Title: Concordance and Reproducibility of a Next Generation mtGenome Sequencing Method for High-Quality Samples using the Illumina MiSeq

Author: Michelle A. Peck Michael D. Brandhagen Charla Marshall Toni M. Diegoli Jodi A. Irwin Kimberly Sturk-Andreaggi

PII: S1872-4973(16)30104-1
DOI: http://dx.doi.org/doi:10.1016/j.fsigen.2016.06.003
Reference: FSIGEN 1531

To appear in: *Forensic Science International: Genetics*

Received date: 29-2-2016
Revised date: 27-5-2016
Accepted date: 3-6-2016

Please cite this article as: Michelle A.Peck, Michael D.Brandhagen, Charla Marshall, Toni M.Diegoli, Jodi A.Irwin, Kimberly Sturk-Andreaggi, Concordance and Reproducibility of a Next Generation mtGenome Sequencing Method for High-Quality Samples using the Illumina MiSeq, Forensic Science International: Genetics http://dx.doi.org/10.1016/j.fsigen.2016.06.003

Concordance and Reproducibility of a Next Generation mtGenome Sequencing Method for High-Quality Samples using the Illumina MiSeq

Michelle A. Peck[a,b,*], Michael D. Brandhagen[c], Charla Marshall[a,b], Toni M. Diegoli[a,b,1,2], Jodi A. Irwin[c], and Kimberly Sturk-Andreaggi[a,b]

[a]Armed Forces DNA Identification Laboratory, 115 Purple Heart Drive, Dover AFB, DE 19902, United States
[b]American Registry of Pathology Sciences, 9210 Corporate Boulevard, Suite 120, Rockville, MD 20850, United States
[c]Federal Bureau of Investigation Laboratory, 2501 Investigation Parkway, Quantico, VA 22135, United States
[1]Present address: Defense Forensic Science Center, 4930 N. 31st Street, Forest Park, GA 30297, United States
[2]Present address: ANSER, 5275 Leesburg Pike, Suite N-5000, Falls Church, VA 22041, United States

[*]Corresponding author at: Armed Forces DNA Identification Laboratory, 115 Purple Heart Dr., Dover AFB, DE 19902, United States. Tel.: +1 302 346 8519. *E-mail address*: michelle.a.peck3.ctr@mail.mil

Highlights

- The full mtGenome of 90 high quality samples were sequenced on the Illumina MiSeq
- Nextera XT library preparation and sequencing was performed at two laboratories
- NGS data was 99.9996% concordant with previously generated Sanger data
- Variant calls were reproducible and variant frequency (VF) differed by only 0.23%
- Replicate analysis resulted in the same variant calls and only 0.01% VF difference

Abstract

Sanger-type sequencing (STS) of mitochondrial DNA (mtDNA), specifically the control region (CR), is routinely employed in forensics in human identification and missing persons scenarios. Yet next-generation sequencing (NGS) has the potential to overcome some of the major limitations of STS processing, permitting reasonable paths forward for full mitochondrial genome (mtGenome) sequencing, while also offering higher-throughput and higher sensitivity capabilities. To establish the accuracy and reproducibility of NGS for the development of mtDNA data, 90 DNA extracts that were previously used to generate forensic quality full mtGenomes using STS were sequenced using Nextera XT library preparation and the Illumina MiSeq. Using the same amplicon product, replicate library sets were generated and sequenced at different laboratories, and analysis was performed in replicate using the CLC Genomics Workbench. Both sequencing sets resulted in 99.998% of positions with greater than 10X coverage when 96 samples (including controls) were multiplexed. Overall, 99.9996% concordance was observed between the NGS data and the STS data for the full mtGenome. The only "discordant" calls involved low level point heteroplasmies, with the differences resulting from stochastic variation and/or the increased sensitivity of NGS. Higher sensitivity also allowed for the detection of a mixed sample previously not detected with STS. Additionally, variant calls were reproducible between sequencing sets and between software analysis versions with the variant frequency only differing by 0.23% and 0.01%,