

# Resolving a human identification case for the Rio de Janeiro Police with massively parallel sequencing of mtDNA using a proposed pipeline

C. Bottino<sup>1,2</sup>, R. Silva<sup>3</sup> and R.S. Moura-Neto<sup>4</sup>

<sup>1</sup> Instituto de Pesquisa e Perícias em Genética Forense, Brasil

<sup>2</sup> Instituto Nacional de Metrologia, Qualidade e Tecnologia, Brasil

<sup>3</sup> Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Brasil

<sup>4</sup> Instituto de Biologia, Universidade Federal do Rio de Janeiro, Brasil

Corresponding author: R.S. Moura-Neto

E-mail: rodrigomouraneto@ufrj.br

Genet. Mol. Res. 20 (1): gmr18757

Received October 25, 2020

Accepted March 10, 2021

Published March 31, 2021

DOI <http://dx.doi.org/10.4238/gmr18757>

**ABSTRACT.** The forensic community is concerned about the quality control of mitotypes reported, as probability estimates may be biased to the disadvantage of a suspect when there is incomplete data. This is particularly important when using new technology such as Massively Parallel Sequencing (MPS), which performs simultaneous analysis of millions of DNA fragments, generating vast amounts of data from small amounts of sample, providing more extensive genetic information. The use of MPS for complex forensic evidence analysis could save time and resources, and provide reliable data for extremely degraded DNA samples. However, data analysis methods can be challenging to implement in a forensic labs routine. Here, we propose a reliable and straightforward pipeline for mtDNA data analysis in forensic samples using MPS, and we used this pipeline to solve a human identification case by Rio de Janeiro Police who had recovered a severely putrefied body that could not be identified by conventional autosomal STR analysis. Mitochondrial DNA (mtDNA) from samples were sequenced on Ion Torrent PGM. Raw data from the sequencing was analyzed in two main pipeline protocols: quality control and forensic analysis. The first includes the analyses of raw

data and contamination control. The second generates haplotypes, allowing haplogroup classification, and biogeographic inference, as well as verification of a matrilineal relationship. This straightforward freeware pipeline for mtDNA data analysis should facilitate the implementation of this type of system in forensic lab routine. Using our pipeline, the definition of the samples being of African origin was more precise when considering the whole mtGenome and not only the control region, which is usually used for routine forensic analysis. The final analysis was consistent with the existence of a matrilineal relationship between the alleged son (bone sample recovered from the sea) and the mother.

**Key words:** Forensic samples; Analysis workflow; Bioinformatics; Corpse identification

## INTRODUCTION

Massively Parallel Sequencing (MPS) performs a simultaneous analysis of millions of DNA fragments, generating vast amounts of data from small amounts of sample. It also allows analysis of more complete genetic information, such as polymorphisms not detected by conventional techniques (for example, low level heteroplasmies in mtDNA) (Scudder et al., 2018).

Recent reports the mtDNA analysis using MPS, evaluated not only the hypervariable regions (HV1, HV2, and HV3) but the whole mitochondrial genome (Parson et al., 2013; Just et al., 2015; Yao et al., 2018; Zhou et al., 2018). However, most of these studies were performed with mock case samples or a few forensic samples, using complex data analysis, which can be challenging to implement on forensic labs routine. Also, the forensic community is particularly concerned with the quality control of mitotypes reported, as probability estimates are usually biased to the disadvantage of a suspect when erroneous data are present (Huber et al., 2018). Data analysis methods are usually complex, which can be challenging to implement on forensic labs routine. Software and parameters are highly variable, including restricted software sometimes. In this work, it is presented a data analysis workflow simple to use and open access. All software and tools are available online or free for download. We offer a reliable straightforward freeware pipeline for mtDNA data analysis in forensic samples using MPS, useful to solve a human identification case on Rio de Janeiro Police, that was not resolved by conventional autosomal STR analysis.

## MATERIAL AND METHODS

In February 2015, a severely decomposed body was recovered from the seacoast of Rio de Janeiro. Police investigation led to a possible mother, but due to the body's advanced state of decomposition, DNA comparison was needed to confirm an identification. Buccal swab sample from the alleged mother was collected upon the signature of an informed consent form and was used as reference sample for comparison. The study was approved by an institutional ethics in human research committee.

Bone fragments from the body (alleged son) were powdered in liquid nitrogen using a SPEX 5500 Freezer Mill (SPEX, US). DNA from two grams of bone powder was prepared using phenol/chloroform (Comey et al., 1994), and a buccal swab reference sample was extracted using Chelex100® resin methods (Walsh et al., 1991). DNA samples were quantified with Quantifiler HP kit (Life Technologies) on ABI Prism 7500 System (Life Technologies), according to manufacturer's instructions. A Degradation Index (DI), based on the ratio between the quantifications of short and long DNA fragments was obtained. Values below 1 indicate a non-degraded sample; values between 1 and 10 indicate moderate degradation and values above 10 indicate severe degradation.

For STR analysis, ~0.5 ng of each sample was amplified using the PowerPlex Fusion kit (Promega Corporation) according to the manufacturer's instructions. Capillary electrophoresis was performed in an ABI 3500 instrument (Life Technologies, 36-cm array, POP-4, Data Collection v1.2, GeneMapper ID-X v1.4). Maternity Indexes were calculated based on Brazilian allelic frequencies published elsewhere (Aguiar et al., 2014; Hessab et al., 2015).

For mtDNA analysis, samples (~1 ng) were amplified using the Early Access Ampliseq™ Mitochondrial Panel kit (kindly provided by Thermo Fischer Scientific). This kit is an early version of the Precision ID mtDNA Whole Genome Panel (Thermo Fischer Scientific), which contains two primer pools, each one containing 81 primer pairs, that allows full mtDNA sequencing. The panel also offers two amplification protocols: Conservative (indicated for high-quality samples) and Full (indicated for degraded samples). Samples were amplified using both protocols.

Library preparation and quantification were performed using HID-Ion Ampliseq Library kit, Ion Xpress Barcode Adapters kit (Thermo Fisher Scientific), and Ion Library Quantitation kit (Thermo Fisher Scientific), on ABI Prism 7500 System (Life Technologies), followed by normalization for 20 pM and pooling. Emulsion PCR, performed with Ion PGM Hi-Q OT2 kit (Thermo Fisher Scientific) was done on Ion OneTouch 2 System (Thermo Fisher Scientific). Samples were sequenced on Ion Torrent PGM™, using a 318 v2 chip and the Ion PGM Hi-Q Sequencing kit (Thermo Fisher Scientific).

FastQ sequencing data reads with mean length of  $\geq 100$  bp and Q-value  $\geq 20$  are considered "pass", and a report for parameters, such as genome coverage, is generated. Reads were then, aligned to the reference genome rCRS\_plus\_80, a mtDNA genome (Accession # NC\_012920.1) that repeats the first 80 nucleotides in the end, to cover all the amplicons from a circular genome. Coverage Analysis was used to evaluate the minimum genome coverage necessary of  $\geq 80\%$ , including entire control region. Subsequently, File Exporter was used to generate files FastQ and Bam/Bai. Lastly, FastQ files were submitted to mtDNA-Server (<https://mtdna-server.uibk.ac.at/index.html>) to check for possible contaminations from others source of DNA. Once the analyses pass all criteria for QC, the following steps are needed for forensic analysis.

The next steps investigate the samples' haplotypes, haplogroup classification, and biogeographic inference, as well the matrilineal relationship. Variant Caller was used to generate VCF files to make a first haplotype assignment (haplotype #1). VCF files were submitted to mitoSAVE (available at <https://www.unthsc.edu/graduate-school-of-biomedical-sciences/molecular-and-medical-genetics/laboratory-faculty-and-staff/mitosave>) to make a second haplotype assignment (haplotype #2) using a minimum read depth of 40x

and a value of 0.1 heteroplasmy as parameters. Discrepancies were inspected and resolved manually using Integrated Genomics Viewer (IGV) software (available at <http://software.broadinstitute.org/software/igv/>) for a definitive haplotype. The final haplotype was submitted to EMPOP (<https://empop.online/>) (Parson & Dür, 2007) and HAPLOGREP (<http://haplogrep.uibk.ac.at/>) (Weissensteiner et al., 2016) to make the haplogroup assignment and biogeographic inference. Lastly, to exclude or not the existence of a maternal relationship, a comparison is made between reference and questioned samples. If no sequence difference is detected, the maternal relationship is stated, if only one nucleotide difference is observed, an inconclusive report is issued. If two or more differing nucleotides are found, no relationship report is released.

## RESULTS

The amount and quality of the DNA (gDNA) recovered from oral swabs were higher than 1.0 ng/uL, and the Degradation Index (DI), showed low levels of degradation (DI=0.59). For DNA preparation of bones, sample was highly degraded (3.9 ng/uL; DI = 204.49), anticipating a challenging analysis.

STR analysis performed on buccal swab resulted in a complete genetic profile. Nevertheless, for the bone sample, only partial profiles were obtained due to sample degradation (Figure S1). This partial profile provided low Maternity Indexes (Likelihood ratio of 6.5), which led to inconclusive results for the existence of a first-degree maternal relationship between the samples.

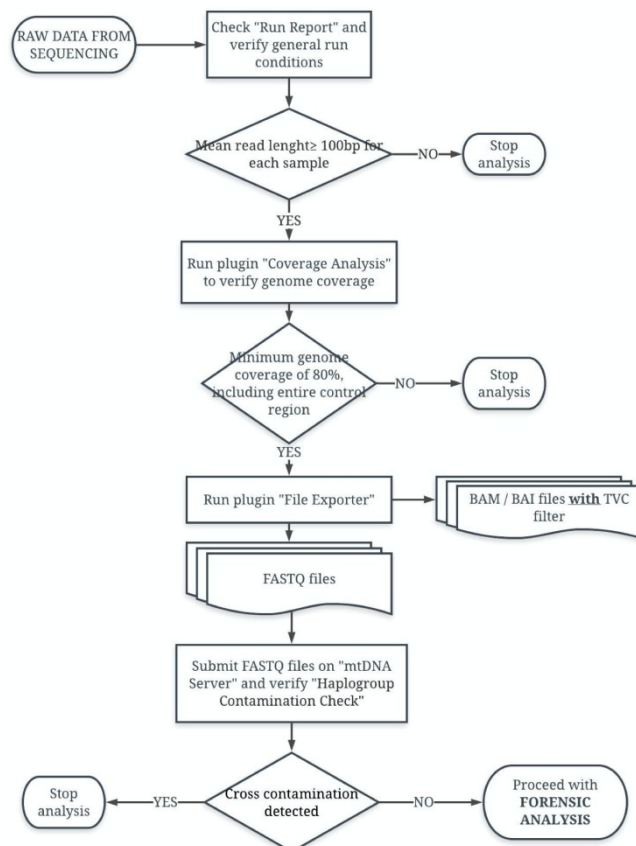
Complete mtDNA genome sequencing performed in bone and reference DNA samples was analyzed according to parameters determined in a two-step workflow. The first one (Figure 1) is dedicated to quality control of the sequencing process, such as number of reads and their average size. Also, minimum genome coverage and sample contamination check should be performed to proceed to the next step, which is the forensic interpretation (Figure 2).

Among all parameters described in the Sequencing Report, provided by the sequencing instrument, we considered a mean read length of at least 100bp for all samples to be the most important, as the maximum expected amplicon size is no longer than 120bp. This was obtained for both bone and reference samples (mean read length: 117bp). Genome coverage for mtDNA sequencing protocol, conservative and full, were comparable and above 90%. Cross-contamination check revealed no signs of contamination in any samples used in this study (data not shown), so we proceeded to forensic interpretation.

The two resulting haplotypes #1 and #2 obtained using both Variant Caller and software mitoSAVE had some inconsistencies, for instances, deletions appeared in positions 513 and 524, known as CA stretch (Figure S2), inspected and edited manually using IGV software. Nevertheless, it is recommended that such deletions in these positions should not be considered for the final haplotype since there is no biological relevance (Parson et al., 2013; Seo et al., 2015).

The final haplotypes obtained for the samples are shown in Table 1, as recommended by the International Society of Forensic Genetics for mtDNA data analysis and FBI SWGDAM recommendations (Parson et al., 2014; [www.swgdam.org/publications](http://www.swgdam.org/publications)). In bold are variant positions in the control region; heteroplasmies were not observed. Maternal relationship between samples could not be

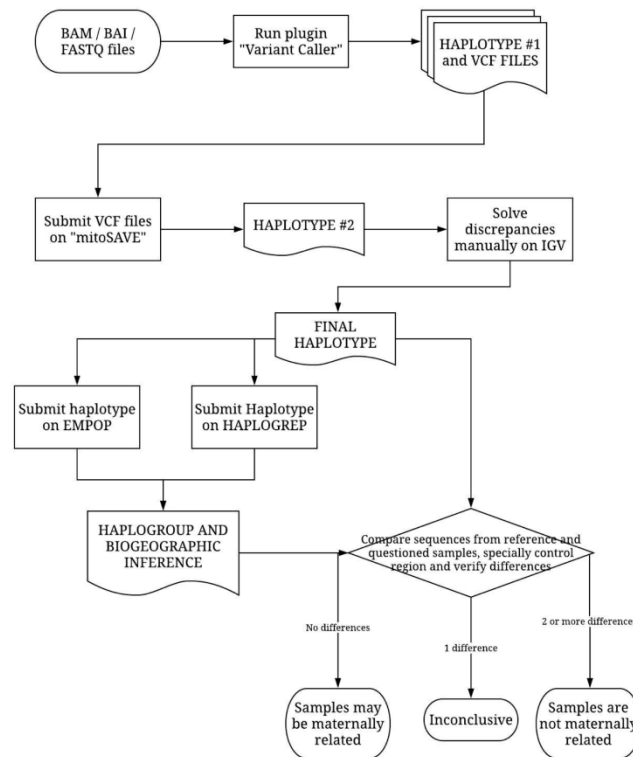
excluded by the comparison of the variants among putative mother and alleged son. The identity between them indicates they share the same matrilineal lineage.



**Figure 1.** First step of Data Analysis: Quality Control. Sequencing reports generated by the Ion Torrent Suite Server provide the Raw data and the alignments with the mtDNA reference genome - rCRS\_plus\_80. File Exporter was used with the TVC filter to download BAM/BAI and FASTQ files after primer trimming. Genome coverage was generated with the Coverage Analysis plugin. FASTQ files were submitted to mtDNA-Server to check for contamination. After all checking, files are ready for the next step of the analysis (Figure 2).

**Table 1.** Final haplotypes obtained for the two samples of this study. Polymorphisms in the control region are in bold type.

Case Sample	Haplotype
<i>Alleged Son (bone)</i>	73G 263G 750G 1438G 2706G 3450T 4769G 5773A 6221C 7028T 8701G
	8860G 9449T 9540C 10086G 10373A 10398G 10873C 11002G 11719A 12001T
	12705T 13105G 13790G 13914A 13980C 14766T 15301A 15311G 15326G
	15824G <b>16124C 16223T 16278T 16311C 16362C 16519C</b>
	73G <b>263G</b> 750G 1438G 2706G 3450T 4769G 5773A 6221C 7028T 8701G
<i>Mother (buccal swab)</i>	8860G 9449T 9540C 10086G 10373A 10398G 10873C 11002G 11719A 12001T
	12705T 13105G 13790G 13914A 13980C 14766T 15301A 15311G 15326G
	15824G <b>16124C 16223T 16278T 16311C 16362C 16519C</b>
	73G <b>263G</b> 750G 1438G 2706G 3450T 4769G 5773A 6221C 7028T 8701G
	8860G 9449T 9540C 10086G 10373A 10398G 10873C 11002G 11719A 12001T



**Figure 2.** Second step of Data Analysis: Forensic Analysis. Haplotype #1 and VCF files were generated with Variant Caller plugin. Haplotype #2 was generated with mitoSAVE after submission of VCF files. Discrepancies were manually resolved with IGV software. Final haplotype was submitted to EMPOP and HAPLOGREP to make haplogroup assignment and biogeographic inference.

Lastly, a phylogeographic analysis was performed (haplogroup classification) of the control region and the whole mtGenome from the samples of both cases, using online databases EMPOP and HAPLOGREP. The samples were defined as of African origin, and the haplogroup classification was concordance between mother/son (L3b1a). It is noteworthy that, for both databases, the inference was more precise when considering the whole mtGenome and not only the control region, as usually performed in routine forensic interpretation.

## DISCUSSION

Here we provide a systematic analysis of the mtDNA using a challenging forensic case of human identification, previously analyzed by Rio de Janeiro Police. Firstly, the conventional autosomal STR analysis with capillary electrophoresis showed inconclusive results. Due to the bones' conditions (exiguous and extremely degraded DNA), it was not possible to establish for sure the existence of a first-degree relationship (maternity) between alleged son and his mother.

We recommend minimum genome coverage of 90%, including the entire control region, to provide better information to compare samples. After haplotype definition and biogeographic inference, the existence of a maternal relationship between samples cannot be excluded. Although recent studies have been published focusing on mtDNA analysis based on Ion Torrent PGM, most of them were performed with mock case samples or just a few forensic samples (Parson et al., 2013; Just et al., 2015; Ambers et al., 2016; Yao et al., 2018; Zhou et al., 2018). So, this work has the value of demonstrating the applicability of the MPS technique on real forensic samples.

Nevertheless, some difficulties need to be pointed out. First, there is no consensus in the literature about the technique's sensibility, especially the threshold values that should be used. These are highly variable and sometimes defined arbitrarily. For read depth, values can vary between 10 and 100X, while for heteroplasmies detection values can range from 1 to 20% (King et al., 2014; Ambers et al., 2016; Churchill et al., 2016; Zhou et al., 2016; Churchill et al., 2018; Ma et al., 2018; Yao et al., 2018). In this work, we used average values (read depth of 40X; heteroplasmies = 10%); but with the spread of technique, a consistent validation study becomes more and more necessary.

Secondly, we need to take into account the performance of Ion Torrent PGM when sequencing homopolymeric regions, especially those that are cytosine-rich. This technical limitation is already well described and showed some improvement with the implementation of Hi-Q kits (Merriman et al., 2012; Bragg et al., 2013; Seo et al., 2015; Churchill et al., 2016). For human mtDNA, some regions are very well described as being challenging to sequence on Ion Torrent PGM due to the presence of such homopolymers: The C-stretches on HV1, HV2 and HV3 (positions 16184-16193; 303-315 and 568-573, respectively); the "AC-stretch" on HV3 (positions 515-524); and around positions 8284 and 11635 (Parson et al., 2013; Seo et al., 2015; Churchill et al., 2016; Zhou et al., 2016; Riman et al., 2017). Usually, it appears on sequencing as false deletions or an increase in strand bias. Some of these artifacts were observed on this study, so we recommend precaution when analyzing those regions. In the proposed workflow, doubtful results usually can be resolved at two points: on manual verification with IGV software; or with haplogroup assignment in EMPOP and/or HAPLOGREP.

Lastly, there is an ethical issue that needs to be pointed out; when sequencing the coding region of the mtDNA genome, there is a possibility to access phenotypic information, such the existence of gene variants associated with some diseases. The increasing use of MPS for forensic genetics may lead to discussions about the depth and range of information that can be accessed, which could conflict with an individual's right to privacy (Beyleveld, 1997; Wienroth et al., 2014). In this context, it is essential to establish how the phenotypical information will be collected, processed, and stored, to avoid ethical, legal, and social conflicts.

The results presented here, along the line of other previously published studies (Parson et al., 2013; Ambers et al., 2016; Zhou et al., 2016), show the applicability of mtDNA analysis with Ion Torrent PGM for forensic cases, particularly those involving challenging samples. Also, we present a reliable straightforward freeware pipeline for mtDNA data analysis in forensic samples using MPS, which should help facilitate implementation in forensic lab routine. Another advantage of our workflow is that data are confirmed more than once in subsequent steps, generating reliable results suitable to be presented in court.



## ACKNOWLEDGMENTS

We thank Claudio Nunes Pereira and César Félix Schmidt for technical support. The authors would like to thank Thermo Fischer Scientific for providing the Early Access mtDNA system and technical support. This work was supported by CAPES Pro-Forensic grant (Proc. n° 23038.006844/2014-46). RS was supported by CAPES Pro-Forensic Grant n° 23038.006844/2014-46, CNPq Grant n° 304.156/2016-7 and FAPERJ Grant n° 202.847/2018. RSMN was supported by CNPq Grant n° 312.905/2018-1 and FAPERJ Grant n° 202.532/2019.

## CONFLICTS OF INTEREST

The authors declare no conflict of interest.

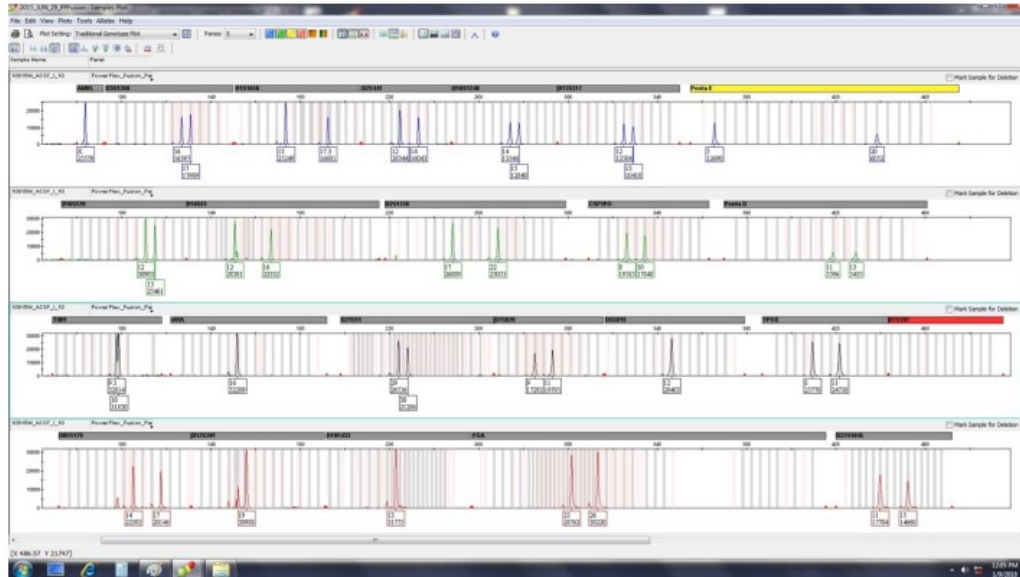
## REFERENCES

- Aguilar VRC, de Castro AM, Almeida VCO, Malta FSV, et al. (2014). New CODIS core loci allele frequencies for 96,400 Brazilian individuals. *Forensic Sci. Int. Genet.* 13: e6-e12.
- Ambers AD, Churchill JD, King JL, Stoljarova M, et al. (2016). More comprehensive forensic genetic marker analyses for accurate human remains identification using massively parallel DNA sequencing. *BMC Genomics.* Oct 17;17(Suppl 9): 750. Erratum in: *BMC Genomics.* 2017 Apr 20;18(1): 312.
- Beyleveld D (1997). Ethical issues in the forensic applications of DNA analysis. *Forensic Sci. Int.* 88: 3-15.
- Bragg LM, Stone G, Butler MK, Hugenholtz P, et al. (2013). Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLoS Comput. Biol.* 9: e1003031.
- Churchill JD, King JL, Chakraborty R and Budowle B (2016). Effects of the Ion PGM Hi-Q sequencing chemistry on sequence data quality. *Int. J. Legal Med.* 130: 1169-1180.
- Churchill JD, Stoljarova M, King JL and Budowle B (2018). Massively parallel sequencing-enabled mixture analysis of mitochondrial DNA samples. *Int. J. Legal Med.* 132: 1263-1272.
- Comey CT, Koons BW, Presley KW, Smerick JB, et al. (1994) DNA Extraction Strategies for Amplified Fragment Length Polymorphism Analysis. *J. Forensic Sci.* 39: 1254-1269.
- Hessab T, Carvalho RM, Souza M, Martha SFP, et al. (2015). Genetic data on 17 STR autosomal loci for a sample population of the State of Rio de Janeiro, Brazil. *Forensic Sci. Int. Genet.* 14: e4-e5.
- Huber N, Parson W and Dür A (2018). Next generation database search algorithm for forensic mitogenome. *Forensic Sci. Int. Genet.* 37: 204-214.
- Just RS, Irwin JA and Parson W (2015). Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. *Forensic Sci. Int. Genet.* 18: 131-139.
- King JL, LaRue BL, Novroski NM, Stoljarova M, et al. (2014). High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. *Forensic Sci. Int. Genet.* 12: 128-135.
- Ma K, Zhao X, Li H, Cao Y, et al. (2018). Massive parallel sequencing of mitochondrial DNA genomes from mother-child pairs using ion torrent personal genome machine (PGM). *Forensic Sci. Int. Genet.* 32: 88-93.
- Merriman B, Ion Torrent R&D Team and Rothberg JM (2012). Progress in Ion Torrent semiconductor chip-based sequencing. *Electrophoresis.* 33: 3397-3417.
- Parson W and Dür A (2007). EMPOP – a forensic mtDNA database. *Forensic Sci. Int. Genet.* 1: 88-92.
- Parson W, Gusmão L, Hares DR, Irwin JA, et al. (2014). DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing. *Forensic Sci. Int. Genet.* 13: 134-142.
- Parson W, Strobl C, Huber G, Zimmermann B, et al. (2013). Reprint of: Evaluation of next generation mtGenome sequencing using the Ion Torrent Personal Genome Machine (PGM). *Forensic Sci. Int. Genet.* 7: 632-639.
- Riman S, Kiesler KM, Borsuk LA and Vallone PM (2017). Characterization of NIST human mitochondrial DNA SRM-2392 and SRM-1292-I standard reference materials by next generation sequencing. *Forensic Sci. Int. Genet.* 29: 181-192.
- Scudder N, McNevin D, Kelty SF, Walsh SJ, et al. (2018). Massively parallel sequencing and the emergence of forensic genomics: Defining the policy and legal issues for law enforcement. *Sci. Just.* 58: 153-158.

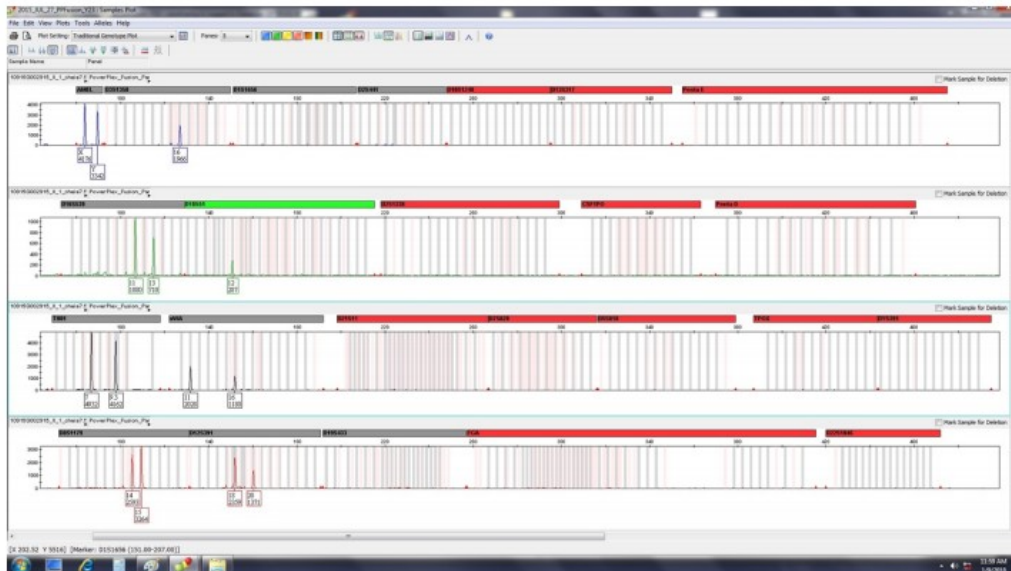


- Seo BS, Zeng X, King JL, Larue BL, et al. (2015). Underlying Data for Sequencing the Mitochondrial Genome with the Massively Parallel Sequencing Platform Ion Torrent PGM. *BMC Genomics*. 16 (Suppl 1): S4.
- Walsh PS, Metzger DA and Higuchi R (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques*. 10: 506-513.
- Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, et al. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res*. 8(44W1): W58-63.
- Wienroth M, Morling N and Williams R (2014). Technological Innovations in Forensic Genetics: Social, Legal and Ethical Aspects. *Recent Adv. DNA Gene Seq*. 8: 98-103.
- Yao L, Xu Z, Zhao H, Tu Z, et al. (2018). Concordance of mitochondrial DNA sequencing method on bloodstains using Ion PGM™. *Legal Med*. 32: 27-30.
- Zhou Y, Guo F, Yu J, Liu F, et al. (2016). Strategies for complete mitochondrial genome sequencing on Ion Torrent PGM™ platform in forensic sciences. *Forensic Sci. Int. Genet*. 22: 11-21.

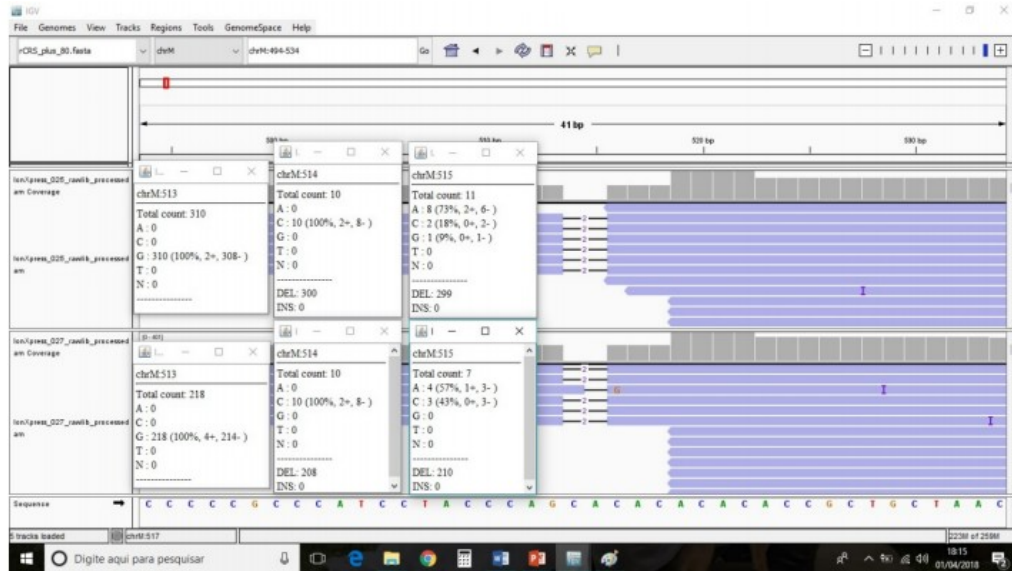
(A) Mother



(B) Alleged Son (Bone)



**Figure S1.** Electropherograms of samples. DNA was amplified with Promega's PowerPlex Fusion, and analyzed at ABI 3500, POP 4 polymer, 36-cm capillary, GeneMapper ID-X v.1.4 software. (A) DNA from the Mother; (B) Bone DNA from the alleged son.



**Figure S2.** False deletions observed on IGV between positions 513 and 515 (CA-stretch). Ion Torrent PGM sequencing technology cannot discriminate cytosine-rich regions successfully, leading to artifacts, such as false deletions.