Analytical Biochemistry 490 (2015) 1–6

Contents lists available at ScienceDirect

Analytical Biochemistry

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

Differentiating between monozygotic twins through next-generation mitochondrial genome sequencing

Zheng Wang [a, b], Ruxin Zhu [a], Suhua Zhang [a, b], Yinnan Bian [a], Daru Lu [b], Chengtao Li [a, *]

[a] Shanghai Key Laboratory of Forensic Medicine, Institute of Forensic Science, Ministry of Justice, P. R. China, Shanghai 200063, China
[b] State Key Laboratory of Genetic Engineering, Institute of Genetics, School of Life Science, Fudan University, Shanghai 200433, China

ARTICLE INFO

Article history:
Received 13 June 2015
Received in revised form
2 August 2015
Accepted 19 August 2015
Available online 29 August 2015

Keywords:
Forensic genetics
Mitochondrial genome (mtGenome)
Monozygotic twins
Massively parallel sequencing (MPS

ABSTRACT

Monozygotic (MZ) twins, considered to be genetically identical, cannot be distinguished from one another by standard forensic DNA testing. A recent study employed whole genome sequencing to identify extremely rare mutations and reported that mutation analysis could be used to differentiate between MZ twins. Compared with nuclear DNA, mitochondrial DNA (mtDNA) has higher mutation rates; therefore, minor differences theoretically exist in MZ twins' mitochondrial genome (mtGenome). However, conventional Sanger-type sequencing (STS) is neither amenable to, nor feasible for, the detection of low-level sequence variants. The recent introduction of massively parallel sequencing (MPS) has the capability to sequence many targeted regions of multiple samples simultaneously with desirable depth of coverage. Thus, the aim of this study was to assess whether full mtGenome sequencing analysis can be used to differentiate between MZ twins. Ten sets of MZ twins provided blood samples that underwent extraction, quantification, mtDNA enrichment, library preparation, and ultra-deep sequencing. Point heteroplasmies were observed in eight sets of MZ twins, and a single nucleotide variant (nt15301) was detected in five sets of MZ twins. Thus, this study demonstrates that ultra-deep mtGenome sequencing could be used to differentiate between MZ twins.

Today, individual identification in forensic casework is mainly based on short tandem repeat (STR) polymorphisms, and a set of 19–24 highly polymorphic STR markers is often genotyped [1,2]. However, monozygotic (MZ) twins are considered to be genetically identical in that they have identical STR profiles [3]. Thus, they cannot be differentiated using STR profiling. This inability causes problems in distinguishing one from another in criminal cases with an MZ twin as a suspect. Recently, small epigenetic differences, especially DNA methylation pattern, between twins have been described [4–6]. However, DNA methylation analysis generally needs bisulfite or enzymatic digestion treatment, which is very labor-intensive and complex and has risk for DNA damage.

In 2014, Weber-Lehmann and coworkers [7] employed whole genome sequencing to search potential single nucleotide polymorphisms (SNPs) for the differentiation between MZ twins. Compared with nuclear DNA, mitochondrial DNA (mtDNA), an extranuclear genome, exhibits higher mutation rates due to the presence of fewer DNA repair mechanisms [8,9]. The 10-fold higher mutation rate, relative to nuclear DNA, helps to introduce more variability in mitochondrial genome (mtGenome); thus, it has been suggested that significant variability presents between individuals and that all individuals would display mtDNA heteroplasmy [10–12]. Although the use of full mtGenome data would no doubt increase the information content and increase its utility in practical forensic casework [13–16], conventional Sanger-type sequencing (STS) is not suitable for the analysis of the full mtGenome and not possible to clearly identify minor heteroplasmic variants.

The recent introduction of massively parallel sequencing (MPS), also termed next-generation sequencing, has revolutionized genomic studies; many targeted regions' sequence information of multiple samples can be obtained much more quickly and cost-efficiently [17]. MPS technology has the potential to increase both sample throughput and overall process efficiency, thereby holding great potential for efforts to expand mtDNA typing beyond current capabilities. Next-generation mtGenome sequencing is currently being used in forensic applications, including mtDNA validation

Abbreviations used: STR, short tandem repeat; MZ, monozygotic; SNP, single nucleotide polymorphism; mtDNA, mitochondrial DNA; mtGenome, mitochondrial genome; STS, Sanger-type sequencing; MPS, massively parallel sequencing; PCR, polymerase chain reaction; rCRS, revised Cambridge Reference Sequence; IGV, Integrative Genomics Viewer; PHP, point heteroplasmy.

* Corresponding author. Fax: +86 21 52352959
E-mail address: lichengtaohla@163.com (C. Li).

http://dx.doi.org/10.1016/j.ab.2015.08.024
0003-2697/

[18–20], larger mtGenome reference data establishment [21,22], and investigation of heteroplasmy occurrence [11,12]. In this study, we explored the use of MPS technology to identify minor differences in mtGenomes between MZ twins. The Illumina HiSeq 2000 Sequencing System uses a reversible terminator-based sequencing by synthesis method capable of producing a massive parallel sequencing environment. For mtDNA sequencing in the forensic context, mtGenome is amplified, fragmented, and modified with adaptors and dual indexes, and it is pooled (for multiplexing), generates DNA cluster by bridge polymerase chain reaction (PCR), and is sequenced. Consequently, the aim of this study was to evaluate whether next-generation mtGenome sequencing analysis will allow for the differentiation between MZ twins.

Materials and methods

Sample preparation

Human blood samples were collected with the approval of the ethics committee of the Institute of Forensic Science, Ministry of Justice, P. R. China. Informed written consent was obtained from 10 sets of MZ twins (25–58 years old). Whole blood samples were collected by venipuncture without anticoagulation treatment. DNA was extracted using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The quantity of extracted DNA was estimated using the Quantifiler Human DNA Quantification Kit (Thermo Fisher, Foster City, CA, USA) on the Applied Biosystems 7500 Real-Time PCR System following the manufacturer's recommendations. Samples were normalized to 1 ng/μl and stored at −20 °C until mtDNA enrichment.

The monozygosity of the twins was confirmed by typing all individuals with the GoldenEye 20A Kit (PeopleSpot, Beijing, China) following the manufacturer's protocol. PCR products were separated on an ABI 3130xl Genetic Analyzer and evaluated using GeneMapper ID version 3.2 software (Thermo Fisher).

Long-range PCR amplification

The entire mtGenome was amplified by long-range PCR in two separate reactions according to the protocol described by Fendt and coworkers [23]. Negative controls (negative amplification control and reagent blank control) were used as controls for potential contamination. PCR products were purified using the QIAquick PCR Purification Kit (Qiagen) and then imaged by agarose gel electrophoresis to confirm successful amplification. The quantity of amplicons was determined using the Qubit dsDNA BR Quantification Kit with the Qubit 2.0 Fluorometer (Thermo Fisher), and then equal quantities of two amplicons were pooled to produce 1.0 ng of DNA for library preparation.

Library preparation

Next-generation sequencing libraries were prepared from blood samples of the MZ twins according to the common guidelines for shotgun library preparation. Briefly, 20 pooled amplicons were prepared using the Nextera XT DNA Sample Preparation Kit (Illumina, San Diego, CA, USA) following the manufacturer's recommendations. Subsequently, Nextera XT products were purified using Agencourt AMPure XP beads (Beckman Coulter, Indianapolis, IN, USA). After cleanup, the libraries were quantified using the Qubit dsDNA BR kit and evaluated for fragment size using the High Sensitivity DNA Kit with the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). All samples were normalized to the same concentration, pooled, and denatured according to the manufacturer's instructions.

Sequencing and data analysis

Sequencing was performed on the Illumina HiSeq 2000 system with chemistry version 3.0 and using the 2 × 100-bp paired-end read mode according to the manufacturer's recommendations. The initial data analysis was carried directly on the HiSeq 2000 system during the run. The FASTQ files comprising the sequence information were mapped to the revised Cambridge Reference Sequence (rCRS; GenBank ID: NC_012920.1) [24] using Burrows–Wheeler Alignment software (BWA version 0.6.2) [25]. Using the SAMtools software package version 1.2 [26], the mapping results were filtered by applying a mapping quality threshold of 20. By multi-sample calling function (mpileup file), a single BAM file containing good-quality unique mapping reads was obtained for each of the twins. VarScan2 software [27] was used to identify variants, and Variant Call Format (VCF) files were generated. The analysis in this study used the manual settings for minimum base call quality (Q30), min coverage ≥500, min reads2 (minor component) ≥200, and min var freq (minor component frequency) ≥0.05. The variant nucleotides from the reference were annotated by base difference and verified by manually viewing BAM files in an Integrative Genomics Viewer (IGV) [28].

Conventional sanger sequencing

PCR amplification for STS was performed using two pairs of primers (F-L16047: 5′-CATGGGGAAGCAGATTTG-3′ and R-H16464: 5′-TTAGCTACCCCCAAGTGT-3′; F-L29: 5′-GGT CTATCACCCTATTAACCAC-3′ and R-H408: 5′-CTGTTAAAACTGCATACCGCCA-3′) to amplify hypervariable regions. Two independent reactions were performed, each in a total volume of 50 μl containing 5 μl of 10 × PCR buffer, 5 μl of 25 mM $MgCl_2$, 2 μl of 10 mM deoxynucleoside triphosphate (dNTP), 5 μl of each primer, 0.4 μl of AmpliTaq Gold Polymerase, and 1 ng of DNA template. PCR was performed in a GeneAmp PCR System 9700 thermal cycler under the following conditions: 94 °C hold for 11 min, 35 cycles of 94 °C for 30 s, 55 °C for 60 s, 72 °C for 90 s, hold for 10 min at 72 °C, and a final hold at 4 °C. Amplified products were purified with the QIAquick PCR purification kit. Purified products were sequenced on an ABI 3730xl capillary sequencer equipped with 50-cm capillaries and POP7 polymer following standard protocols. All sequences were imported into Vector NTI Suite 8.0 (Thermo Fisher) and aligned relative to the hypervariable regions of rCRS (commonly referred to as HV1 and HV2). STS data were analyzed by at least two independent scientists, and mtDNA haplotypes were recorded relative to the rCRS (NC_012920.1).

Results

The monozygosity of the twins was confirmed using standard forensic STR typing with the GoldenEye 20A kit (data not shown). DNA obtained from blood samples of the MZ twins was used for amplification of the entire mtGenome. Two overlapping PCR fragments were confirmed by agarose gel electrophoresis, and no amplification band was observed in the negative control samples. Following library preparation and subsequent whole mtGenome sequencing, the Illumina HiSeq system generates approximately 13.63 gigabases (Gb) of Q30 data. Assuming equal coverage across the mtGenome, each indexed sample would be expected to have more than 40,000 × coverage at each base position of the mtGenome. In practice, coverage was not dispersed evenly among individuals, and Fig. 1 shows the number of nonredundant and uniquely mapped reads for all MZ twins (corresponding mtGenome coverage ranged from 33,078 to 56,130). The production of the raw sequencing data generated required approximately 2 weeks. This