Forensic Science International: Genetics 34 (2018) 186–196

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

Target capture enrichment of nuclear SNP markers for massively parallel sequencing of degraded and mixed samples

Nikhil Bose[a,b], Katie Carlberg[a,c], George Sensabaugh[b,d], Henry Erlich[a], Cassandra Calloway[a,b,*]

[a] Children's Hospital Oakland Research Institute, 5700 Martin Luther King Jr. Way, Oakland, CA 94609, United States
[b] University of California, Davis, Forensic Science Graduate Program, 1909 Galileo Ct. Ste. B, Davis CA 95618, United States
[c] UCSF Benioff Children's Hospital Oakland, Hematology/Oncology Department 747 52nd St, Oakland, CA 94609, United States
[d] University of California, Berkeley, School of Public Health, Berkeley, CA 94720, United States

ARTICLE INFO

Keywords:
Single Nucleotide Polymorphism (SNP)
DNA Probe Capture Target Enrichment
Massively Parallel Sequencing (MPS)
Next Generation Sequencing (NGS)
Degraded DNA

ABSTRACT

DNA from biological forensic samples can be highly fragmented and present in limited quantity. When DNA is highly fragmented, conventional PCR based Short Tandem Repeat (STR) analysis may fail as primer binding sites may not be present on a single template molecule. Single Nucleotide Polymorphisms (SNPs) can serve as an alternative type of genetic marker for analysis of degraded samples because the targeted variation is a single base. However, conventional PCR based SNP analysis methods still require intact primer binding sites for target amplification. Recently, probe capture methods for targeted enrichment have shown success in recovering degraded DNA as well as DNA from ancient bone samples using next-generation sequencing (NGS) technologies. The goal of this study was to design and test a probe capture assay targeting forensically relevant nuclear SNP markers for clonal and massively parallel sequencing (MPS) of degraded and limited DNA samples as well as mixtures. A set of 411 polymorphic markers totaling 451 nuclear SNPs (375 SNPs and 36 microhaplotype markers) was selected for the custom probe capture panel. The SNP markers were selected for a broad range of forensic applications including human individual identification, kinship, and lineage analysis as well as for mixture analysis. Performance of the custom SNP probe capture NGS assay was characterized by analyzing read depth and heterozygote allele balance across 15 samples at 25 ng input DNA. Performance thresholds were established based on read depth ≥500X and heterozygote allele balance within ± 10% deviation from 50:50, which was observed for 426 out of 451 SNPs. These 426 SNPs were analyzed in size selected samples (at ≤75 bp, ≤100 bp, ≤150 bp, ≤200 bp, and ≤250 bp) as well as mock degraded samples fragmented to an average of 150 bp. Samples selected for ≤75 bp exhibited 99-100% reportable SNPs across varied DNA amounts and as low as 0.5 ng. Mock degraded samples at 1 ng and 10 ng exhibited > 90% reportable SNPs. Finally, two-person male-male mixtures were tested at 10 ng in contributor varying ratios. Overall, 85-100% of alleles unique to the minor contributor were observed at all mixture ratios. Results from these studies using the SNP probe capture NGS system demonstrates proof of concept for application to forensically relevant degraded and mixed DNA samples.

1. Introduction

DNA from forensic casework samples such as postmortem tissues, bones, and hairs may sometimes be limited or highly degraded due to cellular and environmental processes [1,2]. DNA samples from mass disasters such as victims of war, commercial airline crashes, and terrorist attacks can be degraded as well as mixed [3–5]. Traditional STR analysis of degraded DNA is more suitable using miniSTR primer sets [6–9], however when DNA template sizes are less than 150 bp, miniSTRs have limited success due to allele imbalances and dropouts [10,11]. Consequently, when DNA is highly degraded, alternative markers such as mitochondrial DNA (mtDNA) and nuclear Single Nucleotide Polymorphisms (SNPs) can provide higher genotyping success [12–19]. More recently, advances in sequencing technologies allow for massively parallel sequencing of the entire mitochondrial genome, therefore further increasing the discriminatory power [20]. However, since mtDNA is inherited only maternally and without recombination, polymorphic sites are not statistically independent [21,22]. Therefore, the product rule cannot be applied for mtDNA analysis, leading to a lower discriminatory power compared to analysis of independent nuclear DNA genetic markers [23].

Nuclear SNP systems which offer a higher discriminatory power can

* Corresponding author at: Cassandra Calloway, Children's Hospital Oakland Research Institute, 5700 Martin Luther King Jr. Way, Oakland, CA 94609, USA.
E-mail address: scalloway@chori.org (C. Calloway).

https://doi.org/10.1016/j.fsigen.2018.01.010
Received 26 October 2017; Received in revised form 14 January 2018; Accepted 31 January 2018
Available online 04 February 2018
1872-4973/ © 2018 Elsevier B.V. All rights reserved.

Table 1
Custom panel of forensically significant SNP markers.

Type of SNP marker	Forensic significance	Currently available panels	# of SNPs	# of unique markers
Identity Informative	Identification	SNPs for a universal individual identification panel [64], Forensic validation of the SNPforID 52-plex assay [65], Developing a SNP panel for forensic identification of individuals [66]	136*	135
Ancestry Informative	Kinship	Progress toward an efficient panel of SNPs for ancestry inference [67]	41*	40
Phenotype Informative		The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA [68], Prediction of eye and skin color in diverse populations using seven SNPs [69]	24	24
X chromosome	Lineage, Mixture detection	Forensic usefulness of a 25X chromosome single nucleotide polymorphism marker set [59]	25	25
Y chromosome		New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree [60]	81	81
Tri-allelic	Identification, Mixture detection	Tri-allelic SNP markers enable analysis of mixed and degraded DNA samples [25]	31	31
Tetra-allelic		Tetra-allelic SNPs: Informative forensic markers compiled from public whole genome sequence data [26]	39	39
Microhaplotype		ALFRED Database [63], Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics [29]	76	36
Total Unique =			451*	411

* One Identity Informative SNP was also a microhaplotype SNP, and one ancestry informative SNP was also a phenotype informative SNP.

be used to analyze degraded DNA samples, since the targeted variation is only a single base compared to multiple bases in STRs [12]. However, since SNPs are predominantly bi-allelic, mixture detection and resolution can be challenging [24]. To improve mixture analysis, multi-allelic SNP markers, such as microhaplotypes, tri-allelic, and tetra-allelic SNPs can be analyzed [25–29]. Another challenge of bi-allelic SNP analysis is that minimally 50–60 high heterozygosity autosomal SNPs must be successfully genotyped to produce a profile as discriminatory as a profile from genotyping 13 core CODIS STR loci [30,31]. While microarray and primer extension assays are available for SNP analysis, the number of SNPs that can be analyzed in parallel is limited due to low multiplexing capabilities [32–34].

Next Generation Sequencing (NGS) technologies now offer a solution for massively parallel, high throughput sequencing of a large number of targets and simultaneous analysis of different types of genetic markers (including length and sequence polymorphisms). The high read depth per base and digital readouts of clonal sequences [35,36] further allow sensitive detection and quantitative resolution of minor variants in mixtures [37]. Recognizing the potential of NGS, several PCR based kits targeting SNPs have been developed for various NGS platforms. These kits target 120–180 nuclear SNPs for identity, kinship, lineage, and phenotype analysis [38–40]. However, PCR based systems require intact forward and reverse primer binding sites in the template DNA and, when applied to DNA samples that are degraded to sizes less than 100 bp result in high dropout rates [41].

Recent studies of a probe capture strategy for target enrichment have demonstrated success in recovering highly degraded DNA, such as ancient DNA, for massively parallel sequencing [42–44]. The probe capture strategy uses short biotinylated DNA or RNA probes to bind to targeted sequences in a DNA shotgun-library. The target sequence-probe complex is then retained using magnetic streptavidin beads while non-target DNA and excess unbound probes are washed away [45,46]. Therefore, highly degraded DNA targets can be enriched from DNA samples independent of fragment length and breakpoint. Previous probe capture NGS systems targeting exome sequences, including the whole human exome, have demonstrated improved enrichment of highly fragmented DNA (50–100 bp) in ancient DNA samples with as low as ~1% endogenous DNA [43,45,47–50]. Custom probe capture NGS systems have also demonstrated success in recovering the whole mitochondrial genome for massively parallel sequencing of highly degraded and ancient DNA samples [51–53]. Commercial DNA probe capture assays, such as the Roche NimbleGen SeqCap, Illumina TruSeq, Illumina Nextera, and Agilent SureSelect, have demonstrated higher sequence coverage of targets than conventional PCR based systems [54]. Furthermore, the NimbleGen SeqCap EZ DNA probe solution uses a high redundancy tiling strategy for capturing targeted bases, which enables efficient recovery of highly polymorphic regions [54–57], such as microhaplotypes and mtDNA [58].

Based on the demonstrated success of the DNA probe capture enrichment strategy employing the Roche NimbleGen SeqCap platform, we developed a solution-based, high redundancy DNA probe capture assay targeting 411 forensically relevant nuclear markers for massively parallel sequencing. The markers include 375 SNPs and 36 microhaplotypes (a total of 451 SNPs) for human identification, kinship, lineage assessment, and mixture detection. We characterized the performance of the custom probe assay based on high read depth and allele balance of targeted SNPs. We then demonstrated proof of concept for analysis of forensically relevant samples by testing the custom capture assay with fragmented and size selected samples, as well as two-person male-male mixtures at varied ratios.

2. Methods

2.1. Probe capture design

Based on a literature survey, panels of eight types of SNP markers were selected for their forensic significance. The compiled panel consisted of 136 identity, 41 ancestry, and 24 phenotype informative SNPs, 31 tri- and 39 tetra-allelic SNPs, 25 X and 81 Y SNPs, and 36 microhaplotype markers (Table 1). The X SNPs selected have a high degree of polymorphism (expressed as genetic diversity of 0.3–0.5) [59], and Y SNPs selected are clade defining or exhibit high genetic diversity (0.4–0.5) [60,61]. X and Y SNPs with high heterozygosity across and within populations are advantageous for identification purposes. The rs numbers (reference SNP ID numbers) and genomic co-ordinates of these SNPs were determined for the human genome version GRCh37/hg19 (Supplement A) using the UCSC Genome Browser (University of California, Santa Cruz), NCBI dbSNP (National Center for Biotechnology Information), and ALFRED (Allele Frequency Database, Yale University) [62,63]. A region of interest (ROI) was established for each probe consisting of the SNP co-ordinates ± 50 bp. The only exceptions to the ROI were the microhaplotypes, where the ROI covered the two to three SNPs of the microhaplotype ± 30 bases. The ROIs were formatted into a BED file and submitted for design using the SeqCap EZ Choice Library Design Submission software (Roche NimbleGen, Madison, WI). The nucleotide chosen at the SNP co-ordinate corresponds to all allelic variants of the nucleotide represented at that site corresponding to GRCh37/hg19 version of the human genome.

The probe design chosen for production covered 41,568 out of 45,372 bases of the ROIs (91.6%) using a high redundancy tiling

Download English Version:

https://daneshyari.com/en/article/6553279

Download Persian Version:

https://daneshyari.com/article/6553279

Daneshyari.com