Forensic Science International: Genetics 15 (2015) 8–15

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Original Research Paper

Massively parallel sequencing of complete mitochondrial genomes from hair shaft samples

Walther Parson [a,b,*], Gabriela Huber [a], Lilliana Moreno [c], Maria-Bernadette Madel [a], Michael D. Brandhagen [c], Simone Nagl [a], Catarina Xavier [a], Mayra Eduardoff [a], Thomas C. Callaghan [c], Jodi A. Irwin [c,*]

[a] Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria
[b] Penn State Eberly College of Science, University Park, PA, USA
[c] FBI Laboratory, Quantico, VA, USA

ARTICLE INFO

Keywords:
Next generation sequencing
mtDNA
mtGenomes
Heteroplasmy
Forensic science

ABSTRACT

Though shed hairs are one of the most commonly encountered evidence types, they are among the most limited in terms of DNA quantity and quality. As a result, DNA testing has historically focused on the recovery of just about 600 base pairs of the mitochondrial DNA control region. Here, we describe our success in recovering complete mitochondrial genome (mtGenome) data (~16,569 bp) from single shed hairs. By employing massively parallel sequencing (MPS), we demonstrate that particular hair samples yield DNA sufficient in quantity and quality to produce 2–3 kb mtGenome amplicons and that entire mtGenome data can be recovered from hair extracts even without PCR enrichment. Most importantly, we describe a small amplicon multiplex assay comprised of sixty-two primer sets that can be routinely applied to the compromised hair samples typically encountered in forensic casework. In all samples tested here, the MPS data recovered using any one of the three methods were consistent with the control Sanger sequence data developed from high quality known specimens. Given the recently demonstrated value of complete mtGenome data in terms of discrimination power among randomly sampled individuals, the possibility of recovering mtGenome data from the most compromised and limited evidentiary material is likely to vastly increase the utility of mtDNA testing for hair evidence.

1. Introduction

Mitochondrial DNA testing in forensic casework has proven extremely useful in those cases for which nuclear DNA is too scarce or degraded to yield forensically informative STR profiles. For particular sample types such as shed hairs – which by some estimates represent up to 90% of the hair samples collected at crime scenes [1] – mtDNA testing is often the only DNA-based testing option. Historically, mtDNA typing has focused on the 600 bp of hypervariable segments (HVS) I and II of the ~1100 bp mtDNA control region (CR). Due to the high substitution rates and the resulting inter-individual variability, the chance that any two randomly sampled individuals match in these regions is generally less than 1% (depending on the population [2]). Nevertheless, there are particular situations in which HVS-I/HVS-II or even CR data do not provide sufficient discriminatory information to resolve distinct maternal lineages. This is the case when common HVS-I/HVS-II and CR haplotypes are encountered (e.g. up to 7% and 4.4%, respectively, for U.S. Caucasians, empop.org [3]) or when questioned and known sequences differ by only a single base. Under these circumstances, examination of CR data alone restricts the power of mtDNA testing in general.

A number of studies over the past decade have demonstrated that by extending mtDNA analysis from the CR to the complete mtGenome, the power of mtDNA testing for identification purposes can be greatly improved. With complete mtGenome data, the most common U.S. Caucasian, African American and Hispanic HVS-I/HVS-II types can be reduced from, in turn, 7 to 2%, 3.6 to 1% and 2.3 to 0.5%, respectively [4,5]. In addition, mtDNA coding region SNPs have been useful for resolving missing persons cases in which more than one reference family share the same CR haplotype [5,6], the sorting and re-association of commingled skeletal remains [5], and increasing statistical support when exclusionary references are unavailable [7].

* Corresponding authors.
E-mail addresses: walther.parson@i-med.ac.at (W. Parson), jodi.irwin@ic.fbi.gov (J.A. Irwin).

http://dx.doi.org/10.1016/j.fsigen.2014.11.009
1872-4973/

More recently, efforts to develop reference population data are beginning to reveal the true value of complete mtGenome data for forensic purposes in terms of inter-individual variation among randomly sampled individuals [8,9]. In a recent study examining individuals from across the United States, HVS-I and HVS-II information provided resolution of 64–76% of the sampled haplotypes (depending on population) and CR data increased those values to 74–80%. However, complete mtGenome data permitted nearly full resolution in the three populations analyzed (94% in Hispanics, 98.5% in Caucasians and 98.8% in African Americans [9].

Despite these developments, coding region information is still not routinely employed in forensic casework. While this may be at least partially due to the policy issues that would need to be addressed before coding region information could be used for identification purposes, there have also been significant technical barriers to generating complete mtGenome sequences from the types of evidentiary material commonly analyzed for mtDNA. Because mtDNA data are primarily sought when the genetic material is severely limited and/or compromised, and because the acquisition of sequence data in most mtDNA casework (i.e. hair and calcified tissue specimens) is generally dependent on the recovery of numerous small amplicons, often in independent amplification reactions, there simply is not enough evidentiary material in most forensic cases to support amplification and Sanger sequencing of the entire mtGenome. Furthermore, the fact that such an approach would be cost and labor prohibitive with currently employed Sanger technologies has discouraged the development of mtGenome assays.

However, as predicted nearly five years ago [10], massively parallel sequencing technologies (MPS) are now finding their way into the forensic arena [8,11–20]. These technologies largely remove the greatest existing technical and cost barriers to mtGenome sequencing of low DNA quality and quantity specimens. In fact, MPS has already been paired with assays developed by the ancient DNA community [21,22] to develop complete mtGenome haplotypes from some of the most compromised unidentified human remains [23,24].

Here, we describe the development of entire mtGenome sequence data from single hair shafts and/or shed hairs lacking follicular tissue. We demonstrate that particular hair samples yield DNA sufficient in quantity and quality to produce larger mtGenome amplicons and that entire mtGenome data can be recovered from hair extracts even without PCR enrichment. Most importantly, we describe a small amplicon multiplex assay that can be routinely applied to the types of compromised hair samples typically encountered in forensic casework.

2. Materials and methods

2.1. Large fragment method

2.1.1. Sample preparation

To both demonstrate the general feasibility of recovering entire mtGenomes from hair samples via PCR enrichment, and explore the potential recovery of mtGenome data from hair extracts that have not undergone enrichment, single hair samples that included the root (but lacked visible tissue when observed under a stereomicroscope) were collected from two donors of west Eurasian ancestry for whom Sanger-based complete mtGenome sequence data were available. Three hairs in total were collected: Two dyed/treated head hairs representing one individual (Samples B1 and B2) and one un-dyed head hair representing the second individual (Sample A). Hair samples B1 and B2 had been stored at 4 °C for approximately 6 months prior to DNA extraction, while hair sample A had been stored at 4 °C for ~3.5 years. All samples were collected with fully informed consent. Approximately 4 cm of each hair's root end were removed, placed into a tube containing UV-irradiated 5% Terg-a-Zyme™ and placed into a sonicating water bath at room temperature for 20 min. Hairs were then briefly rinsed by sequentially transferring the fragments (using sterile forceps) from the 5% Terg-a-Zyme™ tube to a new tube containing UV-irradiated 100% ethanol, and then to a new tube containing UV-irradiated Molecular Biology Grade water. Each hair was extracted using a silica-based method adapted from a Qiagen® user-developed protocol [25,26]. The method employs Qiagen® digestion buffers (Qiagen, Germantown, MD) and the PrepFiler™ Forensic DNA Extraction Kit magnetic particles (ThermoFisher, Waltham, MA).

Extract mtDNA yields were assessed using a custom real-time quantitative PCR (qPCR) assay specific for a 105 bp region of the human mtGenome [27] or the Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA) prior to any further processing.

From these samples, extracts of samples A and B1 were selected for mtDNA enrichment via PCR amplification, while extract B2 was taken directly to library preparation with no intermediate enrichment step.

2.1.2. Large amplicon mtDNA enrichment

Hair extracts that underwent amplification enrichment were processed according to [28]. This mtGenome amplification strategy employs eight overlapping amplicons that range in size from 1 to 3.5 kbp, with the CR amplicon being the only product smaller than 2 kbp. Based on qPCR quantitative values of the two hair extracts, mtDNA inputs per 50 μl amplification reaction were estimated to be ~260,000 and ~130,000 mtGenome equivalents (mtGEs). For the HL60 positive control, total genomic DNA template input was 50 pg, or ~6000 mtGE based on mtDNA qPCR values.

In order to gain some insight into how experimental controls may need to be handled in a forensic MPS workflow, the extraction blanks and PCR negatives processed alongside the hair samples were also taken through the subsequent library preparation and sequencing steps.

2.1.3. Library preparation

Amplification products were quantified using a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). PCR amplicons were purified using Exo-SAP IT (Affymetrix, Santa Clara, CA). All samples, mDNA enriched or otherwise, were prepared for sequencing on the Illumina MiSeq instrument using the NexteraXT kit (Illumina, San Diego, CA).

The eight amplicons for each sample were normalized to 0.2 ng/μl and then pooled prior to library preparation. Library preparation and sample barcoding were performed according to manufacturer recommendations. The samples were multiplexed in the MiSeq run with a positive control, and the respective reagent blanks (RB) and negative amplification controls (NC). In order to maintain the integrity of the controls in terms of them representing the maximum possible contaminant concentration that may be present in an associated sample, the amplification products for the RBs and NCs were pooled based on the largest volume required for each DNA sample amplicon at the normalization/pooling stage. In other words, if 6 μl and 9 μl of a given amplicon from two different samples were required for downstream normalization and pooling, 9 μl (the max volume for that particular amplicon) were used for the pooled RB and NC library.

2.1.4. Shotgun genomic DNA sequencing

The single hair extract selected for shotgun sequencing was concentrated down to 5 μl for input into the Nextera XT library preparation kit. Based on Qubit quantification of the original 50 μl