Document Title:	**Evaluation and Implementation of High Throughput Second Generation Sequencing for Mitochondrial DNA Testing in Missing Persons and Forensic Casework at the UNT Center for Human Identification**
Author(s):	**Bruce Budowle**
Document Number:	**307117**
Date Received:	**June 2023**
Award Number:	**2016-DN-BX-K001**

Federal Agency:	Office of Justice Programs, National Institute of Justice, Department of Justice
Federal Grant:	Award Number 2016-DN-BX-K001
Project Title:	Evaluation and Implementation of High Throughput Second Generation Sequencing for Mitochondrial DNA Testing in Missing Persons and Forensic Casework at the UNT Center for Human Identification
PI:	Bruce Budowle, University of North Texas Health Science Center at Fort Worth, Center for Human Identification, Department of Microbiology, Immunology, and Genetics, Fort Worth, TX 76107 tel: 817-735-2979; email: bruce.budowle@unthsc.edu
Submission Date:	June 05, 2019
DUNS Number:	110091808
EIN:	1756064033A1
Recipient Organization:	University of North Texas Health Science Center 3500 Camp Bowie Blvd., Fort Worth, TX 76107
Institutional Profile Number:	6108502
Project/Grant Period:	03-01-2017 - 06-30-2019
Reporting Period:	03-01-2017 - 06-30-2019
Report:	**Final Summary Report**

Signature: 

Bruce Budowle, Ph.D.

Summary

The Final Technical Report contains the details of the study and thus is not described as explicitly herein. Those interested in more specific technical aspects should consult the Final Technical Report. Instead, a summary of the findings is provided to briefly capture the work performed in this project.

The mitochondrial genome (mtGenome) has a number of characteristics, such as a higher copy number per cell compared with the nuclear genome, a well-characterized phylogeny, a lack of recombination, and is maternally inherited for lineage identification, that provide useful information for forensic investigations. Mitochondrial DNA (mtDNA) has been analyzed by Sanger sequencing [1] and this approach has been applied to forensic casework for more than 20 years [2-3]. While a valid and reliable approach, Sanger sequencing is time-consuming, costly, enables only targeted sequencing of the mtGenome, is low throughput, cannot accommodate analysis of length heteroplasmy, and is not quantitative. In contrast massively parallel sequencing (MPS) may be able to overcome many of the limitations of Sanger sequencing.

MPS technologies now make it feasible for crime laboratories to sequence the entire mtGenome in one run and without consuming additional sample. Sequencing the entire mtGenome expands analysis to a previously untapped resource of far greater variation that resides in the coding region of the mtGenome, which provides a higher discrimination power and higher phylogenetic resolution [4-6]. Additionally, the quantitative nature of MPS provide opportunities for mixture interpretation [7-9].

Even with the noted advantages of MPS, transitioning from current Sanger sequencing technologies to a substantially different workflow with MPS is not a trivial endeavor. The primary goal of this work is to transition whole mtGenome sequencing from a research environment to that of the case working laboratory. Since UNTCHI already performs mtDNA analysis in casework using Sanger sequencing, a baseline performance of the methodology was available to compare how well MPS will perform with that of current capabilities. This implementation process also leveraged the research and validation work being performed at the UNTCHI Research Unit. The Research Unit has been investigating the feasibility of mtGenome sequencing for forensic applications for the past several years [5, 10-19]. The results of that effort demonstrated that the technology was sufficiently mature to consider moving it into casework in the Missing Persons Unit and the Forensic Casework Unit at UNTCHI.

The Ion Torrent S5 system and Ion Chef were sought for these implementation efforts because of several workflow advantages. First and foremost, the Precision ID mtDNA Genome panel amplifies the entire mtGenome with two primer pools of 81 primer pairs each with amplicons of 175 bases in length or less [20]. Therefore, the entire mtGenome can be sequenced from DNA of the limited quality that is typically encountered in human remains. Additionally, the UNTCHI Research Unit already has tested (and presented at national and international meetings) the sensitivity of detection of the system with as little as one pg nuclear DNA input. The final factor for selecting this system is the availability of the Ion Chef, which reduces the manual steps in the Ion Torrent MPS workflow to a mere three pipetting steps, particularly in the library preparation phase. By automating the library preparation and chip loading phases of the workflow, the Ion

Chef reduces the opportunities for user-introduced error, reduces the amount of time an analyst must spend at the lab bench, freeing the analysts up for the more time-consuming analytical steps, and yields more consistent results among users.

Several factors contributed to the success of this project. One of these factors was that UNTHSC's Center for Human Identification (UNTCHI) performs mtDNA casework analysis routinely. Thus, the analysts already had a wealth of experience with mtDNA analyses and were quite familiar with handling mtDNA evidence. This experience aided the progress of the detailed training program developed to familiarize analysts with MPS technologies. The upfront portion of the MPS workflow, such as DNA extraction, quantification, and PCR amplification, are the same for Sanger sequencing methodologies and were steps the analysts did not need to learn *de novo*. The analysts experience with mtDNA analysis also meant they were familiar with the genetic variation in the mtGenome. This knowledge facilitated them in learning to evaluate MPS data.

The routine performance of casework at UNTCHI, an additional factor contributing to the success of this project, meant current capabilities were well known and documented. This experience allowed the performance of MPS to be measured directly with the method primarily used in the crime laboratory (over the last two decades). Thus, the current operational metrics can be collated early on in the proposal. This knowledge of current capabilities and the experience gained during the training program allowed for an effective generation of useful and efficient workflow considerations, standard operating protocols (SOPs), worksheets, and quality assurance practices.

The UNTCHI Research Unit has been validating MPS and mtDNA for a number of years and did not need a gearing up phase to learn the methodology, another factor contributing to success. Thus, UNCHI had someone on staff who had sufficient fundamental knowledge to assist analysts with the necessary training on MPS technologies and address technical difficulties that would inevitably arise during technology transfer. The fruits of that research effort were leveraged so that substantial resources could be dedicated to support the validation and concordance testing. These pre-existing capabilities and efforts enabled implementation in a more cost effective fashion. With an in-house understanding of how the MPS workflow performed, informative validation studies could be planned minimizing the range of parameters that needed to be tested. Therefore, more time could be devoted to the development of bioinformatic tools and data interpretation guidelines.

This project shows the community how MPS substantially can increase the value of mtDNA sequencing for identification of unidentified human remains and analysis of hairs in traditional forensic casework. With this MPS-based workflow, it is possible to characterize the entire mtGenome from highly degraded DNA samples using no more sample input than current Sanger sequencing methods use. Expansion of analysis to the entire mtGenome identified additional single nucleotide polymorphisms (SNPs) that were previously untapped residing in the coding region of the mtGenome due to the need to focus solely on the control region due to the limitations of Sanger sequencing. This additional data increased resolution in haplotype and haplogroup calls.

Data at each stage of this project indicated that the tested MPS workflow generated reliable mtDNA results for the analysis of biological evidence. The numerous resources generated during this process, such as training materials, SOPs, and worksheets, will be made available to the

forensic community so that others may benefit from this project's efforts to facilitate their own implementation of MPS in their respective laboratories.

Account of the Activities

The primary goals of this research were to: 1) to train in a formal manner Technical Leaders and Analysts and document the training process; 2) perform internal validation of the analytical process; 3) develop a working SOP; 4) develop accompanying tools, worksheets, data output reports for the mtGenome sequencing process; 5) institute appropriate Quality Assurance practices; 6) perform concordance testing with MPS and Sanger Sequencing methods; and 7) implement mtGenome sequencing.

Accomplishments

The project was highly successful in all activity areas.

A detailed training program was developed to train the forensic analysts in the Missing Persons Unit and Forensic Unit at UNTCHI. Prior to and during the weekly training sessions, a reading list was provided to the analysts. These papers were selected to increase fundamental knowledge, increase confidence in the use of MPS terminology, and allow analysts to become current on MPS research. After this initial reading list, journal articles on MPS were incorporated into the analysts' regularly scheduled journal club. Weekly training sessions started with 6 one-hour classroom style lectures. The topics of each lecture can be found in the Final Technical Report. These lectures were followed by in-house generated videos of the entire workflow to allow analysts an up-close real-time view of the procedure they had just learned in a classroom environment prior to hands-on use of the instrumentation. Following these didactic lectures, weekly training sessions moved to a computer lab to provide hands-on lessons in analysis of MPS data from the mtGenome. Raw data (BAM, BAI, and VCF files) were provided to the analysts to educate them on how to use the software and assist them in generating mitochondrial haplotypes from MPS data. Finally, analysts progressed to hands-on wet-lab training where teams of two sequenced 30 mtGenomes on the S5 where half of the libraries were prepared manually and half of the libraries were prepared in an automated fashion on the Ion Chef. Analysts' training results were compared to previously generated results on these 30 samples to evaluate performance and concordance. Finally, a subset of analysts was tasked with completing two sequencing runs (one with manually prepared libraries and one with libraries prepared on the Ion Chef) without the direct supervision of the training coordinator. Again, results were compared to previously generated results to assess performance and concordance. Both of these studies yielded results concordant to data previously generated on an orthogonal MPS platform (MiSeq). Training materials used and developed by UNTCHI will be made available to those interested in leveraging UNTCHI's experience.

The validation process for the Precision ID mtDNA Whole Genome Panel, Ion Chef, and Ion S5 included a series of studies that adhered to current SWGDAM validation guidelines [21]. These validation studies included: population studies to evaluate the performance on samples from a wide array of lineages; evaluations of sensitivity and stochastic effects to demonstrate the dynamic range, ideal target range, and limit of detection; reproducibility and repeatability studies to assess the repeatability of results generated by the same operator and the reproducibility of results generated by different operators; a contamination assessment for detection of exogenous DNA in

controls and known samples; evaluation of known and non-probative evidence samples to determine performance with potentially challenging samples; and mixtures to evaluate the capability of the system to detect and resolve mixed haplotypes (even though the primary focus of this validation was for single-source samples). Data from these validation studies support that this MPS workflow yields reliable results for the analysis of biological evidence. The complete set of population data and information on the relative performance of each amplicon in the Precision ID mtDNA Whole Genome Panel can be found in Strobl et al [22]. The remaining validation data will be submitted for publication in the near future.

Next, data and analysts' experience from the training and validation runs were used to draft working SOPs, develop workflow considerations, and generate worksheets for documentation while completing lab work. The working SOPs and worksheets are included in the more detailed Final Technical Report. These resources will be made available to those interested in leveraging UNTCHI's experience and intend to implement mtGenome/MPS analyses into their own laboratories.

The many training and validation runs completed during this implementation process allowed for the development of appropriate Quality Assurance practices. These guidelines included instrument and network maintenance, data storage procedures, inclusion of appropriate controls, evaluation of no template controls (NTCs) and bioinformatic negatives, and a procedure for addressing nuclear mitochondrial DNA segment (NUMTs) that create signal noise when sequencing the mtGenome. Discussions about sequencing artefacts and NUMTs are included in Strobl et al [22]. The process of developing thresholds for analysis of the MPS generated mtGenome data and additional details for addressing NUMTs will be submitted for publication in the near future.

Concordance was evaluated between data generated with the MPS workflow, consisting of the Precision ID mtDNA Whole Genome Panel, Ion Chef, and Ion S5, and UNTCHI's currently used Sanger sequencing methodology. The samples evaluated included hairs, bones, family reference samples, and staff samples. The data were concordant between the common areas sequenced supporting that this MPS workflow yields reliable results for the analysis of biological evidence. The additional information afforded by MPS of the entire mtGenome (approximately 75% of mtGenome variation resides in the coding region [5]) highlighted the potential for an increase in discrimination power and phylogenetic resolution for the samples analyzed. Information specifically addressing concordance with Sanger sequencing results and the evaluation of evidence samples was included in an abstract and submitted for presentation at ISHI in 2019.

There was a six-month delay during this project because the manufacturer redesigned the chemistry. Because of that delay UNTCHI is slightly behind in implementing the MPS workflow for analysis of forensic biological evidence. Implementation is expected in the next quarter.

References

1. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74(12):5463-5467, 1977.
2. Wilson MR, Polanskey D, Butler J, DiZinno JA, Replogle J, Budowle B. Extraction, PCR amplification, and sequencing of mitochondrial DNA from human hair shafts. *BioTechniques* 18:662-669, 1995.
3. Wilson MR, DiZinno JA, Polanskey D, Replogle J, Budowle B. Validation of mitochondrial DNA sequencing for forensic casework analysis. *Int. Journal Leg. Med.* 108:68-74, 1995.
4. Just RS, Scheible MK, Fast SA, Sturk-Andreaggi K, Rock AW, Bush JM, Higginbotham JL, Peck MA, Ring JD, Huber GE, Xavier C, Strobl C, Lyons EA, Diegoli TM, Bodner M, Fendt L, Kralj P, Nagl S, Niederwieser D, Zimmermann B, Parson W, Irwin JA. Full mtGenome reference data: development and characterization of 588 forensic-quality haplotypes representing three U.S. populations. *Forensic Sci. Int. Genet.* 14:141–55, 2015.
5. King JL, LaRue BL, Novroski N, Stoljarova M, Seo SB, Zeng X, Warshauer D, Davis C, Parson W, Sajantila A, Budowle B. High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. *Forensic Sci. Int. Genet.* 12:128-135, 2014.
6. Park S, Cho S, Seo HJ, Lee JH, Kim MY, Lee SD. Entire mitochondrial DNA sequencing on massively parallel sequencing for the Korean population. *J. Korean Med. Sci.* 32:587–92, 2017.
7. Churchill JD, Stoljarova M, King JL, Budowle B. Massively parallel sequencing-enabled mixture analysis of mitochondrial DNA samples. *Int. J. Leg. Med.*132:1263-72, 2018.
8. Holland MM, McQuillan MR, O'Hanlon KA. Second generation sequencing allows for mtDNA mixture deconvolution and high resolution detection of heteroplasmy. *Croat. Med. J.* 52:299–313, 2011.
9. Kim H, Erlich HA, Calloway CD. Analysis of mixtures using next generation sequencing of mitochondrial DNA hypervariable regions. *Croat. Med. J.* 56:208–17, 2015.
10. Seo SB, King J, Warshauer D, Davis C, Ge J, Budowle B. Single nucleotide polymorphism typing with massively parallel sequencing for human identification. *Int. J. Leg. Med*. 127(6):1079-1086, 2013.
11. King JL, Sajantila A, Budowle B. mitoSAVE: mitochondrial sequencing analysis of variants in excel. *Forensic Sci. Genet. Int.* 12:122-125, 2014.
12. Budowle B, Connell ND, Bielecka-Oder A, Colwell RR, Corbett CR, Fletcher J, Forsman M, Kadavy DR, Markotic A, Morse SA, Murch RS, Sajantila A, Schmedes SE, Ternus KL, Turner SD, Minot S. Validation of high throughput sequencing and microbial forensics applications. *BMC Invest. Genet.* 5:9, 2014.
13. Seo SB, Zeng X, King JL, Larue BL, Assidi M, Al-Qahtani MH, Sajantila A, Budowle B. Underlying data for sequencing the mitochondrial genome with the massively parallel sequencing platform Ion Torrent™ PGM™. *BMC Genomics* 16 Suppl 1:S4, 2015.

14. Davis C, Peters D, Warshauer D, King J, Budowle B. Sequencing the hypervariable regions of human mitochondrial DNA using massively parallel sequencing: Improved methods for DNA samples encountered in forensic testing. *Leg. Med*. 17(2):123-127, 2015.
15. Churchill JD, Chang J, Ge J, Rajagopalan N, Lagacé R, Liao W, King JL, Budowle B. Blind study evaluation illustrates utility of the Ion PGM™ System for use in human identity DNA typing. *Croat. Med. J.* 56(3):218-229, 2015.
16. Stoljarova M, King JL, Takahashi M, Aaspõllu A, Budowle B. Whole mitochondrial genome genetic diversity in the Estonian Population sample. *Forensic Sci. Genet. Int.* 130(1):67-71, 2016.
17. Ambers A, Turnbough M, Benjamin R, Gill-King H, King J, Sajantila A, Budowle B. Modified DOP-PCR for improved STR typing of degraded DNA from human skeletal remains and bloodstains. *Leg. Med*. 22:54-63, 2016.
18. Churchill JD, King JL, Chakraborty R, Budowle B. Effects of the Ion PGM™ Hi-Q™ sequencing chemistry on sequence data quality. *Int. J. Leg. Med*. 130(5):1169-80, 2016.
19. Lindberg MR, Schmedes SE, Hewitt C, Ternus K, Kadavy D, Budowle B. A Comparison and integration of MiSeq and MinION platforms for sequencing single source and mixed mitochondrial genomes. *PLoS ONE* 11(12):e0167600, 2016.
20. Thermo Fisher Scientific. Precision ID mtDNA Panels with the HID Ion S5/HID Ion GeneStudio S5 System Application Guide. 2019.
21. Scientific Working Group on DNA Analysis Methods (SWGDAM). SWGDAM Validation Guidelines for DNA Analysis Methods. https://docs.wixstatic.com/ugd/4344b0_813b241e8944497e99b9c45b163b76bd.pdf, 2016.
22. Strobl C, Churchill Cihlar J, Lagacé R, Wootton S, Roth C, Huber N, Schnaller L, Zimmermann B, Huber G, Hong SL, Moura-Neto R, Silva R, Alshamali F, Souto L, Anslinger K, Egyed B, Jankova-Ajanovska R, Casas-Vargas A, Usaquén W, Silva D, Barletta-Carrillo C, Tineo DH, Vullo C, Würzner R, Xavier C, Gusmão L, Niederstätter H, Bodner M, Budowle B, Parson W. Evaluation of mitogenome sequence concordance, heteroplasmy detection, and haplogrouping in a worldwide lineage study using the Precision ID mtDNA Whole Genome Panel. *Forensic Sci. Int. Genet.* 2019; Submitted.

Products Produced

In addition to the research results obtained, numerous presentations and peer-review published papers have been produced that document the work.

Presentations at National and International Meetings that were supported by this work

1. Budowle, B.: The adoption process of MPS into US forensic genetics laboratories: US perspective, Human Identification Solutions Conference, Vienna, Austria, 2017.
2. Churchill, J., Peters, D., Strobl, C., Parson, W., and Budowle, B.: Massively parallel sequencing (MPS) can be considered NGS, i.e., now generation sequencing: implementation of whole genome mitochondrial DNA sequencing into routine casework, 27th Congress of the International Society of Forensic Genetics, Seoul, South Korea, 8/2017.
3. Churchill, J.D., Peters, D., Capt, C., Strobl, C., Parson, W., and Budowle, B.: The road to implementation, 28th International Symposium on Human Identification, Seattle, WA, 10/2017.
4. Churchill JD, Takahashi M, Strobl C, Peters D, Capt C, Parson W, Budowle B. Validation of a Massively Parallel Sequencing Workflow for Mitochondrial DNA Analysis at UNTHSC Center for Human Identification for Missing Persons and Traditional Casework Analyses, 29th Annual International Symposium on Human Identification (ISHI), Phoenix, AZ, 9/2018.
5. Peters, D.: Missing Persons Update, Technical Leaders Meeting, 29th International Symposium on Human Identification, Phoenix, AZ, 9/2018.
6. Churchill Cihlar J, Strobl C, Peters D, Capt C, Parson W, Budowle B. The Implementation Process: A Massively Parallel Sequencing Workflow for Mitochondrial DNA Analaysis. Green Mountain DNA Conference, Burlington, VT, 7/2019
7. Churchill Cihlar J, Peters D, Capt C, Strobl C, Parson W, Budowle B. Massively Parallel Sequencing of the Mitochondrial Genome in Casework-type Samples. 30th Annual International Symposium on Human Identification (ISHI), Palm Springs, CA, 9/2019 (Abstract Submitted)

Publications

Churchill JD, Peters D, Capt C, Strobl C, Parson W, and Budowle B. Working towards implementation of whole genome mitochondrial DNA sequencing into routine casework. Forensic Science International: Genetics Supplement Series 6: e388–e389, 2017.

Strobl C, Churchill Cihlar J, Lagacé R, Wootton S, Roth C, Huber N, Schnaller L, Zimmermann B, Huber G, Hong SL, Moura-Neto R, Silva R, Alshamali F, Souto L, Anslinger K, Egyed B, Jankova-Ajanovska R, Casas-Vargas A, Usaquén W, Silva D, Barletta-Carrillo C, Tineo DH, Vullo C, Würzner R, Xavier C, Gusmão L, Niederstätter H, Bodner M, Budowle B, Parson W. Evaluation of mitogenome sequence concordance, heteroplasmy detection, and haplogrouping in a worldwide lineage study using the Precision ID mtDNA Whole Genome Panel. *Forensic Sci. Int. Genet.* 2019; Submitted.

Additional manuscripts are in process to describe the large amount of data generated during this project.

Invention Report

No patents were submitted related to this project

Participating Scientists and Collaborators

What individuals have worked on the project?

Name: Bruce Budowle
Project role: PI
Nearest person month worked: 1 year
Contribution to Project: Dr. Budowle has provided the overall direction and management of the project as well as participating in the technical research and data analysis.

Name: Jennifer Churchill Cihlar
Nearest person month worked: 1 year
Contribution to Project: Jennifer Churchill Cihlar is a Research Assistant Professor performing the main effort of the work on validation, training, and implementation guidance of mtGenome MPS into casework.

Name: Jonathan King
Nearest person month worked: 1 year
Contribution to Project: Mr. King is the laboratory manager performing daily analysis and evaluation of supporting data accumulation and will be assisting in bioinformatics support.

What other organizations have been involved as partners? none
Have other collaborators or contacts been involved? no

Impact

a. What is the impact of the project on the criminal justice system?

Due to the increased power of discrimination and phylogenetic resolution afforded by being able to analyze the entire mtGenome with the same amount of input DNA as current Sanger sequencing affords, this MPS-based workflow will increase the utility of mtDNA analyses of forensic biological evidence. The outcome is that more biological evidence will be analyzed successfully, which in turn will result in more and better investigative leads to solve crimes and identify human remains. The training materials and supporting data generated during these implementation efforts will be made available to assist other potential users of MPS technologies. In addition, this work will define the advantages and limitations of MPS which policy and decision makers can use for long term planning and funding directions. Lastly, although the focus of the work is on single-source samples, the foundations laid will allow exploitation of the quantitative nature of MPS so that mtDNA may in the near future be used to characterize mixture samples.

b. How has it contributed to crime laboratories?

While the contribution to crime laboratories cannot be measured yet, the overall outcome of this effort is that a more effective tool for characterizing mtDNA evidence is now available. Properly designed amplification panels and MPS technologies now make it feasible for crime laboratories to analyze the entire mtGenome with the same amount of input DNA as current Sanger sequencing methodologies afford, but with the benefit of an increased power of discrimination. UNTCHI's validation and implementation of this MPS workflow will serve as a demonstration of the utility of MPS to other labs that may want to implement MPS technologies or send evidence to UNTCHI to be processed. In addition, training materials will be available for the community to facilitate implementation in those laboratories interested in MPS and mtGenome analyses.

c. What is the impact on technology transfer?

The training materials, experience, and data developed during these implementation efforts will be available for the community to facilitate implementation in those laboratories interested in MPS and mtGenome analyses.

Changes/Problems

There were no challenges that could not be met. The only changes were a no cost extension to ensure that the work could be completed. In late 2017, Thermo Fisher Scientific announced their Ion 520 & Ion 530 Kit-Chef, which is used for templating on the Ion Chef, would be discontinued in 2018. This kit was necessary for the workflow that UNTCHI was validating and implementing. The company's subsequent process of testing which product would be best to replace this kit for Human Identification applications, commercially releasing the new product, and manufacturing limited quantities of this kit in the first few quarters resulted in an approximately six-month delay in UNTCHI's progress of validating and implementing its MPS workflow for whole genome mitochondrial DNA sequencing. During the time period that UNTCHI was unable to acquire reagents for validation studies, time and effort was focused on training and developing worksheets and SOPs.

Proprietary Information

There was no proprietary information related to this work.