Int J Legal Med (2016) 130:73–79
DOI 10.1007/s00414-015-1246-7

ORIGINAL ARTICLE

Inference of biogeographical ancestry across central regions of Eurasia

O. Bulbul[1,2] · G. Filoglu[1] · T. Zorlu[1] · H. Altuncul[1] · A. Freire-Aradas[3] · J. Söchtig[3,4] · Y. Ruiz[3] · M. Klintschar[4] · S. Triki-Fendri[5] · A. Rebai[5] · C. Phillips[3] · M. V. Lareu[3] · Á. Carracedo[3,6] · P. M. Schneider[2]

Received: 17 February 2015 / Accepted: 10 August 2015 / Published online: 20 August 2015

Abstract The inference of biogeographical ancestry (BGA) can provide useful information for forensic investigators when there are no suspects to be compared with DNA collected at the crime scene or when no DNA database matches exist. Although public databases are increasing in size and population scope, there is a lack of information regarding genetic variation in Eurasian populations, especially in central regions such as the Middle East. Inhabitants of these regions show a high degree of genetic admixture, characterized by an allele frequency cline running from NW Europe to East Asia. Although a proper differentiation has been established between the cline extremes of western Europe and South Asia, populations geographically located in between, i.e, Middle East and Mediterranean populations, require more detailed study in order to characterize their genetic background as well as to further understand their demographic histories. To initiate these studies, three ancestry informative SNP (AI-SNP) multiplex panels: the SNPforID 34-plex, Eurasiaplex and a novel 33-plex assay were used to describe the ancestry patterns of a total of 24 populations ranging across the longitudinal axis from NW Europe to East Asia. Different ancestry inference approaches, including STRUCTURE, PCA, DAPC and Snipper Bayes analysis, were applied to determine relationships among populations. The structure results show differentiation between continental groups and a NW to SE allele frequency cline running across Eurasian populations. This study adds useful population data that could be used as reference genotypes for future ancestry investigations in forensic cases. The 33-plex assay also includes pigmentation predictive SNPs, but this study primarily focused on Eurasian population differentiation using 33-plex and its combination with the other two AI-SNP sets.

Keywords Biogeographic ancestry (BGA) · Eurasia · Middle East · Ancestry-informative single nucleotide polymorphisms (AI-SNPs)

Electronic supplementary material The online version of this article (doi:10.1007/s00414-015-1246-7) contains supplementary material, which is available to authorized users.

✉ O. Bulbul
ozlembulbul00@gmail.com

[1] Institute of Forensic Sciences, Istanbul University, Istanbul, Turkey

[2] Institute of Legal Medicine, Faculty of Medicine, University of Cologne, Cologne, Germany

[3] Forensic Genetics Unit, Institute of Forensic Science "Luís Concheiro", University of Santiago de Compostela, Santiago de Compostela, Spain

[4] Department of Forensic Medicine, Hannover Medical School, Hannover, Germany

[5] Bioinformatics Research Group, Centre of Biotechnology of Sfax, Sfax, Tunisia

[6] Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Jeddah, Saudi Arabia

Introduction

When information about a suspect is not available even after STR typing of evidential material, further DNA analyses can provide additional data such as the inference of biogeographical ancestry (BGA) of the DNA donors [1, 2]. Such information can also be useful in cases of missing persons or disaster victim identification. Single-nucleotide polymorphisms (SNPs), the most common genetic variation, are often linked with specific populations due to geographical and cultural isolation or by the action of random genetic drift and selection. These markers can be assembled into sets of ancestry-informative SNPs [1–3], termed AI-SNPs.

Although a range of AI-SNP assays is available to differentiate the major continentally defined population groups [4–6], there are limited studies of Eurasian and Middle Eastern populations—those occupying regions located between the margins of the European, Asian and African continents [7]. Understanding the genetic structure of populations occupying North Africa, Eastern Europe, Middle East and Central South Asia is important, not only from a forensic perspective but also for the interpretation of the clinal patterns of variation running from Europe to Asia. Middle East populations have very low levels of genetic divergence with Europe, while South Asian populations show more divergence [7–10]. Genetic divergence between Middle East and South Asian populations is much lower due to their close geographic proximity as well as thousands of years of trade.

In this study, we have developed a novel SNP assay that combines 22 AI-SNPs and 11 SNPs associated with human pigmentation in a single PCR reaction followed by single-base extension using SNaPshot. This test was specifically developed to be analysed in addition to two ancestry informative panels: the SNPforID 34-plex [11] and Eurasiaplex assays [7] that have been previously established to increase the differentiation of populations in the closely related central regions of Eurasia (Middle East and Eastern Mediterranean regions). The ancestry patterns of 24 populations of Central Eurasia, located along the longitudinal axis from NW Europe to East Asia, were analysed by using the combination of 87 non-overlapping SNPs from these three panels, of which 82 are AI-SNPs.

Materials and methods

Population samples and DNA extraction

Study samples were obtained from 12 Eurasian populations: Germany (N=10), northwest Spain (N=24), Greece (N=42), Turkey (N=100), Morocco (N=41), Egypt (N=37), Kuwait (N=39), Libya (N=35), Yemen (N=36), Azerbaijan (N=43), India (N=23) and Vietnam (N=20). Informed consent was given by all the healthy unrelated individuals participating in the study. Ethical approval was granted from the ethics committee of the Istanbul University, Cerrahpasa Medicine Faculty, Turkey. DNA was extracted from saliva or whole blood using the QIAGEN M48 Biorobot (QIAGEN, Germany) or a standard phenol/chloroform method. SNP genotypes were also collected for a total of 980 samples from 13 1000 Genomes phase 3 populations comprising: Luhya in Webuye, Kenya (LWK) (N=67); Yoruba in Ibadan, Nigeria (YRI) (N=78); Utah residents with European ancestry (CEU) (N=90); Finnish (FIN) (N=36); British (GBR) (N=43); Han Chinese (CHB) (N=68); Southern Han Chinese (CHS) (N=25); Japanese (JPT) (N=84); Bengali in Bangladesh (BEB) (N=86); Gujarati Indian in Houston, Texas (GIH) (N=103); Indian Telugu in the UK (ITU) (N=102); Punjabi in Lahore, Pakistan (PJL) (N=96)and Sri Lankan Tamil in the UK (STU) (N=102). Additionally, data for 306 samples from the CEPH human genome diversity panel (HGDP-CEPH) were collected for 16 population groups: Italy (Sardinian, Bergamo, and Tuscan) (N=20, N=10 and N=8); Russia–Adygei (N=17); Algeria–Mozabite (N=26); Israel (Bedouin, Palestinian and Druze) (N=22, N=40 and N=24); Pakistan (Makrani, Sindhi, Pathan, Kalash, Burusho and Hazara) (N=4, N=24, N=22, N=16, N=15 and N=20); Japan (N=28) and China–Uygur (N=10). Both 1000 Genomes and HGDP-CEPH data were accessed using the SPSmart SNP frequency browser [12]. Genotypes for a further 18 SNPs included in the present study but not available in public databases were obtained by direct genotyping of the same HGDP-CEPH populations (SNP identifiers for these 18 markers are marked in Supplementary File S1).

SNP selection for the novel 33-plex assay

A total of 22 AI-SNPs were selected from the NCBI dbSNP database according to the following criteria: minor allele frequencies in the range 0.01–0.05 in 1000 Genomes Europeans and strongly contrasting allele frequency differences between Europeans and East Asians. In addition, 11 recently published SNPs shown to be associated with human eye, hair, and skin colour variation (some of them are informative for both ancestry and pigmentation) [13–16] were included to build a multiplex of 33 SNPs. Although the pigmentation predictive SNPs included in the 33-plex differ from those of other forensic sets for this purpose [13–16] and formal studies have been made of their predictive power, we focus here on the population differentiation capacity of the AI-SNPs in the 33-plex panel.

Multiplex development and additional AI-SNPs typing

PCR and single-base extension (SBE) primers for a 33-plex PCR and SNaPshot assay were designed with Primer3 Plus software to provide amplicon sizes of 53–148 bp [17]. Primers were checked for primer-dimer interactions and hairpin structures using Autodimer [18] and for homology using BLAST. Nonspecific pigtails were used to space the extension products. PCR assays with a total volume of 10 μL contained 4 μL of 2× QIAGEN Multiplex Master Mix, 3 μL of PCR primer mix (0.1–0.7 μM), 2 μL DNA and (adjustable with) 1 μL water. PCR conditions have been adjusted for efficient amplification of specific targets. PCR used an AB GeneAmp® 9700 thermal cycler with the program: denaturation at 95 °C for 10 min then 35 cycles of 95 °C for 30 s, 60 °C for 50 s, 65 °C for 40 s, and final extension at 65 °C for 6 min. Excess primers and dNTPs were removed with 1 μL ExoSAP-IT (1 U/μL Exonuclease I and Shrimp Alkaline Phosphatase, GE Healthcare) to 2.5 μL PCR product and incubation at

37 °C for 45 min, 85 °C for 15 min. Single-base extension reactions were carried out in 6-μL volumes containing 2.5 μL of SNaPshot™ reaction mix (AB), 1.5 μL of SBE primer mix (0.1–0.2 μM) and 2 μL of purified DNA using the following conditions: 30 cycles of 96 °C for 10 s, 55 °C for 5 s and 60 °C for 30 s. Excess nucleotides were removed with 1 μL SAP (1 U/μL Shrimp Alkaline Phosphatase, GE Healthcare) added to the total volume of extension products incubated at 37 °C for 80 min and 85 °C for 15 min. Capillary electrophoresis was performed using an ABI PRISM 3130 Genetic Analyzer (AB) with GeneMapper IDX or v.4.0 software (Life Technologies, USA). Details of the developed 33 SNPs, as well as their CE bins and panels, can be found in Supplementary Tables S1–S2. The additional 57 AI-SNPs were genotyped using 23-plex plus 34-plex SNaPshot assays as previously described [7, 11].

Statistical analysis

Pairwise F_{ST} values were calculated using Arlequin v. 3.5 [19] and biogeographic ancestry was analysed using STRUCTURE v. 2.3.3 [20]. STRUCTURE parameters comprised: 100,000 burnins (retaining the next 100,000 MCMC), admixture/POPFLAG model, independent allele frequencies, five independent replicates per cluster from K:2 to K:6. Graphics were constructed using CLUMPP v1.1.2 [21] and Distruct v1.1 software [22]. Principal component analysis (PCA) and discriminant analysis of principal components (DAPC) were performed using R v. 2.11.1 [23] and packages SNPassoc [24] and Adegenet [25], respectively. Although PCA is in common use, DAPC is a relatively new methodology providing an efficient description of genetic clusters using synthetic variables (called the discriminant functions). DAPC uses the concept of a training set to create classification rules and a test set to gauge their efficiency. Cross-validation and ancestry assignment likelihoods were measured using the USC Bayesian forensic SNP classifier Snipper [26].

Results

Characteristics of the 33-plex assay

Allele frequency comparisons for Africa, Europe and East Asia for the selected 33-plex AI-SNPs are detailed in Supplementary Fig. S1. An example profile of the 33-plex SNP assay is shown in Supplementary Fig. S2. The relative power to differentiate populations was assessed using Rosenberg divergence values ($\mathbf{I_n}$) for five-population group (African vs. non-African, Europe vs. non-Europe, Middle East vs. non-Middle East, Central South Asia vs. non-Central South Asia and East Asia vs. non-East Asia). Cumulative $\mathbf{I_n}$ values for each multiplex and combined multiplexes are shown in Supplementary Table S3A. Combining data from all three SNP multiplexes provides optimum differentiation of Europeans and East Asians. Values were also compared for 33-plex SNPs plus established 34-plex and Eurasiaplex SNP sets for three-population group (African vs. non-African, Europe vs. non-Europe and East Asia vs. non-East Asia). Comparisons of $\mathbf{I_n}$ values are given in Supplementary Table S3B. Cumulative informativeness values are balanced and reached highest values when three multiplex set was used.

Measure of genetic distances by pairwise F_{ST} estimations

Graphic representation of the analysis of pairwise F_{ST} values, pairwise between-population differences and pairwise within-population differences, in the 87-SNP combination, is depicted in Supplementary Fig. S3. The underlying F_{ST} values are listed in Supplementary Table S4. The most distant population groups were Europeans vs. East Asians, particularly CEU and GBR (average F_{ST} 0.483 and 0.486, respectively), followed by CEPH Italian Sardinian, Bergamo and Tuscans (average F_{ST} 0.475). A gradual but consistent decrease in F_{ST} values is observed from left to right, reflecting the longitudinal cline in allele frequencies across the Eurasian geographical area. We observed Kuwaitis are closer to East Asians than other Middle East samples (average F_{ST} 0.242 vs. 0.370) while also showing the lowest levels of within-population variation. Algeria, Morocco and Central South Asian Uygur show the highest levels of within-population variation.

Analysis of genetic structure

Figure 1 shows STRUCTURE cluster plots for population structure models K:2 to K:5 using 87 SNPs. Reference populations comprised 1000 Genomes African, European, Central South Asian and East Asian plus HGDP-CEPH Middle East populations. Study samples cover a wide range of Eurasian populations: Germany, northwest Spain, Italia (Sardinia, Bergamo and Tuscan), Greece, Turkey, Russia–Adygei, Azerbaijan, Egypt, Kuwait, Libya, Morocco, Yemen, India, Pakistan (Makrani, Sindhi, Pathan, Kalash, Burusho and Hazara), China–Uygur, Vietnam and Japan. The optimum clustering was obtained at K:5. The K:2 analyses separate Africa and East Asia as a single cluster from the three Eurasian regions due to the strongly contrasting allele frequencies differences in the SNPs selected. Then, Africa and East Asia are differentiated, followed by Central South Asia at K:4. Population group clusters at K:5 are almost completely independent for all reference populations. Among study populations, majority of European ancestry is evident in Germany, NW Spain, Italian Bergamo and Tuscans. In contrast, Adygei, Italian Sardinians and Greeks show admixed patterns from co-ancestries of Europe and Middle East at comparable levels but with Europe as the main component. Similar co-ancestry patterns are present

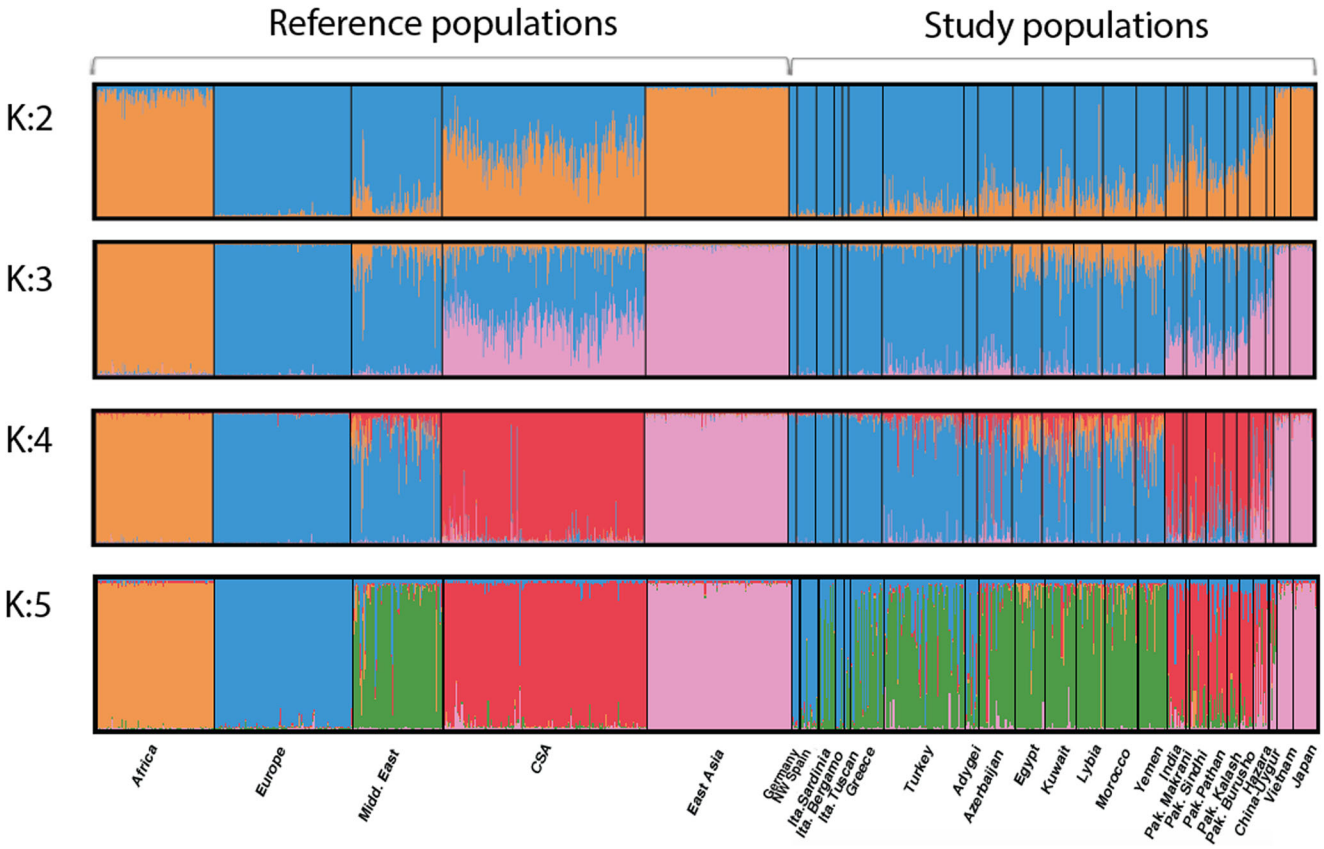

Fig. 1 STRUCTURE cluster plots for the K:2 to K:5 range of population structure models, using 87 SNPs. Reference populations are Africa (*orange*), Europe (*blue*), Middle East (*green*), Central South Asia (*red*) and East Asia (*pink*). Ancestry patterns for the study populations closely match those of the reference population clusters detected in the SNP variation analysed

in Turkish and Azerbaijani samples but in opposite proportions. In several Turkish and Azerbaijani individuals, small proportions of Central South Asian and East Asian co-ancestries were also observed. Individuals from Middle East study populations (Egypt, Kuwait, Libya, Morocco and Yemen) uniformly show the Middle East cluster membership as the major proportion (mean cluster membership proportion: 0.860). A study of South Asian populations (India and six Pakistan regions) displays majority Central South Asian ancestry except Hazara and Uygur. Vietnamese and Japanese show majority East Asian ancestry.

Population divergence assessed with PCA and DAPC analyses

We performed PCA analysing 29 populations (including five reference populations) from Africa, the three Eurasian regions (Europe, Middle East and Central South Asia) and East Asia using 34-plex alone, 34-plex and Eurasiaplex, then 34-plex, Eurasiaplex and 33-plex combined. PCA (Fig. S4) shows increased differentiation among study and reference populations when 34-plex is expanded with additional panels. Central South Asian population separation is improved after adding Eurasiaplex. Furthermore, 33-plex increased separation among Eurasian and East Asian populations. Hence, the best separation of the five groups is obtained using all three SNP assays together since each panel contributes to overall differentiation differently.

Detailed three multiplex combined PCA results are plotted in Fig. 2a. The first principal component (PC1) is 19.31 % of variation and mainly separates Europe from East Asia and Central South Asia from Europe. Although the second PC provides Middle East differentiation (PC2 10.23 %), divergence is much more reduced and some overlapping points are present with the European and Central South Asian clusters. Additionally, some Middle East individuals are slightly displaced towards the African cluster and some outlier populations, notably: Azerbaijani, Greek and Turkish are not positioned in their corresponding Eurasian clusters. In order to improve further separation, a PCA of reduced population comparisons focused on Middle East and European samples (green and blue tonalities, respectively) is shown in Fig. 2b. This PCA gave more distinct differentiation between both groups. Furthermore, Fig. 2b indicates genetic variation is generally discrete between Middle Eastern and European populations. Mediterranean and Eastern European populations

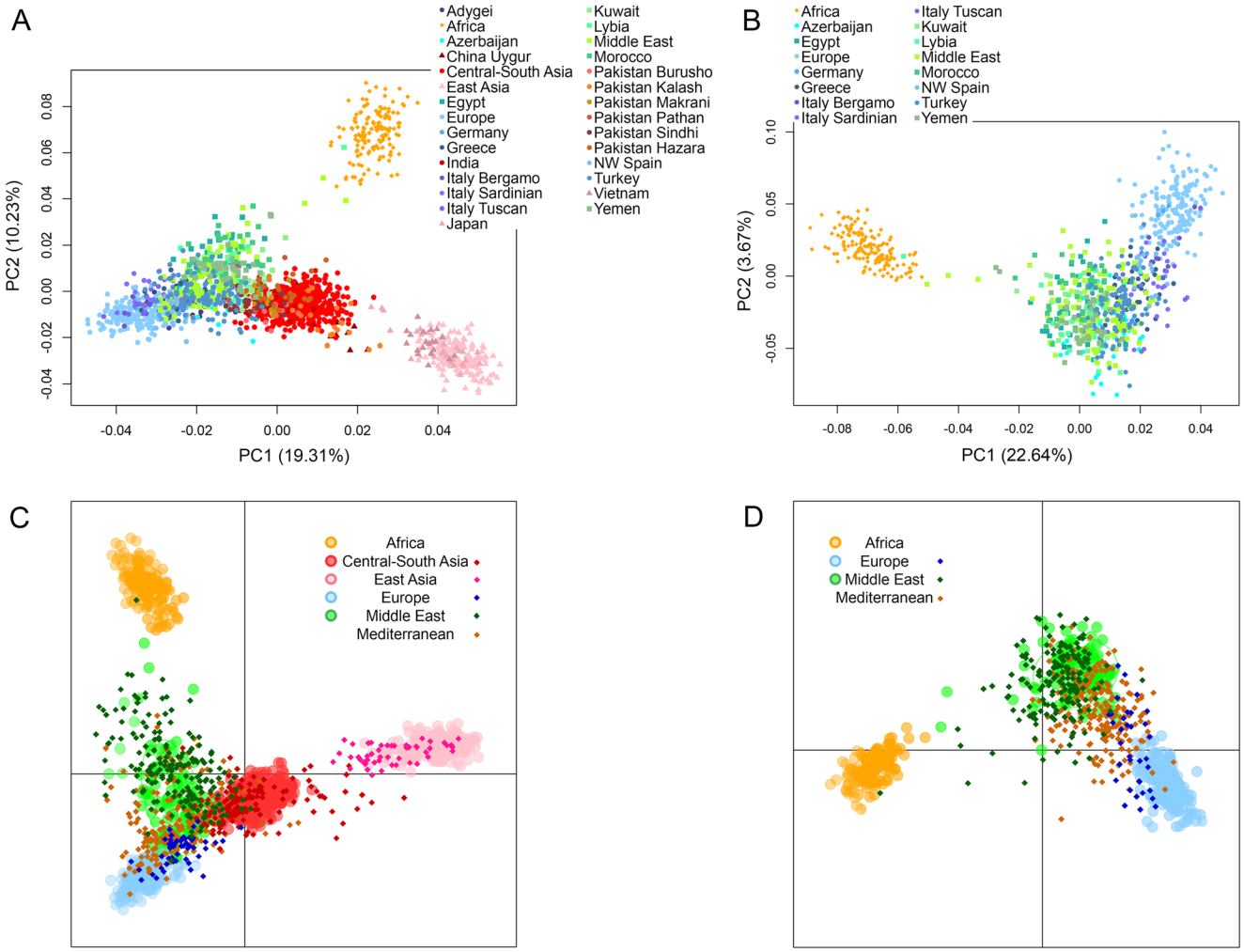

Fig. 2 Principal component analysis and discriminant analysis of principal components analysis of the SNP data. **a** PCA plot for Africa (*orange*), Europe (*blue*), Middle East (*green*), Central South Asia (*red*) and East Asia (*pink*). **b** PCA plot for Middle East vs. Europe and Africa. **c** DAPC plot for Africa, Europe, Middle East, Central South Asia and East Asia populations. Note: *light colours* define reference populations and *dark colours* study populations. **d** DAPC plot for Middle East vs. Europe and Africa

(Turkish, Italians, Greek plus Adygei, Azerbaijani, respectively) cluster between these two groups and some individuals overlap with Middle East or European population clusters. Therefore, genetic variation is more continuous within Middle Eastern and Mediterranean populations and within Mediterranean and European populations.

DAPC analysis of the same 29 populations shows similar results to PCA (Fig. 2c) depicting five uniform clusters. Light and dark colours in Fig. 2c correspond to reference and testing samples, respectively. From the observation of overlapping points for a proportion of Middle East samples, subsequent DAPC analysis removed East Asians and Central South Asians (Fig. 2d), leading to an improved separation of Middle East populations from those of European populations. Eastern-extreme European populations of Azerbaijani, Greek, Turkish and Sardinians are located between European and Middle Eastern population or overlap with these populations.

Bayesian classification analysis

The ancestry informativeness of SNP panel combinations for differentiating 1000 Genomes and HGDP-CEPH reference samples was tested by cross-validation in a five-group model and gave classification success rates of African 100 %, European 95.27 %, Middle East 95.54 %, Central South Asian 100 % and East Asian 100 %. These measurements of classification success/error support the approach used to create five distinct reference population groups for the classification system (Supplementary Table S5). Bayesian ancestry assignments for European vs. Middle East were made with Snipper, and Supplementary Fig. S5 shows a total of 743 individuals from 16 populations assigned as European, Mediterranean or Middle Eastern and ordered by the likelihood ratios displayed. Using a classification threshold of 100 times more likely, Europeans are successfully separated from Middle Eastern samples. However, Mediterraneans show close affinity to either

European or Middle Eastern samples despite displaying a possible separate cluster in DAPC analysis, meaning a higher likelihood threshold must be applied in order to improve the accuracy of the classifications of Eurasians from all regions of Europe and the Middle East.

Discussion

Individuals from the same geographic region or population tend to share similar parental ancestry components [20]. Although diversity and fine-scale population structure exists within Africa, Europe and East Asia, it is difficult to estimate the level of admixture for populations located in intermediate regions between these three continents [27]. East European and Middle East populations are located in the centre of an extensive area that has been the corridor for much human migration into Europe, though detailed studies of the genetic structure in populations of both regions are limited. Recently, Khodjet-el-Khil et al. [28] studied Tunisian, Libyan, Algerian and Moroccan populations with the 34-plex SNPs used in the current study and obtained a progressive clustering of individuals that placed these North African populations into the Middle Eastern cluster. The results we report in this study describe a genetic structure for a larger longitudinal set of populations ranged across Eurasia from NW Europe to East Asia. The STRUCTURE, PCA, DAPC and Snipper-based Bayes analyses outlined in our study all detected a proportion of the Mediterranean cluster (comprising Bergamo, Sardinia and Tuscan Italians; Greeks, Turkish; Adygei and Azerbaijanis), which appeared as a third group of populations between NW European and Middle Eastern clusters. This finding is similar to the analysis of the same populations using 34-plex and Eurasiaplex SNPs, where STRUCTURE results and cross-validation with Snipper indicated distinct patterns of variability from the rest of Europe and a relatively high misclassification rate for incorrectly inferred Middle East ancestry [7]. Such patterns could represent a signal of the original Neolithic migrants from the Middle East, who began moving into Europe more than 8000 years ago, likely from the western part of the Fertile Crescent. These migrants of Middle East origin contributed significantly to variation detected in early European farmers, and this variation is present as detectable genetic cluster components in nearly all modern European populations but to varying degrees [29, 30].

Ancestry assignments for the Mediterranean population group cluster using Snipper showed that admixed ancestry could not be accurately estimated unless the ancestral populations were represented among the reference populations. In order to minimize classification errors between European and Middle Eastern populations, a classification threshold of "100 times more likely" is suggested, as previously indicated by the application of 34-plex and Eurasiaplex SNPs [7]. When attempting to differentiate Mediterranean from Middle East individuals, this probability threshold could be increased further to reduce classification error but the use of 87 SNPs still does not provide a reliable framework for securely differentiating these two population groups. Recent published candidate SNPs could be incorporated in further research to provide additional information on differences among these population groups [31, 32].

Overall, we combined 34-plex and Eurasiaplex with a new 33-plex panel to obtain increased genetic resolution of the Middle East and Eastern European–Mediterranean populations. This set of SNPs can also be run on massively parallel sequencing machines, particularly as high throughput DNA sequencing technology is a promising technology for forensic SNP analysis. Further work to select the best SNPs for improved geographic resolution for forensic applications will form the next phase of our development of optimum panels to analyse Eurasian populations.

Acknowledgments This research was supported by Istanbul University Scientific Research Projects Unit (BAP) under project number 19042. OB was supported by the German Academic Exchange Service (DAAD) under Research Grants for Doctoral Candidates and Young Academics and Scientists Programme and by The Scientific And Technological Research Council Of Turkey (TÜBİTAK) under 2211-Grant Programme. We sincerely thank all the sample contributors. The authors would like to thank Suad Alfadhli from the Department of Medical Laboratory Sciences, Faculty of Allied Health Sciences, Kuwait University, for collecting the Kuwait samples, Department of Criminalistic Investigations DNA Laboratory, Ministry of Internal Affairs of Azerbaijan Republic and Katharina Läer from the Department of Forensic Medicine, Hannover Medical School, Germany for technical assistance.

References

1. Kayser M, de Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. Nat Rev Genet 12:179–192
2. Phillips C (2015) Forensic genetic analysis of bio-geographical ancestry. Forensic Sci. Int. Genet. doi: 10.1016/j.fsigen.2015.05.012 (e-pub ahead of print)
3. Phillips C (2013) Ancestry informative markers. In: J.A. Siegel, P.J. Saukko (Eds.) Encyclopedia of Forensic Sciences, 2nd ed., vol. 1, Academic Press, 323–331
4. Phillips C, Salas A, Sánchez JJ, Fondevila M, Gómez-Tato A, Álvarez-Dios J, Calaza M, Casares de Cal M, Ballard D, Lareu MV, Carracedo Á (2007) The SNPforID consortium, inferring ancestral origin using a single multiplex assay of ancestry—informative marker SNPs. Forensic Sci Int Genet 1:273–280
5. Halder I, Shriver M, Thomas M, Fernandez JR, Frudakis TN (2008) A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. Hum Mutat 29(5):648–658
6. Kersbergen P, van Duijn K, Kloosterman AD, den Dunnen JT, Kayser M, de Knijff P (2009) Developing a set of ancestry-sensitive DNA markers reflecting continental origins of humans. BMC Genet 10:69
7. Phillips C, Freire Aradas A, Kriegel AK, Fondevila M, Bulbul O, Santos C, Serrulla Rech F, Perez Carceles MD, Carracedo Á,

Schneider PM, Lareu MV (2013) Eurasiaplex: a forensic SNP assay for differentiating European and South Asian ancestries. Forensic Sci Int Genet 7:359–366
8. Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. Nature 461:489–494
9. Zerjal T, Spencer Wells R, Yuldasheva N, Ruzibakiev R, Tyler-Smith C (2002) A genetic landscape reshaped by recent events: Y-chromosomal insights into Central Asia. Am J Hum Genet 71: 466–482
10. Spencer Wells R, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S et al (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. Proc Natl Acad Sci U S A 98:10244–10249
11. Fondevila M, Phillips C, Santos C, Freire Aradas A, Vallone PM, Butler JM, Lareu MV, Carracedo Á (2013) Revision of the SNPforID 34-plex forensic ancestry test: assay enhancements, standard reference sample genotypes and extended population studies. Forensic Sci Int Genet 7:63–74
12. J. Amigo, A. Salas, C. Phillips, Á. Carracedo (2008) SPSmart: adapting population based SNP genotype databases for fast and comprehensive web access, BMC Bioinformatics, 9 428. URL: http://spsmart.cesga.es, accessed September 2014
13. Liu F, van Duijn K, Vingerling JR, Hofman A, Uitterlinden AG, Janssens ACJW, Kayser M (2009) Eye color and the prediction of complex phenotypes from genotypes. Curr Biol 19:192–193
14. Walsh S, Liu F, Ballantyne KN, van Oven M, Lao O, Kayser M (2011) Irisplex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci Int Genet 5:170–180
15. Bulbul O, Filoglu G, Altuncul H, Aradas AF, Ruiz Y, Fondevila M, Phillips C, Carracedo Á, Kriegel AK, Schneider PM (2011) A SNP multiplex for the simultaneous prediction of biogeographic ancestry and pigmentation type. Forensic Sci Int Genet Supp Ser 3:500–501
16. Ruiz Y, Phillips C, Gomez-Tato A, Alvarez-Dios J, Casares de Cal M, Cruz RR, Maroñas O, Söchtig J, Fondevila M, Rodriguez-Cid MJ, Carracedo Á, Lareu MV (2013) Further development of forensic eye color predictive tests. Forensic Sci Int Genet 7:28–40
17. Untergrasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities and interfaces. Nucleic Acids Res 40:e115
18. Vallone PM, Butler JM (2004) AutoDimer: a screening tool for primer–dimer and hairpin structures. Biotechniques 37:226–231
19. Excoffier LG, Schneider S (2005) Arlequin v. 3.0: an integrated software package for population genetics data analysis. Evol Bioinformatics Online 1:47–50
20. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population STRUCTURE using multilocus genotype data. Genetics 155: 945–959
21. Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806
22. Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138
23. Core Team R (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
24. Gonzalez JR, Armengol L, Sole X, Guino E, Mercader JM, Estivill X, Moreno V (2007) SNPassoc: an R package to perform whole genome association studies. Bioinformatics 23:644–645
25. Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94
26. The Snipper app suite tool: http:// mathgene.usc.es/Snipper/ Cross validation: http://mathgene.usc.es/Snipper/analysispopfile2_new.html, custom ancestry classifications: Accessed September 2014
27. Jobling M, Hollox E, Hurles M, Kivisild T, Tyler-Smith C (2013) Human evolutionary genetics (2nd ed.). New York: Garland Science. Chapter 14: What happens when populations meet? 443-477 pp
28. Khodjet-el-Khil H, Fadhlaoui-Zid K, Cherni L, Phillips C, Fondevila M, Carracedo Á, Ben Ammar-Elgaaied A (2011) Genetic analysis of the SNPforID 34-plex ancestry informative SNP panel in Tunisian and Libyan populations. Forensic Sci Int Genet 3:e45–e47
29. Zalloua PA, Platt DE, El Siba M, Khalife J, Makhoul N, Haber M, Xue MY, Izaabel H, Bosch E et al (2008) Identifying genetic traces of historical expansions: Phoenician footprints in the Mediterranean. Am J Hum Genet 83:633–642
30. Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG et al (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513:409–413
31. Kidd KK, Speed WC, Pakstis AJ, Furtado MR, Fang R, Madbouly A, Maiers M, Middha M, Friedlaender FR, Kidd JR (2014) Progress toward an efficient panel of SNPs for ancestry inference. Forensic Sci Int Genet 10:23–32
32. Kidd JR, Friedlaender FR, Speed WC, Pakstis AJ, De La Vega FM, Kidd KK (2011) Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples. Investig Genet 2:1