



## Studies of East European populations with a 46-plex ancestry-informative indel set



O. Bulbul<sup>a,\*</sup>, A. Duvenci<sup>a</sup>, T. Zorlu<sup>a</sup>, C. Gurkan<sup>b</sup>, C. Santos<sup>c</sup>, C. Phillips<sup>c</sup>, M.V. Lareu<sup>c</sup>, G. Filoglu<sup>a</sup>

<sup>a</sup> Institute of Forensic Science, Istanbul University, Istanbul, Turkey

<sup>b</sup> Turkish Cypriot DNA Laboratory, Nicosia, North Cyprus, Turkey

<sup>c</sup> Forensic Genetics Unit, Institute of Forensic Science "Luis Concheiro", University of Santiago de Compostela, Spain

### ARTICLE INFO

#### Article history:

Received 25 August 2015

Accepted 7 September 2015

Available online 12 September 2015

#### Keywords:

InDels

Ancestry information markers

East Europe

Turkey

Cyprus

Azerbaijan

### ABSTRACT

There are numerous ancestry informative markers (e.g., SNPs, InDels) used for distinguishing ancestral origins among continental regions of the world. This work presents the 46-plex ancestry informative InDels set data for three East European populations (Turkish, Turkish Cypriot and Azerbaijani).

© 2015 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

Insertion-deletion polymorphisms (InDels) have various characteristics (e.g., low mutation rates, short amplicon sizes and a very straightforward amplification technique) that make them well suited for forensic DNA analysis. InDels are also practical loci for the forensic analysis of biogeographic ancestry as they can show high allele frequency differentiation amongst population groups, while mixed DNA is less likely to go undetected and be misinterpreted as indicating admixed ancestry [1,2]. In this study we investigated patterns of genetic variation from 46 ancestry-informative InDels in Turkish, Turkish Cypriots and Azerbaijani populations.

## 2. Material and methods

40 unrelated individuals from each of the Turkish, Turkish Cypriots and Azerbaijani populations were genotyped. Genotypes were also collected for a total of 857 samples from the CEPH human genome diversity panel (HGDP-CEPH) corresponding to forty-five population groups: seven African ( $n = 105$ ), four Middle

Eastern ( $n = 163$ ), eight European ( $n = 158$ ), nine Central South Asian ( $n = 202$ ) and seventeen East Asian ( $n = 229$ ) populations. The HGDP-CEPH data were accessed using the *SPSmart forInDel* browser. (<http://spsmart.cesga.es/forindel.php>) [3]. PCR amplification and fragment analysis were performed following the protocol in [1]. Allele frequencies, Hardy Weinberg equilibrium (HWE) and pairwise  $F_{ST}$  values were calculated using Arlequin v. 3.5 [4]. Biogeographic ancestry was analyzed using STRUCTURE v. 2.3.4 with a burnin length of 10,000 followed by 10,000 MCMC repetitions [5]. CLUMPAK was applied to create cluster plots [6]. Multiple profile ancestry assignment likelihoods and PCA were measured using the USC Bayesian forensic SNP classifier *Snipper* (<http://mathgene.usc.es/snipper/analysismultipleprofiles.html>).

## 3. Results

The calculated pairwise genetic distances are presented in Table 1. The differences between each pair of the five major populations (Africa, Middle East, Central South Asia and East Asia) were all found to be significant (the level of significance,  $p < 0.001$  after the Bonferroni correction). The Turkish, Turkish Cypriot and Azerbaijani populations show very low  $F_{ST}$  values between each other and indicate very strong genetic similarities.

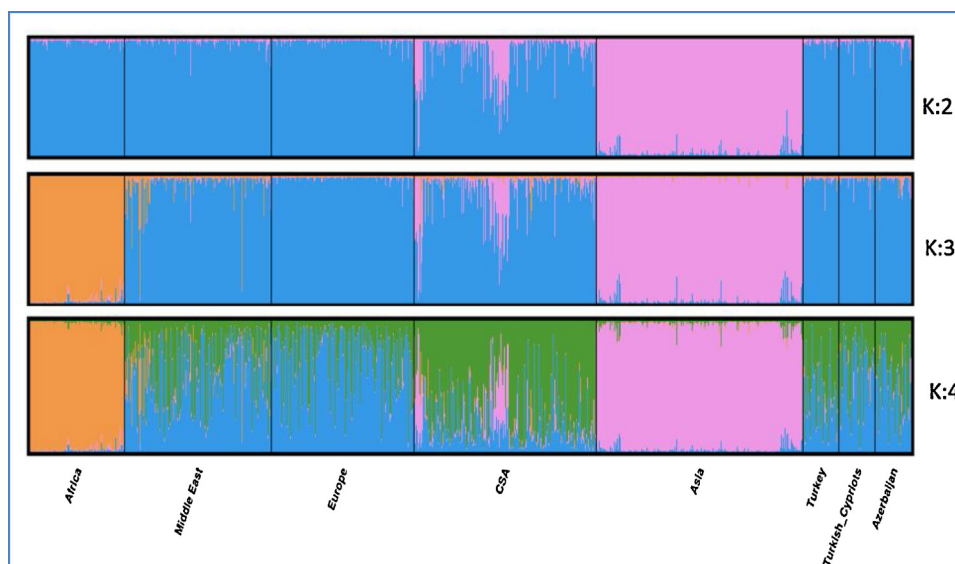
\* Corresponding author Fax: +90 212 5880011.

E-mail address: [ozlembulbul00@gmail.com](mailto:ozlembulbul00@gmail.com) (O. Bulbul).

**Table 1**

Pairwise genetic distance matrix based on the  $F_{ST}$  values.  $P$ -values are shown above and the  $F_{ST}$  values below the diagonal.  $P$ -values in bold/italics indicate non-significant distances between the corresponding population pairs ( $p$ -value  $> 1 \times 10^{-3}$ ).

Pops	Africa	Middle East	CSA	Europe	Easia	Turkey	Turkish Cypriot	Azerbaijan
Africa	*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Middle East	0.2896	*	0.0000	0.0000	0.0000	0.0000	0.0002	0.0003
CSA	0.3654	0.0150	*	0.0000	0.0000	0.0001	0.0000	0.0000
Europe	0.2893	0.0274	0.0353	*	0.0000	0.0000	0.0000	0.0000
Easia	0.3921	0.2514	0.2835	0.1808	*	0.0000	0.0000	0.0000
Turkey	0.3396	0.0118	0.0087	0.0095	0.2480	*	<b>0.0552</b>	<b>0.1592</b>
Turkish Cypriot	0.3361	0.0089	0.0106	0.0242	0.2625	0.0058	*	<b>0.0061</b>
Azerbaijan	0.3300	0.0080	0.0140	0.0127	0.2465	0.0032	0.0089	*



**Fig. 1.** STRUCTURE analysis of grouped reference populations (leftmost), Turkish, Turkish Cypriot and Azerbaijani populations ( $K=2-4$ ) (rightmost). Each vertical bar represents one individual and the colours represent the individual admixture proportions based on  $K$  assumed clusters.

Cluster plots indicate ancestral membership proportions of the eight populations for  $K=2-4$  based on the STRUCTURE results (Fig. 1). The  $K:2$  plot distinguishes Asian from all other population groups. The best differentiation was observed for  $K:3$  with the three main continental population groups (Africa, Europe and East Asia) clustered together, while Middle East and Central South Asia clustered together with Europe. Turkish, Turkish Cypriot and Azerbaijani samples clustered with the European group including Central South Asia and Middle East. The Middle East, Europe and Central South Asian cluster reveal the admixed character of these populations as the reasonable stopping point at  $K:4$ .

Reference HGDP-CEPH diversity panel genetic data from the five population groups (AFR, EUR, ME, CSA and EAS) was used to estimate individual ancestry assignment of each individual from the studied populations. Consistent with the results from the STRUCTURE, the Bayesian ancestry assignments for the tested individuals were assigned as European, Middle Eastern or Central South Asian (*data not shown*).

#### 4. Discussion

Results indicated that the 46 InDel set, though highly efficient in inferring the ancestry of individuals from Africa, Europe and East Asia, did not reveal distinct genetic clusters for Middle Eastern and Central South Asian populations. Populations from Turkey, Cyprus and Azerbaijan all show European, Middle Eastern and Central South Asian ancestry components. Our results strongly corroborate previous studies [1,2].

#### 5. Conclusion

The 46 AIM-Indel markers originally selected to show differences between the major groups African, European, East Asian, Native American and Oceanian [1,2]. Eurasian sub-population groups will be more optimally differentiated by combining the InDels used here with the extended sets of InDels focused on variation within Eurasia as well as with the addition of panels of ancestry-informative SNPs chosen for the same purpose.

This study provided new data for a forensic InDel database based on the SPSmart frequency browser framework: *forInDel* (*Forensic Indel browser*).

#### Conflict of interest

None.

#### Acknowledgments

This research was supported by Istanbul University Scientific Research Projects Unit (BAP) under International Research Projects (IRP) number 40991. CS is supported by funding awarded by the Portuguese Foundation for Science and Technology (FCT) and co-financed by the European Social Fund (Human Potential Thematic Operational Program SFRH/BD/75627/2010). We sincerely thank all the sample contributors. The authors would like to thank Department of Criminalistics Investigations DNA Laboratory, the

Ministry of Internal Affairs of the Azerbaijan Republic for providing DNA samples.

## References

- [1] R. Pereira, C. Phillips, N. Pinto, et al., Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing, *PLoS one* 7 (2012) e29684.
- [2] C. Santos, C. Phillips, F. Oldoni, et al., Completion of a worldwide reference panel of samples for an ancestry informative Indel assay, *Forensic Sci. Int. Genet.* 17 (2015) 75–80.
- [3] J. Amigo, A. Salas, C. Phillips, et al., SPSmart: adapting population based SNP genotype databases for fast and comprehensive web access, *BMC Bioinformatics* 9 (2008) 428.
- [4] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, *Mol. Ecol. Resour.* 10 (2010) 564–567.
- [5] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, *Genetics* 155 (2000) 945–959.
- [6] N.M. Kopelman, J. Mayzel, M. Jakobsson, et al., Clumpak: a program for identifying clustering modes and packaging population structure inferences across K, *Mol. Ecol. Resour.* (2015) .