

II. Material e Métodos.

A área de Biologia e / ou Genética Forense tem vindo a desenvolver esforços no sentido de definir os procedimentos mais adequados para manipulação laboratorial e análise do ADN mt proveniente de vários tipos de amostras (Tully, G, *et al*, 2001; Carracedo, A, *et al*, 2000). Simultaneamente pretende-se verificar a reprodutibilidade da utilização de diferentes métodos, por diferentes laboratórios, na análise deste tipo de genoma (Prieto, L; *et al*, 2003; Alonso, A, *et al*, 2002; Carracedo, A, *et al*, 1998).

Foram propostas algumas directrizes a seguir em relação às práticas laboratoriais mais adequadas, nomenclatura a adoptar e ainda critérios de interpretação de resultados a adoptar para o estudo deste tipo de genoma (Tully, G, *et al*, 2001; Carracedo, A, *et al*, 2000).

Das directrizes respeitantes às práticas laboratoriais, definidas com o intuito de minimizar a ocorrência de contaminações e assegurar que uma contaminação não leva à apresentação de resultados erróneos, salientam-se as enumeradas a seguir:

- Áreas de pré e pós – PCR separadas fisicamente.
- Utilizar, pelo menos em todas as manipulações de amostra / reagentes na fase pré - PCR, pipetas com pontas de pistão ou com pontas descartáveis de filtros resistentes a aerossóis.
- Na medida do possível, assegurar que todo o material plástico a ser usado no pré – PCR está livre de qualquer tipo de ADN contaminante, de preferência por exposição prévia a luz UV (cerca de 20 – 30 min.).
- As soluções usadas no pré – PCR devem ser expostas à luz UV e / ou autoclavadas, para eliminar qualquer ADN contaminante, sempre que tal procedimento seja possível.
- Devem ser incluídos controlos negativos (brancos) em todos os processos de extracção e amplificação por PCR. Estes controlos negativos deverão ser também sujeitos a sequenciação, e sempre que a sequência presente num destes brancos seja igual à de uma amostra o lote de amplificação onde esta se incluía deve ser rejeitado e repetida a sua análise.

- Deve ser introduzido um controlo positivo na fase da amplificação, que também deverá ser sujeito a sequenciação.
- Sempre que a quantidade de amostra seja suficiente e/ou haja dúvidas quanto ao resultado da análise, a amostra deverá ser extraída e analisada em duplicado e, se possível, por técnicos diferentes.
- A sequenciação de determinada amostra deve ser realizada obrigatoriamente em ambos os sentidos, directo e reverso, sempre que haja suspeitas de heteroplasmias, quer estas sejam pontuais ou de comprimento.
- Devem ser determinadas as sequências de ADN mt, HVRI e HVRII, de todo o pessoal que trabalhe no laboratório, para despiste de contaminações das amostras por quem as manipulou.

Relativamente aos critérios de nomenclatura definidos, e adoptados durante a realização deste trabalho, estes são referidos no ponto **10** deste capítulo.

Os critérios de interpretação propostos aplicam-se a estudos de natureza forense, como a identificação de vestígios biológicos pertencentes de determinado dador ou de um ascendente materno. Quando se compara a sequência do ADN mt de duas ou mais amostras há características específicas a ter em conta, como por exemplo:

- O ADNmt possui uma taxa de mutação elevada;
- A ocorrência de heteroplasmias é comum, especialmente em determinados tecidos (ex: cabelos);
- Existem posições nucleotídicas bastante estáveis e outras muito mutáveis;
- As transições são muito mais frequentes do que as transversões, numa proporção de 40:1;
- Inserções e deleções em regiões homopoliméricas são comuns.

1. COLHEITA DAS AMOSTRAS

As amostras estudadas neste trabalho foram colhidas para contribuírem, conjuntamente com outras amostras provenientes de outros Países de Língua Oficial Portuguesa, para vários estudos populacionais, como já foi referido (Corte-Real, 1999).

Após consentimento, obtiveram-se amostras sanguíneas em pessoas naturais dos países estudados neste trabalho da seguinte forma:

- em **Cabo Verde** a colheita foi coordenada pelo Director do Hospital da Praia e pela Directora do respectivo Laboratório e incluiu amostras de naturais ou descendentes de naturais das diferentes ilhas do arquipélago (Santiago, Fogo, S.^{to} Antão, S. Vicente, S. Nicolau, Boavista, Sal, Brava e Maio);

- na **Guiné-Bissau** um elemento do Instituto de Medicina Legal de Coimbra deslocou-se à Maternidade, ao Laboratório e ao Centro de Medicina Tropical do Hospital Nacional Simão Mendes de Bissau, tendo um colaborador realizado colheitas em povoações do interior do País.

As amostras sanguíneas foram colhidas nos centros hospitalares, de forma aleatória, tendo sido eliminados os familiares de pessoas já incluídas na amostragem. Foram feitas manchas sanguíneas em retalhos de tecido de algodão e ou em tiras de papel, deixadas secar à temperatura ambiente e colocadas individualmente em envelopes de papel (Corte-Real, 1999).

Enviou-se do Instituto de Medicina Legal de Coimbra, actual Delegação do Centro do Instituto Nacional de Medicina Legal, juntamente com as instruções de colheita e dos retalhos de tecido de algodão para colocação e secagem de cerca de um mililitro (mL) de sangue periférico, as fichas – protocolo (ver Anexo I) a preencher em cada uma das colheitas, a fim de registar a proveniência de cada amostra e os dados dos ascendentes mais próximos, para um conhecimento mais aprofundado da amostragem.

A origem de cada amostra estudada foi inferida com base nas fichas-protocolo e analisando a linhagem materna de cada amostra. Quando discordantes, foram considerados sempre os dados relativos à ascendente materna mais distante (mãe ou avó), uma vez que, como referido, o ADN mitocondrial apresenta hereditariedade exclusivamente materna.

Infelizmente não foi possível uma recolha de amostras de todas as ilhas do arquipélago de Cabo Verde e de algumas etnias Guineenses, verificando-se também que algumas ilhas (Cabo Verde) e etnias (Guiné-Bissau) possuem uma amostragem pouco representativa.

As amostras da República de **Cabo Verde** (77) sujeitas ao estudo de polimorfismos do ADN mitocondrial são provenientes de diferentes ilhas do arquipélago e pertencentes a diferentes grupos, Negros (N), Mestiços (M) e Brancos (B):

Tabela II.1- Catalogação da amostragem proveniente de Cabo Verde consoante a sua ascendência materna, local onde viviam / nasceram e respectiva raça.

Ilhas	N.º Amostras / raça	
<i>Ilhas de Sotavento:</i>		
Santiago	45	20 N, 18 M, 7 B.
Fogo	14	4 N, 5 M, 5 B.
Brava	4	1 N, 2 M, 1 B.
Maio	2	1 N, 1 B.
Total parcial	65	
<i>Ilhas do Barlavento:</i>		
Santo Antão	4	1 N, 2 M, 1 B.
São Vicente	2	2 B.
São Nicolau	2	1 N, 1B.
Boavista	0	
Sal	0	
Total parcial	8	
Outra (*)	4	
Total	77	

Legenda: N – Negra, M – Mestiça, B – Branca.

(*) **Nota:** A ascendência de 4 amostras não é explícita, embora esteja relacionada com este arquipélago.

Das amostras colhidas em Cabo Verde 3 não foram consideradas uma vez que a ascendência de 2 amostras é desconhecida e a da terceira está relacionada com o continente Africano, nomeadamente, Angola.

Relativamente à amostragem da República da **Guiné-Bissau** (79) verifica-se que apresenta a seguinte proveniência:

Tabela II.2 – Catalogação da amostragem proveniente da Guiné-Bissau consoante a etnia da sua ascendência materna, indicada nas respectivas fichas.

Etnia	N.º Amostras
Papel	18
Manjaco	15
Balanta	13
Mancanha	9
Fula	7
Beafada	5
Bijagós	2
Mandinga	2
Nalu	2
Cobiana	1
Felupe	1
Mansoanca	1
Saracule	1
Sussu	1
Cetua	0
Baiote	0
Banhento ou Banhune	0
Cassanga	0
Total parcial	78
Não sabe / Não refere (*)	1
Total	79

(*) **Nota:** Da amostragem colhida na Guiné-Bissau 1 amostra não possui informação quanto à etnia dos seus ascendentes maternos.

2. EXTRACÇÃO DE ADN

A extracção de ADN foi realizada recorrendo ao uso de Chelex® 100, o que se revelou um método bastante eficiente, rápido, com pouca manipulação da amostra, pois é constituído por poucos passos, diminuindo assim a probabilidade de contaminação da amostra, e permite a obtenção de ADN em qualidade suficiente para amplificação (Walsh P S, *et al*, 1991). No entanto, o ADN obtido a partir deste método, por ser desnaturado, não pode ser utilizado para análise por RFLP (Singer-Sam J, *et al*, 1989).

A basicidade da suspensão de Chelex a 5% (pH10-11) e a exposição a temperaturas próximas dos 100°C resulta na ruptura de membranas e desnaturação do ADN. O ADN libertado é posteriormente amplificado por PCR (Polymerase Chain Reaction), um método de amplificação de pequenas quantidades de sequências de ADN alvo relativamente curtas, que não requer ADN nativo de elevado peso molecular, só a sequência alvo necessita de estar intacta, sendo possível amplificar ADN parcialmente degradado ou desnaturado.

A fervura do ADN revela-se muito útil quando o objectivo é libertar o ADN de uma pequena quantidade de células, sendo que o papel do Chelex na preparação do ADN para a PCR. Walsh P S, *et al*, em 1991 verificaram que a fervura do ADN na ausência de Chelex o torna impróprio (inactivo) para amplificação, o que está de acordo com o papel protector do Chelex referido por Singer-Sam J, *et al*, que em 1989 afirmam que a resina de Chelex® 100 poderia ser usada como um meio de aumentar o sinal do produto da amplificação por PCR de pequenas quantidades de ADN libertadas de tecido.

De facto, o ADN obtido a partir de manchas de sangue pode conter inibidores, verificando-se que as extracções obtidas a partir de Chelex não possuem menos inibidores da reacção de PCR do que extractos obtidos por extracções de phenol-clorofórmio e proteinase K. Trabalhos anteriores (Higuchi R, 1989) sugerem que os compostos de porfirina do sangue são a causa da inibição do PCR. Devido a não ser necessária a utilização de proteinase k, que libertaria o grupo heme da hemoglobina, é possível que nas extracções por Chelex possuam

menos compostos porfirinicos. Sendo, também possível que estes inibidores se liguem à matriz da própria resina. Uma vez que o Chelex® 100 é uma resina quelante com elevada afinidade para iões metálicos polivalentes, a sua presença prevenirá a degradação do ADN durante a fervura, pois quelata os iões metálicos, como o ferro e o magnésio, que possam catalisar a degradação do ADN a elevadas temperaturas em soluções de baixa força iónica.

As propriedades quelantes desta resina podem levar a uma inibição da reacção de PCR, pois esta requer Mg^{2+} , contudo, o Chelex® 100 é fácil de remover para que não interfira com as amplificações subsequentes.

A utilização deste método reduz o número de passos, nos quais há transferência de amostra, necessários para a preparação das amostras havendo, consequentemente, uma menor manipulação das mesmas o que reduz as possibilidades de uma contaminação inadvertida ou de troca de amostras.

A extracção foi realizada de acordo com o protocolo estabelecido por Walsh P S, *et al*, (1991) com ligeiras adaptações, tendo em consideração o suporte físico das manchas de sangue e que estas se encontravam bastante secas (Anexo II).

3. PURIFICAÇÃO DO ADN EXTRAÍDO

Em amostras problemáticas, para as quais não se obteve uma boa amplificação a partir do ADN extraído, foi necessário proceder a esta purificação de forma a melhorar a pureza do produto da extracção e eliminar possíveis inibidores da reacção de PCR.

Este método consiste numa centrifugação diferencial na presença de uma solução de fenol clorofórmio, "*Clorofane*", do ADN extraído e posterior precipitação etanólica das moléculas de ADN.

Numa primeira fase centrifuga-se o produto da extracção misturado com a solução de fenol clorofórmio o que origina a formação de três fases distintas. Em seguida recolhe-se o sobrenadante para outro tubo autoclavado e devidamente identificado, deve-se ter cuidado para não tocar a fase proteica que contem inibidores da reacção de PCR. O precipitado é rejeitado.

Incuba-se então o ADN com etanol absoluto frio e centrifuga-se. As moléculas do ADN precipitam podendo ser observado o pellet no fundo do tubo. Retira-se cuidadosamente o sobrenadante com uma pipeta e deixa-se secar o pellet ao ar.

Ressuspende-se o pellet em 25µl de água desionizada e autoclavada, o que corresponderá não só a uma solução de ADN mais pura, mas, também mais concentrada do que a obtida após a extracção.

O protocolo desta purificação foi adaptado a partir de um protocolo de extracção orgânica de ADN publicado pelo FBI na Internet (<http://www.fbi.gov/>), próprio para situações de âmbito forense com amostras consideradas limite em que a qualidade do produto de extracção é essencial (Anexo II).

4. AMPLIFICAÇÃO DAS REGIÕES HIPERVARIÁVEIS I E II

A PCR (*Polymerase Chain Reaction*) é um método simples de amplificação de pequenas sequências alvo de ADN, usando “*primers*” específicos e a “*TaqDNAPolimerase*”, uma enzima termoestável. A amplificação é selectiva, as sequências dos extremos da região a amplificar (regiões de flanco) são conhecidas e os “*primers*”, oligonucleótidos complementares dessas sequências, ao hibridizarem com o ADN vão delimitar a região a amplificar pela ADN Polimerase. Esta vai adicionando os desoxirribonucleotídeos trifosfatados copiando, assim, o ADN molde, o produto desta reacção consiste num fragmento de ADN de cadeia dupla cujas extremidades correspondem às extremidades 5’ dos “*primers*” utilizados.

A reacção de PCR é composta por três fases, *desnaturação – hibridação – extensão* ou *síntese*, que se repetem por 30 a 40 ciclos. As diferentes temperaturas de cada fase são o “motor” de toda a amplificação, é essencial precisar estas temperaturas. Para controlar a reacção são utilizados termocicladores automáticos que possuem a capacidade de fazer variar a temperatura dos tubos onde ela ocorre num curto espaço de tempo.

Devido à elevada sensibilidade da amplificação por PCR é essencial que exista um cuidado rigoroso durante todo o processamento, de forma a eliminar todas as possíveis fontes de contaminação e a otimizar a amplificação da sequência alvo (Brown T A, 1995; Robertson J M & Walsh-Weller J, 1998).

Primers:

A sequência (complementar à zona de flanco) e comprimento dos “*primers*” são factores críticos para o sucesso da amplificação (Brown, T A, 1995).

Os “*primers*” definem a especificidade da reacção de amplificação. Uma vez conhecidas as sequências das regiões de flanco podem-se desenhar “*primers*” complementares a estas

regiões e para tal devemos ter em atenção vários factores (Robertson J M & Walsh-Weller J, 1998):

- Tamanho: se são muito curtos são pouco específicos, se são muito longos hibridizam mais lentamente, com menos eficiência. Na prática usam-se “*primers*” com uma média de 18 a 30 nucleótidos;
- Concentração: baixas concentrações tornam a PCR pouco eficiente, por outro lado, concentrações muito elevadas elevam o risco de uma hibridização inespecífica dos “*primers*”.
- Sequência de cada “*primer*”: devem ser complementares das regiões de flanco, mas, não semelhantes entre si, de modo a evitar que ocorra a formação de dímeros de “*primers*” por auto-complementariedade entre eles. O conteúdo em bases G+C é normalmente de 40% a 60 %, sendo por vezes necessário aumentar ou diminuir o tamanho do “*primer*” para que esta proporção se verifique.
- Temperatura de “*melting*”: os “*primers*” envolvidos na mesma reacção de amplificação devem ter temperaturas de “*melting*” semelhantes para que a temperatura de hibridação deles não diferencie muito. A temperatura utilizada para a hibridação dos “*primers*” é muito importante, pois interfere na especificidade da reacção: se for demasiado elevada pode não ocorrer hibridação e se for muito baixa pode ocorrer hibridação com zonas não alvo dando origem a produtos de amplificação inespecíficos (Brown, T A, 1995).

Para a amplificação das duas regiões hipervariáveis do ADN mitocondrial foram utilizados os *primers* descritos por Wilson M R, *et al* (1995):

Primers HVR I:

A1 L15997: 5'-CAC CAT TAG CAC CCA AAG CT-3'

B1 H16395: 5'-CAC GGA GGA TGG TGG TCA AG-3'

Primers HVR II:

C1 L047: 5'-CTC ACG GGA GCT CTC CAT GC-3'

D1 H408: 5'-CTG TTA AAA GTG CAT ACC GCC A-3'

A combinação dos “*primers*” A1 e B1 dá origem a um fragmento, produto de PCR, da região HVI com 437 pb (pares de base) de comprimento. Do mesmo modo, a combinação dos “*primers*” C1 e D1 dá origem a um fragmento da região HVII com 402 pb, fig. II.1 (Wilson M R, *et al*, 1995; Butler J M, 1998).

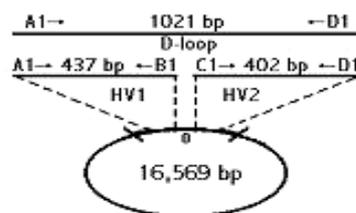


Figura II.1 - Produtos de PCR do ADN mt (região controladora). Tamanhos dos diferentes produtos de PCR obtidos a partir da amplificação da “*D-loop*” do ADN mt usando os primers A1/B1 e C1/D1 (Imagem adaptada de Butler J M, 1998).

Deoxinucleótidos (dNTP's)

São desoxirribonucleótidos trifosfatados (dNTP's) que são incorporados na cadeia de ADN em formação, nas posições complementares às do ADN – molde. São eles: dATP – Adenosina trifosfato; dTTP – Timina trifosfato; dCTP – Citosina trifosfato e dGTP – Guanina trifosfato.

Constituem a matéria-prima da ADN polimerase. É essencial a presença de todos os dNTP's em concentração suficiente para que ocorra a extensão dos fragmentos.

Enzima – ADN polimerase:

A ADN polimerase adiciona “dNTP's” complementares à cadeia molde, o primeiro será adicionado às extremidades 3' de cada primer o segundo a seguir (extremidade 3' do primeiro

“dNTP”) e assim sucessivamente, vai adicionando “dNTP’s” no sentido 5’→3’ à medida que lê a cadeia molde no sentido 3’→5’.

A ADN polimerase utilizada normalmente é uma “Taq DNA polimerase”, termoestável, com uma temperatura óptima de acção de cerca de 72º C. No entanto, verifica-se que ainda apresenta actividade significativa entre os 0º C e os 25 º C (Gelfand D H, 1989) o que pode resultar em produtos de amplificação inespecíficos e por vezes na formação de dímeros de “primers”.

A criação de uma versão modificada da enzima recombinante “*AmpliTaq DNA® polymerase*” que está inactiva até chegar a uma temperatura próxima da temperatura de desnaturação (92º C – 95º C) veio solucionar este problema (Robertson J M & Walsh-Weller J, 1998).

A “*AmpliTaq Gold™*” é fornecida num estado inactivo, sendo posteriormente activada pelo calor. No presente caso para que a enzima fosse completa ou parcialmente activada foi sujeita a um passo de aquecimento pré-PCR, esta activação transforma uma reacção normal de PCR numa “*Hot Start PCR*” a qual permite a mistura dos reagentes de PCR à temperatura ambiente, sem que se corra o risco da formação de produtos não específicos (Perkin Elmer – *PCR Systems, Reagents & Consumables*).

Esta enzima pode substituir a “*AmpliTaq DNA® polymerase*” em quase todas as reacções de PCR pela simples adição do passo de pré-PCR de 9-12 min, a 95º C, procedendo-se assim a uma eficiente reacção de PCR.

Tampão:

Proporciona a força iónica e o pH ideais para que a reacção ocorra. A concentração de iões de magnésio (Mg²⁺), factor crítico de qualquer PCR, requerida para que ocorra amplificação depende da enzima e dos “*primers*” utilizados.

O íão magnésio estabiliza o ADN de cadeia dupla. Concentrações elevadas deste íão podem dar origem a uma hibridização incorrecta dos “*primers*”, o que poderá originar a amplificação de regiões que não a alvo, a baixas temperaturas. Podem também promover a adição extra de “dNTP’s” e, adicionalmente, dificultam a separação das duas cadeias do ADN molde. Concentrações baixas de Mg^{2+} tornam mais específica a reacção de PCR, mas, podem levar a uma diminuição da actividade da ADN polimerase uma vez que estes íões funcionam como cofactores (Robertson J M & Walsh-Weller J, 1998).

Mas, este não é o único factor do tampão a afectar a reacção de PCR, o pH do tampão também a pode afectar; as condições iónicas podem ser ajustadas por forma a que o desempenho dos “*primers*” melhore (Blanchard *et al*, 1993) e a composição salina do tampão pode também interferir, segundo Chamberlin J S, *et al*, 1989, a conjugação de cloreto de potássio (KCl) e cloreto de magnésio ($MgCl_2$) pode ser menos eficiente do que a de sulfato de amónio ($(NH_4)_2SO_4$) e $MgCl_2$, de facto a “*Taq DNA Polymerase*” torna-se inactiva em tampões com uma elevada concentração de KCl (Robertson J M & Walsh-Weller J, 1998).

ADN-molde:

A importância de um ADN molde em boas condições é frequentemente desprezada aquando do desenvolvimento dum protocolo de PCR. Este é tão importante como qualquer outro interveniente na reacção, pois será a base de toda a amplificação.

Se as amostras contiverem inibidores ou ADN “estranho” que também hibridize com os *primers* o rendimento da reacção de amplificação será, obviamente, diminuído. É, portanto, importante que o ADN a utilizar como molde esteja suficientemente puro e em quantidade suficiente (Robertson J M & Walsh-Weller J, 1998). No presente trabalho não se procedeu à quantificação do ADN extraído uma vez que a extracção realizada está de tal forma estandardizada que se pode avançar directamente para uma reacção de PCR. A ser realizada a quantificação esta deve ser específica para ADN humano.

O protocolo de amplificação seguido foi o descrito por Wilson M R, *et al*, em 1995, com ligeiras adaptações (Anexo II).

A amplificação foi realizada nos termocicladores GeneAmp System 9600 e 2700, com os *primers* descritos anteriormente e recorrendo às duas polimerases referidas:

Para iniciar a amplificação é necessário que as amostras sejam incubadas a 95°C durante 1 min. para desnaturar o ADN molde. Quando se utiliza a “AmpliTaq Gold™ Polimerase” este passo é convertido num passo “*Hot Start*” de 11min a 95° C.

Segue-se o ciclo de *desnaturação* (95°C) – *hibridação* (60°C) – *síntese* (72°C) que se repete 32 vezes dando origem a várias centenas de cópias do fragmento de ADN alvo. Se necessário, quando as amostras são problemáticas, pode proceder-se a um aumento do número de ciclos. Na fase de desnaturação a solução é aquecida a 95°C por 10s, separando as duas cadeias de ADN; depois na fase de “annealing” ou hibridação a temperatura baixa para os 60°C durante 30s, permitindo a hibridação dos *primers* às moléculas de ADN molde; por fim na fase de extensão ou síntese a temperatura sobe até aos 72°C, perto da temperatura óptima de acção da “TaqDNAPolimerase” ocorrendo síntese de fragmentos de ADN correspondentes à região alvo por adição sucessiva de “dNTP’s” às extremidades 3’ livres (Wilson M R, *et al*, 1995; Brown, T A, 1995).

5. PURIFICAÇÃO DO PRODUTO AMPLIFICADO

Após a amplificação é necessário remover o excesso de “*primers*” e “dNTP’s” para que estes não venham a interferir com a reacção de sequenciação, para tal foram utilizadas colunas “Autoseq G-50”.

O “AutoSeq G-50” consiste em colunas “MicroSpin™” contendo “Sephadex® G-50” pré-equilibradas em água bidestilada. A “Sephadex® G-50” é uma resina que exclui moléculas com peso molecular superior a 50 KD, conseqüentemente, irá excluir o produto resultante da amplificação que é maior do que os poros desta resina. Os “*primers*” e “dNTP’s” em excesso, bem como pequenos fragmentos ADN resultantes duma amplificação parcial, ficam retidos nos poros da resina, sendo, assim, removidos do produto amplificado.

A purificação das amostras e controlos foi realizada seguindo o protocolo da “*Amersham pharmacia biotech*” (Anexo II).

6. VERIFICAÇÃO E QUANTIFICAÇÃO DO PRODUTO AMPLIFICADO

Quando usamos produto de PCR como molde para sequenciação, a sua concentração e pureza podem influenciar a qualidade da própria sequenciação. A presença de produtos de PCR contaminantes pode originar uma ‘mistura’ de sinais em várias posições ao longo do electroforetograma. Em adição, a não remoção dos *primers* de PCR em excesso pode resultar em produtos de sequenciação não desejados, o que aumenta o ruído de fundo (Butler J M, 1998). Portanto é necessário haver um método pré-sequenciação que possa detectar fragmentos de ADN contaminantes, ou excesso de *primers* de PCR, e que permite verificar a concentração do produto de PCR que servirá de molde (Steighner R J & Holland M, 1998). No presente trabalho esta detecção foi efectuada recorrendo a uma electroforese em gel no aparelho de “PhastSystem”.

O aparelho de “PhastSystem” consiste em duas unidades, uma de Separação e Controlo, e outra de Revelação, permitindo realizar electroforeses rápidas, de elevada resolução e reproduzíveis. A electroforese é um procedimento onde moléculas carregadas migram num campo eléctrico, sendo a taxa de migração determinada pelo tamanho das moléculas e pela sua carga eléctrica (Brock *et al*, 1994).

Este método permite analisar o produto de PCR recorrendo a pequenas quantidades de ADN, 1µl de amostra por aplicação. Para obtermos uma boa resolução visto que estamos a analisar fragmentos de ADN de tamanho reduzido, por volta dos 400pb, recorreremos a um “PhastGel Gradient 10-15” (4,5%T, 3%C), com uma zona de gradiente contínuo 10-15% com 2% de “crosslinking”, e “PhastGel SDS Buffer Strips”.

As amostras e controlos, resultantes da reacção de PCR e após purificação, são aplicados no gel com ajuda de um aplicador, deixando sempre um espaço entre as amostras e os controlos para evitar contaminação entre poços nesta fase e deste modo proceder a uma verificação correcta do produto amplificado. Foi utilizado o aplicador “PhastGel” 12/0,3, que aplica 12 amostras, num volume de aproximadamente 0,3 µl cada, conforme o manual do

“PhastSystem”. Deixamos correr a electroforese a 15°C com voltagem constante de 250V durante cerca de 30 min., tendo os cuidados recomendados para que ocorra a separação sem artefactos. O tempo de corrida não é crucial uma vez que se trata de um gel de prova.

A visualização da electroforese é feita por coloração com prata, que se baseia essencialmente numa reacção do ADN com iões prata (Ag^+) numa solução de nitrato de prata 0,25%, o que ocorre em vários passos na unidade de Revelação do “PhastSystem” (Tabela II.3), seguida de revelação manual: o gel é mergulhado numa solução básica (carbonato de sódio 3%) de formaldeído, seguidamente interrompe-se a revelação com uma solução de ácido acético 5%, e fixa-se com uma solução de ácido acético 10% / glicerol 5%, que previne que o gel enrole ou quebre depois de secar (*Phast System™ Owners Manual*).

Tabela II.3: Método de coloração por prata - programa da câmara de revelação do PhastSystem.

ETOH- Etanol; HAc- Ácido acético; AgNO_3 - Nitrato de prata.

PASSO	SOLUÇÕES	TEMPO	TEMPERATURA
1	50% ETOH / 10% HAc	2 min.	50°C
2	10% ETOH / 5% HAc	2 min.	50°C
3	10% ETOH / 5% HAc	4 min.	50°C
4	Água Milli-Q	6 min.	50°C
5	10% ETOH / 5% HAc	3 min.	50°C
6	10% ETOH / 5% HAc	5 min.	50°C
7	Água Milli-Q	2 min.	50°C
8	Água Milli-Q	2 min.	50°C
9	0,25% AgNO_3	13 min.	40°C
10	Água Milli-Q	0.5 min.	30°C
11	Água Milli-Q	0.5 min.	30°C

Após a revelação o gel é cuidadosamente analisado. Caso se verifique que ainda existe excesso de *primers* nas amostras, o que aparece como bandas que se encontram praticamente na frente do gel devido a terem aproximadamente 20 pb, repete-se a purificação por “AutoSeq G-50”. Quando os controlos aparecerem contaminados a amplificação é repetida, sendo rejeitado o produto desta.

7. SEQUENCIAÇÃO CÍCLICA DAS REGIÕES HVI E HVII

O primeiro método de sequenciação de fragmentos de ADN de tamanho até aproximadamente 500pb foi criado por A M Maxam e W Gilbert, nos finais de 1970. Neste método, as ligações entre os nucleótidos de quatro amostras de fragmentos de ADN marcados são quimicamente quebradas em nucleótidos específicos e diferentes.

Anos mais tarde, F Sanger e os seus colegas desenvolveram um segundo método de sequenciação de ADN, que envolve 2',3'-dideoxynucleótidos trifosfatados (ddNTP's), aos quais falta o grupo hidróxido 3'. Neste método o ADN de cadeia simples funciona como molde para síntese *in vitro*, usando um "primer" sintético. São realizadas quatro reacções separadas de sequenciação, cada uma com um dos quatro "ddNTP's" em concentração reduzida em adição a concentrações superiores de "dNTP's". A incorporação de "ddNTP's" em posições ao acaso termina a sequenciação, devido à ausência do grupo 3'-OH, o que impede a ligação ao nucleótido seguinte. Através duma autoradiografia da electroforese em paralelo a que as reacções são sujeitas pode ler-se a sequência do ADN molde original (Brock *et al*, 1994; Stryer L, 1995; Lodish H, *et al*, 1995).

A estratégia de sequenciação automática fluorescente baseia-se na incorporação nos produtos sintetizados de marcadores fluorescentes, que podem ser "*primers*" marcados com fluorescência na extremidade 5' (*dye primers*) ou "3'-ddNTP's" marcados com fluorocromos (*dye terminators*).

As reacções de sequenciação realizadas com "*primers*" marcados resultam, no electroforetograma, em picos mais regulares com alturas mais semelhantes do que quando são utilizados terminadores marcados. Contudo o método de "*primers*" marcados requer quatro reacções de sequenciação diferentes, em contraste com o de terminadores marcados em que é necessária apenas uma reacção. Adicionalmente existe actualmente ADN polimerases

desenhadas especificamente para sequenciação cíclica que diminuem bastante a disparidade de alturas entre os picos gerados pelos diferentes terminadores, como por exemplo, a AmpliTaq, FS™ da Perkin-Elmer (Parker L T, *et al*, 1995; Korch C & Drabkin H, 1999).

A “AmpliTaQ DNA polimerase, FS” é uma variante mutante da *Thermus aquaticus* (Taq) DNA polimerase e contém uma mutação pontual no seu sítio activo, havendo substituição duma fenilalanina por uma tirosina no resíduo 667, esta mutação resulta numa maior afinidade da enzima relativamente a ddNTPs, requerendo portanto uma menor concentração de terminadores do que a enzima AmpliTaq normal. Também contém uma mutação pontual próximo do seu terminal amino, levando à substituição duma glicina por um aspartato, que lhe retira quase toda a actividade de nuclease 5' → 3', o que elimina artefactos provenientes da sua actividade de exonuclease.

Esta enzima foi formulada com uma Pirofosfatase inorgânica termoestável que cliva o pirofosfato (PPi) libertado durante a síntese de novas cadeias, e impede que ele se acumule durante a sequenciação e interfira com esta (PE Applied Biosystems, 1998; Perkin Elmer, 1995a; PE Applied Biosystems, 1997; Applied Biosystems, 2001).

Com terminadores marcados, cada um dos “ddNTP’s” é marcado com um fluorocromo diferente. O facto de serem marcados os terminadores permite usar *primers* não marcados, que podem ser os mesmos da amplificação por PCR. A reacção de sequenciação pode ser realizada num único tubo, visto que a síntese de uma nova cadeia é simultaneamente terminada e marcada com o fluorocromo correspondente à base. Adicionalmente esta química requer menor quantidade de ADN molde inicial do que qualquer outra química fluorescente (PE Applied Biosystems, 1998; Perkin Elmer, 1995a).

A sequenciação cíclica é um método simples, bastante semelhante ao PCR, em que em cada ciclo de desnaturação, hibridação e síntese resultam numa amplificação linear de produtos de síntese de diferentes tamanhos. Este método requer muito menos ADN molde do que a sequenciação num só passo, e o uso de elevadas temperaturas permite uma síntese mais completa e reduz a hibridação de “*primers*” secundários ao ADN molde (PE Applied Biosystems, 1998).

No presente trabalho foi utilizado o *ABI Prism dRhodamine Terminator ou ABI PRISM BigDye Terminator v3.0 Cycle Sequencing Ready Reaction Kit*, com *AmpliTaQ DNA Polimerase, FS*. Estes “Kits” consistem numa mistura previamente preparada, num único tubo, de terminadores, desoxinucleótidos trifosfatados, enzimas (ADN Polimerases), cloreto de Magnésio e tampão. Esta mistura está pronta a utilizar, para ocorrer a reacção de sequenciação basta adicionar o “primer” específico, o ADN molde (produto amplificado) e água.

Em ambos os “Kits” o desoxinucleótido trifosfatado dGTP foi substituído por dITP que também hibridiza com a Citosina, mas diminui a compressão das duas cadeias de ADN na região onde é inserido. No segundo “Kit”, “BigDye Terminator v3.0” o dTTP é, também, substituído, desta vez por dUTP, melhorando a incorporação do terminador T (timina) o que resulta numa melhora do sinal deste (PE Applied Biosystems, 1997; Applied Biosystems, 2001).

Os terminadores de diclororhodamina (melhoramento dos terminadores rhodamina) foram desenhados com o intuito de originar picos com alturas mais regulares. Caracteriza-se pela obtenção, quando comparados com resultados obtidos por terminadores anteriores, de electroforetogramas com menos “ruído de fundo” (picos inespecíficos de pequenas dimensões); maior uniformidade no sinal dos vários terminadores e redução do padrão comum de picos de Guaninas fracos após os picos de Adeninas (PE Applied Biosystems, 1997). Os terminadores BigDye v3.0 (melhoramento dos anteriores BigDye e BigDye v2.0) caracterizam-se pela obtenção por uma maior uniformidade no sinal (picos) dos vários terminadores ao longo de toda a sequência (Applied Biosystems, 2001).

As reacções de sequenciação das regiões hipervariáveis I e II do ADN mitocondrial foram realizadas, primeiramente, no sentido directo utilizando os “primers” A1 e C1 referidos para a amplificação, respectivamente. Em amostras em que surgiram dúvidas relativamente à sua sequência, nomeadamente heteroplasmias, foi também realizada a sequenciação no sentido reverso, recorrendo aos *primers* B1 e D1, para as regiões respectivas, HVI e HVII (Anexo II).

A reacção de sequenciação foi realizada nos termocicladores GeneAmp System 9600 e 2700. O programa utilizado consiste na repetição por 25 vezes do ciclo: desnaturaçãõ a 95° C (10s) → hibridaçãõ a 56°C (5s) → extensãõ ou síntese a 72°C (4 min).

Foram seguidos os protocolos fornecidos com os “Kits” de sequenciação, dRhodamine™ Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems, 1997) e BigDye™ Terminator v3.0 Cycle Sequencing Ready Reaction Kit (Applied Biosystems, 2001), com ligeiras adaptações.

8. PURIFICAÇÃO DO PRODUTO SEQUENCIADO

Antes de proceder a uma electroforese para analisar os produtos da sequenciação cíclica é necessário remover o excesso de “*primers*” e de terminadores não incorporados, pois estes prejudicam a análise, dado que aparecem como artefactos nos electroforetogramas que prejudicam, por vezes, a visualização das bases correspondentes à real sequência do fragmento de ADN a estudar. Os primeiros fragmentos resultantes da sequenciação a migrar são curtos, possuem um peso molecular reduzido, sendo que os terminadores livres migram juntamente com eles. Estes terminadores apresentam um sinal de fluorescência muito elevado e por vezes potenciam o sinal das bases com que coincidem.

De modo a aumentar a qualidade do produto de sequenciação a analisar e obter melhores resultados procedeu-se a uma purificação do produto de sequenciação com ambos os terminadores por precipitação etanólica, que no caso do resultante da sequenciação com “dRhodamine” se dá na presença de Cloreto de Magnésio ($MgCl_2$). Assim sendo, o produto sequenciado precipita, permanecendo em suspensão os terminadores e “*primers*” não incorporados.

Foram seguidos os protocolos fornecidos com os Kits de sequenciação, dRhodamine™ Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems, 1997) e BigDye™ Terminator v3.0 Cycle Sequencing Ready Reaction Kit (Applied Biosystems, 2001), que diferem ligeiramente entre si. Algumas adaptações, como a supressão duma lavagem final com EtOH a 70%, nomeadamente nos produtos da sequenciação com “BigDye terminators v3.0”, foram realizadas visando melhores resultados (Anexo II).

9. APLICAÇÃO NO SEQUENCIADOR AUTOMÁTICO ABI PRISM 377

A análise num sequenciador automático dum produto de sequenciação depende muito da qualidade do gel usado para a electroforese assim como as placas de vidro utilizadas.

O sequenciador automático ABI Prism 377 possui uma cassette própria para montar e posicionar o gel correctamente. Primeiro é necessário verificar as placas, estas devem ser lavadas cuidadosamente com um detergente específico e enxaguadas com água destilada. Montam-se as placas na cassette e faz-se um “Plate-Check”, primeiro para cada uma delas, limpando qualquer possível interferência, e seguidamente montam-se as duas com os espaçadores e faz-se novo “Plate-Check”, então colocam-se os acessórios necessários para injectar o gel (Watts D, 1998; Perkin Elmer, 1995b).

O gel utilizado foi um Long Ranger, pois foram obtidos melhores resultados do que com géis preparados manualmente, seguindo as instruções do próprio gel. Este é um gel (50ml) de poliacrilamida e bisacrilamida em tampão, que é inicialmente misturada com ureia (6M), e depois com APS (Persulfato de Amónio, 0,05%) e TEMED (0,07%), iniciando-se a polimerização, a concentração destes agentes é crítica. Injecta-se o gel nas placas já preparadas, com cuidado para não formar bolhas, e espera-se que polimerize, o que demora cerca de 2 horas. Prepara-se então o tampão para a electroforese dissolvendo uma solução stock de TBE 10X (Tris Base 890mM, Ácido Bórico 890mM, Na₂EDTA-2H₂O 20mM) em água desionizada de modo a obtermos uma solução de TBE 1X (Perkin Elmer, 1995b; FMC®, 2002).

Antes de iniciar a electroforese faz-se uma “sample sheet”. A sequenciação requer uma matriz e um ficheiro de mobilidade (“*Dye Set / Primer mobility files*”) específicos para o “Kit” utilizado, e convém confirmar também o “run module” a utilizar. A matriz para cada um dos “Kits” utilizados foi criada de acordo com os seus protocolos, os ficheiros de mobilidade utilizados foram os referidos para o sequenciador ABI Prism 377: “*DT {dR Set Any-Primer}*”, específico

para os terminadores de dRodamina, e “DT377 {BDv3.0}”, no caso dos terminadores “BigDye v3,0” (Perkin Elmer, 1995a; PE Applied Biosystems, 1997; Applied Biosystems, 2001).

As amostras e controlos serão corridas num gel desnaturante de modo a obtermos melhor resolução, pelo que precisam ser desnaturadas antes de serem aplicadas. Para tal usa-se formamida que é um agente desnaturante, e aquecem-se as amostras a 95°C para desnaturar as duplas hélices, guardam-se as amostras em gelo até serem aplicadas, para evitar que renaturem (Perkin Elmer, 1995a; PE Applied Biosystems, 1997; Watts D, 1998).

O protocolo, de preparação das amostras para a aplicação no gel, descrito para cada “Kit” é muito semelhante (PE Applied Biosystems, 1997; Applied Biosystems, 2001). A preparação foi realizada da mesma forma para todas as amostras (Anexo II).

Antes de aplicar as amostras deve fazer-se um novo “Plate Check” para certificar que o gel apresenta uma “baseline” baixa e limpa para o detector de fluorescência. Quando é boa pode adicionar-se o tampão ao sistema, e após lavar os poços com tampão, faz-se um “Pré-Run” para elevar a temperatura até à temperatura ideal de corrida (Watts D, 1998).

Seguidamente as amostras são aplicadas, mas somente nas “lanes” ímpares das 36 disponíveis de modo a evitar contaminações “lane” - “lane”, pode aplicar-se 0,75-2µl de cada amostra (Perkin Elmer, 1995a). Inicia-se então a corrida seleccionando o modo de “Run” apropriado, “SeqRun 36E-2400”, que demora 7 horas.

Os fragmentos sequenciados são separados electroforéticamente com base no seu peso molecular, cada fragmento termina no dideoxinucleótido complementar da cadeia molde numa determinada posição e marcado com um fluorocromo específico.

10. DETECÇÃO E ANÁLISE DO PRODUTO SEQUENCIADO

Após sequenciação e finalizada a electroforese os dados obtidos são automaticamente analisados pelo *ABI Prism Collection software*, que recebe as intensidades luminosas de quatro áreas específicas na CCD camera, correspondendo cada uma das áreas a um comprimento de onda específico, funcionando assim como um filtro virtual. As cores exibidas pelos vários terminadores tanto na imagem do gel como nos electroforetogramas obtidos dependem, além do fluorocromo respectivo, do filtro a que a luz por eles emitida está sujeita. O “tracking” das “lanes” do gel é automático, mas é permitido que seja verificado e ajustado.

A detecção da sequência dos fragmentos a analisar é possível porque estes estão marcados com moléculas fluorescentes. No geral a estrutura dum marcador fluorescente envolve um dador derivado de Fluoresceína, por exemplo: 6-carboxifluoresceína (6-FAM), ligado a um receptor de diclororhodamina (dRhodamine). O máximo de excitação de cada marcador coincide com o do dador de fluoresceína, e o espectro de emissão é o do receptor de dRhodamina. A ligação confere à transferência energética entre dador e receptor uma extrema eficiência (PE Applied Biosystems, 1998).

O sequenciador ABI PRISM 377 detecta a fluorescência dos quatro fluorocromos diferentes que são usados para identificar as diferentes bases (Tabela II.4). Cada marcador fluorescente emite luz num comprimento de onda específico quando excitado por um laser iónico de argon; a cor correspondente é determinada pelo filtro receptor de fluorescência, sendo assim criada uma imagem do gel cujas cores das bandas apresentadas irão corresponder aos diferentes nucleótidos. As quatro cores, e consequentemente as quatro bases, podem ser detectadas distintamente numa única “lane” do gel. A correspondência entre os nucleótidos (terminadores) e as cores dos picos representados nos electroforetogramas depende do ficheiro de mobilidade e da matriz seleccionados (Watts D, 1998; PE Applied Biosystems, 1998).

Tabela II.4 – Fluorocromos de cada terminador, específicos para cada “Kit”, e correspondente cor exibida no gel de electroforese no sequenciador “ABI Prism 377”.

Terminadores	“Kit dRhodamine”		“Kit BigDye v3.0”	
	Fluorocromo	Imagem no Gel	Fluorocromo	Imagem no Gel
A	dicloro [R6G]	Verde	V3 Dye 2	Verde
C	dicloro [TAMRA]	Amarelo	V3 Dye 4	Vermelho
G	dicloro [R110]	Azul	V3 Dye 1	Azul
T	dicloro [ROX]	Vermelho	V3 Dye 3	Amarelo

A análise dos multicomponentes dos dados da sequenciação é também realizada automaticamente pelo *DNA Sequencing Analysis Software™ v.3.7*, que aplica cálculos matemáticos de matrizes a todas as amostras.

A sequência de nucleótidos obtida em cada electroforetograma é alinhada e comparada com a “*Cambridge Reference Sequence*” (CRS), pelo *SeqScape® Software v.2.0*, sendo assinalados os pontos em que ocorre divergência. Esta sequência padrão corresponde à primeira sequência de ADN mitocondrial descrita (Andrews R M, *et al*, 1999, Anderson S *et al*, 1981).

A análise da sequência dos electroforetogramas obtidos numa reacção de sequenciação cíclica requer, por vezes, a atribuição manual de vários pares de bases. Os problemas típicos que obrigam a este procedimento incluem a remoção incompleta dos terminadores marcados que não foram incorporados durante a sequenciação cíclica, a existência de diferentes produtos de PCR que foram sequenciados e a ocorrência de heteroplasmias do ADN mitocondrial.

A análise também terá que ter em conta algumas limitações da química dos Kits utilizados para a reacção de sequenciação: a presença de picos de tamanho variável devido a certos padrões, específicos dos Kits, de incorporação de determinados terminadores e a velocidades de migração dos terminadores diferentes (Parker L T, *et al*, 1995; Metzker M L, *et al*, 1995).

É importante manter presente a noção de que a intensidade de sinal, altura e área do pico no electroforetograma, corresponde à quantidade de marcador fluorescente e não à real quantidade de ADN ou de qualquer base que tenha sido sequenciada (Butler J M, 1998).

Os resultados obtidos foram interpretados de acordo com as considerações da “*DNA Commission of the International Society for Forensic Genetics*” –ISFG – (Carracedo A, *et al*, 2000) e do “*European DNA profiling group*” –EDNAP- (Tully G, *et al*, 2001).

A análise de uma amostra é realizada em comparação com a sequência reportada por Anderson S, *et al*, em 1981, designada por *CRS*, sendo reportadas as diferenças, observadas na sequência da amostra, relativamente a esta sequência padrão:

- *Substituições (Transições e Transversões).*

Sempre que apareça um nucleótido diferente do reportado por Anderson S, *et al* para uma determinada posição, essa diferença será representada pelo número correspondente à posição em causa seguido pela base observada na sequência da amostra nessa posição. Por exemplo, se na posição 280 aparecer um T onde na *CRS* se encontra um C, esta diferença é evidenciada e a nomenclatura a adoptar é 280T.

- *Delecções.*

Se o nucleótido da posição x da *CRS* não aparece na amostra analisada, estamos na presença duma delecção, neste caso esta diferença será representada pelo número correspondente à posição em causa seguido da letra “d”. Por exemplo, se a base existente entre a posição 245 e 247 não aparece, na sequência da amostra analisada, a nomenclatura para esta diferença é 246d.

- *Inserções.*

Se aparecer um nucleótido a mais em determinada posição, este será assinalado com “.1” após o n.º referente à posição mais baixa das bases entre as quais se verifica a inserção. Por exemplo, se for observada uma Adenosina adicional entre as posições 245 e 246 esta será referenciada como 245.1A. Se a inserção ocorrer num poli – C, por exemplo, a sua localização exacta é desconhecida, nesta situação assume-se que

ocorreu na posição mais elevada. Por exemplo, se for observada uma inserção de uma Citosina entre as posições 302 e 310 (Poli – C) a nomenclatura a adoptar será 309.1C; duas inserções, no mesmo poli – C, serão designadas por 309.1C e 309.2C.

- *Heteroplasmias pontuais.*

Se for clara a presença de duas bases distintas, com picos de intensidade semelhantes, na mesma posição nucleotídica estamos presentes uma heteroplasma e pode-se usar a designação IUB (International Code for Unresolved Bases) apropriada, ver tabela II.5, ou, em alternativa a designação de C~T, por exemplo. As designações IUB foram definidas em 1984 pelo 'Nomenclature Committee of the International Union of Biochemistry (NC-IUB)', e publicadas em várias revistas da especialidade nos anos subsequentes, NC-IUB, 1986; NC-IUB, 1985; Cornish-Bowden, A, 1985, entre outras.

Tabela II.5 – Designações possíveis para diferentes heteroplasmias e exemplos.

Heteroplasma	Designação IUB	Razão da designação
G ~ A	R	puRine
T ~ C	Y	pYrimidine
A ~ C	M	aMino
G ~ T	K	Keto
G ~ C	S	Strong interaction (3 ligações H)
A ~ T	W	Weak interaction (2 ligações H)
A ~ C ~ T	H	não-G, H é a letra a seguir ao G
G ~ T ~ C	B	não-A, B é a seguir ao A
G ~ C ~ A	V	não-T (não-U), V é a seguir ao U
G ~ A ~ T	D	não-C, D é a seguir ao C
G ~ A ~ T ~ C	N	aNy (qualquer base)
Exemplo:	✓ Se na posição 16261 se observar a presença, em igual proporção, das bases C e T, verifica-se a existência de uma heteroplasma que pode ser designada de duas formas: 16261 Y ou 16261 C~T.	

Uma situação de heteroplasmia deve ser sempre comprovada pela sequenciação de ambas as cadeias do ADN mt. Só se considera a presença de uma heteroplasmia se o sinal referente a ambas as bases se destacar claramente do ruído de fundo.

Caso se verifique, após sequenciação em ambos os sentidos, uma predominância de uma base relativamente à outra, pode aplicar-se a notação C>T, por exemplo, em que C será a base maioritária e T a base que se encontra em menor número de moléculas de ADN mt. Num estudo populacional não é vulgar a utilização desta notação, um vez que a presença de uma base minoritária é desprezada, por norma.

Se não for possível identificar as bases presentes numa determinada posição, nem com a segunda reacção de sequenciação, essa posição é designada como ambígua e assinalada com um "N".

- *Heteroplasmias de comprimento.*

Como referido anteriormente, na região controlo do ADN mt existem zonas homopoliméricas, *i. e.*, Poli – C's, tanto na região HVI (da posição 16183 a 16194) como na região HVII (da posição 302 a 310), onde, frequentemente, se observa a ocorrência de heteroplasmias de comprimento, inserção diferencial de bases. A análise deste tipo de heteroplasmia pode ser realizada com base nas mesmas considerações que foram apresentadas no ponto anterior.

Geralmente a presença de uma heteroplasmia de comprimento caracteriza-se por um desfasamento na sequência nucleotídica a jusante da região homopolimérica. Por vezes torna-se difícil a determinação do número exacto de resíduos de Citosina presentes no poli-C heteroplásmico, mesmo recorrendo a sequenciações múltiplas em ambos os sentidos, directo e reverso. Se o número de Citosinas puder ser determinado a notação a usar pode ser do tipo: 309.1C > 309.2C, no caso de uma inserção de uma Citosina num poli-C (302-310) e de uma inserção de duas Citosinas no mesmo poli-C, mas numa família de moléculas de ADN mt menor do que a que alberga a primeira inserção.

Quando estamos na presença de deleções ou inserções que levam a uma alteração no comprimento da região controlo do ADN mt, deve ter-se em consideração ainda as recomendações de Wilson, M R, *et al* (2002) a quando da sua análise. Estas recomendações são (por ordem hierárquica):

1. Devemos optar pela situação em que temos que registar o menor número de diferenças relativamente à sequência de referência;
2. Quando podemos emparelhar a sequência em análise com a de referência de várias maneiras, mantendo o menor número de diferenças, devemos seguir a seguinte ordem de prioridades quanto às diferenças a registar:
 - a. Inserções e/ou deleções;
 - b. Transições;
 - c. Transversões.
3. As inserções e deleções devem ser sempre reportadas como colocadas no extremo **3'** da sequência, tendo como referência a cadeia L (leve / light) do ADN mt.

No que toca às heteroplasmias de comprimento a sua análise deve ser particularmente cuidada, recorrendo-se sempre à sequenciação no sentido directo e reverso, com o intuito de melhor esclarecer que tipo de heteroplasmia, inserção / deleção, estamos a analisar. Por vezes é necessário ainda adoptar outras estratégias de modo a melhor compreender o fenómeno heteroplásmico em questão, podendo passar por uma amplificação específica e direccionada da zona em questão (Rasmussen E M, *et al*, 2002).

11. ANÁLISE ESTATÍSTICA

Esta análise foi efectuada considerando, somente, as amostras em que foi possível determinar a sequência das duas regiões hipervariáveis (HVI e HVII) do ADN mitocondrial. Na análise intra-populacional, das populações estudadas neste trabalho, e na análise comparativa destas duas populações foram considerados os intervalos de posições nucleotídicas entre 16024 e 16391 para a HVI e entre as posições 58 e 391 para a HVII.

Todos os polimorfismos observados, na amostragem considerada, foram analisados inferindo-se a sua frequência e qual a razão da ocorrência de determinados fenómenos mutacionais relativamente a outros. Foram contabilizados o total de polimorfismos de sequência para cada uma das regiões hipervariáveis estudadas e dentro destes foi determinada, para os resultantes de fenómenos de Substituição, a razão entre Transições e Transversões, contabilizando-se os vários tipos de Transversões. Foram ainda determinadas as frequências de ocorrências de fenómenos de Inserção e Delecção. Outros polimorfismos contabilizados, cujas frequências também foram inferidas, foram as Heteroplasmias de Sequência e de Comprimento.

Para a comparação com outras populações foram realizadas duas análises em separado, uma tendo em conta apenas a região HVI (16090-16365), uma vez que muitos estudos se resumem a esta região. Para as populações com informação relativa à região HVII (073-340) esta também foi tida em consideração numa nova análise conjunta das duas regiões hipervariáveis HVI + HVII.

A análise estatística intra-populacional e inter-populacional foi realizada recorrendo ao software **Arlequin 3.11** (Excoffier L, *et al*, 2005). Nesta análise não foram tidas em consideração inserções, deleções nas regiões homopoliméricas e heteroplasmias de comprimento.

Recorrendo a este software foram calculados vários parâmetros de diversidade molecular e *mismatch distributions*, para as populações consideradas nas diferentes análises

realizadas: Intra – populacionais (Cabo Verde e Guiné-Bissau – HVI; HVII e HVI+HVII) e Inter – populacional, entre estas duas populações (HVI, HVII e HVI+HVII) e com outras populações (HVI e HVI + HVII). Os Parâmetros de diversidade molecular considerados foram os seguintes:

- **N** – tamanho da amostra, n^o de indivíduos estudados;
- **K** – número de sequências diferentes encontradas / haplótipos;
- **A** – número de posições nucleotídicas que apresentam polimorfismo;
- **π** – diversidade nucleotídica, em que:

$$\pi = (n/n-1) \cdot (1/l) \cdot \sum_{i=1}^l (1-x_i^2)$$

n – tamanho da amostragem; l – comprimento da sequência em estudo; x_i – frequência relativa de cada nucleótido na posição i (Nei M & Tajima F, 1981).

- **J** – diversidade genética de sequências / haplótipos:

$$J = (n/n-1) \cdot (1-\sum_{i=1}^k p_i^2)$$

n – tamanho da amostragem; p_i – frequência relativa de cada sequência diferente i ; K – número de sequências diferentes (Nei M & Tajima F, 1981).

As matrizes de distâncias obtidas pelo método de *Pairwise difference* foram utilizadas para traçar as árvores filogenéticas das populações, recorrendo ao programa **Neighbor** (Neighbor-Joining/UPGMA method version 3.573c), incorporado no **Phylip 3.5** (Felsenstein J, 1993), para tratar os dados, e de seguida ao programa **TreeView 1.5** para visualizar a representação gráfica dos dados na forma de árvores (Page R, 1996).

12. CLASSIFICAÇÃO EM HAPLOGRUPOS

A classificação em haplogrupos de ADN mt é normalmente realizada com recurso a análise de Polimorfismos do Comprimento dos Fragmentos de Restrição (PLFR's ou RFLP's) que se localizam em várias regiões do ADN mt (por exemplo, Easton R D *et al*, 1996). Vários autores complementam a informação retirada desta análise com a informação obtida a partir das sequências das regiões hipervariáveis do ADN mt ou utilizam somente esta informação para classificar as amostras em haplogrupos e sub-haplogrupos, vulgarmente a HVI (exemplos: Jackson B A *et al*, 2005; Moraga M *et al*, 2005; Brehm A *et al*, 2003; Brehm A *et al*, 2002; Malyarchuk B A *et al*, 2002; Salas, A, *et al*, 2002, Pereira L *et al*, 2001; Quintana-Murci L *et al*, 1999; Watson E *et al*, 1997). Por vezes é conjugada a informação da HVI com a posição 73 da HVII para fazer uma melhor correspondência entre Haplogrupos e sequências e / ou complementar a informação sobre determinada população (Tolk H-V *et al*, 2001; Pereira L *et al*, 2000; Macaulay V *et al*, 1999). Alguns autores já optam por considerar também toda a região HVII (Barbosa A B G *et al*, 2008; Berniell-Lee G *et al*, 2008; Beleza S *et al*, 2005; Jin H J *et al*, 2006; Brandstätter A *et al*, 2004; Poetsch M *et al*, 2003; Derbeneva O A *et al*, 2002; Bandelt H-J *et al*, 2001; Torroni A *et al*, 1996).

Esta classificação é muito relevante quer a nível antropológico, quer numa análise forense, uma vez que se sabe que há haplogrupos característicos de determinados continentes ou regiões do mundo. Os haplogrupos L0, L1, L2 e L3 são tipicamente originários de África, os haplogrupos H, I, J, K, T, U, V, W e X estão associados à Europa; nas populações da Ásia são frequentes os haplogrupos A, B, C, D, E, F, G, M e N; os haplogrupos M e N encontram-se também na Oceânia; no Continente Americano encontramos os haplogrupos A, B, C, D e X, este último na América do Norte (ver figura II.2).

Do ponto de vista forense a classificação em haplogrupos, em complemento da análise das regiões HVI e HVII, permite validar a informação que advem da sequenciação destas regiões, uma vez que a existência de erros, provenientes do processamento e/ou análise das

amostras e em sequências publicadas e disponíveis em base de dados, pode ser evidenciada e assim detectada (Barbosa A B G, 2006).

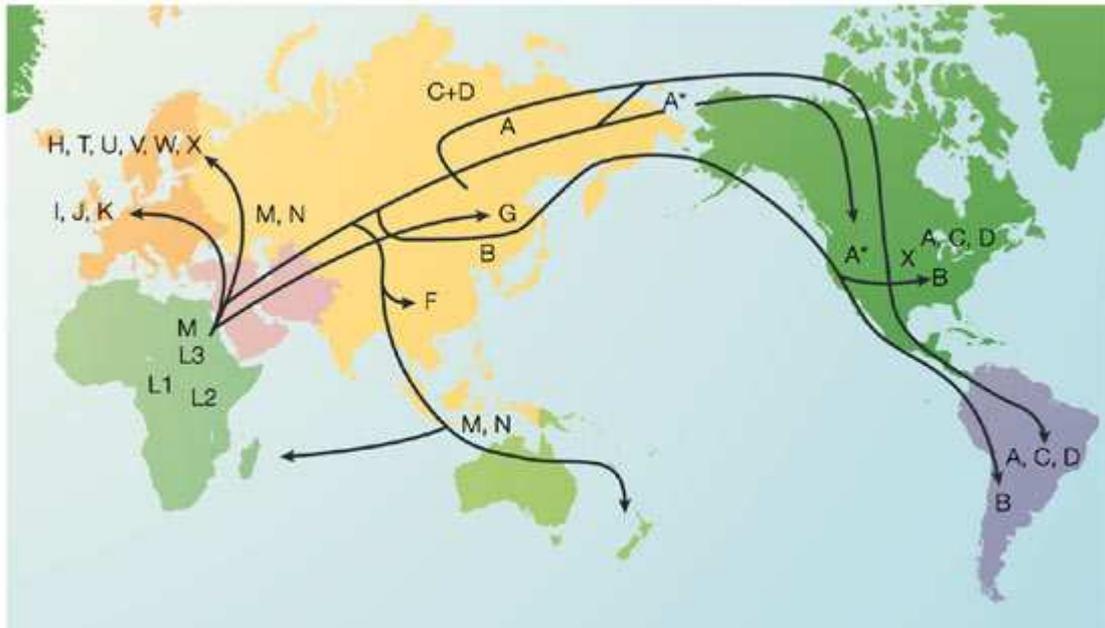


Figura II.1 – Distribuição geográfica dos diferentes Haplogrupos e respectivas migrações (Imagem de Shriver M D & Kites R A, 2004).

No presente trabalho, a classificação em haplogrupos e sub-haplogrupos foi efectuada com base nas sequências das duas regiões hipervariáveis do ADN mt, HVI e HVII, recorrendo a uma ferramenta para uso forense disponível on-line: 'mtDNAManager – forensic mtDNA database' – <http://mtmanager.yonsei.ac.kr> – (Anexo III). A classificação obtida foi revista tendo em consideração as classificações publicadas por Berniell-Lee G, *et al* (2008), Behar D M, *et al* (2007), Barbosa A B G (2006), Brandstätter A, *et al* (2004), Salas A, *et al* (2004 e 2002), Pereira L, *et al* (2001), Bandelt H-J, *et al* (2001), Watson E, *et al* (1997), Torroni A, *et al* (1996).