Forensic Science International: Genetics 7 (2013) 632–639

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Reprint of: Evaluation of next generation mtGenome sequencing using the Ion Torrent Personal Genome Machine (PGM)☆

Walther Parson [a,b,*], Christina Strobl [a], Gabriela Huber [a], Bettina Zimmermann [a], Sibylle M. Gomes [c], Luis Souto [c], Liane Fendt [a,d], Rhena Delport [e], Reina Langit [f], Sharon Wootton [f], Robert Lagacé [f], Jodi Irwin [g]

[a] Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria
[b] Penn State Eberly College of Science, University Park, PA, USA
[c] Department of Biology, University of Aveiro, Campus de Santiago, Aveiro, Portugal
[d] Division of Human Genetics, Innsbruck Medical University, Innsbruck, Austria
[e] Department of Chemical Pathology, School of Medicine, University of Pretoria, South Africa
[f] Life Technologies, Foster City, CA, USA
[g] FBI Laboratory, Quantico, VA, USA

ARTICLE INFO

Keywords:
Next Generation Sequencing
mtDNA genomes
Heteroplasmy
Sanger-type sequencing
PGM
Forensic science

ABSTRACT

Insights into the human mitochondrial phylogeny have been primarily achieved by sequencing full mitochondrial genomes (mtGenomes). In forensic genetics (partial) mtGenome information can be used to assign haplotypes to their phylogenetic backgrounds, which may, in turn, have characteristic geographic distributions that would offer useful information in a forensic case. In addition and perhaps even more relevant in the forensic context, haplogroup-specific patterns of mutations form the basis for quality control of mtDNA sequences. The current method for establishing (partial) mtDNA haplotypes is Sanger-type sequencing (STS), which is laborious, time-consuming, and expensive. With the emergence of Next Generation Sequencing (NGS) technologies, the body of available mtDNA data can potentially be extended much more quickly and cost-efficiently. Customized chemistries, laboratory workflows and data analysis packages could support the community and increase the utility of mtDNA analysis in forensics. We have evaluated the performance of mtGenome sequencing using the Personal Genome Machine (PGM) and compared the resulting haplotypes directly with conventional Sanger-type sequencing. A total of 64 mtGenomes (>1 million bases) were established that yielded high concordance with the corresponding STS haplotypes (<0.02% differences). About two-thirds of the differences were observed in or around homopolymeric sequence stretches. In addition, the sequence alignment algorithm employed to align NGS reads played a significant role in the analysis of the data and the resulting mtDNA haplotypes. Further development of alignment software would be desirable to facilitate the application of NGS in mtDNA forensic genetics.

1. Introduction

Mitochondrial (mt)DNA is present in higher copy number in the cell than nuclear (n)DNA and thus the likelihood of recovering useable DNA data is increased in forensic samples that fail to yield useful nDNA typing results. Due to its maternal mode of inheritance and lack of recombination [1] the discriminatory power of mtDNA is somewhat restricted, however and dependent on mtDNA databases [2]. This limitation is even more evident in routine forensic applications where, to date, analysis has been restricted to the non-coding control region (CR, or its hypervariable segments) for legal and technical reasons. While the use of coding region data outside of the control region would no doubt increase the information content of this genetic marker and increase its utility in practical forensic casework, conventional Sanger-type sequencing (STS) is neither amenable to, nor feasible for, the analysis of the full mtGenome in minute forensic samples.

Yet, outside of evidentiary testing, the forensic field is already making use of full mtGenomes. They are a prerequisite for the phylogenetic assignment of mtDNA haplotypes [3–5] and they form the basis for quality control of novel mtDNA data [6–8]. Most

DOI of original article: http://dx.doi.org/10.1016/j.fsigen.2013.06.003

☆ 'A publishers' error resulted in this article appearing in the wrong issue. The article is reprinted here for the reader's convenience and for the continuity of the special issue. For citation purposes, please use the original publication details: Forensic Science International: Genetics 7(5), 543–549.

* Corresponding author at: Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria. Tel.: +43 512 9003 70640; fax: +43 512 9003 73640.
E-mail address: walther.parson@i-med.ac.at (W. Parson).

1872-4973/$ – see front matter © 2013 The authors. Published by Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.fsigen.2013.09.007

available mtGenomes have been generated with STS. However, this is a laborious, time-consuming, and expensive endeavor, particularly when high quality data with redundant sequence coverage are required [9].

Next Generation Sequencing (NGS) technologies have the potential to significantly increase both sample throughput and overall process efficiency, thereby facilitating the establishment of larger mtGenome databases in relatively short terms [10]. However, careful validation of these new technologies is required to maintain quality and consistency with the established datasets and technologies [11]. To date, only a few publications are available that describe the application of NGS technology to mtDNA testing in the forensic context. Mikkelsen et al. [12] conducted an early study on pyrosequencing with the FLX (454, Roche). Holland et al. [13] investigated the detection of point heteroplasmy (PHP) with the smaller Roche instrument (454 junior) and Loreille et al. [14] described mtGenome sequencing of highly degraded skeletal remains using the Illumina chemistry. In this study, the Personal Genome Machine (PGM [15]) was used to sequence complete mtGenomes, and the NGS results were evaluated by direct comparison with STS derived consensus haplotypes.

2. Materials and methods

2.1. Samples and DNA extraction

A total of 42 samples were voluntarily provided under informed consent. These included peripheral blood samples from five indigenous Khoe-San individuals from Angola [16] six paraffin-embedded tissue samples from Tyrol, Austria [17] buccal swab samples from eight individuals of the Democratic Republic of Timor-Leste in Dili (present study) and buccal swab samples from 23 individuals from Tyrol, Austria (present study). DNA extraction was performed either as previously described for the pre-existing samples, or using the Chelex protocol as detailed in [18].

2.2. PCR amplification

The entire mtDNA molecule was amplified with two overlapping 8.5 kilo base pair (kbp) fragments according to the protocol described in Ref. [9]. Both amplicons were purified using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) and quality-controlled/quantified using non-denaturating polyacrylamide gel electrophoresis with silver staining [19] and the Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) following the manufacturer's recommendations.

2.3. Sanger-type sequencing

Earlier published mtGenomes from five Khoe-San individuals [16] and six individuals from Tyrol [17] were Sanger-type sequenced following the protocol outlined in the respective publications. For the remaining 31 novel mtGenomes generated in this study, some sequencing primers were replaced (Table S1). All sequences were imported into Sequencher 5.0 (Gene Codes Corporation, Ann Arbor, MI, USA) and aligned relative to the revised Cambridge Reference Sequence for human mitochondrial DNA (rCRS [20]) using the phylogenetic alignment rules detailed in Ref. [21]. STS data were analyzed by at least two independent scientists and the final consensus haplotypes were based on redundant sequence coverage over all positions (at least two independent sequence reads).

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2013.09.007.

2.4. Library construction for the PGM

The construction of the library involved the following three steps: enzymatic shearing, ligation of the adapters and size selection. The quantity of amplified DNA was determined with a Nanodrop spectrophotometer (Nanodrop Products, Wilmington, DE, USA). Both 8.5 kbp fragments were normalized to a quantity of 100 ng and then pooled. The amplicons were enzymatically sheared into suitable sized fragments using the Ion Xpress Plus Fragment Library Kit (Life Technologies (LT), Foster City, CA, USA) following the manufacturer's recommendations. For the 100 bp sequencing kit, incubation times were set to 25 min to yield fragments with sizes of approx. 130 bp. For the 200 bp sequencing kit shearing times were reduced to 7 min to yield fragments around 260 bp. Size and quality of fragmented DNA were determined with the Agilent DNA High Sensitivity Kit on the Bioanalyzer (Agilent) following the manufacturer's recommendations. Specific Ion Torrent compatible adapters were ligated onto the 5′ and 3′ ends of each fragment and linked by nick translation. For the barcoded libraries, the Ion P1 Adapter and the Ion Xpress barcode *X* adapter (LT, *X* = number of the used barcode) were applied to allow for sequencing multiple samples simultaneously. The fragmented and adapter ligated libraries were size selected using the E-Gel SizeSelect Agarose Gel (Invitrogen Corporation, Carlsbad, CA, USA) following the manufacturer's recommendations, and then batches of 4, 12, 15 and 31 samples were subsequently loaded onto 316 chips.

2.5. Template preparation

The quantity of the size selected library was determined by a real-time PCR approach using the Ion Library Quantitation Kit (LT) following the manufacturer's recommendations, with the template dilution factor calculated for a final concentration of ~26 pM per target. Targets were then subjected to emulsion PCR using the Ion One Touch (LT) following the manufacturer's recommendations. For clonal amplification, DNA was localized to Ion Sphere particles (LT), which were automatically enriched with the Ion OneTouch ES system (LT). Quality was assessed using the Qubit 2.0 Fluorometer (Invitrogen Corporation) following the manufacturer's recommendations.

2.6. PGM sequencing

Next Generation Sequencing (NGS) was performed using the Personal Genome Machine (PGM, LT). Before initializing the PGM Sequencer, a cleaning protocol was performed that started with a chlorite cleaning solution and was followed by a wash with 18 MΩ water (Elga Purelab Flex 3 Water Purification, Veolia Water Systems, Austria). After initialization, the chip was washed with 100% isopropanol and annealing buffer (from sequencing kit) and then tested for its functionality on the PGM. Sequencing primer and Control Ion Spheres of the Ion PGM sequencing kit were added to the library. After the annealing step sequencing polymerase was added and the sample was loaded onto the chip.

2.7. PGM data analysis

All PGM sequences were analyzed with the Ion Torrent Software Suite (Vs. 3.2) using the plug-in variant caller (Vs. 3.2.43647) that employed a TMAP Smith–Waterman alignment optimization [22]. The output of the variant caller was presented in tabular format, as a list of differences to the rCRS without a graphical display of the aligned reads. At this time, graphical displays of the TMAP alignment could only be visualized with separate tools for alignment and assembly viewing, such as