Chapter 12

Analysis of Mitochondrial Control Region Using Sanger Sequencing

David Ballard

Abstract

The analysis of mitochondrial DNA (mtDNA) is an established forensic tool and has been used extensively to aid with both the identification of human remains and evidence recovered from scenes of crime. The biology of mtDNA confers both advantages and disadvantages when using it as a tool for identification. It benefits from a high copy number, which facilitates analysis from samples with highly degraded DNA or trace amounts of DNA, but the maternal mode of inheritance restricts its power of discrimination. With Next Generation Sequencing being used in research and some forensic casework laboratories the scope of mtDNA analysis in forensic casework may expand in the near future. Currently, however, most casework laboratories rely on Sanger sequencing and an established method for analyzing the hypervariable sequence regions is described.

Key words Mitochondrial DNA, Hypervariable sequence region I (HVS-I), Hypervariable sequence region II (HVS-II), Hypervariable sequence region III (HVS-III)

1 Introduction

Mitochondria are distinct organelles within the cell cytoplasm that play a crucial role in cellular respiration. Mitochondria have their own DNA content independent of the cell's nuclear DNA, comprising of approximately 16,569 base pairs arranged in a circular double stranded configuration $[1]$. The first mitochondrial sequence was published in [1](#page-11-0)981 by Anderson et al. [1], and can be found referred to as either the Anderson sequence or Cambridge Reference Sequence (CRS) while a subsequent reanalysis in 1999 resulted in 11 minor changes to this reference sequence. All human mitochondrial genomes are now compared with this Revised Cambridge Reference Sequence (rCRS) [\[2](#page-11-0)].

Unlike nuclear DNA, which is generally present in cells in just two copies (one copy of each chromosome from each parent), the copy number of mitochondrial DNA is much higher. In 1991, Satoh and Kuroiwa [3] experimentally determined that each

William Goodwin (ed.), *Forensic DNA Typing Protocols*, Methods in Molecular Biology, vol. 1420, DOI 10.1007/978-1-4939-3597-0_12, © Springer Science+Business Media New York 2016

mitochondrion within a cell contained 1–15 copies of the mitochondrial genome, at an average of 4–5 copies. Additionally each cell contains multiple mitochondria, giving an average copy number of 200–1700 [4] mitochondrial genomes per cell, although this is known to be tissue specific, varying between 50 and 75 copies in a spermatozoon $\lceil 5 \rceil$ up to the order of 100,000 in a maturing primary oocyte $[6]$. The presence of multiple copies of the mitochondrial genome helps to explain why it can often be analyzed in situations where nuclear DNA has degraded (e.g., ancient bones). Furthermore, the cellular location of the mitochondrial DNA (in the mitochondria organelles) also has a strikingly protective function in regard to DNA degradation [\[7](#page-12-0)].

In a forensic scenario, mitochondrial analysis is most commonly performed in cases where nuclear DNA is unlikely to be present, notably in analysis of shed hairs and bones. Given the strictly matrilineal inheritance of the mitochondrial genome $[5, 8]$ $[5, 8]$ $[5, 8]$, it can also be used to help resolve complex relationship cases where there is a common maternal link.

The more bases within the mitochondrial genome that are analyzed, the more discriminating the results will be. The advent of newer, so called, next generation sequencing technologies (e.g., Illumina's MiSeq or Life Technologies' Ion PGM) means that it is now relatively easy to sequence the entire mitochondrial genome of an individual, for example, using the PCR strategy described by Lyons et al. [9] (also *see* Chapter [12](http://dx.doi.org/10.1007/978-1-4939-3597-0_12)). Amplification of such long PCR amplicons relies on having reasonable quality DNA; however, this is still a feasible approach for some case samples, in addition to being a valid strategy when analyzing reference samples or for use in complex relationship cases.

A disproportionate amount of variation is centered within the 1.1 kb control region of the mitochondrial DNA and is even more concentrated around three areas within this control region termed hypervariable region I (HVS-I), hypervariable region II (HVS-II), and hypervariable region III (HVS-III). This is the area most traditionally sequenced for forensic analysis because a reasonable amount of information can be obtained while only sequencing a smaller fraction of the mitochondrial genome, saving time and money when using a traditional Sanger sequencing method, and providing a more relevant amplification target for degraded DNA samples.

While it is possible to amplify HVS-I alone, or additionally HVS-II and HVS-III, it is of most benefit to analyze the entire control region due to the increased information content contained within the whole genome. The amplification strategy for this will depend on the quality of the mitochondrial DNA extracted. Listed below are two protocols for amplifying the control region either in a single 1.1 kb amplicon or in 5 ~ 300 bp overlapping amplicons. Should the DNA be too degraded for even this approach to produce results, then it is possible to amplify the region in ten

amplicons ranging in size from 101 to 190 bp, details of which can be found in Eichmann et al. [[10](#page-12-0)].

- (d) L314 - ccg ctt ctg gcc aca gca ct
- (e) H16175 tgg att ggg ttt tta tgt a
- 3. Illustra ExoStar 1-Step (VWR US77705).
- 4. 125 mM EDTA—made up from EDTA disodium salt dihydrate (MW 372.24) powder (Sigma-Aldrich E4884). Dissolve 93 mg in 2 ml H_2O .
- 5. 100 % ethanol.
- 6. 70 % ethanol—made by combining absolute ethanol (100 %) with water in a ratio of 7:3.
- 7. Deionized formamide (Life Technologies).

3 Methods

An initial PCR must be performed to amplify the region of interest: one of the two following PCR strategies should be followed. *3.1 Initial PCR*

> If DNA is believed to be of reasonable quality then the full control region strategy (*see* Subheading 3.1.1) should be followed—this amplifies the entire control region in a single reaction with a 1.1 kb amplicon. If DNA is believed to be of poor quality, or the full control region strategy fails, then the Midi-Mito strategy (*see* Subheading [3.1.2](#page-4-0)) should be followed—this requires the use of two multiplex PCRs that amplify the control region in five smaller, overlapping, fragments.

> It is possible to specifically quantify the amount of mitochondrial DNA present within the sample to analyze $[11]$; however, in the absence of a commercial solution, implementation of this method can be challenging. While specific mitochondrial DNA quantification is desirable, a viable alternative is to proceed to PCR without quantification, adding variable amounts of DNA depending on the sample type and expected cleanliness of the extract. It is assumed here that mitochondrial quantification has not been performed.

- 2. The required components and quantities for setting up the PCR are listed in Table [1.](#page-4-0) Reactions are performed in a 20 μl final volume.
- 3. A master mix should be prepared containing all the components except the DNA and water; 11.6 μl of this master mix can then be aliquoted into each PCR tube or well.

 Table 1

 Polymerase chain reaction reagent concentrations and volumes for amplifi cation of the control region in one reaction

- 4. Up to 8.4 μl of extracted DNA can then be added to the reaction. For good quality samples or those likely to contain inhibitors, adding 2 μl of DNA and 6.4 μl of water is recommended.
- 5. A positive and negative control should be included with each batch: for the positive control add 1 ng of 2800 M DNA (*see* **Note 5**) while for the negative control substitute sample DNA with DNA-free water.
- 6. PCR should be carried out in a PCR machine with a heated lid using the following program: an initial activation of 95 °C for 3 min followed by 36 cycles of 94 °C for 15 s, 56 °C for 1 min, and 72 °C for 2 min.

1. Two multiplex PCR reactions must be setup for each sample. PCR should be carried out using DNA-free consumable in a pre-PCR laboratory environment. Reactions can be prepared in 0.2 ml PCR tubes or 96-well PCR plates as desired. *3.1.2 Midi-Mito Strategy*

- 2. The required components and quantities for setting up Midi-Mito PCR A are listed in Table [2.](#page-5-0) Reactions are performed in a $20 \mu l$ final volume.
- 3. A master mix should be prepared containing all the components for Midi-Mito PCR A except the DNA and water; the appropriate volume (14.8 μl) of this master mix can then be aliquoted into each PCR tube or well.
- 4. Up to 5.2 μl of extracted DNA can be added to the reaction. For good quality samples or those likely to contain inhibitors, adding 2 μl of DNA and 3.2 μl of water is recommended.
- 5. For each sample, repeat **steps 2 4** above to setup Midi-Mito PCR B by using the PCR recipe detailed in Table [3](#page-5-0) that produces a master mix volume of 13.2 μl and allows a maximum of 6.8 μl of DNA to be added for each sample.

 Table 2

 Polymerase chain reaction reagent concentrations and volumes for Midi-Mito PCR A amplifying three of the five Midi-Mito reactions

Table 3

 Polymerase chain reaction reagent concentrations and volumes for Midi-Mito PCR B amplifying two of the five Midi-Mito reactions

- 6. A positive and negative control should be included with each batch for both PCR A and PCR B: for the positive control add 1 ng of 2800 M DNA while for the negative control substitute sample DNA with DNA-free water.
- 7. PCR should be carried out in a PCR machine with a heated lid using the following program: an initial activation of 95 °C for 3 min followed by 34 cycles of 94 °C for 15 s, 57 °C for 10 s, and $72 \degree$ C for 20 s .

Before proceeding to the sequencing stage, PCR products are first visualized on an agarose gel to check PCR success (*see* **Note 6**).

3.2 PCR Quality Check

- 1. Make a 0.5× TBE working solution by diluting the 10× TBE stock 1 in 20 with water.
- 2. Prepare a 1.5 % agarose gel by dissolving analytical grade agarose into 0.5× TBE, e.g., to prepare 100 ml gel, dissolve 1.5 g agarose in 100 ml 0.5× TBE. Add 1 μl of Gel Red stain per 10 ml, e.g., in the above example add 10 μl of Gel Red stain.
- 3. Heat in a microwave on a medium-high power setting until all the agarose has dissolved.
- 4. Pour into a gel cast that has been sealed at both ends, add a comb (a comb producing 1 mm length wells will be sufficient). Leave to set (*see* **Note 7**).
- 5. Once the gel has set, unseal the ends, remove the comb, and place in a gel tank containing 0.5× TBE.
- 6. For each PCR prepared in Subheading [3.1,](#page-3-0) load 8 μl onto the gel in individual wells. In a separate well, add 5 μl of DNA size ladder.
- 7. Run the gel at 100 V for 45 min with an unlimited current.
- 8. Visualize the gel under UV light. The DNA size ladder should be sharp, well separated, and bright against the background of the gel. For the entire control region PCR, a visible band at about 1100 bp indicates successful amplification while if the shorter amplicon protocol has been followed then bands should be visible around the 300 bp mark (there should be three bands present for the first Midi-Mito reaction and two present for the second Midi-Mito reaction; however, due to the similarity in amplicon size between two of the products in the first Midi reaction it is probable that only two bands will be easily differentiated).
- 9. Should there be bands in the negative control reactions matching the size expected for the product then all relevant PCR products should be disposed of and the protocol restarted from the PCR step.
- 10. Providing the PCR positive has worked, should it be evident that PCR has failed for one or more samples then processing of those particular failed PCRs should stop and the PCR stage be repeated using the Midi-Mito strategy for those samples.
- 1 For each reaction shown to have viable amplified products in the previous step, add 1.5 μl of ExoStar 1-Step to the remaining 10–12 μl of PCR product (*see* **Note 8**). Additionally process the negative PCR control in the same way. *3.3 PCR Cleanup*
	- 2 In a thermal cycler, incubate at 37 °C for 45 min followed by 80 °C for 15 min.
- Sequencing reactions can be undertaken using PCR tubes; however, it is more convenient to use a semiskirted 96-well plate that can be directly loaded onto a capillary electrophoresis machine following the sequencing cleanup stage. *3.4 Sequencing*
	- 1. Separate sequencing reactions in the forward and reverse direction must be setup for every sample. A single primer is included in each sequencing reaction, and hence multiple sequencing reactions are needed for every PCR amplicon. The sequencing primers used depend on the PCR strategy undertaken, such that:
		- (a) Full control region strategy—seven sequencing reactions are carried out on each single 1.1 kb PCR product using primers L15977, H16175, H16401, L16450, H274, L314, and H599.
		- (b) Midi-Mito strategy:
			- Midi-Mito PCR A—six sequencing reactions are carried out using primers L15989, H16248, L16450, H180, L317, and H619.
			- Midi-Mito PCR B—four sequencing reactions are carried out using primers L16197, H16509, L109, and H389.
	- 2. Sequencing reactions should be setup in accordance with the components and volumes detailed in Table 4 using the primers specified above such that there are seven sequencing reactions for each sample amplified using the full control region strategy and ten sequencing reactions for each sample amplified using the Midi-Mito strategy. Sequencing reactions are carried out in a total reaction volume of 6.8 μl.
	- 3. Multiple master mixes can be prepared and aliquoted into the plate for components that are identical between multiple wells, e.g., a different master mix could be prepared for each sequenc-

Table 4

Sequencing reaction reagent concentrations and volumes

ing primer such that each master mix contains the BigDye ready reaction mix, BigDye dilution buffer, a single specific sequencing primer, and water.

- 4. By default, 1 μl of purified PCR product is added to the sequencing reaction. Should the gel electrophoresis stage have shown poor amplification (weak bands) then additional purified PCR product can be added to the reaction at the expense of the water. Purified PCR products should not be brought into a pre-PCR area.
- 5. The positive and negative controls from the PCR stage should be taken the whole the way through the process, and hence the purified PCR products for the positive and negative PCR controls should be sequenced now alongside the rest of the samples.
- 6. Seal the plate with either PCR caps or a PCR seal.
- 7. Sequencing should be carried out in a PCR machine with a heated lid using the following program: an initial activation of 96 °C for 4 min followed by 25 cycles of 96 °C for 15 s, 50 °C for 10 s, and 60 \degree C for 2 min.
- 1. Add 1.7 μ l of 125 mM EDTA to each amplified sequencing reaction in the plate. Vortex the plate.
- 2. Add 20 μl of 100 % ethanol into the same wells now containing sequencing product and EDTA (*see* **Note 9**).
- 3. Vortex briefly then leaves at room temperature for a minimum of 15 min and a maximum of 24 h to precipitate sequencing products.
- 4. Centrifuge the plate at $2250 \times g$ for 30 min.
- 5. Remove all supernatant by unsealing the plate, inverting it onto a wad of folded paper towels and then centrifuging in this inverted orientation for 30 s at 400 rpm.
- 6. Add 120 μl of 70 % ethanol, reseal the plate, and spin in a normal orientation for 10 min at maximum speed (or 4000 rpm if this is lower) (*see* **Note 10**).
- 7. Repeat **step 5** to remove the supernatant.
- 8. Leave the plate at room temperature until all residual ethanol has evaporated (under normal circumstances, 5 min should suffice).
- 9. Add 10 μl of deionized formamide to each well containing sequencing product.
- 10. Heat at 95 °C for 3 min and then snap-cool (for example, by putting the plate in ice).
- 11. The sequencing products can now be run on a capillary electrophoresis instrument such as the 3130xl or 3500 (both Life Technologies).

3.5 Sequencing Cleanup and Electrophoresis

The mitochondrial control region sequence obtained from each sample needs to be compared to the revised Cambridge Reference Sequence (rCRS) and any differences between the reference sequence and the sample noted. To accomplish this, the multiple sequences produced for each sample from the different sequencing reactions must first be aligned to the reference sequence. There are various software solutions that will facilitate this alignment, such as SeqScape (Life Technologies) or Sequencer (Gene Codes). Once this alignment is complete, there are some important points to note. *3.6 Sequence Analysis*

- 1. It is ideal to have sequence coverage in both the forward and reverse direction for the entire control region, although this may be difficult to achieve in some circumstances, such as following homopolymeric stretches.
- 2. In almost all samples, there will be a difference to the reference sequence at base 263 $(A > G)$ and an insertion of a C base following nucleotide 315. Full details regarding the nomenclature to use for reporting such changes can be found in the International Society for Forensic Genetics' guidelines $[13-17]$.
- 3. The mitochondrial control region sequence of the 2800 M positive control has changes with respect to the rCRS at the following positions: 16519 (T > C), 152 (T > C), 263 (A > G), $315.1C, 477 (T > C).$
- 4. Length heteroplasmy (the presence of multiple similar sequences that differ only in length due to the insertion or deletion of a base) is a common occurrence in mitochondrial analysis and is especially problematic in three regions:
	- (a) If there is $T > C$ change at base 16189 then this creates a poly-C stretch between bases 16184 and 16193 and almost inevitably results in multiple mitochondrial sequences with differing numbers of C nucleotides in this stretch. The consequence of this is that any sequencing reactions covering this area will be unreadable following the poly-C stretch because there will be multiple different sequences (offset by a few bases) displayed concurrently.
	- (b) There are two more poly-C stretches between 303–309 and 311–315, and it is possible to observed length heteroplasmy here as well.
	- (c) There is another poly-C stretch between bases 568 and 573, and in some individuals, there can be length heteroplasmy following this stretch of C nucleotides if additional C bases have been inserted into the sequence. This can result in failure of the H599 or H619 sequences.
- 5. Point heteroplasmy (two different bases at a single position) can also be observed, albeit at a lower frequency than length heteroplasmy. More detail on heteroplasmy can be found in the literature $[17]$.
- 6. It is possible to observe mitochondrial sequence differences (usually at a single nucleotide position) between different tissues in an individual, and even between different hairs from the same individual. This usually manifests as point heteroplasmy but can be seen as a complete base change.
- 7. Using a web-based portal called EMPOP (www.empop.org), it is possible to obtain a frequency for the generated sequence (*see* **Note 11**).
- 8. Phylogenetic analysis can help to identify any errors in the final result; however, this can be difficult to undertake manually using phylotree (http://www.phylotree.org/tree/main.htm) unless relatively experienced. Alternatively, when examining the results from EMPOP, the program will display all those sequences that match closely, but not exactly, to your queried sequence. Here, for example, it may highlight that most similar sequences have a change at position 489 while your queried sequence has a change at 498, raising the possibility of a transcription error with the last two digits.

4 Notes

- 1. The DNA extraction method chosen will depend on the substrate, for example, specialized extraction methods are required for hair or bone. Normal DNA quantification methods are not very useful due to the types of samples used—total DNA quantification may be predominantly comprised of bacterial DNA while human specific methods target only the nuclear DNA whose concentration may bear no relation to the mitochondrial DNA concentration.
- 2. We notice no significant base misincorporation when using the MyTaq HS premix; however, it is also possible to use a proofreading polymerase such as Advantage HF 2 (Clontech).
- 3. It is possible to use any DNA sample as a positive control if it contains a consistent level of mitochondrial DNA and the mitochondrial sequence is known. 2800 M is stated here in order to provide a known set of mutations that can be referred to when undertaking analysis for the first time.
- 4. This is a nontoxic, although more expensive, alternative to using ethidium bromide to visualize the DNA bands. Ethidium bromide staining would also be acceptable, as would other nontoxic dyes.
- 5. Any characterized positive control can be used .
- 6. It is recommended to perform the PCR quality check stage (using gel electrophoresis) in order to gauge whether there is sufficient product to proceed with sequencing and estimate how much product to use in the sequencing reaction; however, it is possible to skip this stage and move straight to the sequencing stage entering an arbitrary amount of PCR amplicon into the sequencing reaction.
- 7. The time taken for the gel to set will depend on a number of factors including the ambient temperature and thickness of the gel; however, most gels will set within 20 min.
- 8. The PCR cleanup is performed to remove any unincorporated dNTPs and primers left over from the PCR stage. The method employed here utilizes two enzymes to accomplish this task; however, it is alternatively possible to achieve this end using a column cleanup approach.
- 9. The EDTA must be added to the amplified sequencing reaction before the 100 % ethanol because any direct contact between the EDTA and ethanol will result in precipitation of the EDTA—if a white precipitate/deposit is observed in any wells then either this EDTA/ethanol step was performed incorrectly or the EDTA used was too concentrated.
- 10. Care should be taken not to disturb the pellet when adding the 70 % ethanol, and the plate should not be vortexed after the ethanol has been added. The use of a 70 % ethanol concentration is important to keep the sequencing products precipitated, and hence 70 % ethanol should be made up fresh from absolute ethanol if the previous dilution is over 2 weeks old.
- 11. Reports should state the sequence range analyzed (e.g., 16,024–576) and all changes from the rCRS. Sequence frequency estimations can be obtained from EMPOP along with a 95 % confidence interval, for example, the 2800 M control region sequence is currently observed ten times out of 26,127 samples giving a frequency of 0.00038 with a 95 % confidence interval for this frequency between 0.00018 and 0.00070. Depending on the particular case circumstances it may be appropriate to report the frequency within the entire database, or more typically report the frequencies in one or more metapopulations (e.g., African, Westeurasian, and East Asian).

References

- 1. Anderson S, Bankier AT, Barrell BG, De Bruijn MHL, Coulson AR, Drouin J et al (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465
- 2. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N

(1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA [5]. Nat Genet 23:147

 3. Satoh M, Kuroiwa T (1991) Organization of multiple nucleoids and DNA molecules in

mitochondria of a human cell. Exp Cell Res 196:137–140

- 4. Robin ED, Wong R (1988) Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. J Cell Physiol 136:507–513
- 5. Torroni A, D'Urbano L, Rengo C, Scozzari R, Sbracia M, Manna C et al (1998) Intracytoplasmic injection of spermatozoa does not appear to alter the mode of mitocondrial DNA inheritance. Hum Reprod 13:1747–1749
- 6. Jenuth JP, Peterson AC, Fu K, Shoubridge EA (1996) Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. Nat Genet 14:146–151
- 7. Foran DR (2006) Relative degradation of nuclear and mitochondrial DNA: an experimental approach. J Forensic Sci 51:766–770
- 8. Bandelt H-J, Kong Q, Parson W, Salas A (2005) More evidence for non-maternal inheritance of mitochondrial DNA? J Med Genet 42:957–960
- 9. Lyons EA, Scheible MK, Sturk-Andreaggi K, Irwin JA, Just RS (2013) A high-throughput Sanger strategy for human mitochondrial genome sequencing. BMC Genomics 14:881
- 10. Eichmann C, Parson W (2008) 'Mitominis': multiplex PCR analysis of reduced size amplicons for compound sequence analysis of the entire mtDNA control region in highly degraded samples. Int J Legal Med 122:385–388
- 11. Ward RH, Frazier BL, Dew-Jager K, Pääbo S (1991) Extensive mitochondrial diversity

within a single Amerindian tribe. Proc Natl Acad Sci U S A 88:8720–8724

- 12. Brandstätter A, Peterson CT, Irwin JA, Mpoke S, Koech DK, Parson W et al (2004) Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database. Int J Legal Med 118:294–306
- 13. Parson W, Bandelt H (2007) Extended guidelines for mtDNA typing of population data in forensic science. Forensic Sci Int Genet 1:13–19
- 14. Tully G, Bär W, Brinkmann B, Carracedo A, Gill P, Morling N et al (2001) Considerations by the european DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. Forensic Sci Int 124:83-91
- 15. Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Holland M et al (2000) DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. Int J Legal Med 113:193–196
- 16. Carracedo A, Bär W, Lincoln P, Mayr W, Morling N, Olaisen B et al (2000) DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. Forensic Sci Int 110:79–85
- 17. Irwin JA, Saunier JL, Niederstätter H, Strouss KM, Sturk KA, Diegoli TM et al (2009) Investigation of heteroplasmy in the human mitochondrial DNA control region: a synthesis of observations from more than 5000 global population samples. J Mol Evol 68:516–527