

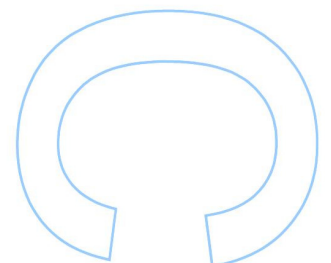
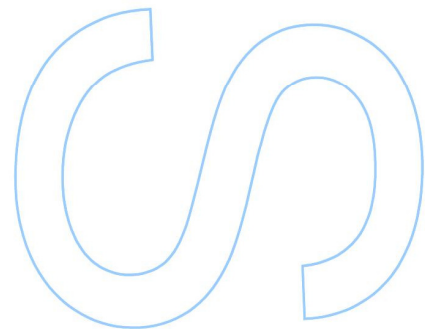
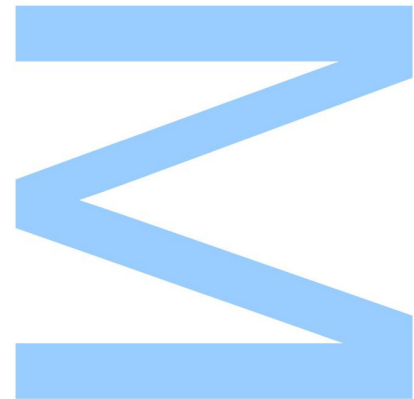


Sequenciação completa do mtDNA por MPS e avaliação da deteção de heteroplasmias

Sílvia Daniela Lobo Magalhães
Mestrado em Genética Forense
Departamento de Biologia
2015

Orientador

Doutora Ana Goios, PhD, Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP)

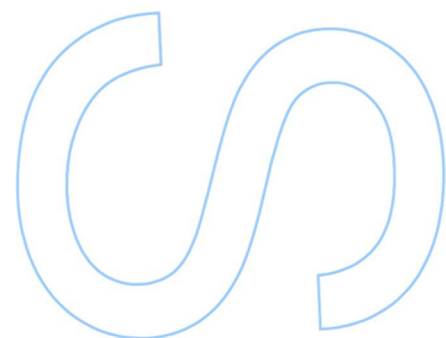
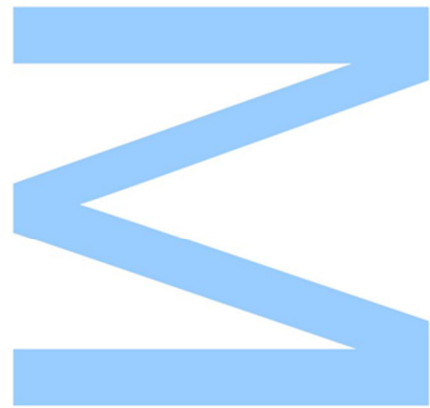




Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ____/____/____



AGRADECIMENTOS

Expresso aqui o mais profundo agradecimento a todos aqueles que tornaram a realização deste trabalho possível.

Ao Professor Amorim, agradeço a oportunidade que tive em frequentar este mestrado que muito contribuiu para o enriquecimento da minha formação académica e científica.

À minha orientadora, Ana Goios, quero agradecer por me ter integrado no seu projeto, por toda a orientação que me deu no desenvolvimento deste trabalho, pelo conhecimento transmitido e ajuda, por todos os esclarecimentos científicos, correções e sugestões.

À Sofia M. agradeço a pronta disponibilidade, partilha de conhecimentos e simpatia.

Ao Luís, Cíntia, Filipe P. e Filipe C. agradeço as valiosas contribuições para este trabalho.

Aos meus Pais e à minha irmã agradeço o apoio incondicional prestado neste percurso. Espero um dia retribuir-vos.

Aos meus colegas de mestrado, em especial à Tatiana e Filipa, muito obrigada pela vossa amizade.

RESUMO

A heteroplasma define-se como a presença de dois ou mais tipos de DNA mitocondrial (mtDNA) num indivíduo. Os níveis heteroplásmicos têm interesse tanto como marcadores genéticos, podendo aumentar o poder de discriminação em análises forenses, como em diagnóstico genético de doenças mitocondriais, já que o fenótipo da doença apenas se torna evidente quando a percentagem de moléculas mutantes existentes excedem um determinado limiar. A sequenciação do genoma mitocondrial completo através do método de Sanger para o estudo de heteroplasmas é dispendiosa, morosa e não é sensível o suficiente para a deteção a níveis inferiores a 15-20%. A Sequenciação Paralela Massiva (MPS) em diversas plataformas, e em particular na plataforma Ion Torrent PGM™, permite superar grande parte dessas limitações possibilitando uma análise mais sensível do genoma mitocondrial para a identificação das misturas com níveis reduzidos de heteroplasma, uma vez que são sequenciadas moléculas resultantes da amplificação monoclonal de uma única cadeia modelo, produzindo diversas *reads* independentes para cada posição.

A análise quantitativa das moléculas amplificadas de mtDNA permitiu-nos criar misturas de DNA de dois indivíduos distintos, com haplótipos previamente determinados, simulando heteroplasmas em proporções específicas. Foi utilizada a amplificação do genoma mitocondrial completo por Long-Range PCR (LR-PCR) para se tentar proporcionar uma cobertura uniforme. A análise dos dados de MPS indicou que até ao nível de 5% as variantes podiam ser detetadas e quantificadas com grande precisão ($R^2 = 0,9998$). Foi visível a problemática já conhecida da análise de regiões homopoliméricas por MPS. Em todas as proporções das misturas foram detetados falsos positivos o que sugere uma menor especificidade do método, embora a maioria deles surgissem associados a verdadeiras variantes ou regiões homopoliméricas, o que facilita a exclusão destas posições da análise.

ABSTRACT

Heteroplasmy is defined by the existence of two or more mitochondrial DNA (mtDNA) types within an individual. Heteroplasmic level is of particular interest as genetic marker, increasing the power of discrimination when applied to forensic cases, and in the genetic diagnosis of mitochondrial disease, since the disease phenotype becomes evident only when the percentage of mutant molecules exceeds a critical threshold value. Sequencing the whole mitochondrial genome searching for heteroplasmy using Sanger methodology is expensive, time-consuming and not sensitive enough to detect levels below 15-20%. Massively Parallel Sequencing (MPS) technology in several platforms, and particularly in the Ion Torrent PGM™ System, overcomes several of these limitations allowing for a more sensitive analysis of the mitochondrial genome for identification of low-level mixtures, since clonal populations of molecules originating from a single template strand are sequenced generating many independent reads for each position.

The quantitative analysis of amplified mtDNA molecules allowed us to create mixtures of DNA from two different individuals, with previously determined haplotypes, simulating heteroplasmy in defined proportions. The entire mitochondrial genome amplification by Long-Range PCR (LR-PCR) was used in attempt to provide uniform coverage. Analysis of MPS data indicated that minor variants could be detected and quantified to 5% level with great accuracy ($R^2=0,9998$). We could observe that the well characterized problem of homopolymeric stretches is indeed problematic when performing MPS. False positives were detected at all mixtures ratios suggesting a reduced specificity of the method, although most of them were associated to true variant sites or homopolymeric stretches, which allows for an eased exclusion of these positions from the analysis.

PALAVRAS-CHAVE

mtDNA, heteroplasma, MPS, PGM, sequenciação Sanger, Long-Range PCR

KEYWORDS

mtDNA, heteroplasmy, MPS, PGM, Sanger sequencing, Long-Range PCR

ÍNDICE

Agradecimentos.....	iii
Resumo	iii
Abstract.....	iv
Palavras-Chave	v
Keywords	v
Índice	vi
Lista de Tabelas	viii
Lista de Figuras	x
Abreviaturas.....	xiii
1. Introdução.....	15
1.1. DNA Mitocondrial	16
1.1.1. Origem e estrutura.....	16
1.1.2. Herança materna e ausência de recombinação.....	17
1.1.3. Número de cópias	18
1.1.4. Taxa de mutação.....	18
1.1.5. Heteroplasmia	19
1.1.6. Aplicações em Genética Forense.....	19
1.1.7. Aplicações na prática clínica	21
1.2. Tecnologias utilizadas na sequenciação do mtDNA	24
1.2.1. Sequenciação de Sanger	24
1.2.2. Sequenciação Paralela Massiva	26
2. Objetivos.....	30
3. Materiais e Métodos	32
3.1. Visão global da estratégia experimental e seleção de amostras	33
3.2. Extração do DNA	34
3.3. Amplificação das amostras por PCR/PCR multiplex e sequenciação pelo método de Sanger.....	35
3.3.1. Amplificação das amostras.....	35
3.3.2. Sequenciação do mtDNA pelo método de Sanger.....	37
3.3.3. Visualização e análise das sequências	40
3.4. Amplificação das amostras por Long-Range PCR e sequenciação pela plataforma Ion Torrent PGM	41
3.4.1. Amplificação das amostras por Long-Range PCR	41

3.4.2. Purificação dos amplificados	43
3.4.3. Quantificação dos amplificados purificados	44
3.4.4. Preparação da biblioteca.....	44
3.4.5. Preparação do <i>template</i>	48
3.4.6. Limpeza e inicialização do PGM	50
3.4.7. Preparação e carregamento do chip	50
3.4.8. Sequenciação no Ion Torrent PGM	51
3.4.9. Análise bioinformática dos resultados	52
4. Resultados.....	54
4.1. Haplótipos obtidos por sequenciação de Sanger.....	55
4.2. Amplificação por Long-Range PCR	56
4.2.1. Cobertura das amostras de referência	56
4.3. Sequenciação no Ion Torrent PGM.....	57
4.3.1. Sequenciação das amostras de referência	57
4.3.2. Relatório da corrida das amostras com heteroplasmias artificiais	57
4.3.3. Cobertura das amostras com heteroplasmias artificiais.....	60
4.4. Deteção e quantificação de variantes	64
4.5. Precisão da quantificação das heteroplasmias.....	68
4.6. Análise das misturas pelo método de Sanger.....	68
5. Discussão	70
5.1. Amplificação por Long-Range PCR	71
5.2 Haplótipos e Haplogrupos das amostras de referência	72
5.3. Cobertura	73
5.4. Sensibilidade, especificidade e precisão da quantificação de heteroplasmias..	75
5.5. Comparação da sensibilidade com o método de Sanger	77
5.6. Considerações finais.....	79
Referências Bibliográficas	80
Anexos.....	85

LISTA DE TABELAS

Tabela 1 - Proporções das 7 misturas para análise por MPS.....	33
Tabela 2 - Proporções das 7 misturas para análise pelo método de Sanger.	34
Tabela 3 - Localização da posição de ligação e sequência dos 12 pares de <i>primers</i> utilizados para amplificação da molécula de mtDNA.....	36
Tabela 4 - Localização da posição de ligação e sequência dos <i>primers</i> utilizados para sequenciar a molécula de mtDNA.	39
Tabela 5 - Sequência dos <i>primers</i> utilizados para amplificar a molécula de mtDNA... ..	41
Tabela 6 - Haplogrupos das amostras T3, T4 e T5 obtidos com recurso à ferramenta HaploGrep e a precisão dessa atribuição (qualidade), tendo por base os haplótipos completos obtidos pela sequenciação das duas amostras.	55
Tabela 7 - Características das <i>reads</i> segundo os diferentes níveis de qualidade de alinhamento.	60
Tabela 8 - Tabela parcial dos resultados do <i>plug-in Variant Caller</i> da mistura 1(T4):99(T5).....	86
Tabela 9 - Tabela parcial dos resultados do <i>plug-in Variant Caller</i> da mistura 5(T4):95(T5).....	88
Tabela 10 – Tabela parcial dos resultados do <i>plug-in Variant Caller</i> da mistura 10(T4):90(T5).....	90
Tabela 11 – Tabela parcial dos resultados do <i>plug-in Variant Caller</i> da mistura 50(T4):50(T5).....	92
Tabela 12 – Tabela parcial dos resultados do <i>plug-in Variant Caller</i> da mistura 90(T4):10(T5).....	94
Tabela 13 – Tabela parcial dos resultados do <i>plug-in Variant Caller</i> da mistura 95(T4):5(T5).....	96

Tabela 14 – Tabela parcial dos resultados do *plug-in Variant Caller* da mistura 99(T4):1(T5)..... 98

LISTA DE FIGURAS

Figura 1 - Representação esquemática do efeito de gargalo mitocondrial. (Adaptado de Taylor & Turnbull., 2005).	23
Figura 2 - Representação esquemática do método de terminação de cadeia (Adaptado de: Lehninger <i>et al.</i> , 2008).....	25
Figura 3 - Sequenciação no Ion Torrent PGM: A) Imagem de um chip semiconductor; B) Esquema representativo do interior de um chip; C) Esquema representativa da sequenciação por síntese que ocorre no interior de cada poço (Adaptado de: Calandro, 2012).....	28
Figura 4 - Condições para a amplificação do mtDNA completo por PCR multiplex.	37
Figura 5 - Condições para a reação de sequenciação do mtDNA.....	40
Figura 6 - Esquema representativo das posições de ligação dos <i>primers</i> MT_16555L e MT_16554H na molécula de mtDNA.	42
Figura 7 - Condições de amplificação por Long-Range PCR do mtDNA.	42
Figura 8 - Esquema representativo do método de purificação por Agencourt®: A) Produtos finais do PCR (amplificados e resíduos); B) Adição de Agencourt contendo <i>beads</i> ; C) Ligação dos amplificados às <i>beads</i> que são atraídas pelo suporte magnético; D) Lavagem com etanol; E) Libertação dos amplificados das <i>beads</i> ; F) Solução purificada após separação das <i>beads</i> (Adaptada de: Beckman Coulter, 2013).....	43
Figura 9 - Esquema representativo de todos os passos incluídos na preparação da biblioteca (Adaptada de: Life Technologies, 2013).....	44
Figura 10 - Esquema representativo das posições da strip do sistema <i>AB Library Builder™</i> (Adaptada de: Life Technologies, 2013).	45
Figura 11 - Imagem do <i>E-Gel® SizeSelect™</i> (Retirada de: Thermo Fisher, 2014).....	46
Figura 12 - Condições para a amplificação da biblioteca.	47

- Figura 13** - Esquema representativo do processo do emPCR, No interior de cada bolha podem encontrar-se dispostos ao acaso fragmentos de DNA e ISPs. Após o processo de amplificação podemos obter uma variedade de produtos para além dos desejados (Adaptado de: Inger Jonasson). 49
- Figura 14** - Esquema representativo do processo de enriquecimento: **A)** Biotina liga-se ao adaptador X; **B)** A biotina liga-se às esferas de estreptavidina *Dynabeads*; **C)** As *Dynabeads* são atraídas para um íman e as impurezas são removidas; **D)** O conjugado ISP-*template* é desnaturado com NaOH (Adaptado de: ThermoFisher). 50
- Figura 15** - Verificação da amplificação por Long-Range PCR das amostras T4 e T5 em gel de agarose a 0,8%. Legenda: M – Marcadores de peso molecular; T4 – Amostra T5; T5 – Amostra T5; C- – Controlo negativo. 56
- Figura 16** - Cobertura da amostra T4 amplificada em um só fragmento de aproximadamente 16,5 kbs **(A)** e em dois fragmentos sobreponíveis de aproximadamente 8,5 kbs **(B)**. O eixo das abcissas representa a posição dos nucleótidos ao longo da molécula de mtDNA e o eixo das ordenadas a cobertura. Destacado a azul a zona de ligação dos *primers* da estratégia de amplificação em um só fragmento e a roxo as zonas onde se sobrepõe os dois fragmentos da estratégia de Fendt e colaboradores. 57
- Figura 17** - Esquemas do relatório pré-alinhamento: **A)** Imagem ilustrativa do carregamento dos poços do chip ao longo da sua superfície; **B)** Resumo do estado das ISPs; **C)** Histograma do comprimento das *reads*. 59
- Figura 18** - Relatório de alinhamento: **A)** Gráfico do alinhamento das *reads*; **B)** Precisão do alinhamento das *reads*. 60
- Figura 19** - Gráficos da cobertura [em número de *reads* (eixo das ordenadas)] das posições assinaladas como variantes (eixo das abcissas) nas amostras com mistura: **A)** 1(T4):99(T5); **B)** 5(T4):95(T5); **C)** 10(T4):90(T5). 62
- Figura 20** - Gráficos da cobertura [em número de *reads* (eixo das ordenadas)] das posições assinaladas como variantes (eixo das abcissas) nas amostras com mistura: **A)** 50(T4):50(T5); **B)** 90(T4):10(T5); **C)** 95(T4):5(T5); **D)** 99(T4):1(T5). 63
- Figura 21** - Gráficos da cobertura alélica e frequência [em número de *reads* e percentagem (eixo das ordenadas esquerdo e direito, respetivamente)] das posições

assinaladas como variantes (eixo das abcissas) nas amostras com mistura: **A)** 1(T4):99(T5); **B)** 5(T4):95(T5); **C)** 10(T4):90(T5)..... 65

Figura 22 - Gráficos da cobertura alélica e frequência [em número de reads e percentagem (eixo das ordenadas esquerdo e direito, respetivamente)] das posições assinaladas como variantes (eixo das abcissas) nas amostras com mistura: **A)** 50(T4):50(T5); **B)** 90(T4):10(T5); **C)** 95(T4):5(T5); **D)** 99(T4):1(T5)..... 66

Figura 23 - Frequência da heteroplasmas [em percentagem (eixo das ordenadas)] nas posições onde se esperavam serem detetadas variantes (eixo das abcissas) nas 7 amostras com mistura. 67

Figura 24 - Correlação entre os níveis de heteroplasma esperados (eixo das abcissas) e observados (eixo das ordenadas). O círculo e a barra de erro representam, respetivamente, a média e o desvio padrão das 82 posições de cada uma das 7 amostras. 68

Figura 25 - Visualização de algumas heteroplasmas (destacadas a azul) das 7 amostras com mistura analisadas pelo método de Sanger no *software SeqScape*..... 69

Figura 26 – Imagem de um gel com produtos da amplificação por LR-PCR num só fragmento onde se pode visualizar uma amostra (destacada a azul) com bandas adicionais (A. Goios, comunicação pessoal). 72

Figura 27 - Gráficos da cobertura (eixo das ordenadas) de todas as posições dos nucleótidos ao longo da molécula de mtDNA (eixo das abcissas) nas 7 amostras com mistura visualizadas no *software IGV*: **A)** 1(T4):99(T5); **B)** 5(T4):95(T5); **C)** 10(T4):90(T5); **D)** 50(T4):50(T5); **E)** 90(T4):10(T5); **F)** 95(T4):5(T5); **G)** 99(T4):1(T5). 100

ABREVIATURAS

ATP - Trifosfato de Adenosina

bp – Pares de Bases

ddNTP – Didesoxirribonucleotídeo fosfatado

D-loop – *Displacement Loop*

DNA - Ácido Desoxirribonucleico

dNTP – Desoxirribonucleotídeo Fosfatado

emPCR – *Emulsion Polymerase Chain Reaction*

g – Força-g

HVR – Região Hipervariável

Indel – Inserção/Deleção

ISP – *Ion Sphere Particles*

Kbs – Kilo Pares de Bases

M – Milhões

Mb – Megabases

Mbp – Mega Pares de Bases

min – Minutos

ml - Mililitros

MPS – Sequenciação Massiva Paralela (*Massive Paralell Sequencing*)

mRNA - Ácido Ribonucleico Mensageiro

mtDNA - Ácido Desoxirribonucleico Mitocondrial

nDNA - Ácido Desoxirribonucleico Nuclear

ng - Nanogramas

NGS – Sequenciação de Nova Geração (*Next-Generation Sequencing*)

nt – Nucleótidos

OXPHOS – Fosforilação Oxidativa

PBS - Tampão Fosfato Salino

PCR – *Polymerase Chain Reaction*

pg - Picogramas

PGM – Personal Genome Machine

rCRS - *revised Cambridge Reference Sequence*

RNA - Ácido Ribonucleico

rRNA - Ácido Ribonucleico Ribossomal

seg – Segundos

tRNA - Ácido Ribonucleico de transferência

µL - Microlitros

1. INTRODUÇÃO

1.1. DNA Mitocondrial

1.1.1. Origem e estrutura

As mitocôndrias são organelos presentes em quase todas as células eucarióticas. São constituídas por uma membrana externa e uma membrana interna que possui várias invaginações – cristas mitocondriais - que aumentam a sua área de superfície. No seu interior, encontra-se a matriz mitocondrial onde estão dispersas enzimas, ribossomas, RNA e DNA.

A sua principal função é a produção de energia a partir da hidrólise de hidratos de carbono e de ácidos gordos, os quais são convertidos em ATP por um processo designado fosforilação oxidativa (OXPHOS) (Andersson *et al.*, 2003; da Fonseca *et al.*, 2008).

A teoria endossimbiótica defende que as mitocôndrias derivam da endocitose de um procarionte ancestral aeróbio por um eucarionte unicelular anaeróbio. Esta relação terá persistido por ser mutuamente benéfica. A célula procariótica proporcionava ao eucarionte a capacidade de realizar a fosforilação oxidativa, permitindo-lhe sobreviver em ambientes diversos, e a célula anaeróbia fornecia nutrientes e proteção ao endossimbionte. Ao longo do tempo, a célula procariótica tornou-se parte integrante da célula hospedeira, culminando com a evolução para células eucarióticas atuais. Esta teoria é suportada por vários argumentos, nomeadamente por as mitocôndrias serem muito semelhantes a alguns procariontes em dimensão e estrutura, e por possuírem DNA próprio, o que lhes confere autonomia relativamente à célula hospedeira por lhes permitir dividir-se independentemente desta (Margulis, 1981).

Desde a origem da célula eucariótica até ao DNA mitocondrial (mtDNA) atual de mamíferos, muitos dos genes que constituíam o genoma mitocondrial foram perdidos e/ou transferidos para o DNA nuclear (nDNA). Da mesma forma que estes genes foram transferidos para o núcleo, alguns fragmentos de mtDNA têm sido inseridos nos cromossomas nucleares durante a nossa história evolutiva, sob a forma de pseudogenes (não-codificantes). Algumas destas inserções de mtDNA no genoma nuclear (NUMTs) são remotas no tempo e partilhadas por todos os primatas, enquanto outras são mais recentes, variando mesmo a sua presença ou não entre populações (Ricchetti *et al.*, 2004).

O DNA mitocondrial de mamíferos é uma molécula circular de cadeia dupla constituída por 37 genes que codificam duas moléculas de rRNA, 22 moléculas de tRNA e 13

polipéptidos envolvidos no processo de fosforilação oxidativa (OXPHOS) e na produção de energia (Andrews *et al.*, 1999). O DNA mitocondrial humano tem cerca de 16569 pares de bases, podendo este comprimento variar devido a mutações de um pequeno número de nucleótidos, que podem ser inserções ou deleções (Anderson *et al.*, 1981; Andrews *et al.*, 1999). A cadeia pesada ou H (*heavy*) é mais rica em guaninas e a cadeia leve ou L (*light*) possui mais citosinas na sua composição (Butler, 2012).

O mtDNA apresenta uma organização bastante económica: a maioria dos seus genes são contíguos, não possuem intrões e praticamente todo o mRNA é traduzido. Contudo, há uma região não codificante que se estende desde a posição 16024 até à posição 576 – região controlo ou *D-loop* – que contém os promotores de transcrição e a origem de replicação (Andrews *et al.*, 1999). Esta área também inclui pequenos segmentos que, quando comparados com o resto do genoma, apresentam sequências altamente variáveis – as regiões hipervariáveis (HVR): a região hipervariável 1 (HVR1) que corresponde ao segmento entre as posições 16024 e 16365; a região hipervariável 2 (HVR2), que está localizado entre as posições 73 e 340, e a região hipervariável 3 (HVR3) que corresponde à zona entre as posições 438 e 574 (Butler, 2012; Lutz *et al.*, 1997).

1.1.2. Herança materna e ausência de recombinação

O mtDNA humano tem transmissão exclusivamente materna (Giles *et al.*, 1980; Taylor *et al.*, 2003). Durante a fecundação, as mitocôndrias paternas ficam retidas na cauda do espermatozoide e o zigoto herda somente as mitocôndrias que se encontram no citoplasma do oócito. Foi sugerido que, caso algumas mitocôndrias paternas conseguissem penetrar no oócito, estas ficariam extremamente diluídas em comparação com as maternas e que seriam marcadas com ubiquitina para posterior degradação (Sutovsky *et al.*, 2000). Foi apenas reportado um caso de herança mitocondrial paterna (Schwartz & Vissing, 2002).

Consequentemente, este genoma não sofre recombinação e é transmitido “em bloco” (haploide). Assim, o mtDNA pode ser herdado intacto através das gerações a não ser que ocorra uma mutação. O haplótipo mitocondrial materno será passado a todos os filhos e filhas no entanto, apenas estas os transmitirão aos seus descendentes.

1.1.3. Número de cópias

Uma mitocôndria contém entre 2 a 10 cópias de mtDNA e cada célula pode encerrar entre dezenas a milhares destes organelos (Cavelier *et al.*, 2000). Assim, enquanto a maioria das células apenas possui duas cópias de DNA nuclear, podem conter entre 1000 e 100000 cópias de mtDNA, dependendo do tipo de célula e dos seus requerimentos energéticos (Chinnery & Hudson, 2013).

1.1.4. Taxa de mutação

O mtDNA é herdado ao longo das gerações mas, durante esse processo, podem ocorrer mutações que alteram a sua sequência. Muitas dessas mutações são mantidas ao longo do tempo e não apresentam qualquer efeito adverso para o seu portador, aparecendo em pelo menos 1% da população – polimorfismos.

Nos polimorfismos do mtDNA destacam-se os SNPs (*Single Nucleotide Polymorphisms*) que surgem quando um par de bases é substituído por outro - mutação pontual.

O genoma mitocondrial exibe uma acelerada taxa de mutação que poderá ser cerca de 5 a 10 vezes superior à taxa mutacional média do nDNA (Butler, 2012). A taxa de substituições na região controladora encontra-se entre 5×10^{-7} e 5×10^{-6} substituições por posição por ano (Parsons *et al.*, 1997) e para a região codificante $1,26 \times 10^{-8}$ substituições por posição por ano (Mishmar *et al.*, 2003).

Esta elevada taxa de mutação, em comparação com o nDNA, deve-se a:

- à acumulação de radicais livres, que podem danificar o DNA, resultantes da atividade da mitocôndria;
- à ausência de histonas na molécula de mtDNA;
- à frequente replicação do mtDNA;
- e à inexistência de sistemas eficientes de reparação de DNA (Birky, 2001).

1.1.5. Heteroplasmia

A ocorrência de uma nova mutação ou polimorfismo numa determinada molécula de mtDNA leva a que, dentro de uma célula, passe a existir uma mistura de sequências de DNA diferentes, uma condição denominada heteroplasmia (Lightowlers *et al.*, 1997). Quando uma célula heteroplásmica se replica, os dois tipos de moléculas são segregadas de forma aleatória para as células-filhas e o acaso determina a proporção do tipo de DNA em cada uma. Apesar de a maioria das células herdar uma mistura dos dois tipos de DNA, algumas células podem receber apenas um deles – homoplasmia. Assim, um único indivíduo pode ter diferentes tipos de mtDNA num só tecido, ter um tipo de mtDNA num tecido e outro tipo noutra amostra e/ou apresentar heteroplasmia numa amostra de um tecido e homoplasmia em outra amostra do mesmo tecido (Carracedo *et al.*, 2000) ou até mesmo apresentar heteroplasmia numa única mitocôndria (Reiner *et al.* 2010).

As heteroplasmias podem ser heteroplasmias de sequência (ou pontuais) ou heteroplasmias de comprimento. As heteroplasmias de sequência caracterizam-se pela presença de dois nucleótidos diferentes na mesma posição. As heteroplasmias de comprimento ocorrem, normalmente, junto às expansões de citosinas nas regiões homopoliméricas da HVR1 (da posição 16184 até à 16193) e da HVR2 (da posição 303 até à 310) (Parson & Bandelt, 2007; Stewart *et al.* 2001).

Estudos recentes apontam para que quase todos os indivíduos possuam algum estado de heteroplasmia mesmo que em níveis muito reduzidos (Naue *et al.*, 2015; Payne *et al.*, 2013). A incidência de heteroplasmia de sequência em tecidos com elevada atividade metabólica (por exemplo, músculo e fígado) parece ser frequentemente maior do que em tecidos como o sanguíneo, o ósseo e o epitélio bucal (Naue *et al.*, 2015; Ramos *et al.*, 2013).

1.1.6. Aplicações em Genética Forense

De uma maneira generalizada, pode-se afirmar que todos os humanos são relacionados por via materna pois descendem todos de um tipo ancestral de mtDNA comum presente na população da África subsariana há cerca de 150000-200000 anos (Wallace *et al.*, 1999). As mutações que se acumularam no genoma mitocondrial desde então, aliadas às diferentes migrações das populações pelo planeta e outras pressões seletivas, são os geradores das diferenças existentes entre indivíduos e populações.

A utilização das regiões variáveis do genoma é útil quando aplicada na identificação humana (Butler, 2012). No entanto, por não sofrer recombinação e ser transmitido por via materna, o mtDNA não permite distinguir um indivíduo dos outros pertencentes à mesma linhagem materna e, assim, possui um poder de discriminação inferior ao do nDNA (Holland & Parsons, 1999).

Ainda assim, a herança materna permite o aumento do número de potenciais amostras de referência de um indivíduo, quando amostras deste ou dos seus familiares diretos não são acessíveis para comparação. A amostra de referência poderá, então, provir de qualquer familiar da linhagem materna, mesmo que separado por várias gerações (Budowle *et al.*, 2003).

Além disto, a proteção pela dupla membrana da mitocôndria e a forma circular da molécula de DNA, que o torna menos vulnerável à degradação por exonucleases, faz com que a possibilidade de ser recuperado intacto, após um longo período de tempo, seja mais elevada do que o nDNA. Da mesma forma, a pequena molécula e as muitas cópias de mtDNA em cada célula, tornam possível a obtenção de DNA a partir de amostras biológicas muito pequenas, como as encontradas em desastres em massa ou em investigações de pessoas desaparecidas.

No início dos anos 90, a análise das HVR ou, por vezes, de toda a região controlo tornou-se o método mais comum de análise de mtDNA por serem estas as zonas que apresentam mais variabilidade. Atualmente, as análises forenses de rotina do mtDNA ainda consistem exclusivamente na análise da região controlo (Butler, 2012). Apenas depois da passagem do século, surgiram os primeiros estudos que contemplavam a sequenciação completa de genomas mitocondriais de várias populações mundiais de forma a aumentar o poder de discriminação (Ingman *et al.* 2000; Parsons & Coble 2001).

Tendo em conta as técnicas mais utilizadas na maioria dos laboratórios forenses, a presença de heteroplasma pode complicar a interpretação dos resultados, como veremos mais à frente. Contudo, este fenómeno pode também aumentar o valor de uma identificação quando se encontra a mesma heteroplasma na amostra de referência e na amostra a testar (Coble *et al.*, 2009).

Por outro lado, como já referido anteriormente, amostras provenientes de diferentes tecidos que podem não apresentar o mesmo padrão de heteroplasma. Quando uma amostra referência apresenta este fenómeno e a amostra teste não apresenta, ou vice-versa, gera-se uma dificuldade na interpretação dos resultados. Isto ocorre frequentemente quando a amostra de um local de crime é, por exemplo, cabelo

(frequência de heteroplasma de 36.7% a 30.2%) e a amostra colhida ao suspeito é saliva ou sangue (19.8% a 16.2%) (Naue *et al.*, 2015).

1.1.7. Aplicações na prática clínica

As doenças mitocondriais são disfunções da cadeia respiratória, de origem genética, que afetam a fosforilação oxidativa. São doenças multissistémicas mas afetam, sobretudo, os tecidos com elevadas necessidades energéticas como o sistema nervoso central, os músculos esqueléticos, o coração, os rins e o sistema endócrino.

As citopatias mitocondriais podem ser causadas por alterações nos genes mitocondriais ou nos genes nucleares relacionados com a estrutura e função mitocondrial (Chinnery & Hudson, 2013). As mutações do mtDNA são responsáveis por 5-40% das doenças mitocondriais na idade adulta e até 25% dos casos na criança (Thorburn, 2004a & 2004b). Apesar da maioria das alterações no mtDNA serem polimorfismos neutros (Ingman *et al.*, 2000), a primeira mutação patogénica foi identificada em 1988 (Holt *et al.*, 1988; Wallace *et al.*, 1988). Desde aí, mais de 270 mutações causadoras de doença foram registadas (Lott *et al.*, 2013) e originam um espectro variado de doenças (MacFarland *et al.*, 2007).

O modo de transmissão complexo (materno, no caso de mutações no mtDNA, e mendeliano, no caso de mutações no nDNA) e a heterogeneidade clínica das doenças mitocondriais resultam muitas vezes em diagnósticos demorados ou mal concebidos. A prevalência das doenças mitocondriais é, também por isso, difícil de determinar. Estimativas cautelosas apontam para cerca de 1 em 5000/10000 indivíduos afetados (Chinnery *et al.*, 2000; Schaefer *et al.*, 2008; Skladal *et al.*, 2003). Como a penetrância é reduzida, a prevalência nos adultos poderá ser ainda mais elevada. Um estudo afirma mesmo que 1 em cada 200 adultos saudáveis possui uma mutação pontual no mtDNA (Elliott *et al.*, 2008).

As mutações de mtDNA podem ser pontuais ou rearranjos do DNA (grandes deleções ou inserções). As mutações pontuais são geralmente herdadas e frequentemente heteroplásmicas (Tuppen *et al.*, 2009). Dada a elevada taxa de mutação do mtDNA e a presença frequente de polimorfismos, a distinção entre mutações neutras ou causadoras de doença pode ser problemática. Os rearranjos de mtDNA são geralmente grandes deleções que variam entre 1,3 a 8 kbs e incluem diversos genes (Schon *et al.*,

1989). Costumam ocorrer de forma esporádica no início do desenvolvimento e de forma constante em todas as células do mesmo tecido (Chen *et al.*, 1995).

1.1.7.1. Efeito limiar

A manifestação clínica de doença mitocondrial e os defeitos a nível bioquímico ocorrem quando a fosforilação oxidativa fica enfraquecida, o que só acontece quando se atinge um número mínimo crítico de moléculas de mtDNA mutado - efeito limiar. Este limiar varia com o tipo de mutação e de tecido, sendo o seu nível mais reduzido nos tecidos com maiores requerimentos energéticos (Tuppen *et al.*, 2009).

A expressão clínica de uma heteroplasmia patogénica no mtDNA está diretamente correlacionada com a proporção de mtDNA mutado. Parece ser necessária uma proporção superior a 60% mtDNA mutado para ser observado um fenótipo, variando os níveis acima deste valor de acordo com o tipo de mutação e de tecido afetado (Rossignol *et al.*, 2003).

Para o diagnóstico de doenças mitocondriais têm sido utilizados tecidos que possuem um elevado número de mitocôndrias, tal como o músculo esquelético. Uma biópsia deste tecido é o ideal para se detetarem variantes patogénicas que podem não ser detetadas numa amostra comum de sangue através das tecnologias mais comuns num laboratório de diagnósticos.

1.1.7.2. Modo de transmissão e efeito de gargalo mitocondrial

A transmissão de doenças mitocondriais entre gerações é bastante complexa devido à interação dos dois genomas, mitocondrial e nuclear, na manifestação de determinado fenótipo. Para além disto, como muitos fatores clínicos dependem da proporção de mtDNA mutado, torna-se difícil estimar a probabilidade de uma mãe que apresente doença mitocondrial ter filhos afetados (Carling *et al.*, 2011; Cree *et al.*, 2008).

Alguns estudos apontam para o facto de mães com heteroplasmia pontual terem uma descendência com herança de mtDNA variada. Este fenómeno está associado ao efeito de gargalo mitocondrial. No início da oogénese, há uma redução de moléculas de mtDNA que são distribuídas para cada oócito primordial. À maturação dos oócitos está associada uma rápida replicação do mtDNA. Este evento de restrição seguida de

amplificação das moléculas de mtDNA pode levar a uma mudança brusca na proporção de moléculas mutadas nas diferentes gerações e é a responsável pelos níveis variados de mtDNA mutado observado nos descendentes afetados de mães com mutações de mtDNA patogénicas (Carling *et al.*, 2011) (figura 1). Consequentemente, a variabilidade na percentagem de moléculas de mtDNA mutadas observadas na descendência de uma progenitora com heteroplasmia provém, pelo menos em parte, da seleção de apenas um subconjunto de mtDNA durante a oogénese.

Uma exceção à herança materna mtDNA mutante ocorre quando a progenitora tem heteroplasmia provocada por uma grande deleção no mtDNA. Por razões desconhecidas, as moléculas com remoção de mtDNA, geralmente, não são transmitidas das mães clinicamente afetadas para a sua descendência. Progenitoras não afetadas que tenham um filho com este tipo de mutação têm pouca probabilidade de ter um segundo filho afetado. Já as progenitoras afetadas correm o risco de 1 para 24 (Chinnery *et al.*, 2004).

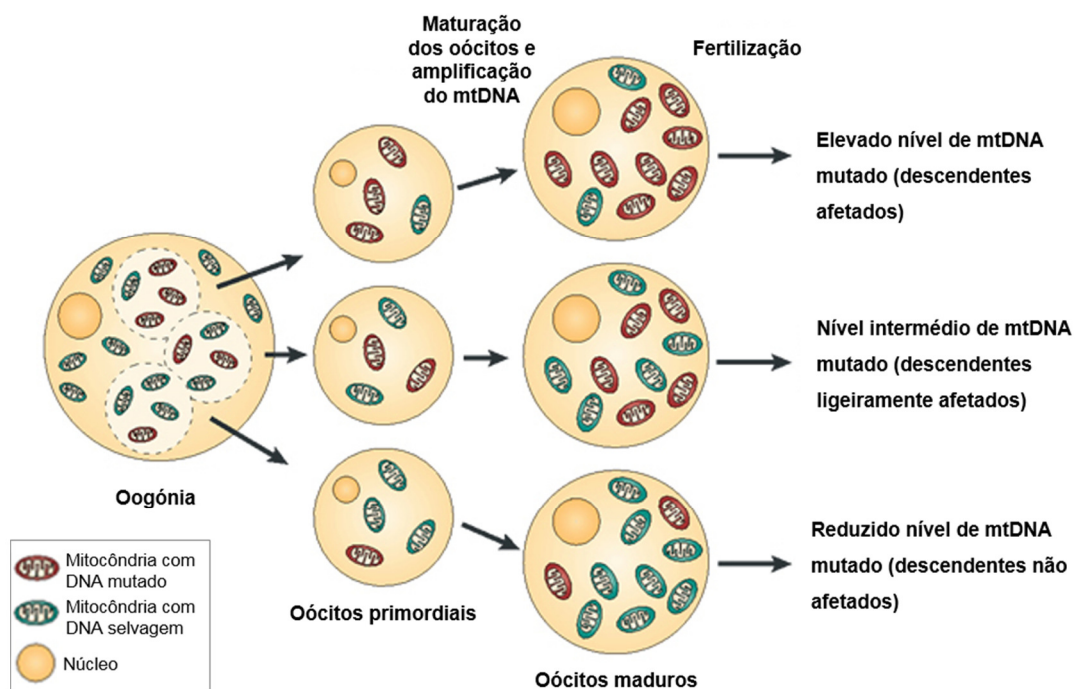


Figura 1 - Representação esquemática do efeito de gargalo mitocondrial. (Adaptado de Taylor & Turnbull., 2005).

1.2. Tecnologias utilizadas na sequenciação do mtDNA

A sequenciação permite-nos desvendar a sequências de nucleótidos de determinada molécula de DNA. Diversas técnicas de sequenciação têm sido utilizadas para a detecção de heteroplasmias devido ao crescente interesse no campo da genética forense, principalmente para o aumento do poder de discriminação de haplótipos, e na prática clínica, onde é necessário quantificar a heteroplasmia e pesquisá-la em vários tecidos para poder ser associada a doença.

Os primeiros estudos em mtDNA foram efetuados por RFLP (*Restriction Fragment Length Polymorphism*). Mais tarde, esta técnica foi aliada à PCR (*Polymerase Chain Reaction*) originando uma técnica simples e conveniente para a detecção de SNPs (Ota *et al.*, 2007). As endonucleases de restrição reconhecem sequências específicas no DNA onde efetuam um corte, originando fragmentos com comprimentos conhecidos. A presença de um polimorfismo, cria ou elimina o local de corte de uma endonuclease de restrição, resultando em fragmentos com comprimentos distintos, detetados num gel sujeito a eletroforese. A detecção de heteroplasmias através desta técnica pode ser afetada por diversos fatores e níveis inferiores a 10% não conseguem mesmo ser detetados (Bai & Wong, 2004).

1.2.1. Sequenciação de Sanger

Em 1975, Frederick Sanger e os seus colaboradores, introduziram uma metodologia de sequenciação de DNA baseada no PCR convencional, designada método de terminação de cadeia (Sanger & Coulson, 1975; Sanger *et al.*, 1977). Devido à sua enorme facilidade de execução, este método tornou-se o padrão para a sequenciação entre a comunidade científica durante mais de três décadas.

De acordo com esta técnica, o DNA é replicado na presença de didesoxirribonucleótidos (ddNTPs), que não possuem o grupo hidroxilo do carbono 3' e atuam como inibidores sintéticos que competem com os seus análogos naturais dNTPs. Para que a síntese do DNA ocorra, a DNA polimerase tem que catalisar a reação entre do hidroxilo do nucleótido antecessor e o fosfato do carbono 5' do nucleótido a ser adicionada à cadeia. Assim, quando um ddNTPs é incorporado numa cadeia a sua extensão é interrompida pois a DNA polimerase não consegue executar a sua função.

A reação de síntese ocorre na presença de concentrações planeadas dos dois tipos de nucleótidos de forma a, no final de vários ciclos, se conseguir obter diversos fragmentos de DNA de comprimentos variados, que terminam com um dos inibidores.

De seguida, cabe separar esses fragmentos por ordem de forma a revelar a sequência total. Hoje em dia, a técnica mais utilizada para este fim é a eletroforese capilar. Nesta, utilizam-se os ddNTPs marcados com fluoróforos e fazem-se migrar os diferentes fragmentos ordenadamente no interior dos capilares. À medida que os terminadores são excitados por um feixe de laser, emitem luz nos diferentes comprimentos de onda que correspondem a cada um dos quatro ddNTPs terminadores que são detetados pelos sensores (figura 2).

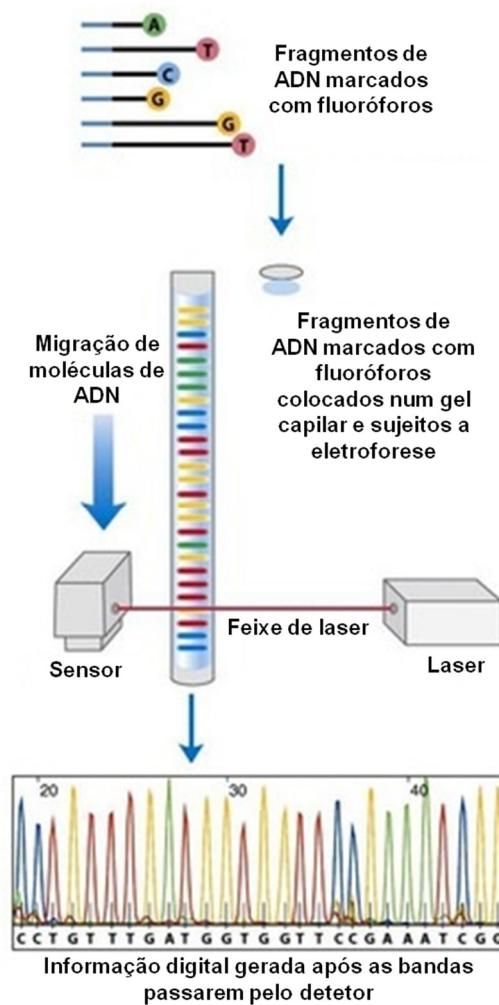


Figura 2 - Representação esquemática do método de terminação de cadeia (Adaptado de: Lehninger *et al.*, 2008).

A probabilidade de ser atribuída uma base errada a uma posição, através desta metodologia, pode resultar de erros que ocorrem na amplificação, por contaminação da

amostra ou devido ao *slippage* da polimerase em regiões de *short tandem repeat* e homopolímeros. Ainda assim, a média final de erros na leitura é muito baixa (1 erro por cada 10000-100000 nt) (Ewing & Green, 1998).

Contudo, apesar de o método de Sanger ser amplamente utilizado para a sequenciação, apresenta algumas limitações. A maioria dos laboratórios apenas possui aparelhos que atingem um máximo de cerca 6 Mb de seqüências de DNA por dia, produzindo leituras da ordem de 600 a 1000 nucleótidos, com um custo a rondar os 500 dólares por cada 1Mb (Kircher & Kelso, 2010). Assim, este método demonstrou-se demasiado moroso e dispendioso para ser empregue em projetos de sequenciação de larga escala.

A detecção de heteroplasma pontual é feita pela observação de dois picos na mesma posição do eletroforegrama, pelo que “ruído de fundo” criado por aparelhos e corantes, podem não permitir a sua distinção e dar origem a falsos positivos. Além disso, a sequenciação de Sanger não é um método quantitativo, não possui a sensibilidade para detetar heteroplasma com níveis inferiores a 15-20% e não deteta grandes deleções no genoma mitocondrial (Rohlin et al., 2009; Chinault et al., 2009; Wong et al., 2008).

1.2.2. Sequenciação Paralela Massiva

Mais recentemente, avanços na tecnologia de sequenciação, levaram à produção de plataformas capazes de gerar informação muitas vezes superior à das tecnologias até então existentes. Estas tecnologias de Sequenciação Paralela Massiva (MPS) ou Sequenciação de Nova Geração (NGS) são capazes de conceber informação sobre milhões de pares de bases numa única corrida e a um custo bastante inferior, quando comparadas com a metodologia de Sanger (Just *et al.*, 2015).

Com tecnologias desta rapidez torna-se fácil imaginar a análise do mtDNA completo em análises de rotina em vez de apenas se sequenciar as zonas HVR ou a região controlo.

Estas tecnologias de MPS originam *reads* (pequenas seqüências de nucleótidos obtidos por sequenciação) de pequeno tamanho (com 30 a 400 bp) em prol do aumento da cobertura da sequenciação quando comparadas com as plataformas de sequenciação de Sanger por eletroforese capilar. A cobertura é definida como o número de *reads* alinhadas com a referência que representam uma determinada região, isto é, o número de vezes que uma posição nucleotídica é sequenciada. Quanto maior o número de vezes que essa região é sequenciada (ou seja, a intensidade de cobertura) e a

uniformidade da cobertura maior é a confiança e a sensibilidade do processo de identificação de variantes. A intensidade de cobertura pode ser influenciada pela própria natureza da sequência a investigar, do método de enriquecimento, do tipo de variante a ser analisada e se pertence à linhagem somática ou germinativa (Life Technologies, 2014). Zhang e colaboradores, comparam as coberturas obtidas por diferentes estratégias de PCR na amplificação do genoma mitocondrial completo. Aquela cujo genoma foi amplificado num único fragmento obteve uma intensidade de cobertura mais elevada e mais uniforme (Zhang *et al.*, 2012).

Como estas tecnologias geram produtos de leitura pequenos, tornaram possível a obtenção de sequências de amostras ancestrais de DNA, que normalmente se encontram muito fragmentadas para serem analisadas com as técnicas tradicionais (Green *et al.*, 2008; Miller *et al.*, 2008).

Desde 2005 são muitas as plataformas de MPS que têm surgido com base em tecnologias que utilizam leitura ótica. Mais recentemente, surgiu uma plataforma de sequenciação que, em vez de se basear nas emissões de luz geradas no processo de identificação de bases, apoia-se na modificação do pH do suporte onde ocorre o processo de sequenciação – o Ion Torrent *Personal Genome Machine*TM (PGM - Life Technologies). Esta tecnologia segue um princípio básico do processo bioquímico da síntese de DNA: quando um nucleótido é incorporado na cadeia de DNA em extensão, é libertado um protão. Como consequência, o pH da solução altera-se e essa variação é detetada por sensores e transformada em informação digital (Rothberg *et al.*, 2011). Como não requer fluorescência nem sistema de deteção por câmara de leitura ótica, torna-se numa equipamento de menor tamanho, muito mais rápido e mais económico (~1000 dólares) (Quail *et al.*, 2012).

O Ion Torrent PGMTM tem por base a sequenciação por síntese num chip semicondutor. Este chip possui milhões de poços que captam informação química. O processo de sequenciação inicia-se quando o DNA a ser estudado é fragmentado em milhões de fragmentos. Cada fragmento é, então, ligado a uma *bead* (esfera) e é copiado até que toda a superfície desta *bead* esteja coberta por cópias do fragmento original. Estas *beads* são, de seguida, carregadas individualmente, num dos poços do chip. O chip vai sendo inundado faseada e ordenadamente por soluções contendo cada um dos quatro nucleótidos. Se o nucleótido for incorporado na cadeia em extensão, há alteração do pH do poço, se não for tal não se verifica. Esta alteração de pH é transformada em voltagem

e traduzida em informação digital e é diretamente proporcional ao número de nucleótidos adicionados (Flusberg *et al.*, 2010; Rothberg *et al.*, 2011) (figura 3).

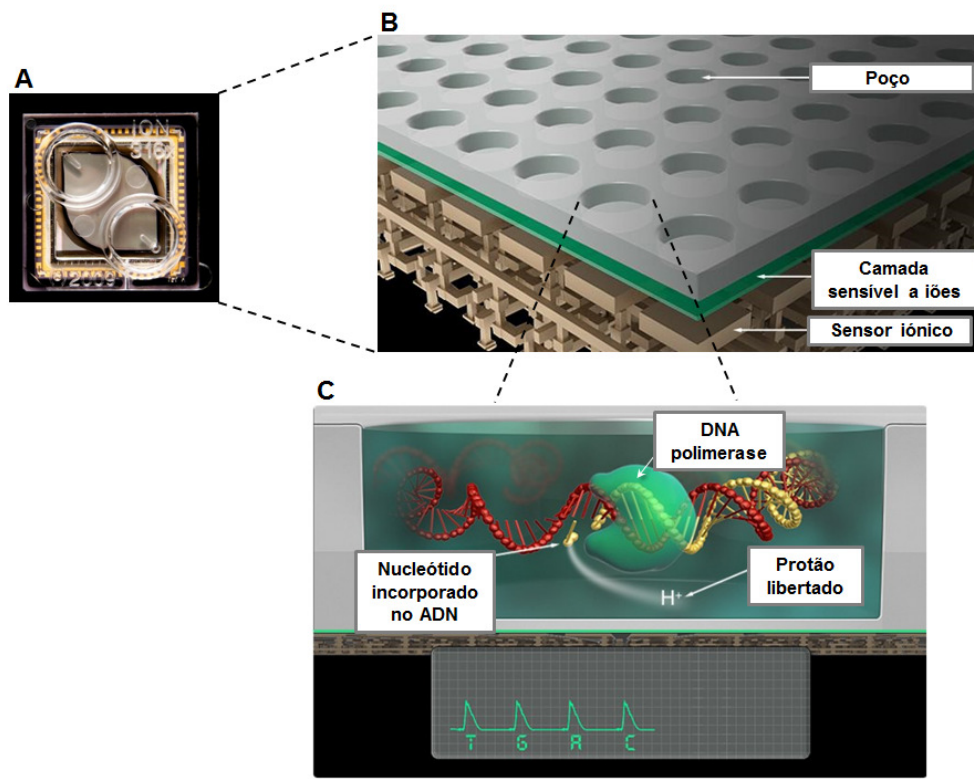


Figura 3 - Sequenciação no Ion Torrent PGM: **A)** Imagem de um chip semicondutor; **B)** Esquema representativo do interior de um chip; **C)** Esquema representativa da sequenciação por síntese que ocorre no interior de cada poço (Adaptado de: Calandro, 2012).

Este método parece ser o ideal para aplicar em projetos de sequenciação mais pequenos tal como a obtenção do genoma de microorganismos, *targeted sequencing*, sequenciação de amplificadores e para determinação de variações e alterações em populações de microorganismos (Diaz-Sanchez *et al.*, 2013).

Contudo, como cada *read* obtida por esta metodologia é menor, há mais fragmentos a necessitar de tratamento e o processo de *assembly* é mais complicado e moroso, requerendo uma análise bioinformática mais extensa. A probabilidade de ser atribuída uma base errada a uma determinada posição é maior (0,1 a 1% por *read*) do que a tradicional sequenciação de Sanger (Ross *et al.*, 2013).

Ainda assim, as tecnologias de MPS têm vindo a ser utilizadas com sucesso em diversos estudos desde o seu aparecimento, nomeadamente, em estudos de genética forense (Churchill *et al.*, 2015; Weber-Lehmann *et al.*, 2014; Young *et al.*, 2014) e para o

diagnóstico de doenças mitocondriais (Collins *et al.*, 2013; Chung *et al.*, 2015; Li *et al.*, 2014; Yeung *et al.*, 2014).

A utilização das tecnologias MPS para a detecção/quantificação de heteroplasmias no mtDNA tem sido ultimamente muito estudada. A avaliação da precisão dos haplótipos obtidos por sequenciação de mtDNA completo na plataforma Ion Torrent PGM, por comparação com aqueles obtidos pelo método de Sanger, obteve um elevado nível de concordância entre as duas técnicas (Parson *et al.*, 2013). Tang e colaboradores, elaboraram misturas a partir de amostras com haplótipos conhecidos, em diferentes proporções de modo a obterem diferentes níveis de heteroplasmias em cada mistura. De seguida, sujeitaram essas amostras a sequenciação na plataforma Illumina *Genome Analyzer* obtendo uma sensibilidade de 5% na detecção das heteroplasmias artificiais (Tang & Huang, 2010).

Os valores de sensibilidade destas técnicas permitem discriminar os indivíduos através da detecção de reduzidos níveis de heteroplasmia nos seus haplótipos e permitem um diagnóstico de doença mitocondrial mais detalhado através de exames não invasivos, ao contrário da biópsia muscular, já que apenas requerem para análise uma simples amostra de sangue.

2. OBJETIVOS

O objetivo principal deste estudo foi desenvolver, implementar e validar uma metodologia de sequenciação paralela massiva (MPS) do mtDNA, com análise de heteroplasmias, para aplicação nos campos clínico e de identificação, recorrendo à plataforma Ion Torrent PGM.

Os objetivos específicos traçados para este projeto foram:

- Desenvolver uma estratégia de Long-Range PCR para a amplificação de amostras de forma a uniformizar o mais possível a cobertura da MPS e permitir a deteção de grandes deleções no mtDNA;
- Avaliar a sequenciação do mtDNA humano completo na plataforma Ion Torrent PGM por comparação com os resultados obtidos pela sequenciação de Sanger;
- Determinar a sensibilidade, especificidade e precisão na deteção de diferentes níveis de heteroplasmias da plataforma Ion Torrent PGM através da preparação de misturas de amostras de referência de forma a criar heteroplasmias artificiais.

3. MATERIAIS E MÉTODOS

3.1. Visão global da estratégia experimental e seleção de amostras

A seleção de amostras para efetuar misturas que simulassem diferentes proporções de heteroplasmias englobou várias fases. Recolheram-se quatro amostras (T2, T3, T4 e T5) de quatro dadores voluntários saudáveis. Inicialmente, todas foram sequenciadas para a região controlo para determinação do haplogrupo, tendo-se obtido as seguintes classificações:

- T2: U5b1c;
- T3: U4a1;
- T4: H2a2a;
- T5: L1b1a12.

Para as três amostras que apresentavam haplogrupos mais distintos (T3, T4 e T5) efetuou-se a sequenciação por Sanger do genoma mitocondrial completo para determinação do haplótipo. Para essas três amostras foi, igualmente, efetuada a amplificação por Long Range-PCR, testando a estratégia da obtenção de dois fragmentos e a de um de fragmento. Após se verificar que a estratégia de amplificação em um fragmento ocorreu com sucesso, selecionaram-se os amplificados das duas amostras que apresentavam valores de concentração mais elevados (T4 e T5). Estes foram sequenciados por MPS e misturados, depois de normalizadas as suas concentrações, nas proporções indicadas na tabela 1. Os amplificados correspondentes à estratégia de dois fragmentos da amostra T4 foram também sujeitos a MPS de forma a poderem comparar-se as coberturas resultantes.

Tabela 1 - Proporções das 7 misturas para análise por MPS.

Amostras	Proporções (%)						
T4	1	5	10	50	90	95	99
T5	99	95	90	50	10	5	1

Para comparar a detecção de heteroplasmias pelas duas metodologias de sequenciação também se realizou a mistura das duas amostras para análise pelo método de Sanger. Devido aos custos e período de tempo associados à sequenciação de toda a molécula

de mtDNA pelo método de terminação de cadeia, optou-se por se analisar somente uma pequena parte da molécula de mtDNA de cada uma das amostras. Depois de amplificadas e quantificadas as concentrações de cada uma das amostras por *NanoDrop 2000c Spectrophotometer* (Thermo Scientific) e da normalização das mesmas, misturaram-se ambas de forma a se obter as proporções assinaladas na tabela 2.

Tabela 2 - Proporções das 7 misturas para análise pelo método de Sanger.

Amostras	Proporções (%)						
T4	5	10	25	50	75	90	95
T5	95	90	75	50	25	10	5

3.2. Extração do DNA

Procedeu-se à recolha de saliva de cada um dos respetivos dadores e à extração do respetivo DNA aplicando o protocolo adaptado para saliva do *DNeasy® Blood & Tissue Kit* (Qiagen). A 1 ml de saliva adicionaram-se 4 ml de PBS e centrifugou-se a 1800 *g* durante 5 min. Descartou-se o sobrenadante obtido e ressuspendeu-se o *pellet* em 180 μ l de PBS. Adicionaram-se 25 μ l de uma solução com proteinase K, 200 μ l de *Buffer AL* e misturou-se com recurso ao agitador *vortex*. De seguida, incubou-se a 56°C durante 10 min. Ao fim desse tempo, adicionaram-se 200 μ l de etanol (96-100%) a cada uma das amostras e misturou-se no *vortex*. Pipetou-se toda a mistura obtida no passo anterior para uma coluna *DNeasy Mini* colocada sobre um tubo de recolha e centrifugou-se a ≥ 6000 *g* durante 1 min. Posteriormente, descartou-se o tubo de recolha com o seu conteúdo e colocou-se a coluna sobre um novo tubo de recolha. Adicionaram-se 500 μ l de *Buffer AW1* e centrifugou-se a ≥ 6000 *g* durante 1 min. Descartou-se novamente o tubo de recolha com o seu conteúdo e colocou-se a coluna sobre um novo tubo de recolha. Adicionaram-se 500 μ l de *Buffer AW2* e centrifugou-se a 20000 *g* durante 3 min. Descartou-se de novo o tubo de recolha com o seu conteúdo e colocou-se a coluna sobre um *ependorf*. Por fim, o DNA foi eluído da coluna com 100 μ l de *Buffer AE*, após incubação à temperatura ambiente durante 1 min e centrifugação a ≥ 6000 *g* durante 1 min, e mantido a -20°C até à sua utilização.

3.3. Amplificação das amostras por PCR/PCR multiplex e sequenciação pelo método de Sanger

3.3.1. Amplificação das amostras

- Estratégia para a amplificação da região controlo:

Para cada uma das quatro amostras utilizaram-se os pares de *primers* MTT1 e MTT2 (tabela 3), em reações separadas, para amplificar a região controlo.

- Estratégia para a amplificação do mtDNA completo:

Para cada amostra (T3, T4 e T5) utilizaram-se 12 pares de *primers* num PCR *multiplex* cujos fragmentos resultantes se sobrepõem de modo a garantir a cobertura total da molécula de mtDNA (tabela 3) (Ramos *et al.*, 2009 & 2011). Os 12 pares de *primers* foram combinados em duas *primer mix* de forma a poder amplificar-se simultaneamente 6 fragmentos em cada reação. Uma vez que os fragmentos amplificados são sobreponíveis, e de forma a evitar a amplificação de fragmentos mais pequenos que o desejado, os pares de *primers* foram distribuídos alternadamente pelas duas *mix*. Assim, na *primer mix* 1 colocaram-se os pares de *primers* MTT1, MTT3A, MTT4A, MTT5A, MTT6 e MTT8 e na *primer mix* 2 os pares de *primers* MTT2, MTT3B, MTT4B, MTT5B, MTT7 e MTT9. Em cada *mix*, colocou-se 2 μ l de cada *primer* e completou-se com água desionizada/*RNAse free* até perfazer 100 μ l de volume final, de forma a cada *primer* ficar a uma concentração de 2 μ M.

- Estratégia para a amplificação do fragmento para as misturas:

Para cada amostra (T4 e T5) utilizou-se o par de *primers* MTT1 para amplificar um fragmento para ser misturado em diferentes proporções.

Tabela 3 - Localização da posição de ligação e sequência dos 12 pares de *primers* utilizados para amplificação da molécula de mtDNA.

Pares	Nome	Localização	Sequência
MTT1	MTT1_F1	14898_F	5'-TAGCCATGCACTACTCACCAGA-3'
	MTT1_R1	151_R	5'- GGATGAGGCAGGAATCAAAGAC-3'
MTT2	MTT2_F1	16488_F	5'- CTGTATCCGACATCTGGTTCCT-3'
	MTT2_R1	1677_R	5'- GTTTAGCTCAGAGCGGTCAAGT-3'
MTT3A	MTT3_F1	1404_F	5'- ACTTAAGGGTCTGAAGGTGGATT-3'
	MTT3_R3	2801_R	5'- TAATGCAGGTTTGGTAGTTTA-3'
MTT3B	MTT3_F3	2646_F	5'- GGTTTCAGCTGTCTCTTACTTT-3'
	MTT3_R1	3947_R	5'- TCGATGTTGAAGCCTGAGACTA-3'
MTT4A	MTT4_F1	3734_F	5'- AAGTCACCCTAGCCATCATTCTA-3'
	MTT4_R3	5571_R	5'- AAGTATTGCAACTTACTGAGG-3'
MTT4B	MTT4_F4	5468_F	5'- CACGCTACTCCTACCTATCTC-3'
	MTT4_R1	6739_R	5'- GATATCATAGCTCAGACCATACC-3'
MTT5A	MTT5_F1	6511_F	5'- CTGCTGGCATCACTATACTACTA-3'
	MTT5_R3	8000_R	5'- CAACGTCAAGGAGTCGCAGGT-3'
MTT5B	MTT5_F3	7713_F	5'- TCCTAACACTCACAAACAAAAC-3'
	MTT5_R1	9220_R	5'- GATTGGTGGGTCATTATGTGTTG-3'
MTT6	MTT6_F1	8910_F	5'- CTTACCACAAGGCACACCTACA-3'
	MTT6_R1	10648_R	5'- GGCACAATATTGGCTAAGAGGG-3'
MTT7	MTT7_F1	10360_F	5'- GTCTGGCCTATGAGTGACTACA-3'
	MTT7_R1	12226_R	5'- CAGTTCTTGTGAGCTTTCTCGG-3'
MTT8	MTT8_F1	11977_F	5'- CTCCTCTACATATTTACCACAAC-3'
	MTT8_R1	13830_R	5'- AAGTCCTAGGAAAGTGACAGCGA-3'
MTT9	MTT9_F1	13477_F	5'- GCAGGAATACCTTTCCTCACAG-3'
	MTT9_R1	15349_R	5'- GTGCAAGAATAGGAGGTGGAGT-3'

- Amplificação:

Para a amplificação, utilizaram-se dois tubos distintos para cada amostra. A cada tubo adicionaram-se 15 µl de água desionizada/*RNAse free*, 5 µl de par de *primer/primer mix* e, por fim, 25 µl de *Qiagen Multiplex PCR Kit* (Qiagen). De seguida, foram adicionados 5 µl de DNA da amostra nos dois tubos, perfazendo assim 50 µl de volume de reação.

Para cada corrida de amplificação foi efetuado um controlo negativo que continha a mesma mistura reacional mas sem DNA adicionado.

Todos os passos, à exceção da adição do DNA, foram executados numa sala de pré-PCR de forma a evitar contaminações. A amplificação foi executada num termociclador programado com as condições apresentadas na figura 4. A partir da amplificação o procedimento decorreu no laboratório de pós-PCR.

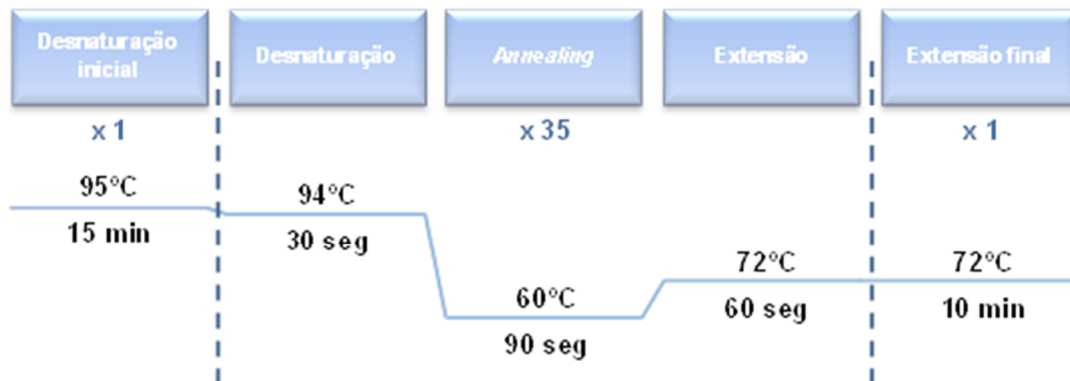


Figura 4 - Condições para a amplificação do mtDNA completo por PCR multiplex.

Para a deteção dos produtos de PCR, os mesmos foram sujeitos a eletroforese horizontal através de gel de poliacrilamida T9C5. A visualização das bandas foi alcançada pela coloração do gel por nitrato de prata (Budowle *et al.* 1991).

3.3.2. Sequenciação do mtDNA pelo método de Sanger

A partir deste momento, para cada fragmento a ser sequenciado realizaram-se reações independentes. Após a amplificação, os produtos de PCR foram submetidos a uma purificação por *Exosap-IT* (GE Healthcare) para a remoção dos dNTPs e *primers* não utilizados na reação de PCR, de modo a evitar que interferissem na reação de sequenciação. Utilizou-se 1 µl de *Exosap-IT* para cada 1,5 µl de amostra. Esta reação ocorreu num termociclador durante 15 min a 37°C e seguiu-se um passo para a inativação da enzima durante 15 min a 80°C.

Aos 2,5 µl de amostra já purificada adicionou-se 1 µl de *BigDye Terminator v3.1 Cycle Sequencing Kit* (Applied Biosystems®), 1 µl de *Buffer* (Applied Biosystems®), e 0,5 µl do *primer* (2 µM) correspondente ao fragmento a sequenciar. De modo a sequenciar a

molécula completa de mtDNA foi utilizada uma estratégia com 31 fragmentos sobreponíveis (tabela 4) (Ramos *et al.*, 2009 & 2011). Todos os fragmentos foram sequenciados no sentido *forward* (cadeia L) e apenas no sentido *reverse* (cadeia H) quando, após análise dos resultados, se detetava a falha da sequenciação da primeira estratégia, normalmente, devido à presença de regiões homopoliméricas.

Tabela 4 - Localização da posição de ligação e sequência dos *primers* utilizados para sequenciar a molécula de mtDNA.

Nome	Local.	Sequência	Nome	Local.	Sequência
MTT1_F1	14898_F	5'-TAGCCATGCACTACTCACCAGA-3'	MTT5_F2	7111_F	5'-ACACCCTAGACCAAACCTACG-3'
MTT1_F2	15416_F	5'-TACACAATCAAAGACGCCCTC-3'	MTT5_F3	7713_F	5'-TCCTAACACTCACAACAAAAC-3'
MTT1_F3	15966_F	5'-AGTCTTTAACTCCACCATTAG-3'	MTT5_F4	8311_F	5'-TAGCATTAACTTTTAAGTTA-3'
MTT1_R1	151_R	5'-GGATGAGGCAGGAATCAAAGAC-3'	MTT5_R1	9220_R	5'-GATTGGTGGGTCATTATGTGTTG-3'
MTT1_R2	16281_R	5'-GTTGGTATCCTAGTGGGTGAG-3'	MTT5_R2	8600_R	5'-AGAATGATCAGTACTGCGGCG-3'
MTT1_R3	15825_R	5'-GTGAAGTATAGTACGGATGCT-3'	MTT5_R3	8000_R	5'-CAACGTCAAGGAGTCGCAGGT-3'
MTT2_F1	16488_F	5'-CTGTATCCGACATCTGGTTCCT-3'	MTT5_R4	7359_R	5'-CTACTATTAGGACTTTTCGCT-3'
MTT2_F2	411_F	5'-CGGTATGCACTTTTAAACAGTC-3'	MTT6_F1	8910_F	5'-CTTACCACAAGGCACACCTACA-3'
MTT2_F3	909_F	5'-GATTAACCCAAGTCAATAGAA-3'	MTT6_F2	9393_F	5'-CGAGAAAGCACATACCAAGGC-3'
MTT2_R1	1677_R	5'-GTTTAGCTCAGAGCGGTCAAGT-3'	MTT6_F3	9874_F	5'-TAATATTTCACTTTACATCCA-3'
MTT2_R2	1159_R	5'-TAAGCTGTGGCTCGTAGTGT-3'	MTT6_R1	10648_R	5'-GGCACAATATTGGCTAAGAGGG-3'
MTT2_R3	638_R	5'-GGTGATGTGAGCCCGTCTAAA-3'	MTT6_R2	10154_R	5'-TTCTATGTAGCCGTTGAGTTG-3'
MTT3_F1	1404_F	5'-ACTTAAGGGTGAAGGTGGATT-3'	MTT6_R3	9647_R	5'-AGCTCAGGTGATTGATACTCC-3'
MTT3_F2	2028_F	5'-GATAGAATCTTAGTCAACTT-3'	MTT7_F1	10360_F	5'-GTCTGGCCTATGAGTGACTACA-3'
MTT3_F3	2646_F	5'-GGTTCAGCTGTCTCTACTTT-3'	MTT7_F2	10892_F	5'-ATCAACAACAACCTATTTAGC-3'
MTT3_F4	3239_F	5'-GCAGAGCCCGTAATCGCATA-3'	MTT7_F3	11461_F	5'-ACTCTTAAACTAGGCGGCTA-3'
MTT3_R1	3947_R	5'-TCGATGTTGAAGCCTGAGACTA-3'	MTT7_R1	12226_R	5'-CAGTTCTTGTGAGCTTTCCTCGG-3'
MTT3_R2	3382_R	5'-TTCGTTCCGGTAAGCATTAGGA-3'	MTT7_R2	11673_R	5'-GTTTGGATGAGAATGGCTGTT-3'
MTT3_R3	2801_R	5'-TAATGCAGGTTTGGTAGTTTA-3'	MTT7_R3	11163_R	5'-CGGGTGATGATAGCCAAGGTG-3'
MTT3_R4	2263_R	5'-GGGTGTGAGGAGTTCAGTTAT-3'	MTT8_F1	11977_F	5'-CTCCCTCTACATATTTACCACAAC-3'
MTT4_F1	3734_F	5'-AAGTCACCCTAGCCATCATTCTA-3'	MTT8_F2	12500_F	5'-TGTGCCTAGACCAAGAAGTTA-3'
MTT4_F2	4346_F	5'-GAACCCATCCCTGAGAATCCA-3'	MTT8_F3	12988_F	5'-CTAGCAGCAGCAGGCAAATCA-3'
MTT4_F3	4896_F	5'-TACCAAATCTCTCCCTCACTA-3'	MTT8_R1	13830_R	5'-AAGTCCTAGGAAAGTGACAGCGA-3'
MTT4_F4	5468_F	5'-CACGCTACTCCTACCTATCTC-3'	MTT8_R2	13297_R	5'-GGTTGATGCCGATTGTAACTA-3'
MTT4_F5	5995_F	5'-CAGCTCTAAGCCTCCTTATTC-3'	MTT8_R3	12763_R	5'-CGATGAACAGTTGGAATAGGT-3'
MTT4_R1	6739_R	5'-GATATCATAGCTCAGACCATAACC-3'	MTT9_F1	13477_F	5'-GCAGGAATACCTTTCCTCACAG-3'
MTT4_R2	6154_R	5'-GGAAGTGTAGTCAGTTGCCAAAG-3'	MTT9_F2	13950_F	5'-CTATCTAGGCCTTCTTACGAG-3'
MTT4_R3	5571_R	5'-AAGTATTGCAACTTACTGAGG-3'	MTT9_F3	14440_F	5'-ATACTCCTCAATAGCCATCGC-3'
MTT4_R4	5017_R	5'-GAGGAGTATGCTAAGATTTTG-3'	MTT9_R1	15349_R	5'-GTGCAAGAATAGGAGGTGGAGT-3'
MTT4_R5	4421_R	5'-CTTATTTAGCTGACCTTACTT-3'	MTT9_R2	14838_R	5'-CATCATGCGGAGATGTTGGAT-3'
MTT5_F1	6511_F	5'-CTGCTGGCATCACTATACTACTA-3'	MTT9_R3	14325_R	5'-AACTTTAATAGTGTAGGAAGC-3'

A reação de sequenciação foi executada num termociclador que apresentava a programação descrita na figura 5.

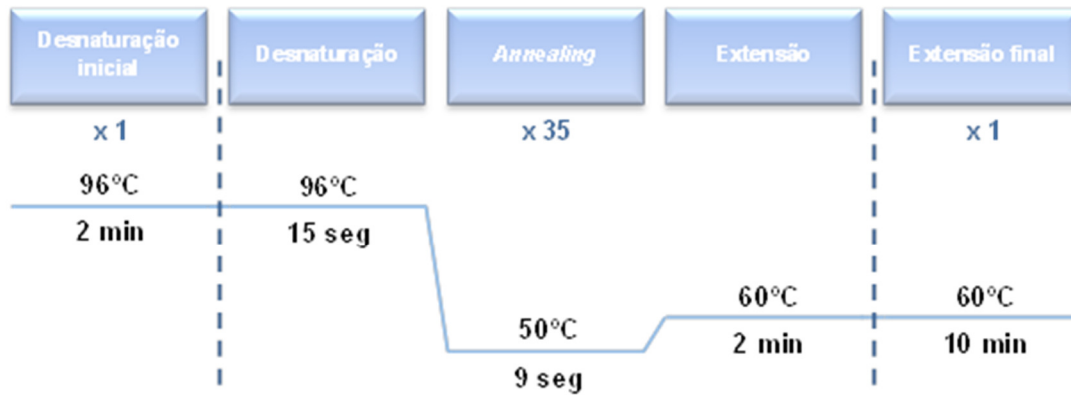


Figura 5 - Condições para a reação de sequenciação do mtDNA.

No final, os produtos de sequenciação foram submetidos a uma nova purificação de modo a remover os dNTPs não utilizados na reação de sequenciação. Colocaram-se colunas apropriadas em tubos de recolha sobre as quais se pipetaram 750 µl de Sephadex™ G50 (GE Healthcare) e centrifugou-se a 4400 rpm durante 4 min. De seguida, as colunas foram transferidas para *ependorfs* e pipetou-se a totalidade das amostras sobre estas colunas. Procedeu-se a uma nova centrifugação a 4400 rpm durante 4 min. A cada tubo *ependorf* com amostra purificada adicionaram-se 10 µl de formamida Hi-Di™ Formamide (Applied Biosystems). Por fim, todos os fragmentos foram sujeitos à eletroforese capilar desnaturante nos equipamentos Applied Biosystems Genetic Analyzer 3130 ou Applied Biosystems Genetic Analyzer 3130xl (Applied Biosystems®).

3.3.3. Visualização e análise das sequências

A visualização dos eletroferogramas obtidos foi realizada com recurso ao *software* Geneious Pro v.5.5.8 (Biomatters) e SeqScape v.2.5 (Life Technologies) que alinham e comparam as amostras com a sequência de referência do mtDNA – a rCRS (*revised Cambridge Reference Sequence*) (Andrews *et al.*, 1999). Isto leva à identificação de variantes presentes nas amostras e, conseqüentemente, aos haplótipos das mesmas utilizando a nomenclatura adotada pela ISFG (*International Society for Forensic Genetics*) (Carracedo *et al.*, 2000).

Para efeitos de comparação entre as duas metodologias de sequenciação, as misturas foram analisadas em ambos os *software* referidos com o limite de deteção de heteroplasmias definido para valores superiores a 5% (figura 25).

A classificação em haplogrupos foi feita através da utilização da ferramenta *web HaploGrep* (Kloss-Brandstaetter *et al.*, 2010) que se baseia na mais atual versão da árvore filogenética das variações mitocondriais *PhyloTree* (van Oven M. *et al.*, 2009).

3.4. Amplificação das amostras por Long-Range PCR e sequenciação pela plataforma Ion Torrent PGM

3.4.1. Amplificação das amostras por Long-Range PCR

Para cada uma das amostras procedeu-se à amplificação completa da molécula de mtDNA em dois fragmentos de ~8,5 kbs (Fendt *et al.*, 2009) e num só fragmento de ~16,5 kbs. Para a estratégia de um fragmento, utilizaram-se primers (tabela 5) desenhados com recurso ao *software Primer-BLAST* da NCBI (National Center for Biotechnology Information: <http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) (Ye *et al.*, 2012).

Tabela 5 - Sequência dos *primers* utilizados para amplificar a molécula de mtDNA.

<i>Primers</i>	Sequência
MT_16555L_LRF	5'-TAAGACATCACGATGGATCACAGGT-3'
MT_16554H_LRR	5'-TTTAAGGGGAACGTGTGGGCTATTT-3'

As posições de ligação dos *primers* na molécula de mtDNA (16554 e 16555) foram escolhidas por se encontrarem numa zona muito conservada, posicionada entre os segmentos HVRI e HVRII. Os dois *primers* iniciavam-se em posições adjacentes de forma a amplificarem a molécula de mtDNA completa num só fragmento (figura 6).

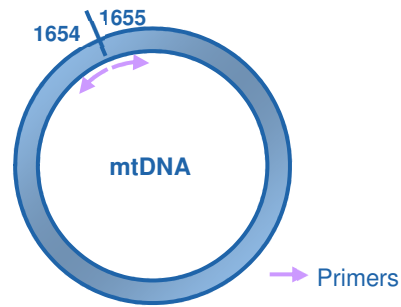


Figura 6 - Esquema representativo das posições de ligação dos *primers* MT_1655L e MT_1654H na molécula de mtDNA.

Para a amplificação, adicionaram-se, a cada tubo de cada amostra, 10 μ l de *5x Ranger Reaction Buffer* (Bioline), 2 μ l de cada *primer* (a 20 μ M cada um), 1 μ l de enzima *Ranger DNA Polymerase* (Bioline), 3 μ l de DNA de cada uma das amostras e 32 μ l de água *DNase/RNase free* de modo a se obter 50 μ l de volume reacional. Para cada corrida de amplificação foi necessário um controlo negativo que contém a mesma mistura reacional mas sem DNA.

Todos os passos foram executados numa sala de pré-PCR de forma a evitar contaminações.

A reação de amplificação por Long-Range PCR foi executada num termociclador programado com as condições descritas na figura 7.

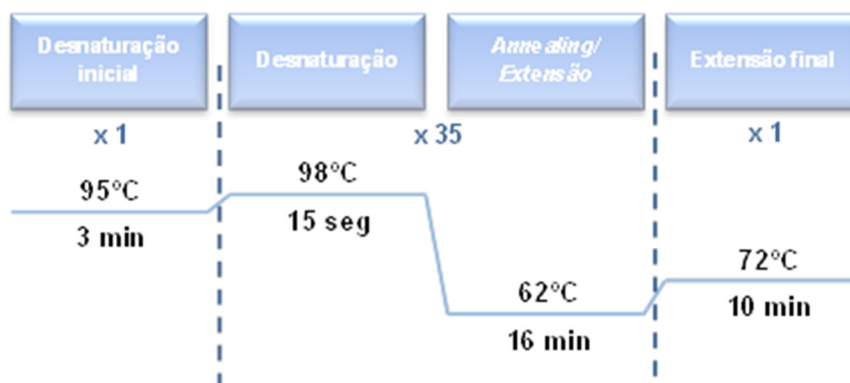


Figura 7 - Condições de amplificação por Long-Range PCR do mtDNA.

A partir da amplificação o procedimento decorreu no laboratório de pós-PCR.

Para a detecção dos produtos de Long-Range PCR, os mesmos foram sujeitos a eletroforese horizontal através de gel de agarose a 0,8% durante uma a uma hora e meia.

3.4.2. Purificação dos amplificados

Os produtos de PCR foram purificados através da utilização de *Agencourt® AMPure® XP Kit* (Beckman Coulter) (figura 8). A um volume de 20 µl de cada amostra amplificada adicionaram-se 36 µl de *Agencourt®* e misturou-se a solução, pipetando *up and down*. Deixou-se incubar à temperatura ambiente durante 5 min. De seguida, colocaram-se os tubos num suporte magnético durante 3 min (ou até que a solução ficasse limpa). Retirou-se o sobrenadante cuidadosamente, de forma a não perturbar as *beads* depositadas. Logo a seguir, adicionaram-se 500 µl de etanol (70%) a cada tubo e incubou-se durante 30 seg, rodando duas vezes os tubos no suporte. Retirou-se de novo o sobrenadante e realizou-se uma segunda lavagem com etanol, tal como a descrita anteriormente. Retirou-se, novamente, o sobrenadante e, realizou-se um *spin down* na centrífuga. Voltou a colocar-se os tubos no suporte, retirou-se o etanol residual e deixaram-se secar ao ar num período inferior a 5 min. Retiraram-se os tubos do suporte magnético, adicionou-se 25 µl de *LowTE* e pipetou-se *up and down* de forma a ressuspender as *beads*. Colocou-se no agitador *vortex* e efetuou-se um *spin down* na centrífuga. Voltou a colocar-se os tubos no suporte magnético até que a solução ficasse limpa. Por fim, retirou-se (apenas até 20 µl de modo a não perturbar as *beads*) e guardou-se o sobrenadante.

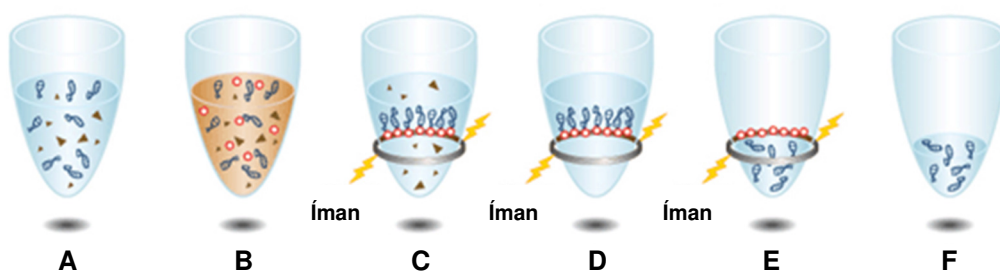


Figura 8 - Esquema representativo do método de purificação por *Agencourt®*: **A)** Produtos finais do PCR (amplificados e resíduos); **B)** Adição de *Agencourt* contendo *beads*; **C)** Ligação dos amplificados às *beads* que são atraídas pelo suporte magnético; **D)** Lavagem com etanol; **E)** Libertação dos amplificados das *beads*; **F)** Solução purificada após separação das *beads* (Adaptada de: Beckman Coulter, 2013).

3.4.3. Quantificação dos amplificadores purificados

Quantificou-se o DNA das amostras utilizando o fluorímetro *Qubit® 2.0* (Life Technologies). Para isso, preparou-se uma solução contendo 199 μ l de *HS Buffer* e 1 μ l de fluoróforo por amostra. Nos tubos do fluorímetro colocou-se 199 μ l da solução preparada e 1 μ l da amostra. Misturou-se no *vortex* e incubou-se durante 2 min e efetuou-se a leitura.

3.4.4. Preparação da biblioteca

A preparação da biblioteca inicia-se com a digestão enzimática do DNA para se obterem fragmentos de DNA de igual tamanho. De seguida, dá-se a ligação de adaptadores a cada uma das extremidades dos fragmentos de DNA, contendo ambas zonas de ligação de *primers*. O adaptador P1 é complementar à cadeia de DNA presente nas esferas ISP (*Ion Sphere Particles*), que vão ser utilizadas mais à frente no PCR de emulsão. Aos adaptadores A vão-se ligar diferentes *barcodes*, de modo a permitir a sequenciação de várias amostras distintas em simultâneo – adaptador X (figura 9).

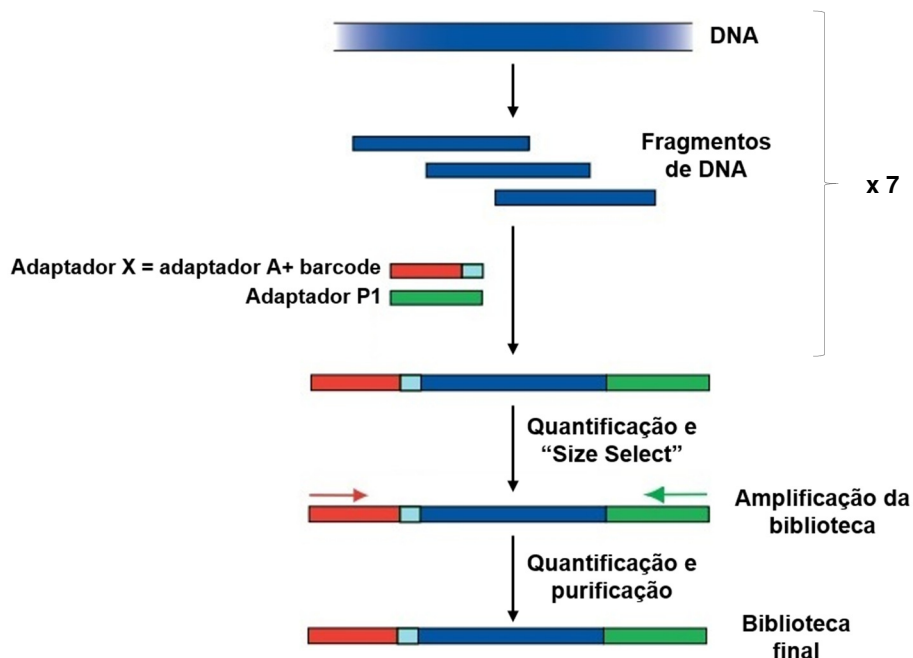


Figura 9 - Esquema representativo de todos os passos incluídos na preparação da biblioteca (Adaptada de: Life Technologies, 2013).

Para a realização de todos estes passos foi utilizado o sistema semi-automático *AB Library Builder™* (Applied Biosystems). Este sistema necessita da preparação de uma *strip* por cada amostra, que contém já alguns dos poços preenchidos com reagentes e que vai ser inserida no equipamento de forma a este desempenhar as suas funções (figura 10).

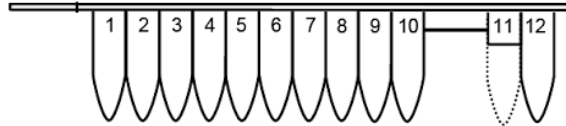


Figura 10 - Esquema representativo das posições da *strip* do sistema *AB Library Builder™* (Adaptada de: Life Technologies, 2013).

Os valores de concentração, obtidos no passo anterior, foram inseridos numa folha *Excel* para calcular os valores para a diluição de cada amostra (amplificados purificados), em *E1 Buffer*, de forma a se obterem 100 ng em 50 μ l. Antes de se preencherem os poços, juntou-se, ainda, num tubo 48 μ l de *E1 Buffer* com 1 μ l de adaptador P1 por amostra (mais um adicional). Adicionou-se 49 μ l da solução anterior aos tubos de *LdN Buffer* fornecidos (um por amostra). Estes tubos foram, depois, marcados com o *barcode* a acrescentar e adicionou-se 1 μ l do *barcode* correspondente a cada tubo. Cada *strip* foi, depois, preenchida da seguinte maneira:

- No poço 12 foram colocados os volumes, anteriormente calculados, de amostra e *E1 Buffer*;
- No poço 11 colocaram-se os tubos (destapados) com os *barcodes*;
- O poço 10 foi furado e adicionaram-se 500 μ l de *Agencourt®*.

Este sistema também requiere o preenchimento de um suporte de reagentes do equipamento com tubos de reagentes fornecidos. Esse suporte foi preenchido com tubos vazios para a eluição das amostras, pontas e *Ion End Repair Enzyme Mix do Plus Library Kit*.

De seguida, colocou-se o suporte e as *strips* no equipamento e ligou-se o aparelho programado para a obtenção de fragmentos com cerca de 260 bp. No final, retiraram-se os tubos das amostras do suporte e colocaram-se no suporte magnético de forma a remover algum *Agencourt*.

- Quantificação das amostras fragmentadas e normalização das amostras:

Efetou-se uma quantificação das amostras fragmentadas de igual forma ao processo descrito na secção “Quantificação dos amplificados purificados”. Com base nos valores de concentração obtidos, juntaram-se todas as amostras, marcadas com os *barcodes*, em proporções equimolares numa mesma *pool*.

- “Size Select”:

Para a seleção de fragmentos apenas com o tamanho desejado utilizou-se gel de agarose *E-Gel® SizeSelect™* (Invitrogen Corporation). Para isso, diluíram-se 2 µl de *Marker* em 18 µl de água. Colocaram-se 20 µl da biblioteca num dos poços da fila superior do gel e 10 µl do *Marker* diluído no poço superior correspondente à posição “M”. De seguida, adicionaram-se 25 µl de água nos restantes poços superiores e em todos os poços inferiores – poços de recolha (no poço “M” apenas cabem 10 µl). Colocou-se a tampa e executou-se o programa “SizeSelect 2%” durante 12 a 14 min. Decorrido este tempo, aplicou-se novamente 10 µl de água nos poços inferiores. Voltou a executar-se o programa até que a banda mais intensa do *Marker* entrasse por completo no poço de recolha. Recolheu-se a amostra do poço de recolha correspondente, 10 µl por cada duas vezes (figura 11). Por fim, lavou-se o poço com 10 µl de água e voltou a recolher-se a amostra.



Figura 11 - Imagem do *E-Gel® SizeSelect™* (Retirada de: Thermo Fisher, 2014).

- Amplificação da biblioteca:

De forma a incrementar o número de fragmentos de tamanho desejado realizou-se uma amplificação por PCR. Para isso, juntaram-se 100 µl de *Platinum PCR SuperMix High Fidelity* (Thermo Fisher), 5 µl de *Library Amplification Primer* (Thermo Fisher) e 25 µl da biblioteca preparada anteriormente. Colocou-se no termociclador com a programação descrita na figura 12.

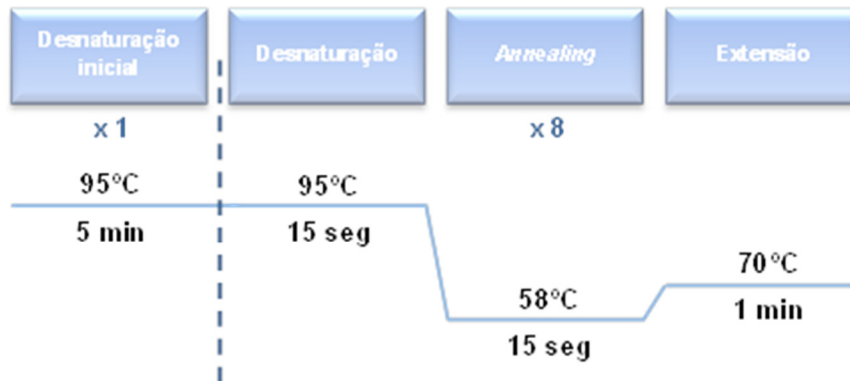


Figura 12 - Condições para a amplificação da biblioteca.

- Purificação da biblioteca:

A biblioteca amplificada foi purificada através da utilização de *Agencourt® AMPure® XP Kit* (Beckman Coulter) de modo a remover-se produtos indesejados resultantes do PCR. Adicionou-se um volume de 130 µl da biblioteca amplificada a 195 µl de *Agencourt®* e misturou-se a solução pipetando *up and down*. Deixou-se incubar à temperatura ambiente durante 5 min. De seguida, colocaram-se os tubos num suporte magnético durante 3 min (ou até que a solução ficasse limpa). Retirou-se o sobrenadante cuidadosamente, de forma a não perturbar as *beads* depositadas. Logo a seguir, adicionaram-se 500 µl de etanol (70%) a cada tubo e incubou-se durante 30 seg, rodando duas vezes os tubos no suporte. Retirou-se de novo o sobrenadante e realizou-se uma segunda lavagem com etanol, tal como a descrita anteriormente. Retirou-se, novamente, o sobrenadante e, de forma a não ficar etanol residual, realizou-se um *spin down* na centrífuga. Voltou a colocar-se os tubos no suporte e deixaram-se secar ao ar num período inferior a 5 min. Retiraram-se os tubos do suporte magnético, adicionaram-se 20 a 25 µl de *LowTE* e pipetou-se *up and down* de forma a ressuspender as *beads*. Colocou-se no agitador *vortex* e efetuou-se um *spin down* na centrífuga. Voltou a colocar-se os tubos no suporte magnético até que a solução ficasse limpa. Por fim,

retirou-se (apenas até 20 µl de modo a não perturbar as *beads*) e guardou-se o sobrenadante.

- Quantificação e diluição da biblioteca:

Efetou-se uma quantificação das amostras fragmentadas de igual forma ao processo descrito na secção “Quantificação dos amplificadores purificados”. Com base nos valores de concentração obtidos, calculou-se o fator de diluição necessário para obter uma concentração de 0,6 nM. - TDF (*Template Dilution Factor*) – utilizando uma folha *Excel*. De acordo com o valor obtido, diluiu-se a biblioteca e agitou-se no *vortex*.

3.4.5. Preparação do *template*

Neste passo, cada fragmento de DNA irá ficar ligado a uma única ISP (*Ion Sphere™ Particle*) com adaptador P1. É, ainda, realizada a amplificação clonal, de forma a que cada ISP fique revestida por várias cópias do mesmo fragmento de DNA (figura 13). Ao mesmo tempo, o adaptador X é marcado com biotina para ser possível realizar-se o enriquecimento das ISPs. Durante o enriquecimento, o lado do conjugado ISP-*template* marcado com biotina vai ligar-se a esferas magnéticas revestidas com estreptavidina, que vão ser atraídas por um íman. Todas as ISPs sem fragmentos de DNA vão ser removidas por lavagem e as que ISP com amplificadores vão ser desnaturadas (figura 14).

-PCR de emulsão:

Neste passo, realizou-se a amplificação clonal utilizando-se o sistema semi-automático *OneTouch™* (Thermo Fisher). Assim, preparou-se a solução de amplificação para um volume total de 900 µl juntando 25 µl de água, 500 µl de *Reagent Mix*, 300 µl de *Reagent B*, 50 µl de *Enzyme Mix* e 25 µl da biblioteca diluída. Agitou-se no *vortex* e fez-se um *spin down*. Prepararam-se as ISPs, adicionaram-se 100 µl destas à solução de amplificação preparada anteriormente e agitou-se no *vortex*. Por fim, carregou-se o filtro do equipamento, de forma lenta e contínua, com 1000 µl de solução de amplificação, de seguida, com 1000 µl de *Reaction Oil* e, como uma nova ponta, mais 500 µl de *Reaction Oil*. Inverteu-se o filtro e inseriu-se no aparelho.

No final da corrida, efetuou-se uma centrifugação final de cerca de 10 min. Retiraram-se os tubos e, em cada tubo, retirou-se cuidadosamente o sobrenadante com uma pipeta de 1000 µl do lado oposto ao sedimento, deixando apenas cerca de 50 µl no fundo dos tubos. Com uma nova ponta, ressuspendeu-se o sedimento nos 50 µl de cada tubo, pipetando *up and down*. Transferiram-se as soluções de ambos os tubos para um novo *ependorff*.

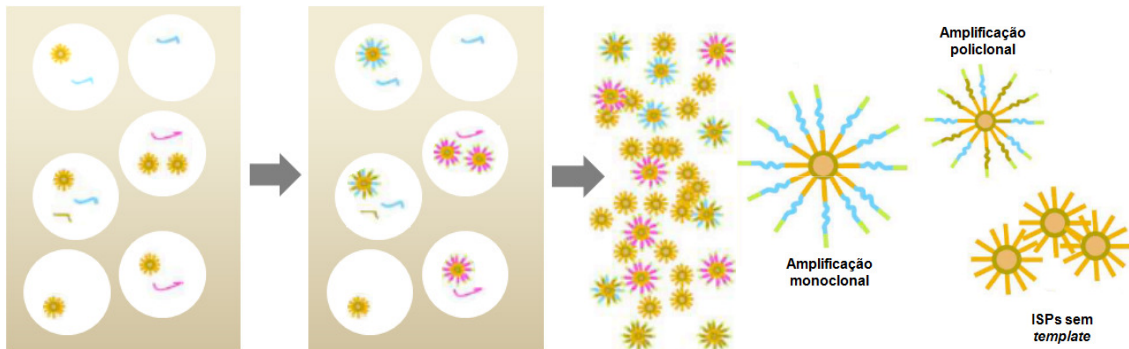


Figura 13 - Esquema representativo do processo do emPCR. No interior de cada bolha podem encontrar-se dispostos ao acaso fragmentos de DNA e ISPs. Após o processo de amplificação podemos obter uma variedade de produtos para além dos desejados (Adaptado de: Inger Jonasson).

- Enriquecimento do *template*:

Os conjugados *ISP-template* resultantes do passo anterior foram, então, enriquecidos utilizando o sistema *OneTouch™ES* no qual se colocou, no início, uma nova ponta e um tubo de recolha de 200 µl contendo 10 µl de *Neutralization Solution*. Preparou-se a solução *Melt-Off* juntando 280 µl de *Tween* com 40 µl de NaOH (1M). Agitou-se no *vortex* o tubo contendo *Dynabeads® MyOne™ Streptavidin C1* (Life Technologies) e retiraram-se 13 µl para um novo *ependorff*. De seguida, colocou-se o *ependorff* no suporte magnético durante 2 min, retirou-se cuidadosamente o sobrenadante e adicionaram-se 130 µl de *MyOne Beads Wash Solution*. Por fim, ressuspendeu-se e agitou-se no *vortex* a solução.

Preencheu-se a *strip* de 8 poços do *Ion OneTouch™ES* na seguinte ordem:

- Poço 1: todo o volume de *template* (cerca de 100 µl);
- Poço 2: 130 µl das *Dynabeads* preparadas;
- Poços 3 a 5: 300 µl de *OneTouch Wash Solution* em cada poço;

- Poço 6: mantém-se vazio;
- Poço 7: 300 µl da solução *Melt-Off* preparada;
- Poço 8: mantém-se vazio.

Colocou-se a *strip* no equipamento e iniciou-se o aparelho.

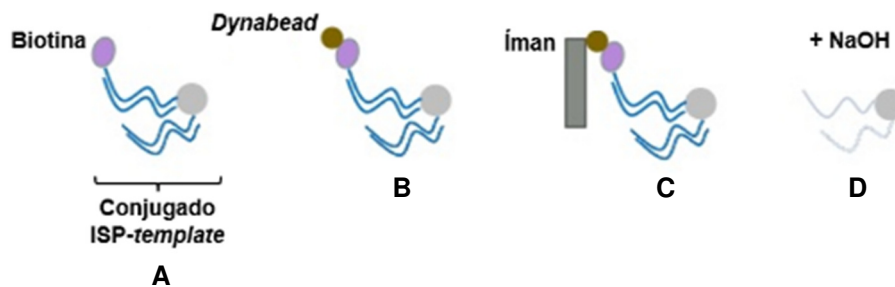


Figura 14 - Esquema representativo do processo de enriquecimento: **A)** Biotina liga-se ao adaptador X; **B)** A biotina liga-se às esferas de estreptavidina *Dynabeads*; **C)** As *Dynabeads* são atraídas para um íman e as impurezas são removidas; **D)** O conjugado *ISP-template* é desnaturado com NaOH (Adaptado de: ThermoFisher).

3.4.6. Limpeza e inicialização do PGM

Antes da inicialização do equipamento Ion Torrent PGM foi realizada uma limpeza do mesmo de acordo com o manual fornecido (*Ion PGM™ Hi-Q™ Sequencing Kit*). De seguida, planeou-se a corrida no *Torrent Browser*.

3.4.7. Preparação e carregamento do chip

Após preparação das *Control Ion Sphere Particles* (para controlo de qualidade), adicionaram-se 5 µl destas ao volume total do *template* de ISPs enriquecido. Misturou-se pipetando *up and down* e centrifugou-se durante 2 min a 15500 *g*. De seguida, retirou-se cuidadosamente o sobrenadante deixando 15 µl no tubo. Adicionaram-se 12 µl de *sequencing primer* e confirmou-se o volume final de 27 µl. Misturou-se pipetando *up and down*. Colocou-se no termociclador a 95°C durante 2 min seguido de 37°C durante 2 min para o *annealing* do *sequencing primer* às ISPs. No final adicionaram-se 3 µl de *Ion PGM Sequencing Polymerase* às ISPs, para um volume final de 30 µl. Pipetou-se *up and down* e incubou-se à temperatura ambiente durante 5 min.

Posteriormente, testou-se o funcionamento do chip no PGM, segundo as instruções do aparelho.

Sem utilizar luvas, retirou-se o chip do PGM e inclinou-se o chip a 45°, com a porta de carregamento para baixo, inseriu-se a pipeta na porta de carregamento e, com a pipeta perpendicular ao chip, removeu-se o máximo possível de líquido. De seguida, colocou-se o chip virado para baixo no suporte da centrífuga, com o *tab* (marcação no chip) apontando para o centro. Centrifugou-se durante 5 seg. A seguir, retirou-se o chip e pousou-se virado para cima no suporte. Aspirou-se todo o volume da amostra e inseriu-se a ponta da pipeta na porta de carregamento, com a pipeta na vertical. Inseriu-se, rodando o êmbolo da pipeta, de forma lenta e contínua, as ISPs no chip (deixando um pequeno volume na pipeta no final). Centrifugou-se o chip durante 30 seg. Com a pipeta regulada para o volume total da amostra, inclinou-se o chip a 45° com a porta de carregamento para baixo, e pipetou-se lentamente a amostra para fora e para dentro do chip 3 vezes, sempre de forma lenta e sem aspirar todo o líquido de forma a não criar bolhas. Voltou a centrifugar-se o chip durante 30 seg. Repetiu-se a agitação utilizando a pipeta e voltou-se a centrifugar durante 30 seg. Inclinou-se, novamente, o chip a 45° com a porta de carregamento para baixo, e lentamente aspirou-se o líquido através da porta de carregamento rodando o êmbolo da pipeta. Realizou-se um novo *spin* na centrífuga e removeu-se algum líquido que ainda permanecia no chip.

Cada chip tem inúmeros poços que vão ser preenchidos com uma ISP revestida com DNA amplificado de modo a gerar um sinal suficiente para a deteção da incorporação de cada base durante a sequenciação. Cada poço só pode ser preenchido por uma ISP, caso contrário será eliminado na análise bioinformática, e alguns poderão permanecer vazios.

3.4.8. Sequenciação no Ion Torrent PGM

Seguir as instruções no ecrã do PGM, selecionar a corrida planeada e introduzir o chip quando solicitado. Nesta fase, ocorreu a sequenciação por síntese, onde milhões de poços do chip continham apenas um conjugado ISP-*template* desnaturado. O chip foi sendo impregnado com os diferentes nucleótidos faseadamente. Cada vez que um nucleótido foi inserido numa cadeia em extensão, libertou-se um próton que alterou o pH do poço correspondente e essa alteração foi registada com a informação que o nucleótido foi incorporado. Quando nucleótidos não foram incorporados, não houve

alteração de pH. Assim, conseguiu obter-se a sequência de forma paralela e massiva de todo o genoma mitocondrial das amostras (figura 3).

3.4.9. Análise bioinformática dos resultados

Os resultados obtidos são transmitidos a partir do *PGM™ System* para o *Torrent Suite v.4.4.3* que traduz os resultados em ficheiros BAM (*Binary Alignment Map*) que podem ser alinhados com a rCRS e visualizados no *software Integrative Genomics Viewer v.2.3.55*.

O *plug-in Variant Caller v.4.4.23.-1* foi definido para “Generic – PGM – Somatic – Low Stringency” e com a “Minimum allele frequency” definida para valores superiores a 0,1%. Nos resultados deste *plug-in* vêm somente as diferenças em relação à rCRS em formato de tabela (anexo I). Esta tabela foi analisada no *Microsoft Office Excel* onde também se realizou a análise estatística dos dados obtidos.

Para a análise da cobertura elaboraram-se gráficos (figuras 19 e 20) recorrendo aos valores referentes à cobertura *forward* e à cobertura *reverse* em cada variante reportada para cada mistura.

Para análise da frequência das variantes foram elaborados gráficos (figuras 21 e 22) para cada mistura onde se diferenciaram as variantes esperadas (das amostras T4 e T5) dos valores dos falsos positivos. Para facilitar a visualização das amostras correspondentes a cada um dos componentes da mistura, os valores de frequência das variantes da amostra T5 e SNPs comuns entre as duas amostras de referência foram subtraídos ao valor de 100% (rCRS). Nestes gráficos incluiu-se ainda os valores referentes à cobertura alélica para cada posição reportada das variantes comuns e da amostra T5. Esta foi calculada através da subtração do valor da cobertura alélica *forward* ao valor da cobertura total *forward* e o mesmo foi feito para os valores da cobertura *reverse*. Para as variantes da amostra T4 e falsos positivos foram utilizados os valores da cobertura alélica *forward* e *reverse* originais. Desta forma, para elaboração dos gráficos, a amostra T5 converteu-se na nossa “referência”.

Para análise das frequências observadas das variantes esperadas foi elaborado um gráfico (figura 23) no qual constavam os valores de todas as 7 amostras com mistura. Às variantes esperadas que não foram registadas pelo *software*, isto é, aos falsos negativos, foram atribuídos os valores de 0%, caso fossem iguais à rCRS, ou de 100%,

caso diferissem desta. Assim, os valores apontados no gráfico apresentam-se como a proporção da amostra T4 em cada uma das misturas reportadas como diferenças em relação à amostra T5.

Para apurar a precisão da deteção das frequências esperadas elaborou-se um gráfico (figura 24) que comparava os valores de frequência observados com os valores esperados das variantes em cada mistura. Da média dos valores das 7 amostras resulta uma linha de regressão que nos dará o valor da correlação entre as duas variáveis. O cálculo do desvio padrão e a sua representação no gráfico permite-nos visualizar a desproporção entre as duas variáveis em cada mistura.

4. RESULTADOS

4.1. Haplótipos obtidos por sequenciação de Sanger

Os haplótipos obtidos e os haplogrupos definidos com a ajuda da ferramenta *online HaploGrep* podem ser visualizados na tabela 6.

Tabela 6 - Haplogrupos das amostras T3, T4 e T5 obtidos com recurso à ferramenta HaploGrep e a precisão dessa atribuição (qualidade), tendo por base os haplótipos completos obtidos pela sequenciação das duas amostras.

Amostra	Haplogrupo	Qualidade (%)	Haplótipo
T3	U4a1a1	99,6	73G, 152C, 195C, 263G, 309.1C, 315.1C, 499A, 524.1A, 524.2C, 750G, 961C, 965.1C, 1438G, 1811G, 2706G, 4646C, 4769G, 5999C, 6047G, 7028T, 8818T, 8860G, 11332T, 11467G, 11719A, 11722C, 12308G, 12372A, 12937G, 14620T, 14766T, 15326G, 15693C, 16134T, 16356C, 16519C
T4	H13a1a	95,3	263G, 315.1C, 750G, 1438G, 2259T, 4745G, 4769G, 8860G, 12061T, 13680T, 14872T, 15326G
T5	L1b1a12a	99,6	73G, 152C, 182T, 185T, 195C, 198T, 247A, 263G, 315.1C, 357G, 523d, 524d, 709A, 710C, 750G, 769A, 825A, 1018A, 1462A, 1738C, 2352C, 2706G, 2758A, 2768G, 2885C, 3308C, 3594T, 3666A, 3693A, 4104G, 4769G, 5036G, 5046A, 5393C, 5655C, 6548T, 6827C, 6989G, 7028T, 7055G, 7146G, 7256T, 7389C, 7521A, 7867T, 8248G, 8468T, 8655T, 8701G, 8860G, 9540C, 10398G, 10688A, 10810C, 10873C, 11002G, 11719A, 12519C, 12705T, 13105G, 13506T, 13650T, 13789C, 13880A, 14178C, 14203G, 14560A, 14766T, 14769G, 15115C, 15326G, 16126C, 16187T, 16189C, 16223T, 16264T, 16270T, 16278T, 16293G, 16311C, 16400T, 16519C

4.2. Amplificação por Long-Range PCR

A estratégia de Long-Range PCR foi utilizada para amplificar o genoma mitocondrial num só fragmento, não sobreponível, através do uso de um par de *primers*. Os produtos amplificados foram sujeitos a eletroforese num gel de agarose 0,8%. Nas duas amostras foram encontrados fragmentos do tamanho desejado (cerca de 16,5 kbs) (figura 15).

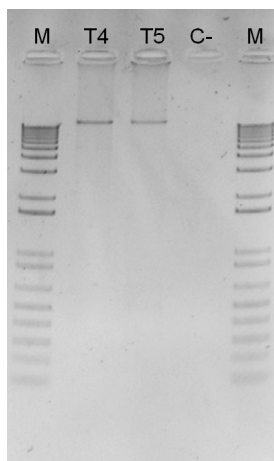


Figura 15 - Verificação da amplificação por Long-Range PCR das amostras T4 e T5 em gel de agarose a 0,8%.
Legenda: M – Marcadores de peso molecular; T4 – Amostra T4; T5 – Amostra T5; C- – Controlo negativo.

4.2.1. Cobertura das amostras de referência

A cobertura das amostras de referência amplificadas por Long-Range PCR, quer num só fragmento ou em dois amplicões sobreponíveis, pode ser visualizada na figura 16. Os resultados das sequenciações efetuadas pelo Ion Torrent PGM analisados pelo *software Torrent Suite v.4.4.3* e convertidos em ficheiros BAM que foram visualizados no *software Integrative Genomics Viewer*. Verificou-se que as coberturas são muito semelhantes com ambas as estratégias.

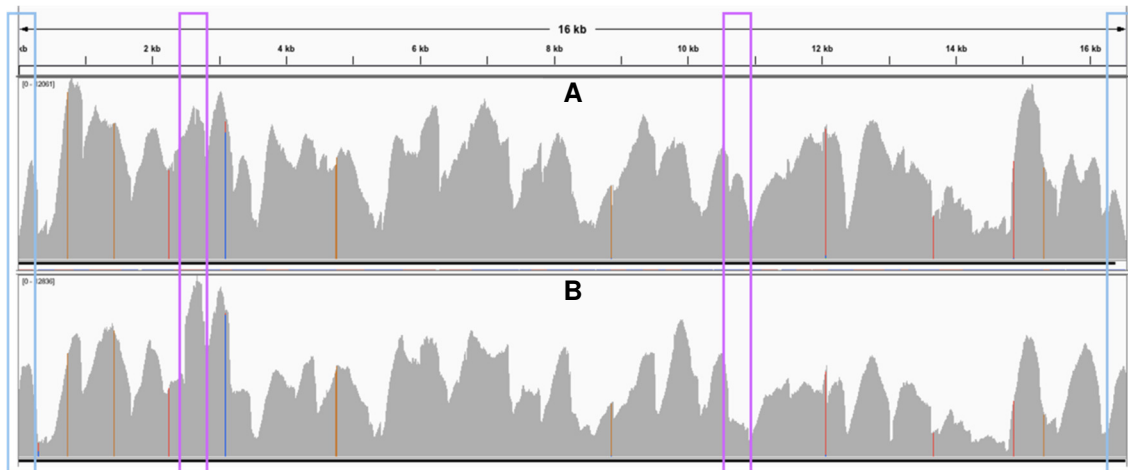


Figura 16 - Cobertura da amostra T4 amplificada em um só fragmento de aproximadamente 16,5 kbs (A) e em dois fragmentos sobreponíveis de aproximadamente 8,5 kbs (B). O eixo das abcissas representa a posição dos nucleótidos ao longo da molécula de mtDNA e o eixo das ordenadas a cobertura. Destacado a azul a zona de ligação dos primers da estratégia de amplificação em um só fragmento e a roxo as zonas onde se sobrepõe os dois fragmentos da estratégia de Fendt e colaboradores.

4.3. Sequenciação no Ion Torrent PGM

4.3.1. Sequenciação das amostras de referência

As variantes obtidas pelas duas metodologias foram na sua maioria concordantes. Contudo, na sequenciação pela plataforma Ion Torrent PGM, surgiram algumas diferenças, relacionadas essencialmente com diferenças de nomenclatura:

- Na amostra T4 a inserção de uma citosina na posição 315 (Sanger) foi assinalada na posição 311 (MPS);
- Na amostra T5 as deleções nas posições 523 e 524 de uma citosina e de uma adenina (Sanger) foram assinaladas na posição 514 (MPS);
- Na amostra T5 as substituições nas posições 709 e 710 (Sanger) foram reportadas como uma dupla substituição de uma guanina e uma timina para uma adenina e uma citosina na posição 709 (MPS).

4.3.2. Relatório da corrida das amostras com heteroplasmas artificiais

O relatório final elaborado pelo *software Torrent Suite v.4.4.3* permite analisar a qualidade da corrida.

O relatório pré-alinhamento indica a densidade de carregamento das ISPs, um resumo do estado em que estas se encontravam e o comprimento das *reads* geradas. Na figura 17A podemos observar uma imagem ilustrativa da densidade de ISPs (77%). A densidade de ISP é medida pela alteração de pH dos poços, que depende da presença de ISPs com DNA. O número total de bases registadas no final foi 370 milhões. Na figura 17B podemos observar um gráfico cuja percentagem das colunas inferiores dizem respeito ao valor a azul da coluna imediatamente acima. A primeira coluna do gráfico indica-nos densidade de poços ocupados/vazios. 100% das ISPs desses poços são ISPs que continham DNA da biblioteca, não havendo poços sem DNA *template*. Apesar de se pretender uma amplificação monoclonal do DNA em cada ISP, 50% destas continham dois ou mais *templates*, ou seja, eram policlonais. O *software* faz esta distinção com base em valores invulgares de pH detetados nos poços das ISPs policlonais. Da percentagem de ISPs monoclonais, 83% das *reads* resultantes ultrapassaram os filtros para detetar fragmentos policlonais, de fraca qualidade ou dímeros de *primers* e foram utilizadas para o alinhamento. Em comparação com o número de ISPs que continham DNA da biblioteca (4854154) as *reads* utilizáveis correspondem a 42% (2026375). O número total de *reads* é o primeiro indicador para se descobrir se um número de dados obtido será razoável. Uma variação brusca entre os números de *reads* reportados nos ficheiros de cada amostra pode indicar um problema com a qualidade da biblioteca. A grande maioria das *reads* geradas tinha aproximadamente 250 bp de comprimento (figura 17C).

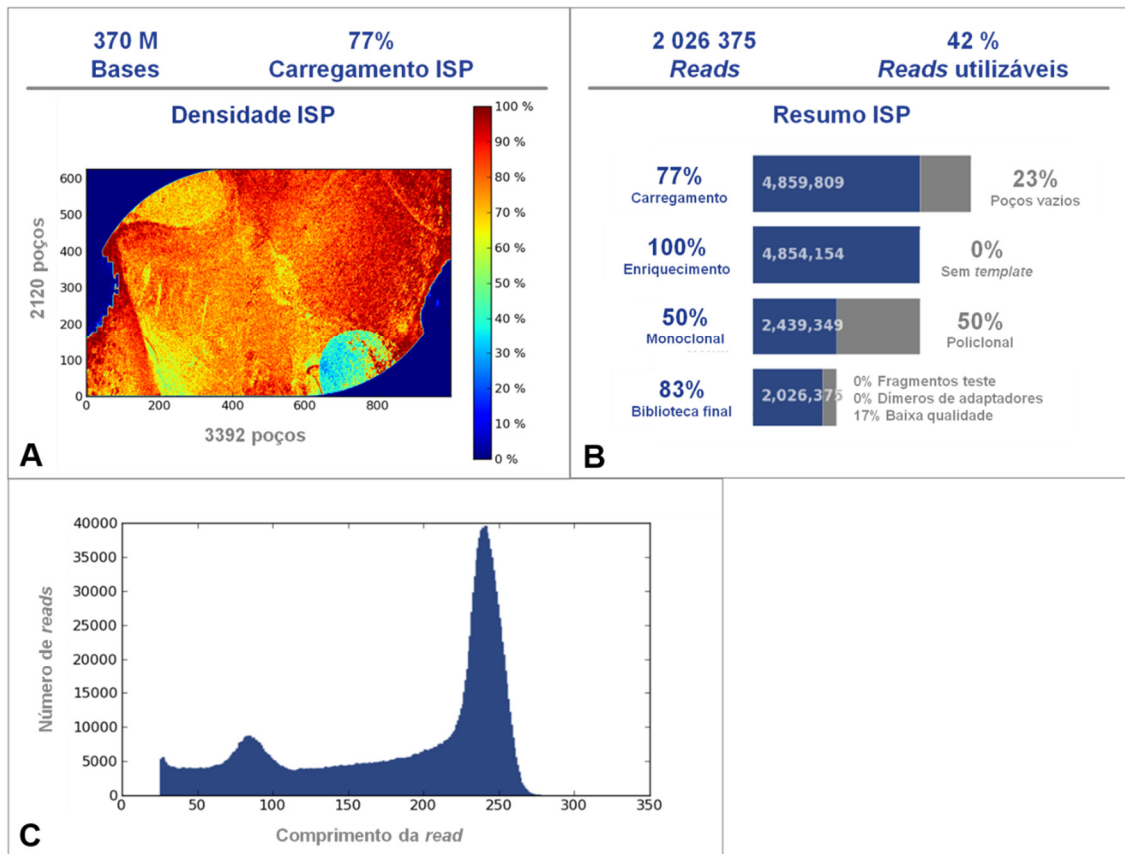


Figura 17 - Esquemas do relatório pré-alinhamento: **A)** Imagem ilustrativa do carregamento dos poços do chip ao longo da sua superfície; **B)** Resumo do estado das ISPs; **C)** Histograma do comprimento das *reads*.

O relatório de alinhamento com a rCRS mostra-nos um sumário do alinhamento das *reads* e da sua qualidade. Na figura 18A, podemos ver o número total de bases alinhadas (332 milhões) assim como número total de bases alinhadas em relação ao número de bases da sequência de referência (rCRS). Isto é, a cobertura da referência foi efetuada 20045,4 vezes pelas bases nas *reads*. Do número total de *reads* geradas durante o processo de *basecalling* (atribuição de bases) 92% encontram-se alinhadas com a sequência de referência. Na figura 18B pode observar-se a percentagem de precisão do alinhamento para cada posição das *reads* (98,7%). O alinhamento é útil para verificar a qualidade da sequenciação e da biblioteca subjacente. Os parâmetros do alinhamento (tabela 7) são apontados de acordo com a tolerância do *software* a bases erradas. Neste caso, o *Torrent Suite* permite taxas de erro de 1% (AQ20) e 2% (AQ17) (Life Technologies, 2014).

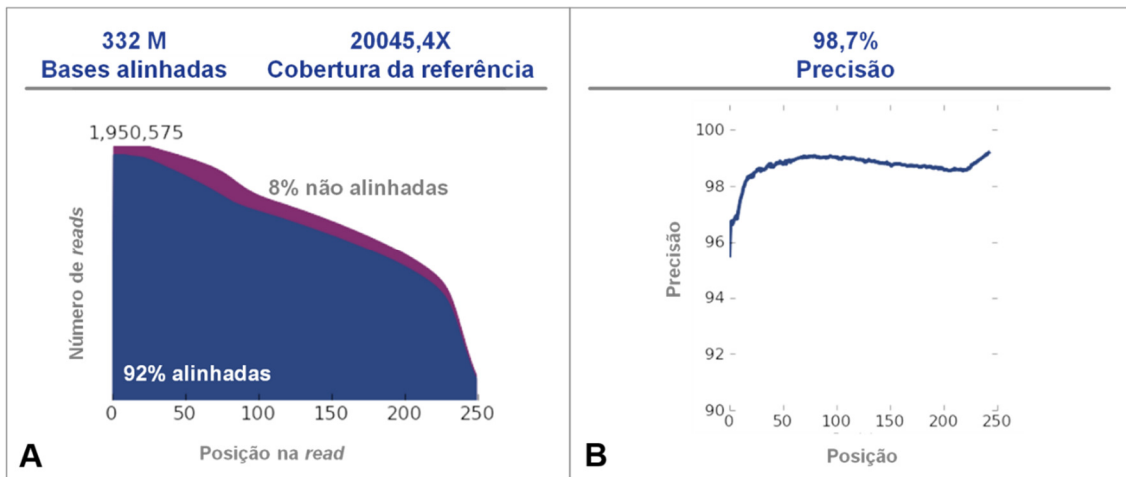


Figura 18 - Relatório de alinhamento: A) Gráfico do alinhamento das reads; B) Precisão do alinhamento das reads.

Tabela 7 - Características das reads segundo os diferentes níveis de qualidade de alinhamento.

	AQ17	AQ20	Perfeito
Número total de bases (Mbp)	299 M	262 M	174 M
Comprimento médio (bp)	178	165	117
Alinhamento mais longo (bp)	296	296	286
Intensidade média de cobertura (x)	18084.0	15835.5	10538.6

Definindo o nosso nível de qualidade para 1% podemos dizer que o número total de pares de bases foi 262 milhões. O comprimento médio das reads foi 165 bp, com o alinhamento mais longo registado chegando a 296 bp (tabela 7).

4.3.3. Cobertura das amostras com heteroplasmias artificiais

Os resultados foram analisados com o *plug-in Variant Caller* para “Generic – PGM – Somatic – Low Stringency” e o parâmetro “Minimum allele frequency” foi definido para apresentar apenas as variantes para percentagens superiores a 0,1%. Os valores recomendados para detetar SNPs encontram-se entre 0,01 (1%) e 0,2 (20%) e para a

identificação de indels são recomendados valores entre 0,05 (5%) e 0,2 (20%) (Life Technologies, 2014).

A cobertura variou de intensidade ao longo do genoma em todas as amostras apesar das variações serem semelhantes nas 7 amostras e às encontradas previamente nas amostras de referência (secção 4.2.1). Nas figuras 19 e 20 podemos observar a intensidade de cobertura nas posições identificadas como variantes pelo *plug-in Variant Caller*. Algumas posições apresentaram-se consistentemente com cobertura reduzida (inferior a 1000 vezes) em todas as amostras:

- da posição 310 à posição 525;
- posição 3594;
- posição 5393;
- posições 8468 e 8655;
- posições 10873 e 11002;
- da posição 13650 à posição 14770;
- da posição 16187 à posição 16223.

Importa referir, que na observação dos gráficos das figuras 19 e 20 poderão ser encontrados falsos negativos (posições em branco). Nomeadamente, nas zonas registadas entre a posição 310 e a 525 e entre a posição 13650 e a 14770, inclusivamente. De um modo geral, estas zonas encontram-se com cobertura reduzida em todas as amostras. Algumas variantes esperadas podem não aparecer em todas as amostras, principalmente naquelas cujas proporções são mais desiguais. Isto acontece porque o *plug-in* apenas regista a cobertura das posições que interpreta como variantes (relativamente à rCRS). Quando a variante atinge uma proporção tal que deixa de ser detetada pelo *plug-in*, deixamos de ter informação sobre a cobertura dessa posição.

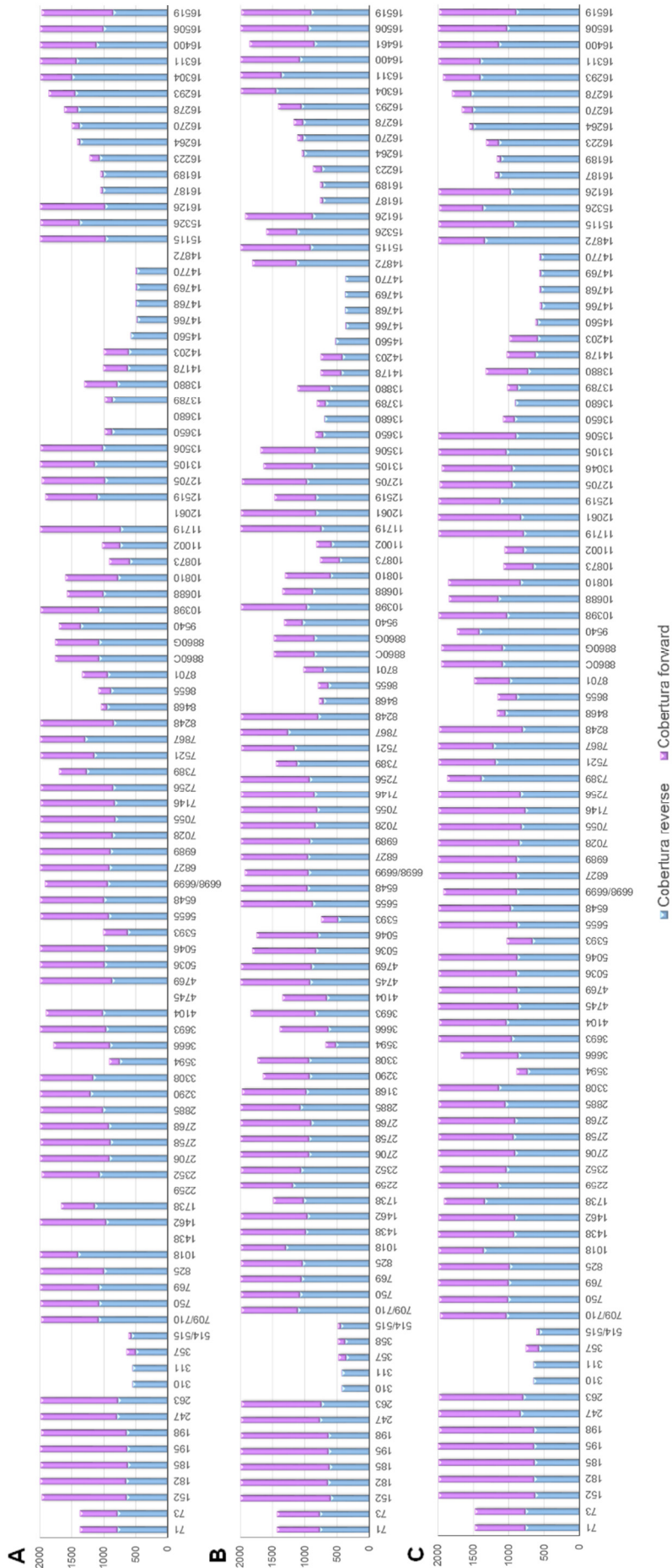


Figura 19 - Gráficos da cobertura [em número de reads (eixo das ordenadas)] das posições assinaladas como variantes (eixo das abscissas) nas amostras com mistura: **A)** 1(T4):99(T5); **B)** 5(T4):95(T5); **C)** 10(T4):90(T5).

Sequenciação completa do mtDNA por MPS e avaliação da deteção de heteroplasmias



Figura 20 - Gráficos da cobertura [em número de reads (eixo das ordenadas)] das posições assinaladas como variantes (eixo das abscissas) nas amostras com mistura: **A** 50(T4):50(T5); **B** 90(T4):10(T5); **C** 95(T4):5(T5); **D** 99(T4):1(T5).

4.4. Deteção e quantificação de variantes

Os haplótipos das duas amostras de referência diferem em 82 posições, sendo outras 6 variantes comuns entre as duas mas distintas da rCRS.

Nas figuras 21 e 22, apresentam-se os gráficos individuais de cada uma das 7 amostras correspondentes às diferentes misturas, com o registo das frequências das variantes das duas amostras de referência, da cobertura alélica (forward e reverse) e dos falsos positivos e falsos negativos (em branco).

A posição 315.1C, comum às duas amostras de referência, apresentou sempre valores de frequência mais elevados do que as restantes posições comuns, que se mantiveram sempre próximas de 0% (figuras 21 e 22).

Foram detetados falsos positivos em todas as 7 amostras, numa média de 7 a 12 posições em cada uma, sempre a níveis inferiores a 5%. O número de falsos positivos surge de modo aleatório nas amostras de diferentes proporções, no entanto, a maioria das posições onde se encontram é detetada de forma recorrente nas diferentes misturas. As posições que surgem mais frequentemente como falsos positivos são: 71, 310, 520, 3290, 6698/6699, 8249, 8860C, 14768, 14770 e 16506.

Na figura 23 podemos observar as frequências das 82 variantes representadas como diferenças à amostra T5.

Na mistura 1(T4):99(T5) nenhuma das variantes esperadas para a amostra T4 foi registada. Em todas as outras proporções, as mesmas variantes foram sempre detetadas. As variantes da amostra T5 foram quase sempre identificadas com exceção da mistura com a proporção 99(T4):1(T5), onde somente a posição 16159 foi identificada. Nas restantes amostras verificou-se a ausência das posições:

- 523/524, 16187 e 16189 na mistura com a proporção 90(T4):10(T5);
- 523/524, 10873, 16189 e da posição 16264 à 16400 na mistura com a proporção 95(T4):5(T5).

Sequenciação completa do mtDNA por MPS e avaliação da detecção de heteroplasmias

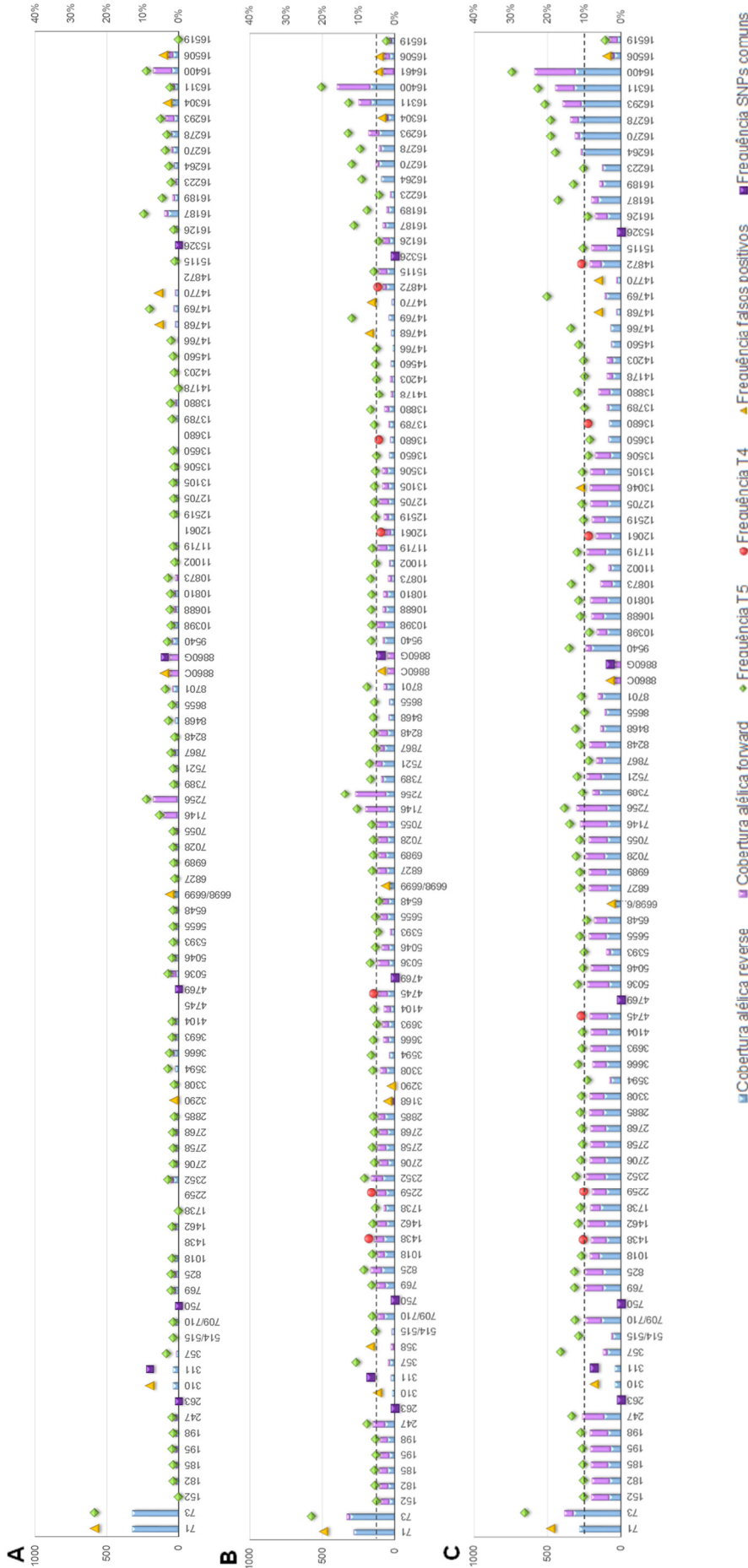


Figura 21 - Gráficos da cobertura alélica e frequência [em número de reads e percentagem (eixo das ordenadas esquerdo e direito, respetivamente)] das posições assinaladas como variantes (eixo das abcissas) nas amostras com mistura: **A**) 1 (T4):99(T5); **B**) 5(T4):95(T5); **C**) 10(T4):90(T5).

Sequenciação completa do mtDNA por MPS e avaliação da detecção de heteroplasmias



Figura 22 - Gráficos da cobertura alélica e frequência [em número de reads e percentagem (eixo das ordenadas esquerdo e direito, respetivamente)] das posições assinaladas como variantes (eixo das abcissas) nas amostras com mistura: **A)** 50(T4):50(T5); **B)** 90(T4):10(T5); **C)** 95(T4):5(T5); **D)** 99(T4):1(T5).

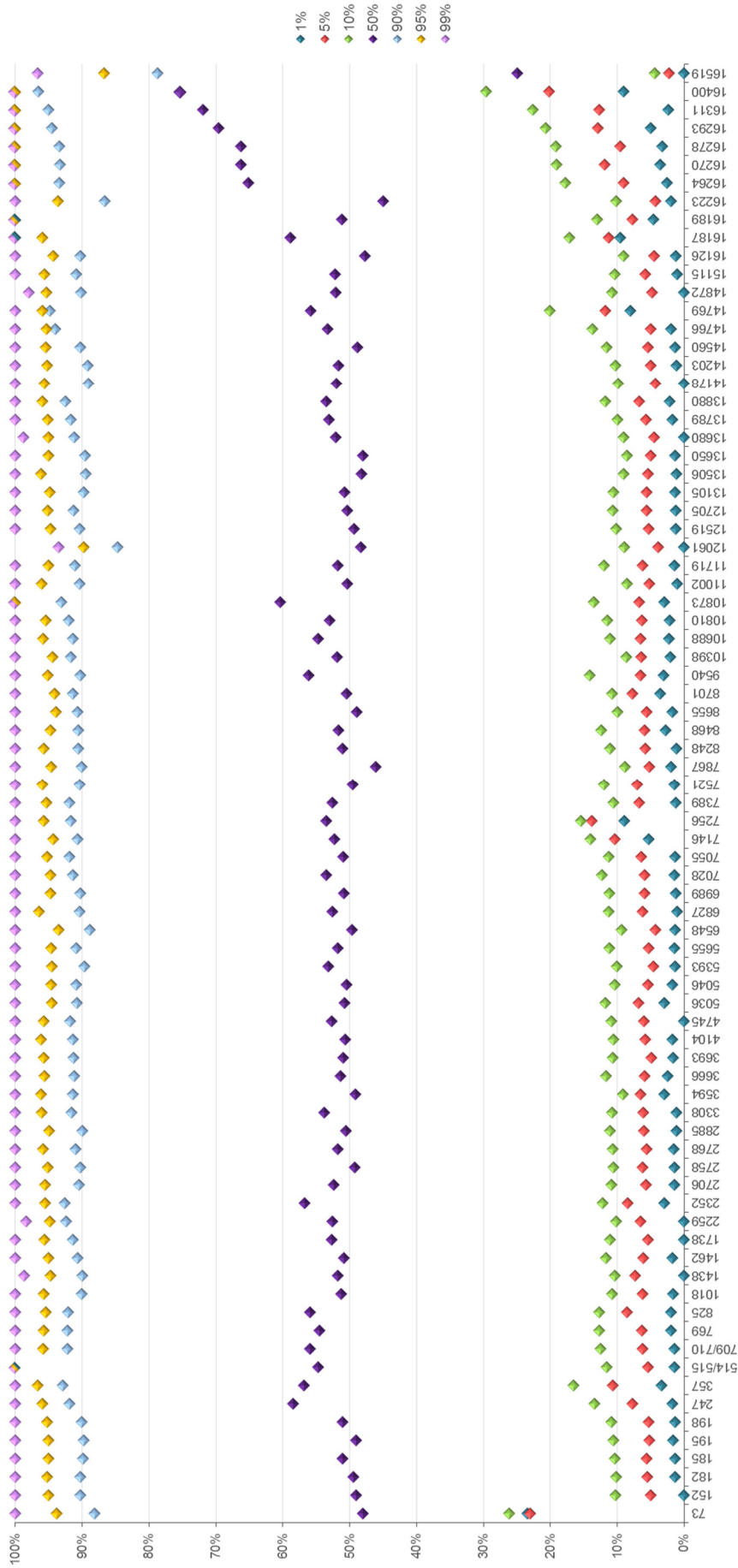


Figura 23 - Frequência da heteroplasmias [em percentagem (eixo das ordenadas)] nas posições onde se esperavam serem detetadas variantes (eixo das abcissas) nas 7 amostras com mistura.

4.5. Precisão da quantificação das heteroplasmas

Na figura 24 podemos examinar o gráfico correspondente à correlação entre os níveis de heteroplasma esperados e observados nas 7 amostras, onde se pode confirmar que a quantificação das heteroplasmas é bastante correta e precisa.

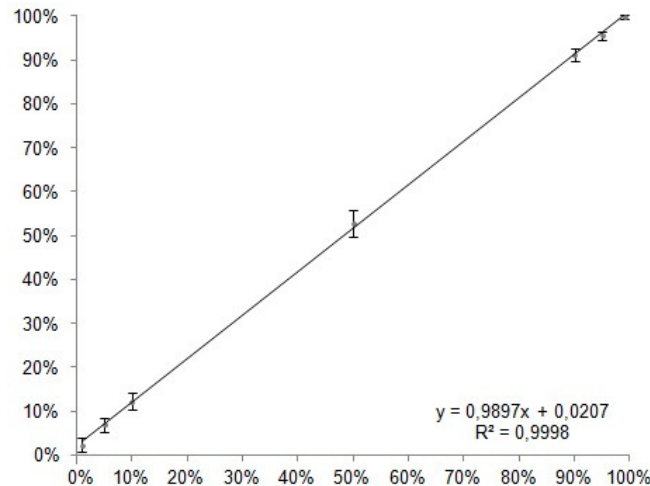


Figura 24 - Correlação entre os níveis de heteroplasma esperados (eixo das abcissas) e observados (eixo das ordenadas). O círculo e a barra de erro representam, respetivamente, a média e o desvio padrão das 82 posições de cada uma das 7 amostras.

4.6. Análise das misturas pelo método de Sanger

Na figura 25 podemos observar algumas das variantes detetadas por sequenciação de Sanger, numa pequena zona do fragmento MTT1, das 7 amostras em diferentes proporções. Verifica-se que a mistura só é detetada com clareza, ou seja, claramente distinta do ruído de fundo, a partir das proporções de 25-75% apesar do limite de deteção do *software* ter sido definido para 5%.

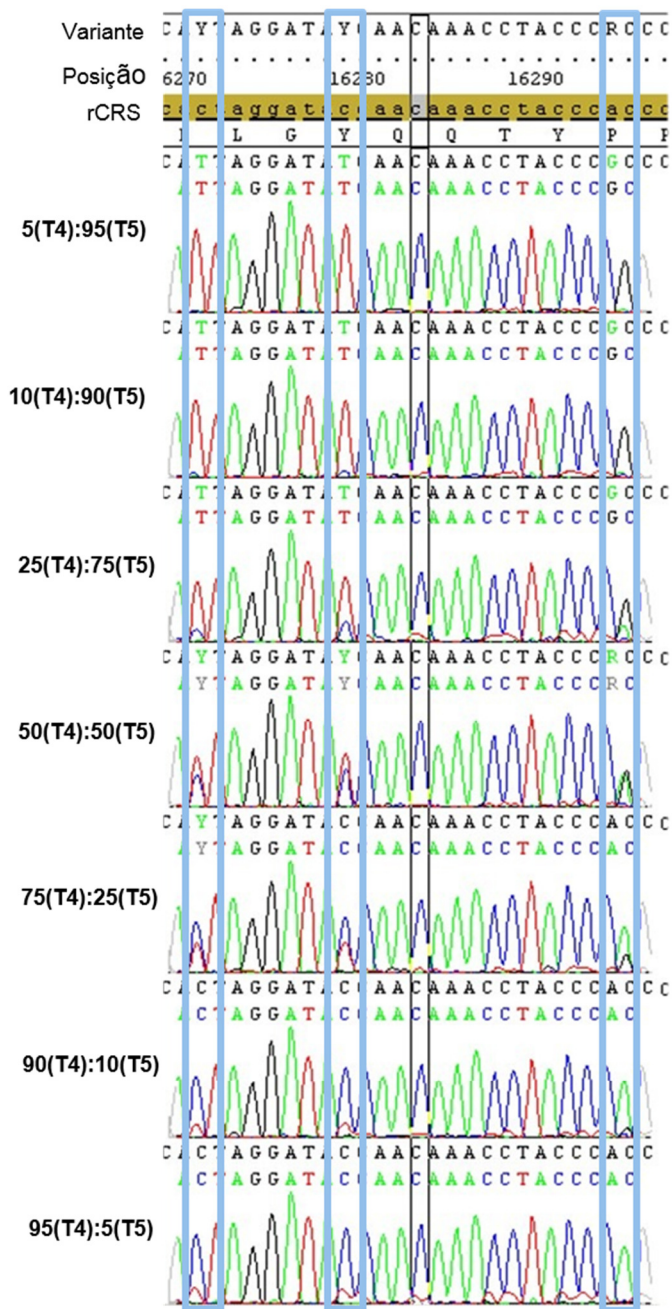


Figura 25 - Visualização de algumas heteroplasmias (destacadas a azul) das 7 amostras com mistura analisadas pelo método de Sanger no *software SeqScape*.

5. DISCUSSÃO

5.1. Amplificação por Long-Range PCR

A escolha do Long-Range PCR para a amplificação baseou-se em estudos já elaborados por parecer a mais apropriada para o objetivo de detetar/quantificar heteroplasmias por MPS.

A estratégia de se utilizarem vários pares de *primers* para a amplificação do mtDNA (tal como a que utilizamos para a amplificação da molécula de DNA mitocondrial para a sequenciação de Sanger) é eficaz em amplificar toda a molécula. Contudo, e apesar de poder demonstrar uma cobertura uniforme em cada fragmento após MPS, entre os diferentes fragmentos a intensidade de cobertura varia de forma brusca e nas zonas de sobreposição desses fragmentos atinge valores extremos de cobertura (Zhang *et al.*, 2012).

A estratégia de amplificação por 2 pares de *primers* cria dois fragmentos sobreponíveis, com cerca de 8,5 kbs de comprimento cada um (Fendt *et al.*, 2009) que também amplificam toda a molécula de mtDNA. Na figura 16, podemos comparar a cobertura obtida por esta estratégia e a alcançada na amplificação num só fragmento, após o LR-PCR da amostra T4. Em ambas apenas se observou um padrão ligeiramente assimétrico da cobertura, típico na sequenciação do mtDNA com o sistema Ion Torrent PGM, mas sempre com uma intensidade de cobertura satisfatória em todas as posições da molécula de mtDNA. Embora a intensidade cobertura pudesse ser suficiente para se quantificar devidamente o nível de heteroplasmias, outras complicações podem surgir na estratégia da amplificação em dois fragmentos. A presença de variantes na zona de ligação dos *primers* ou outras pequenas desigualdades nas condições de reação, podem resultar numa eficiência de PCR desigual entre os dois fragmentos e, por isso, as variantes registadas e a quantificação do nível de heteroplasmias devem ser interpretadas com cautela.

A estratégia de amplificação em um só fragmento apresenta ainda duas vantagens adicionais, uma vez que evita a co-amplificação de pseudogenes de mtDNA dispersos no DNA nuclear (Cui *et al.*, 2013) e permite identificar a presença de grandes deleções (Cui *et al.*, 2013) que, por vezes, aparecem associadas a doenças mitocondriais (Chinnery & Hudson, 2013).

A estratégia de amplificação do mtDNA num fragmento demonstrou-se eficaz. No gel de verificação da amplificação detetaram-se fragmentos com cerca de 16,5 kbs (figura 15). Caso existissem grandes deleções seriam visualizadas como bandas adicionais, de menor tamanho, no gel tal como se pode observar no exemplo da figura 26.

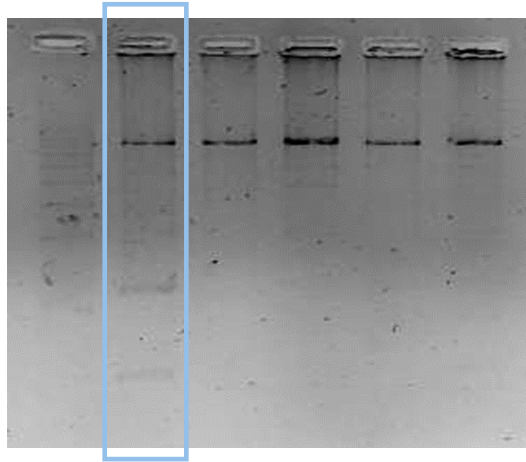


Figura 26 – Imagem de um gel com produtos da amplificação por LR-PCR num só fragmento onde se pode visualizar uma amostra (destacada a azul) com bandas adicionais (A. Goios, comunicação pessoal).

A zona de ligação dos *primers* foi escolhida por ser uma zona muito conservada. No entanto, se houver necessidade de sequenciar a região de alinhamento dos primers ou se por amplificação ineficiente se desconfiar de alterações na zona de ligação dos mesmos deverá recorrer-se a um novo PCR com *primers* diferentes.

5.2 Haplótipos e Haplogrupos das amostras de referência

Os resultados da sequenciação completa do genoma mitocondrial das amostras de referência por MPS foram confrontados com os obtidos por sequenciação de Sanger. De um modo geral, observou-se um elevado nível de concordância das informações obtidas entre ambas as metodologias.

As discrepâncias observadas correspondem essencialmente a posições onde ocorrem inserções ou deleções, ou substituições em posições consecutivas, por exemplo:

- Na amostra T4, verificou-se que a inserção identificada na posição 311 pela plataforma Ion Torrent PGM correspondia na anotação por Sanger à inserção da posição 315. Esta observação pode ser explicada por estas duas posições se inserirem numa região poli-C no interior da região HVR2, e uma das metodologias reporta a inserção no início do homopolímero enquanto a outra reporta na posição final. Uma elevada taxa de erro na sequenciação de indels nas regiões HVR, por algumas plataformas de MPS foi, também, anteriormente

apontada devido, provavelmente, a falhas no alinhamento de sequências nas regiões homopoliméricas (Parson *et al.*, 2013; Luo *et al.*, 2011);

- Nas posições 523 e 524, detetadas por sequenciação de Sanger na amostra T5. Tal deve-se a estas se encontrarem numa região de repetição dinucleotídica AC, tratando-se, tal como no caso anterior, de uma diferença de alinhamento/nomenclatura. Assim, as deleções foram reportadas em conjunto na posição 514;
- Na posição 709 da amostra T5, registada na sequenciação por Ion Torrent PGM, notamos que a substituição do segundo nucleótido (timina) é o mesmo reportado pela sequenciação de Sanger na posição 710. Assim, e visto que as duas posições são adjacentes, podemos concluir que os dois polimorfismos foram reportados em conjunto na análise pelo Ion Torrent PGM.

Após a obtenção de consenso no registo dos haplótipos das amostras (ver tabela 6) avaliamos a precisão dessas sequências no contexto filogenético. Para isso, recorreu-se à ferramenta *online HaploGrep*, que atribui uma lista de classificação dos haplogrupos a que a amostra possa pertencer. As classificações que obtenham resultados superiores a 90% (van Oven & Kayser, 2009) têm uma maior confiança na precisão da atribuição. Quando a cotação é inferior a esse valor, pode denotar, por exemplo, a falta de alguns polimorfismos típicos do haplogrupo e assim não corresponder na totalidade a essa atribuição ou indiciar erros de sequenciação. Na análise das nossas amostras este programa atribuiu a amostra T4 ao haplogrupo H13a1a (com uma qualidade de precisão de 95,27%) e a amostra T5 ao haplogrupo L1b1a12a (com uma precisão de 99,6%). Em ambas as amostras foram registados todos os polimorfismos esperados para esses haplogrupos e, cada uma, apenas exibiu uma mutação não associada ao haplogrupo declarado.

Os haplótipos das duas amostras diferem entre si em 82 polimorfismos no genoma mitocondrial o que torna as amostras ideais para efetuar misturas para testes de detecção de heteroplasmia.

5.3. Cobertura

A intensidade média de cobertura obtida na nossa corrida foi de 15835,5 vezes (com uma taxa de erro de 1%) (tabela 7). Para o mtDNA, é recomendada uma intensidade

média de cobertura de 20000 vezes para se detetarem, de modo fidedigno, heteroplasmias a uma proporção de 1,33% (Zhang *et al.*, 2012).

Embora o padrão ideal fosse uma cobertura equilibrada ao longo de todas as posições, a cobertura obtida não foi uniforme, tendo variado ao longo do genoma em todas as amostras. Todavia, o mapa de cobertura foi bastante semelhante nas 7 amostras estudadas. Dado a estratégia de Long-Range PCR gerar um só amplicão de toda a molécula de mtDNA, a variação na cobertura não pode ser explicada pelo PCR. As diferenças de cobertura podem dever-se, muito provavelmente, à preparação da biblioteca e/ou ao processo de sequenciação. Quail e colaboradores apontaram que a enzima utilizada no processo de fragmentação da biblioteca pode influenciar na cobertura (Quail *et al.*, 2012), sendo também possível que estas diferenças tenham origem no passo de amplificação da biblioteca já fragmentada.

Esta observação pode também ser explicada pela presença de extensão de citosinas em algumas zonas da molécula de mtDNA, particularmente nas junto às posições 310 e 16187. Foi anteriormente registada uma elevada taxa de erro, por algumas plataformas de MPS, devido, provavelmente, a falhas no alinhamento de sequências nas regiões homopoliméricas (Parson *et al.*, 2013; Luo *et al.*, 2011).

Outra explicação para as regiões com menor cobertura poderá ser a ocorrência de discrepâncias entre a sequenciação das cadeias *forward* e *reverse* - *strand bias*. Em teoria, as duas cadeias, *forward* e *reverse*, deveriam ser sequenciadas o mesmo número de vezes, no entanto há regiões onde apenas cadeias de um dos sentidos são sequenciadas, como podemos ver em algumas posições apresentadas nas figuras 19 e 20. Apesar de o *strand bias* não significar necessariamente fraca qualidade na detecção de variantes, uma representação equilibrada dará uma maior confiança que a variante apresentada foi a correta. Quando uma cadeia está sobre-representada poderá surgir uma variante incorreta, em particular quando se trata da detecção de inserções/deleções. Algumas das posições supracitadas, devido à reduzida intensidade da cobertura, estão representadas de forma exclusiva pela cadeia *reverse* (da posição 311 à 316 e da posição 14560 à 14770). Porém, podemos observar outras zonas com menor cobertura, que não apresentam tanta disparidade entre a representação da cadeia *forward* e *reverse*.

5.4. Sensibilidade, especificidade e precisão da quantificação de heteroplasmias

A correta identificação e quantificação do nível de heteroplasmia é essencial para um diagnóstico clínico preciso e para o aumento do poder de discriminação em genética forense. Assim, torna-se importante o estudo da capacidade das novas tecnologias nesta área.

As variantes esperadas foram detetadas em todas as misturas com exceção das variantes da amostra T4 na mistura com a proporção 1(T4):99(T5) e de quase todas as variantes da T5 na mistura com a proporção 99(T4):1(T5).

As variantes relacionadas com inserções ou deleções nas regiões homopoliméricas ou de repetições dinucleotídicas (por exemplo a 315.1C e 523/524del) apresentam as dificuldades de sequenciação por MPS já discutidas nas secções anteriores (Parson *et al.*, 2013; Luo *et al.*, 2011). Pelos mesmos motivos, as frequências apresentadas para estas posições são também as que mais diferem dos valores esperados nas misturas realizadas (figuras 21 e 22).

Na figura 23 podemos observar que a grande maioria dos valores de frequência das variantes observadas nas 7 amostras se encontram nos níveis esperados. Nas variantes localizadas no segmento HVR1 observa-se um incremento acentuado dos valores de frequência devido à região homopolimérica presente.

Ao comparar-se as amostras com proporções opostas 1(T4):99(T5) e 99(T4):1(T5) (também para as posições entre a 16264 e 16400 na amostra 95(T4):5(T5)), observou-se que não existiam variantes detetadas entre as percentagens de 0% a 1% nem de 99% a 100%, respetivamente, apesar de o limite de deteção ter sido definido para 0,1%. Como os valores observados nas outras amostras foram, algumas vezes, ligeiramente inferiores ou superiores aos esperados, torna-se curioso que o mesmo não acontecesse nestas. Não é claro o que poderá estar a causar esta observação, mas poderá estar relacionado com alguma definição avançada do *plug-in* utilizado para a determinação das variantes poder estar a filtrar as variantes presentes a menos de 1% como mecanismo para evitar falsos positivos. No entanto, esta possibilidade não foi confirmada uma vez que nos parâmetros das definições do *plug-in* que são editáveis através do *Torrent Browser* não se encontrou nenhum que pudesse ser a justificação.

A maioria dos falsos positivos foram detetados junto a regiões homopoliméricas ou em posições adjacentes a variantes reais, o que pode sugerir erros de alinhamento, como por exemplo as posições: 310, 358, 525, 6698/6699, 8249, 9531/9532, 14768 e 14770.

Estas posições surgem sempre a níveis inferiores a 5% o que na maior parte dos casos permite diferenciá-las das verdadeiras variantes. Contudo, quando se avaliam heteroplasmias a níveis muito reduzidos os falsos positivos podem ser um problema e devem ser tidos em conta, e avaliados conforme a sua localização, isto é, caso se encontrem nestas zonas esperadas será mais fácil determinar que são falsas variantes.

Assim, podemos concluir que a sequenciação por Ion Torrent PGM apresenta, no geral, uma boa sensibilidade e especificidade na quantificação de substituições em heteroplasmia a níveis superiores a 5%. O mesmo não acontece com indels em zonas homopoliméricas ou de repetição dinucleotídica onde, por vezes, não foram detetadas as variantes ou foram-no com coberturas ou frequências baixas.

De modo a verificar o grau de precisão da nossa análise, comparamos as frequências das 82 substituições observadas com valores esperados, nas diferentes 7 amostras (figura 24). A correlação entre as percentagens observadas e esperadas foi quase excelente ($R^2=0,9998$; $y = 0.9897x + 0.0207$). Isto revela uma precisão notável na detecção dos níveis de heteroplasmia.

Estudos para outra plataforma de MPS – Illumina GA – observaram os mesmos valores de sensibilidade ($R^2= 0,9997$) (Tang & Huang, 2010) e atingiram mesmo 1,33% como limite de detecção com um grau de confiança de 99,9% ($R^2=1$) (Zhang *et al.*, 2012). Tang e Huang, conseguiram sempre detetar todas as substituições esperadas nas amostras com níveis superiores a 5% ao contrário das nossas observações. Conseguiram ainda identificar, até este nível, as deleções nas posições 523 e 524 a uma frequência bastante próxima da esperada o que não aconteceu no nosso estudo com o Ion Torrent PGM . Estas duas observações parecem indicar uma maior sensibilidade da plataforma Illumina GA em comparação com o Ion Torrent PGM. Além disto, até ao nível de 5%, a análise desempenhada pelos mesmos autores, não detetou quaisquer falsos positivos, o que demonstra uma maior especificidade do Illumina em relação ao Ion Torrent PGM que registou sempre entre 7 a 12 falsos positivos (embora em baixa frequência). A plataforma Illumina detetou ainda a grande maioria das substituições esperadas em amostras com níveis de heteroplasmia de 1%, o que não aconteceu com o Ion Torrent PGM, no entanto o elevado número de falsos positivos obtidos resultou numa baixa especificidade.

5.5. Comparação da sensibilidade com o método de Sanger

De modo a fazer uma pequena comparação da MPS com a sequenciação de Sanger, efetuamos a mistura de um pequeno fragmento amplificado das duas amostras de referência em 7 misturas de proporções, adaptadas às percentagens apontadas como limite de detecção de níveis de heteroplasmia através desta técnica.

Na análise pelo método de Sanger, podemos notar que o *software SeqScape* apenas anotou a ocorrência de heteroplasmias nas proporções 50(T4):50(T5) e 75(T4):25(T5) nas três variantes visíveis (figura 25). Também com o *software Geneious* as substituições que constavam na zona analisada foram todas detetadas até aos 25%, surgindo somente cerca de metade destas a 10%. No entanto, nesta amostra (10%) mais de 50 falsos positivos foram registados num fragmento de apenas 1822 bp. Apesar de não podermos definir um limite de detecção podemos afirmar que, este método de sequenciação, apresenta um grau de especificidade muito baixo para identificar com sucesso heteroplasmias a níveis muito reduzidos.

Estes resultados vêm de encontro ao que já tem sido descrito para a sequenciação de Sanger. Uma vez que a detecção de heteroplasmias com esta tecnologia baseia-se apenas na altura dos picos do eletroferograma, o *software* de análise pode não conseguir distinguir as verdadeiras variantes do “ruído de fundo” a níveis inferiores a 15-20% gerando o aparecimento de falsos positivos (Rohlin *et al.*, 2009; Chinault *et al.*, 2009; Wong *et al.*, 2008).

5.6. Considerações finais

Neste estudo, apresentamos um potencial método para análises do mtDNA em áreas distintas, como a identificação humana e a clínica, com recurso à recente e cada vez mais utilizada área das MPS. Os resultados obtidos foram de encontro com a esperada detecção de heteroplasmias a níveis mais reduzidos do que as técnicas mais convencionais de sequenciação por Sanger. Demonstramos a utilidade e viabilidade da amplificação do mtDNA por Long-Range PCR e determinamos os limites para a detecção de variantes heteroplásmicas. Apesar de apresentar algumas dificuldades na identificação de variantes em regiões homopoliméricas e alguns falsos positivos, o Ion Torrent PGM consegue de forma eficaz detetar as variantes de uma amostra com mistura/heteroplasmia, sendo bastante preciso na determinação da frequência de cada variante.

O método aqui apresentado não permite avaliar o que ocorre quando se parte de heteroplasmias reais, onde a amplificação por PCR vai alterar as proporções das variantes presentes. Para podermos avaliar isto, seria necessário quantificar o mtDNA (isolado do nDNA) em cada amostra, e depois misturar as amostras e só então proceder ao PCR. A quantificação de mtDNA isolado é um processo complicado e está longe de ser fiável. Além disto, sabe-se que a amplificação exponencial das moléculas de DNA presentes no PCR é altamente aleatória (Wagner *et al.*, 1994), pelo que em amostras com baixas frequências de heteroplasmia será sempre difícil obter resultados reprodutíveis num estudo deste género.

Assim, consideramos que este estudo é um documento importante para a validação da metodologia de MPS no Ion Torrent PGM aplicada a mtDNA com heteroplasmias/misturas controlando, na medida do possível, as variantes em análise.

REFERÊNCIAS BIBLIOGRÁFICAS

- Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., *et al.* (1981). Sequence and organization of the human mitochondrial genome. *Nature*, 290(5806): 457-65.
- Andersson, S.G., Karlberg, O., Canback, B. & Kurland, C.G. (2003). On the origin of mitochondria: a genomics perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358(1429): 165-79.
- Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. & Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics*, 23(2): 147.
- Bai, R.K. & Wong, L.J. (2004). Detection and quantification of heteroplasmic mutant mitochondrial DNA by real-time amplification refractory mutation system quantitative PCR analysis: a single-step approach. *Clinical Chemistry*, 50: 996-1001.
- Bannwarth, S., Procaccio, V., Lebre, A.S., Jardel, C., Chaussonot, A., Hoarau, C., *et al.* (2013). Prevalence of rare mitochondrial DNA mutations in mitochondrial disorders. *Journal of Medical Genetics*, 50(10): 704-714.
- Beckman Coulter (2013). *Instructions For Use Agencourt AMPure XP PCR Purification*. [Consult. 2 out. 2015] Disponível em: <https://www.beckmancoulter.com/wsrportal/techdocs?docname=B37419AA>
- Birky, C.W. Jr. (2001). The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annual Review of Genetics*, 35: 125-48.
- Brown, D.T., Herbert, M., Lamb, V.K., Chinnery, P.F., Taylor, R.W., Lightowlers, R.N., *et al.* (2006). Transmission of mitochondrial DNA disorders: possibilities for the future. *The Lancet*, 368: 87-9.
- Budowle, B., Allard, M.W., Wilson, M.R. & Chakraborty, R. (2003). Forensics and mitochondrial DNA: applications, debates, and foundations. *Annual Review of Genomics and Human Genetics*, 4: 119-41.
- Budowle, B., Chakraborty, R., Giusti, *et al.* (1991). Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. *American Journal of Human Genetics*, 48(1):137-144.
- Butler, J.M. (2012). *Advanced topics in forensic DNA typing: methodology*. Waltham, Elsevier Academic Press.
- Calandro, L. (2012). *Human Identification Applications on Ion Torrent PGM™*. [Consult. 14 set. 2015] Retirado de: http://pt.slideshare.net/Lifetech_HID/human-identification-applications-on-ion-torrent-pgm
- Carling, P.J., Cree, L.M. & Chinnery, P.F. (2011). The implications of mitochondrial DNA copy number regulation during embryogenesis. *Mitochondrion*, 11: 686-92.
- Carracedo, A., Bar, 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 Science International*, 110: 79-85.
- Cavelier, L., Jazin, E., Jalonen, P. & Gyllensten, U. (2000). MtDNA substitution rate and segregation of heteroplasmy in coding and noncoding regions. *Human Genetics*, 107: 45-50.
- Chen, X., Prosser, R., Simonetti, S., Sadlock, J., Jagiello, G. & Schon, E.A. (1995). Rearranged mitochondrial genomes are present in human oocytes. *American Journal of Human Genetics*, 57: 239-247.

- Chinault, A.C., Shaw, C.A., Brundage E.K., Tang, L.Y. & Wong, L.J. (2009). Application of dual-genome oligonucleotide array-based comparative genomic hybridization to the molecular diagnosis of mitochondrial DNA deletion and depletion syndromes. *Genetics in Medicine*, 11: 518-526.
- Chinnery, P.F., DiMauro, S., Shanske, S., Schon, E.A., Zeviani, M., Mariotti, C., *et al.* (2004). Risk of developing a mitochondrial DNA deletion disorder. *The Lancet*, 364: 592-596.
- Chinnery, P.F., Johnson, M.A., Wardell, T.M., Singh-Kler, R., Hayes, C., Brown, D.T., *et al.* (2000). The epidemiology of pathogenic mitochondrial DNA mutations. *Annals of Neurology*, 48(2): 188-93.
- Chinnery, P.F. & Hudson, G. (2013). Mitochondrial genetics. *British Medical Bulletin*, 106: 135-59.
- Chung, W.K., Martin, K., Jalas, C., Braddock, S.R., Juusola, J., Monaghan, K.G., *et al.* (2015). Mutations in *COQ4*, an essential component of coenzyme Q biosynthesis, cause lethal neonatal mitochondrial encephalomyopathy. *Journal of Medical Genetics*, 52(9): 627-635.
- Churchill, J.D., Chang, J., Ge, J., Rajagopalan, N., Wootton, S.C., Chang, C.W., *et al.* (2015). Blind study evaluation illustrates utility of the Ion PGM system for use in human identity DNA typing. *Croatian Medical Journal*, 56(3): 218-229.
- Coble, M.D., Loreille, O.M., Wadhams, M.J., Edson, S.M., Maynard, K., Meyer, C.E., *et al.* (2009). Mystery Solved: The Identification of the Two Missing Romanov Children Using DNA Analysis. *PLoS ONE*, 4(3): e4838.
- Collins, D.W., Gudiseva, H.V., Trachtman, B.T., Jerrehian, M., Gorry, T., Merritt, W.T. 3rd, *et al.* (2013). Mitochondrial Sequence Variation in African-American Primary Open-Angle Glaucoma Patients. *PLoS ONE*, 8(10): e76627.
- Cree, L.M., Samuels, D.C., de Sousa Lopes, S.C., Rajasimha, H.K., Wonnapijit, P., Mann, J.R., *et al.* (2008). A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. *Nature Genetics*, 40:249-54.
- Cui, H., Li, F., Chen, D., Wang, G., Truong, C.K., Enns, G.M., Graham, B., *et al.* (2013). Comprehensive next-generation sequence analyses of the entire mitochondrial genome reveal new insights into the molecular diagnosis of mitochondrial DNA disorders. *Genetics in Medicine*, 15: 388-394.
- da Fonseca, R.R., Johnson, W.E., O'Brien, S.J., Ramos, M.J. & Antunes, A. (2008). The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics*, 9: 119.
- Diaz-Sanchez, S., Hanning, I., Pendleton, S. & D'Souza, D. (2013). Next-generation sequencing: the future of molecular genetics in poultry production and food safety. *Poultry Science*, (2): 562-572.
- Elliott, H.R., Samuels, D.C., Eden, J.A., Relton, C.L. & Chinnery, P.F. (2008). Pathogenic Mitochondrial DNA Mutations Are Common in the General Population. *American Journal of Human Genetics*, 83(2): 254-260.
- Ewing, B. & Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, 8(3): 186-194.
- Fendt, L., Zimmermann, B., Daniaux, M. & Parson, W. (2009). Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences. *BMC genomics*, 10(1): 139.
- Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., *et al.* (2010). Direct detection of DNA methylation during single-molecule, realtime sequencing. *Nature Methods*, 7(6): 461-65.
- Giles, R.E., Blanc, H., Cann, H.M. & Wallace, D.C. (1980). Maternal inheritance of human mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 77(11), 6715-9.

- Gorman, G.S., Schaefer, A.M., Ng, Y., Gomez, N., Blakely, E.L., Alston, C.L., et al. (2015). Prevalence of nuclear and mtDNA mutations related to adult mitochondrial disease. *Annals of Neurology*, 77(5): 753-9.
- Govindaraj, P., Khan, N.A., Gopalakrishna P., Chandra, R.V., Vanniarajan, A., Reddy, A.A., et al. (2011). Mitochondrial dysfunction and genetic heterogeneity in chronic periodontitis. *Mitochondrion*, 11: 504-12
- Green, R.E., Malaspina, A.S., Krause, J., Briggs, A.W., Johnson, P.L., Uhler, C., et al. (2008). A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell*, 134: 416-426.
- Holland, M.M. & Parsons, T.J. (1999). Mitochondrial DNA sequence analysis - validation and use for forensic casework. *Forensic Science Review*, 11: 21-50.
- Holt, I.J., Harding, A.E. & Morgan-Hughes, J.A. (1988). Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathy. *Nature*, 331: 717-719.
- Inger Jonasson. Next Generation Sequencing on AB SOLiD™ & Ion Torrent PGM™. [Consult. 03 out. 2015] Disponível em: <http://uppnex.se/twiki/pub/Courses/CM1205/WebHome/SolidIonSeq.pdf>
- Ingman, M., Kaessmann, H., Pääbo, S. & Gyllenstein, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature*, 408, 708-13.
- Just, R.S., Irwin, J.A., & Parson, W. (2015). Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. *Forensic Science International. Genetics*, 18: 131-39.
- Kircher, M. & Kelso, J. (2010). High-throughput DNA sequencing - concepts and limitations. *Bioessays*, 32: 524-36.
- Kloss-Brandstatter, A., Pacher, D., Schoenherr S., Weissensteiner, H., Binna, R., Specht, G., et al. (2010). HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Human mutation*, 32(1): 25-32.
- Lehninger, A., Nelson, D.L. & Cox, M.M. (2008). *Lehninger Principles of Biochemistry*. 5ª edição. Nova Iorque, W. H. Freeman.
- Li, L.H., Kang, T., Chen, L., Zhang, W., Liao, Y., Chen, J., et al. (2014). Detection of mitochondrial DNA mutations by high-throughput sequencing in the blood of breast cancer patients. *International Journal of Molecular Medicine*, 33: 77-82.
- Life Technologies (2013). *Ion Xpress™ Plus and Ion Plus Library Preparation for the AB Library Builder™ System*. [Consult. 2 out. 2015] Disponível em: https://tools.thermofisher.com/content/sfs/manuals/MAN0006946_IonXpressLibraryPrepforABLB_UG.pdf
- Life Technologies (2014). *Torrent Suite Software User Documentation*. [Consult. 25 Set. 2015] Disponível em: <http://www.fert.org.uk/ion-docs/>
- Lightowlers, R.N., Chinnery, P.F., Turnbull, D.M. & Howell, N. (1997). Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. *Trends in Genetics*, 13: 450-5.
- Lott, M.T., Leipzig, J.N., Derbeneva, O., Xie, H.M., Chalkia, D., Sarmady, M., et al. (2013). MtDNA variation and analysis using MITOMAP and MITOMASTER. *Current Protocols in Bioinformatics*, 1(123): 1.23.1-1.23.26.
- Luo, L., Boerwinkle, E. & Xiong, M. (2011). Association studies for next-generation sequencing. *Genome Research*, 21: 1099-108.
- Lutz, S., Weisser, H.J., Heizmann, J. & Pollak, S. (1997). A third hypervariable region in the human mitochondrial D-loop. *Human Genetics*, 101: 384.
- Margulis, L. (1981). *Symbiosis in Cell Evolution: Life and Its Environment on the Early Earth*. São Francisco, W. H. Freeman.

- McFarland, R., Taylor, R.W. & Turnbull, D.M. (2007). Mitochondrial disease-its impact, etiology, and pathology. *Current Topics in Developmental Biology*, 77: 113-155.
- Miller, W., Drautz, D.I., Ratan, A., Pusey, B., Qi, J., Lesk, A.M., Tomsho, L.P., *et al.* (2008). Sequencing the nuclear genome of the extinct woolly mammoth. *Nature*, 456: 387-390.
- Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., *et al.* (2003). Natural selection shaped regional mtDNA variation in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 100(1): 171–6.
- Naue, J., Horer, S., Sanger, T., Strobl, C., Hatzer-Grubwieser, P., Parson, W., *et al.* (2015). Evidence for frequent and tissue-specific sequence heteroplasmy in human mitochondrial DNA. *Mitochondrion*, 20: 82-94.
- Ota, M., Fukushima, H., Kulski, J.K. & Inoko, H. (2007). Single nucleotide polymorphism detection by polymerase chain reaction-restriction fragment length polymorphism. *Nature Protocols*, 2 (11): 2857-64.
- Parson, W. & Bandelt, H.J. (2007). Extended guidelines for mtDNA typing of population data in forensic science. *Forensic Science International: Genetics*, 1: 13-19.
- Parson, W., Strobl, C., Strobl, C., Huber, G., Zimmermann, B., Gomes, *et al.* (2013). Evaluation of next generation mtGenome sequencing using the Ion Torrent Personal Genome Machine (PGM). *Forensic Science International Genetics*, 7: 632–639.
- Parsons, T.J. & Coble, M.D. (2001). Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome. *Croatian Medical Journal*, 42: 304-309.
- Parsons, T.J., Muniec, D.S., Sullivan, K., Woodyatt, N., Alliston-Greiner, R., Wilson, M.R., *et al.* (1997). A high observed substitution rate in the human mitochondrial DNA control region. *Nature Genetics*, 15(4): 363-8.
- Payne, B.A.I., Wilson, I.J., Yu-Wai-Man, P., Coxhead, J., Deehan, D., Horvath, R., *et al.* (2013). Universal heteroplasmy of human mitochondrial DNA. *Human Molecular Genetics*, 22(2): 384-90.
- Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., *et al.* (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, 13: 341.
- Ramos, A., Santos, C., Alvarez, L., Nogués, R. & Aluja, M.P. (2009). Human mitochondrial DNA complete amplification and sequencing: a new validated primer set that prevents nuclear DNA sequences of mitochondrial origin co-amplification. *Electrophoresis*, 30: 1587–1593.
- Ramos, A., Santos, C., Barbena, E., Mateiu, L., Alvarez, L., Nogués, R., *et al.* (2011). Validated primer set that prevents nuclear DNA sequences of mitochondrial origin co-amplification: a revision based on the New Human Genome Reference Sequence (GRCh37). *Electrophoresis*, 32: 782–783.
- Ramos, A., Santos, C., Mateiu, L., Gonzalez, M. del M., Alvarez, L., Azevedo, L., *et al.* (2013). Frequency and Pattern of Heteroplasmy in the Complete Human Mitochondrial Genome. *PLoS ONE*, 8(10): e74636.
- Reiner, J.E., Kishore, R.B., Levin, B.C., Albanetti, T., Boire, N., Knipe, A., *et al.* (2010). Detection of Heteroplasmic Mitochondrial DNA in Single Mitochondria. *PLoS ONE*, 5(12): e14359.
- Ricchetti, M., Tekai, F. & Dujon, B. (2004). Continued Colonization of the Human Genome by Mitochondrial DNA. *PLoS Biology*, 2(9): e273.
- Roberts, K.A. & Calloway, C. (2011). Characterization of mitochondrial DNA sequence heteroplasmy in blood tissue and as a function of hair morphology in four population groups. *Journal of Forensic Sciences*, 56(1): 46-60.

- Rohlin, A., Wernersson, J., Engwall, Y., Wiklund, L., Björk, J. & Nordling, M. (2009). Parallel sequencing used in detection of mosaic mutations: comparison with four diagnostic DNA screening techniques. *Human Mutation*, 30: 1012–20.
- Ross, M.G., Russ, C., Costello, M., Hollinger, A., Lennon, N.J., Hegarty, R., *et al.* (2013). Characterizing and measuring bias in sequence data. *Genome Biology*, 14(5): R51.
- Rossignol, R., Faustin, B., Rocher, C., Malgat, M., Mazat, J.P. & Letellier, T. (2003). Mitochondrial threshold effects. *The Biochemical Journal*, 370: 751–62.
- Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., *et al.* (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475: 348-352.
- Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12): 5463-5467.
- Sanger, F. & Coulson, A.R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, 94: 441-448.
- Schaefer, A.M., McFarland, R., Blakely, E.L., He, L., Whittaker, R.G., Taylor, R.W., *et al.* (2008). Prevalence of mitochondrial DNA disease in adults. *Annals of Neurology*, 63: 35-9.
- Schwartz, M. & Vissing, J. (2002). Paternal Inheritance of Mitochondrial DNA. *New England Journal of Medicine*, 347: 576-80.
- Skladal, D., Halliday, J. & Thorburn, D.R. (2003). Minimum birth prevalence of mitochondrial respiratory chain disorders in children. *Brain: a journal of neurology*, 126(Pt 8):1905-12.
- Stewart, J.E.B., Fisher, C.L., Aagaard, P.J., Wilson, M.R., Isenberg, A.R., Polansky, D., *et al.* (2001). Length variation in HV2 of the human mitochondrial DNA control region. *Journal of Forensic Sciences*, 46(4): 862-70.
- Sutovsky, P., Moreno, R.D., Ramalho-Santos, J., Dominko, T., Simerly, C. & Schatten, G. (2000). Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. *Biology of Reproduction*, 63(2): 582-90.
- Tang, S. & Huang, T. (2010). Characterization of mitochondrial DNA heteroplasmy using a parallel sequencing system. *Biotechniques*, 48(4): 287-96.
- Taylor, R.W. & Turnbull, D.M. (2005). Mitochondrial DNA mutations in Human disease. *Nature Reviews: Genetics*, 6(5): 389-402.
- Taylor, R.W., McDonnell, M.T., Blakely, E.L., Chinnery, P.F., Taylor, G.A., Howell, N., *et al.* (2003). Genotypes from patients indicate no paternal mitochondrial DNA contribution. *Annals of Neurology*, 54(4): 521-4.
- Thermo Fisher (2014). *E-Gel® Technical Guide*. [Consult. 27 out. 2015] Disponível em: <https://tools.thermofisher.com>
- Thermo Fisher. *Dynabeads® MyOne™ Streptavidin C1*. [Consult. 05 out. 2015] Disponível em: <https://www.thermofisher.com/order/catalog/product/65001>
- Thorburn, D.R. (2004a). Mitochondrial disorders: prevalence, myths and advances. *Journal of Inherited Metabolic Disease*, 27: 349-362.
- Thorburn, D.R. (2004b). Mitochondrial diseases: not so rare after all. *Internal Medicine Journal*, 34: 3-5.
- Tuppen, H.A., Blakely, E.L., Turnbull, D.M. & Taylor, R.W. (2009). Mitochondrial DNA mutations and human disease. *Biochimica et Biophysica Acta*, 1797(2):113-28.
- van Oven, M. & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, 30(2): e386-e394. <http://www.phylotree.org>

- Wagner, A., Blackstone, N., Cartwright, P., Dick, M., Misof, B., Snow, P., *et al.* (1994). Surveys of Gene Families Using Polymerase Chain Reaction: PCR Selection and PCR Drift. *Systematic Biology*, 43(2): 250-61.
- Wallace, D.C., Brown, M.D. & Lott, M.T. (1999). Mitochondrial DNA variation in human evolution and disease. *Gene*, 238: 211-30.
- Wallace, D.C., Singh, G., Lott, M.T., Hodge, J.A., Schurr, T.G., Lezza, A.M., *et al.* (1988). Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science*, 242: 1427-1430.
- Weber-Lehmann, J., Schilling, E., Gradl, G., Richter, D.C., Wiehler, J., Rolf, B. (2014). Finding the needle in the haystack: differentiating 'identical' twins in paternity testing and forensics by ultra-deep next generation sequencing. *Forensic Science International: Genetics*, 9(1): 42-46.
- Wong, L.J. C. (2013). Challenges of Bringing Next Generation Sequencing Technologies to Clinical Molecular Diagnostic Laboratories. *Neurotherapeutics*, 10(2): 262–272.
- Wong, L.J., Dimmock, D., Geraghty, M.T., Quan, R., Lichter-Konecki, U., Wang, J., *et al.* (2008). Utility of oligonucleotide array-based comparative genomic hybridization for detection of target gene deletions. *Clinical Chemistry*, 54: 1141-8.
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13:134.
- Yeung, K.Y., Dickinson, A., Donoghue, J.F., Polekhina, G., White, S.J., Grammatopoulos, D.K., *et al.* (2014). The identification of mitochondrial DNA variants in glioblastoma multiforme. *Acta Neuropathologica Communications*, 2:1.
- Young, J.M., Weyrich, L.S. & Cooper, A. (2014). Forensic soil DNA analysis using high-throughput sequencing: a comparison of four molecular markers. *Forensic Science International: Genetics*, 13: 176-84.
- Zhang, W., Cui, H. & Wong, L.J. (2012). Comprehensive one-step molecular analyses of mitochondrial genome by massively parallel sequencing. *Clinical Chemistry*, 58: 1322–31.

ANEXOS

I. Tabelas de resultados do *plug-in Variant Caller***Tabela 8** – Tabela parcial dos resultados do *plug-in Variant Caller* da mistura 1(T4):99(T5).

Position	Ref	Variant	Freq.	Quality	Type	Cov.	Cov+	Cov-	Al. Cov	Al. Cov+	Al. Cov-	S. Bias
73	A	G	76,6	27333,3	SNP	1378	590	788	1055	590	465	0,5
71	GTATGC	TAGCTA	23,4	27333,3	MNP	1378	590	788	323	0	323	0,5
152	T	C	100	43533,7	SNP	1977	1333	644	1977	1333	644	0,5
182	C	T	98,6	43257	SNP	1999	1339	660	1972	1325	647	0,5023
185	G	T	98,6	43082,6	SNP	1991	1356	635	1964	1339	625	0,5008
195	T	C	98,3	42971,3	SNP	1999	1358	641	1965	1335	630	0,5001
198	C	T	98,6	42916,2	SNP	1983	1333	650	1955	1320	635	0,5034
247	G	A	98,2	42689,4	SNP	1991	1194	797	1956	1170	786	0,5016
263	A	G	100	44119,3	SNP	1989	1204	785	1989	1204	785	0,5
310	T	C	8,1	7864,65	SNP	557	5	552	45	0	45	0,5
311	-	C	91,9	7864,65	INS	557	5	552	512	5	507	0,5
357	A	G	96,6	12706,1	SNP	655	152	503	633	147	486	0,5002
514	CA	-	98,5	12595,9	DEL	620	45	575	611	45	566	0,5039
709	GT	AC	98,5	30658,9	MNP	1980	895	1085	1950	885	1065	0,5018
750	A	G	100	44109,7	SNP	1990	912	1078	1990	912	1078	0,5
769	G	A	98	42669,3	SNP	1997	918	1079	1957	899	1058	0,5003
825	T	A	98	42470,4	SNP	1991	988	1003	1951	965	986	0,5016
1018	G	A	98,3	42827	SNP	1996	594	1402	1962	588	1374	0,5025
1462	G	A	98,2	42842,6	SNP	1999	1036	963	1963	1017	946	0,5002
1738	T	C	100	36633,3	SNP	1674	530	1144	1674	530	1144	0,5
2352	T	C	97	41553,6	SNP	1978	909	1069	1919	885	1034	0,5016
2706	A	G	98,5	43124,9	SNP	1998	1079	919	1969	1061	908	0,5012
2758	G	A	98,5	43031,1	SNP	1995	1097	898	1965	1082	883	0,5008
2768	A	G	98,4	43096,8	SNP	1999	1070	929	1968	1051	917	0,5012
2885	T	C	98,8	43254,9	SNP	1995	977	1018	1971	966	1005	0,5004
3290	T	C	1,3	42,96	SNP	1997	785	1212	25	1	24	0,9391
3308	T	C	98,8	43439,2	SNP	2000	831	1169	1977	822	1155	0,5003
3594	C	T	97	19274,5	SNP	925	170	755	897	169	728	0,5076
3666	G	A	97,5	37679,8	SNP	1786	876	910	1741	864	877	0,5058
3693	G	A	98,3	42948,9	SNP	1997	1034	963	1964	1021	943	0,5021
4104	A	G	98,2	40919,8	SNP	1910	896	1014	1876	880	996	0,5
4769	A	G	100	44227,9	SNP	1996	1125	871	1996	1125	871	0,5
5036	A	G	97	41920,1	SNP	2000	1007	993	1939	963	976	0,5069
5046	G	A	98,2	42902,9	SNP	1998	1010	988	1963	992	971	0,5002
5393	T	C	98,6	21418,4	SNP	1020	391	629	1006	387	619	0,5014
5655	T	C	98,5	43030,3	SNP	1995	1070	925	1965	1054	911	0,5
6548	C	T	98,6	43130,6	SNP	1997	997	1000	1969	979	990	0,502
6698	AG	GA	2,5	104,7	MNP	1927	985	942	49	6	43	0,8821
6827	T	C	99	43554,1	SNP	2000	1082	918	1979	1075	904	0,5022
6989	A	G	98,7	43299,4	SNP	1999	1095	904	1974	1081	893	0,5002
7028	C	T	98,5	43158,9	SNP	2000	1131	869	1970	1117	853	0,5015

7055	A	G	98,6	42582,7	SNP	1992	1171	821	1965	1153	812	0,5011
7146	A	G	94,7	40076	SNP	1989	1153	836	1884	1058	826	0,5185
7256	C	T	91,1	37708,2	SNP	1994	1138	856	1816	971	845	0,5364
7389	T	C	98,7	36816	SNP	1709	439	1270	1687	432	1255	0,501
7521	G	A	98,5	42563,2	SNP	1993	840	1153	1964	830	1134	0,5012
7867	C	T	98	42738,7	SNP	1999	708	1291	1959	691	1268	0,5016
8248	A	G	98,9	43113,4	SNP	1993	1146	847	1971	1134	837	0,5003
8468	C	T	97,2	21446,9	SNP	1050	99	951	1021	96	925	0,5008
8655	C	T	98,2	22893	SNP	1094	203	891	1074	200	874	0,5011
8701	A	G	96,4	27501,8	SNP	1343	404	939	1295	396	899	0,5059
8860	A	C	3,9	37733,1	SNP	1761	685	1076	68	67	1	0,5
8860	A	G	96,1	37733,1	SNP	1761	685	1076	1693	618	1075	0,5
9540	T	C	96,9	35550,7	SNP	1706	341	1365	1653	333	1320	0,5024
10398	A	G	97,9	42583,9	SNP	1995	919	1076	1954	906	1048	0,503
10688	G	A	97,7	33181,1	SNP	1576	563	1013	1540	547	993	0,5022
10810	T	C	97,8	34033,5	SNP	1609	828	781	1574	816	758	0,5038
10873	T	C	97	19949,5	SNP	929	334	595	901	315	586	0,5108
11002	A	G	99	21998,8	SNP	1038	291	747	1028	289	739	0,501
11719	G	A	98,5	43113,7	SNP	1998	1259	739	1968	1242	726	0,501
12519	T	C	98,7	41509,1	SNP	1919	817	1102	1894	807	1087	0,5003
12705	C	T	98,7	44339	SNP	1976	994	982	1950	981	969	0,5
13105	A	G	98,6	43212	SNP	1999	852	1147	1971	840	1131	0,5
13506	C	T	98,8	43157,1	SNP	1990	971	1019	1966	956	1010	0,5017
13650	C	T	98,6	21413,3	SNP	996	119	877	982	117	865	0,5008
13789	T	C	98,2	21220,3	SNP	990	124	866	972	120	852	0,5041
13880	C	A	97,8	27449,6	SNP	1313	519	794	1284	506	778	0,5013
14178	T	C	100	21795,6	SNP	1017	388	629	1017	388	629	0,5
14203	A	G	98,8	21335,8	SNP	1012	397	615	1000	394	606	0,5018
14560	G	A	98,6	12442,2	SNP	588	7	581	580	7	573	0,5035
14766	C	T	98	10282,1	SNP	499	22	477	489	20	469	0,5196
14768	A	T	5,4	161,44	SNP	515	22	493	28	5	23	0,8296
14769	A	G	92	9534,28	SNP	511	23	488	470	16	454	0,5722
14770	C	G	5,5	161,79	SNP	511	23	488	28	7	21	0,876
15115	T	C	99	43483	SNP	2000	1027	973	1979	1018	961	0,5009
15326	A	G	100	44145	SNP	1992	610	1382	1992	610	1382	0,5
16126	T	C	98,7	43217,3	SNP	1996	1010	986	1970	1001	969	0,5021
16187	C	T	90,4	19371	SNP	1062	48	1014	960	21	939	0,6791
16189	T	C	95,4	18730,1	SNP	1062	57	1005	1013	39	974	0,5862
16223	C	T	98	25762,2	SNP	1229	163	1066	1205	159	1046	0,5015
16264	C	T	97,4	29336,3	SNP	1418	50	1368	1381	47	1334	0,5092
16270	C	T	96,4	30830,9	SNP	1507	125	1382	1452	111	1341	0,5222
16278	C	T	96,7	33596,4	SNP	1621	217	1404	1567	209	1358	0,5011
16293	A	G	95	37209,1	SNP	1864	418	1446	1770	358	1412	0,5327
16304	T	A	3,1	245,52	SNP	1995	503	1492	62	15	47	0,5137
16311	T	C	97,6	42460,4	SNP	1999	569	1430	1952	553	1399	0,5017
16400	C	T	91	37836,1	SNP	1997	876	1121	1818	743	1075	0,5307
16506	T	C	4,3	416,27	SNP	1997	984	1013	85	44	41	0,5249

16519	T	C	100	43437,4	SNP	1972	1113	859	1972	1113	859	0,5
-------	---	---	-----	---------	-----	------	------	-----	------	------	-----	-----

Tabela 9 – Tabela parcial dos resultados do *plug-in Variant Caller* da mistura 5(T4):95(T5).

Position	Ref	Variant	Freq.	Quality	Type	Cov.	Cov+	Cov-	Al. Cov	Al. Cov+	Al. Cov-	S. Bias
73	A	G	77	27092	SNP	1441	671	770	1109	645	464	0,5026
71	GTATGC	TAGCTA	19,7	27092	MNP	1441	671	770	284	1	283	0,96
152	T	C	95	40499,1	SNP	1998	1392	606	1898	1332	566	0,5061
182	C	T	94,5	40216,8	SNP	1999	1353	646	1890	1290	600	0,5065
185	G	T	94,4	40119,5	SNP	1999	1374	625	1887	1313	574	0,5099
195	T	C	94,8	40377,5	SNP	1999	1359	640	1895	1294	601	0,5035
198	C	T	94,7	40134,3	SNP	1992	1354	638	1886	1297	589	0,5092
247	G	A	92,3	38585,9	SNP	1997	1228	769	1843	1143	700	0,5056
263	A	G	100	44113,5	SNP	1987	1242	745	1987	1242	745	0,5
310	T	C	4,8	6089,59	SNP	440	6	434	21	0	21	0,5
311	-	C	93,4	6089,59	INS	440	6	434	411	5	406	0,5
357	A	G	89,3	8264,38	SNP	488	119	369	436	105	331	0,5041
358	A	G	6,8	214,99	SNP	502	122	380	34	28	6	0,9355
514	CA	-	94,6	9322,48	DEL	497	51	446	470	46	424	0,5131
709	GT	AC	93,8	27881,6	MNP	1982	869	1113	1859	818	1041	0,5016
750	A	G	100	44154	SNP	1993	912	1081	1993	912	1081	0,5
769	G	A	93,7	39623,7	SNP	2000	943	1057	1874	881	993	0,5014
825	T	A	91,5	37913,8	SNP	1990	953	1037	1820	872	948	0,5002
1018	G	A	93,8	39597,8	SNP	1995	700	1295	1871	649	1222	0,5044
1438	A	G	7,3	1032,8	SNP	1999	1013	986	145	69	76	0,5309
1462	G	A	93,9	39605,5	SNP	1993	1026	967	1871	963	908	0,5001
1738	T	C	94,6	29619,3	SNP	1495	470	1025	1414	449	965	0,5037
2259	C	T	6,5	869,04	SNP	1999	808	1191	130	67	63	0,6105
2352	T	C	91,6	38059,6	SNP	1989	929	1060	1822	846	976	0,5028
2706	A	G	94,3	40088,2	SNP	1999	1061	938	1886	995	891	0,5032
2758	G	A	93,8	39622	SNP	1998	1057	941	1874	995	879	0,5019
2768	A	G	94,4	40147,6	SNP	2000	1094	906	1888	1033	855	0,5001
2885	T	C	94	39810,3	SNP	1998	930	1068	1878	879	999	0,5026
3168	C	T	2,1	69,75	SNP	1978	999	979	42	36	6	0,8545
3290	T	C	1	28,72	SNP	1650	716	934	17	2	15	0,8514
3308	T	C	93,9	34395,9	SNP	1738	801	937	1632	752	880	0,5001
3594	C	T	93,5	13360,6	SNP	692	166	526	647	156	491	0,5017
3666	G	A	94,1	27501,5	SNP	1396	755	641	1314	718	596	0,5056
3693	G	A	95,1	37301	SNP	1843	1005	838	1753	959	794	0,5018
4104	A	G	94,2	26735,8	SNP	1358	695	663	1279	646	633	0,5067
4745	A	G	6	773,38	SNP	2000	1078	922	121	69	52	0,5316
4769	A	G	100	44301,5	SNP	1996	1099	897	1996	1099	897	0,5
5036	A	G	93,2	35688,9	SNP	1822	992	830	1699	910	789	0,5089
5046	G	A	94,6	35006,5	SNP	1749	952	797	1654	896	758	0,5026
5393	T	C	95,4	15268,4	SNP	756	264	492	721	250	471	0,5027
5655	T	C	94,7	40167,7	SNP	1993	1110	883	1887	1053	834	0,5011

6548	C	T	95,7	40881,1	SNP	1993	1027	966	1907	986	921	0,5017
6698	AG	GA	2,6	99,08	MNP	1932	984	948	50	5	45	0,9031
6827	T	C	93,8	39680,5	SNP	1997	1040	957	1874	976	898	0,5
6989	A	G	94,1	39813	SNP	1993	1074	919	1876	1018	858	0,5038
7028	C	T	94,1	39935,7	SNP	2000	1158	842	1882	1091	791	0,5007
7055	A	G	93,6	39159,7	SNP	1994	1181	813	1867	1106	761	0,5001
7146	A	G	89,6	36736,7	SNP	1989	1130	859	1782	974	808	0,5218
7256	C	T	86,2	34807,9	SNP	2000	1072	928	1725	858	867	0,5386
7389	T	C	93,3	28334,5	SNP	1456	336	1120	1359	317	1042	0,5035
7521	G	A	93	38798,2	SNP	1982	815	1167	1843	764	1079	0,5034
7867	C	T	94,8	40343,8	SNP	1998	734	1264	1894	701	1193	0,503
8248	A	G	94,2	39736,9	SNP	1995	1200	795	1880	1136	744	0,5029
8468	C	T	94,1	15624,6	SNP	791	80	711	744	72	672	0,5122
8655	C	T	94,4	16046	SNP	806	165	641	761	157	604	0,5024
8701	A	G	92,3	19377,8	SNP	1030	314	716	951	291	660	0,5013
8860	A	C	3,8	31672,3	SNP	1483	624	859	56	56	0	0,5
8860	A	G	96,2	31672,3	SNP	1483	624	859	1427	568	859	0,5
9540	T	C	93,5	25864,5	SNP	1332	293	1039	1245	278	967	0,5048
10398	A	G	93,6	39435,7	SNP	1992	1019	973	1865	960	905	0,5032
10688	G	A	93,5	26407,8	SNP	1356	471	885	1268	443	825	0,5022
10810	T	C	93,7	25555,6	SNP	1311	700	611	1228	669	559	0,5109
10873	T	C	93,3	15348,4	SNP	774	308	466	722	277	445	0,515
11002	A	G	94,8	16718,5	SNP	832	240	592	789	228	561	0,5006
11719	G	A	93,8	39713,6	SNP	1999	1252	747	1875	1182	693	0,5044
12061	C	T	3,9	368,25	SNP	1995	1161	834	78	48	30	0,5347
12519	T	C	94,7	29586,4	SNP	1482	652	830	1404	618	786	0,5002
12705	C	T	94,4	41286	SNP	1978	1002	976	1867	935	932	0,5058
13105	A	G	94,4	32750,5	SNP	1643	765	878	1551	718	833	0,5027
13506	C	T	94,6	33913	SNP	1696	858	838	1604	816	788	0,5028
13650	C	T	95	16925,8	SNP	844	117	727	802	115	687	0,5098
13680	C	T	4,5	169,27	SNP	710	13	697	32	1	31	0,6336
13789	T	C	94,3	16392,8	SNP	825	146	679	778	137	641	0,5015
13880	C	A	93,3	21520,4	SNP	1122	510	612	1047	476	571	0,5001
14178	T	C	95,7	15499,4	SNP	763	326	437	730	307	423	0,5069
14203	A	G	95	15389,8	SNP	766	341	425	728	320	408	0,5057
14560	G	A	94,6	10567	SNP	539	12	527	510	11	499	0,5081
14766	C	T	95	7272,13	SNP	379	9	370	360	9	351	0,5132
14768	A	T	7,1	180,38	SNP	395	9	386	28	4	24	0,8772
14769	A	G	88,2	6693,87	SNP	389	9	380	343	6	337	0,5709
14770	C	G	6,5	155,9	SNP	386	9	377	25	3	22	0,8509
14872	C	T	4,8	485,79	SNP	1815	681	1134	88	30	58	0,5372
15115	T	C	94,2	39945,1	SNP	1998	1083	915	1882	1021	861	0,5005
15326	A	G	100	35277,9	SNP	1603	481	1122	1603	481	1122	0,5
16126	T	C	95,5	39289,3	SNP	1924	1047	877	1837	997	840	0,5015
16187	C	T	88,7	13960,2	SNP	776	50	726	688	23	665	0,6657
16189	T	C	92,3	13130,3	SNP	775	55	720	715	39	676	0,5697
16223	C	T	95,7	18051,4	SNP	884	140	744	846	127	719	0,5158

16264	C	T	90,9	19118,9	SNP	1054	41	1013	958	38	920	0,5051
16270	C	T	88,1	19645,7	SNP	1121	91	1030	988	68	920	0,5445
16278	C	T	90,4	21681,9	SNP	1182	155	1027	1069	137	932	0,5066
16293	A	G	87,1	24805,4	SNP	1417	365	1052	1234	289	945	0,5315
16304	T	A	3,6	321,91	SNP	1996	554	1442	72	26	46	0,5953
16311	T	C	87,3	35532	SNP	1999	635	1364	1745	545	1200	0,5062
16400	C	T	79,8	31146,3	SNP	1998	920	1078	1595	686	909	0,5307
16461	A	G	4,6	458,61	SNP	1862	1005	857	85	79	6	0,9181
16506	T	C	4,3	414,69	SNP	1998	1051	947	85	48	37	0,5389
16519	T	C	97,7	42085,3	SNP	1988	1086	902	1943	1061	882	0,5002

Tabela 10 – Tabela parcial dos resultados do *plug-in Variant Caller* da mistura 10(T4):90(T5).

Position	Ref	Variant	Freq.	Quality	Type	Cov.	Cov+	Cov-	Al. Cov	Al. Cov+	Al. Cov-	S. Bias
73	A	G	73,8	26123,9	SNP	1487	716	771	1097	652	445	0,5019
71	GTATGC	TAGCTA	19,2	26123,9	MNP	1487	716	771	286	0	286	0,9999
152	T	C	89,7	36974,4	SNP	1997	1361	636	1792	1241	551	0,5128
182	C	T	89,8	37037,4	SNP	1997	1345	652	1793	1220	573	0,5079
185	G	T	89,6	36941,8	SNP	1998	1357	641	1790	1239	551	0,5151
195	T	C	89,4	36837,5	SNP	1999	1347	652	1787	1208	579	0,5025
198	C	T	89,1	36485,6	SNP	1987	1340	647	1770	1213	557	0,5125
247	G	A	86,6	34953,9	SNP	1993	1154	839	1726	1004	722	0,5027
263	A	G	100	43972,7	SNP	1983	1179	804	1983	1179	804	0,5
310	T	C	7,4	9019,45	SNP	665	7	658	49	0	49	0,5
311	-	C	92,6	9019,45	INS	665	7	658	616	7	609	0,5
357	A	G	83,5	12256	SNP	772	195	577	645	163	482	0,5002
514	CA	-	88,4	10566,8	DEL	620	50	570	548	37	511	0,5478
709	GT	AC	87,5	24563,9	MNP	1974	936	1038	1728	825	903	0,5033
750	A	G	100	43749,4	SNP	1987	977	1010	1987	977	1010	0,5
769	G	A	87,3	35574	SNP	2000	992	1008	1747	864	883	0,5014
825	T	A	87,3	35339,7	SNP	1991	1004	987	1738	876	862	0,5002
1018	G	A	89,2	36582,3	SNP	1995	639	1356	1779	573	1206	0,5021
1438	A	G	10,4	1807,6	SNP	2000	1070	930	209	106	103	0,5278
1462	G	A	88,3	36053,6	SNP	1997	1084	913	1763	960	803	0,5017
1738	T	C	88,9	34960,8	SNP	1921	581	1340	1707	510	1197	0,5044
2259	C	T	10,2	1728,74	SNP	1999	852	1147	203	103	100	0,581
2352	T	C	87,8	35563,3	SNP	1982	948	1034	1741	817	924	0,5091
2706	A	G	89,1	36664,4	SNP	2000	1083	917	1782	974	808	0,5051
2758	G	A	89,4	36728,9	SNP	1998	1064	934	1786	964	822	0,5073
2768	A	G	89,3	36768,1	SNP	2000	1082	918	1786	973	813	0,5038
2885	T	C	88,9	36533,2	SNP	1998	943	1055	1777	843	934	0,5024
3308	T	C	89,2	36741,3	SNP	2000	861	1139	1785	763	1022	0,5031
3594	C	T	90,8	16723,7	SNP	896	158	738	814	144	670	0,501
3666	G	A	88,3	30286,9	SNP	1685	816	869	1488	720	768	0,5004
3693	G	A	89,3	36624,5	SNP	1995	1038	957	1782	929	853	0,501
4104	A	G	89,4	36576,2	SNP	1985	945	1040	1775	834	941	0,5062

4745	A	G	10,9	1922,31	SNP	2000	1134	866	218	127	91	0,5159
4769	A	G	100	44020,4	SNP	1989	1104	885	1989	1104	885	0,5
5036	A	G	88,2	36088,4	SNP	1998	1106	892	1762	954	808	0,5122
5046	G	A	89,6	36921,9	SNP	1997	1112	885	1789	987	802	0,5052
5393	T	C	89,9	18749	SNP	1039	364	675	934	332	602	0,5056
5655	T	C	88,8	36346,5	SNP	1994	1109	885	1770	986	784	0,5009
6548	C	T	90,6	37594,9	SNP	1997	1021	976	1810	931	879	0,5031
6698	AG	GA	2,8	96,33	MNP	1923	1034	889	53	14	39	0,7641
6827	T	C	88,7	36394,6	SNP	1998	1103	895	1773	968	805	0,5061
6989	A	G	88,8	36485,5	SNP	1999	1104	895	1776	983	793	0,5012
7028	C	T	87,7	35811,2	SNP	1999	1149	850	1754	1017	737	0,5052
7055	A	G	88,7	35653,2	SNP	1995	1174	821	1770	1042	728	0,5002
7146	A	G	86	34560,2	SNP	1994	1224	770	1714	1038	676	0,5087
7256	C	T	84,6	33801,9	SNP	1998	1158	840	1691	950	741	0,5181
7389	T	C	89,4	34366,7	SNP	1869	485	1384	1671	434	1237	0,5003
7521	G	A	88	35799	SNP	1992	805	1187	1753	701	1052	0,5044
7867	C	T	91,2	38017,2	SNP	1999	777	1222	1824	728	1096	0,5109
8248	A	G	88,9	36146,8	SNP	1989	1181	808	1769	1065	704	0,5086
8468	C	T	87,6	20512,7	SNP	1175	126	1049	1029	100	929	0,5274
8655	C	T	90	21203,6	SNP	1165	270	895	1049	255	794	0,5156
8701	A	G	89,2	27167,2	SNP	1495	504	991	1334	470	864	0,5168
8860	A	C	3	42355,6	SNP	1959	871	1088	59	58	1	0,5
8860	A	G	97	42355,6	SNP	1959	871	1088	1900	813	1087	0,5
9540	T	C	85,9	30013,4	SNP	1739	320	1419	1494	277	1217	0,5023
10398	A	G	91,4	38097,8	SNP	1997	966	1031	1826	892	934	0,5048
10688	G	A	88,9	33697	SNP	1850	703	1147	1644	618	1026	0,5043
10810	T	C	88,5	33525,7	SNP	1854	1013	841	1640	897	743	0,5006
10873	T	C	86,5	18531,3	SNP	1080	427	653	934	344	590	0,5286
11002	A	G	91,5	19767,9	SNP	1065	275	790	974	255	719	0,5047
11719	G	A	88	35930,7	SNP	1996	1201	795	1757	1066	691	0,5052
12061	C	T	8,9	1404,68	SNP	1999	1169	830	177	107	70	0,5204
12519	T	C	89,8	37099,4	SNP	1998	876	1122	1795	782	1013	0,5028
12705	C	T	89,3	37510,5	SNP	1978	1023	955	1766	903	863	0,5059
13046	-	C	11,1	1662,32	INS	1949	999	950	216	203	13	0,9369
13105	A	G	89,4	36801,6	SNP	1998	963	1035	1786	862	924	0,5007
13506	C	T	91	37654,4	SNP	1992	1089	903	1812	981	831	0,5053
13650	C	T	91,5	20065	SNP	1087	167	920	995	162	833	0,5172
13680	C	T	9,1	679,69	SNP	924	16	908	84	2	82	0,5805
13789	T	C	90	18505,8	SNP	1025	152	873	922	132	790	0,5103
13880	C	A	88,2	23678,3	SNP	1336	605	731	1178	524	654	0,5081
14178	T	C	90,1	18767,4	SNP	1038	412	626	935	367	568	0,5046
14203	A	G	89,7	18449,7	SNP	997	404	593	894	359	535	0,5038
14560	G	A	88,4	11108,3	SNP	631	34	597	558	28	530	0,5188
14766	C	T	86,3	9585,6	SNP	569	31	538	491	26	465	0,5075
14768	A	T	6,2	226,95	SNP	583	31	552	36	6	30	0,7807
14769	A	G	79,9	8797,59	SNP	577	31	546	461	18	443	0,5829
14770	C	G	6,2	228,05	SNP	579	33	546	36	8	28	0,8253

14872	C	T	10,8	1900,34	SNP	1995	658	1337	216	82	134	0,5542
15115	T	C	89,6	36930,3	SNP	1997	1064	933	1790	954	836	0,5002
15326	A	G	100	44062,7	SNP	1990	627	1363	1990	627	1363	0,5
16126	T	C	90,9	37758,1	SNP	1996	1018	978	1815	931	884	0,5029
16187	C	T	82,8	19287,8	SNP	1201	65	1136	994	15	979	0,7888
16189	T	C	87	17117,7	SNP	1175	59	1116	1022	29	993	0,6442
16223	C	T	89,8	24207,6	SNP	1326	180	1146	1191	168	1023	0,5111
16264	C	T	82,2	24968,3	SNP	1560	58	1502	1283	46	1237	0,5094
16270	C	T	80,9	26176,7	SNP	1666	158	1508	1347	122	1225	0,5127
16278	C	T	80,8	28551,4	SNP	1808	276	1532	1460	215	1245	0,5106
16293	A	G	79,3	29827,6	SNP	1934	534	1400	1534	403	1131	0,517
16311	T	C	77,4	29756	SNP	1998	600	1398	1546	467	1079	0,5021
16400	C	T	70,4	25974,2	SNP	1998	852	1146	1406	569	837	0,5224
16506	T	C	3,9	347,11	SNP	1999	970	1029	77	31	46	0,5831
16519	T	C	95,6	40448,7	SNP	1985	1089	896	1898	1027	871	0,5076

Tabela 11 – Tabela parcial dos resultados do *plug-in Variant Caller* da mistura 50(T4):50(T5).

Position	Ref	Variant	Freq.	Quality	Type	Cov.	Cov+	Cov-	Al. Cov	Al. Cov+	Al. Cov-	S. Bias
73	A	G	52	14182,2	SNP	1415	733	682	736	397	339	0,5214
152	T	C	50,9	16592,9	SNP	1995	1231	764	1015	677	338	0,5542
182	C	T	50,5	16452,7	SNP	1998	1245	753	1008	675	333	0,5508
185	G	T	48,9	15751	SNP	1996	1209	787	976	655	321	0,5705
195	T	C	50,9	16603,9	SNP	1995	1204	791	1015	672	343	0,5628
198	C	T	48,9	15742,4	SNP	1993	1205	788	975	628	347	0,542
247	G	A	41,5	12621,1	SNP	1999	1130	869	830	468	362	0,5015
263	A	G	100	43937	SNP	1983	1128	855	1983	1128	855	0,5
310	T	C	6,1	9342,91	SNP	720	1	719	44	0	44	0,5
311	-	C	92,6	9342,91	INS	720	1	719	667	1	666	0,5
357	A	G	43,1	5097,38	SNP	775	189	586	334	84	250	0,5102
514	CA	-	45,3	3904,29	DEL	592	39	553	268	14	254	0,5613
709	GT	AC	44,1	8308,83	MNP	1989	948	1041	878	437	441	0,5211
750	A	G	100	44002,8	SNP	1988	987	1001	1988	987	1001	0,5
769	G	A	45,5	14302,9	SNP	2000	1040	960	910	505	405	0,5351
825	T	A	44,1	13660,1	SNP	1995	983	1012	880	426	454	0,5086
1018	G	A	48,7	15664,9	SNP	1996	667	1329	973	324	649	0,5013
1438	A	G	51,8	17034,7	SNP	1998	1018	980	1034	519	515	0,5076
1462	G	A	49,1	15781,9	SNP	1994	1033	961	979	511	468	0,5039
1738	T	C	47,3	12357,1	SNP	1654	535	1119	783	256	527	0,504
2259	C	T	52,6	17417,6	SNP	1999	1011	988	1051	559	492	0,5261
2352	T	C	43,3	13289,5	SNP	1990	1032	958	861	405	456	0,5481
2706	A	G	47,6	15195,8	SNP	1998	1052	946	951	492	459	0,5092
2758	G	A	50,7	16538,5	SNP	1996	1051	945	1012	540	472	0,5071
2768	A	G	48,2	15496,8	SNP	2000	1055	945	965	538	427	0,5302
2885	T	C	49,4	15958,8	SNP	1997	937	1060	987	473	514	0,5101
3290	T	C	0	10,48	SNP	1978	871	1107	0	0	0	0,5

Sequenciação completa do mtDNA por MPS e avaliação da deteção de heteroplasmias

3308	T	C	46,2	14586,7	SNP	1999	900	1099	923	394	529	0,5237
3594	C	T	50,8	6690,05	SNP	813	164	649	413	77	336	0,5244
3666	G	A	48,6	12019,6	SNP	1552	753	799	755	365	390	0,5017
3693	G	A	49	15670,2	SNP	1991	1063	928	975	514	461	0,5067
4104	A	G	49,3	14318,2	SNP	1805	915	890	889	457	432	0,5071
4745	A	G	52,7	17442,2	SNP	1998	1111	887	1052	589	463	0,5039
4769	A	G	100	44077,2	SNP	1992	1156	836	1992	1156	836	0,5
5036	A	G	49,2	15805,8	SNP	1997	1199	798	983	558	425	0,5337
5046	G	A	49,5	16020,5	SNP	1995	1197	798	988	570	418	0,5238
5393	T	C	46,8	7561,66	SNP	1054	474	580	493	217	276	0,5097
5655	T	C	48,2	15437,6	SNP	1998	1064	934	964	535	429	0,5226
6548	C	T	50,3	16355,6	SNP	1994	1049	945	1003	536	467	0,5083
6827	T	C	47,4	15104,3	SNP	1999	1088	911	947	517	430	0,5017
6989	A	G	49,1	15835,5	SNP	1997	1168	829	980	556	424	0,5179
7028	C	T	46,5	14738,5	SNP	1999	1173	826	930	551	379	0,5059
7055	A	G	49	15292,8	SNP	1989	1201	788	974	601	373	0,5139
7146	A	G	47,7	15161,3	SNP	1997	1175	822	952	549	403	0,512
7256	C	T	46,5	14634,4	SNP	1996	1121	875	928	504	424	0,5187
7389	T	C	47,4	12716,7	SNP	1694	393	1301	803	201	602	0,525
7521	G	A	50,4	16412,5	SNP	1995	854	1141	1006	420	586	0,5108
7867	C	T	53,9	18005,8	SNP	2000	726	1274	1077	463	614	0,5696
8248	A	G	48,9	15722,5	SNP	1996	1123	873	977	564	413	0,5149
8468	C	T	48,3	8131,4	SNP	1086	186	900	524	80	444	0,5342
8655	C	T	51	8314,23	SNP	1033	229	804	527	119	408	0,5059
8701	A	G	49,5	10204,9	SNP	1298	456	842	643	219	424	0,5118
8860	A	C	3,6	34347,5	SNP	1604	745	859	58	56	2	0,5
8860	A	G	96,4	34347,5	SNP	1604	745	859	1546	689	857	0,5
9540	T	C	43,9	10965,3	SNP	1632	404	1228	716	157	559	0,5395
10398	A	G	48,1	15412,9	SNP	1999	998	1001	961	489	472	0,5096
10688	G	A	45,3	12648,1	SNP	1789	659	1130	810	313	497	0,5192
10810	T	C	47	12504,3	SNP	1684	917	767	792	430	362	0,5016
10873	T	C	39,6	5351,65	SNP	1015	454	561	402	128	274	0,634
11002	A	G	49,6	9651,64	SNP	1231	430	801	610	213	397	0,5001
11719	G	A	48,2	15461,4	SNP	1999	1264	735	964	608	356	0,5017
12061	C	T	48,3	15418,3	SNP	1993	1226	767	962	573	389	0,5204
12519	T	C	50,6	14394,7	SNP	1750	748	1002	886	373	513	0,5066
12705	C	T	49,6	15822,5	SNP	1993	982	1011	989	498	491	0,5108
13046	-	C	10,4	1397,18	INS	1791	957	834	187	178	9	0,9451
13105	A	G	49,2	15920,6	SNP	1998	1004	994	984	468	516	0,5269
13506	C	T	51,8	16494,8	SNP	1940	1082	858	1004	553	451	0,507
13650	C	T	52	7078,03	SNP	833	71	762	433	63	370	0,6463
13680	C	T	52,1	6788,83	SNP	797	35	762	415	28	387	0,6117
13789	T	C	46,9	7666,2	SNP	1065	178	887	499	76	423	0,5276
13880	C	A	46,5	10458,3	SNP	1444	671	773	672	302	370	0,5154
13961	T	C	0	6,88	SNP	1128	542	586	0	0	0	0,5
14155	C	T	0,1	6,79	SNP	922	391	531	1	0	1	0,9888
14178	T	C	48	7164,32	SNP	934	399	535	448	202	246	0,524

Sequenciação completa do mtDNA por MPS e avaliação da deteção de heteroplasmas

14203	A	G	48,3	7112,54	SNP	919	411	508	444	205	239	0,5146
14560	G	A	51,1	4604,41	SNP	568	33	535	290	14	276	0,5487
14622	C	A	3,2	86,34	SNP	657	43	614	21	19	2	0,9926
14766	C	T	46,7	3430,16	SNP	486	25	461	227	7	220	0,6302
14768	A	T	3,2	65,72	SNP	495	24	471	16	1	15	0,5667
14769	A	G	44,2	3274,06	SNP	495	24	471	219	6	213	0,644
14770	C	G	3,1	59,71	SNP	486	24	462	15	1	14	0,5789
14872	C	T	52,1	17136,5	SNP	1996	713	1283	1039	348	691	0,5246
15115	T	C	47,8	15283	SNP	1996	1059	937	955	509	446	0,5024
15326	A	G	100	41725	SNP	1889	650	1239	1889	650	1239	0,5
16126	T	C	52,3	17305,6	SNP	1999	1174	825	1046	598	448	0,516
16187	C	T	41,1	6126,68	SNP	983	188	795	404	14	390	0,8682
16189	T	C	48,8	4961,59	SNP	896	125	771	437	26	411	0,7193
16223	C	T	55	8485,25	SNP	918	158	760	505	92	413	0,5173
16264	C	T	34,8	8444,8	SNP	1718	67	1651	598	17	581	0,581
16270	C	T	33,7	8801,45	SNP	1861	150	1711	628	49	579	0,5088
16278	C	T	33,7	9484,26	SNP	1995	305	1690	673	98	575	0,5143
16293	A	G	30,4	8007,2	SNP	1939	582	1357	590	141	449	0,5773
16311	T	C	28,1	7369,21	SNP	1996	613	1383	560	178	382	0,5125
16400	C	T	24,7	6195,07	SNP	1998	879	1119	494	199	295	0,538
16506	T	C	4	368,43	SNP	1999	1083	916	80	42	38	0,5168
16519	T	C	75,1	28205,3	SNP	1982	1216	766	1489	872	617	0,529

Tabela 12 – Tabela parcial dos resultados do *plug-in Variant Caller* da mistura 90(T4):10(T5).

Position	Ref	Variant	Freq.	Quality	Type	Cov.	Cov+	Cov-	Al. Cov	Al. Cov+	Al. Cov-	S. Bias
73	A	G	11,9	2175,03	SNP	1981	995	986	236	109	127	0,5404
152	T	C	9,7	1603,6	SNP	1997	1193	804	193	147	46	0,6829
182	C	T	9,7	1614,27	SNP	2000	1164	836	194	146	48	0,686
185	G	T	10,1	1704,13	SNP	1998	1170	828	201	142	59	0,6301
195	T	C	10,2	1752,06	SNP	2000	1184	816	205	148	57	0,6415
198	C	T	9,9	1637,99	SNP	1998	1176	822	197	143	54	0,6492
247	G	A	8,1	1216,94	SNP	1996	1168	828	161	104	57	0,564
263	A	G	100	44158,3	SNP	1990	1190	800	1990	1190	800	0,5
310	T	C	7,2	9156,45	SNP	718	4	714	52	0	52	0,5
311	-	C	91,4	9156,45	INS	718	4	714	656	4	652	0,5
357	A	G	7,1	387,56	SNP	771	170	601	55	13	42	0,5225
520	C	A	7,1	332,59	SNP	679	96	583	48	4	44	0,6442
525	C	G	3,1	86,98	SNP	667	88	579	21	0	21	0,9992
709	GT	AC	7,8	341,97	MNP	1989	949	1040	155	81	74	0,5454
750	A	G	100	43825,2	SNP	1981	990	991	1981	990	991	0,5
769	G	A	7,8	1146,18	SNP	1998	1000	998	155	81	74	0,5221
825	T	A	7,9	1172,35	SNP	1988	1000	988	157	73	84	0,538
1018	G	A	9,9	1658,51	SNP	1992	611	1381	197	53	144	0,5459
1438	A	G	90	37202,4	SNP	1997	1021	976	1798	910	888	0,5051
1462	G	A	9,3	1511,27	SNP	1997	1036	961	186	93	93	0,5188

Sequenciação completa do mtDNA por MPS e avaliação da deteção de heteroplasmias

1738	T	C	8,6	1061,18	SNP	1611	460	1151	138	40	98	0,5053
2259	C	T	92,4	38716,5	SNP	1995	925	1070	1844	873	971	0,5098
2352	T	C	7,4	1055,26	SNP	1998	1125	873	147	71	76	0,5797
2706	A	G	9,5	1564,74	SNP	2000	1030	970	190	103	87	0,5272
2758	G	A	9,7	1619,03	SNP	1998	1003	995	194	102	92	0,5238
2768	A	G	9	1430,04	SNP	1998	1031	967	179	88	91	0,5244
2885	T	C	10	1690,97	SNP	1998	904	1094	200	95	105	0,5227
3290	T	C	0	12,57	SNP	1967	932	1035	0	0	0	0,5
3308	T	C	8,4	1308,39	SNP	2000	976	1024	169	84	85	0,509
3594	C	T	8,6	512,28	SNP	766	146	620	66	17	49	0,5957
3666	G	A	8,8	1024,64	SNP	1502	716	786	132	77	55	0,6058
3693	G	A	8,7	1288,38	SNP	1875	964	911	164	98	66	0,5839
4104	A	G	8,6	1158,24	SNP	1750	814	936	150	74	76	0,5282
4745	A	G	91,8	38365,7	SNP	2000	1122	878	1835	1032	803	0,5014
4769	A	G	100	44006,7	SNP	1991	1126	865	1991	1126	865	0,5
5036	A	G	9,2	1482,19	SNP	1997	1211	786	184	102	82	0,5533
5046	G	A	9,1	1431,39	SNP	1958	1180	778	178	101	77	0,5362
5393	T	C	10,3	780,55	SNP	885	352	533	91	30	61	0,5732
5655	T	C	9,1	1453,61	SNP	2000	1082	918	181	94	87	0,5217
6548	C	T	11,1	1974,35	SNP	1992	1053	939	222	116	106	0,5061
6827	T	C	9,6	1571,26	SNP	2000	1138	862	191	107	84	0,5089
6989	A	G	9,7	1604,05	SNP	1998	1149	849	193	103	90	0,5418
7028	C	T	8,6	1332,93	SNP	2000	1189	811	171	110	61	0,5516
7055	A	G	8,1	1174,62	SNP	1974	1219	755	159	93	66	0,534
7146	A	G	9,3	1506,42	SNP	1999	1171	828	186	102	84	0,538
7256	C	T	8,3	1265,71	SNP	1998	1101	897	166	81	85	0,5629
7389	T	C	8,1	993,11	SNP	1647	380	1267	133	36	97	0,5531
7521	G	A	9,6	1593,27	SNP	1993	860	1133	192	101	91	0,5939
7867	C	T	9,9	1622,8	SNP	1943	558	1385	193	66	127	0,5633
8248	A	G	9,4	1508,8	SNP	1997	1117	880	187	113	74	0,5461
8249	G	C	0,9	22,06	SNP	1986	1095	891	17	9	8	0,5221
8468	C	T	9,4	751,42	SNP	965	124	841	91	9	82	0,5732
8655	C	T	9,3	687,27	SNP	901	160	741	84	19	65	0,5751
8701	A	G	8,6	735,8	SNP	1127	398	729	97	35	62	0,5084
8860	A	C	3,6	32228	SNP	1510	650	860	54	49	5	0,5
8860	A	G	96,4	32228	SNP	1510	650	860	1456	601	855	0,5
9531	AC	CA	6,3	98,41	MNP	1347	338	1009	85	3	82	0,9014
9540	T	C	9,7	1180,45	SNP	1505	400	1105	146	25	121	0,6366
10398	A	G	8,3	1285,43	SNP	2000	990	1010	167	88	79	0,5319
10688	G	A	8,6	1086,7	SNP	1648	541	1107	141	39	102	0,561
10810	T	C	8	868,99	SNP	1481	772	709	118	55	63	0,555
10873	T	C	6,9	454,11	SNP	954	441	513	66	21	45	0,6481
11002	A	G	9,6	757,99	SNP	948	269	679	91	34	57	0,6009
11719	G	A	8,9	1405,41	SNP	1999	1225	774	177	119	58	0,5645
12061	C	T	84,7	33576,2	SNP	1988	1194	794	1683	974	709	0,5226
12519	T	C	9,6	1359,68	SNP	1735	775	960	166	79	87	0,5294
12705	C	T	8,7	1357,1	SNP	1990	973	1017	174	76	98	0,5523

Sequenciação completa do mtDNA por MPS e avaliação da deteção de heteroplasmias

13046	-	C	14,1	1886,9	INS	1585	885	700	224	209	15	0,9168
13105	A	G	10,2	1459,1	SNP	1702	917	785	173	82	91	0,5645
13506	C	T	10,5	1496,33	SNP	1670	892	778	175	95	80	0,5088
13650	C	T	10,4	639,48	SNP	729	23	706	76	8	68	0,7831
13680	C	T	91,2	14172	SNP	753	65	688	687	64	623	0,5209
13789	T	C	8,3	602,41	SNP	936	152	784	78	9	69	0,5978
13880	C	A	7,5	710,61	SNP	1351	634	717	101	44	57	0,5339
14178	T	C	10,9	789,52	SNP	824	298	526	90	43	47	0,6175
14203	A	G	10,8	748,76	SNP	798	306	492	86	40	46	0,583
14560	G	A	9,7	433,3	SNP	554	9	545	54	0	54	0,9996
14766	C	T	6	183,67	SNP	499	31	468	30	0	30	0,9994
14769	A	G	5,2	139,37	SNP	481	29	452	25	0	25	0,9992
14872	C	T	90,2	37139,9	SNP	1992	714	1278	1796	641	1155	0,5017
15115	T	C	9,1	1470,45	SNP	1993	1085	908	182	99	83	0,5005
15326	A	G	100	38556	SNP	1754	551	1203	1754	551	1203	0,5
16126	T	C	9,7	1602,09	SNP	2000	1112	888	193	86	107	0,6091
16223	C	T	13,4	978,99	SNP	764	126	638	102	9	93	0,6711
16264	C	T	6,6	858,23	SNP	1910	72	1838	127	5	122	0,5113
16270	C	T	6,7	901,3	SNP	1983	153	1830	133	13	120	0,5644
16278	C	T	6,6	891,36	SNP	1995	277	1718	132	26	106	0,6034
16293	A	G	5,5	647,92	SNP	1938	613	1325	107	24	83	0,6154
16311	T	C	5	557,93	SNP	1995	653	1342	99	31	68	0,5163
16400	C	T	3,4	282,46	SNP	1999	833	1166	67	29	38	0,5165
16519	T	C	21,3	5003,5	SNP	1993	1225	768	425	216	209	0,6068

Tabela 13 – Tabela parcial dos resultados do *plug-in Variant Caller* da mistura 95(T4):5(T5).

Position	Ref	Variant	Freq.	Quality	Type	Cov.	Cov+	Cov-	Al. Cov	Al. Cov+	Al. Cov-	S. Bias
73	A	G	6,2	783,24	SNP	1982	1001	981	122	49	73	0,6032
152	T	C	5	557,42	SNP	1997	1165	832	99	65	34	0,5772
182	C	T	4,8	528,31	SNP	1998	1157	841	96	63	33	0,5812
185	G	T	5	566,65	SNP	1999	1192	807	100	69	31	0,6011
195	T	C	5	556,92	SNP	2000	1190	810	99	63	36	0,5436
198	C	T	4,8	527,35	SNP	1995	1183	812	96	63	33	0,5672
247	G	A	4,1	403,56	SNP	1997	1126	871	82	54	28	0,5987
263	A	G	100	43977,6	SNP	1985	1148	837	1985	1148	837	0,5
316	G	C	0,9	8204,09	SNP	867	11	856	8	0	8	0,5
311	-	C	99,1	8204,09	INS	867	11	856	859	11	848	0,5
357	A	G	3,3	140,11	SNP	965	198	767	32	12	20	0,6991
520	C	A	8,6	596,79	SNP	887	130	757	76	20	56	0,6753
709	GT	AC	4,2	54,74	MNP	1990	910	1080	83	46	37	0,596
750	A	G	100	43940,3	SNP	1987	956	1031	1987	956	1031	0,5
769	G	A	4,3	447,52	SNP	2000	1013	987	87	53	34	0,603
825	T	A	4,6	493,73	SNP	1991	1009	982	92	40	52	0,5719
1018	G	A	4,3	441,41	SNP	1994	686	1308	86	30	56	0,5053
1438	A	G	94,7	40295,2	SNP	2000	993	1007	1893	936	957	0,502

Sequenciação completa do mtDNA por MPS e avaliação da deteção de heteroplasmias

1462	G	A	5	561,43	SNP	1998	1057	941	100	58	42	0,5514
1738	T	C	4,4	437,63	SNP	1950	625	1325	85	24	61	0,5452
2259	C	T	94,8	40329	SNP	1996	977	1019	1893	929	964	0,5013
2352	T	C	4,5	470,09	SNP	1994	1113	881	90	47	43	0,5361
2706	A	G	4,5	474,59	SNP	1998	1084	914	90	48	42	0,5093
2758	G	A	4,9	532,07	SNP	1997	1063	934	97	54	43	0,5246
2768	A	G	4,2	430,02	SNP	2000	1094	906	85	51	34	0,554
2885	T	C	5,1	578,39	SNP	1996	930	1066	101	57	44	0,5976
3168	C	T	4	287,32	SNP	1963	1025	938	78	72	6	0,9164
3290	T	C	1,4	41,18	SNP	1998	938	1060	28	2	26	0,9197
3308	T	C	4	393,69	SNP	2000	967	1033	81	38	43	0,5144
3594	C	T	3,9	167,02	SNP	862	187	675	34	8	26	0,5262
3666	G	A	4,4	408,36	SNP	1789	901	888	79	41	38	0,5154
3693	G	A	4,3	435,36	SNP	1996	1060	936	86	42	44	0,5426
4104	A	G	3,9	370,34	SNP	1998	1006	992	78	44	34	0,5606
4745	A	G	95,7	41029,3	SNP	1999	1187	812	1913	1147	766	0,506
4769	A	G	100	44242	SNP	1995	1175	820	1995	1175	820	0,5
5036	A	G	5,5	656,46	SNP	1997	1224	773	109	64	45	0,5268
5046	G	A	5,4	645,24	SNP	1992	1212	780	108	59	49	0,5634
5393	T	C	5,5	378,92	SNP	1164	482	682	64	24	40	0,5408
5655	T	C	5,4	633,73	SNP	1998	1054	944	107	65	42	0,5809
6548	C	T	6,5	870,06	SNP	1995	1019	976	130	65	65	0,5108
6827	T	C	3,6	313,81	SNP	1997	1114	883	71	39	32	0,5086
6989	A	G	5,3	614,42	SNP	1999	1153	846	105	54	51	0,5628
7028	C	T	5,3	623,98	SNP	2000	1186	814	106	48	58	0,6377
7055	A	G	4,8	523,14	SNP	1982	1177	805	95	40	55	0,6678
7146	A	G	5,7	688,36	SNP	1999	1217	782	113	66	47	0,5257
7256	C	T	4,3	431,39	SNP	1999	1155	844	86	52	34	0,5278
7389	T	C	4,7	510,49	SNP	2000	491	1509	94	33	61	0,6244
7521	G	A	4,1	404,37	SNP	1996	845	1151	82	30	52	0,5599
7867	C	T	5,4	643,58	SNP	1999	572	1427	108	41	67	0,6042
8248	A	G	4,3	448,11	SNP	1991	1068	923	86	61	25	0,6783
8249	G	C	0,8	18,96	SNP	1994	1069	925	15	6	9	0,634
8468	C	T	5,3	353,26	SNP	1155	196	959	61	7	54	0,6119
8655	C	T	6,1	427,96	SNP	1121	234	887	68	23	45	0,6595
8701	A	G	5,9	502,1	SNP	1364	467	897	81	32	49	0,5564
8860	A	C	3,6	38620,2	SNP	1799	774	1025	65	61	4	0,5
8860	A	G	96,4	38620,2	SNP	1799	774	1025	1734	713	1021	0,5
9540	T	C	4,9	500,15	SNP	1852	525	1327	90	17	73	0,6294
10398	A	G	5,6	673,31	SNP	1999	950	1049	111	58	53	0,5472
10688	G	A	4,2	413,01	SNP	1999	752	1247	83	34	49	0,535
10810	T	C	4,6	460,63	SNP	1856	1069	787	86	49	37	0,5063
11002	A	G	4	224,16	SNP	1150	402	748	46	20	26	0,5887
11719	G	A	5	557,34	SNP	1999	1221	778	99	61	38	0,5056
12061	C	T	89,8	36756,1	SNP	1987	1246	741	1785	1081	704	0,5227
12519	T	C	5,3	614,65	SNP	1998	955	1043	105	56	49	0,5552
12705	C	T	4,9	535,37	SNP	1984	1089	895	97	54	43	0,5079

Sequenciação completa do mtDNA por MPS e avaliação da deteção de heteroplasmias

13046	-	C	15	2616,99	INS	1945	1070	875	292	275	17	0,9297
13105	A	G	5,2	614,23	SNP	2000	1077	923	105	53	52	0,5338
13506	C	T	3,9	371,9	SNP	1992	1070	922	78	39	39	0,5371
13650	C	T	5	257,8	SNP	898	20	878	45	4	41	0,8107
13680	C	T	95	18744,1	SNP	926	64	862	880	62	818	0,5052
13789	T	C	4,9	323,55	SNP	1215	238	977	59	13	46	0,5371
13880	C	A	4,1	330,58	SNP	1632	767	865	67	34	33	0,5374
14178	T	C	4,4	222,98	SNP	949	362	587	42	21	21	0,6185
14203	A	G	4,8	246,95	SNP	920	376	544	44	23	21	0,613
14560	G	A	4,6	146,81	SNP	613	44	569	28	1	27	0,676
14766	C	T	4,7	139,38	SNP	555	41	514	26	1	25	0,6659
14769	A	G	4,1	106,8	SNP	539	38	501	22	2	20	0,5686
14872	C	T	95,3	40709	SNP	1998	691	1307	1905	653	1252	0,5034
15115	T	C	4,4	448,06	SNP	1997	1129	868	87	44	43	0,5597
15326	A	G	100	43973,7	SNP	1985	711	1274	1985	711	1274	0,5
16126	T	C	5,7	693,36	SNP	1998	1150	848	113	71	42	0,5549
16187	C	T	4,1	227,26	SNP	1086	289	797	45	2	43	0,8861
16223	C	T	6,4	403,36	SNP	934	170	764	60	10	50	0,5266
16519	T	C	13,3	2569,31	SNP	1993	1303	690	266	147	119	0,6045

Tabela 14 – Tabela parcial dos resultados do *plug-in Variant Caller* da mistura 99(T4):1(T5).

Position	Ref	Variant	Freq.	Quality	Type	Cov.	Cov+	Cov-	Al. Cov	Al. Cov+	Al. Cov-	S. Bias
263	A	G	100	44224	SNP	1993	1058	935	1993	1058	935	0,5
310	T	C	8,2	18227,1	SNP	1390	4	1386	114	0	114	0,5
311	-	C	91,4	18227,1	INS	1390	4	1386	1271	4	1267	0,5
520	C	A	9,8	965,4	SNP	1217	158	1059	119	5	114	0,7728
525	C	G	4,6	291,98	SNP	1190	149	1041	55	1	54	0,8852
750	A	G	100	43766	SNP	1982	961	1021	1982	961	1021	0,5
1438	A	G	98,7	43281,5	SNP	2000	1034	966	1973	1019	954	0,5005
2259	C	T	98,4	43021,4	SNP	1999	891	1108	1967	878	1089	0,5007
3290	T	C	1,2	32,62	SNP	1997	820	1177	24	1	23	0,9408
4745	A	G	100	43651,6	SNP	1982	1169	813	1982	1169	813	0,5
4769	A	G	100	44450,2	SNP	1998	1198	800	1998	1198	800	0,5
7816	C	T	0	6,26	SNP	1991	948	1043	0	0	0	0,5
8249	G	C	0,7	15,22	SNP	1989	1045	944	14	7	7	0,5254
8860	A	C	3,3	43163,9	SNP	1996	816	1180	66	64	2	0,5
8860	A	G	96,7	43163,9	SNP	1996	816	1180	1930	752	1178	0,5
12061	C	T	93,5	39059,9	SNP	1982	1172	810	1853	1052	801	0,5242
13046	-	C	12,2	1928,83	INS	1943	1060	883	237	225	12	0,9398
13680	C	T	98,8	27814	SNP	1302	61	1241	1287	57	1230	0,5147
14153	T	C	4	136,67	SNP	1328	392	936	53	47	6	0,9491
14872	C	T	98	42618,7	SNP	1999	669	1330	1959	658	1301	0,5014
15326	A	G	100	43978,5	SNP	1989	592	1397	1989	592	1397	0,5
16519	T	C	3,3	267,49	SNP	1997	1323	674	65	35	30	0,6272

II. Cobertura das 7 amostras com mistura

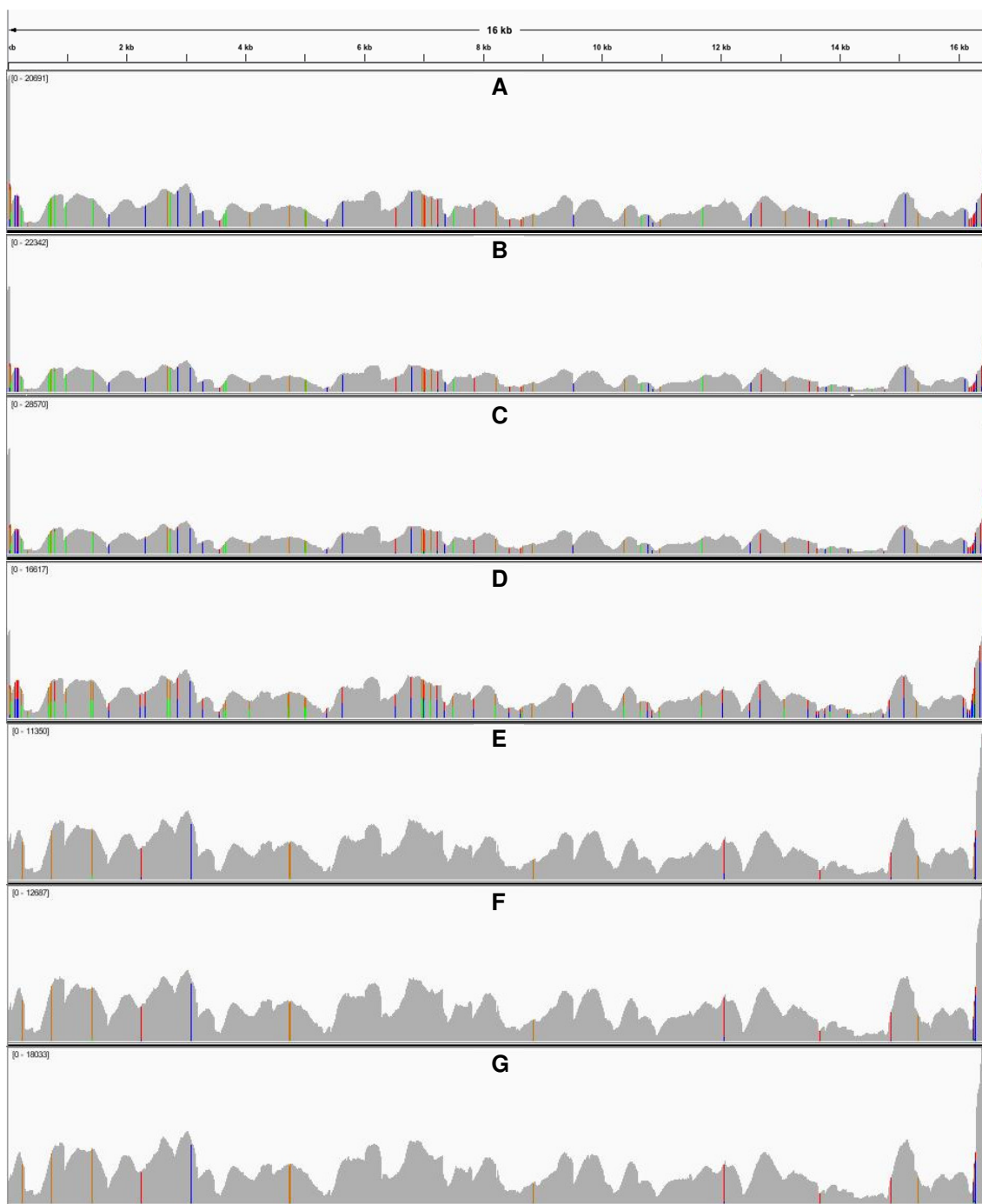


Figura 27 - Gráficos da cobertura (eixo das ordenadas) de todas as posições dos nucleótidos ao longo da molécula de mtDNA (eixo das abcissas) nas 7 amostras com mistura visualizadas no *software* IGV: **A)** 1(T4):99(T5); **B)** 5(T4):95(T5); **C)** 10(T4):90(T5); **D)** 50(T4):50(T5); **E)** 90(T4):10(T5); **F)** 95(T4):5(T5); **G)** 99(T4):1(T5).

III. Ata de congresso da apresentação oral no “26th Congress of the International Society for Forensic Genetics” realizado em Cracóvia, Polónia de 31 de agosto a 5 de setembro de 2015



Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/FSIGSS



Evaluation of heteroplasmy detection in the Ion Torrent PGM

S. Magalhães^{a,b,c}, S.L. Marques^{a,b}, C. Alves^{a,b}, A. Amorim^{a,b,c}, L. Alvarez^{a,b}, A. Goios^{a,b,*}

^aInstituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal

^bInstitute of Molecular Pathology and Immunology of the University of Porto (Ipatimup), Porto, Portugal

^cFaculdade de Ciências da Universidade do Porto, Porto, Portugal

ARTICLE INFO

Article history:

Received 26 August 2015

Accepted 7 September 2015

Available online xxx

Keywords:

mtDNA

Heteroplasmy

MPS

PGM

Long-range PCR

ABSTRACT

The use of mitochondrial DNA (mtDNA) heteroplasmy in forensic genetics casework and databasing has been limited due to the technical limitations of the PCR/Sanger sequencing technology. Massively Parallel Sequencing (MPS) has been suggested to be more sensitive and accurate in detecting and evaluating heteroplasmy levels than Sanger sequencing, particularly when contamination is controlled and the amplification strategy is carefully designed.

In this work, we present the results of a sensitivity study of heteroplasmy detection and quantification with Ion Torrent PGM, using control samples of predetermined haplotypes belonging to different haplogroups. The amplification strategy used for the PGM sequencing is based on a single long-range PCR, in order to even the proportions of the mixed bases throughout the molecule. To simulate different levels of heteroplasmy we used mixtures of the amplified control samples (amplicons) at variable proportions.

The PGM sequencing strategy accurately detected all artificial heteroplasmy levels $\geq 5\%$, with the exception of variants at repetitive regions. At the 1% level, some of the mixed bases were detected although the false positive background noise does not allow for a trustworthy variant call.

1. Introduction

The use of mitochondrial DNA (mtDNA) heteroplasmy in forensic genetics has been limited, however recent guidelines illustrate the importance of thoroughly detecting heteroplasmy both in casework and databasing [1].

Sanger sequencing technology is not sensitive enough to detect low-level heteroplasmy ($\leq 10\text{--}20\%$) and it does not allow for accurate allele quantification. It has been demonstrated that generally Massively Parallel Sequencing (MPS) methods have higher sensitivity to detect low-level heteroplasmy, and depending on the amplification strategy it has the additional potential to quantify the heteroplasmy level.

Presently, several MPS platforms are available, with different chemistries that should be evaluated and compared when it comes to heteroplasmy detection. The Illumina GA was demonstrated to detect all $\geq 5\%$ heteroplasmy with virtually no false positives [2]. We here present the results of a sensitivity study of heteroplasmy detection and quantification with Ion Torrent

PGM, LifeTechnologies (LT), using control samples of predetermined haplotypes belonging to different haplogroups.

2. Material and methods

DNA samples of two healthy donors, with previously determined complete mtDNA haplotypes (belonging to haplogroups H13a1a (T4) and L1b1a12a (T5)), were extracted from 1 mL saliva samples through a user-developed protocol using DNeasy[®] Blood & Tissue Kit (Qiagen). In both samples, the complete mtDNA molecule was amplified in a single long-range PCR using the home-designed primers MT_16555L (5'-TAAGACATCAGGATGGAT-CACAGGT-3') and MT_16554H (5'-TTTAAGGGGAACGTGTGGC-TATTT-3'), with the enzyme mix Ranger DNA Polymerase (Bioline) following the manufacturer's instructions. Amplicons were quantified with Qubit 2.0 Fluorometer (Invitrogen Corporation) and normalized to the same concentration. Mixtures of the two amplicons were performed at different proportions: 1:99, 5:95, 10:90, 50:50, 90:10, 95:5, 99:1. Each mixture was further processed as an independent sample.

The 200 bp sequencing chemistry was used. Enzymatic shearing and Ion Xpress (LT) adapter ligation were carried out in the AB Library Builder (LT) according to the manufacturer's instructions. After Qubit quantification, the 7 different barcoded samples, corresponding to different mixtures, were pooled

* Corresponding author at: Institute of Molecular Pathology and Immunology of the University of Porto (Ipatimup), Porto, Portugal. Tel.: +351 225570700; fax: +351 225570799.

E-mail address: aalmeida@ipatimup.pt (A. Goios).

<http://dx.doi.org/10.1016/j.fsigs.2015.09.006>

1875-1768/© 2015 Elsevier Ireland Ltd. All rights reserved.

together at equimolar concentrations, and the library was size-selected (fragments ~260 bp) using the E-Gel SizeSelect Agarose Gel (Invitrogen Corporation), following manufacturer's recommendations. The library was further subjected to 8 amplification cycles using the Platinum PCR SuperMix High Fidelity.

Template was prepared using the Ion One Touch (LT) and enriched in positive Ion Sphere Particles with the Ion One Touch ES system (LT), all according to manufacturer's recommendations. MPS was performed with the Ion Torrent PGM (LT) after the cleaning protocol, as recommended by the manufacturer.

All sequences were analysed with the Ion Torrent Software Suite using the plug-in Variant Caller V.4.4.3.3 (VC), based on the Generic – PGM – Somatic – Low Stringency parameters, producing a table with the list of differences to the rCRS (revised Cambridge Reference Sequence).

3. Results and discussion

The long-range amplification strategy proved to be efficient and specific when applied to reference samples. The primers were designed to anneal in a very conserved segment between HV1 and HV2, so that it is unlikely that a SNP would occur. Should there be the need to type these positions, these would need to be targeted in an independent PCR. When comparing the coverage map obtained from the long-range amplicons with those obtained from an alternative amplification of the mtDNA molecule in two ~8 kb fragments [3] there were no significant differences observed, presenting generally an uneven coverage profile typical of the PGM sequencing of mtDNA genomes. However, when there is interest in quantifying heteroplasmy/mixture levels throughout the molecule, it is useful to work with a single amplicon so that there are no differences in PCR efficiency in the different regions.

A 316 chip with 7 samples corresponding to 7 different mixtures was run, resulting in a total of 1950575 reads with 96.7% aligned to the reference genome, and an average coverage depth of 15835.5x (predicted error rate of 1%). When analyzing the coverage variant sites reported by VC, there are positions that recurrently show lower coverage, which mostly (but not exclusively) are due to

strand bias. This is the case of the region near the 310 heteroplasmic stretch and around positions 14560–14770, both showing exclusively sequences in the minus sense.

The VC SNP detection threshold was set to 0.1% to test the detections in the 1:99 mixtures. Erroneous variant calls are found in all mixture proportions (7–12 per sample) and recurrently in the same positions, most of which are near true variant sites or homopolymers. Typical positions/substitutions presenting these results are 71, 310, 520, 525, 8860C, 13046, 14768, 14770 and 16506. These positions are mostly found at very low frequencies (<5%), however when analyzing low-level heteroplasmies (<5%), they result in false positives.

For an easier visualization, a graphical display of the results is presented (Fig. 1), where all the variant positions are represented as differences to one of the samples (T5). Variants with known mixtures not called by the software were assigned a value of 0% or 100%, depending on whether they are equal or different from the rCRS. All artificial heteroplasmies were detected down to the 5 (T4):95(T5) mixture ratio, with the exception of indels at homopolymeric or dinucleotide repetitive regions. The method was not sensitive enough to detect the minor component of the 1 (T4):99(T5) or 99(T4):1(T5) mixtures. These two symmetric mixtures do not show exactly the same behavior. In fact, we observe that, the variants present in sample T4 (haplogroup H13a1a) are not detected in the 1(T4):99(T5) mixture and are under-represented in the 99(T4):1(T5) mixture, and the variants characteristic of sample T5 (L1b1a12a) present an inverse behavior in the same mixtures. This effect is visible also to a lesser extent in some control region variant positions that are not called in the 95 (T4):5(T5) mixture ratio. This suggests that the algorithm employed by the VC plugin appears to be biased towards the reference sequence used (rCRS). This is probably related with one or more default settings defined in the Generic – PGM – Somatic – Low Stringency parameters, since the plugin does not seem to detect variants at frequencies lower than 1% although the threshold was set to 0.1%.

Although some regions present an overestimation of the heteroplasmic frequencies, visible in particular in the region of HVR1 (Fig. 1), generally the observed frequencies were consistent

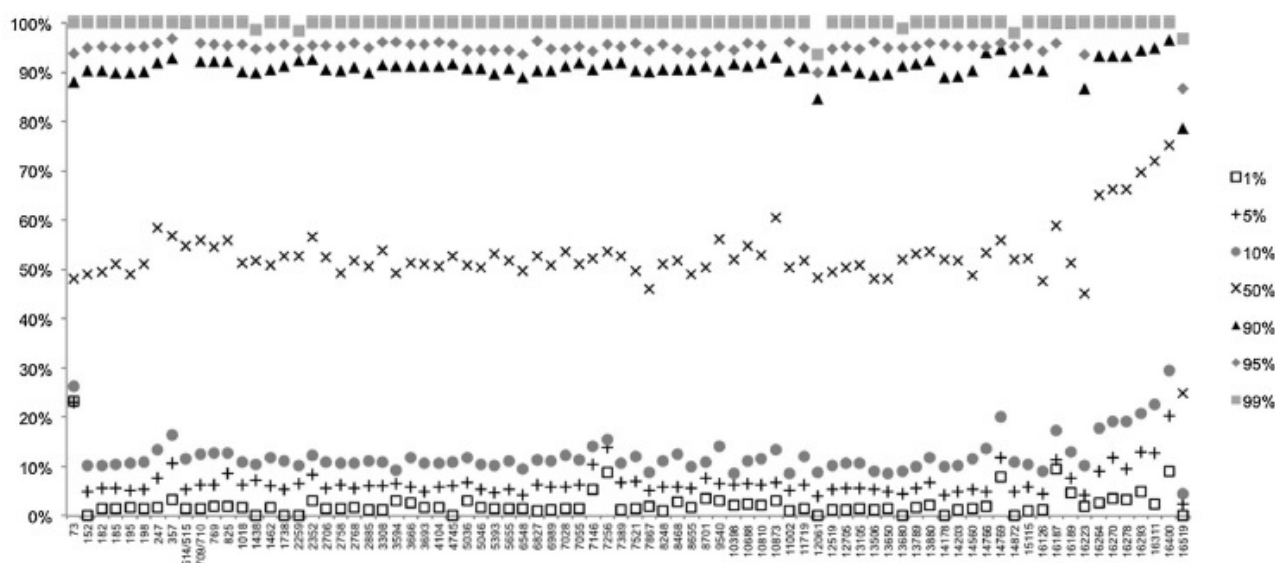


Fig. 1. Detected heteroplasmy levels for the expected substitutions present in the artificial heteroplasmic samples, reported as differences to sample T5. The different symbols represent the relative proportion of sample T4 in the mixture.

with the expected values for each mixture ratio ($y = 0.9897x + 0.0207$; $R^2 = 0.9998$) revealing a high accuracy at detecting heteroplasmy levels.

4. Conclusions

The Ion Torrent PGM MPS technology presents limitations in that (1) the coverage obtained throughout the mtDNA molecule is variable with some regions presenting extreme strand bias, and (2) false positives are mostly generated by alignment problems in the analysis algorithms. However, it proved to be sensitive and accurate at detecting and quantifying mixture/heteroplasmy fractions >5% in artificial heteroplasmic samples, when compared to Sanger sequencing. Note that in true mixtures/heteroplasmies the PCR reactions will alter the initial proportions of each component. By using a single long-range PCR this effect is minimized, and equalized throughout the mtDNA molecule.

Conflict of interest

None.

Acknowledgements

IPATIMUP integrates the i3S Research Unit, which is partially supported by FCT, (Portuguese Foundation for Science and Technology). This work is funded by FEDER funds through COMPETE (Operational Programme for Competitiveness Factors) and National Funds through the FCT, under projects PEst-C/SAU/LA0003/2013, PTDC/ATP-DEM/4545/2012, NORTE-07-0162-FEDER-00018 and NORTE-07-0162-FEDER-000067. LA and AG are supported by FCT fellowships SFRH/BPD/65000/2009, SFRH/BPD/43646/2008, respectively.

References

- [1] W. Parson, L. Gusmão, D.R. Harese, 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.
- [2] S. Tang, TH, uang, Characterization of mitochondrial DNA heteroplasmy using a parallel sequencing system, *Biotechniques* 48 (2010) 287–296.
- [3] LB, M.D. Zimmermann, W.P. aniaux, Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences, *BMC Genomics* 10 (2009) 193.