



Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: [www.elsevier.com/locate/FSIGSS](http://www.elsevier.com/locate/FSIGSS)



## A custom software solution for forensic mtDNA analysis of MiSeq data

Mitchell Holland\*, Jennifer McElhoe

Forensic Science Program, Pennsylvania State University, University Park, PA 16802, United States

### ARTICLE INFO

Article history:  
Available online xxx

Keywords:  
NextGENe®  
Heteroplasmy  
Length-variants  
DNA motifs

### ABSTRACT

The use of a next generation sequencing (NGS) approach for mitochondrial (mt) DNA analysis in forensic casework is imminent. A feature of the NGS approach is the ability to detect and report mtDNA heteroplasmy, which will significantly enhance the discrimination potential of the testing method. A single software solution that allows for robust and user-friendly analysis of NGS-derived mtDNA sequence is not readily available, especially when considering heteroplasmy. This communication outlines the desired features of a software package for forensic applications, and the progress made towards those aims with the development of a custom version of NextGENe®.

© 2015 Elsevier Ireland Ltd. All rights reserved.

### 1. Introduction

With the development of next generation sequencing (NGS) approaches, and the increased resolution provided by this technology, the forensic community is presented with an opportunity to identify, report, and take advantage of mtDNA heteroplasmic variants observed in casework [1,2]. A number of software tools have been developed for the analysis of mtDNA sequence data generated on an NGS platform. Some are simple Excel-spreadsheet-based solutions [3] that help convert data into haplotypes using phylogenetically-derived nomenclature, and allow for consistent assignment of homopolymeric stretches and sequence motifs [4,5], but do not allow for analysis of the individual sequencing reads; commonly referred to as the *pileup*. Others involve the reconstruction of mtgenomes from total genomic DNA [6], which is currently beyond the scope of forensic applications. Still others have toolboxes from which pipelines can be created to assist in the process of data management [7–9], but lack the proper nomenclature conversion, analysis of the *pileup*, and can be challenging to use.

The commercially available software package, NextGENe® (SoftGenetics, Inc.) is user friendly, includes numerous user defined parameters, and allows for detailed analysis of the *pileup*. It has been used successfully to analyze mtDNA sequence data on various NGS platforms [1,2], but initially lacked some desired

features. Those features include: (1) alignment to a circular version of the mtgenome so that data properly spans the transition point in the mtgenome numbering system [10], (2) alignment and nucleotide numbering consistent with the revised mtgenome sequence [11], (3) recognition and proper assignment of motifs and insertions/deletions (INDELs) consistent with phylogenetic and forensic considerations [4,5], (4) robust identification of heteroplasmic sequences, and (5) export of reports that address forensic considerations and allow for seamless import into tertiary analysis tool such as the new version of EMPop; [www.empop.org](http://www.empop.org), v3/R11. This communication reports on progress made towards the development of a custom, forensic version of NextGENe® that includes the features above.

### 2. Materials and methods

Data sets containing hundreds of mtDNA sequences from random individuals in the population have been analyzed in our laboratory with NextGENe®. As new versions have been developed, mtDNA data sets were run through the software to ensure that the outcomes were concordant with previous analyses. The most recent, commercially available, version of NextGENe® being used in our laboratory is v2.4.0.1, which was brought online in June 2015. Alpha versions (currently v2.4.2) of the custom software addressing the needs of the forensic mtDNA community have been tested in our laboratory on existing data sets. A beta version of the software is available by contacting SoftGenetics, Inc.

\* Corresponding author. Fax: +1 814 863 8372.  
E-mail address: [mmh20@psu.edu](mailto:mmh20@psu.edu) (M. Holland).

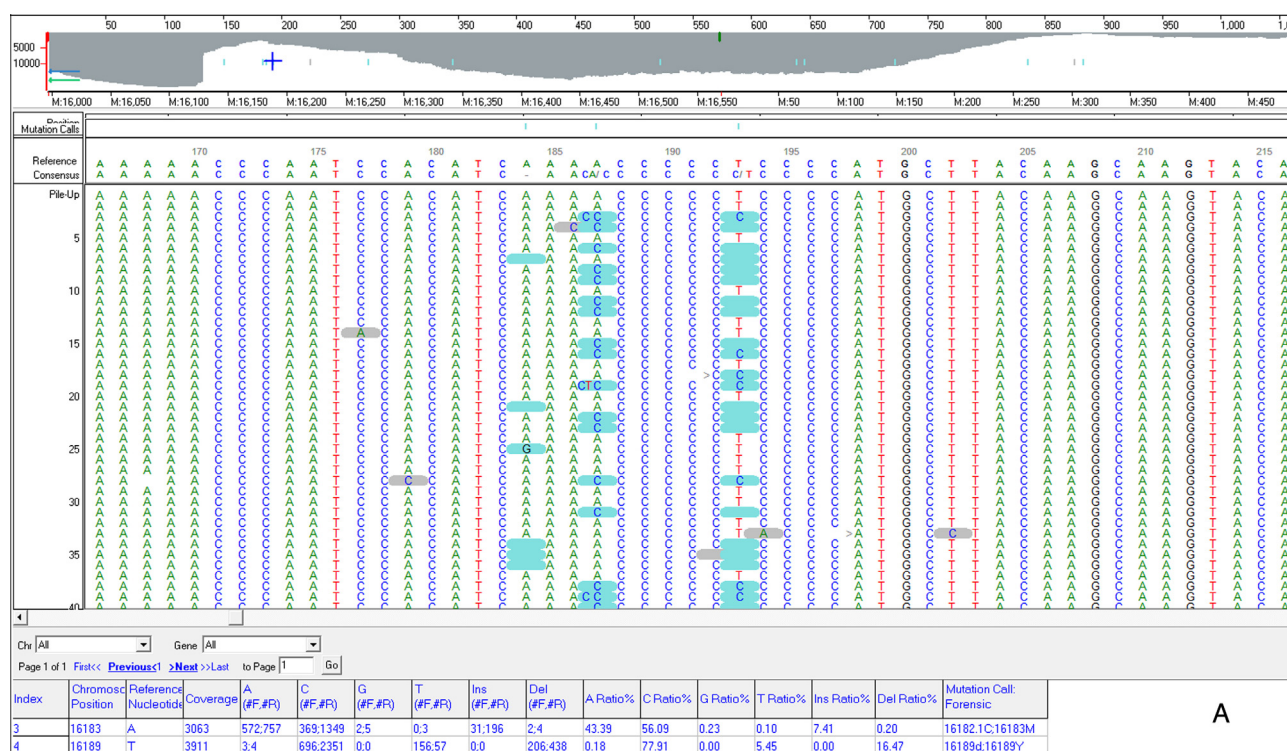


Fig. 1. Screenshots of the traditional (A) and beta (B) versions of NextGENe<sup>®</sup> software for the analysis of mtDNA NGS data.

### 3. Results and discussion

The haplotypes for all samples analyzed with new versions of the NextGENe<sup>®</sup> software have been concordant with previous results, and for those tested, are concordant with Sanger-derived haplotypes [7]. The newest forensic version of the software addresses the primary needs of the forensic community; alignment to a circular version of the mtgenome, appropriate use of nomenclature, alignment of INDELs, and reporting outputs. Fig. 1 illustrates some of these features, and compares the new interface to the traditional version of the software.

The traditional version of NextGENe<sup>®</sup> displays three different numbering systems (Fig. 1A), none of which are numbered according to [11]. The region of sequence being analyzed includes familiar polymorphic sites; A16183C and T16189C. However, the calls reported in the table below the pileup are incorrect; 16.182.1C, 16.183 M, 16.189d, 16.189Y. This is due to the misalignment of this region of sequence. On the contrary, in Fig. 1B, the top strand is the revised reference sequence with the correct numbering system. The second strand is a major haplotype consensus sequence that can be imported as a sequence string for searching purposes in EMPOP. The data is represented as hash-marks, instead of nucleotide calls, and is clearly better aligned to the reference sequence. This is illustrated in the table, as the polymorphic sites are correctly identified. As with the traditional version, metadata such as read coverage for the pileup, distribution of reads across all potential nucleotides, and an assessment of INDELs is provided. The user is able to export reports containing the major allele consensus sequence string, haplotype information, and heteroplasmy information. The generation of these reports streamlines the analysis workflow for the user and removes the introduction of transcription errors.

### 4. Conclusions

The introduction of an NGS approach for forensic mtDNA sequence analysis will require the use of a software package that enables the examiner to easily navigate through the data, reliably report the findings, and use the outputs to effectively run database searches. A custom, forensic version of NextGENe<sup>®</sup> has been developed and was evaluated to assess whether it meets these goals, which it does; a beta version of the software is available through SoftGenetics, Inc. The software remains in development, and will require additional iterations that address common bugs and refinements that address the needs of the forensic community. A more thorough analysis of the commercially available version of the new software will be the subject of a future communication.

### Conflict of interest

The authors of this article have no relevant financial relationships with commercial interests to disclose.

### Acknowledgements

The authors gratefully acknowledge the support of Jonathan Liu, John McGuigan and John Fosnacht from SoftGenetics for development of new versions of NextGENe<sup>®</sup>. This work was supported in part by grant 2014-DN-BX-K022 from the National Institute of Justice (NIJ). The points of view in this document are those of the authors and do not represent the official position or policies of the U.S. Department of Justice.

B

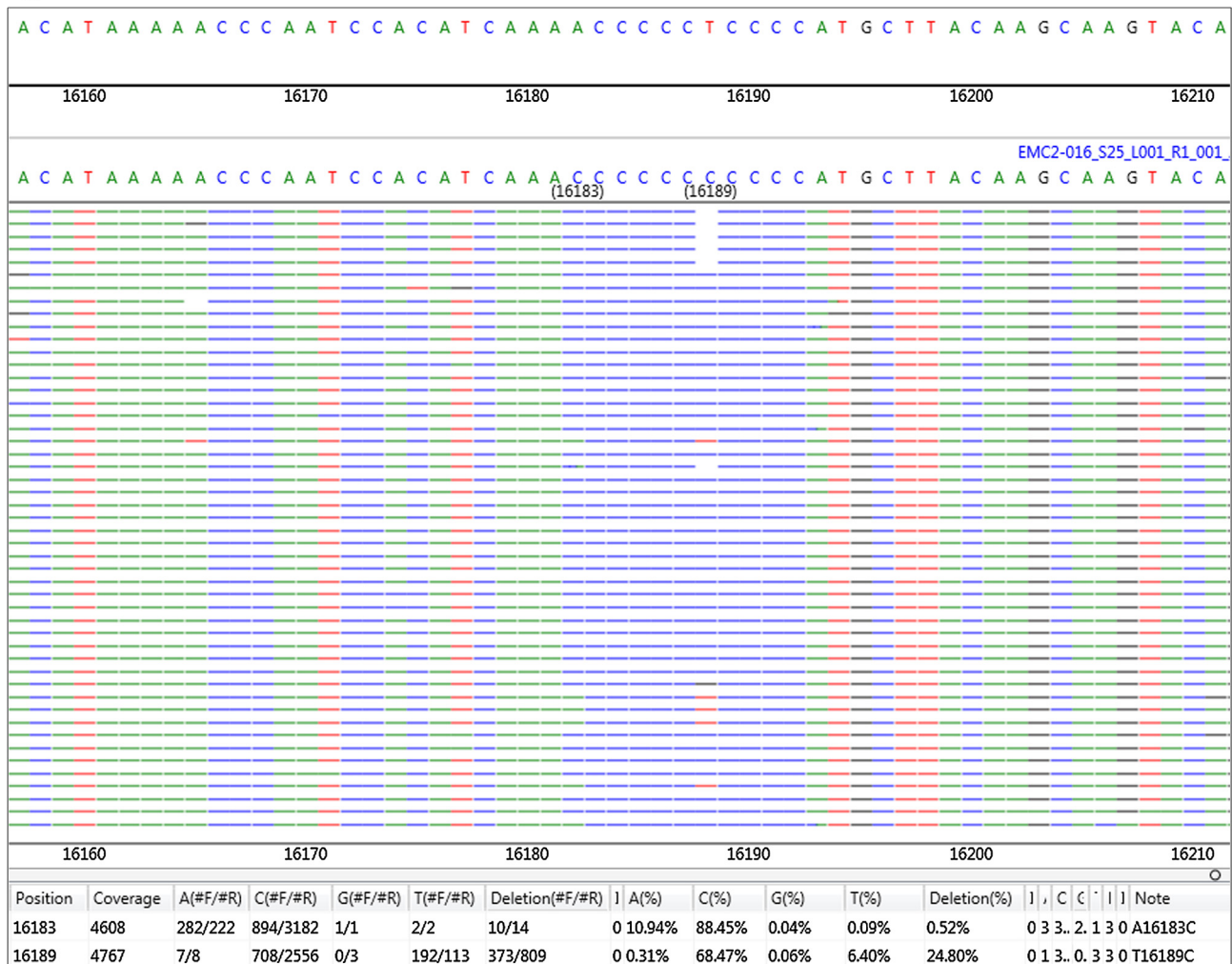


Fig. 1. (Continued)

## References

- [1] M.M. Holland, M.R. McQuillan, K.A. O'Hanlon, Second generation sequencing allows for mtDNA mixture deconvolution and high resolution detection of heteroplasmy, *Croat. Med. J.* 52 (2011) 299–313.
- [2] J. McElhoe, M. Holland, K. Makova, et al., Development and assessment of an optimized next-generation DNA sequencing approach for the mtgenome using the Illumina MiSeq, *Forensic Sci. Int. Genet.* 13 (2014) 20–29.
- [3] J.L. King, A. Sanjantila, B. Budowle, mitoSAVE: mitochondrial sequence analysis of variants in Excel, *FSIG 12* (2014) 122–125.
- [4] W. Parson, L. Gusmao, D.R. Hares, et al., DNA Commission of the International Society for Forensic Genetics: revised and extended guidelines for mitochondrial DNA typing, *Forensic Sci. Int. Genet.* 13 (2014) 134–142.
- [5] H.-J. Bandelt, W. Parson, Consistent treatment of length variants in the human mtDNA control region: a reappraisal, *Int. J. Legal Med.* 122 (2008) 11–21.
- [6] L. Hahn, Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach, *Nucl. Acids Res.* 29 (2013) 1–9.
- [7] B. Rebolledo-Jaramillo, M.S.-W. Su, N. Stoler, et al., Maternal age effect and severe germ-line bottleneck in the inheritance of human mitochondrial DNA, *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014) 15474–15479.
- [8] C. Calabrese, D. Simone, M.A. Diroma, et al., MToolBox: a highly automated pipeline for heteroplasmy annotation and prioritization analysis of human mitochondrial variants in high-throughput sequencing, *Bioinformatics* 30 (2014) 3115–3117.
- [9] J. Y. Guo, C. Li Li, et al., MitoSeek: extracting mitochondria information and performing high-throughput mitochondria sequencing analysis, *Bioinformatics* 29 (2013) 1210–1211.
- [10] S. Anderson, A.T. Bankier, B.G. Barrell, et al., Sequence and organization of the human mitochondrial genome, *Nat. Genet.* 290 (1981) 457–465.
- [11] R.M. Andrews, I. Kubacka, P.F. Chinnery, et al., Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, *Nat. Genet.* 23 (1999) 147.