Forensic Science International: Genetics 7 (2013) 359–366

Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

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

Eurasiaplex: A forensic SNP assay for differentiating European and South Asian ancestries

C. Phillips [a,*], A. Freire Aradas [a,1], A.K. Kriegel [b,1], M. Fondevila [a], O. Bulbul [c], C. Santos [a], F. Serrulla Rech [d], M.D. Perez Carceles [e], Á. Carracedo [a], P.M. Schneider [a], M.V. Lareu [a]

[a] Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Santiago de Compostela, Galicia, Spain
[b] Institute of Legal Medicine, Medical Faculty, University of Cologne, Cologne, Germany
[c] Institute of Forensic Science, Istanbul University, Istanbul, Turkey
[d] Forensic Anthropology Unit, Institute of Legal Medicine of Galicia, Spain
[e] Department of Legal Medicine, University of Murcia, Murcia, Spain

ARTICLE INFO

Article history:
Received 7 August 2012
Received in revised form 17 February 2013
Accepted 23 February 2013

Keywords:
SNPs
Eurasia
Central-South Asia
East Asia
Ancestry inference
HGDP-CEPH

ABSTRACT

We have selected a set of single nucleotide polymorphisms (SNPs) with the specific aim of differentiating European and South Asian ancestries. The SNPs were combined into a 23-plex SNaPshot primer extension assay: *Eurasiaplex*, designed to complement an existing 34-plex forensic ancestry test with both marker sets occupying well-spaced genomic positions, enabling their combination as single profile submissions to the Bayesian *Snipper* forensic ancestry inference system. We analyzed the ability of *Eurasiaplex* plus 34plex SNPs to assign ancestry to a total 1648 profiles from 16 European, 7 Middle East, 13 Central-South Asian and 21 East Asian populations. Ancestry assignment likelihoods were estimated from *Snipper* using training sets of five-group data (three Eurasian groups, East Asian and African genotypes) and four-group data (Middle East genotypes removed). Five-group differentiations gave assignment success of 91% for NW European populations, 72% for Middle East populations and 39% for Central-South Asian populations, indicating Middle East individuals are not reliably differentiated from either Europeans or Central-South Asians. Four-group differentiations provided markedly improved assignment success rates of 97% for most continental Europeans tested (excluding Turkish and Adygei at the far eastern edge of Europe) and 95% for Central-South Asians, despite applying a probability threshold for the highest likelihood ratio above '100 times more likely'. As part of the assessment of the sensitivity of *Eurasiaplex* to analyze challenging forensic material we detail *Eurasiaplex* and 34-plex SNP typing to infer ancestry of a cranium recovered from the sea, achieving 82% SNP genotype completeness. Therefore, *Eurasiaplex* provides an informative and forensically robust approach to the differentiation of European and South Asian ancestries amongst Eurasian populations.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Indian subcontinent populations, often termed South Asian, occupy a well-defined geographic region south of the Himalayas but are normally grouped as Eurasians with European and Middle Eastern populations [1,2]. The complex history of South Asians is just beginning to be understood from high-density single nucleotide polymorphism (SNP) genotyping [3,4]. This data, combined with uni-parental variability, indicate a relatively long human occupation of the Indian sub-continent compared to East Asia and Europe. There is little or no detectable common ancestry between ancestral South Asians and these neighboring groups, to east and west, for tens of thousands of years [3–7]. Additionally, South Asian variation is overlaid by high levels of population substructure, particularly amongst Indian populations, due to endogamy and founder effects within castes: a long-standing system of socially imposed stratification [8–11]. The most in-depth SNP study of South Asia by Reich et al. [4] found strong evidence for two ancient, divergent founding populations. Reich's study successfully modeled development of present South Asian SNP variability indicating populations of mainland India have a close relationship with Ancestral North Indian (ANI) founding groups, detecting 39–71% ANI ancestry overall, with higher ANI proportions in upper castes and Indo-European linguistic groups. The other founding group reconstructed, Ancestral South Indian (ASI), may no longer be detectable as a major ancestry component in mainland populations, appearing confined to vestigial populations with little ANI ancestry, e.g. Andaman Islanders. Therefore most South Asian populations have higher pairwise within-group Fst

* Corresponding author. Tel.: +34 981 582 327; fax: +34 981 580 336.
E-mail address: c.phillips@mac.com (C. Phillips).
[1] These authors contributed equally to the study.

1872-4973/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.fsigen.2013.02.010

than Europeans or East Asians, while lack of divergence between most mainland South Asians and other Eurasians is largely due to inferred genetic similarity between ANI ancestral groups and Europeans, Central Asians (north of the Himalayas) and Middle Eastern populations. Although South Asians show a much longer occupation of their region than Europeans or Middle East populations, the overall pattern of autosomal SNP variability across Eurasia is an allele frequency cline running from NW Europe to SE South Asia, with a similar west-east clinal pattern discernible in studies of Y chromosome variation in Central Asia [12].

Against this backdrop of recently mapped Eurasian variability we sought to collect autosomal ancestry informative SNPs (AIM-SNPs) into one forensic multiplex capable of reliably differentiating South Asians from Europeans. Differentiation of Middle Eastern populations, while desirable, was considered unrealistic using small-scale SNP sets required by forensic analysis. Middle East populations show lower levels of divergence between Europe and South Asia and this extensive geographic area is positioned in the middle of the NW-SE Eurasia cline. To construct the final assay we shortlisted candidate SNPs showing maximum European-South Asian divergence from allele frequencies of the CEPH Human Genome Diversity Panel (HGDP-CEPH) study of 650,000 SNPs by Stanford University [2]. From ~60 candidates, 23 SNPs were combined in a primer extension assay we named *Eurasiaplex*. The *Eurasiaplex* SNP set is designed to complement an existing forensic 34-plex AIM-SNP assay [13,14]. We measured ancestry assignment success using all 57 markers, analyzing relevant populations from HGDP-CEPH, HapMap and 1000 Genomes, plus study populations from Europe, Middle East and India. Population sets were initially used inter-changeably as training sets and test sets to gauge routine classification performance using *Snipper* – a Bayesian online classifier previously developed for the 34-plex assay [13]. Finally, we report a forensic casework application of SNP-based ancestry inference where differentiation of East Asian, European, Middle Eastern and South Asian ancestries using 34-plex and *Eurasiaplex* SNPs contrasted with interpretations of morphometric data made to help identify a skull washed ashore in NW Spain.

2. Materials and methods

2.1. Population samples and SNP genotype databases accessed

SNP variability in Eurasian and East Asian populations was collected from four sources: (i) 650,000 SNPs genotyped in 51 HGDP-CEPH populations by Stanford University [2], (ii) HapMap SNP genotypes for 1218 samples from 11 populations (3 Eurasian), (iii) 1000 Genomes SNP genotypes for 1093 samples from 14 populations (5 Eurasian), and (iv) our study populations comprising: 38 NW Spanish; 31 Germans; 44 Moroccans; 23 Indians from Murcia, Spain; 35 Vietnamese; 42 Greeks; 68 Turkish; 16 Afghans; 38 Iraqis and 8 Iranians. HGDP-CEPH genotypes for eight European, four Middle East/North African, nine Central-South Asian and seventeen East Asian populations, were accessed with SPSmart [http://spsmart.cesga.es]. HGDP-CEPH Central-South Asians comprise 8 Pakistani populations plus Uygur from extreme western China. SPSmart also accessed HapMap Gujarati Indians from Houston (GIH), though seventeen 34-plex SNPs are not genotyped in GIH.

2.2. SNP selection and multiplex construction

Sixty candidate SNPs were shortlisted showing highest divergence by estimating Rosenberg's ancestry informativeness metric: I_n [15] comparing HGDP-CEPH Central-South Asians vs. HGDP-CEPH Europeans (excluding Adygei, Sardinian outlier populations). From candidates 23 SNPs were selected with optimum multiplex performance. *Eurasiaplex* loci show good genomic separation from 34-plex SNPs (genomic map in Supplementary Fig. S1). We combined *Eurasiaplex* SNPs into single PCR and primer extension (EXT) multiplexes using SNaPshot (AB: Applied Biosystems, Foster City, CA, US) with ~4–5 bp spacing using pigtails. Primer sequences, reaction mix ratios and reaction conditions are given in Supplementary Table S2 and 34-plex reactions were as originally described [13]. Capillary electrophoresis preparation combined 1 μl purified extension products, 10 μl AB HiDi™ formamide and 0.25 μl AB LIZ-120. Separations used AB 3130xl detectors, POP-4™, 36 cm capillaries, then AB GeneMapper ID v. 3.2.

To test *Eurasiaplex* sensitivity fifteen typical challenging forensic casework samples and five highly degraded DNAs were analyzed in parallel to standard SEFiler and Identifiler STR genotyping.

2.3. Assessment of population divergence and ancestry assignment performance of Eurasiaplex and 34-plex SNPs

Geographically close-sited populations are difficult to assess for clear divergence patterns when they show very close relationships. We used principal component analysis (PCA) to reveal patterns of similarity and divergence amongst 60 populations, progressively decreasing complexity by step-wise reduction of data-points: removing Middle East data and colorizing outlier populations. PCA used R v. 2.11.1 [16] and the SNPassoc package [17]. HapMap GIH was treated as an un-admixed Indian population as they show no divergence from other Central-South Asian populations. Outlying PCA positions corresponded to geographically or genetically distinct populations. These included: Tuscan, Turkish, Sardinian and Adygei Europeans (the latter two found to be outliers previously [13]), Hazara and Kalash Central-South Asians (Kalash recognized as atypical in [1]), plus Yakut from northeast Siberia.

Ancestry assignment performance of the combined 57 SNPs was measured by constructing differently composed training sets for the *Snipper* online Bayesian classification system [13]. *Snipper* comprises an online training set/profile submission portal generating ancestry likelihoods for single SNP profiles. User-defined SNP sets up to 400 loci plus corresponding reference data (maximum 2000 individuals, 20 populations) can be uploaded as customized training sets to analyze forensic samples of unknown ancestry. This system largely parallels the widely used Structure algorithm, but *Snipper* benefits from an ability to classify single profiles in real time, typifying forensic analyses. We assessed Eurasian ancestry comparisons with two training sets: five group classifications using European, Middle East, Central-South Asian, African and East Asian data, and; four-group classifications excluding Middle East assignment. We chose a central Middle East training set of HGDP-CEPH Druze, Palestinian, Bedouin populations. Training sets and test sets were largely interchangeable, e.g. HGDP-CEPH Pakistani training sets gave similar likelihoods for Indians (Murcia), as Indian training sets for Pakistanis. A North African training set analyzing the other Middle East populations gave the greatest contrast in assignment probabilities and success compared to HGDP-CEPH Druze-Palestinian-Bedouin testing North Africans. Therefore all Middle East classifications used the HGDP-CEPH Druze-Palestinian-Bedouin training set. Other training sets selected gave consistently high assignment probabilities in the widest number of test individuals, consisting of: 1000 Genomes-YRI Africans, 1000 Genomes-CEU Europeans; 1000 Genomes-CHB East Asians and Central-South Asian Indians of Murcia. Five- and four-group training sets used are provided in Supplementary Data Files S3A/B.

A key factor when using *Snipper* is proper assessment of assignment likelihoods. We observed amongst study samples that

Download English Version:

https://daneshyari.com/en/article/98797

Download Persian Version:

https://daneshyari.com/article/98797

Daneshyari.com