inner Mongolia China (Current Study) and Baluchi population from Pakistan (Current Study). Evolutionary relationships between the Hazara population of Pakistan and these five populations were also inferred from the Neighbor-joining tree based on the F*st* values (Figure 3). LDA plot showed the association between 4 populations (Hazara Pakistan, Hazara Afghanistan, Kazakh of Xinjiang and Mongol of Inner Mongolia China) while Uyghur was placed on the right side of LDA and Baluchi population was placed on left side (Figure 4).

Evolutionary relationships between the Hazara population of Pakistan and other reference populations were inferred from the Neighbor-joining tree based on the R_{ST} values (Figure 5). The similarities between the Hazara and the Kazak population were consistent with historical records, which established that present Kazak population was an admixture of Central Asian, Mongolian, Uyghur and other populations formed around the 13th century[30].

INFERENCE OF ANCESTRY BASED ON Y STRs

Y chromosomal and mtDNA variations are not interrupted by recombination, so is conserved in both lineages and associates strongly with continental regions. The Hazara samples exhibited two main haplogroups M217 in 71 samples and M420 in 21 samples (147 out of 153 Hazara Pakistan samples used for this analysis). M217 is found at high frequencies among Central Asian populations (Mongolians), indigenous Siberians (Kazakhs) and M420 which is distributed in a large region in Eurasia, extending from Scandinavia and Central Europe to southern Siberia and South Asia[3,31,32]. These results support the view that Hazaras of Pakistan may have Mongol[1,32] and Kazakh origin. The median joining network based on 23 Y STR loci shows the extent of variation within Hazara population (Figure 6).

LDA was performed on the Hazara population of Pakistan, Central Asian, Middle Eastern, south west Asian and East Asian population samples to explore the potential for markers that are ancestry-informative. Figure 2A shows all individual samples plotted on the two LDA factors (F1 and F2). The first factor (DYS635) explained the majority (46.52%) of the variation while second factor (DYS389II) explained the minority (34.79%) of variations. The markers DYS635 and DYS389II had the largest correlation coefficient (0.7064 and 0.5188) with the first and second factor, respectively (Figure 2B). LDA Plot showed the association of Hazara population of Pakistan with East and Central Asian populations.

PHYSICAL CHARACTERIZATION OF DYS448 DELETIONS

Most of human Y chromosomes have a single DYS448 allele, we observed that some have carried null alleles. Null alleles usually can occur due to deletions within the target region or primer binding sites or by primer binding site mutations that destabilize hybridization of at least one of the primers flanking the target region[33–36]. By using GoldenEye Y26 system null types at DYS448 were observed from 26 individuals (Figure 7) in Hazara population of Pakistan. The phenomena of null allele was previously reported, in which other commercial kits were used[37–42]. Current population study represents the high frequencies of null allele at DYS448 when compared with previously reported population to date (Table 8). When we used Microreader Y29 system the results were same as with GoldenEye Y26. Then finally we used Yfiler™ Plus kit which also showed null alleles in these 25 individuals (out of 26).

One individual who shown allele 20 at the DYS448 locus was selected and sequenced to serve as a reference for alignment. All samples who showed null allele at DYS448 (25 samples), also showed a specific pattern of deletion which might be specific to Hazara population of Pakistan (Table 9). To confirm this phenomenon more studies are required on Hazara population of Pakistan and Afghanistan. Frequency of null allele at DYS448 is more common in Asia more specifically in East and Central Asia when it compared to rest of the world[18,38]. Commercial companies should give special attention while designing DYS448 primers.

CONCLUDING REMARKS

In current study we have genotyped 153 Hazara individuals for 26 Y STRs, and 45 Baluchi & Mongol individuals for 29 Y STRs. The genetic variation in Hazara population and its comparison to several relevant populations were done using various statistical tests. This showed

that the Hazara population has close affinity to the Central Asian and Mongol populations. In previous study[3] star-cluster profile was 10-16-25-10-11-13-14-12-11-11-11-12-8-10-10, for the loci DYS389I-DYS389b-DYS390-DYS391-DYS392-DYS393-DYS388-DYS425-DYS426-DYS434-DYS435-DYS436-DYS437-DYS438-DYS439. In current study most frequent haplotype was 15-13-29-24-10-11-13-14-11-12 for loci DYS19-DYS389I-DYS389II-DYS390-DYS391-DYS392-DYS393-DYS437-DYS438-DYS439 found in 43 individuals while 14-13-29-24-8-11-13-14-11-11 found in 9 individuals and 15-13-29-24-11-11-13-14-11-12 found in 8 individuals. These haplotypes are most frequent haplotype in Mongols and Kazakhs[43]. Allelic Frequencies on these 10 Y STRs are almost similar to Kazak43 population of Kazakhstan Central Asia and somehow similar to Mongol population of Inner Mongolia. Our results contributed towards deciphering the origin of Hazara population of Pakistan and supported the view that they had Kazakh and Mongol origins. The forensic parameter calculations showed high discriminatory power with potential applications for forensic casework. We detected 25 null alleles at the locus DYS448 that equated to 16% of null alleles at this locus. This was the highest percentage of DYS448 null alleles in a population to date. The null allele sequencing results showed a specific deletion patterns, which need to be further, investigated in the Afghani Hazaras and also Mongol populations for assessing the impact of the sequence.

ACKNOWLEDGEMENTS

We are very grateful to the volunteers in our study in particular Fengrui Li, Mannis Van Ovan, Muhammad Farhan, Ammar Sabir Cheema, Xiaoni Zhan and Fatima. This project is supported by the National Natural Science Foundation of P. R. China (NSFC, No. 81471826).

AUTHOR CONTRIBUTIONS

H.P. and A.A. developed the idea and designed the experimental approach; A.A., A.R., A.N. and S.N. had collected the samples, A.A. performed the experiments, supported by J.X., Jf. X.; A.A. analyzed the data, A.A. wrote the initial manuscript, which is modified by A.A., H.P., S.H., A.R. and A.N. All authors reviewed the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

FIGURES

Figure 1:

Two-dimensional plot from multi-dimensional scaling analysis of R*st*-values based on Yfiler haplotypes for Hazara population of Pakistan with reference to 27 other populations.

Figure 2:

A) LDA Analysis between Hazara, Central Asia, South Asia, Russia and East Asian populations.
B) Correlation coefficient of different Y STR markers and the two LDA factors

Figure 3:

A) LDA Analysis between Hazara Pakistan, Hazara Afghanistan, Baluchi Pakistan, Kazakh Xinjiang, Uyghur Xinjiang and Mongolian population of Inner Mongolia China populations.
B) Correlation coefficient of different Y STR markers and the two LDA factors

Figure 4:

Neighbour-joining phylogenetic tree of the Hazara population and the Central Asia, South Asia, Russia and East Asian populations based on a distance matrix of R*st*.

Figure 5:

Neighbour-joining phylogenetic tree of the Hazara population and the Central Asia, South Asia, Russia and East Asian populations based on a distance matrix of F*st*.

Figure 6:

The median joining network of Hazara population of Pakistan based on 23 Y STR

Figure 7:

Null types at DYS448 were observed from 25 to 26 individuals based on 3 different kits (GoldenEye Y 26, Microreader Y 29 and Yfiler Plus)

TABLES:

Table 1:

Allele Frequencies of Hazara population of Pakistan on 30 Y-STRs

Table 2:

Allelic frequencies of Baluchi population of Pakistan based on 29 Y STRs

Table 3:

Allelic frequencies of Mongol population of Inner Mongolia, China based on 29 Y STRs

Table 4:

Forensic statistical parameters in Hazara population of Pakistan on 8 levels

Table 5:

Forensic statistical parameters in Baluchi population of Pakistan on 8 levels

Table 6:

Forensic statistical parameters in Mongol population of Inner Mongolia China on 8 levels

Table 7:

Pairwise R*st* values between Hazara Population of Pakistan and other 27 reference populations

Table 8:

Frequencies of null allele at DYS448 in various ethnic groups across continents

Table 9:

Sequence in the relevant flanking and repeat region of the DYS448 locus for normal and null alleles

SUPPLIMENTRY TABLES

Table S1A

The haplotype distributions and haplotype frequencies of Hazara population from Pakistan (n = 153).

Table S1B

The haplotype distributions and haplotype frequencies of Baluchi population from Pakistan (n = 15).

Table S1C

The haplotype distributions and haplotype frequencies of Mongol population from Inner Mongolia China (n = 30).

Tables:

Table 1: Allele Frequencies of Hazara population of Pakistan on 30 Y-STRs

Allele	***DYS19***	***Allele***	***DYS518***	***Allele***	***DYS533***	***Allele***	***DYS390***	***Allele***	***DYS439***	***Allele***	***DYS643***	***Allele***	***DYS576***
12	0.0065	37	0.0638	10	0.0131	19	0.0196	8	0.0065	8	0.0196	15	0.0065
13	0.0458	38	0.0638	11	0.1177	22	0.0458	9	0.0261	9	0.0784	16	0.1242
14	0.2157	39	0.6170	12	0.6013	23	0.0850	10	0.1307	10	0.7582	17	0.5621
15	0.6275	40	0.1489	13	0.0654	24	0.6732	11	0.1961	11	0.0850	18	0.1961
16	0.0784	41	0.0851	14	0.1438	25	0.1569	12	0.5882	12	0.0523	19	0.0850
17	0.0261	42	0.0213	15	0.0588	27	0.0196	13	0.0523	13	0.0065	20	0.0261
AC	6	AC	6	AC	6	AC	6	AC	6	AC	6	AC	6
GD	0.5544	GD	0.5939	GD	0.5999	GD	0.5155	GD	0.5989	GD	0.4113	GD	0.6263
Allele	**DYS388**	***Allele***	**DYS389I**	***Allele***	**DYS389II**	***Allele***	**DYS438**	***Allele***	**DYS391**	***Allele***	**DYS549**	***Allele***	**DYS456**
12	0.2418	12	0.0588	28	0.0392	9	0.0588	8	0.1307	10	0.0327	13	0.0523
13	0.5490	13	0.8235	29	0.7451	10	0.1699	9	0.0327	11	0.1765	14	0.0654
14	0.1046	14	0.0915	30	0.1111	11	0.7059	10	0.6078	12	0.7124	15	0.2484
15	0.0784	11,14	0.0131	31	0.0719	9,11	0.0065	11	0.2222	13	0.0719	16	0.1111
16	0.0261	15	0.0131	32	0.0327	12	0.0588	12	0.0065	14	0.0065	17	0.5229
AC	5	AC	5	AC	5	AC	5	AC	5	AC	5	AC	5
GD	0.6264	GD	0.3117	GD	0.4275	GD	0.4690	GD	0.5667	GD	0.4580	GD	0.6498

Allele	DYS460	Allele	DYS392	Allele	DYS570	Allele	GATA_H4	Allele	DYS393	Allele	DYS437	Allele	DYS448
7	0.0065	10	0.0065	15	0.0131	10	0.0196	12	0.1242	14	0.8628	Null	0.1699
9	0.6013	11	0.8431	16	0.1046	11	0.6797	13	0.8562	15	0.1373	18	0.0065
10	0.1307	12	0.0261	17	0.0850	12	0.1895	14	0.0196	AC	2	19	0.3072
11	0.2418	13	0.0850	18	0.1503	13	0.1111	AC	3	GD	0.2384	20	0.2614
12	0.0196	14	0.0131	19	0.4837	AC	4	GD	0.2528			21	0.1307
AC	5	15	0.0196	20	0.1438	GD	0.4925					22	0.1111
GD	0.5661	16	0.0065	21	0.0196							23	0.0065
		AC	7	AC	7							24	0.0065
		GD	0.2824	GD	0.7087							AC	8
												GD	0.7840
Allele	**DYS635**	**Allele**	**DYS449**	**Allele**	**DYS458**	**Allele**	**DYS481**	**Allele**	**DYS627**	**Allele**	**DYS387S1**	**Allele**	**DYS385a,b**
18	0.0131	26	0.0261	12	0.0065	20	0.0131	16	0.0213	35,35	0.0213	10,11	0.0065
19	0.0065	27	0.0654	14	0.0327	22	0.0980	17	0.0638	36,36	0.0638	10,12	0.0065
20	0.1895	28	0.5686	15	0.0784	23	0.1634	18	0.0638	36,37	0.0213	10,13	0.0065
21	0.4837	29	0.0654	16	0.2418	24	0.0523	19	0.0213	36,38,40	0.0213	11,11	0.0131
22	0.1111	30	0.0980	17	0.4967	25	0.0523	20	0.3830	36,38,39	0.0213	11,12	0.0065
23	0.1503	31	0.0327	18	0.0915	26	0.0458	21	0.2128	36,39	0.0213	11,13	0.0065
24	0.0392	32	0.0327	18.2	0.0131	27	0.0980	22	0.1277	36,40	0.0213	11,14	0.1307
25	0.0065	33	0.0719	19	0.0261	28	0.4052	23	0.0851	37,38,39	0.4043	11,15	0.0196
AC	8	34	0.0327	20	0.0131	29	0.0719	24	0.0213	37,38,40	0.0213	11,16	0.0065
GD	0.6980	27,39	0.0065	AC	9	AC	9	AC	9	37,38	0.0638	11,17	0.0131
		AC	10	GD	0.6826	GD	0.7821	GD	0.7919	37,39	0.1064	12,13	0.0654
		GD	0.6537							37,40	0.0638	12,14	0.4183
										38,39	0.0213	12,15	0.1046
										38,40	0.0213	12,16	0.0261
										39,39	0.0851	12,18	0.0131
										39,40	0.0213	12,19	0.0065
										GD	0.813	13,14	0.0654
												13,15	0.0065
												13,16	0.0196
												13,17	0.0131
												13,18	0.0065
												13,19	0.0065
												13,20	0.0065
												14,16	0.0196
												14,15	0.0065
												GD	0.9600

AC= Allele count GD= Genetic diversity

Table 2:

Allelic frequencies of Baluchi population of Pakistan based on 29 Y STRs

Allele	DYS19	Allele	DYS549	Allele	DYS643	Allele	DYS389I	Allele	DYS391	Allele	DYS437	Allele	DYS438
14	0.2667	11	0.2000	10	0.6667	12	0.3333	10	0.7333	14	0.2667	9	0.2667
15	0.5333	12	0.3333	11	0.2000	13	0.6000	11	0.2000	15	0.4667	10	0.4000
16	0.2000	13	0.4667	12	0.1333	14	0.0667	12	0.0667	16	0.2667	11	0.3333
GD	0.6476	GD	0.6762	GD	0.5333	GD	0.5619	GD	0.4476	GD	0.6857	GD	0.7048
Allele	**DYS635**	**Allele**	**GATA_H4**	**Allele**	**DYS533**	**Allele**	**DYS460**	**Allele**	**DYS385b**	**Allele**	**DYS390**	**Allele**	**DYS392**
21	0.5333	11	0.6000	11	0.3333	10	0.3333	14	0.2000	22	0.3333	11	0.5333
23	0.2000	12	0.2667	12	0.6000	11	0.6000	15	0.1333	23	0.1333	12	0.0667
24	0.2667	13	0.1333	13	0.0667	12	0.0667	16	0.2667	24	0.4000	13	0.0667
GD	0.6476	GD	0.5905	GD	0.5619	GD	0.5619	18	0.4000	25	0.1333	14	0.3333
								GD	0.7619	GD	0.7429	GD	0.6381
Allele	**DYS393**	**Allele**	**DYS439**	**Allele**	**DYS456**	**Allele**	**DYS458**	**Allele**	**DYS481**	**Allele**	**DYS448**	**Allele**	**DYS518**
11	0.2000	10	0.6000	13	0.2667	15	0.3333	21	0.0667	18	0.1333	36	0.1333
12	0.3333	11	0.1333	14	0.2000	16	0.1333	22	0.1333	19	0.4000	37	0.1333
13	0.2667	12	0.1333	15	0.2667	17	0.2000	23	0.6000	20	0.1333	38	0.4000
14	0.2000	13	0.1333	16	0.2667	18	0.3333	24	0.2000	21	0.2000	41	0.2667
GD	0.7905	GD	0.6286	GD	0.8000	GD	0.7714	GD	0.6191	22	0.1333	42	0.0667
										GD	0.8000	GD	0.7810

Allele	DYS627	Allele	DYF387S1a	Allele	DYF387S1b	Allele	DYS576	Allele	DYS449	Allele	DYS385a	Allele	DYS389II
17	0.2000	35	0.1333	37	0.2667	14	0.1333	27	0.0667	9	0.2000	27	0.0667
18	0.2000	36	0.5333	38	0.3333	15	0.4000	28	0.0667	10	0.0667	28	0.2667
19	0.0667	37	0.2000	39	0.0667	16	0.0667	29	0.2000	11	0.2000	29	0.3333
20	0.4000	38	0.0667	40	0.2667	17	0.1333	30	0.2000	12	0.0667	30	0.1333
22	0.1333	39	0.0667	41	0.0667	18	0.1333	31	0.4000	13	0.4000	31	0.0667
GD	0.7905	GD	0.6952	GD	0.7905	19	0.1333	33	0.0667	15	0.0667	39	0.1333
						GD	0.8191	GD	0.8000	GD	0.8000	GD	0.8286

Allele	DYS570
13	0.1333
14	0.1333
15	0.2000
16	0.1333
17	0.0667
18	0.1333
19	0.0667
20	0.1333
GD	0.9238

GD= Genetic diversity

Table 3:

Allelic frequencies of Mongol population of Inner Mongolia, China based on 29 Y STRs

Allele	DYS391	Allele	DYS393	Allele	*DYS460*	Allele	DYS456	Allele	DYS437	Allele	DYS438	Allele	*DYS19*
9	0.0667	12	0.3000	9	0.0667	14	0.3000	13	0.2000	9	0.0667	14	0.1667
10	0.8000	13	0.3333	10	0.6667	15	0.5667	14	0.6667	10	0.8000	15	0.4333
11	0.1333	14	0.3667	11	0.2667	16	0.1333	15	0.1333	11	0.1333	16	0.2667
GD	0.3494	GD	0.6874	GD	0.4966	GD	0.5908	GD	0.5149	GD	0.3494	17	0.1333
												GD	0.7195
Allele	**DYS385a**	**Allele**	**DYS389I**	**Allele**	**DYS390**	**Allele**	**DYS439**	**Allele**	**GATA_H4**	**Allele**	**DYS549**	**Allele**	**DYS643**
11	0.2667	11	0.0333	22	0.0333	10	0.2667	10	0.1000	11	0.0667	9	0.1333
12	0.4000	12	0.4000	23	0.3667	11	0.4667	11	0.4000	12	0.6333	10	0.4667
13	0.2000	13	0.4000	24	0.3000	12	0.1667	12	0.4667	13	0.2667	11	0.2667
14	0.1333	14	0.1667	25	0.3000	13	0.1000	13	0.0333	14	0.0333	12	0.1333
GD	0.7356	GD	0.6736	GD	0.7081	GD	0.6966	GD	0.6322	GD	0.5402	GD	0.6989
Allele	**DYS635**	**Allele**	**DYS533**	**Allele**	**DYS389II**	**Allele**	**DYS392**	**Allele**	**DYS448**	**Allele**	**DYS458**	**Allele**	**DYS576**
19	0.0333	9	0.0667	27	0.0667	7	0.0333	17	0.0333	14	0.0333	15	0.2000
20	0.0667	10	0.0667	28	0.2667	11	0.2000	18	0.1667	15	0.0667	16	0.1333
21	0.2667	11	0.5000	29	0.5000	12	0.1667	19	0.2667	16	0.2667	17	0.3333
22	0.3333	12	0.3000	30	0.0667	13	0.4333	20	0.2667	17	0.2333	18	0.2333

23	0.3000	13	0.0667	31	0.0333	14	0.1333	21	0.1000	18	0.3333	19	0.0667
GD	0.7471	GD	0.6690	32	0.0667	16	0.0333	22	0.1667	19	0.0667	20	0.0333
				GD	0.6874	GD	0.7494	GD	0.8184	GD	0.7793	GD	0.7977
Allele	**DYS481**	**Allele**	**DYS570**	**Allele**	**DYF387s1b**	**Allele**	**DYS627**	**Allele**	**DYF387s1a**	**Allele**	**DYS385b**	**Allele**	**DYS449**
20	0.0667	15	0.0333	36	0.0333	19	0.1667	34	0.0333	11	0.0667	26	0.0667
21	0.1000	16	0.2000	37	0.1333	20	0.1333	35	0.1000	12	0.2000	27	0.3333
22	0.1000	17	0.1000	38	0.1667	21	0.2333	36	0.3333	13	0.3333	28	0.1333
23	0.1333	18	0.1333	39	0.3000	22	0.3000	37	0.3000	14	0.0333	29	0.1667
24	0.1667	19	0.4000	40	0.2333	23	0.1000	38	0.0667	16	0.0333	30	0.0667
25	0.4333	20	0.1333	41	0.1333	24	0.0333	39	0.1000	17	0.1667	31	0.0667
GD	0.7678	GD	0.7793	GD	0.6989	25	0.0333	41	0.0667	18	0.0333	32	0.0333
						GD	0.8253	GD	0.5402	20	0.0333	33	0.0667
										21	0.1000	34	0.0333
										GD	0.8299	35	0.0333
												GD	0.8506

Allele	***DYS518***
33	0.0333
35	0.0667
36	0.1000
37	0.2000
38	0.2667
39	0.1667
40	0.0333
41	0.0333
42	0.0333
43	0.0333
44	0.0333
GD	0.8690

GD= Genetic diversity

Table 4:

Forensic statistical parameters in Hazara population of Pakistan on 8 levels

Haplotypes	***MH 9 Loci***	***SWGDAM 11 loci***	***PowerPlex Y12 loci***	***Y-filer 17 loci***	***PowerPlex Y23 loci***	***26-Y Loci***	***Y-filer Plus 27 loci***	***30 Y Loci***
Sample size	***153***	***153***	***153***	***153***	***153***	***153***	***47***	***47***
Haplotype diversity	0.9316 +/- 0.0157	0.9485 +/- 0.0129	0.9485 +/- 0.0129	0.9942 +/- 0.0019	0.9981 +/- 0.0012	0.9985 +/- 0.0012	0.9991 +/- 0.0046	1.0000 +/- 0.0044
RMP	0.07442128	0.0577134	0.0577134	0.01231458	0.00835716	0.00800872	0.02223081	0.0213234
DC	0.470588235	0.529411765	0.529411765	0.764705882	0.91503268	0.941176471	0.978723404	1
Haplotypes	72	81	81	117	140	144	46	47
Unique Haplotype	54	63	64	97	132	140	45	47
% of Unique Haplotype	0.75	0.777777778	0.790123457	0.829059829	0.942857143	0.972222222	0.97826087	100

Table 5:

Forensic statistical parameters in Balochi population of Pakistan on 8 levels

	MH 9 Loci	***SWGDAM 11 loci***	***PowerPlex Y12 loci***	***Y-filer 17 loci***	***PowerPlex Y23 loci***	***26-Y Loci***	***Y-filer Plus 27 loci***	***29-Y Loci***
Sample size	15	15	15	15	15	15	15	15
Gene diversity	0.9810 +/- 0.0308	1.0000 +/- 0.0243	1.0000 +/- 0.0243	1.0000 +/- 0.0243	1.0000 +/- 0.0243	1.0000 +/- 0.0243	1.0000 +/- 0.0243	1.0000 +/- 0.0243
RMP	0.0844	0.0667	0.0667	0.0667	0.0667	0.0667	0.0667	0.0667
DC	0.866	1	1	1	1	1	1	1
Haplotypes	13	15	15	15	15	15	15	15
Unique Haplotype	11	15	15	15	15	15	15	15
% of Unique Haplotype	0.85%	100%	100%	100%	100%	100%	100%	100%

Table 6:

Forensic statistical parameters in Mongol population of Inner Mongolia China on 8 levels

	MH 9 Loci	***SWGDAM 11 loci***	***PowerPlex Y12 loci***	***Y-filer 17 loci***	***PowerPlex Y23 loci***	***26-Y Loci***	***Y-filer Plus 27 loci***	***29-Y Loci***
Sample size	30	30	30	30	30	30	30	30
Gene diversity	0.9931 +/- 0.0105	0.9931 +/- 0.0105	0.9931 +/- 0.0105	0.9977 +/- 0.0094	1.0000 +/- 0.0086	1.0000 +/- 0.0086	1.0000 +/- 0.0086	1.0000 +/- 0.0086
RMP	0.04	0.04	0.04	0.0356	0.0332667	0.0332667	0.0332667	0.0332667
DC	0.9	0.9	0.9	0.966666667	1	1	1	1
Haplotypes	27	27	27	29	30	30	30	30
Unique Haplotype	24	24	24	28	30	30	30	30
% of Unique Haplotype	0.888888889	0.888888889	0.888888889	0.965517241	100	100	100	100

Table 7:

Pairwise R*st* values between Hazara Population of Pakistan and other 27 reference populations

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]
[1]	-	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
[2]	0.013	-	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
[3]	0.026	0.058	-	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
[4]	0.113	0.149	0.140	-	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
[5]	0.094	0.147	0.143	0.006	-	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
[6]	0.147	0.152	0.142	0.010	0.004	-	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
[7]	0.030	0.113	0.161	0.056	0.057	0.066	-	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
[8]	0.038	0.038	0.073	0.068	0.059	0.072	0.054	-	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
[9]	0.036	0.095	0.090	0.012	0.016	0.025	0.047	0.027	-	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
[10]	0.023	0.033	0.057	0.048	0.042	0.053	0.042	0.006	0.013	-	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
[11]	0.051	0.023	0.046	0.094	0.093	0.098	0.080	0.032	0.050	0.017	-	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
[12]	0.058	0.290	0.289	0.071	0.090	0.082	0.110	0.183	0.112	0.142	0.187	-	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
[13]	-0.005	0.000	0.072	0.127	0.126	0.131	0.083	0.029	0.071	0.019	0.027	0.270	-	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
[14]	0.014	0.079	0.177	0.177	0.193	0.191	0.093	0.085	0.132	0.083	0.089	0.320	0.065	-	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
[15]	0.016	0.028	0.098	0.191	0.187	0.194	0.145	0.045	0.123	0.064	0.069	0.342	0.040	0.084	-	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
[16]	0.016	0.203	0.164	0.126	0.153	0.136	0.242	0.161	0.110	0.114	0.122	0.272	0.194	0.319	0.263	-	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
[17]	-0.004	-0.004	0.062	0.156	0.158	0.161	0.126	0.026	0.089	0.030	0.017	0.325	0.005	0.083	0.009	0.184	-	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
[18]	0.045	0.095	0.101	0.161	0.181	0.172	0.181	0.115	0.112	0.087	0.043	0.259	0.097	0.169	0.161	0.109	0.084	-	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
[19]	-0.005	0.054	0.106	0.092	0.091	0.088	0.111	0.045	0.051	0.024	0.018	0.255	0.050	0.156	0.112	0.114	0.054	0.038	-	-1	-1	-1	-1	-1	-1	-1	-1	-1
[20]	0.047	0.054	0.059	0.184	0.196	0.186	0.189	0.092	0.129	0.082	0.028	0.325	0.085	0.158	0.101	0.186	0.039	0.040	0.078	-	-1	-1	-1	-1	-1	-1	-1	-1
[21]	0.060	0.070	0.089	0.138	0.153	0.148	0.150	0.086	0.094	0.065	0.022	0.242	0.078	0.139	0.131	0.127	0.056	0.006	0.032	0.026	-	-1	-1	-1	-1	-1	-1	-1
[22]	0.005	0.139	0.134	0.306	0.319	0.302	0.326	0.183	0.232	0.166	0.119	0.536	0.193	0.274	0.148	0.371	0.130	0.129	0.271	0.085	0.139	-	-1	-1	-1	-1	-1	-1
[23]	0.027	0.263	0.316	0.302	0.338	0.314	0.358	0.235	0.261	0.223	0.203	0.550	0.287	0.382	0.282	0.450	0.244	0.234	0.389	0.207	0.199	0.668	-	-1	-1	-1	-1	-1
[24]	0.003	0.107	0.108	0.267	0.279	0.260	0.283	0.155	0.199	0.135	0.084	0.490	0.149	0.230	0.135	0.310	0.098	0.091	0.206	0.061	0.099	-0.008	0.630	-	-1	-1	-1	-1
[25]	0.010	0.037	0.069	0.210	0.205	0.200	0.210	0.072	0.141	0.076	0.052	0.422	0.068	0.185	0.038	0.285	0.020	0.122	0.118	0.054	0.101	0.103	0.379	0.083	-	-1	-1	-1
[26]	0.010	0.053	0.076	0.213	0.217	0.208	0.214	0.092	0.147	0.085	0.041	0.420	0.089	0.181	0.076	0.278	0.035	0.092	0.127	0.027	0.073	0.065	0.382	0.039	0.014	-	-1	-1
[27]	0.020	0.067	0.103	0.237	0.241	0.230	0.240	0.119	0.177	0.108	0.059	0.427	0.105	0.192	0.099	0.279	0.058	0.097	0.124	0.049	0.086	0.048	0.392	0.019	0.031	-0.006	-	-1
[28]	0.029	0.028	0.059	0.196	0.193	0.187	0.197	0.071	0.139	0.072	0.035	0.371	0.060	0.164	0.053	0.242	0.014	0.100	0.096	0.035	0.076	0.083	0.310	0.055	0.001	0.004	0.020	-

[1]Hazara,Balochistan, Pakistan; [2] Afghanistan [Afghan]; [3] Afghanistan [Pathan]; [4] Beijing, China [Han];[5] China [Han]; [6]Shanghai, China [Han];[7] Xinjiang, China [Kazakh]; [8] Inner Mongolia, China [Mongolian]; [9]Liaoning, China [Mongolian]; [10] Ordos-Western Inner Mongolia, China [Mongolian]; [11] Xinjiang, China [Uighur];[12] East Kazakhstan, Kazakhstan [Kazakh]; [13] Kazakhstan [Kazakh]; [14] South Kazakhstan, Kazakhstan [Kazakh];[15] Ulaanbaatar, Mongolia [Mongolian]; [16] Swat and Dir District, Pakistan [Gujjar]; [17]Balochistan, Pakistan [Hazara]; [18] Azad Kashmir, Pakistan [Kashmiri]; [19] Swat and Dir District, Pakistan [Kohistani]; [20] Pakistan [Pathan]; [21] Punjab, Pakistan [Punjabi]; [22]Swat and Dir District, Pakistan [Tharklani, Pashtun] ;[23] Swat and Dir District, Pakistan [Uthmankheil, Pashtun]; [24] Swat and Dir District, Pakistan [Yousafzai, Pashtun]; [25] Archangelsk, Russian Federation [Russian]; [26] Lipezk, Russian Federation [Russian]; [27] Pensa, Russian Federation [Russian]; [28] Russian Federation [Russian]

Table 8:

Frequencies of null allele at DYS448 in various ethnic groups across continents

Continent	***Population***	***Number of samples***	***No of del***	***%***	***Reference***
Asia	Hazara	153	26	16.59%	Current Study
	Korean	708	6	0.85%	Myung Jin Park et al .,2005
	Kalmykia	99	7	7.07%	Roewer et al.,2007
	Japan	1079	10	0.92%	Mizuno et al.,2007
	Malaysia	980	3	0.30%	Chang et al.,2007
	Nepal	769	3	0.39%	Parkin et al.,2007
	Tajikistan	124	3	2.41%	Balaresque et al.,2008
	Kyrgyzstan	87	9	10.34	Balaresque et al.,2008
	China	130	3	2.30%	Balaresque et al.,2008
	Asian	330	2	0.61%	AmpFlSTR® Yfiler™ database
Europe	Spain	247	1	0.40%	Sanchez et al.,2007
Africa	Egypt	208	1	0.48%	Omran et al.,2008
Americas	Mexico	326	1	0.30%	Gutierrez-Alarcon et al.,2007
	African American	985	2	0.20%	AmpFlSTR® Yfiler™ database
	Caucasian (USA)	1276	2	0.16%	AmpFlSTR® Yfiler™ database
		7501	**79**	**1.05%**	

Table 9:

Sequence in the relevant flanking and repeat region of the DYS448 locus for normal and null alleles

H-53 (GenBank Accession MH200577)
GCTAAGTCT**A**CAG**G**GTTCAGA**GA**GTGGGTT**A**T**G**GTTGGC**A**ATGAGACTTCCATG GGAGAG**A**CAAG**A**ATCCAA**G**TAAAGAA**G**AGAGA**G**GTGTCAAAGAG**T**TTCA**G**TGG A**T**ATTAGA**G**A**C**AGAGAT**A**G**T**GAGA**T**AGAAAG**AT**AGATAGAGA**T**A**GA**GATA**G**AG **TT**AGATAGAGATAGAG**G**T(2bp del)**G**GAT**TA**AGATAGAGATA-**[146 bp del relative to allele 20-GAGAT(AGAGAT$_5$+N$_{42}$+AGAGAT$_8$)-AGAGAGGTAAAGATAGAGATA]-** AA**C**TTCC**T**GAC**A**G**C**CCAGAAA**G**ATG**G**GAAAATTC**CT**GTTAAGGAA**A**AAAAGGT CTCCAACA**T**AG**C**GAATCA**G**GAAATAA**T**CTCCACTCAGGAAAGAGGCTAGAAAAA AGGG**G**ACCCAGCTA**AT**GGGAG**T**CAGGGAAACA
H-57(GenBank Accession MH200578)
GCTAAGTCT**A**CAG**G**GTTCAGA**GA**GTGGGTT**A**T**G**GTTGGC**A**ATGAGACTTCCATG GGAGAG**A**CAAG**A**ATCCAA**G**TAAAGAA**G**AGAGA**G**GTGTCAAAGAG**T**TTCA**G**TGG A**T**ATTAGA**G**A**C**AGAGAT**A**G**T**GAGA**T**AGAAAG**AT**AGATAGAGA**T**A**GA**GATA**G**AG **TT**AGATAGAGATAGAG**G**T(2bpdel)**G**GAT**TA**AGATAGAGATA-**[146 bp del relative to allele 20-GAGAT(AGAGAT$_5$+N$_{42}$+AGAGAT$_8$)- AGAGAGGTAAAGATAGAGATA]-** AA**C**TTCC**T**GAC**A**G**C**CCAGAAA**G**ATG**G**G**A**AAATTC**CT**GTTAAGGAA**A**AAAAGGT CTCCAACA**T**AG**C**GAATCA**G**GAAATAA**T**CTCCACTCAGGAAAGAGGCTAGAAAAA AGGG**G**ACCCAGCTA**AT**GGGAG**T**CAGGGAAACA
H-73 (Genbank accession #MH200579)
AGG***C***CTAAGTC***A***(1bpdel)CAG***G***GTTCAGAAG(1bpdel)TGGGTTTTAGTTGGCTATGA GACTTCCAT***T***GGAGAGGCAAG***A***AT***A***CAAATAAAGAACAGAGA***G***GT***T***TCAAAGAG CTTCAATG***A***AGATTAGA***G***ATAGAGAT***AAT***GGGA***T***AGAA(4bp del)AGATAGAG(2bp del)AT (AGAGAT)$_5$AGAGA***A***(AGAGAT)$_2$A***C***AGATAGAGATA***C***AGATAGAGATAG(4bp del)ATA***A***AGATAGAGA-**[96bp del relative to allele 20 (N29+ AGAGAT$_8$)-AGAGAGGTAAAGATAGAGA]-** T***T***AA***C***TTCC***T***GAC***A***GGC***T***AGAAA***G***ATG***G***G***A***AAATT***C***C***A***TTAAGGAA***A***AAAAG(1bp del)TCTCCAACA***T***AGGGAATCA***G***GAAATAA***T***CTCCACTCAGGAAAGAG***A***CTA(1bp del)AAAAAAGGGAACCCAGCTA(1bp del)A***T***GGAG***T***CAGGGAAACA
H-74 (GenBank Accession MH200580)
[269 bp del relative to allele 20(N$_{162}$+AGAGAT$_{12}$+N$_{35}$)]-TAGAGA**C**AGAGATAG(2bp add)AGATA**A**A**A**G**A**T**A**GATAGAGATAGAGATAGAG**TT**AGA**T**ATAGA(2bp add)GAGGT**G**(4bpadd)AA**T**ATAGA**T**ATAAA**C**TTCC**T**GAC**A**G**C**CCA**A**AAA**G**ATG**G**G**A** AAATTC**CT**GTTA**TT**GA**T**GAAA(1bpdel)GGTCTCCAACA**T**AG**C**GAATC**TC**GAAATA A**T**CTCCACTCAGGAAAGAGGCTAGAAAAAAGGG**G**ACCCAGCTA(1bp del)A**T**GGA(1bp del)GCAGGGAAACA
H-76 (GenBank accession # MH200581)

GGCCTAAGTCT***A***CAG***G***GTTCAGA***GA***GTGGGTT***A***T***G***GTTGGC***A***ATGAGACTTCCAT
GGGAGAG***A***CAAG***A***ATCCAA***G***TAAAGAA***G***AGAGA***G***GTGTCAAAGAG***T***TTCA***G***TG
GA***T***AT-**[157bp del relative to allele 20(N_{47}+AGAGAT$_{12}$+N_{36})]-**
TAGAGA***C***AGAGATAG***TG***AGATAGA***A***A***G***A***T***AGATAGAGATAGAGATAGAG***TTAG***
ATAGAGATAGAG***GT***G***G***A***TT***AAGATAGAGATAAA***C***TTCC***T***GAC***A***G***C***CCAGAAA***G***A
TG***G***G***A***AAATTC***CT***GTTAAGGAA***A***AAAAGGTCTCCAACA***T***AG***C***GAATCA***G***GAAAT
AA***T***CTCCACTCAGGAAAGAGGCTAGAAAAAAGGGGACCCAGCTA***(1bp del)***ATGGAGTCAGGGAAACAGG

H-82 (GenBank accession # MH200582)

GAAGGTGGGTTTTAGTTGGCTATGAGACTTCCATGGGAGAGGCAAGGATCCAAA
TAAAGAACAGAGAAGTGTCAAAGAGCTTCAATGGAGATTAGAAATAGAGATCGC
GAGACAGAAAGGGAGATAGAGACATGGATAA(AGAGAT)$_{12}$-N$_{42}$-(AGAGAT)$_8$
AGAGAGGTAAAGATAGAGATAAATTTCCAGACCGGCCAGAAATATGAGGAAATT
CACGTTAAGGAAGAAAAGGTCTCCAACACAGGGAATCA

H-84 (GenBank Accession MH200583)

[272 bp del relative to allele 20(N_{162}+AGAGAT$_{12}$+N_{38})]-AGA**C**AGAGATAG(2bp add)AGATA**A**A**A**A**G**A**T**AGATAGAGATAGAGATAGAG**T**TAGA**T**ATAGA**(2bp add)**GAGGT**G(4bp add)**AA**T**ATAGA**T**ATAAA**C**TTCC**T**GAC**A**G**C**CCA**A**AAA**G**ATG**G**G**A**AAATTC**CT**GTT
A**TT**GA**T**GAAA(1 bp
del)GGTCTCCAACA**T**AG**C**GAATC**TC**GAAATAA**T**CTCCACTCAGGAAAGAGGCTAG
AAAAAAGGG**G**ACCCAGCTA(1 bp del)A**T**GGA-GCAGGGAAACA

H-85(GenBank Accession MH200584)

GCTAAGTCT**A**CAG**G**GTTCAGA**GA**GTGGGTT**A**T**G**GTTGGC**A**ATGAGACTTCCATG
GGAGAG**A**CAAG**A**ATCCAA**G**TAAAGAA**G**AGAGA**G**GTGTCAAAGAG**T**TTCA**G**TGG
A**T**ATTAGA**G**A**C**AGAGAT**A**G**T**GAGA**T**AGAAAG**AT**AGATAGAGA**T**A**GA**GATA**G**AG
TTAGATAGAGATAGAG**G**T(2bpdel)**G**GAT**TA**AGATAGAGATA-**[146 bp del relative to allele 20-GAGAT(AGAGAT$_5$+N$_{42}$+AGAGAT$_8$)- AGAGAGGTAAAGATAGAGATA]-**
AA**C**TTCC**T**GAC**A**G**C**CCAGAAA**G**ATG**G**G**A**AAATTC**CT**GTTAAGGAA**A**AAAAGGT
CTCCAACA**T**AG**C**GAATCA**G**GAAATAA**T**CTCCACTCAGGAAAGAGGCTAGAAAAA
AGGG**G**ACCCAGCTA**AT**GGGAG**T**CAGGGAAACA

H-86 (Genbank accession MH200585)

GGCTAAGTC**A**(1bpdel)CAG**G**GTTCAGAAG(1bpdel)TGGGTTTTAGTTGGCTATGAGA
CTTCCAT**T**GGAGAGGCAAG**A**AT**A**CAAATAAAGAACAGAGA**G**GT**T**TCAAAGAGCT
TCAATG**A**AGATTAGA**G**ATAGAGAT**AAT**G**G**GA**T**AGAA(4bpdel)AGATAGAG(2bpdel
)AT(AGAGAT)$_5$AGAGA**A**(AGAGAT)$_2$A**C**AGATAGAGATA**C**AGATAGAGATAG(4bpde
l)ATA**A**AGATAGAGA-**[96 bp del relative to allele 20 (N29+ AGAGAT$_8$)-
AGAGAGGTAAAGATAGAGA]-**
T**T**AA**C**TTCC**T**GAC**A**GGC**T**AGAAA**G**ATG**G**G**A**AAATTC**C**C**A**TTAAGGAA**A**AAAAG(
1bpdel)TCTCCAACA**T**AGGGAATCA**G**GAAATAA**T**CTCCACTCAGGAAAGAG**A**CTA(
1bpdel)AAAAAAGGGAACCCAGCTA**D**A**T**GGAG**T**CAGGGAAACAG

H-89 (GenBank Accession MH200586)

[269 bp del relative to allele 20 (N_{162}+AGAGAT$_{12}$+N_{35})]-
TAGAGA**C**AGAGATAG**T**GATA**T**AGA**A**AGA(2bp
add)GAT(AGAGAT)$_2$AGAG**T**TAGA**T**ATAGA(2bp add)GAGGT**G**(4bp add)

AA**T**ATAGAGATAAA**C**TTCC**T**GAC**A**G**C**CCA**A**AAA**G**ATG**G**G**A**AAATTC**CT**GTTAAG GAA**A**AAAAGGTCTCCAACA**T**AG**C**GAATC**TC**GAAATAA**T**CTCCACTCAGGAAAGA GGCTAGAAAAAAGGG**G**ACCCAGCTA(1bp del)A**T**GGAG**T**CAGGGAAACA
H-93 (Genbank Accession MH200587)
GGCTAA**A**GTCT**A**CAG**G**GTTCA**A**GAA**A**GTGGGTT**A**T**G**GTTGGC**A**ATGAGACTTCC ATGGGAGAG**A**CAAG**A**ATCCAA**G**TAAAGAA**G**AGAGA**G**GTGTCAAAGAG**T**TTCA**G** TGGA**T**ATTAGA**G**A**C**AGAGAT**A**G**T**GAGA**T**AGAAAG**AT**AGATAGAGA**T**A**GA**GATA **G**AG-(2bp add)-AGATAGAGATAGAG**G**T-(2bp del)-**G**GAT**TA**AGATAGAGATA-**[146 bp del relative to allele 20 (GAGAT+(AGAGAT)$_5$+N$_{42}$+(AGAGAT)$_8$AGAGAGGTAAAGATAGAGATA]-** AA**C**TTCC**T**GAC**A**G**C**CCAGAAA**G**ATG**G**G**A**AAATTC**CT**GTTAAGGAA**A**AAAAGGT CTCCAACA**T**AG**C**GAATCA**G**GAAATAA**T**CTCCACTCAGGAAAGAGGCTAGAAAAA AGGG**G**ACCCAGCTA-A**T**GGAG**T**CAGGGAAACA
H-94 (Genbank accession MH200588)
GGCTAAGTC**A**(1bpdel)CAG**G**GTTCAGAAG(1bpdel)TGGGTTTTAGTTGGCTATGAGA CTTCCAT**T**GGAGAGGCAAG**A**AT**A**CAAATAAAGAACAGAGA**G**GT**T**TCAAAGAGCT TCAATG**A**AGATTAGA**G**ATAGAGAT**AAT**GGGA**T**AGAA(4bpdel)AGATAGAG(2bpdel) AT(AGAGAT)$_5$AGAGA**A**(AGAGAT)$_2$A**C**AGATAGAGATA**C**AGATAGAGATAG(4bpdel)ATA**A**AGATAGAGA-**[96bp del relative to allele 20 (N29+ AGAGAT$_8$)-AGAGAGGTAAAGATAGAGA]-** T**T**AA**C**TTCC**T**GAC**A**GGC**T**AGAAA**G**ATG**G**G**A**AAATTC**C**C**A**TTAAGGAA**A**AAAAG(1bpdel)TCTCCAACA**T**AGGGAATCA**G**GAAATAA**T**CTCCACTCAGGAAAGAG**A**CTA(1bpdel)AAAAAAGGGAACCCAGCTA(1bpdel)A**T**GGAG**T**CAGGGAAACAG
H-95 (GenBank Accession MH200589)
GTCT**A**C**G**G**G**GTTCAGA**GA**GTGGGTT**A**T**G**G**G**TGGC**A**ATGAGACTTCCATGGGAGA G**A**CAAG**A**ATCCA**TG**TAAAGAA**G**AGAGA**G**GTGTCAAAGAG**T**TTCA**G**TGGA**T**ATT AGA**G**A**C**AGAGAT**A**G**T**GAGA**T**AGAAAG**AT**AGATAGAGA**T**A**GA**GATA**G**AG**TT**AG ATAGAGATAGAG**G**T(2bp del)GGAT**TA**AGATAGAGATA-**[145 bp del relative to allele 20 (GAGAT+(AGAGAT)$_5$+N$_{42}$+(AGAGAT)$_8$AGAGAGGTAAAGATAGAGAT]-** AA**C**TTCC**T**GAC**A**G**C**CCAGAAA**G**ATG**G**G**A**AAATTC**CT**GTTAAGGAA**A**AAAAGGT CTCCAACA**T**AG**C**GAATCA**G**GAAATAA**T**CTCCACTCAGGAAAGAGGCTAGAAAAA AGGG**G**ACCCAGCTA-A**T**GGAG**T**CAGGGAAACA
H-96 (GenBank Accession MH200590)
[269 bp del relative to allele 20(N$_{162}$+AGAGAT$_{12}$+N$_{35}$)]-TAGAGACAGAGATAG(2bp-add) AGATA**A**AA**A**G**A**TAGATAGAGATAGAGATAGAG**T**TAGA**T**ATAGA(2bp add)GAGGT**G**(4bp-add)AA**T**ATAGA**T**ATAAA**C**TTCC**T**GAC**A**G**C**CCA**A**AAA**G**ATG**G**G**A**AAATTC**CT**GTT A**TT**GA**T**GAAA(1bp-del)GGTCTCCAACA**T**AG**C**GAATC**TC**GAAATAA**T**CTCCACTCAGGAAAGAGGCTAG AAAAAAGGG**G**ACCCAGCTA-A**T**GGA(1bp-del)GCAGGGAAACA
97 (Genbank Accession MH200591)
GGCTAAGTC**A**(1bpdel)CAGGGTTCAGAAG(1bpdel)TGGGTTTTAGTTGGCTATGAGA CTTCCAT**T**GGAGAGGCAAG**A**AT**A**CAAATAAAGAACAGAGA**G**GT**T**TCAAAGAGCT TCAATG**A**AGATTAGA**G**ATAGAGAT**AAT**G**G**GA**T**AGAA(4bpdel)AGATAGAG(2bpdel)AT**A**GA**G**A**T**AGAGATAGAGATAGAGATAGAGATAGAGA**A**AGAGATAGAGATA**C**

AGATAGAGATA**C**AGATAGAGATAG(4bpdel)ATA**A**AGATAGAGA-**[92bp del relative to allele 20 (N29+ AGAGAT$_8$)- AGAGAGGTAAAGATA]-** GAGAT**T**AA**C**TTCC**T**GAC**A**GGC**T**AGAAA**G**ATG**G**G**A**AAATTC**C**C**A**TTAAGGAAAA AAAG(1bpdel)TCTCCAACA**T**AGGGAATCA**G**GAAATAA**T**CTCCACTCAGGAAAGAG **A**CTA(1bpdel)AAAAAAGGGAACCCAGCTA(1bpdel)A**T**GGAG**T**CAGGGAAACAGG**A** AG
100 (Genbank Accession MH200592)
GGCTAAGTC**A**(1bpdel)CAG**G**GTTCAGAAG(1bpdel)TGGGTTTTAGTTGGCTATGAGA CTTCCAT**T**GGAGAGGCAAG**A**AT**A**CAAATAAAGAACAGAGA**G**GT**T**TCAAAGAGCT TCAATG**A**AGATTAGA**G**ATAGAGAT**AAT**G**G**GA**T**AGAA(4bpdel)AGATAGAG(1bpdel)AT(AGAGAT)$_5$AGAGA**A**(AGAGAT)$_2$A**C**AGATAGAGATA**C**AGATAGAGATAG(4bpde l)ATAAAGATAGAGA-**[96bp del relative to allele 20 (N29+ AGAGAT$_8$)- AGAGAGGTAAAGATAGAGA]-** T**T**AACTTCC**T**GAC**A**GGC**T**AGAAA**G**ATG**G**G**A**AAATTC**C**C**A**TTAAGGAA**A**AAAAG(1bpdel)TCTCCAACA**T**AGGGAATCA**G**GAAATAA**T**CTCCACTCAGGAAAGAG**A**CTA(1bpdel)AAAAAAGGGAACCCAGCTA(1bpdel)A**T**GGAG**T**CAGGGAAACA
102 (Genbank Accession MH200593)
CTAAGTC**A**(1bpdel)CAG**G**GTTCAGAAG(1bpdel)TGGGTTTTAGTTGGCTATGAGACT TCCAT**T**GGAGAGGCAAG**A**AT**A**CAAATAAAGAACAGAGA**G**GT**T**TCAAAGAGCTTC AATG**A**AGATTAGA**G**ATAGAGAT**AAT**G**G**GA**T**AGAA(4bpdel)AGATAGAG(1bpdel)A T**A**GA**G**A**T**AGAGATAGAGATAGAGATAGAGATAGAGA**A**AGAGATAGAGATA**C**AG ATAGAGATA**C**AGATAGAGATAG(4bpdel)ATA**A**AGATAGAGA-**[96bp del relative to allele 20 (N29+ AGAGAT$_8$)- AGAGAGGTAAAGATAGAGA]-** T**T**AA**C**TTCC**T**GAC**A**GGC**T**AGAAA**G**ATG**G**G**A**AAATTC**C**C**A**TTAAGGAA**A**AAAAG(1bpdel)TCTCCAACA**T**AGGGAATCA**G**GAAATAA**T**CTCCACTCAGGAAAGAG**A**CTA(1bpdel)AAAAAAGGGAACCCAGCTA(1bpdel)A**T**GGAG**T**CAGGGAAACA
H-103(GenBank Accession MH200594)
AAGTCT**A**CAG**G**G**G**TTCAGA**GA**GTGG**TG**TT**AT**GT**GTG**TG**T**GC**A**AT**A**GAGACTTCC ATGGG**G**AGAGACAAG**A**ATCCAA**G**TAAAGAA**G**AGAGA**G**GTGTCAAAGAG**T**TTCA **G**TGGA**T**ATTAGAGA**C**AGAGAT**A**G**T**GAGA**T**AGAAAG**AT**AGATAGAGA**T**A**GA**GAT AG**A**G**TT**AGATAGAGATAGAG**G**T(2bp del)GGAT**TA**AGATAGAGATA-**[146 bp del relative to allele 20 (GAGAT+(AGAGAT)$_5$+N$_{42}$+(AGAGAT)$_8$AGAGAGGTAAAGATAGAGATA]-** AA**C**TTCC**T**GAC**A**G**C**CCAGAAA**G**ATG**G**G**A**AAATTC**CT**GTTAAGGAA**A**AAAAGGT CTCCAACA**T**AG**C**GAATCA**G**GAAATAA**T**CTCCACTCAGGAAAGAGGCTAGAAAAA AGGG**G**ACCCA**T**CTA(1bp del)A**T**GGAG**T**CAGGGAAACA
H105 (GenBank Accession MH200595)
[269 bp del relative to allele 20 (N162+AGAGAT12+N35)]- TAGAGA**C**AGAGATAG**T**GATA**T**A**A**A**A**AGA(2bp add)GATAGAGATAGAGATAGAG**T**TAGA**T**ATAGA(2bp add)GAGGT**G**(4bp add)AA**T**ATAGAGATAAA**C**TTCC**T**GAC**A**G**C**CCA**A**AAA**G**ATG**G**G**A**AAATTC**CT**GTT AAGGAA**A**AAAAGGTCTCCAACA**T**AG**C**GAATC**TC**GAAATAA**T**CTCCACTCAGGAA AGAGGCTAGAAAAAAGGG**G**ACCCAGCTA(1bp del)A**T**GGAG**T**CAGGGAAACAG
H-106 (GenBank Accession MH200596)

AGTCT**A**CAG**G**GTTCAGA**GA**GTGGGTT**A**T**G**GTTGGC**A**ATGAGACTTCCATGGGAG
AG**A**CAAG**A**ATCCAA**G**TAAAGAAGAGAGA**G**GTGTCAAAGAG**T**TTCA**G**TGGA**T**AT
TAGA**G**A**C**AGAGAT**A**G**T**GAGA**T**AGAAAG**AT**AGATAGAGA**T**A**GA**GATA**G**AG**TT**AG
ATAGAGATAGAG**G**T(2bp del)GGAT**TA**AGATAGAGATA-**[145 bp del relative to allele 20 (GAGAT+(AGAGAT)$_5$+N$_{42}$+(AGAGAT)$_8$AGAGAGGTAAAGATAGAGAT]-**
AACTTCCTGACAGCCCAGAAAGATGGGAAAATTCCTGTTAAGGAAAAAAAGGTC
TCCAACATAGCGAATCAGGAAATAATCTCCACTCAGGAAAGA
GGCTAGAAAAAAGGGGACCCAGCTA(1bp del)ATGGAGTCAGGGAAACAGACA

108 (Genbank Accession MH200597)

CAG**G**GTTCAGAAG(1bpdel)TGGGTTTTAGTTGGCTATGAGACTTCCAT**T**GGAGAGG
CAAG**A**AT**A**CAAATAAAGAACAGAGA**G**GT**T**TCAAAGAGCTTCAATG**A**AGATTAGA
GATAGAGAT**AAT**G**G**GA**T**AGAA(4bpdel)AGATAGAG(2bpdel)AT(AGAGAT)$_5$AGAGA
A(AGAGAT)$_2$A**C**AGATAGAGATA**C**AGATAGAGATAG(4bpdel)ATA**A**AGATAGAGA-
[96bp del relative to allele 20 (N29+ AGAGAT$_8$)-AGAGAGGTAAAGATAGAGA]-
T**T**AA**C**TTCC**T**GAC**A**GGC**T**AGAAA**G**ATG**G**G**A**AAATTC**C**C**A**TTAAGGAA**A**AAAAG(
1bpdel)TCTCCAACA**T**AGGGAATCA**G**GAAATAA**T**CTCCACTCAGGAAAGAG**A**CTA(
1bpdel)AAAAAAGGGAACCCAGCTA(1bpdel)A**T**GGAG**T**CAGGGAAACAG

125 (Genbank Accession MH200598)

CTAAGTC**A**(1bpdel)CAG**G**GTTCAGAAG(1bpdel)TGGGTTTTAGTTGGCTATGAGACT
TCCAT**T**GGAGAGGCAAG**A**AT**A**CAAATAAAGAACAGAGA**G**GT**T**TCAAAGAGCTTC
AATG**A**AGATTAGA**G**ATAGAGAT**AAT**G**G**GA**T**AGAA(4bpdel)AGATAGAG(2bpdel)A
T**A**GA**G**A**T**AGAGATAGAGATAGAGATAGAGA**A**AGAGATAGAGATA**C**AG
ATAGAGATA**C**AGATAGAGATAG(4bpdel)ATA**A**AGATAGAGA-**[96bp del relative to allele 20 (N29+ AGAGAT$_8$)- AGAGAGGTAAAGATAGAGA]-**
T**T**AA**C**TTCC**T**GAC**A**GGC**T**AGAAA**G**ATG**G**GAAAATTC**C**C**A**TTAAGGAA**A**AAAAG(
1bpdel)TCTCCAACA**T**AGGGAATCA**G**GAAATAA**T**CTCCACTCAGGAAAGAG**A**CTA(
1bpdel)AAAAAAGGGAACCCAGCTA(1bpdel)A**T**GGAG**T**CAGGGAAACA

H-150 (GenBank Accession MH200599)

GTCT**A**C**G**G**G**GTTCAGA**GA**GTGGGTT**A**T**G**G**G**TGGC**AGA**GA**C**ACTTCCA**G**GGGAG
AG**A**CAA**AA**ATCCA**TG**TAAAGAA**G**AGAGA**G**GTGTCAAAGAG**T**TTC**TG**TGGA**T**AT
TAGA**G**A**C**AGAGAT**A**G**T**GAGA**T**AGAAAG**AT**AGATAGAGA**T**A**GA**GATA**G**AG**TT**AG
ATAGAGATAGAG**G**T(2bp Del)GGAT**TA**AGATAGAGATA-**[146 bp del relative to allele 20 (GAGAT+(AGAGAT)$_5$+N$_{42}$+(AGAGAT)$_8$AGAGAGGTAAAGATAGAGATA]-**
AA**C**TTCC**T**GAC**A**G**C**CCAGAAA**G**ATG**G**G**A**AA**T**TTC**CT**GTTAAGGAA**AG**AAAGGT
CTCCAACA**T**AG**C**GAATCA**G**GAAATAA**T**CTCCACTCAGGAAAGAGGCTAGAAAAA
AGGG**G**ACCCAGCTA(1bp del)A**T**GGAG**T**CAGGGAAACA

151 (Genbank Accession MH200600)

GCTAAGTC**A**(1bpdel)CAG**G**GTTCAGAAG(1bpdel)TGGGTTTTAGTTGGCTATGAGAC
TTCCAT**T**GGAGAGGCAAG**A**AT**A**CAAATAAAGAACAGAGA**G**GT**T**TCAAAGAGCTT
CAATG**A**AGATTAGA**G**ATAGAGAT**AAT**G**G**GA**T**AGAA(4bpdel)AGATAGAG(2bpdel)
AT(AGAGAT)$_5$AGAGA**A**(AGAGAT)$_2$A**C**AGATAGAGATA**C**AGATAGAGATAG(4bpdel
)ATA**A**AGATAGAGA-**[96bp del relative to allele 20 (N29+ AGAGAT$_8$)-**
AGAGAGGTAAAGATAGAGA]-
T**T**AA**C**TTCC**T**GAC**A**GGC**T**AGAAA**G**ATG**G**G**A**AAATTC**C**C**A**TTAAGGAA**A**AAAAG(
1bpdel)TCTCCAACA**T**AGGGAATCA**G**GAAATAA**T**CTCCACTCAGGAAAGAG**A**CTA(
1bpdel)AAAAAAGGGAACCCAGCTA(1bpdel)A**T**GGAG**TC**CAGGGAAACAGG

H-152 (GenBank Accession MH200601)
[269 bp del relative to allele 20(N_{162}+$AGAGAT_{12}$+N_{35})]-TAGAGA**C**AGAGATAG(2bp add)AGATAGA**A**A**G**A**T**AGATAGAGATAGAGATAGAG(4bp add)ATAGAGATAGAG**GT**G**G**A(2bp add)AAGATAGAGATAAA**C**TTCC**T**GAC**A**G**C**CCA**A**AAA**G**ATG**G**G**A**AAATTC**CT**GTT AAGGAA**A**AAAAGGTCTCCAACA**T**AG**C**GAATCA**G**GAAATAA**T**CTCCACTCAGGA AAGAGGCTA**A**AAAAAAGGG**G**ACCCAGCTA-A**T**GGAG**T**CAGGGAAACA
H-153 (Genbank Accession MH200602)
CAG**G**GTTCAGAA**A**GTGGGTTTTAGTTGGCTATGAGACTTCCAT**T**GGAGAGGCAA G**A**AT**A**CAAATAAAGAACAGAGA**G**GT**T**TCAAAGAGCTTCAATG**A**AGATTAGA**G**AT AGAGAT**AAT**G**G**GA**T**AGAA(4bpdel)AGATAGAG(2bpdel)AT(AGAGAT)$_5$AGAGA**A**(A GAGAT)$_2$A**C**AGATAGAGATA**C**AGATAGAGATAG(4bpdel)ATAAAGATAGAGA-**[96bp del relative to allele 20 (N29+ $AGAGAT_8$)- AGAGAGGTAAAGATAGAGA]**-T**T**AA**C**TTCC**T**GAC**A**GGC**T**AGAAA**G**ATG**G**G**A**AAATTC**C**C**A**TTAAGGAA**A**AAAAG(1bpdel)TCTCCAACA**T**AGGGAATCA**G**GAAATAA**T**CTCCACTCAGGAAAGAG**A**CTA(1bpdel)AAAAAAGGGAACCCAGCTA(1bpdel)A**T**GGAG**T**CAGGGAAACA

REFERENCE:

[1] Qamar, R., Ayub, Q., Mohyuddin, A., Helgason, A., Mazhar, K., Mansoor, A., Zerjal, T., Tyler-Smith, C., Mehdi, S. Q., *Am. J. Hum. Genet.* 2002, *70*, 1107–1124.

[2] Siddique, A., *Afghanistan's Ethnic Divides*, 2012.

[3] Zerjal, T., Xue, Y., Bertorelle, G., Wells, R. S., Bao, W., Zhu, S., Qamar, R., Ayub, Q., Mohyuddin, A., Fu, S., Li, P., Yuldasheva, N., Ruzibakiev, R., Xu, J., Shu, Q., Du, R., Yang, H., Hurles, M. E., Robinson, E., Gerelsaikhan, T., Dashnyam, B., Mehdi, S. Q., Tyler-Smith, C., *Am. J. Hum. Genet.* 2003, *72*, 717–721.

[4] Adnan, A., Ralf, A., Rakha, A., Kousouri, N., Kayser, M., *Forensic Sci. Int. Genet.* 2016, *25*, 45–51.

[5] Adnan, A., Rakha, A., Noor, A., Oven, M. van, Ralf, A., Kayser, M., *Int. J. Legal Med.* 2017.

[6] Adnan, A., Rakha, A., Lao, O., Kayser, M., *Forensic Sci. Int. Genet.* 2018.

[7] Jobling, M. A., Samara, V., Pandya, A., Fretwell, N., Bernasconi, B., Mitchell, R. J., Gerelsaikhan, T., Dashnyam, B., Sajantila, A., Salo, P. J., Nakahori, Y., Disteche, C. M., Thangaraj, K., Singh, L., Crawford, M. H., Tyler-Smith, C., *Hum. Mol. Genet.* 1996, *5*, 1767–1775.

[8] Jobling, M. A., Lo, I. C. C., Turner, D. J., Bowden, G. R., Lee, A. C., Xue, Y., Carvalho-Silva, D., Hurles, M. E., Adams, S. M., Chang, Y. M., Kraaijenbrink, T., Henke, J., Guanti, G., McKeown, B., Oorschot, R. A. H. van, Mitchell, R. J., Knijff, P. de, Tyler-Smith, C., Parkin, E. J., *Hum. Mol. Genet.* 2007, *16*, 307–316.

[9] Repping, S., Skaletsky, H., Brown, L., Daalen, S. K. M. van, Korver, C. M., Pyntikova, T., Kuroda-Kawaguchi, T., Vries, J. W. A. de, Oates, R. D., Silber, S., Veen, F. van der, Page, D. C., Rozen, S., *Nat. Genet.* 2003, *35*, 247–251.

[10] Bosch, E., Jobling, M. A., *Hum. Mol. Genet.* 2003, *12*, 341–347.

[11] Repping, S., Daalen, S. K. M. van, Brown, L. G., Korver, C. M., Lange, J., Marszalek, J. D., Pyntikova, T., Veen, F. van der, Skaletsky, H., Page, D. C., Rozen, S., *Nat. Genet.* 2006, *38*, 463–467.

[12] Verma, R. S., Rodriguez, J., Dosik, H., *J. Hered.* 1982, *73*, 236–238.

[13] Affara, N. A., Ferguson-Smith, M. A., Tolmie, J., Kwok, K., Mitchell, M., Jamieson, D., Cooke, A., Florentin, L., *Nucleic Acids Res.* 1986, *14*, 5375–5387.

[14] Bernstein, R., Wadee, A., Rosendorff, J., Wessels, A., Jenkins, T., *Hum. Genet.* 1986, *74*, 223–229.

[15] Page, D. C., *Cold Spring Harb. Symp. Quant. Biol.* 1986, *51 Pt 1*, 229–235.

[16] Budowle, B., Aranda, X. G., Lagace, R. E., Hennessy, L. K., Planz, J. V., Rodriguez, M., Eisenberg, A. J., *Int. J. Legal Med.* 2008, *122*, 421–427.

[17] Westen, A. A., Kraaijenbrink, T., Clarisse, L., Grol, L. J. W., Willemse, P., Zuniga, S. B., Robles de Medina, E. A., Schouten, R., Gaag, K. J. van der, Weiler, N. E. C., Kal, A. J., Kayser, M., Sijen, T., Knijff, P. de, *Forensic Sci. Int. Genet.* 2015, *14*, 174–181.

[18] Balaresque, P., Bowden, G. R., Parkin, E. J., Omran, G. A., Heyer, E., Quintana-Murci, L., Roewer, L., Stoneking, M., Nasidze, I., Carvalho-Silva, D. R., Tyler-Smith, C., Knijff, P. de, Jobling, M. A., *Hum. Mutat.* 2008, *29*, 1171–1180.

[19] Kayser, M., Roewer, L., Hedman, M., Henke, L., Henke, J., Brauer, S., Krüger, C., Krawczak, M., Nagy, M., Dobosz, T., Szibor, R., Knijff, P. de, Stoneking, M., Sajantila, A., *Am. J. Hum. Genet.* 2000, *66*, 1580–1588.

[20] Larmuseau, M. H. D., Vanderheyden, N., Van Geystelen, A., Decorte, R., *Forensic Sci. Int. Genet.* 2014, *11*, 214–219.
[21] Davis, C., Ge, J., Sprecher, C., Chidambaram, A., Thompson, J., Ewing, M., Fulmer, P., Rabbach, D., Storts, D., Budowle, B., *Forensic Sci. Int. Genet.* 2013, *7*, 204–208.
[22] Consortium, T. Y. C., *Genome Res.* 2002, *12*, 339–348.
[23] Prinz, M., Boll, K., Baum, H., Shaler, B., *Forensic Sci. Int.* 1997, *85*, 209–218.
[24] Zhang, S., Tian, H., Wang, Z., Zhao, S., Hu, Z., Li, C., Ji, C., *Forensic Sci. Int. Genet.* 2014, *13*, 112–120.
[25] Nei, M., *Molecular Evolutionary Genetics*, Columbia University Press, New York, 1997.
[26] Kumar, S., Stecher, G., Tamura, K., *Mol. Biol. Evol.* 2016, *33*, 1870–1874.
[27] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria., 2015.
[28] Quintana-Murci, L., Chaix, R., Wells, R. S., Behar, D. M., Sayar, H., Scozzari, R., Rengo, C., Al-Zahery, N., Semino, O., Santachiara-Benerecetti, A. S., Coppa, A., Ayub, Q., Mohyuddin, A., Tyler-Smith, C., Qasim Mehdi, S., Torroni, A., McElreavey, K., *Am. J. Hum. Genet.* 2004, *74*, 827–845.
[29] Haber, M., Platt, D. E., Ashrafian Bonab, M., Youhanna, S. C., Soria-Hernanz, D. F., Martínez-Cruz, B., Douaihy, B., Ghassibe-Sabbagh, M., Rafatpanah, H., Ghanbari, M., Whale, J., Balanovsky, O., Wells, R. S., Comas, D., Tyler-Smith, C., Zalloua, P. A., Genographic Consortium, *PloS One* 2012, *7*, e34288.
[30] Dulik, M. C., Osipova, L. P., Schurr, T. G., *PLoS ONE* 2011, *6*, e17548.
[31] Derenko, M., Malyarchuk, B., Grzybowski, T., Denisova, G., Dambueva, I., Perkova, M., Dorzhu, C., Luzina, F., Lee, H. K., Vanecek, T., Villems, R., Zakharov, I., *Am. J. Hum. Genet.* 2007, *81*, 1025–1041.
[32] Rakha, A., Fatima, Peng, M.-S., Adan, A., Bi, R., Yasmin, M., Yao, Y.-G., *Forensic Sci. Int. Genet.* 2017, *30*, e1–e5.
[33] Chang, C.-W., Mulero, J. J., Budowle, B., Calandro, L. M., Hennessy, L. K., *J. Forensic Sci.* 2006, *51*, 344–348.
[34] Collins, F. S., Brooks, L. D., Chakravarti, A., *Genome Res.* 1998, *8*, 1229–1231.
[35] Fredman, D., *Nucleic Acids Res.* 2004, *32*, 516D – 519.
[36] NCBI Resource Coordinators, Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bollin, C., Bolton, E., Bourexis, D., Brister, J. R., Bryant, S. H., Canese, K., Cavanaugh, M., Charowhas, C., Clark, K., Dondoshansky, I., Feolo, M., Fitzpatrick, L., Funk, K., Geer, L. Y., Gorelenkov, V., Graeff, A., Hlavina, W., Holmes, B., Johnson, M., Kattman, B., Khotomlianski, V., Kimchi, A., Kimelman, M., Kimura, M., Kitts, P., Klimke, W., Kotliarov, A., Krasnov, S., Kuznetsov, A., Landrum, M. J., Landsman, D., Lathrop, S., Lee, J. M., Leubsdorf, C., Lu, Z., Madden, T. L., Marchler-Bauer, A., Malheiro, A., Meric, P., Karsch-Mizrachi, I., Mnev, A., Murphy, T., Orris, R., Ostell, J., O'Sullivan, C., Palanigobu, V., Panchenko, A. R., Phan, L., Pierov, B., Pruitt, K. D., Rodarmer, K., Sayers, E. W., Schneider, V., Schoch, C. L., Schuler, G. D., Sherry, S. T., Siyan, K., Soboleva, A., Soussov, V., Starchenko, G., Tatusova, T. A., Thibaud-Nissen, F., Todorov, K., Trawick, B. W., Vakatov, D., Ward, M., Yaschenko, E., Zasypkin, A., Zbicz, K., *Nucleic Acids Res.* 2018, *46*, D8–D13.
[37] Chang, Y. M., Perumal, R., Keat, P. Y., Kuehn, D. L. C., *Forensic Sci. Int.* 2007, *167*, 70–76.

[38] Park, M. J., Shin, K.-J., Kim, N. Y., Yang, W. I., Cho, S.-H., Lee, H. Y., *J. Forensic Sci.* 2008, *53*, 331–334.
[39] Parkin, E. J., Kraayenbrink, T., Opgenort, J. R. M. L., Driem, G. L. van, Tuladhar, N. M., Knijff, P. de, Jobling, M. A., *Forensic Sci. Int.* 2007, *166*, 176–181.
[40] Mizuno, N., Nakahara, H., Sekiguchi, K., Yoshida, K., Nakano, M., Kasai, K., *Forensic Sci. Int.* 2008, *174*, 71–76.
[41] Roewer, L., Krüger, C., Willuweit, S., Nagy, M., Rodig, H., Kokshunova, L., Rothämel, T., Kravchenko, S., Jobling, M. A., Stoneking, M., Nasidze, I., *Forensic Sci. Int.* 2007, *173*, 204–209.
[42] Sánchez, C., Barrot, C., Xifró, A., Ortega, M., Aranda, I. G. de, Huguet, E., Corbella, J., Gené, M., *Forensic Sci. Int.* 2007, *172*, 211–217.
[43] Tarlykov, P. V., Zholdybayeva, E. V., Akilzhanova, A. R., Nurkina, Z. M., Sabitov, Z. M., Rakhypbekov, T. K., Ramanculov, E. M., *Croat. Med. J.* 2013, *54*, 17–24.