



Research paper

The redesigned Forensic Research/Reference on Genetics-knowledge base, FROG-kb

Kenneth K. Kidd^{a,*}, Usha Soundararajan^a, Haseena Rajeevan^{a,b}, Andrew J. Pakstis^a, Katherine N. Moore^c, Jeri D. Roper-Miller^c

^a Department of Genetics, Yale University School of Medicine, New Haven, CT 06520, USA

^b Center for Medical Informatics, Yale University School of Medicine, New Haven, CT 06520, USA

^c Center for Forensic Sciences, RTI International, Research Triangle Park, NC 27709, USA

ARTICLE INFO

Keywords:

Forensic resource

Web site

Knowledge base

SNPs

Ancestry inference

Individual identification

ABSTRACT

The Forensic Resource/Reference on Genetics-knowledge base (FROG-kb) web site <<https://frog.med.yale.edu/FrogKB/>> was introduced in 2011 and in the five years since the previous publication ongoing research into how the database can better serve forensics has resulted in extensive redesign of the database interface and functionality. Originally designed as a prototype to support forensic use of single nucleotide polymorphisms (SNPs), FROG-kb provides a freely accessible web interface that facilitates forensic practice and can be useful for teaching and research. Based on knowledge gained through its use, the web interface has been redesigned for easier navigation through the multiple components. The site also has functional enhancements, extensive new documentation, and new reference panels of SNPs with new curated data. FROG-kb focuses on single nucleotide polymorphisms (SNPs) and provides reference population data for several published panels of individual identification SNPs (IISNPs) and several published panels of ancestry inference SNPs (AISNPs). For each of the various marker panels with reference population data, FROG-kb calculates random match probabilities (RMP) and relative likelihoods of ancestry for a user-entered genotype profile (either completely or partially specified). Example genotype profiles are available and the User's Manual presents interpretation guidelines for the calculations. The extensive documentation along with ongoing updates makes FROG-kb a comprehensive tool in facilitating use of SNPs in forensic practice and education. An overview of the new FROG-kb with examples and material explaining the results of its use are presented here.

1. Introduction

In 2011 we introduced FROG-kb (Forensic Resource/Reference on Genetics-knowledge base) (<https://frog.med.yale.edu/FrogKB/>), an open access web tool, as a reference database of population allele frequencies for Single Nucleotide Polymorphisms (SNPs) likely to be used in forensics. Thus, the focus has been on di-allelic markers as distinct from the standard multiallelic short tandem repeat (STR) polymorphisms (STRPs) traditionally used in forensic sciences. FROG-kb allows viewing and retrieval of forensically relevant data as well as calculation of statistics on several forensically relevant published sets of SNPs and one panel of Insertion-Deletion polymorphisms (InDels) [1]. Since the introduction of FROG-kb, SNPs have gained in importance in forensic sciences. Consequently, FROG-kb has considerably changed from the original description [1] involving new functionalities and expanded data. These results of ongoing research into database and interface design as well as the newly incorporated population genetic data

warrant this description of the current version of FROG-kb.

As background, we note that the ability of DNA genotyping to be of use in forensic sciences is completely dependent on the existence of reference data. A random match probability (RMP) is calculated using the frequencies of a subject's alleles in a population; for which population the RMP is calculated is an issue that can be of relevance and will be case dependent. The need for web-based tools and databases to predict population affiliations by allowing calculation of random match probabilities in forensic cases is well recognized. Many databases exist for standard sets of Short Tandem Repeat (STR) Polymorphisms (STRPs), e.g., STRBase (<http://www.cstl.nist.gov/strbase/>) [2], the European Network of Forensic Science Institute's (ENFSI) DNA working group database STRidER (STRs for identity ENFSI Reference database, <http://strider.online/>) [3,4], and PopAffiliator (<http://cracs.fc.up.pt/popaffiliator/>). The same requirement exists for database(s) with reference allele frequencies for di-allelic markers of forensic interest, SNPs and InDels. In many ways, multiple reference populations are

* Corresponding author at: Department of Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520-8005, USA.
E-mail address: kenneth.kidd@yale.edu (K.K. Kidd).

<https://doi.org/10.1016/j.fsigen.2017.11.009>

Received 16 June 2017; Received in revised form 13 October 2017; Accepted 13 November 2017

Available online 14 November 2017

1872-4973/© 2017 The Authors. Published by Elsevier Ireland Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

more important for SNPs than for the standard forensic STR markers because the very high mutation rates and global heterozygosity of STRPs result in relatively low levels of global differentiation [5] whereas SNPs can have the maximum difference of alternative alleles fixed in different populations.

The enhancements to the database and the redesign of the interface to FROG-kb have involved many that are individually small, but helpful and/or important in standardization. Some of them are mentioned here, but for those planning to actually use FROG-kb, more detail is given in the supplemental material and in the online User's Manual. We have also included here material to help in both the understanding of potential uses of FROG-kb and the interpretation of results of the calculations made possible through the FROG-kb web site.

2. Basic redesign and update of FROG-kb

The original purpose of FROG-kb was to be a prototype that, from a forensic perspective, could serve as a tool facilitating use of SNPs in forensic practice and for teaching and research. FROG-kb focuses on individual identification SNPs (IISNPs) and ancestry inference SNPs (AISNPs). For those two types of markers the interface allows the user to query the reference data for many different panels of SNPs for a multisite genotype of an individual. The web site returns the probability of that genotype in each of the reference populations and the likelihood ratio of the most probable population compared to each alternative specified population, all based on the data in the underlying database. Through the connections into ALFRED, the ALlele FREquency Database (<https://alfred.med.yale.edu/>), the “knowledge base” component of FROG-kb provides details on the population frequency data and the molecular definitions of the polymorphisms. This paper focuses on the web site of FROG-kb and what functionalities are available; the original paper provides a description of the underlying database structure and bioinformatics aspects [1].

3. Basic navigation of FROG-kb

A more intuitive user-friendly interface has been designed. All pages have a series of buttons across the top for top-level navigation through the web site (Fig. 1). The Home Page text gives a brief summary of the functions available in FROG-kb.

Sub-menus exist, specific to each button. For example, selecting the **Documentation** button on the top-level navigation opens a sub-menu (Fig. 2) with options for more detailed information. This is recommended as a first step for all new users because it leads to the User's manual. The **Manual** button opens the downloadable comprehensive user manual with text and graphical elements designed to make FROG-kb navigation easier for the user. The manual provides navigation pointers to our graphical user interface. It is the ultimate resource for information on the database/web interface and we welcome input on improving it. The manual explains the buttons for all of the various sub-

menus.

4. Reference panels available

In Fig. 1, there are two buttons for SNP panels (**IISNP**, **AISNP**) that open pages with the respective panels of each type available for use (Table 1). For each of the IISNP and AISNP panels there are (1) a list of the specific SNPs available for calculations with url links to ALFRED and dbSNP, (2) a list of the reference populations with links to ALFRED and (3) a link for data entry. Explanatory information, examples, and the reference allele frequencies used are also available for each of the panels.

For each panel, the number of reference populations currently available is given. Relevant references for the panels are contained within the database. The specific reference populations for each panel are listed as part of the information for each panel with links to their definitions in ALFRED. Data for many of the various reference populations are the result of data collection on the populations studied in the Kidd Lab.

The phenotype inference (PISNP) button currently links only to the 6-SNP Irisplex [6] calculation; the eye color prediction in FROG-kb uses the formula from the publication. We note that this formula may not be accurate in all parts of the world [7].

Besides these SNP sets in FROG-kb, additional IISNP and AISNP panels are also available from the ALFRED SNP Sets page under the **Search** tab on the ALFRED homepage. These have not, so far, had sufficient population data to have priority for entry into FROG-kb.

5. New panels addressing the ‘empty matrix’ issue

One of the user requests following the initial release of FROG-kb was for a way to calculate statistics using SNPs from more than one published panel. Meeting that request has been difficult because of the empty matrix problem: different SNP panels have been studied on different populations [8]. The likelihood comparisons that are fundamental to forensic ancestry inference require that all SNPs have allele frequency data for all reference populations. We have added two AISNP panels that partially address the empty matrix issue: “Overlap set of AISNPs” (Overlap set) and the “Combined panel of 192 AISNPs” (Combined panel). The “Overlap set” is comprised of 44 SNPs of the 46 SNPs that occur in three or more of 21 different published AI panels involving 1397 markers in total [8]. Two of the SNPs have data for only a few of the populations; 44 of the 46 SNPs have complete data for 72 reference populations. Unfortunately, these 44 SNPs do not allow biogeographic resolution by STRUCTURE [9] beyond about five “quasi-continental” regions. The “Combined panel” is comprised of subsets of the KiddLab-55 [10,11], SeldinLab-128 [12,13] and SNPforID 34-plex [14] AISNPs. 79 reference populations have data for 192 of the union of 200 SNPs. With this integrated combined panel, any user-defined subset of the 192 SNPs can be used to calculate the likelihoods of a sample



Welcome to the New FROG-kb

FROG-kb seeks to make allele frequency data for panels of SNPs more useful in a forensic setting. The primary objective of FROG-kb is to provide a web interface that, from a forensic perspective, is useful for teaching and research and can serve as a tool facilitating forensic practice. The underlying data are housed in the already extensively used and referenced ALlele FREquency Database (ALFRED).

Fig. 1. Screen shot of the main navigation buttons.



Fig. 2. Screen shot showing the buttons for options available under Documentation.

Table 1

The specific SNP panels available for IISNPs (1a) and AISNPs (1b). For each panel the number of reference populations currently available is given. Relevant references for the panels are contained within the database. The specific reference populations for each panel are listed as part of the information for each panel with links to their definitions in ALFRED.

(1a) Summary of IISNP panels in FROG-kb. IISNP panels and number of Populations Included for likelihood calculations	
KiddLab - 45 Unlinked IISNP	45
KiddLab - List of 86 IISNPs	45
SNPforID 52-plex	20
Qiagen Investigator DIPplex kit	28
(1b) Summary of AISNP panels in FROG-kb. AISNP panels and number of Populations Included for likelihood calculations	
Seldin's list of 128 AISNPs	90
SNPforID 34-plex	53
KiddLab – Set of 55 AISNPs	139
Kayser's set of 24 AI Markers	73
Daniele Podini's list of 32 AISNPs	111
Eurasiaplex 23 SNP Panel	76
Nievergelt's Set of 41AIMs	123
Li's panel of 74AIMs	78
Overlap set of AISNPs	72
Combined panel of 192 SNPs	79
Pacifiplex	34
EUROFORGEN Global ancestry-informative SNP panel	40

originating from any of the 79 populations.

6. Ongoing data curation and entry

The usefulness of any database is dependent on the quality of its contents. New published panels and additional reference population data for existing SNP panels are systematically added to FROG-kb. Data are added from scanning of published literature, from Kidd Lab data collection, from collaborators, and through data submissions by researchers. SNPs included in FROG-kb are made consistent to represent alleles on the forward strand.

The reference frequency data entered for each of these panels has supporting population information. Web site links exist to pages in ALFRED for more details and allele frequency data tables for specific populations. The comprehensive set of reference populations available for most of these panels includes the 26 population samples from the Phase III 1000 Genomes consortium (1000 Genomes Project Consortium) [15,16]. When a new population has data for all SNPs in a panel, it becomes a reference population sample included in the computations in the FROG-kb interface. Data often exist on additional populations for individual SNPs and are accessible through ALFRED.

7. Data availability

Because it is important to document the exact allele frequency estimates used in the likelihood calculations, tables of the values used for the various panels are available for download through **Frequencies Download** in the **Documentation** sub-menu and the sub-menu available when each panel is selected.

KiddLab - 45 Unlinked IISNPs

Population likelihoods based on 45 SNPs and 45 reference populations for : Korean Individual

Print Table Format View SNPs Used

● Indicates the values are within an order of magnitude of the highest likelihood.

Population (Region, Sample Size 2N)	Probability of Genotype in each Population	Likelihood Ratio
Koreans (EastAsia,132)	● 1.224E-18	
Finns (Europe,72)	● 8.83E-19	1.39
Japanese (EastAsia,112)	● 3.943E-19	3.1
Hungarians (Europe,184)	● 3.204E-19	3.82
Irish (Europe,232)	● 2.729E-19	4.49
Mixed Europeans (Europe,190)	● 2.506E-19	4.88
Danes (Europe,102)	● 1.594E-19	7.68
Adygei (Europe,108)	● 1.516E-19	8.07
Hakka (EastAsia,86)	1.169E-19	10.5
Lao Long (Asia,238)	6.975E-20	17.5
Ashkenazi Jews (Europe,166)	6.677E-20	18.3

Fig. 3. Screenshot of the result page showing calculation result displayed as three columns for an IISNP dataset of a Korean individual.

8. Interpreting the results from FROG-kb

8.1. The statistical results

For each SNP panel FROG-kb calculates the probability of the user-entered multi-locus genotype in each of the reference populations. If a SNP is not included in the data entered, it is not used in the calculation. The results of the calculation are displayed as a table with three columns: each line contains 1) the name of the reference population sample with its geographic region and sample size, 2) the probability of the entered genotype occurring in that population, and 3) the likelihood ratio of the most probable population to the specific reference population (Fig. 3).

The populations are ordered by their probabilities of generating the entered genotype from highest to lowest. Note that the example in Fig. 3 uses an IISNP panel with loci selected for little variation in allele frequencies. Thus, very distant populations have very similar probabilities of generating the specific multi-locus genotype found in this Korean individual.

8.2. Random match probability

The probability of the entered genotype is equivalent to a random match probability (RMP) assuming no deviation from Hardy-Weinberg ratios in the population. The results for one of the IISNP panels (Fig. 3) provides an indication of how rare the genotype is globally. The largest value, the one listed at the top, provides an upper bound for the RMP among the populations tested. In this example the relative likelihoods for the populations in Fig. 3 do not provide useful ancestry information because the SNPs in the IISNP panels were generally chosen to have similar allele frequency values around the world. The results for an AISNP panel can also be interpreted simply as an indication of the upper bound for the RMP among the reference populations.

8.3. Inference of ancestry

In the case of ancestry inference each probability can also represent the likelihood that the specific population is the origin of the entered genotype. Fig. 4 is an example. In ancestry inference the absolute value has no meaning; only the relative likelihoods are meaningful. The population with the highest probability is the most likely ancestral population among the set of reference populations. Dividing the highest likelihood by those for the other populations yields likelihood ratios representing how many times more likely the entered genotype is in the most likely population compared to occurring in the specific population. These range from 1 to progressively larger numbers for the less likely populations of origin. More detailed information on *Results of the Calculations* can be found in Section 5.2 of the User's Manual.

Multiple factors affect how one interprets the results from FROG-kb for one of the AISNP panels. Many are discussed in [17]. One of the first is that none of the panels contains reference populations truly representative of the human species. The inference of ancestry for an unknown DNA sample (individual) can only be as good as the global coverage of the reference population samples. If the true population of origin is not among the reference populations, the results cannot identify it. In Fig. 4, even a separate sample of Japanese is a less good fit to the JPT individual than a sample of Han Chinese. Were there no Japanese reference samples, Chinese and Vietnamese would be the most likely ancestries. If the unknown comes from one of the closely related reference populations, any distinction is questionable *a priori* because the true population of origin may not be the most likely or significantly different from the most likely. Moreover, those issues can be different for different sets of SNPs.

Using the likelihood framework makes it clear that the “most likely” may not be meaningfully different from other highly likely populations. A very relevant point is that there is a finite probability of the unknown

genotype arising in almost every population in the world. Thus, the “most likely” is simply that, the most likely, and others are less likely to extremely unlikely. If the likelihood ratio among the more likely populations is within a factor of 10 of the most likely, there is no meaningful basis for distinguishing among those potential ancestral populations. Even a ratio of up to 100 includes populations that cannot be meaningfully excluded from possibly being ancestral for the specific genotype.

The fact that the “most likely” ancestral population cannot be interpreted as the true ancestral population may be easier to understand when one considers the fact that the SNPs being used are polymorphic and hence different individuals in the same population will have different genotypes. An example for two individuals from Kerala in India is elaborated in [17]. For one individual, the likelihoods favor populations from South India. For the other individual, a Pakistani population is the most likely and it is not possible to exclude other more northern groups in India.

9. Examples and exercises in ancestry inference using FROG-kb

Part of the objective of FROG-kb as a tool for forensic sciences is facilitating the understanding of the results of the calculations. Details of the FROG-kb calculations and more text pertaining to the interpretation of the results are available in the online User Manual. To further help with the understanding of the FROG-kb likelihood results we have included in Supplemental Materials specific examples to illustrate important aspects such as the dependence of results on the specific SNPs used. We have also included in the supplemental material some exercises to help with the inference of ancestry using FROG-kb

10. Conclusion

FROG-kb is a unique web site offering access to reference data for many published SNP panels that have forensic relevance and the ability to calculate relevant statistics for an unknown forensic sample when it has genotypes for the SNPs in one of those panels. The web site has undergone significant redesign with enhancements in functionality and user friendliness since the original version was put online in 2011. In addition to the major reorganization of the interface, new published SNP panels have been added with their reference population data. The redesign of the interface to FROG-kb has also involved many individually small changes that are helpful and/or important in standardization. Some of the more important changes are briefly mentioned above with additional information given in supplemental data. An online User's Manual has been developed and updated to be more useful for those planning to actually use FROG-kb. The text in the User's Manual and example data provided for each panel are designed to help forensic scientists understand the results of the calculations. As part of ongoing curation of the database efforts will be made to increase the reference panels and populations and to enhance the educational value of FROG-kb.

Conflicts of interest

None.

Acknowledgments

The underlying database and the FROG-kb and ALFRED web interfaces are supported by grant 2016-DN-BX-0162 to K.K. Kidd by the U.S. National Institute of Justice. Web site redesign was partially supported by the Forensic Technology Center of Excellence (2011-DN-BX-K564) awarded by the U.S. National Institute of Justice, Office of Investigative Sciences. The opinions, findings, and conclusions or recommendations expressed in this publication/program/exhibition are those of the author(s) and do not necessarily reflect those of the United

Combined panel of Kiddlab-55, Seldin's-128, and SNPforID34-plex AISNPs (192 SNPs)

Population likelihoods based on 192 SNPs and 79 reference populations for : JPT example

[Print Table Format](#) [View SNPs Used](#)

● Indicates the values are within an order of magnitude of the highest likelihood.

Population (Region, Sample Size 2N)	Probability of Genotype in each Population	Likelihood Ratio
Japanese(JPT) (EastAsia,208)	● 1.378E-59	
Han Chinese(CHS) (EastAsia,210)	7.263E-62	190.0
Japanese (EastAsia,112)	3.627E-62	380.0
Vietnamese(KHV) (EastAsia,202)	1.867E-62	738.0
San Francisco Chinese (EastAsia,124)	6.297E-63	2190.0
Taiwanese Han (EastAsia,100)	4.169E-63	3310.0
Han Chinese(CHB) (EastAsia,206)	9.405E-64	14700.0
Cambodian (EastAsia,52)	5.373E-64	25600.0
Koreans (EastAsia,132)	4.803E-64	28700.0
Lao Long (Asia,238)	1.611E-64	85500.0
Hakka (EastAsia,86)	1.607E-64	85700.0
Chinese Dai(CDX) (EastAsia,186)	1.57E-65	878000.0
Armi (EastAsia,80)	3.054E-67	4.51E7
Malaysians (EastAsia,24)	3.687E-71	3.74E11
Yakut (Siberia,102)	2.235E-72	6.17E12
Micronesians (Oceania,78)	1.261E-74	1.09E15
Samoans (Oceania,22)	1.044E-76	1.32E17
Kachari (Asia,36)	7.647E-81	1.8E21
Bengali(BEB) (Asia,172)	2.218E-81	6.209999999999999E21
Atavai (EastAsia,84)	2.18E-82	6.32E22
Khamti (Asia,100)	1.754E-84	7.86E24
Keralite (Asia,60)	3.532E-86	3.9E26
Sri Lankan Tamil(STU) (Asia,204)	1.524E-86	9.04E26
Telugu(TU) (Asia,204)	2.624E-87	5.25E27
Thoti (Asia,28)	2.927E-88	4.71E28
Peruvian(PEL) (SouthAmerica,170)	2.779E-88	4.96E28
Gujarati(GIH) (Asia,206)	4.493E-89	3.07E29
Punjabi(PJL) (Asia,192)	1.289E-89	1.07E30
Maya Yucatec (NorthAmerica,106)	2.424E-96	5.68E36

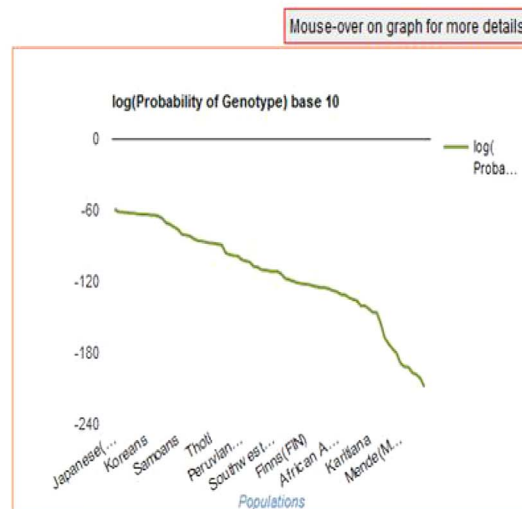


Fig. 4. An example of ancestry inference for a JPT individual using the Combined Panel of 192 SNPs. This screen shot also included the graph of the log likelihoods showing the full range of nearly 100 orders of magnitude.

States Department of Justice.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.fsigen.2017.11.009>.

References

[1] H. Rajeevan, U. Soundararajan, A.J. Pakstis, K.K. Kidd, Introducing the Forensic Research/Reference on Genetics knowledge base, FROG-kb, *Investig. Genet.* 3 (1) (2012) 18, <http://dx.doi.org/10.1186/2041-2223-3-18>.

[2] C.M. Ruitberg, D.J. Reeder, J.M. Butler, STRBase: a short tandem repeat DNA database for the human identity testing community, *Nucleic Acids Res.* 29 (1) (2001) 320–322.

[3] P. Gill, L. Foreman, J.S. Buckleton, C.M. Triggs, H. Allen, A comparison of adjustment methods to test the robustness of an STR DNA database comprised of 24 European populations, *Forensic Sci. Int.* 131 (2–3) (2003) 184–196.

[4] L.A. Welch, P. Gill, C. Phillips, R. Ansell, N. Morling, W. Parson, et al., European network of forensic science institutes (ENFSI): evaluation of new commercial STR multiplexes that include the European standard set (ESS) of markers, *Forensic Sci. Int. Genet.* 6 (6) (2012) 819–826, <http://dx.doi.org/10.1016/j.fsigen.2012.03.005>.

[5] B.F. Algee-Hewitt, M.D. Edge, J. Kim, J.Z. Li, N.A. Rosenberg, Individual identifiability predicts population identifiability in forensic microsatellite markers, *Curr. Biol.* 26 (7) (2016) 935–942, <http://dx.doi.org/10.1016/j.cub.2016.01.065>.

[6] S. Walsh, F. Liu, K.N. Ballantyne, M. van Oven, O. Lao, M. Kayser, IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information, *Forensic Sci. Int. Genet.* 5 (3) (2011) 170–180, <http://dx.doi.org/10.1016/j.fsigen.2010.02.004>.

[7] F. Liu, S. Walsh, M. Kayser, Of sex and IrisPlex eye colour prediction: a reply to Martinez-Cadenas, et al, *Forensic Sci. Int. Genet.* 9 (2014) e5–6, <http://dx.doi.org/10.1016/j.fsigen.2013.06.006>.

[8] U. Soundararajan, L. Yun, M. Shi, K.K. Kidd, Minimal SNP overlap among multiple panels of ancestry informative markers argues for more international collaboration, *Forensic Sci. Int. Genet.* 23 (2016) 25–32, <http://dx.doi.org/10.1016/j.fsigen.2016.01.013>.

[9] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, *Genetics* 155 (2) (2000) 945–959.

[10] K.K. Kidd, W.C. Speed, A.J. Pakstis, M.R. Furtado, R. Fang, A. Madbouly, et al., Progress toward an efficient panel of SNPs for ancestry inference, *Forensic Sci. Int. Genet.* 10 (2014) 23–32, <http://dx.doi.org/10.1016/j.fsigen.2014.01.002>.

[11] A.J. Pakstis, L. Kang, L. Liu, Z. Zhang, T. Ji, E.L. Grigorenko, et al., Increasing the reference populations for the 55 AISNP panel: the need and benefits, *Int. J. Legal Med.* 131 (2017) 913–917, <http://dx.doi.org/10.1007/s00414-016-1524-z>.

[12] R. Kosoy, R. Nassir, C. Tian, P.A. White, L.M. Butler, G. Silva, et al., Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America, *Hum. Mutat.* 30 (1) (2009) 69–78, <http://dx.doi.org/10.1002/humu.20822>.

[13] J.R. Kidd, F.R. Friedlaender, W.C. Speed, A.J. Pakstis, F.M. De La Vega, K.K. Kidd, Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples, *Investig. Genet.* 2 (1) (2011) 1, <http://dx.doi.org/10.1186/2041-2223-2-1>.

[14] C. Phillips, A. Salas, J.J. Sanchez, M. Fondevila, A. Gomez-Tato, J. Alvarez-Dios, et al., Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs, *Forensic Sci. Int. Genet.* 1 (3–4) (2007) 273–280, <http://dx.doi.org/10.1016/j.fsigen.2007.06.008>.

[15] 1000 Genomes Consortium, G.R. Abecasis, A. Auton, L.D. Brooks, M.A. DePristo, R.M. Durbin, R.E. Handsaker, et al., An integrated map of genetic variation from 1,092 human genomes, *Nature* 491 (7422) (2012) 56–65, <http://dx.doi.org/10.1038/nature11632>.

[16] P.H. Sudmant, T. Rausch, E.J. Gardner, R.E. Handsaker, A. Abyzov, J. Huddleston, et al., An integrated map of structural variation in 2,504 human genomes, *Nature* 526 (7571) (2015) 75–81, <http://dx.doi.org/10.1038/nature15394>.

[17] K.K. Kidd, Thoughts on estimating ancestry. Chapter 7, in: Antonio Amorim, Bruce Budowle (Eds.), *Handbook of Forensic Genetics: Biodiversity and Heredity in Civil and Criminal Investigation*, 2016.