International Congress Series 1261 (2004) 233–235

www.ics-elsevier.com

Population specific single nucleotide polymorphisms

C. Phillips[a,*], M. Lareu[a], A. Salas[a], M. Fondevila[a], G. Berniell Lee[a], A. Carracedo[a], N. Morling[b], P. Schneider[b], D. Syndercombe Court[b]

[a] *Institute of Legal Medicine, University of Santiago de Compostela, Calle San Francisco S/N, Galicia, 15705 Santiago de Compostela, Spain*
[b] *The SNP for ID Consortium, Spain*

Abstract. A search has been made of the Celera SNP database for markers exhibiting population specific polymorphism, where variation is only shown by one of the two populations used for validation. Using the SNPs found, assays are being developed for population-of-origin tests.

Keywords: Population specific; SNP; African; European; SNaPshot; Forensic analysis

1. Introduction

Using polymorphic markers to ascertain the population-of-origin of an unknown sample has long been an area of interest in forensic analysis. Unfortunately, most markers have alleles which are shared between each population, or have population specific alleles at frequencies too low to be useful. SNPs display more differentiated variation between populations owing to the age and stability of nucleotide substitutions. This study aimed to find population specific SNPs, validated to a sufficiently high level, and to develop genotyping assays for these markers. SNP loci exhibiting variation exclusively in one of two population groups (described as African-Americans and Caucasians) were collected from the Celera CDS database comprising 4.8 million SNPs [1]. All the selected markers had allele frequencies estimated to two decimal places using 45 individuals from each population. Initial genotyping of a small set of loci indicates the expected population linked polymorphism in each case.

2. Materials and methods

Suitable autosomal SNPs were selected from the Celera human RefSNP database. For African linked SNPs, allele frequency was set to 0.4–0.5 and combined using intersection [AND] with Caucasians set to 0. This was reversed for European linked SNPs and

* Corresponding author. Tel.: +34-981582327; fax: +34-981580336.
E-mail address: chrisp@usc.es (C. Phillips).

0531-5131/
doi:10.1016/j.ics.2003.12.041

extended, in both populations, to lower allele frequency ranges. An additional subset of nonspecific SNPs (exhibiting shared alleles with minimal variability in Africans) was found using the following frequency ranges: Caucasian, 0.4–0.5; African, 0–0.05. Genotyping was performed on 50 Galician (N.W. Spain) and 50 Mozambican samples. PCR primers were designed to give products in the size range 70–110 bp. Amplification and mini-sequencing reactions were as previously described (in this volume, in: "Non-Binary Single Nucleotide Polymorphisms").

3. Results and discussion

A total of 4712 SNPs were found with complete population linked variation. Four African linked SNPs taken from the above set have been genotyped and allele frequency estimates indicate complete population linkage in each case. Eighty-two SNPs exhibited polymorphism with minor allele frequencies 0.4–0.5 (average Dp 62.2%): of these, 81 were African linked and 1 European linked. A further 219 exhibited polymorphism with minor allele frequencies 0.3–0.4 (average Dp 60.2%); of these, 218 were African linked and 1 European linked. Such figures reflect the differing demographic histories and levels of variability that characterize each population. To increase the pool of SNPs with European linked polymorphism, markers were collected showing maximum variability in Europeans (0.4–0.5 minor allele frequency) and minimum variability in Africans (0.01–0.05 minor allele frequency). Including these markers in an assay can provide a

Table 1
Allele frequency distributions and information content of 10 European linked SNP loci

CDS no.	Population	Minor allele	Major Allele	% Informative	% Non-informative	Cumulative non-informative	Cumulative individuals with same genotypes
hCV1140336	Eur.	0.42	0.58	66.36	33.64	33.6400	1 in 3
	Afr.	0	1				
hCV16237183	Eur.	0.33	0.67	55.11	44.89	15.1010	1 in 7
	Afr.	0	1				
hCV2177452	Eur.	0.28	0.72	48.16	51.84	7.8284	1 in 13
	Afr.	0	1				
hCV11250782	Eur.	0.28	0.72	48.16	51.84	4.0582	1 in 25
	Afr.	0	1				
hCV3135019	Eur.	0.28	0.72	48.16	51.84	2.1038	1 in 48
	Afr.	0	1				
hCV151291	Eur.	0.27	0.73	46.71	53.29	1.1211	1 in 89
	Afr.	0	1				
hCV11857041	Eur.	0.27	0.73	46.71	53.29	0.5974	1 in 167
	Afr.	0	1				
hCV2677956	Eur.	0.27	0.73	46.71	53.29	0.3184	1 in 314
	Afr.	0	1				
hCV2039416	Eur.	0.26	0.74	45.24	54.76	0.1743	1 in 574
	Afr.	0	1				
hCV1048567	Eur.	0.25	0.75	43.75	56.25	0.0981	1 in 1020
	Afr.	0	1				

Table 2
Allele frequency distributions and information content of 13 African linked SNP loci

CDS no.	Population	Minor allele	Major Allele	% Informative	% Non-informative	Cumulative non-informative	Cumulative individuals with same genotypes
hCV3095460	Afr.	0.5	0.5	75	25	25.000000	1 in 4
	Eur.	0	1				
hCV1969450	Afr.	0.5	0.5	75	25	6.250000	1 in 16
	Eur.	0	1				
hCV971171	Afr.	0.5	0.5	75	25	1.562500	1 in 64
	Eur.	0	1				
hCV958204	Afr.	0.5	0.5	75	25	0.390625	1 in 256
	Eur.	0	1				
hCV608420	Afr.	0.5	0.5	75	25	0.097656	1 in 1024
	Eur.	0	1				
hCV608417	Afr.	0.5	0.5	75	25	0.024414	1 in 4096
	Eur.	0	1				
hCV2460535	Afr.	0.5	0.5	75	25	0.006104	1 in 16,384
	Eur.	0	1				
hCV11377586	Afr.	0.49	0.51	74	26	0.001588	1 in 62,991
	Eur.	0	1				
hCV15905778	Afr.	0.49	0.51	74	26	0.000413	1 in 242,181
	Eur.	0	1				
hCV12003211	Afr.	0.49	0.51	74	26	0.000107	1 in 931,105
	Eur.	0	1				
hCV3111622	Afr.	0.49	0.51	74	26	0.000028	1 in 3.6 million
	Eur.	0	1				
hCV2960082	Afr.	0.49	0.51	74	26	0.000007	1 in 13.8 million
	Eur.	0	1				
hCV3270591	Afr.	0.49	0.51	74	26	0.000002	1 in 53 million
	Eur.	0	1				
	Afr.	0	1				

likelihood ratio indicating population of origin in a similar way to STR loci [2]. Tables 1 and 2 list the most informative loci that could form a population of origin test: the best African linked and the best European linked. The predictive power of these loci in combination is given by the cumulative ''non-informative'' result probability: the probability of not finding the population specific allele in an individual originating from that population. Clearly, combining $\sim$ 8 SNPs from each group could provide a powerful predictive test for distinguishing Africans and Europeans.

References

[1] http://www.celeradiscoverysystem.com/.
[2] D. Syndercombe Court, et al., Advances in Forensic Genetics 9 (2003) 67–69.