Evaluation of 30 insertion/deletion polymorphisms as forensic markers in the Kuwaiti population

Mahdi Haidar[1,2], Hussain Alsaleh[1,2], Penelope R. Haddrill[*1]

[1] Centre for Forensic Science, Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow, Scotland, UK

[2] Kuwait Identification DNA Laboratory (KIDL), General Department of Criminal Evidence, Ministry of Interior, Kuwait

* Correspondence:

Mahdi Haidar, Centre for Forensic Science, Department of Pure and Applied Chemistry, University of Strathclyde, Royal College, 204 George Street, Glasgow G1 1XW.
Email: Mahdi.haidar@strath.ac.uk

Abstract:

Insertion/deletion (indel) polymorphism markers have been developed as a supplementary genotyping system in forensic DNA testing. The only commercial indel genotyping kit is the Qiagen Investigator® DIPplex kit, which includes 30 loci that can be typed using PCR-capillary electrophoresis (PCR-CE). We show that these 30 markers can be used to genetically assign Kuwaiti individuals to their population of origin when compared to other distant populations, using principal components analysis (PCA) and STRUCTURE. We also demonstrate a relationship between geographic and genetic distance using allele frequencies for the 30 markers from 26 populations.

Keywords:

Kuwait, indels, DIPplex, forensic genetics, PCA, *STRUCTURE.*

1. Introduction

Although short tandem repeat (STR) markers are used as standard for human forensic DNA testing worldwide, other genetic markers such as insertion and deletion (indel) markers address some of the limitations of STRs. These di-allelic markers have low mutation rates

and can be assayed using short amplicon sizes and traditional PCR-CE methods, improving the success rate for degraded DNA. These advantages have increased interest in using indel markers as a supplementary genotyping system in forensics. The Qiagen Investigator® DIPplex kit includes 30 indel loci and the Amelogenin sex marker. We have used DIPplex genotyping data from 150 Kuwaiti individuals to determine whether these markers can be used to assign samples to the correct population of origin. We also examine how the Kuwaiti population clusters in relation to other populations for which DIPplex data are available, to determine the utility of these markers for genetic differentiation of distant populations.

2. Materials and Methods

469 DIPplex profiles were used comprising 150 Kuwaitis (this study), 80 Danes [1], 119 Chinese (She) [2], and 120 South Koreans [3]. Thirty Kuwaiti samples were randomly selected and assigned as 'unknown'. Snipper (http://mathgene.usc.es/snipper/) was used to assign unknown individuals to their population of origin based on a Bayesian algorithm, which gives a likelihood ratio (LR) for population membership. Principal components analysis (PCA) was conducted in R software and results visualised using Microsoft Excel. STRUCTURE analysis was carried out using the following parameters: admixture model, 50,000 Markov Chain Monte Carlo steps per run, inferred K=2-6. At each K, the analysis was repeated three times, and results visualised using *CLUMPAK* [8].

To assess how genetically distant the Kuwaiti population is from other global populations, PCA was conducted based on allele frequencies obtained from published data for 26 different populations: Kuwait (N=150, this study), Iraq (N=101) [4], Turkey (N=93) [4], Lithuania (N=110) [4], Slovenia (N=157) [4], Iran (N=253) [5], Spain (N=71) [6], Poland (N=122) [7], Italy (N=200) [8], Finland (N=151) [9], Greece (N=177) [10], Portugal (N=108)

[11], Japan (*N*=251) [12], Spain (Basque Country) (*N*=60) [6], Czech Republic (*N*=55) [13], Somalia (*N*=175) [9], Taiwan (*N*=126) [7], Uruguay (*N*=131) [14], Korea (*N*=373) [3], Brazil (*N*=519) [15], Germany (*N*=80) [16], Han Chinese (*N*=565) [2], US Caucasian (*N*=262) [17], US Asian (*N*=50) [17], US Hispanic (*N*=140) [17] and US African (*N*=260) [17]. The final data set consisted of 4,707 individuals typed for the DIPplex loci. PCA results were visualised in a 3-dimensional plot using the *plot3D* package in R.

3. Results and Discussion

PCA analysis of four populations (Supplementary Figure 1) showed samples clustered according to their population of origin, with the Chinese and South Korean samples overlapping. Snipper analysis of 30 unknown (Kuwaiti) individuals correctly assigned 28 individuals to their population of origin, with LR values for 19 of them being three times greater than for other input populations, and nine individuals having LRs<3. STRUCTURE analysis indicated the most likely number of populations (*K)* was four, which was the actual number of input populations used in the analysis [18]. Figure 1 shows there was a clear signal of genetic differentiation between the populations, suggesting there is genetic structuring between populations, in agreement with the PCA analysis.
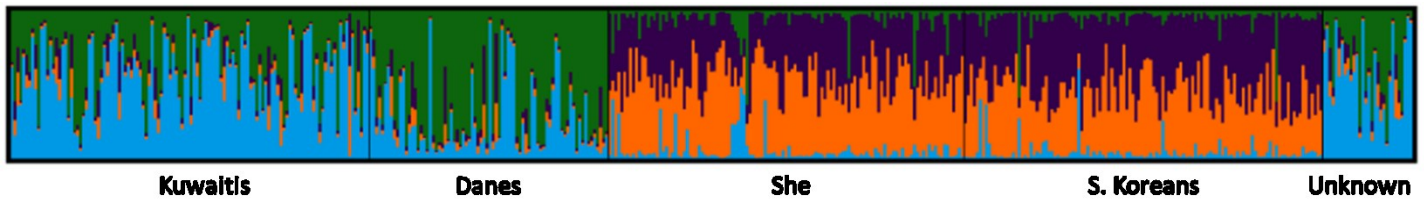
Figure 1. STRUCTURE analysis showing the most likely number of populations (*K*=4). Each column represents an individual in the data set, and the amount of each colour in each column represents the proportion of their genetic ancestry that has been assigned to each population.

To determine how genetically distant the Kuwaiti population is compared to other global populations, PCA was carried out using DIPplex marker allele frequencies from 27 different global populations [2-17], including Kuwait. This revealed five distinct clusters representing different continents, namely Asia, Europe, South America, Africa and the Middle East region (Supplementary Figure 2). These clusters indicate that geographic distance between locations restricts gene flow between populations. Thus, there is more gene flow between neighbouring countries, leading to similar allele frequency distributions among them. For example, the Kuwaiti population is more genetically similar to surrounding countries such as Iran, Iraq, and Turkey, than to other countries on different continents. Overall, it can be seen that DIPplex marker allele frequencies can be used to successfully separate geographically distant global populations on the basis of their genetic distance, making them a good choice for detecting genetic differentiation between populations.

4. Conclusion

We show that population structure can be inferred from the 30 indel markers in the DIPplex kit using Snipper as well as the Bayesian-based clustering approaches implemented in STRUCTURE software. This could be useful for identifying the ancestral origin of samples, especially in cases such as mass disasters, or where degraded DNA is encountered. Comparison of the DIPplex loci in 27 global human populations indicates that populations can be clustered geographically on the basis of their allele frequencies, and shows that the Kuwaiti population clusters with other Middle Eastern populations.

Conflict of interests

None.

References

1. Friis, S.L., et al., *Typing of 30 insertion/deletions in Danes using the first commercial indel kit—Mentype® DIPplex.* Forensic Science International: Genetics, 2012. **6**(2): p. e72-e74.
2. Wang, Z., et al., *Population genetics of 30 insertion–deletion polymorphisms in two Chinese populations using Qiagen Investigator® DIPplex kit.* Forensic Science International: Genetics, 2014. **11**: p. e12-e14.
3. Seong, K.M., et al., *Population genetics of insertion–deletion polymorphisms in South Koreans using Investigator DIPplex kit.* Forensic Science International: Genetics, 2014. **8**(1): p. 80-83.

4.   Tomas, C., et al., *Thirty autosomal insertion-deletion polymorphisms analyzed using the Investigator® DIPplex Kit in populations from Iraq, Lithuania, Slovenia, and Turkey.* Forensic Science International: Genetics, 2016. **25**(Supplement C): p. 142-144.

5.   Poulsen, L., et al., *Population and forensic data for three sets of forensic genetic markers in four ethnic groups from Iran: Persians, Lurs, Kurds and Azeris.* Forensic Sci Int Genet, 2015. **17**: p. 43-46.

6.   Martín, P., et al., *Population genetic data of 30 autosomal indels in Central Spain and the Basque Country populations.* Forensic Science International: Genetics, 2013. **7**(2): p. e27-e30.

7.   Pepinski, W., et al., *Population genetics of 30 INDELs in populations of Poland and Taiwan.* Molecular Biology Reports, 2013. **40**(7): p. 4333-4338.

8.   Turrina, S., G. Filippini, and D. De Leo, *Forensic evaluation of the Investigator DIPplex typing system.* Forensic Science International: Genetics Supplement Series, 2011. **3**(1): p. e331-e332.

9.   Neuvonen, A.M., et al., *Discrimination power of Investigator DIPplex loci in Finnish and Somali populations.* Forensic Sci Int Genet, 2012. **6**(4): p. e99-102.

10.   Tomas, C., et al., *Results for five sets of forensic genetic markers studied in a Greek population sample.* Forensic Science International: Genetics, 2015. **16**(Supplement C): p. 132-137.

11.   da Silva, C.V., et al., *Genetic portrait of south Portugal population with InDel markers.* Forensic Science International: Genetics, 2013. **7**(4): p. e101-e103.

12.   Nunotani, M., et al., *Analysis of 30 insertion-deletion polymorphisms in the Japanese population using the Investigator DIPplex(R) kit.* Leg Med (Tokyo), 2015. **17**(6): p. 467-70.

13.   Zidkova, A., et al., *Application of the new insertion–deletion polymorphism kit for forensic identification and parentage testing on the Czech population.* International Journal of Legal Medicine, 2013. **127**(1): p. 7-10.

14.   Saiz, M., et al., *Allelic frequencies and statistical data from 30 INDEL loci in Uruguayan population.* Forensic Science International: Genetics, 2014. **9**: p. e27-e29.

15.   Ferreira Palha, T.d.J.B., et al., *Population genetic analysis of insertion–deletion polymorphisms in a Brazilian population using the Investigator DIPplex kit.* Forensic Science International: Genetics, 2015. **19**: p. 10-14.

16.   *Qiagen supplementary material: Population data for analysis of results from the Investigator DIPplex kit. Available at* http://www.qiagen.com/literature/render.aspx?id=200253.

17.   Fondevila, M., et al., *Forensic performance of two insertion-deletion marker assays.* Int J Legal Med, 2012. **126**(5): p. 725-37.

18.   Evanno, G., S. Regnaut, and J. Goudet, *Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study.* Mol Ecol, 2005. **14**(8): p. 2611-20.