Forensic Science International: Genetics 7 (2013) 305–312

Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

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

A 21 marker insertion deletion polymorphism panel to study biogeographic ancestry

Daniel Zaumsegel *, Markus A. Rothschild, Peter M. Schneider

Institute of Legal Medicine, Faculty of Medicine, University of Cologne, Melatengürtel 60/62, 50823 Köln, Germany

ARTICLE INFO

Article history:
Received 28 July 2012
Received in revised form 21 November 2012
Accepted 26 December 2012

Keywords:
DIP
InDel
Insertion–deletion polymorphism
AIM
Ancestry informative marker
Forensic genetics

ABSTRACT

Insertion/deletion polymorphisms have recently received increased interest in the forensic genetics community. This class of markers combines the advantageous genetic properties of single nucleotide polymorphisms (i.e., low mutation rate, genetic stability, and short amplicon size) with the technical advantage of short tandem repeat markers (simple detection by fluorescence-labelled PCR and capillary electrophoresis). For a large number of indel markers significant differences in allele frequencies between the major populations have been reported, making this class of markers suitable for the analysis of biogeographic ancestry. We have developed a multiplex PCR assay designed to establish the biogeographic ancestry of forensic DNA samples based on insertion/deletion polymorphisms. A panel of 21 short indels with allele frequency differences between three major population groups (European, African and Asian) was selected to be incorporated into a single-tube multiplex PCR assay. The assay is highly sensitive, requiring less than 0.5 ng of genomic DNA for successful typing. Due to the short fragment lengths below 200 bp, the assay is ideally suited for the typing of challenging forensic genetic case work samples. A population genetic study has been performed proving the performance of the assay in inferring the ancestral population of individuals. The chosen 21 markers are sufficient to distinguish between three major global population groups. Furthermore, the assay design leaves room for an extension in order to cover additional population groups.

1. Introduction

Forensic genetic research has largely focused on the use of genetic polymorphisms for the identification of individuals in criminal casework. The method of choice is typing of short tandem repeat polymorphisms (STRs), due to their high power of discrimination and the availability of allele frequency data for a large variety of populations [1,2].

Recently, different forensic applications have moved into focus, such as the prediction of the biogeographic ancestry of an unknown stain donor. For this application, however, STRs are less useful due to their mutational instability and resulting high intra-population variability compared to relatively low inter-population variability [3]. Single nucleotide polymorphisms (SNPs) have emerged as the marker class of choice for this task [4,5].

SNPs combine a variety of characteristics required for the use as ancestry informative markers (AIMs), such as low mutation rate, high density of distribution throughout the genome and a full range of allele frequency patterns across populations [3], as well as robustness for typing highly degraded DNA samples [6,7]. However, being sequence polymorphisms in contrast to fragment length polymorphisms like the commonly used forensic STR markers, the use of SNPs poses some technical difficulties for the routine forensic laboratory due to the lack of adequate SNP typing equipment [5].

Bridging the gap between these established forensic methods, insertion/deletion polymorphisms (indels) have received increased attention during the last ten years. Indels are abundant in the genome with at least one indel every 7.2 kb according to the landmark study of Mills et al. [8]. Since indels derive from a single mutation event occurring with a low frequency, they are genetically quite stable [9] and may show significantly different allele frequency distributions between distant populations making them ideal candidates for ancestry informative markers [10,11].

The discovery of a large number of short indels by Mills et al., further improved by the follow-up study from 2011 [12], makes these markers especially interesting for forensic applications, since short indels can be analysed in the routine forensic laboratory employing standard techniques such as PCR with fluorochrome-labelled primers and capillary electrophoresis, and offer considerable multiplexing capabilities as well as potential for incorporation into automated high-throughput genotyping systems. In addition, short indels can greatly improve amplification success even with

* Corresponding author. Tel.: +49 221 478 88335/88222; fax: +49 221 478 88370.
E-mail address: daniel.zaumsegel@uk-koeln.de (D. Zaumsegel).

1872-4973/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.fsigen.2012.12.007

Table 1
Markers selected for inclusion in the AIM indel assay, including chromosomal position, location in the genome, reported alleles and fluorochrome label.

rs number	Chromosome	Position (bp)	Alleles	Expected amplicon length	Fluorochrome
rs4646006	1	15,717,609	–/CTCA	61–65	6-FAM
rs140864	1	36,391,662	–/TTC	91–94	HEX
rs140858[a]	1	96,836,326	–/CT	86–88	HEX
rs2308026	4	119,404,855	–/CA	90–92	6-FAM
rs33948716	4	123,994,263	–/CCT	158–161	6-FAM
rs1610963	5	112,274,982	–/ATAACTAA	163–171	6-FAM
rs3834371	8	130,940,151	–/GAGT	108–112	HEX
rs140847	9	12,617,325	–/GCTT	152–156	HEX
rs35906376	11	36,007,532	–/AGGACT	114–120	HEX
rs2307666	11	64,486,500	–/GTTAC	97–102	6-FAM
rs33972805	11	126,288,872	–/CT	126–128	6-FAM
rs2308171	13	43,778,155	–/TCTG	132–136	HEX
rs2308036	15	65,207,011	–/CC	98–100	HEX
rs3069460	16	88,362,823	–/AGTACTG	70–77	6-FAM
rs16711	17	20,023,011	–/TTTCTTCCTA	164–174	HEX
rs5828358	19	53,833,222	–/CAGA	67–71	HEX
rs11471448	20	17,363,020	–/GCA	129–132	6-FAM
rs34785121	20	57,744,778	–/TGGA	136–140	6-FAM
rs6481	22	34,031,900	–/GTGGA	147–152	6-FAM
rs34123598	22	35,599,490	–/ATCT	116–120	6-FAM
rs3218285	22	35,866,670	–/CAACCAT	80–87	6-FAM
rs4253631	22	44,933,760	–/TTT	144–147	HEX

[a] Removed from final marker panel.

highly degraded DNA samples often encountered in forensic casework. Indel markers have already been used in a variety of studies ranging from the analysis of genetic structure in human populations [13–15], inferring biogeographic ancestry [11,16], assessing individual admixture [17] and identification of individuals [18].

The study described here aims at the development of a robust multiplexed PCR assay designed to predict the biogeographic ancestry of forensic casework DNA samples based on insertion/deletion polymorphisms. The PCR-based multiplex typing assay contains 21 short indels with allele length variations between 2 and 10 bp and sufficient allele frequency differences between three major population groups predominantly relevant for forensic casework in Central Europe (European, sub-Saharan African and Asian). Assay design was keyed towards flexibility by only using two fluorescent labels, leaving room for further extension of the assay. Short fragment lengths below 200 bp for each marker improve amplification success in degraded samples, and the design as a single-tube reaction with high sensitivity allows for the successful analysis of routine forensic samples with low DNA content.

2. Material and methods

2.1. Marker selection

An initial set of candidate markers was selected from the available online database of the United States National Center for Biotechnology Information, dbSNP [19] and the Marshfield Clinic diallelic insertion/deletion database [20] based on the following criteria: (i) biallelic autosomal indels, (ii) non-coding, but in close proximity to genes to take advantage of selection effects, (iii) allele length variation 2–10 bp, and (iv) a large allele frequency difference in one of the major population groups compared to the other two with the main focus on Europe, Africa, and Asia. Where markers are located on the same chromosome, care was taken to select these as far apart from each other as possible to avoid loss of information resulting from linkage between these markers. Flanking sequences of the selected indels were checked for sequence variants and repeat structures likely to interfere with primer design or to disrupt analysis. In total, 22 indels with allele length variations between 2 and 10 bp were selected from an initial candidate list of 60 markers for incorporation into the assay (detailed information in Table 1). However, one of the selected markers (rs140858) posed considerable technical difficulties in typing as observed during the data validation process, and was subsequently removed from the final panel.

2.2. Primer and assay design

Primer design was performed using the Primer3 Plus Web Interface [21,22] aiming for an amplicon size between 50 and 200 bp, an optimum T_m of 60 ± 2 °C and an optimum GC content of ≈50%.

During primer design, the markers were assigned for labelling with two fluorochromes (6-FAM and HEX) according to amplicon size with at least 4 bp gaps between neighbouring amplicons. For primer pairs presenting with strong −1 bp artefacts caused by incomplete adenylation of the PCR product, the reverse primer was extended by a 5′ tail of GTTTCTT to promote full adenylation [23]. Other primer pairs were extended with unspecific tails as closely related to the GTTTCTT tag sequence as possible where necessary for incorporation in the multiplex assay to ensure even spacing of the amplicons.

All primer pairs obtained were checked for unspecific binding using the Basic Local Alignment Search Tool of the NCBI [24,25] against the whole human genome. Primer pairs were also checked for hairpin and primer dimer formation using the AutoDimer software [26]. If not stated otherwise, all PCR amplifications were performed using the Qiagen Multiplex PCR kit (Qiagen, Hilden, Germany) in a total volume of 10 μl containing 5 μl of 2× Multiplex Master Mix, 1.5 μl of primer mix (details regarding the primer sequences and concentrations in the primer mix are available in Supplementary table 1) and 0.25–0.5 ng genomic DNA filled up to 10 μl with deionised water.

Thermocycling conditions were as follows: initial denaturation at 95 °C for 15 min, 28 cycles of 30 s at 95 °C, 90 s at 63 °C and 90 s at 72 °C followed by a final extension step of 60 min at 68 °C. Amplification products were purified by gel filtration using Sephadex™ G-50 (GE Healthcare, Munich, Germany) and subse-