Forensic Science International: Genetics 23 (2016) 153–158

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Evaluating a subset of ancestry informative SNPs for discriminating among Southwest Asian and circum-Mediterranean populations

Ozlem Bulbul[a,b,*], Lotfi Cherni[c,d], Houssein Khodjet-el-khil[c], Haseena Rajeevan[a,e], Kenneth K. Kidd[a]

[a] Department of Genetics, Yale University School of Medicine, New Haven, CT 06520 USA
[b] Institute of Forensic Science, Istanbul University, Istanbul, 34098, Turkey
[c] Laboratory of Genetics, Immunology and Human Pathology, Science Faculty of Tunis, University Tunis El Manar, 2092 Tunis, Tunisia
[d] High Institute of Biotechnology, University of Monastir, 5000 Monastir, Tunisia
[e] Center for Medical Informatics, Yale University School of Medicine, New Haven, CT 06520 USA

ARTICLE INFO

Article history:
Received 28 August 2015
Received in revised form 22 April 2016
Accepted 26 April 2016
Available online 27 April 2016

Keywords:
AIMs
55 AISNP set
Southwest Asia
Mediterranean

ABSTRACT

Many different published sets of single nucleotide polymorphisms (SNPs) and/or insertion-deletion polymorphisms (InDels) can serve as ancestry informative markers (AIMs) to distinguish among continental regions of the world. For a focus on Southwest Asian ancestry we chose to start with the Kidd Lab panel of 55 ancestry-informative SNPs (AISNPs) because it already provided good global reference data (FROG-kb: frog.med.yale.edu) in a set of 73 population samples distinguishing at least 8 biogeographic clusters of populations. This panel serves as a good first tier ancestry panel. We are now interested in identifying region-specific second tier panels for more refined distinction among populations within each of the global regions. We have begun studying the global region centered on Southwest Asia and the region encompassing the Mediterranean Sea. We have incorporated 10 populations from North Africa, Turkey and Iran and included 31 of the original 73 populations and eleven 1000 Genomes Phase3 populations for a total of 3129 individuals from 52 populations, all typed for the 55 AISNPs. We have then identified the subset of the 55 AISNPs that are most informative for this region of the world using Heatmap, Fst, and Informativeness analyses to eliminate those SNPs essentially redundant or providing no information among populations in this region, reducing the number of SNPs to 32. STRUCTURE and PCA analyses show the remaining 32 SNPs identify the North African cluster and appropriately include the Turkish and Iranian samples with the Southwest Asian cluster. These markers provide the basis for building an improved, optimized panel of AISNPs that provides additional information on differences among populations in this part of the world. The data have also allowed an examination of the accuracy of the ancestry inference based on 32 SNPs for the newly studied populations from this region. The likelihood ratio approach to ancestry inference embodied in FROG-kb provides highly significant population assignments within one order of magnitude for each individual in the Turkish, Iranian, and Tunisian populations.

1. Introduction

Ancestry informative markers (AIMs) have been used to estimate ancestry of an unknown individual as an aid to forensic investigations when there is no suspect even after searches of databases. Similarly, accurate determination of the ancestry of unidentified human remains can be a great aid in the eventual identification of the person [1,2]. Many different panels of ancestry informative single nucleotide polymorphisms (AISNPs) have been published to distinguish among biogeographic origins [3–10]. Most studies have tested a limited number of populations or populations from a single region of the world. Additionally, practical issues such as methodologies for multiplexing SNP assays and financial considerations generally limit the number of SNPs in a forensic panel. In forensic applications accurate inference of individual ancestry requires availability and use of relevant reference population data [7]. There are two forensic web-based ancestry estimation calculators for individual ancestry inference: SNIPPER

* Corresponding author at: Department of Genetics, SHM I-348, Yale University School of Medicine, P.O. Box 208005, 333 Cedar Street, New Haven, CT 06520-8005 USA.
E-mail address: ozlembulbul00@gmail.com (O. Bulbul).

http://dx.doi.org/10.1016/j.fsigen.2016.04.010

(http://mathgene.usc.es/snipper/) and Forensic Resource/Reference on Genetics–knowledge base (FROG-kb) (http://frog.med.yale.edu) [11]. The SNIPPER website holds a limited number of specific panels but it allows custom sets of reference population data to be uploaded by the user. These can be genotypes from the user's laboratory or collected online as relevant to any AIM panel of interest. On the other hand the FROG-kb website has eight specific AI SNP panels already available with existing sets of 53–125 reference populations, depending on the panel.

Our area of focus in this study, Southwestern Asia and coastal Mediterranean, is not cleanly divided by geographical borders. Historical events, including population migrations and inland and sea trade, have caused some gene flow across this area [12,13]. Our favored approach to ancestry inference is a two-tier approach with an initial screening for the best worldwide geographic resolution followed by a region-specific second-tier panel to refine ancestry inference among populations within the geographic region [14]. Thus, our ultimate goal is a second-tier panel of SNPs that will distinguish among the populations in Southwestern Asia and around the Mediterranean Sea. The 55 AISNP forensic panel now distinguishes at least 8 biogeographic clusters among 125 population samples from around the world [15] including distinction between Southwest Asian and European populations. Our initial aim is to identify the best subset of those 55 SNPs for inferring ancestry in this predefined region. This initial step has identified a smaller panel of markers useful for ancestry inference in this region of the world while still providing useful distinctions elsewhere in the world. This subset is a starting point for a robust second tier panel for this region. We have also investigated the statistical accuracy of these 32 SNPs for ancestry inference within the region and find that reasonable accuracy exists even among closely related populations.

2. Material and methods

We genotyped the 55 SNPs on 492 new individuals using TaqMan® assays (Life Technologies, California, USA) following the protocols used in previous studies [7]. The ten new populations involved were two general population samples, one from Turkey and one from Iran, and eight specific Tunisian/Libyan population samples (Nebeur, Kesra, Kairoun, Sousse, Mehdia, Kerkennah, Smar, and Libyans). We included 31 population samples previously reported [7] and 11 populations from 1000 Genomes Phase3 (http://www.1000genomes.org/) all with full data for all 55 AISNPs. This final set of 52 population samples comprises the dataset analyzed; geographically these populations are located from South Asia, through Southwest Asia, North and East Africa, and Europe into Northwest Siberia. Supplemental Table S1 provides the name, source of the data and sample size for each of them. Overall, we analyzed 3129 individuals from these 52 populations. Moreover, to determine the usefulness of the 32 AISNPs on a global scale we combined data on the 52 populations in the current study with data on an additional 34 world populations from the study by Pakstis et al. [15]. A total of 4972 individuals were included in the global dataset for STRUCTURE analysis.

Allele frequencies of the 55 AISNPs were estimated by simple gene counting for each of the 52 populations, assuming full co-dominance of each SNP. We used Fst, Rosenberg's Informativeness [16], and heatmap analyses to evaluate each of the 55 AISNPs across the populations. Fst was calculated across all 52 populations for the allele frequencies using the simple formula of Wright [17]. Heatmaps of the populations by allele frequencies were generated using R v. 3.1.2.

Principal components analysis (PCA) for population sample allele frequencies was calculated using XLSTAT 2015 (http://www.

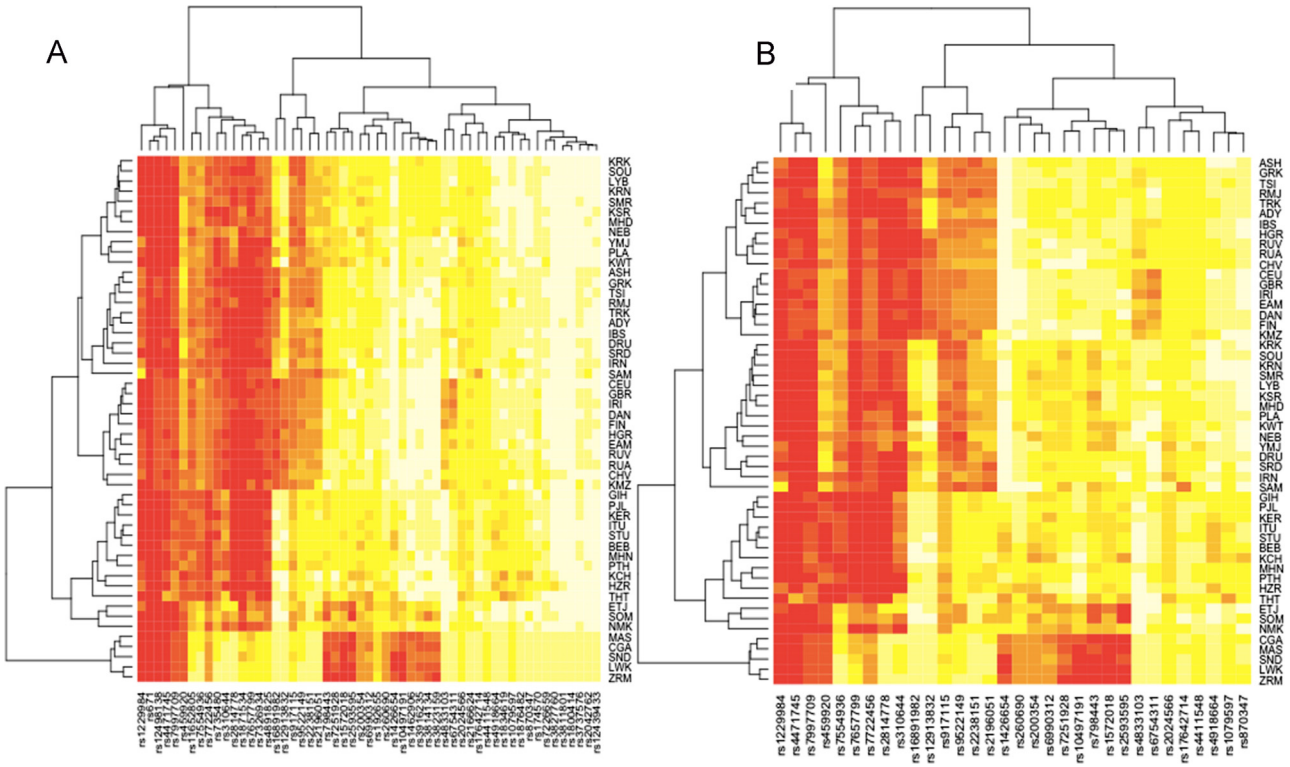

Fig. 1. (A) The heatmap of the full set of 55 AISNPs. (B) The heatmap of the subset of 32 selected AISNPs after removing relatively uninformative and highly correlated SNPs. The highest branches of the population dendogram (left side of the graphic) show that five population blocks (Mediterranean, Europe, Middle East, Central South Asia and Africa, respectively) occur among these 32 AISNPs. Heatmaps are based on allele frequencies for the same 52 populations in both figures.

xlstat.com/en/about-us/company.html); PCA of individuals' genotypes was performed using R v. 3.1.2 and packages SNPassoc [18]. We also used STRUCTURE v. 2.3.4 (http://pritchardlab.stanford.edu/structure.html) to evaluate and illustrate the population and individual ancestral proportions [1]. The STRUCTURE parameters were: 10,000 burnins and 10,000 MCMC; admixture model; correlated allele frequencies; 20 independent replicates per cluster from K = 3 to K = 7. Results were displayed for the analysis with the highest value of -ln P(D) for each K value. Graphics were constructed using CLUMPP v1.1.2 and Distruct v1.1 software [1,19].

Individual ancestry assignment (by relative likelihoods) was inferred based on 32 AISNPs for each individual from Turkey, Iran, and the eight populations from Tunisia/Libya (total 492 samples) using FROG-kb [11], which currently includes 125 reference populations for the 55 AISNP panel, including the 10 new populations presented here [15]. The underlying data (the population frequencies of each SNP) used in the FROG-kb calculations are taken from the referenced ALlele FREquency Database, ALFRED. From those frequencies FROG-kb calculates relative likelihoods of ancestry from different populations for user-specific genotypes using data extracted from ALFRED [11]. To test how well the individuals in the ten new populations were assigned to their own populations, we modified the frequencies used in the on-line FROG-kb calculations by removing the tested individual from their population and re-estimating the population's allele frequencies. Thus, for each individual of these ten populations we computed the likelihood of it coming from the known population of origin by first calculating an unbiased allele frequency estimate for that population leaving out the target individual. Frequencies for other populations were unchanged.

3. Results

3.1. Selection of the markers

The allele frequencies for the 55 AISNPs for 52 population samples are publicly available in ALFRED (http://alfred.med.yale.edu) and the genotype frequencies used in the FROG-kb calculations can be downloaded for each of the AISNP panels (http://frog.med.yale.edu/FrogKB/freqDownload.jsp). The 52-population Fst results for the 55 AISNPs are listed in Supplemental Table S2, ranked by Fst. Rosenberg's I_n is also listed and the correlation between the two is 0.978. rs2814778 within the DARC gene (aka Duffy Blood Group) has the highest Fst value (Fst = 0.682), reflecting the near fixation of alternate alleles between sub-Saharan Africans and the rest of the world. Other high Fst SNPs also distinguish sub-Saharan African populations from European populations. The well-known pigmentation SNPs associated with European ancestry–rs16891982 in SLC45A2 (Fst = 0.582), rs1426654 in SLC24A5 (Fst = 0.563), and rs12913832 in HERC2 (Fst = 0.441)–have among the highest Fst values. The two lowest Fst values are East Asian specific SNPs: rs671 in ALDH2 (Fst = 0.024) and rs1800414 in OCA2 (Fst = 0.021). SNP rs12498138 in GOLGB1 (Fst = 0.036) and rs2042762 (Fst = 0.036) are also among the SNPs with the lowest Fst values. These findings are expected because these SNPs primarily distinguish East Asian and/or Native American populations from the rest of the world. These and other SNPs among the 55 are simply not very informative for the more western populations in this analysis. However, STRUCTURE and heatmap analyses of these 52 more western populations using all 55 SNPs both show that five clusters can be distinguished.

The heatmap of the 55 AISNPs based on the allele frequencies of all populations is shown in Fig. 1A. Shorter "sister" branches of the SNP dendogram indicate SNPs with more highly correlated allele frequencies across the 52 populations.

Similarly, shorter pairs of branches in the population dendogram indicate genetically closer populations. The branches highest in the dendogram of the populations show five different population clusters (North Africa-Middle East, Mediterranean, Europe, Central South Asia and Sub-Saharan Africa, respectively). Based on this result, the SNPs that showed essentially no variation among the 52 populations were clustered at the right and all were then eliminated. Among the remaining SNPs those pairs with the shortest branches were identified. Since such pairs are the most

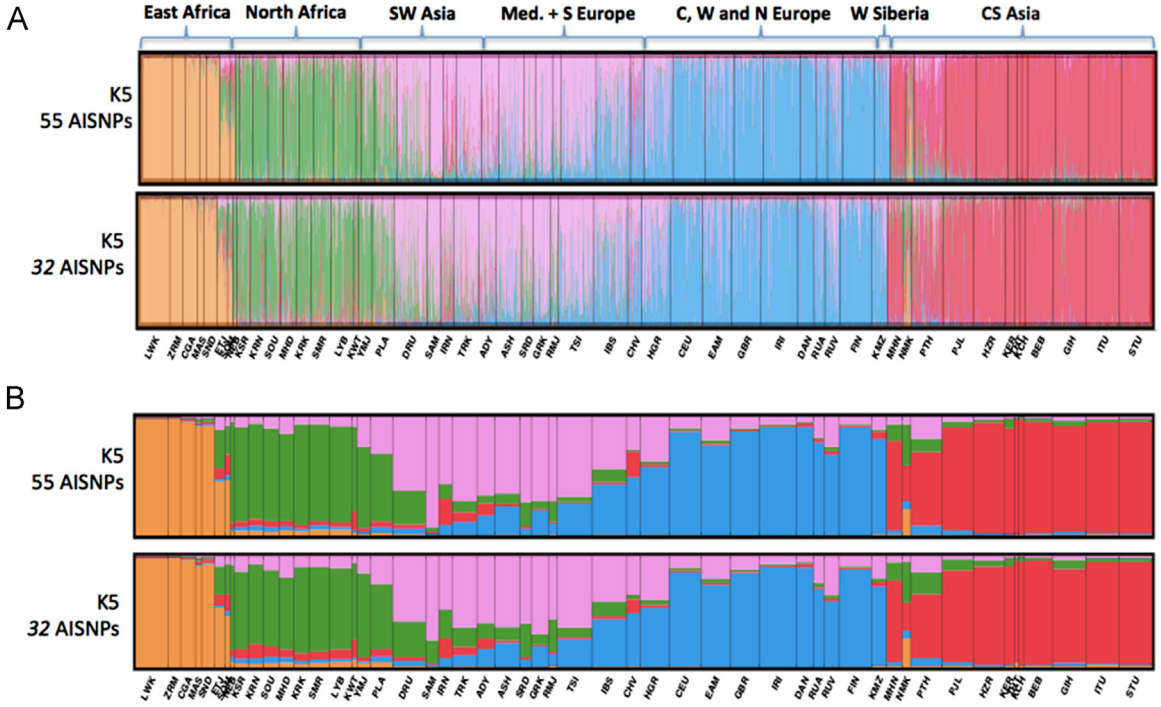

Fig. 2. The most likely of the 20 STRUCTURE analyses at optimum K = 5 for the 55 and 32 AIMs datasets for individuals (A) and populations (B). Population average clusters are very similar showing five biogeographic clusters: Africa, Middle East, Mediterranean, Europe and Central South Asia regions.

correlated, one of each redundant pair was removed, leaving only those SNPs that show higher variation among western populations.

The resulting 32 SNPs provide differentiation of several population groups similar to that obtained with 55 SNPs (Fig. 1B). However, different SNPs contribute differently to population distinctions and slight differences in clustering of populations occur as a result of the SNPs used in the calculations. Thus, Druze, Sardinians, Iranians, and, Samaritans as a branch and Hungarians, Russians, and Chuvash as a second branch clustered slightly differently in the second heatmap.

3.2. Population clustering analysis

STRUCTURE results with the highest likelihood at each K value for 55 and 32 AI sets, K = 3–7, are shown in Supplemental Fig.S1 and S2 . Five distinct clusters [(1) East Africa, (2) North Africa to Middle East, (3) SW Asia to Mediterranean and Southern Europe, (4) Central, Western and Northern Europe and (5) Central South Asia, respectively] are obtained from both AIM sets at K = 5. The likelihoods of the 32 AI set began to plateau by the K = 5 (Supplemental Fig.S3). The pattern of the individual runs (a total of 20 runs for each K) is quite consistent at K = 5 (Supplemental Fig. S2). Although the 55 AISNP set shows very similar results at higher numbers of clusters, the consistency of the patterns and highest likelihoods of the replicate runs are starting to scatter (Supplemental Fig.S1 and S3). Thus, the patterns and the likelihoods of the individual runs signify a statistically reasonable stopping point at K = 5 for both sets. We compared the highest likelihood runs at K = 5 for the 55 and 32 AI sets to inspect individual (upper plots) and population (lower plots) assignment level of differences (Fig. 2 K = 5). Although many individuals are not identically allocated to clusters, on a population level the results for the two datasets appear nearly identical. The notable, but slight, differences between 32 and 55 AISNP results are that in the 32 SNP results (Fig. 2) the Middle Eastern pattern (green color) increased more than 10% (10–22%) in the Samaritan, Roman Jewish, Sardinian and Iranian population samples whereas the Mediterranean pattern (pink color) decreased overall.

The STRUCTURE analyses of the global dataset were run for K = 3–9 and the best solutions (K = 3–8) are illustrated in Supplemental Fig.S4. The 32 AISNPs set provides clear clusters of the broad regions like Africa, North Africa to Middle East, SW Asia, Europe, Central South Asia, East Asia, Pacific and Americas for K = 8 (Supplemental Fig. S4). The results at K = 7, Africa to Central Asia, are similar to the results in Fig. 2 where the five clusters appear. Although these 32 AISNPs successfully differentiate eight major regions of the world, the smaller panel of 32 SNPs provides less information compared to the whole set of 55 [7,15].

We performed a Principal Component Analysis (PCA) based on the individual data (genotypes for each individual) as well as population data (allele frequencies for each population) for the 32 and 55 AISNPs (Supplemental Fig.S5). Overall, PC1 and PC2 mainly reflect the geographic distribution of the individuals or populations, with similar results with both sets of SNPs. The PCA of individuals highlights the genetic variation among individuals while the population PCA shows the clear separation of the populations. The first principal component (PC1) reflects major genetic variation between African and non-African individuals or populations. Northern African populations are separated from Western European populations. PC2 reflects genetic variation from Europe through Central South Asia and separates Central South Asia from other populations. Mediterranean individuals occupy the space between Europe, Southwest Asia, and Central South Asia.

Overall, heatmap and Fst analyses led to a good selection of SNPs for distinguishing among the biogeographic groups of populations of this study. The 55 AISNP panel is an efficient and globally useful panel of ancestry informative markers with a reference database of 125 populations [15]. Obviously, ancestry specific markers that distinguish East Asian and Native American populations from others are unlikely to provide much differentiation among Western Asian and European populations. Similarly, redundant SNPs do not add extra information on the differentiation of the populations. Therefore, we eliminated a total of 23 SNPs from the 55 AI panel without losing any significant ancestry information for this part of the world. Supplemental Table S1 presents the 32 AISNP subset of the 55 AISNP panel.

3.3. Individual ancestry assignment analysis

These analyses show the similarities in results of the subset of 32 SNPs to those of the full 55 SNPs with respect to clusters of populations. However, they also raise two concerns about the use of the reduced panel of 32 AISNPs to determine ancestry: the robustness to missing data and the accuracy of the smaller panel. The issue of missing data is always a concern but it is difficult to address because it is completely dependent on both what marker (s) are missing data and what population is the origin of the sample. However, we can begin to address the basic forensic question: How well would these reference data assign a new individual to its correct population of origin? This question is especially interesting for populations that are part of the same cluster of populations. One approach to evaluating ancestry assignment is to use likelihood ratios for populations based on the probability of the query genotype occurring in each population. A revised standalone java application was used to access directly the underlying data of FROG-kb and run the ancestry likelihood function to assemble the matrix of likelihoods of the 125 potential ancestral populations for each of the 492 individuals in the ten new populations. The data were then summarized in two different ways. First, the summaries of the populations with highest likelihoods were assembled for each of the ten populations for the individuals in that population. Many different populations were “most likely” for at least one individual in each population (Supplemental Table S3). Up to twelve different populations occurred as the most likely and the frequency that the population of the individual’s origin is the most likely varies considerably among populations, from 0% for some Tunisian populations up to 39% for Iranians. Noteworthy is that the value never even approaches 50% when there are many closely related populations involved.

Because the analyses in FROG-kb calculate and list likelihoods for all reference populations, the alternative populations that are not significantly less likely than the most likely are listed and those within one order of magnitude are actually flagged. While there can be argument over what likelihood ratio is required to conclude a significant difference, a standard in use for many decades in various genetic analyses is that there is no significant difference for a likelihood ratio of one order of magnitude between the best to an alternative hypothesis. In the present situation, the different possible ancestral populations are the different hypotheses. Thus, we consider that tabulating the best is less meaningful than considering all alternatives with a likelihood ratio of up to 10. That summary is presented in Supplemental Table S4. What is now clear is that the population of origin occurs among the top populations more often than any other population, with the exception of the relatively close Tunisian populations: Kerkennah, Kairun, Mehdia.

4. Discussion

For a forensic panel to provide accurate biogeographic ancestry assignment it needs (1) a sufficient number of SNPs to be effective for its purpose and (2) reference data on the relevant and

appropriate populations [7,9,10]. The SNPs most frequently identified by the various published forensic panels allow us to distinguish among five continental regions: Africa, Europe (including Southwest Asia), South Asia, East Asia (including Oceania), and the Americas [14]. Populations in Southwest Asia and other coastal Mediterranean areas share not only this geographic region but also a very long intertwined history; it is not surprising that they also share genetic information with each other and with populations in more northerly European populations. This area has been interesting to many researchers for different purposes e.g. biomedical genetic association, evolutionary and forensic studies [6,12,20–22]. The 55 AISNP panel is an efficient and globally useful panel of ancestry informative markers that is comprised of highly informative SNPs for differentiation of at least eight ancestry groups including Europe, Southwest Asia, and South Asia [7,15]. With new data on other populations in this broad region we developed a dataset of 3129 individuals in 52 populations. We used STRUCTURE and PCA analyses to visualize genetic variation among individuals and populations. The heatmap, STRUCTURE, and PCA analyses outlined in our study indicate that 32 SNPs out of these 55 give us nearly identical results with those given by the full 55 AISNP panel. This subset successfully separates our study populations into five clusters. However, we believe this set of SNPs can be improved with other population discriminating SNPs from other and/or future studies. This set of 32 SNPs is considered simply the starting point for development of an even more discriminating second tier panel of SNPs for this region of the world.

To examine the accuracy of 32 SNPs for closely related populations included in one cluster using STRUCTURE and PCA we have used the likelihood ratio approach embodied within FROG-kb. For the Turkish individuals the populations within one order of magnitude are generally from the Mediterranean region. The high-ranking populations for the Iranian individuals are mostly from the Middle Eastern, Mediterranean and South Asian regions. Tunisian individuals clustered together or with populations from the nearby Middle East. Thus, considering the populations with non-significant likelihood ratios for ancestry assignment is more accurate than using only the single most likely population. We have found that those other populations have similar geographic relatedness with the origin population. Moreover, on average the true population of origin occurs among the most likely more frequently than any other population.

Accurate prediction of the individual genetic ancestry is the main purpose of forensic studies such as this one. We used FROG-kb for individual ancestry assignment and summarized it at the population level as well. The results from FROG-kb provide relative likelihoods of ancestry from different populations for user-specified genotypes. Such data are probabilistic in principle and these results should not be taken as absolute values [7,11]. Our population assignment data reveal that frequently the population of origin is not the most likely, especially when there are several closely related populations among the reference populations. However, the population of origin will very frequently be among those populations for which the best is only better by a likelihood ratio of ten or less. This cluster of populations is suggested as a minimum confidence range. Thus, it is not possible in many cases (especially where the people are not isolated geographically or genetically) to give a single population level assignment with a limited number of AISNPs and without a large reference database. This is not particularly surprising when one recognizes that most of the markers being used are polymorphic and hence individuals with all three genotypes are expected to exist within each population. Even if the allele frequencies are not similar, an individual may have genotypes at multiple loci that are "uncommon" for their population but more common in a nearby population.

Many forensic panels have focused on small but efficient sets of AISNPs for ancestry assignment to three to five main continental populations. However, we think it is important to develop second-tier panels for significantly more accurate prediction of the ancestry of individuals within regions. Thus, we can now move forward to select new SNPs that provide additional information on differences among populations in this part of the world. For a second-tier AISNP panel focused on Mediterranean and Southwestern Asian populations we do not need to worry or consider what information the SNPs contribute on the relationships among Central, East, and Southeast Asian populations, among Pacific populations, or among Native American populations. There are many good candidates but they have not yet been typed on this set of population samples nor any comparable set of populations from this region. Once we add new markers to the dataset, reiterating the refinement procedures with the new markers added should result in an improved panel of AISNPs that provides additional information on differences among populations in this part of the world. With more SNPs with variation in this region of the world the regional accuracy is expected to increase. However, with this set of closely related populations it is unreasonable to expect perfect inference of the actual population of origin.

Conflicts of interest

The authors declare no conflicts of interest.

Ethical considerations

All samples newly tested in this study are anonymous and were previously collected with informed consent. The other data are from public resources with data from previous studies.

Acknowledgments

OB was supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK) under 2219-Grant Programme. This study was supported in part by NIJ Grant 2013-DN-BX-K023 to KKK awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Points of view in this presentation are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice. Some of the cell lines were obtained from the National Laboratory for the Genetics of Israeli Populations at Tel Aviv University. Special thanks are due to the many hundreds of individuals who volunteered to give blood samples for studies of gene frequency variation and to the many colleagues who helped us collect the samples. The authors would like to thank Eva Haigh for technical assistance, Cai-xia Li, Andrew J. Pakstis, William C. Speed, and Judith R. Kidd for their comments on the manuscript.

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.2016.04.010.

References

[1] N.A. Rosenberg, J.K. Pritchard, J.L. Weber, H.M. Cann, K.K. Kidd, L.A. Zhivotovsky, M.W. Feldman, Genetic structure of human populations, Science 298 (5602) (2002) 2381–2385.

[2] J.M. Butler, B. Budowle, P. Gill, K.K. Kidd, C. Phillips, P.M. Schneider, P.M. Vallone, N. Morling, Report on ISFG SNP panel discussion, Forensic Sci. Int. 1 (1) (2008) 471–472.

[3] C. Phillips, A. Salas, J.J. Sanchez, M. Fondevila, A. Gomez-Tato, J. Alvarez-Dios, M. Calaza, M.C. de Cal, D. Ballard, M.V. Lareu, 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.
[4] I. Halder, M. Shriver, M. Thomas, J.R. Fernandez, T. Frudakis, A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications, Hum. Mutat. 29 (5) (2008) 648–658.
[5] R. Kosoy, R. Nassir, C. Tian, P.A. White, L.M. Butler, G. Silva, R. Kittles, M.E. Alarcon-Riquelme, P.K. Gregersen, J.W. Belmont 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.
[6] C.M. Nievergelt, A.X. Maihofer, T. Shekhtman, O. Libiger, X. Wang, K.K. Kidd, J.R. Kidd, Inference of human continental origin and admixture proportions using a highly discriminative ancestry informative 41-SNP panel, Invest. Genet. 4 (1) (2013) 13.
[7] K.K. Kidd, W.C. Speed, A.J. Pakstis, M.R. Furtado, R. Fang, A. Madbouly, M. Maiers, M. Middha, F.R. Friedlaender, J.R. Kidd, Progress toward an efficient panel of SNPs for ancestry inference, Forensic Sci. Int. Genet. 10 (2014) 23–32.
[8] C. Phillips, W. Parson, B. Lundsberg, C. Santos, A. Freire-Aradas, M. Torres, M. Eduardoff, C. Borsting, P. Johansen, M. Fondevila, Building a forensic ancestry panel from the ground up: the EUROFORGEN Global AIM-SNP set, Forensic Sci. Int. Genet. 11 (2014) 13–25.
[9] P. Kersbergen, K. van Duijn, A.D. Kloosterman, J.T. den Dunnen, M. Kayser, P. de Knijff, Developing a set of ancestry-sensitive DNA markers reflecting continental origins of humans, BMC Genet. 10 (2009) 69.
[10] U. Rogalla, E. Rychlicka, M.V. Derenko, B.A. Malyarchuk, T. Grzybowski, Simple and cost-effective 14-loci SNP assay designed for differentiation of European, East Asian and African samples, Forensic Sci. Int. Genet. 14 (2015) 42–49.
[11] H. Rajeevan, U. Soundararajan, A.J. Pakstis, K.K. Kidd, Introducing the forensic research/reference on genetics knowledge base, FROG-kb, Invest. Genet. 3 (1) (2012) 18.
[12] X. Yang, S. Al-Bustan, Q. Feng, W. Guo, Z. Ma, M. Marafie, S. Jacob, F. Al-Mulla, S. Xu, The influence of admixture and consanguinity on population genetic diversity in Middle East, J. Hum. Genet. 59 (11) (2014) 615–622.
[13] G. Hellenthal, G.B. Busby, G. Band, J.F. Wilson, C. Capelli, D. Falush, S. Myers, A genetic atlas of human admixture history, Science 343 (6172) (2014) 747–751.
[14] 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.
[15] A.J. Pakstis, E. Haigh, L. Cherni, A.B. ElGaaied, A. Barton, B. Evsanaa, A. Togtokh, J. Brissenden, J. Roscoe, O. Bulbul, 52 additional reference population samples for the 55 AISNP panel, Forensic Sci. Int. Genet. 19 (2015) 269–271.
[16] N.A. Rosenberg, L.M. Li, R. Ward, J.K. Pritchard, Informativeness of genetic markers for inference of ancestry, Am. J. Hum. Genet. 73 (6) (2003) 1402–1422.
[17] 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, Invest. Genet. 2 (1) (2011) 1.
[18] R.C. Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2014.
[19] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics 155 (2) (2000) 945–959.
[20] C. Phillips, A. Freire Aradas, A.K. Kriegel, M. Fondevila, O. Bulbul, C. Santos, F. Serrulla Rech, M.D. Perez Carceles, A. Carracedo, P.M. Schneider, Eurasiaplex: a forensic SNP assay for differentiating European and South Asian ancestries, Forensic Sci. Int. Genet. 7 (3) (2013) 359–366.
[21] O. Bulbul, G. Filoglu, T. Zorlu, H. Altuncul, A. Freire-Aradas, J. Sochtig, Y. Ruiz, M. Klintschar, S. Triki-Fendri, A. Rebai, Inference of biogeographical ancestry across central regions of Eurasia, Int. J. Legal Med. 130 (1) (2016) 73–79.
[22] L. Cherni, A.J. Pakstis, S. Boussetta, S. Elkamel, S. Frigi, H. Khodjet-El-Khil, A. Barton, E. Haigh, W.C. Speed, A. Ben Ammar Elgaaied, et al., Genetic variation in Tunisia in the context of human diversity worldwide, Am. J. Phys. Anthropol. (April 22) (2016) (in press).