Forensic Science International: Genetics 20 (2016) 89–100

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

EurEAs_Gplex—A new SNaPshot assay for continental population discrimination and gender identification

P. Daca-Roszak[a], A. Pfeifer[b,c], J. Żebracka-Gala[b], B. Jarząb[b], M. Witt[a,d], E. Ziętkiewicz[a,*]

[a] Institute of Human Genetics, Polish Academy of Sciences, Strzeszynska 32, 60-479 Poznan, Poland
[b] Maria Sklodowska-Curie Memorial Cancer Centre and Institute of Oncology, Gliwice Branch, Poland
[c] Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology, Gliwice, Poland
[d] International Institute of Molecular and Cell Biology, Warsaw, Poland

ARTICLE INFO

Article history:
Received 1 July 2015
Accepted 15 October 2015
Available online 19 October 2015

Keywords:
Biogeographic origin
Forensic test
Multiplex SNP genotyping
Low-quantity DNA template
Allele drop-out/drop-in
Population SNaPshot

ABSTRACT

Assays that allow analysis of the biogeographic origin of biological samples in a standard forensic laboratory have to target a small number of highly differentiating markers. Such markers should be easy to multiplex and the assay must perform well in the degraded and scarce biological material. SNPs localized in the genome regions, which in the past were subjected to differential selective pressure in various populations, are the most widely used markers in the studies of biogeographic affiliation. SNPs reflecting biogeographic differences not related to any phenotypic traits are not sufficiently explored.

The goal of our study was to identify a small set of SNPs not related to any known pigmentation/phenotype-specific genes, which would allow efficient discrimination between populations of Europe and East Asia. The selection of SNPs was based on the comparative analysis of representative European and Chinese/Japanese samples (B-lymphocyte cell lines), genotyped using the Infinium HumanOmniExpressExome microarray (Illumina). The classifier, consisting of 24 unlinked SNPs (*24-SNP classifier*), was selected. The performance of a 14-SNP subset of this classifier (*14-SNP subclassifier*) was tested using genotype data from several populations. The *14-SNP subclassifier* differentiated East Asians, Europeans and Africans with ~100% accuracy; Palestinians, representative of the Middle East, clustered with Europeans, while Amerindians and Pakistani were placed between East Asian and European populations.

Based on these results, we have developed a SNaPshot assay (*EurEAs_Gplex*) for genotyping SNPs from the *14-SNP subclassifier*, combined with an additional marker for gender identification. Forensic utility of the *EurEAs_Gplex* was verified using degraded and low quantity DNA samples. The performance of the *EurEAs_Gplex* was satisfactory when using degraded DNA; tests using low quantity DNA samples revealed a previously not described source of genotyping errors, potentially important for any SNaPshot-based assays.

1. Introduction

Determination of the biogeographic origin of human samples has recently become an important part of forensic science investigation, mostly due to the increasing threat of world terrorism. Important factors, which frequently impede application of DNA genotyping in such settings, include the scarcity of genetic material in available traces of biological samples, and its degradation. In the recent years, there is a trend for development of assays that allow analysis of population affiliation to be performed by a standard-equipped forensic laboratory. Such assays have to target a small number of highly differentiating markers, which are easy to multiplex and perform well in the degraded biological material.

Single nucleotide polymorphisms (SNPs) are easy to multiplex and their amplification requires only short DNA templates (<150 bp), which significantly increases the chance of successful genotyping of a degraded DNA [1,2]. SNP markers characterized by significant differences in allele frequencies between populations are the best ancestry informative markers (AIMs) and indicators of population affiliation [3–5]. However, a single SNP is not sufficient to achieve the required discrimination power. Identification of a limited set of SNPs markers that would allow efficient discrimination between populations is a challenging task. The majority of

Abbreviations: SBE, single base extension; AIM, ancestry informative marker.

* Corresponding author. Fax: +48 618233235.
E-mail address: zietkiee@man.poznan.pl (E. Ziętkiewicz).

http://dx.doi.org/10.1016/j.fsigen.2015.10.004
1872-4973/

genetic variance is shared by all major human groups, reflecting the relatively young evolutionary age of our species (e.g. [6]). African population is the only exception, since in addition to the common part of the genetic variance, it also possesses African-specific polymorphisms [7,8]. The variance among non-African populations is mostly due to the differences in allele frequencies of common, world-wide shared, polymorphic loci.

For the construction of population-specific tests, it is essential to know whether populations to be compared share a very recent common origin or if they represent independent lineages; this distinction is usually reflected by geographic distances separating these populations (e.g. intra- versus inter-continental). While discrimination of populations inhabiting neighboring geographic regions requires a large number of SNPs to be used in order to achieve the desired discrimination power (e.g. [9–11]), a number of studies have demonstrated that a small set of SNPs is sufficient for the discrimination of continental populations (e.g. Europeans vs. Africans [12–14]). Usually, such panels (classifiers) are implemented in a form of SNaPshot minisequencing assays (Life Technologies), which are based on the cost-effective multiplexed single base extension (SBE) technology [15,16]. The majority of published SNaPshot assays used for the biogeographic identification consist of SNPs selected from the human genome databases (e.g. HapMap, 1000 Genomes, dbSNP, Ensembl, Alfred) and/or literature data (e.g. [17–21]). The most widely used SNPs are localized in the genome regions, which in the past were subjected to differential selective pressure in various populations, for example SNPs linked with the genes involved in the metabolism of xenobiotics, immune response, fertility or pigmentation [22,23]. At the same time, there are many SNPs, which have the characteristics of AIMs, but are not related to the genes subjected to selection. Their differential allele frequencies reflect past demographic processes (such as migrations or population size fluctuations) that have enhanced differentiating effects of the genetic drift in separated populations. Such SNPs, allowing biogeographic discrimination not related to any phenotypic traits, are not sufficiently employed in the design of the existing AIM assays.

The goal of our study was to identify a possibly small set of SNP markers, not related to any known phenotype-specific genes, which would allow efficient discrimination of two populations: European and East Asian. The selection of SNPs was based on the comparative analysis of representative European and Chinese/Japanese samples (B-lymphocyte cell lines), genotyped using the Infinium HumanOmniExpressExome microarray (Illumina). The classifier, consisting of 24 unlinked SNPs (*24-SNP classifier*), was selected. The performance of the 14-SNP subset of this classifier (*14-SNP subclassifier*) was analyzed using a number of populations. The SNaPshot assay (*EurEAs_Gplex*) for genotyping 14 SNPs and allowing gender identification was developed. To verify forensic utility of the *EurEAs_Gplex* assay, a series of tests using low quantity and/or degraded DNA were performed.

2. Material and methods

2.1. DNA samples

DNA samples were isolated either from unrelated human male B-lymphocyte cell lines (Coriell Cell Repositories), or from the samples of peripheral blood; some DNA samples were obtained as a gift from the existing DNA collections. Thirty two DNA samples from unrelated cell lines representing European and Chinese populations were used in microarrays analysis, which led to the selection of differentiating SNPs. An independent set of unrelated DNA samples representing several world populations was used for testing the SNaPshot assay and genotyping selected SNPs (for details see Table 1). Genotypes of unrelated individuals representing various populations used to perform population statistics tests were extracted from public databases using the SPsmart tool, which allows accessing data from the HapMap and/or 1000 Genomes projects.

2.2. Microarray analysis

Genotyping of SNPs in 32 samples of European and Chinese origin was performed using the Infinium HumanOmniExpressExome

Table 1
DNA samples used in this study.

Continent	Population origin	Population id	Number of individuals	DNA source
Microarray analysis – SNP selection: $N = 32$				
Europe	European from Utah (Northern and Western ancestry)	CEU-0	16	B-lymphocyte cell lines from Coriell Cell Repository
East-Asia	Han Chinese from Beijing	CHB-0	16	B-lymphocyte cell lines from Coriell Cell Repository
SNaPshot testing* and unknown samples genotyping: $N = 211$				
Europe	European from Utah (Northern and Western ancestry)	CEU-1*	38	B-lymphocyte cell lines from Coriell Cell Repository
Europe	Polish	POL	42	Authors' collection (blood)
Middle-East	Palestinian	PAL	39	A gift from prof. Jaruzelska (blood)
East-Asia	Han Chinese from Beijing	CHB-1*	40	B-lymphocyte cell lines from Coriell Cell Repository
South-Asia	Pakistan	PAK	22	A gift from prof. Labuda (blood)
America	Ojibwa	OJI	30	A gift from prof. Labuda (blood)
The 1000 Genomes data for population statistics tests: $N = 742$				
Europe	European from Utah (Northern and Western ancestry)	CEU-2	85	NA
Europe	British from England and Scotland	GBR	88	NA
Europe	Finns from Finland	FIN	93	NA
Europe	Iberians from Spain	IBS	14	NA
Europe	Italians from Toscani	TSI	98	NA
East-Asia	Han Chinese from Beijing	CHB-2	70	NA
East-Asia	Southern China	CHS	100	NA
East-Asia	Japanese from Tokyo	JPT	94	NA
Africa	African American – Yoruba from Ibadan	YRI	28	NA
Africa	African American – Luhya from Webuye, Kenya	LWK	48	NA
Africa	African American – African Americans from Southwest US	ASW	24	NA

NA, not applicable.
* Samples used at the stage of SNaPshot testing.