Forensic Science International: Genetics 28 (2017) 138–145

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Evaluation of the Precision ID Ancestry Panel for crime case work: A SNP typing assay developed for typing of 165 ancestral informative markers

Vania Pereira[1,*], Helle S. Mogensen[1], Claus Børsting[1], Niels Morling[1]

Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, DK-2100 Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 11 November 2016
Received in revised form 20 February 2017
Accepted 23 February 2017
Available online 24 February 2017

Keywords:
Forensic genetics
Ancestry informative markers (AIMs)
Ancestry prediction
Massive parallel sequencing (MPS)
Ion Torrent PGM

ABSTRACT

The application of massive parallel sequencing (MPS) methodologies in forensic genetics is promising and it is gradually being implemented in forensic genetic case work. One of the major advantages of these technologies is that several traditional electrophoresis assays can be combined into one single MPS assay. This reduces both the amount of sample used and the time of the investigations.

This study assessed the utility of the Precision ID Ancestry Panel (Thermo Fisher Scientific, Waltham, USA) in forensic genetics. This assay was developed for the Ion Torrent PGM™ System and genotypes 165 ancestry informative SNPs. The performance of the assay and the accompanying software solution for ancestry inference was assessed by typing 142 Danes and 98 Somalis. Locus balance, heterozygote balance, and noise levels were calculated and future analysis criteria for crime case work were estimated. Overall, the Precision ID Ancestry Panel performed well, and only minor changes to the recommended protocol were implemented. Three out of the 165 loci (rs459920, rs7251928, and rs7722456) had consistently poor performance, mainly due to misalignment of homopolymeric stretches. We suggest that these loci should be excluded from the analyses.

The different statistical methods for reporting ancestry in forensic genetic case work are discussed.

1. Introduction

Autosomal Ancestry Informative Markers (AIMs) have become the main markers for the investigation of individual ancestry. The interest in AIMs in forensic genetics has increased during the last decade [1]. AIMs are genetic polymorphisms, mainly SNPs that exhibit large allele frequency divergences among major ethnic groups to better capture the genetic differences among them [2–4]. Besides tracing back individual genealogies, AIMs can have a prominent role in identification of missing persons and victims of mass disasters, as well as in association studies where AIMs may identify possible false associations due to substructure in case/control studies.

Forensic case work often deals with trace samples that have low amounts of DNA and/or that are very degraded. It is therefore pivotal to efficiently use these samples to retrieve as much genetic information as possible from each trace. Traditionally, samples are analysed by PCR and capillary-electrophoresis (CE) with simultaneous amplification of markers. However, there is a limit on how many markers it is feasible to investigate in one CE run, due to the size separation of the fragments and the number of fluorophores available.

Modern DNA sequencing technologies (massively parallel sequencing or MPS) have the potential to revolutionize the forensic genetic field. One of the major advantages of MPS is that most of the existing PCR-CE assays may be combined into one single MPS assay, which may reduce both the amount of sample used and the time of the investigations. Forensic genetic laboratories are currently assessing whether MPS methods can replace CE methods [5–17]. At the same time, kits specifically made for the forensic genetic community that combine all the DNA markers currently in use in forensic genetics into one single assay are being developed.

* Corresponding author at: Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Frederik V's Vej 11, DK-2100 Copenhagen, Denmark.
E-mail address: Vania.pereira@sund.ku.dk (V. Pereira).
[1] These authors contributed equally.

http://dx.doi.org/10.1016/j.fsigen.2017.02.013
1872-4973/

In 2014, Thermo Fisher Scientific (Waltham, USA) released two SNP typing assays for the Ion Torrent PGM™ System, one for human identification (Precision ID Identity Panel) and one for ancestry inference (Precision ID Ancestry Panel). These panels can type in a single run 124 and 165 SNPs, respectively, in a single run. Data are analysed with platform-specific software – the HID-SNP Genotyper plug-in (Thermo Fisher Scientific, Waltham, USA).

For the ancestry inference analysis, and besides the usual parameters of coverage or data quality, the software also predicts admixture and the population likelihood of DNA profiles. The admixture prediction analysis of the Precision ID Ancestry Panel estimates the overall proportion of ancestry of an individual based on seven continental populations. The population likelihood calculation is a more refined analysis that tries to identify the population of origin of an individual based on the allele/genotype frequencies from 65 populations.

The concept of likelihood is particularly relevant in forensic genetics. Ideally, the presentation of forensic evidence in court includes an estimate of the uncertainty of the evidence. The forensic community has recommended that the weight of the evidence should be assessed using the likelihood ratio principle that statistically compares the probabilities of an observation under two mutually exclusive scenarios or hypotheses. This study assessed the utility of the Precision ID Ancestry Panel in forensic genetics. The performance of the assay was evaluated by typing selected samples from Danes and Somalis.

2. Materials and methods

2.1. Samples, DNA extraction, and DNA quantification

Blood samples or buccal swabs collected on FTA cards (Whatman Inc., Clifton, NJ) from 142 self-declared Danes with parents and grandparents born in Denmark were used. FTA cards from 98 unrelated Somalis with self-declared Somali ancestry were selected from Somali individuals who wanted to immigrate to Denmark. The DNA from blood samples were extracted with the BioRobot EZ1 Workstation (Qiagen, Hilden, Germany) using the manufacturer's recommendations.

The concentration of DNA in all extracts was measured by the Qubit™ dsDNA High Sensitivity assay and the Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, California) following the manufacturer's recommendations.

The work was approved by the Danish ethical committee (H-1-2011-081).

2.2. Library preparation

DNA libraries were constructed using the Ion AmpliSeq™ Library Kit 2.0 (Thermo Fisher Scientific, Waltham, USA) and the Precision ID Ancestry Panel (Thermo Fisher Scientific, Waltham, USA). For DNA extracts, the input DNA was in the range of 500–1000 pg. For the buccal swabs, a 1.2 mm FTA card punch was washed for 30 min with 190 μl molecular grade water, dried for 60 min in a laminar flow bench, and used directly in the amplification mix [18].

The PCR contained 4 μl of 5x Ion AmpliSeq™ HiFi Mix, 10 μl Precision ID Ancestry Panel, and 6 μl of sample. For amplification of FTA card punches, 6 μl of water was added to reach a total volume of 20 μl. The PCR programme was initiated by simultaneous enzyme activation and denaturation at 99 °C for 2 min and 15 s, and an annealing/extension step at 60 °C for 4 min. For the population samples, 25 cycles were used. All PCRs were finalized with a hold at 10 °C.

For the libraries in the PCR cycle experiments (see below) and the first replica of the population samples, 2 μl of FuPa reagent (Thermo Fisher Scientific, Waltham, USA) was added. For the second replica of the population samples, 1.7 or 1.5 μl of FuPa reagent was added. The viscosity of the FuPa reagent makes precise pipetting difficult. The volume of FuPa reagent provided with the Ion AmpliSeq™ Library Kit 2.0 is often insufficient to digest the recommended number of samples. The reduction of the volume of FuPa reagent to 1.5 μl did not influence library concentration or the number of reads per sample (data not shown). The reactions were incubated at 50 °C for 10 min, 55 °C for 10 min, and 60 °C for 20 min. All libraries were barcoded using the Ion Xpress™ Barcode Adaptors (Thermo Fisher Scientific, Waltham, USA). A total of 4 μl Switch Solution, 2 μl barcode, and adapter mix, and 2 μl DNA Ligase (Thermo Fisher Scientific, Waltham, USA) was added to the FuPa digested PCR products, and incubated at 22 °C for 30 min and 72 °C for 10 min.

All libraries were purified using the Agencourt AMPure XP reagents (Beckman Coulter Inc., CA, USA) on a Biomek® 3000 Laboratory Automation Workstation (Beckman Coulter Inc., CA, USA) with an in-house customized script (available upon request). The purified libraries were quantified using the Agilent 2200 TapeStation system with the D1000 ScreenTape system (Agilent Technologies, Santa Clara, USA) or by using the Qubit™ dsDNA High Sensitivity assay protocol and the Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, California). After quantification, libraries were pooled in equimolar amounts.

2.3. Emulsion PCR, loading of sequencing chip, and sequencing

Emulsion PCR, enrichment of templated Ion Sphere™ Particles (ISPs), and loading of sequencing chips were performed using the Ion Chef™ instrument (Thermo Fisher Scientific, Waltham, USA) and the Ion PGM™ IC 200 Kit (Thermo Fisher Scientific, Waltham, USA) as described by the manufacturer. Sequencing was performed on the Ion PGM™ using the Ion PGM™ Sequencing 200 Kit v2 reagents (Thermo Fisher Scientific, Waltham, USA).

2.4. Data analysis

The primary sequence analyses of DAT-files were performed on the Torrent Suite Server (v4.2.1, v4.4.2 or v 4.6) (Thermo Fisher Scientific, Waltham, USA). Secondary sequence analysis was performed on the generated BAM-files using the HID-SNP Genotyper v4.2 plug-in (Thermo Fisher Scientific, Waltham, USA) with the corresponding target and hotspot BED-files to specify the location of the loci within the hg19 reference genome (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/). The default settings were used for analysis: minimum allele frequency = 0.1, minimum coverage = 6, minimum coverage on either strand = 0, and maximum strand bias = 1. The CSV-files generated by the plug-in were used for all further analyses.

The locus balance was calculated as the number of reads for a locus divided by the average number of reads per locus. The heterozygote balances were calculated as the number of reads for one nucleotide divided by the number of reads for the other nucleotide in the order A, C, G, and T. Noise was calculated as the number of reads that were different from the genotype divided by the total number of reads. Simple calculations and plots of kit performance parameters were done by MS™ Access and the statistical software R [19].

Allele frequencies for all markers were estimated by direct counting.

Deviations from Hardy–Weinberg expectations (HWE) were investigated with the Arlequin v.3.5 software [20] using 1,000,000 Markov chain steps.

Statistically significant association between pairs of loci was tested with Arlequin v3.5 using an exact test of linkage