



Research paper

Strategies for complete mitochondrial genome sequencing on Ion Torrent PGM™ platform in forensic sciences



Yishu Zhou^a, Fei Guo^{b,c}, Jiao Yu^b, Feng Liu^b, Jinling Zhao^b, Hongying Shen^b, Bin Zhao^b, Fei Jia^b, Zhu Sun^b, He Song^a, Xianhua Jiang^{a,b,*}

^a China Medical University School of Forensic Medicine, No. 77, Puhe Road, Shenyang North New Area, Shenyang, Liaoning 110122, PR China

^b Criminal Science and Technology Institute of Liaoning Province, No. 2, Qishan Middle Road, Huanggu District, Shenyang, Liaoning 110032, PR China

^c Department of Forensic Medicine, National Police University of China, No. 83, Tawan Street, Huanggu District, Shenyang, Liaoning 110854, PR China

ARTICLE INFO

Article history:

Received 21 April 2015

Received in revised form 30 December 2015

Accepted 8 January 2016

Available online 15 January 2016

Keyword:

Next generation sequencing (NGS)

Ion Torrent PGM™

Mitochondrial genome (mtGenome)

Northern Chinese Han

Forensic sciences

ABSTRACT

Next generation sequencing (NGS) is a time saving and cost-efficient method to detect the complete mitochondrial genome (mtGenome) compared to Sanger sequencing. In this study we focused on developing strategies for mtGenome sequencing on the Ion Torrent PGM™ platform and NGS data analysis. With our experience, 4, 15 and 30 samples could be loaded onto Ion 314™, Ion 316™ and Ion 318™ chips respectively at a pooling concentration of 26 pM, achieving to sufficient average coverage of $\geq 1500\times$ and well strand balance of 1.05. Data processing software is essential to NGS mega data analysis. The in-house Perl scripts were developed for primary data analysis to screen out uncertain positions and samples from variant call format (VCF) reports and for pedigree study to perform pairwise comparisons. The Integrative Genomic Viewer (IGV) and the NextGENE software were introduced to secondary data analysis. The mthap and EMMA were employed for haplogroup assignment. The dataset was reviewed and approved by the EMPOP as the final version, which showed 2.66% error rate generated from the Torrent Variant Caller (TVC). Across the mtGenome, 4022 variants were found at 725 nucleotide positions, where ratio of transitions to transversions was estimated at 20.89:1 and 22.18% of variants was concentrated at hypervariable segments I and II (HVS-I and HVS-II). Totally, 107 complete mtGenome haplotypes were observed from 107 Northern Chinese Han and assigned to 88 haplogroups. The random match probability (RMP) of complete mtGenome was calculated as 0.009345794, decreasing 26.19% by comparison to that of HVS-I only, and the haplotype diversity (HD) was evaluated as 1, increasing 0.33% by comparison to that of HVS-I only. Principal component analysis (PCA) showed that our population was clustered to East and Southeast Asians. The strategies in this study are suitable for complete mtGenome sequencing on Ion Torrent PGM™ platform and Northern Chinese Han (EMP00670) is the first complete mtGenome dataset contributed to the EMPOP from East Asian.

© 2016 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The human mitochondrial genome (mtGenome) is an approximately 16,569 base pairs (bp) long extra-chromosomal genome presenting in the subcellular organelle. The 15,447 bp long coding region (CodR) is economical construction, which encodes 37 genes correlate to oxidative phosphorylation. The 1,122 bp long non-coding control region (CR) has proven to be informatics, which

comprises an origin of replication along with three hypervariable segments (HVS) [1].

Compared with nuclear DNA, mitochondrial DNA (mtDNA) possesses some unique characteristics including simplistic inheritance and great copy number per cell [2]. It is considered to be inherited strictly from our mothers and to be transmitted without recombination across many generations. While specimens like highly decomposed remains, bones or hair shafts are hard to obtain full short tandem repeat (STR) profiles, mtDNA typing can provide an alternative clue to investigation by comparison with known maternal relatives [3]. Thus, it is commonly used in human evolutionary studies [4] and forensic practice [5].

Previous studies often restrict to sequence the HVS-I, II and III of CR [6,7] as well as some specific single nucleotide polymorphisms (SNPs) of CodR [8]. Such partial information may limit the

* Corresponding author at: China Medical University School of Forensic Medicine, No. 77, Puhe Road, Shenyang North New Area, Shenyang, Liaoning 110122, PR China. Tel.: +86 24 31939433; fax: +86 24 31939433.

E-mail address: jiangxianhua_2006@aliyun.com (X. Jiang).

polymorphism information content of this genetic marker and hinder its application in practical forensic casework. Undoubtedly, the complete mtGenome sequencing will provide new insights beyond current capabilities. However, most available mtGenomes are generated with conventional Sanger sequencing technology, which is a time-consuming and expensive task particularly when large mtGenome databases are established or high quality data are required with redundant sequence coverage. Fortunately, next generation sequencing (NGS) technology has the potential to dramatically increase sample throughput, workflow efficiency and detection resolution, therefore facilitating to obtain reliable and accurate complete mtGenome information. In this study, we have developed strategies for complete mtGenome sequencing on the Ion Torrent Personal Genome™ Machine (PGM™) platform and investigated mtGenome features of the Northern Chinese Han population to evaluate the application in forensic sciences.

2. Materials and methods

2.1. Samples and DNA extraction

A total of 125 peripheral blood samples of Northern Han were collected from Shenyang City, Liaoning Province, Northeast China after informed consent, including 107 samples from unrelated individuals and 18 samples from five maternal genealogies (four generations in three families and three generations in two families). DNA was extracted on the EZ1® Biorobot workstation using the EZ1® DNA Investigator Extraction Kit (QIAGEN, Hilden, Germany). Purified DNA was stored at -20°C .

2.2. PCR amplification

The mtGenome was amplified in two independent reactions using primers described in [9], namely L644, H8982, L8789, and H877 (Table S1). Two amplicons, fragment A (8.3 kilo base pairs, kbp) and fragment B (8.6kbp), were partially overlapped at positions 622–898 and 8764–9003. Long range PCRs were carried out using the SequalPrep™ Long PCR Kit (Thermo Fisher Scientific, MA, USA) according to the manufacturer's recommendations with two modifications: (1) the total DNA input was 30–50 ng per reaction. Fragment A was dealt with 2 μl enhancer A and 2 μl enhancer B; fragment B was done with 1 μl enhancer A. (2) amplification was performed on the GeneAmp® 9700 System thermal cycler (Thermo Fisher) with the thermal cycling condition for both reactions: denaturation for 2 min at 94°C , amplification for 32 cycles of 10 s at 94°C , 45 s at 60°C , 9 min at 68°C , final extension for 5 min at 72°C , and hold at 4°C . PCR products (5 μl) were qualified and quantified by 1.0% agarose gel electrophoresis with GoodView Nucleic Acid Stain (SBS Genetech, Beijing, China) to confirm successful amplification.

2.3. Library construction

Library construction included three steps as follows: fragmentation, adapter ligation and size selection. (1) Fragmentation: fragments A and B were quantified on the Qubit® 2.0 fluorometer using the Qubit® dsDNA BR Assay kit (Thermo Fisher) and were pooled with each of 50 ng. Then enzymatic shearing was performed using the Ion Shear™ Plus Reagent (Thermo Fisher) following the manufacturer's instruction. In order to yield fragments with a median size of 200 bp, shearing condition was adjusted to 8 min at 37°C ; (2) Adapter ligation: both ends of each fragment were ligated with the Ion P1 Adapters and the Ion Xpress Barcode Adapters according to the manufacturer's recommendations; (3) Size selection: it was conducted on the E-Gel® iBase™ Power System with E-Gel® SizeSelect™ 2% agarose gels (Thermo

Fisher) according to the manufacturer's recommendations. The selected libraries were amplified for 8 cycles with the reagents supplied in Ion Plus Fragment Kit (Thermo Fisher). Purification was an essential step after each process (fragmentation, adapter ligation and libraries amplification), using the Agencourt AMPure® XP PCR Purification system following the manufacturer's recommendations.

2.4. Template preparation and PGM™ sequencing

Quantification and qualification for the amplified libraries were examined on the Agilent 2100 Bioanalyzer System with the Agilent High Sensitivity DNA Kit (Agilent Technologies, CA, USA) according to the manufacturer's recommendations. All barcoded libraries were pooled in equimolar amounts of 26 pM to ensure equal representation of each barcoded library in the sequencing run. Pooled libraries were then subjected to emulsion PCR on the Ion OneTouch™ 2 instrument (Thermo Fisher) with the Ion PGM™ Template OT2 200 Kit (Thermo Fisher), and followed by template-positive Ion Sphere Particles (ISPs) enrichment on the Ion OneTouch™ ES instrument (Thermo Fisher) according to the manufacturer's recommendations.

NGS was performed on PGM™ platform (Thermo Fisher) with three size chips (Ion 314™ Chip, Ion 316™ Chip and Ion 318™ Chip) using the Ion PGM™ Sequencing 200 v2 Kit. All preparations were conducted to the protocol released by the manufacturer: chlorite cleaning and 18.2 M Ω water cleaning, initialization the PGM™ system and chip check. Enriched template-positive ISPs were mixed with control ISPs, sequencing primers and sequencing polymerase in a consecutive way, and then loaded onto an Ion chip carefully.

2.5. Data analysis

All raw reads were mapped with the revised Cambridge Reference Sequence (rCRS, NC_012920.1) [10] by the Ion Torrent Suite Software (TSS) version 4.0 based on the TMAP Smith-Waterman alignment algorithm. Variants were called using the Torrent Variant Caller (TVC) plugin v.4.0 with somatic-low stringency and filed into variant call format (VCF) reports. The in-house Perl scripts (Boxes 1–3) were compiled for primary data analysis. Base by base investigation was performed using the Integrative Genomic Viewer (IGV) package v.2.3.34 [11] and the NextGENe version 2.3.4.5 (SoftGenetics, PA, USA), where binary alignment map (BAM) and binary alignment index (BAI) files were visualized based on the Torrent mapping alignment program (TMAP) using the IGV, and FASTQ-converted-FASTA files were re-aligned based on the Burrow–Wheeler transform (BWT) alignment algorithm using the NextGENe. Parameters of NextGENe were described in [12]. The preliminary haplogroup assignment was chosen from the highest ranked haplogroup from mthap (<http://dna.jameslick.com/mthap>) and EMMA [13] in the European DNA Profiling Group mtDNA Population (EMPOP) database (www.empop.org) [14]. The final haplogroup assignment was determined according to rCRS-oriented version of the mtDNA tree Build 16 (released 19 Feb., 2014) [15]. Homopolymers in rCRS and each sample were detected using mreps 2.6 software [16]. Profiles were compared among maternal genealogies using Perl script (Box 4) for pedigree study.

2.6. Sanger sequencing

Sanger sequencing was employed to confirm variants in discordance between IGV and NextGENe. Primers for Sanger sequencing were listed in Table S1, where five primers were described in [17] and one primer was self-designed in this study.

Download English Version:

<https://daneshyari.com/en/article/98701>

Download Persian Version:

<https://daneshyari.com/article/98701>

[Daneshyari.com](https://daneshyari.com)