SNPs associated with physical traits: A valuable tool for the inference of biogeographical ancestry

Runa Daniel[1]*, Juan J Sanchez[2], Najah T Nassif[1], Alexis Hernandez[2], Simon J Walsh[3]

[1]*Centre for Forensic Science, University of Technology Sydney, NSW, Australia*
[2]*National Institute of Toxicology and Forensic Services, Tenerife, Spain*
[3]*Forensic and Technical Services, Australian Federal Police, ACT, Australia*

Abstract
Sixteen autosomal SNPs were selected to differentiate major populations in Australia and a multiplex PCR assay, based on the selected SNPs, was developed for the inference of biogeographical ancestry.
Keywords: SNPs; biogeographical ancestry; physical traits; inference tools

1. Introduction

The ability to infer an individual's biogeographical ancestry can be instrumental in narrowing a pool of suspects in the initial stages of a criminal investigation. Such an inference tool can provide intelligence to forensic investigators when STR profiling has been unsuccessful, or when the STR profile does not match a database profile and when no eyewitnesses are available.

SNPs are reliable indicators of biogeographical ancestry due to their low mutation rate. A multiplex assay for the inference of biogeographical ancestry has been developed and is currently being validated. Furthermore, preliminary assessment of a SNP associated with pigmentation has also been conducted.

2. Materials and Methods

Biogeographical ancestry over three generations, as well as eye and hair colour information was self-declared by participants. Collected samples (N=242) were divided into six groups; Asian (*N*=51), Sub-Continental Asian (*N*=33), Middle Eastern (*N*=25),

* Corresponding author. Tel: +61-417597493
E-mail address: runa.daniel@student.uts.edu.au

Caucasian (*N*=79), North African (*N*=20) and Sub-Saharan African (N=34). All participants' parents were from the same sub-population group as the participants.

Self-administered buccal swabs were collected, transferred onto FTA® card (Whatman, United Kingdom) and 2mm^2 FTA discs were used to optimize the multiplex assay.

Based on extensive literature surveys, 16 autosomal SNPs were selected to distinguish the major Australian populations. PCR and SNP primers were designed by Juan J Sanchez[2]. Primer extension reactions were achieved using the ABI Prism® SNaPshot™ Multiplex Reaction Kit (Applied Biosystems).

3. Results and Discussion

The θ estimates for each pair of populations is shown above the diagonal and the genetic distance based on those θ values (d=-ln(1- θ)) are shown below the diagonal (Table 1).

Table 1 – Pairwise genetic distance matrix for all populations

	Asian	Sub-Continental Asian	North African	Middle Eastern	Sub-Saharan African	Caucasian
Asian		0.0875	0.2844	0.1706	0.3699	0.2721
Sub-Continental Asian	0.0916		0.1580	0.0694	0.3108	0.1582
North African	0.3348	0.1719		0.1100	0.2218	0.1742
Middle Eastern	0.1870	0.0720	0.1165		0.3294	0.0522
Sub-Saharan African	0.4618	0.3723	0.2508	0.3996		0.3920
Caucasian	0.3176	0.1723	0.1915	0.0536	0.4976	

A neighbour-joining tree was constructed from θ estimates for the full set of populations using GDA software (1). A close genetic relationship exists between the Middle Eastern and Caucasian, Sub-Continental Asian and Middle Eastern and Sub-Continental Asian and Asian populations as indicated by genetic distance values of 0.05, 0.07 and 0.09 respectively (Figure 1). A considerable degree of diversity is observed between the Caucasian and Asian populations, which are two populations of principal importance in the Australian context. Pairwise θ estimates in the order of 27% (Asian-Caucasian) are akin to those observed at non-autosomal markers (2).

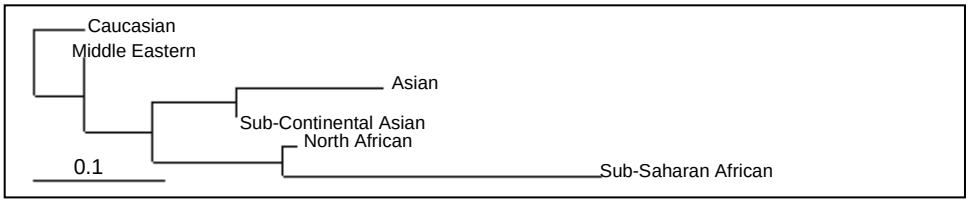

Figure 1 – Neighbour joining tree for all populations (*N*pops=6, *N*=242)

Population assignment was achieved using log-likelihood estimates calculated by GenAlEx software (3) following the method of Paetkau *et al* (4). In this analysis, an individual sample is removed from the dataset before it is assigned to a population. A high degree of accuracy was achieved for the inference of Asian (90% accuracy), Caucasian (91%), North African (90%) and Sub-Saharan African (100%) populations using these 16 SNPs. An intermediate degree of accuracy was observed for the Middle Eastern and Sub-Continental Asian populations with 60 and 79% accuracy respectively. Similar results for population assignment were obtained using the Snipper web portal (5) as shown in Table 2. These results indicate the alternate populations to which samples are being incorrectly assigned. For example, 83% of Sub-Continental Asian samples are assigned to the correct population, however, 7% of the samples are incorrectly assigned to the Asian and Middle Eastern populations. Similarly, 19% of Middle Eastern samples are incorrectly assigned to the Caucasian population and 10% to the Sub-Continental population. Due to the high level of interrelatedness, the Sub-Continental Asian and Middle Eastern populations are difficult to differentiate. Additional population specific SNPs will enable more effective separation between these populations.

		Assigned population				
Population of origin	Asian	Sub-Continental Asian	Middle Eastern	Caucasian	North African	Sub-Saharan African
Asian	93.5	4.35	2.17	0	0	0
Sub-Continental Asian	6.9	82.76	6.9	3.45	0	0
Middle Eastern	0	9.52	71.43	19.05	0	0
Caucasian	0	0	6.76	91.89	1.35	0
North African	0	0	0	0	94.12	5.88
Sub-Saharan African	0	0	0	0	0	100.00

Table 2 - Population assignment using log-likelihood estimates (Snipper)

One of the 16 SNPs used is Leu374Phe located in the *MATP* (*AIM-1*, *SLC45A2*) gene which is a major determinant of pigmentation. The Leu374 SNP has exhibited associations with dark hair, eye and skin colour in Europeans (6). As a crude preliminary assessment, 11 hair and eye colour combinations from 187 samples were divided into three broad groups; dark, medium and light. The GenAlEx assignment technique was used, as described previously and results are shown in table 3 below.

Phenotype	Correctly assigned	Incorrectly assigned	Accuracy (%)
Dark	94	9	91.3
Medium	0	31	0
Light	52	1	98.1

These results indicate that the Leu374Phe SNP may be useful for phenotypic inferences. Given the polygenic nature of pigmentation, this SNP would be used in combination with other pigmentation-associated SNPs to make reliable phenotypic inferences.

The described SNP multiplex assay has been developed for the inference of biogeographical ancestry. Additional population specific SNPs will be investigated to further separate the populations of interest and to facilitate reliable inferences of samples

with mixed ancestry. The ability to infer biogeographical ancestry, as well as phenotype, will provide valuable intelligence to forensic investigators.

References

1. Lewis, P.O. and Zaykin, D.. Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). Free program distributed by the authors over the internet from http://lewis.eeb.uconn.edu/lewishome/software.html.
2. Jorde, L.B., et al., American Journal of Human Genetics, 2000. **66**: 979-88.
3. GenAlEx ??
4. Paetkau, D., Slade, R. et al. Molecular Ecology, 2004. **13:** 55-65.
5. Snipper??
6.Graf et al, 2005

Acknowledgements
This project is funded by the Centre for Forensic Science, University of Technology Sydney and the Australian Federal Police. We wish to thank Dr. Chris Phillips for his advice and assistance.