Forensic Science International: Genetics Supplement Series 7 (2019) 189–190

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/fsigss

Population genetics of 30 insertion-deletion polymorphism in polish Populations

Eliza Michalak[a,*], Monica Abreu-Głowacka[a], Magdalena Konarzewska[b], Witold Pepiński[c], Małgorzata Skawrońska[c], Monika Wójcik[c], Ireneusz Sołtyszewski[d], Czesław Żaba[a]

[a] Department of Forensic Medicine, Poznan University of Medical Sciences, Poznan, Poland
[b] Department of Medical Genetics, Warsaw Medical University, Warsaw, Poland
[c] Department of Forensic Medicine, Medical University of Bialystok, Bialystok, Poland
[d] Department of Criminalistics and Forensic Medicine, University of Warmia and Mazury in Olsztyn, Poland

ARTICLE INFO

Keywords:
InDel polymorphism
Investigator DIPplex
Poland
Population genetics
Forensic efficiency

ABSTRACT

The Investigator DIPplex® kit (Qiagen) contains components for the simultaneous amplification and analysis of 30 biallelic autosomal InDels and amelogenin. The objective of this study was to estimate the diversity of the 30 markers in a Polish population sample (N = 389) and to evaluate their usefulness in forensic genetics. The DIPplex genotype distributions showed no significant deviation from Hardy-Weinberg rule expectations (Bonferroni corrected). The mean observed heterozygosity value is 0.4611, and the combined Matching Probability value is $1.08x10^{-13}$. The investigated marker set has been confirmed as a potential extension to standard STR - based kits or a separate informative system for individual identification and kinship analysis.

1. Introduction

InDels (insertion-deletion) or DIPs (deletion-insertion polymorphisms) are short length diallelic polymorphisms, consisting of the presence or absence of short sequences (typically 1–50 bp). They are relatively common throughout the human genome representing 15–20% of all polymorphisms [1] with the total number estimated at about 2 million [2]. InDels combine the advantages of both STR and SNP markers in forensic genetics applications, including short amplicon size (50–150 bp) and low mutation rate ($< 2 \times 10^{-8}$). They can be conveniently typed using a single multiplexed PCR (30–40 markers) with fluorescently labeled primers followed by capillary electrophoresis (a current technology for human identification) [3–5].

The Investigator DIPplex® kit (Qiagen) contains components for the simultaneous amplification and analysis of 30 biallelic autosomal InDels and amelogenin. The InDels are distributed over 19 autosomes at the minimum distance of 10 Mbp to routinely used STR and SNP markers. The allele length variations of the kit markers are between 4 and 22 bp, and all amplicons are shorter than 160 bp.

Populations and DNA extraction: Anonymized buccal swabs were anonymized and collected from 389 unrelated volunteers inhabiting western (N = 168), central (N = 113) and northeastern (N = 108) regions of Poland along with information on the birthplace and ethnicity of the donor. Signed informed consents were obtained from all the participants and this study complied with the protocol approved by the Ethical Committee of Poznan University of Medical Sciences (Ref: 139/13). The extraction of genomic DNA was carried out using QIAamp® DNA Mini Kit (Qiagen). The quantitation was performed using Quantifiler™ Human DNA Quantification Kit on a 7500 Real-Time PCR System (Applied Biosystems) according to the manufacturer's specifications.

Amplification and genotyping: PCR conditions were applied according to the protocol recommended by the manufacturer of the Investigator DIPplex Kit (Qiagen) in PCR System 9700 (Applied Biosystems, USA). The amplification was performed with 30 PCR cycles. Electrophoresis and typing were performed in 3130 Genetic Analyzer (Applied Biosystems, USA) using BTO 550 (Qiagen) was used as the internal lane standard spanning fragments from 60 to 550 bps. The data were collected using Data Collection v3.0 software. GeneMapper® ID-X v1.1.1 software was used for the InDels classification.

Statistical analysis: AMOVA and population differentiation exact test were calculated with the Arlequin v.3.5 software [6]. For subsequent analyses frequencies from the three respective population samples were pooled. Estimates for genetic diversity (allele frequencies, heterozygosity), conformance to expectations of the Hardy-Weinberg

* Corresponding author.
E-mail address: emichalak@ump.edu.pl (E. Michalak).

https://doi.org/10.1016/j.fsigss.2019.09.074
Received 23 September 2019; Accepted 24 September 2019
Available online 14 October 2019

equilibrium (HWE) and for independence (Linkage Disequilibrium, LD) were obtained using GDA v1.0 software [7]. For multiple comparisons, the original significance levels achieved (*P*-values) were transformed by the Bonferroni correction procedure [8], i.e. 30 markers per database yield an actual significance level of 0.0016667. Forensic informativeness was estimated by calculating discrimination power (DP), match probability (MP), polymorphic information content (PIC), typical paternity index (TPI), and power of paternity exclusion (PE) using Powerstats v1.2 spreadsheet (Promega) [9]. Comparison of allele frequency distributions was performed by means of a pairwise population comparison test (RxC contingency test; G. Carmody, Ottawa, Canada).

2. Results and discussion

Wright's F_{ST} was analysed to measure population substructure effects [10]. AMOVA results revealed that most of the molecular variation was due to variation within the three regional population samples (99.88%) rather than among them, with average fixation index value of 0.0012. Our findings correspond to those presented by other authors who used AMOVA to compare the allelic frequencies for each DIPplex locus in populations of Europe, Africa, Asia and North America [11,12]. Moreover, in our analysis individual InDels displayed noticeable disparities in fixation index spanning from -0.0034 (HLD118) to 0.0177 (HLD67). The individual mutation rate of a locus is one of the factors that may explain the observed discrepancy [13]. However, when compared with mutation rates of 10^{-3} to 10^{-5} for STRs [14,15], SNPs have essentially mutation rates estimated at as low as 10^{-8} [16]. From the point of view of forensic genetics, markers with high heterozygosity and very low F_{ST} are potentially advantageous due to relatively high discrimination efficiency irrespective of population of origin [16,17]. High heterozygosity enhances the polymorphism information at each SNP and low F_{ST} diminishes the chance of interpopulation effects. Some SNPs are reported to have remarkably little variation in allele frequency around the world [18]. On the other hand, ancestry informative single-nucleotide polymorphisms (AISNPs) are required to show low heterozygosity and high allele frequency divergence between different ancestral or geographically distant populations (F_{ST} values). These genetic markers are especially useful in establishing the high probability of an individual's biogeographical ancestry [19,20]. Previously we have selected eight InDels (HLD131, HLD111, HLD118, HLD99, HLD122, HLD64, HLD81, HLD39) with F_{ST} higher than 0.1 between Poles and Taiwanese as potential AISNPs for further analyses [21]. It is noteworthy that the same loci significantly accounted for diversity between Caucasian and Asian samples, based on North American datasets published elsewhere [12].

In the Polish population sample the InDels frequency distributions showed no deviations from HWE (Bonferroni corrected, $0.0019 < P < 1.0000$) evaluated by randomization procedure (10,000 cycles). Pairwise comparison using the exact test disequilibrium analysis yielded departures from independence for 26 out of 435 pairs of InDels under the analysis ($0.0000 < P < 0.0012$) (data not shown). Observed heterozygosity for all the systems ranged from 0.4139 (HLD56) to 0.5425 (HLD122), with an average of 0.4611, which corresponds to the values reported for Czech [22], German [23], Danish [24], Finnish [25], Central Spain, and the Basque Country populations [11], and being higher than those reported for Asians, Asian-Americans, and African-Americans [23,26,27]. Based on data of the 30 InDels the combined MP value among Poles is 1.09×10^{-13} and the combined PE value is 0.9938. Both parameters indicate a favourable value of a random match comparable with that of 10 microsatellite markers (AmpFlSTR SGM kit) [28]. A pairwise testing for heterogeneity using the χ^2-test was applied to compare allelic distributions. Minor or no significant differences were found between the Polish sample and Czech [22], Danish [24], Finnish [25], and American-Caucasian [12] data sets.

References

[1] J.L. Weber, D. David, J. Heil, Y. Fan, C. Zhao, G. Marth, Human diallelic insertion/deletion polymorphisms, Am. J. Hum. Genet. 71 (2002) 854–862.
[2] J.M. Mullaney, R.E. Mills, W.S. Pittard, S.E. Devine, Small insertions and deletions (INDELs) in human genomes, Hum. Mol. Genet. 19 (2010) R131–R136.
[3] C. Li, S. Zhao, S. Zhang, L. Li, et al., Genetic polymorphism of 29 highly informative InDel markers for forensic use in the Chinese han population, Forensic Sci. Int. Genet. 5 (2011) e27-e30.
[4] J.R. Pimenta, S.D.J. Pena, Efficient human paternity testing with a panel of 40 short insertion–deletion polymorphisms, Genet. Mol. Res. 9 (2010) 601–607.
[5] M.W. Nachman, S.L. Crowell, Estimate of the mutation rate per nucleotide in humans, Genetics 156 (2000) 297–304.
[6] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resour. 10 (2010) 564–567.
[7] P.O. Lewis, D. Zaykin, Genetic Data Analysis: Computer Program for the Analysis of Allelic Data. Version 1.0 (d16c), Free program distributed by the authors over the internet from: (2011) http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php.
[8] B.S. Weir, Multiple tests, Genetic Data Analysis vol. 109, Sunderland: Sinauer Associates Inc, 1990, pp. 109–110.
[9] A. Tereba, Tools for Analysis of Population Statistics. PowerStats Program Version 1.2. Promega Corporation, (1999).
[10] B. Weir, C.C. Cockerham, Estimating F-statistics for the analysis of population structure, Evolution 38 (1984) 1358–1370.
[11] P. Martín, H.B. García, I. Yurrebaso, A. Aguirre, A. Alonso, Population genetic data of 30 autosomal indels in Central Spain and the Basque Country populations, Forensic Sci. Int. Genet. 7 (2013) e27–30, https://doi.org/10.1016/j.fsigen.2012.10.003.
[12] B.L. LaRue, J. Ge, J.L. King, B. Budowle, A validation study of the Qiagen Investigator DIPplex® kit; an INDEL-based assay for human identification, Int. J. Legal Med. 126 (2012) 533–540.
[13] P. Wolanska-Nowak, On the subtle southern Polish population subdivision, Problems of Forensic Sciences 71 (2007) 265–279.
[14] Q.Y. Huang, F.H. Xu, H. Shen, et al., Mutation patterns at dinucleotide microsatellite loci in humans, Am. J. Hum. Genet. 70 (2002) 625–634.
[15] B.M. Dupuy, M. Stenersen, T. Egeland, B. Olaisen, Y-chromosomal microsatellite mutation rates: differences in mutation rate between and within loci, Hum. Mutat. 23 (2004) 117–124.
[16] A.J. Pakstis, W.C. Speed, J.R. Kidd, K.K. Kidd, Candidate SNPs for a universal individual identification panel, Hum. Genet. 121 (2007) 305–317.
[17] B. Budowle, A. van Daal, Forensically relevant SNP classes, BioTechniques 44 (2008) 603–610.
[18] K.K. Kidd, A.J. Pakstis, W.C. Speed, et al., Developing a SNP panel for forensic identification of individuals, Forensic Sci. Int. 164 (2006) 20–32.
[19] T. Frudakis, K. Venkateswarlu, M.J. Thomas, Z. Gaskin, S. Ginjupalli, S. Dunturi, et al., A classifier for the SNP-based inference of ancestry, J. Forensic Sci. 48 (2003) 771–782.
[20] C. Phillips, A. Salas, J.J. Sanchez, et al., Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs, Forensic Sci. Int. Genet. 1 (2007) 273–280.
[21] W. Pepinski, M. Abreu-Glowacka, M. Koralewska-Kordel, et al., Population genetics of 30 INDELs in populations of Poland and Taiwan, Mol. Biol. Rep. 40 (2013), https://doi.org/10.1007/s11033-013-2521-7 4333-38.
[22] A. Zidkova, A. Horinek, V. Kebrdlova, M. Korabecna, Application of the new insertion-deletion polymorphism kit for forensic identification and parentage testing on the Czech population, Int. J. Leg. Med. 127 (2013) 7–10.
[23] Qiagen supplementary material: population data for analysis of results from the Investigator DIPplex kit (2010) Qiagen.
[24] S.L. Friis, C. Børsting, E. Rockenbauer, et al., Typing of 30 insertion/deletions in Danes using the first commercial indel kit - Mentype® DIPplex, Forensic Sci. Int. Genet. 6 (2012) e72-e74.
[25] A.M. Neuvonen, J.U. Palo, M. Hedman, A. Sajantila, Discrimination power of Investigator DIPplex loci in Finnish and Somali populations, Forensic Sci. Int. Genet. 6 (2012) e99-e102.
[26] H. Zhang, G. He, J. Guo, et al., Genetic diversity, structure and forensic characteristics of Hmong–Mien-speaking Miao revealed by autosomal insertion/deletion markers, Mol. Genet. Genomics (2019), https://doi.org/10.1007/s00438-019-01591-7.
[27] K.M. Seong, J.H. Park, S.H. Young, P.W. Kang, et al., Population genetics of insertion-deletion polymorphisms in South Koreans using investigator DIPplex kit, Forensic Sci. Int. Genet. 8 (2014) 80–83.
[28] A. Tucholska-Lenart, J. Wujec, J. Samborski, E. Jakubowska, Allele frequencies for 10 STR loci in a population from central Poland, Forensic Sci. Int. 129 (2002) 131–133.