

The study of novel sequence alleles for Qatari population using ForenSeq™ DNA ki

Almohammed. E ^{1*,2*}; Dragana Zgonjanin ^{3*}; Hadi. S^{1*}

¹School of Forensic and Applied Sciences, University of Central Lancashire, Preston, UK; ²Ministry of Interior of Qatar, Doha, Qatar;

³Faculty of Medicine, University of Novi Sad, Novi Sad, Serbia

*Corresponding author: eida-k-al@hotmail.com

ABSTRACT

For the last three decades, Short Tandem Repeat (STR) markers and capillary electrophoresis-based DNA sequencers have been the gold standard technology for human identification testing in the forensic field. MPS technology has enabled sequencing of several types of genetic loci in one multiplex including single nucleotide polymorphisms (SNPs) and STRs. One hundred and fifty (150) reference samples were profiled using the ForenSeq™ DNA Signature kit. PCR Primer Mix B for this kit was used containing autosomal/Y/X STRs and AISNPs, iSNPs and pSNPs loci. The analyses of the sequence of STR alleles in STRait Razor software were able to determine novel alleles in the autosomal, Y and X STR as well as SNPs loci. This increased number of alleles had a clear advantage in forensic casework. Several of these sequenced-based alleles had not been reported earlier in the literature. This study also confirmed some of the previously determined alleles in several loci. In conclusion, the results have clearly demonstrated the potential use of MPS methods to study the genetics of Qatari population.

Keywords: MPS, ForenSeq™; MiSeq FGX; GlobalFiler™; Forensic; Qatar.

1. Introduction

1.1 Massively Parallel Sequencing (MPS)

Short tandem repeat (STR) analysis using capillary electrophoresis (CE) has been the gold standard method in forensic genetics. However, the number of different analysable loci with overlapping size ranges that can be included in a single assay is limited (1). In addition, CE can only detect length polymorphisms. Advancements in Massively Parallel Sequencing (MPS) led to the development of the ForenSeq™ kit by Illumina (2). The kit enables MPS of 230 genetic markers including STRs and SNP in the same multiplex library preparation run. The ForenSeq kit covers most loci of forensic interest (27 autosomal STR, 24 Y-chromosome STR loci, 95 iSNPs, 56 aSNPs and 22 pSNPs). The ForenSeq™ Universal Analysis Software (UAS) developed by Illumina performs both primary and secondary analyses of the data. In addition, tertiary analysis was done using STRait Razor software. This is the first study for the Qatari population using MPS methods and the results would be subsequently useful for casework purposes in the forensic laboratory.

1.2 Ancestry of Qatari population

Qatar occupies the land of the Qatar Peninsula, covering a surface of 11,427 km², which extends Northward from the Eastern edge of the Arabian Peninsula into the Arabic Gulf (8), (11). Previous study by Rodriguez-Flores *et al.* (2016), described that the natives of the Arabian Peninsula can be divided into at least three genetic subpopulations that can reveal the historical migration patterns in the region: Bedouin, Persian-South Asian, and African (10), (3), (6).

2. Materials

2.1 Collection of Qatari Population Samples

A total of 150 saliva samples were collected from unrelated Qatari males. An informed consent was obtained from all volunteers providing the samples. All samples used in this study were consented for the purpose of the research and were obtained from the Ministry of Interior of Qatar. Ethical approval was granted by

Ministry of Interior of Qatar and provided by University of Central Lancashire “Ref STEM454”.

2.2 DNA Extraction

The samples were extracted using the QIAamp® extraction DNA Mini protocol (Qiagen Ltd, West Sussex, UK). The extraction procedure was followed in accordance with Manufacturer’s recommendation.

2.3 ForenSeq™ DNA Signature Prep Kit (Beta version)

The beta version of the ForenSeq™ kit provided PCR primer mixes for the targeted amplification of STRs and SNPs. The sizes of the targeted amplicons for SNPs ranged from 64 to 231 base-pairs (bp), and the sizes of the targeted amplicons for STRs range from 61 to 430 bp (4), (5).

2.3.1 Analysis Using STRait Razor Software

Short Tandem Repeat Allele Identification software (STRait Razor) was used to characterize the haplotypes of short tandem repeats (STRs) from the MPS data. STRait Razor v3.0 has been used to identify the STR allele sequences and has led to the identification of novel alleles (4), (12). Alleles were reported using nomenclature recommended by Parson *et al.* (2016). The nomenclature used herein was based on practices in the literature but used standards for variant allele naming likely to be established by the International Society of Forensic Genetics (ISFG), especially if novel variants were observed where permutations might exist for allele nomenclature (7).

3. Results

3.1 Primary and Secondary Analysis Using Universal Analysis Software (UAS)

The sample representation and the quality metrics were developed using the UAS for all the three runs, namely mean number of reads per run or mean sample representation (SP), cluster density, percentage of pass filter (PF) clusters and the percentages of phasing and pre-phasing. These were above 1407 k/mm² for all three runs. The results for the ForenSeq™ DNA Signature kit showed that on the average, over 90% of all possible loci were amplified.

3.2 ForenSeq™ DNA Signature Kit Data Using STRait Razor Software

In this study, STRait Razor v3.0 was used in the analysis to identify both the alleles and the novel alleles (4), (12). It calculates the depth of coverage by summing all effective reads within the loci. It is designed to detect forensically relevant STR alleles in FASTQ sequence data, based on allelic length.

3.2.1 Tertiary analysis using ForenSeq™ Universal Analysis Software and STRait Razor Software

This study identified several intra-allelic sequence variants at 19 of 59 STR loci (16 autosomal STRs, 2 X-STRs, and 11 Y-STRs). A number of these had not been reported earlier so these were novel results. The largest percent increases in heterozygosity due to the identification of new variants was observed for the four auSTR loci D2S441, D12S391, D22S1338 and D4S2408. UAS generated the alleles for STR loci and this was based on length and sequence both and the displays were easy to interpret. Seven of the thirteen auSTRs (D4S2408, D13S317, D21S11, D8S1179, D3S1358, D12S391 and D9S1122) showed increase as expected heterozygosity with sequence-based alleles compared with length-based alleles. A few sequence variants were also identified at the DXS10135 and DXS10103 loci. DXS10135 has four novel sequences that have not been reported in previous publication. Only twelve autosomal loci (CSF1PO, D10S1248, D17S1201, D18S51, D18S51, D19S433, D22S1045, D6S1043, PentaD, Penta E, TH01 and TPOX) and five Y-STR loci (DYS19, DYS439, DYS505, DYS549 and DYS643) did not have any effective increase in sequence based (SB) alleles. DYF387S1 has two novel SB alleles, which were 33 and 43.2 (Fig 1- Supplementary). These have not been reported before. The novel variants were compared to the human genome reference sequence (GRCh38) (Tables 1, 2, 3, 4- Supplementary). The repeat pattern variants were compared to the existing database published literature and the database STRBase (strbase.nist.gov) [Accessed Date: 30-09-2019].

3.2.2 Genetic Analysis of Qatari population (n=150) using 27 autosomal STRs

Allele frequencies for 150 samples of Qatari population were calculated using ForenSeq™ signature. All loci were highly polymorphic as expected. Tests for HWE were performed separately for Length-Based (LB) alleles for 27 auSTR loci in Qatari population. For autosomal loci, the LB (Length-Based) allele data showed departures from HWE expectations ($p < 0.05$) at four STRs out of 27 (D18S51, D19S433 and Penta D). A significant ($p < 0.05$) departure from HWE was observed and even after Bonferroni correction at the three STR loci. These significant departures from HWE may be due to chance or population substructure (8).

The statistical calculations for AuSTRs, iSNPs, and ASNPs were done using GenAIEx 6.5 and Strait Razor V3. These results showed that the Qatari population had high observed and expected heterozygosity. For X-STRs (LB), analysis of DXS10074 shows the highest number of alleles 7, 13, 16 and 20. All other X STRs had only two loci. The data showed that most frequencies were above 0.5 indicating the forensic value of the markers. AuSTR and iSNP RMPs for both length match probability and sequence match probabilities were calculated using STRait Razor software (Supplementary material). For auSTRs, these values were 2.21×10^{-31} and 1.29×10^{-35} . For iSNPs, these were 9.5×10^{-38} and 8.7×10^{-41} (Fig 2- Supplementary). The data showed that there was an increase of RMP for all auSTR markers except eight au STRs. These were D12S391, D18S51, D1S1656, D2S1338. In contrast, Penta E showed a quite low increase due to minimal increases even with the SB alleles (9). Surprisingly these loci include FGA, which is included in all commercial kits due to its high discrimination potential. Four loci D17S1301, D22S1358, D2S441, D4S2408, and D5S818 showed the greatest increased in RMP when SB alleles were included. The iSNPs RMPs were calculated and the average single-locus RMP value was between 0.471 and 0.131 arranged from 0.375 (rs10776839) to 0.909 (rs2056277) (Fig 2- Supplementary). Loci rs2056277, rs722098, rs97118 and rs938283 showed the highest increases in RMP due to the inclusion of SB alleles in the Qatari populations. Such data have not been reported earlier. The locus rs2056277 was previously reported by Poulsen *et al.* (2011) to be relatively uninformative for identity testing purposes. The iSNP full profile RMP was calculated to be 1.08×10^{-32} . STRait Razor was used to generate the combined MPs for auSTRs and iSNPs and the combined match probabilities for LB and LB+SB alleles were 2.9×10^{-68} and 1.12×10^{-75} .

4. Discussion

MPS technology is either standalone or additional technology which has the potential to overcome the limitations of CE-based STR typing. Further MPS enables the simultaneous analysis of a large number of various types of markers. The information on sequence variation allows for the discrimination of isometric alleles which are defined as two or more typed alleles with the same length. The Qatari population had not been studied before using the MPS technology and the ForenSeq™ DNA Signature kit allowing for the collection of population data for the markers which are included in the kit. Numerous isoalleles (9), (12), as well as SNP sites in the flanking regions of the repeat structures, have already been documented. These provides the capability to distinguish between two individuals with the same allele call for a locus, thus increasing the discriminatory value of the allele and also aiding the deconvolution of mixtures. This would be beneficial for identity testing in populations to increase the genetic diversity. These data serve as the first Middle Eastern population assessment using the ForenSeq™ panel and highlight the value of employing sequence-based alleles for forensic DNA typing to increase heterozygosity. Their contributions to sequence-based allele designation were included in this study. This was shown for example by increasing of haplotype diversity in LB and SB allelic data and ranged from (0.9524-0.9912) in Y-STRs. The markers containing the sequence variants that can increase diversity compared to STRs lacking sequence variation which are particularly desirable, since they should amplify faster and reduce preferential amplification. Previously published sequence data were used to check the known sequence variants per locus and alleles were classified as previously observed in either the literature or publications. These loci capture similar diversity in a smaller allele spread, minimizing the occurrence of large size differences between alleles of heterozygotes. Few sequence variants were also identified at the DXS10135 and DXS10103 loci. DXS10135 has the highest novel sequences a total number of four which have not been reported previously. The loci that carry sequence-based alleles are important for future STR population panels with STRs enhanced diversity loci. The obtained results of this study showed that AuSTRs and SNPs used in the ForenSeq™ Signature kit were very informative for forensic purposes in Qatari population. It has been demonstrated that MPS can reveal micro-heterogeneity of STR alleles (1). The possibility to detect various STRs and SNPs in a single sequencing run allowed for immense increase in discrimination power from DNA recovered from crime scenes.

5. Conclusion

This is the first study to show set of MPS data for Qatari population using the ForenSeq™ DNA Signature Prep kit. The auSTR profiles were concordant to those of Globalfiler™ kit. STRs. The study confirmed the high genotyping accuracy and resolution of MPS method in comparison to CE technology. It has been demonstrated that MPS can reveal micro-heterogeneity of STR alleles (1). The possibility to detect various STRs and SNPs in a single sequencing run has allowed for immense increase in discrimination power from DNA recovered from crime scenes. This was estimated as length combined match probability 2.9×10^{-68} and sequence combined match probability 1.12×10^{-75} for auSTR and ID SNPs combined for the Qatari population. Overall, the combined RMPs were lower by at least two orders of magnitude for every population using a SB approach. This highlighted the value of SB allele designations for statistical purposes.

6. References

1. BØRSTING, C. & MORLING, N. 2015. Next generation sequencing and its applications in forensic genetics. *Forensic Science International: Genetics*, 18, 78-89.
2. BUERMANS, H. P. J. & DEN DUNNEN, J. T. 2014. Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1842, 1932-1941.
3. FARIAS-HESSON, E., ERIKSON, J., ATKINS, A., SHEN, P., DAVIS, R. W., SCHARFE, C. & POURMAND, N. 2010. Semi-automated library preparation for high-throughput DNA sequencing platforms. *BioMed Research International*, 2010.
4. KING, J. L., WENDT, F. R., SUN, J. & BUDOWLE, B. 2017. STRait Razor v2s: Advancing sequence-based STR allele reporting and beyond to other marker systems. *Forensic Science International: Genetics*, 29, 21-28.
5. MISEQFGX™MANUAL 2015. *MiSeq FGx™ Instrument, Reference Guide*. www.illumina.com/systems/miseq-fgx.html.
6. OMBERG, L., SALIT, J., HACKETT, N., FULLER, J., MATTHEW, R., CHOUCANE, L., RODRIGUEZ-FLORES, J. L., BUSTAMANTE, C., CRYSTAL, R. G. & MEZEY, J. G. 2012. Inferring genome-wide patterns of admixture in Qataris using fifty-five ancestral populations. *BMC genetics*, 13, 1.
7. PARSON, W., BALLARD, D., BUDOWLE, B., BUTLER, J. M., GETTINGS, K. B., GILL, P., GUSMÃO, L., HARES, D. R., IRWIN, J. A., KING, J. L., KNIJFF, P. D., MORLING, N., PRINZ, M., SCHNEIDER, P. M., NESTE, C. V., WILLUWEIT, S. & PHILLIPS, C. 2016. Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. *Forensic Science International: Genetics*, 22, 54-63.
8. PÉREZ-MIRANDA, A. M., ALFONSO-SÁNCHEZ, M. A., PEÑA, J. A. & HERRERA, R. J. 2006. Qatari DNA variation at a crossroad of human migrations. *Human heredity*, 61, 67-79.
9. ROCKENBAUER, E., HANSEN, S., MIKKELSEN, M., BØRSTING, C. & MORLING, N. 2014. Characterization of mutations and sequence variants in the D21S11 locus by next generation sequencing. *Forensic Science International: Genetics*, 8, 68-72.
10. RODRIGUEZ-FLORES, J. L., FAKHRO, K., AGOSTO-PEREZ, F., RAMSTETTER, M. D., ARBIZA, L., VINCENT, T. L., ROBAY, A., MALEK, J. A., SUHRE, K. & CHOUCANE, L. 2016a. Indigenous Arabs are descendants of the earliest split from ancient Eurasian populations. *Genome research*.
11. TEXASESA&M 2017. Qatar Map.
12. WARSHAUER, D. H., KING, J. L. & BUDOWLE, B. 2015. STRait Razor v2.0: The improved STR Allele Identification Tool – Razor. *Forensic Science International: Genetics*, 14, 182-186.