

U. PORTO



INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

**Diversidade mitocondrial no Sudão: aplicações à genética
forense e à genética populacional**

CARLA ALICE AFONSO

Dissertação de Mestrado em Medicina Legal

2008

CARLA ALICE AFONSO

Diversidade mitocondrial no Sudão: aplicações à genética forense e à genética populacional

Dissertação de Candidatura ao grau de
Mestre em Medicina Legal submetida ao
Instituto de Ciências Biomédicas de
Abel Salazar da Universidade do Porto

Orientadora - Doutora Luísa Maria de Sousa Mesquita Pereira
Investigadora do IPATIMUP-Instituto de
Patologia e Imunologia Molecular da
Universidade do Porto

AGRADECIMENTOS

Vou tentar exprimir por palavras o sentimento de profundo agradecimento para com todos aqueles que contribuíram para a concretização deste trabalho.

Agradeço à minha Amiga e Orientadora, Doutora Luísa Pereira pelo seu apoio, empenho e disponibilidade incondicionais desde o primeiro momento em que me propus elaborar este trabalho. Encontrei junto dela um sincero voto de confiança nas minhas capacidades, a motivação para a actividade científica, estímulo e força de vontade necessárias para, apesar das dificuldades, não deixar de concluir o objectivo proposto. Mesmo quando as forças faltaram para continuar, conseguiu com a sua compreensão, imensa paciência e esforço despertar em mim a vontade de conseguir levar até ao fim este trabalho. Foi para mim um verdadeiro privilégio tê-la como Orientadora a partilhar novamente comigo mais uma etapa da vida académica!

À Prof. Doutora Maria José Carneiro de Sousa, coordenadora do Mestrado, agradeço a disponibilidade e colaboração, sempre que solicitadas, o incentivo e motivação para a realização deste trabalho.

A realização da parte prática deste trabalho só foi possível devido ao apoio do IPATIMUP, instituição a que, na pessoa do seu Presidente, o Prof. Doutor Sobrinho Simões, eu agradeço pela disponibilidade, na íntegra, de todos os recursos necessários à execução deste trabalho.

Agradeço o apoio financeiro concedido pela Fundação para a Ciência e Tecnologia para a realização prática deste trabalho, através do projecto PTDC/ANT/66275/2006 - "Reavaliação com ferramentas genéticas de elevada resolução das origens geográficas e das rotas de dispersão dos primeiros homens modernos e dos primeiros agricultores da bacia mediterrânica".

À Doutora Farida Alshamali, agradeço a colaboração pelo envio das amostras usadas neste estudo.

Quero manifestar um agradecimento muito especial aos elementos do grupo de trabalho, Diversidade Genética e Bioinformática onde fui inserida: Fernando, Joana, Marta, Sandra e Verónica, que com o seu carinho, apoio e cooperação muito contribuíram para um excelente ambiente de trabalho. Conseguiram com a sua atitude positiva de alegria no trabalho e espírito de camaradagem, "dispersar a nuvem negra" do meu cansaço, tornando

o grupo muito “acolhedor” para mim, onde realmente me senti bem-vinda, apoiada e compreendida.

Aos colegas de Mestrado, particularmente à Ana Lerenó e Ana Raquel Monteiro, agradeço o apoio, o convívio e os laços de amizade sincera criados ao longo deste trabalho: quando nos apoiamos mutuamente parece que tudo se torna mais fácil.

Agradeço de forma especialmente sincera ao Tozé pelo imenso e carinhoso apoio que me dedicou desde o início do trabalho e que foi contributo essencial para a manutenção da minha motivação e auto-confiança.

Também quero agradecer à Zélia, minha colega de trabalho e amiga pelas palavras encorajamento constantes e pelos nossos mútuos “desabafos”!

Aos meus pais e irmã agradeço a especial atenção e compreensão dispensadas ao longo da realização do Mestrado.

RESUMO

As características peculiares do DNA mitocondrial (mtDNA), como transmissão uniparental, ausência de recombinação, o facto de constituir $\frac{1}{4}$ do efectivo populacional relativamente aos marcadores autossómicos, presença em várias cópias por célula, torna-o uma das ferramentas genéticas mais usadas e informativas em genética populacional e forense, mas requer um estudo prévio cuidado e uma conceitualização formal diferente da clássica genética autossómica.

A nível forense, são necessárias bases de dados de efectivo amostral considerável, porque a probabilidade de partilha ou *match* de haplótipos pode ser elevada para uma pequena proporção de haplótipos, sendo a extensa maioria dos haplótipos única, como se comprovou para as populações euroasiáticas. É também aconselhável um estudo filogenético dos haplótipos, como ferramenta de controlo de qualidade, de edição das sequências e troca de amostras entre análises independentes.

Neste trabalho efectuou-se a caracterização da diversidade mitocondrial nas regiões hipervariáveis I e II (HVRI e HVRII) numa amostra de 102 indivíduos do Sudão, contribuindo para a construção de uma base de dados para populações do leste de África, obedecendo aos critérios de qualidade forense. A probabilidade de *match* de haplótipos no Sudão é muito baixa: 2% em HVRI; 4% em HVRII; 1,8% em HVRI+HVRII. Mesmo dentro dos vários haplogrupos, os haplótipos são bastante diversos. A caracterização da sequência completa de mtDNA em 21 haplótipos L3(xM,N) do Sudão e a sua comparação com 94 sequências completas permitiu avaliar as taxas de mutação relativas das diversas posições neste haplogrupo, informação essencial ao desenvolvimento de kits SNaPshot para aplicação forense e populacional.

Na área da genética populacional, os primeiros dados do mtDNA vieram apoiar o modelo Out-of-Africa para a origem do Homem Moderno, datando a origem única há cerca de 200.000 anos e localizando-a, com elevada probabilidade, no leste de África. Os dados de sequenciação completa estão a permitir afinar datações e inferir possíveis rotas de migração, estando actualmente em debate uma rota de migração mais antiga (cerca de 65.000 anos) através do Corno de África, pelo Mar Vermelho, sul da Península Arábica e em direcção ao sudeste asiático versus a rota mais recente (cerca de 45.000 anos) através do Levante. A caracterização da sequência completa de mtDNA em 21 haplótipos L3(xM,N) do Sudão e a sua comparação com 94 sequências completas permitiu, neste trabalho, contribuir para uma afinação da datação das linhagens L3 na origem de toda a diversidade

fora de África. A idade estimada para o haplogrupo L3 foi de 69.600 ± 5.900 anos. Comprovou-se a considerável multifurcação a partir da raiz de L3, apresentando os subhaplogrupos originados uma idade da ordem de grandeza do aparecimento do haplogrupo, atestando a forte explosão demográfica ocorrida nesta altura. Não se encontraram linhagens M e N basais no Sudão, pelo que estas devem ter sido originadas já fora de África.

SUMMARY

Mitochondrial DNA (mtDNA) is one of the most used and informative tools in population and forensic genetics, due to its peculiar characteristics of uniparental transmission, absence of recombination, $\frac{1}{4}$ of the autosomal effective size, presence in several copies per cell. But it implies a reformulation of concepts comparatively to the classical autosomal genetics, and a careful preliminary study should be performed before application to casework.

In forensics, it is necessary to have databases with a considerable size because the probability of match can be high for a few haplotypes, while the major proportion is made of unique haplotypes, as has been estimated in euroasiatic populations. It is also advisable to perform a phylogenetic study, as a tool of quality control of sequences edition and mismatch of samples between independent analyses.

In this work, we characterised the mtDNA diversity for hypervariable regions I and II (HVRI e HVRII) in a sample of 102 Sudanese. This is a contribution to the construction of a mtDNA database for East African populations, obeying to the quality criteria in forensics. The match probability in Sudan is low: 2% in HVRI; 4% in HVRII; 1.8% in HVRI+HVRII. Even intra-haplogroups, the haplotypes are diverse. The screening of the complete mtDNA sequence in 21 L3(xM,N) haplotypes and its comparison with 94 published sequences allowed to evaluating the mutation rates per position in this haplogroup, an essential information to the development of SNaPshot kits for forensic and population application.

In population genetics, the first mtDNA data supported the model Out-of-Africa for the origin of modern humans, dating this event at 200,000 years ago and locating it in East Africa. The data from complete mtDNA sequencing are now allowing to refining the dating and evaluation of possible migration routes. It is in debate an earlier (~65,000 years ago) southern route, through the Horn of Africa, Red Sea, south Arabian Peninsula and in direction to southeast Asia versus the later (~45,000 years ago) route through the Levant. In this work, the screening of the complete mtDNA sequence in 21 L3(xM,N) haplotypes and its comparison with 94 published sequences allowed to refining the dating of the L3 lineages, in the origin of all non-African diversity. The estimated age for L3 haplogroup is $69,600 \pm 5,900$ years. The considerable multifurcation from the L3 root was confirmed, and all sub-haplogroups have an age of the same order of magnitude as L3, attesting the strong demographic expansion by this time. Basal M and N lineages were not detected in Sudan, supporting the hypothesis that they were originated already outside Africa.

RÉSUMÉ

Les caractéristiques particulières du DNA mitochondrial (mtDNA), comme transmission uniparental, absence de recombinaison, il comprends $\frac{1}{4}$ de l'effectif des populations à l'égard des marqueurs autosomiques, présence dans plusieurs copies par cellule, rend l'un des outils génétiques plus usés et informatifs aux études de la génétique des populations et dans la génétique criminelle, mais ça va impliquer une étude préalable soignée et une contextualisation différente de la classique génétique autosomique.

Au niveau de la médecine légale, on a besoin des bases de données avec un grand format parce que la probabilité de partage ou *matches* de haplotypes peut se révéler élevée pour un petit nombre de haplotypes, pendant que la grande portion de ces haplotypes reste unique comme si elle a été vérifiée pour les populations euroasiatiques. Il est aussi recommandé de faire une étude phylogénétique des haplotypes, comme un moyen de contrôler la qualité de l'édition des séquences et la discordance de résultats pour les mêmes échantillons mais dans des études indépendantes.

Dans ce projet, nous avons caractérisé la diversité du mtDNA pour les régions hypervariables I et II (HVRI et HVRII) dans un échantillon de 102 Soudanais. Ça c'est une contribution pour la construction d'une base de données du DNA mitochondrial pour les populations d'Est d'Afrique, en obéissant aux critères de qualité de la génétique criminelle. La probabilité de *matches* de haplotypes au Soudan est très basse: 2% dans HVRI; 4% dans HVRII; 1,8% dans HVRI+HVRII. Même à l'intérieur des plusieurs haplogroupes les haplotypes sont suffisamment diversifiés. La caractérisation de la séquence complète de mtDNA dans 21 haplotypes L3(xM, N) du Soudan et sa comparaison avec 94 séquences complètes a permis d'évaluer les taux de mutation relatifs des diverses positions dans ce haplogroupe, informations essentielles dans le développement de kits SNaPshot pour l'application dans le domaine de la génétique criminelle et des populations.

Dans le domaine de la génétique des populations les premières données du mtDNA sont venues soutenir le modèle Out-of-Africa pour l'origine de l'Homme Moderne, datant l'origine seule d'environ 200.000 ans et en la localisant, avec une élevée probabilité, dans l'est d'Afrique. Les données de séquençage complet permettent aujourd'hui d'aiguiller des datations et d'inférer des itinéraires de migration possibles, étant actuellement en débat une route de migration plus ancienne (environ 65.000 ans) à travers la Corne d'Afrique, Mer Rouge, Péninsule Arabique et dans la direction à la sud-est asiatique versus la route plus récente (environ 45.000 ans) à travers le Levant. La caractérisation de la séquence complète

du mtDNA dans 21 haplotypes L3(xM, N) et sa comparaison avec 94 séquences complètes a permis, dans ce travail, contribuer à un raffinement de la datation des lignages L3 dans l'origine de toute à la diversité non-Africain. L'âge estimé pour le haplogoupe L3 est 69.600 ± 5.900 ans. S'est vérifiée la considérable multifurcation depuis la racine de L3, en présentant tous les sub-haplogroupes un âge de l'ordre de grandeur du haplogroupe certifiant la forte explosion démographique produite à cette occasion.

Elles ne se trouvent pas des lignages M et N basal au Soudan, par lequel celles-ci doivent avoir été données lieu déjà dehors Afrique.

ABREVIATURAS

A-Adenina

ATP- Adenosina trifosfato

bp- pares de bases

C- Citosina

CoQ- Coenzima Q-10

COX1- Citocromo c oxidase subunidade I

COX2- Citocromo c oxidase subunidade II

COX3- Citocromo c oxidase subunidade III

CRS- *Cambridge Reference Sequence*

D-Loop- *Displacement Loop* da região controlo

DNA- Ácido desoxirribonucleico

G- Guanina

HVRI- Região hipervariável I

HVRII- Região hipervariável II

mtDNA- Ácido desoxirribonucleico mitocondrial

MN- Aglutinogénios M e N

mRNA- Ácido ribonucleico mensageiro

NADH- Dinucleotídeo de nicotinamida adenina hidrogenase

ND1- NADH desidrogenase subunidade 1

ND2- NADH desidrogenase subunidade 2

ND3- NADH desidrogenase subunidade 3

PCR- *Polimerase Chain Reaction*

RFLP- *Restriction Fragment Length Polymorphism*

Rh- Rhesus

RNA-Ácido ribonucleico

rRNA – Ácido ribonucleico ribossómico

SNP- *Single Nucleotide Polymorphism*

STR- *Short Tandem Repeats*

tRNA- Ácido ribonucleico de transferência

T- Timina

VNTR- *Variable Number of Tandem Repeats*

ÍNDICE

Resumo.....	4
Summary.....	6
Résumé.....	7
Abreviaturas.....	9
I – INTRODUÇÃO.....	12
1- DNA e genética forense.....	12
2 - Marcadores não recombinantes e de transmissão uniparental	15
3 - DNA mitocondrial.....	17
4 - Contributo do DNA mitocondrial para a Genética Populacional Humana.....	19
5 – Sudão e leste de África.....	24
6 – Objectivos.....	29
II - MATERIAL E MÉTODOS.....	31
1 - Amostras estudadas.....	31
2 - Extracção de DNA.....	31
3 - Amplificação do DNA por PCR.....	31
4 - Electroforese nativa em gel de poliacrilamida e visualização por coloração com nitrato de prata.....	34
5 - Sequenciação automática.....	34
6 - Afiliação em haplogrupos.....	36
7 - Populações usadas para comparação.....	38
8 - Análise estatística.....	39
III - RESULTADOS e DISCUSSÃO.....	44
1 - Diversidade haplotípica em HVRI e HVRII.....	44
2 - Probabilidade de <i>match a priori</i>	50
3 - Estruturação populacional baseada na diversidade haplotípica.....	54
4 - Diversidade de haplogrupos.....	57
5 - Filogenia das linhagens L3(xM,N) com base na sequência completa de mtDNA.....	66
6 - A recorrência na região codificante do mtDNA no haplogrupo L3(xM,N) e implicações para a sua aplicação forense.....	70

IV – CONCLUSÕES.....	74
1 - Aplicações à genética forense – a diversidade mitocondrial em populações do leste de África.....	74
2 - Aplicações à genética populacional – inferências para o Out-of-Africa.....	75
V - REFERÊNCIAS BIBLIOGRÁFICAS.....	79
VI-ANEXOS.....	89
Anexo 1: Sequências completas ou região codificante de L3(xM,N) e L4 para comparação com as obtidas neste trabalho.....	90
Anexo 2: Posições variáveis, por comparação com CRS, para HVRI (menos 16000) e HVRII.....	93
Anexo 3: Posições variáveis, por comparação com CRS, para a molécula inteira do mtDNA.....	97
Anexo 4: Espectro de taxas de mutação para a região codificante em 155 sequências L3(xM,N) e para a região controlo em apenas 74 dessas sequências.....	100

I - INTRODUÇÃO

1 - DNA e genética forense

De um modo análogo às inúmeras aplicações de diagnóstico molecular, a genética está a revolucionar as ciências forenses, permitindo acrescentar informação essencial à investigação forense quando existem provas biológicas.

A primeira abordagem de identificação genética consistia na análise de marcadores imunológicos, como o sistema sanguíneo ABO, por metodologias de imunologia. De facto, desde inícios do século XX que a descoberta deste sistema sanguíneo e a sua caracterização em diversos grupos populacionais mostrou a sua informatividade para a identificação genética dos indivíduos. Adicionalmente, foram identificados outros sistemas imunológicos, como o Rh e o MN.

Seguidamente, verificou-se ser possível o estudo de marcadores proteicos, cuja distinção entre isoformas era possível devida a diferenças de mobilidade electroforética consoante o pH do sistema de tipagem.

Ambos estes marcadores eram contudo limitados, possuindo uma variabilidade reduzida. A limitação da variabilidade genética foi contornada com a tipagem directa de DNA, cuja aplicação no campo das ciências forenses ocorreu por volta de meados dos anos 80. Em 1985, Alec Jeffreys descreveu a tecnologia do DNA *fingerprinting*, a qual é altamente informativa para a identificação humana (Jeffreys *et al.*, 1985). Esta técnica tira partido da elevada variabilidade existente entre os indivíduos em algumas regiões genómicas designadas por minisatélites ou VNTRs (abreviatura de *Variable Number of Tandem Repeats*). Os VNTRs são regiões em que uma sequência de bases (podem ir de 10 a cerca de 100 pares de bases) está repetida sequencialmente várias vezes (em *tandem*), o número das quais é altamente variável entre os indivíduos. Quando o DNA de um determinado indivíduo é submetido ao corte por várias enzimas de restrição, vai ficar fragmentado em inúmeros blocos de tamanho diverso devido aos polimorfismos dos VNTRs e da própria actuação polimórfica das enzimas de restrição, gerando um padrão de bandas quando a amostra é separada em electroforese. Os polimorfismos de restrição são designados por RFLPs (de *Restriction Fragment Length Polymorphism*). A técnica do DNA *fingerprinting* permitia, assim, uma avaliação qualitativa por comparação visual dos padrões de bandas entre suspeitos e prova biológica.

Estas primeiras técnicas moleculares tinham as desvantagens de serem fastidiosas e ser necessária uma quantidade elevada de DNA, num bom estado de conservação. Estes factores limitavam grandemente a aplicação forense onde, na extensa maioria dos casos, a prova biológica se encontra em quantidade vestigial e em condições de degradação e contaminação.

Em 1985, Kary Mullis e co-autores apresentaram uma técnica conhecida por PCR – *Polimerase Chain Reaction*, que torna possível obter milhões de cópias de um determinado fragmento de DNA, a partir de uma pequena quantidade de DNA inicial. Esta metodologia revolucionou todas as áreas de diagnóstico e investigação em genética. No que concerne às ciências forenses, com o PCR iniciou-se a aplicação de um novo tipo de marcadores genéticos altamente polimórficos, os microsátélites também conhecidos por STRs (de *Short Tandem Repeats*), possuindo uma taxa de mutação de cerca de 10^{-3} por *locus* por ano (Brinkmann *et al.*, 1998). Os STRs são um tipo particular de VNTRs, em que o motivo repetitivo tem cerca de 2 a 6 bp, geralmente repetidos em número inferior a 50. Deste modo, os fragmentos analisáveis por PCR são pequenos, contendo aproximadamente entre 100 a 300bp, o que possibilita a obtenção de resultados mesmo a partir de amostras degradadas. Com o desenvolvimento concomitante dos métodos de detecção de DNA após electroforese, como seja a marcação por fluorescência, foi possível desenvolver sistemas multiplex de STRs, em que vários marcadores são estudados simultaneamente. Entre 1990 e 1996 apareceram os primeiros kits comerciais multiplex para STRs, que permitem a amplificação de até 15 loci, que reduzem substancialmente a quantidade de amostra necessária para a análise e minimiza o tempo necessário para a obtenção de resultados, bem como o seu custo e a possibilidade de erro por troca de amostras.

Recentemente, há um interesse crescente na utilização de outro tipo de marcadores genéticos, os SNP (*Single Nucleotide Polymorphism*) no campo da genética forense. Estes caracterizam-se por serem substituições nucleotídicas pontuais (sendo por vezes também incluídas inserções/deleções de uma a duas bases), as quais apresentam uma taxa de mutação de 10^{-8} por base por ano, para o genoma humano (Nachman *et al.*, 2000). A sua menor taxa de mutação, faz com que seja necessário o estudo de muitos mais marcadores para obter o mesmo poder discriminatório de 15 loci de STRs (Amorim e Pereira, 2005), mas têm a vantagem de não serem tão susceptíveis a mutação recorrente. Também permitem ser analisados em pequenos amplicons, crítico para o sucesso da análise de material degradado. Novos métodos de genotipagem estão a ser desenvolvidos para os SNP, como

as tecnologias de minisequenciação, (Sobrinho *et al.*, 2005), os microarrays ou a espectrometria de massa.

Os marcadores autossômicos permitem uma identificação individual porque um determinado perfil genético é específico de um só indivíduo, sendo apenas partilhado por gémeos monozigóticos.

Na validação de um dado sistema genético a ser implementado para aplicação forense é necessário fazer um estudo populacional prévio para caracterização das frequências genéticas. É indicado que cada laboratório caracterize a população na qual vai realizar o seu trabalho de investigação pericial ou, na impossibilidade, use as frequências de uma população próxima.

Esta caracterização da população permite a quantificação da prova genética: quantas vezes é mais provável que um dado perfil genético de uma prova pertença ao suspeito do que a um indivíduo ao acaso da população.

Os vários marcadores autossômicos estudados devem ter transmissão independente, devendo estar localizados em cromossomas diferentes ou em braços diferentes do mesmo cromossoma. Assim sendo, são passíveis de aplicação da regra do produto: os índices (sejam de paternidade ou de identidade entre prova e suspeito) individuais podem ser multiplicados para a obtenção de um índice global. E é possível determinar frequências genotípicas esperadas a partir das frequências alélicas observadas, pela aplicação do produto do binómio, logo que os marcadores estejam em equilíbrio de Hardy-Weinberg.

Curiosamente, se os marcadores autossômicos são poderosos para a identificação individual, já não possuem poder distintivo para a identificação de grupos populacionais. Os primeiros estudos de genética populacional, iniciados na década de 70 por Cavalli-Sforza e colaboradores, mostraram que já não existem alelos específicos das populações, havendo como que um pool genético contínuo, com gradientes de frequências dos alelos. Quando se avaliou a hierarquização da diversidade genética verificou-se que para STRs nucleares, 93,2% da variabilidade existe intra-população, e apenas 2,5% entre populações do mesmo grupo e 4,3% entre os grandes grupos populacionais (Belledi *et al.*, 2000).

2 - Marcadores não recombinantes e de transmissão uniparental

A região não recombinante do cromossoma Y (compreendendo 95% da sua extensão) e o DNA mitocondrial (mtDNA) constituem uma outra categoria de marcadores genéticos, caracterizada por ausência de recombinação, haploidia e transmissão uniparental. O cromossoma Y é de transmissão unicamente masculina, definindo linhagens paternas e o mtDNA de transmissão unicamente feminina, definindo linhagens maternas.

Como tal, os marcadores uniparentais não permitem identificação individual. Todos os indivíduos aparentados por via masculina partilharão a mesma linhagem de cromossoma Y, enquanto todos os indivíduos aparentados por via feminina partilharão a mesma linhagem de mtDNA, salvo nos casos de ocorrência de mutações durante a transmissão. Ademais, a ausência de recombinação faz com que todos os marcadores genéticos no cromossoma Y e todos os do mtDNA sejam transmitidos, em bloco, ligados, o que impossibilita a previsão de frequências genotípicas e a aplicação da regra do produto. Existe alguma polémica em volta da quantificação da prova genética com base em marcadores genéticos não recombinantes. Geralmente, opta-se por uma posição prudente em que se avalia qualitativamente a identidade ou *match* entre linhagens da prova e suspeito e se informa que essa linhagem está presente na população referência numa frequência X em N indivíduos analisados. Mais do que permitir identificação, os marcadores uniparentais são muito informativos para a exclusão de suspeitos.

Existem bases de dados genéticas para os marcadores uniparentais, de cariz científico mas com cuidados mínimos de controlo de qualidade para que os dados possam ser usados como referência em pesquisas forenses. A mais antiga é a base de dados de STRs do cromossoma Y, que começou por ser europeia (Roewer *et al.*, 2001) e que se alargou a mundial (URL: www.yhrd.org). Os novos desenvolvimentos pretendem incluir marcadores SNPs do cromossoma Y. Para o mtDNA, a base EMPOP está finalmente a ser lançada, após a passagem por controlos apertados de qualidade dos dados aí depositados (URL: www.empop.org). Há também uma base de dados de mtDNA do FBI (Monson *et al.*, 2002), para a população dos Estados Unidos da América, estruturada nos grandes grupos populacionais existentes nesse país.

Apesar da grande desvantagem de não permitir identificação individual, é muitas vezes necessário recorrer ao estudo dos marcadores uniparentais nas ciências forenses. Quando a quantidade de material biológico é muito pequena (como talo de cabelo sem raiz, ossos antigos) ou está altamente degradado, a análise de mtDNA pode ser a única fonte

possível de resultados. A razão para tal é a existência de um número elevado de cópias de mtDNA por célula, desde 700-1200 nos espermatozóides (Díez-Sánchez *et al.*, 2003) até 100.000 nos oócitos (Jenuth *et al.*, 1996).

Em genética forense, é comum aparecerem misturas de DNA devido à contribuição de vários indivíduos ou a contaminações. É o caso das amostras colhidas após agressões sexuais, geralmente contendo uma mistura de DNA masculino e feminino. Utilizando os marcadores do cromossoma Y, pode ser determinado o perfil do agressor. Os marcadores para o cromossoma Y são também usados em testes de paternidade, investigações históricas ou na investigação de pessoas desaparecidas, quando só existem parentes afastados do lado paterno para serem dadores de material de referência.

O facto do efectivo populacional para os marcadores uniparentais ser apenas de $\frac{1}{4}$ relativamente ao dos marcadores autossómicos, torna-os muito mais sensíveis aos efeitos demográficos, como efeitos fundadores, *bottlenecks*, expansões e deriva genética. Os primeiros estudos de marcadores uniparentais efectuados em ampla escala geográfica atestaram a sua mais elevada estruturação geográfica, permitindo uma maior capacidade de distinção entre populações. Belledi *et al.* (2000) mostraram que a percentagem de variação observada entre populações dum mesmo grande grupo populacional era de 12,3% para uma região do mtDNA, 20,6% para STRs do cromossoma Y e 16,9% para SNPs do cromossoma Y, e entre os grandes grupos populacionais de 16,4%, 20,3% e 36,9%, respectivamente.

Deste modo, quando em genética forense se procura alguma informação sobre a afiliação populacional de uma dada amostra pode-se recorrer à caracterização da linhagem materna e, nos indivíduos do sexo masculino, da linhagem paterna. Com as devidas ressalvas de o mtDNA e o cromossoma Y representarem uma porção mínima do genoma total do indivíduo.

Os contributos dos marcadores uniparentais foram importantíssimos na genética populacional. Finalmente dispunha-se de marcadores que permitiam distinguir grupos populacionais e que eram ferramentas informativas para a reconstrução de filogenias, inferência e datação de migrações. De facto, a ausência de recombinação mantém a história evolutiva nos haplótipos, quer do cromossoma Y quer do mtDNA.

Desde a década de 90 que os estudos populacionais efectuados se alicerçaram no poder informativo dos marcadores uniparentais. Era possível trazer achegas genéticas a hipóteses arqueológicas e antropológicas de migrações pré-históricas e históricas humanas e levar à formulação de novas hipóteses. O mtDNA podia mesmo ser estudado em

indivíduos de populações extintas, como os Neandertais (Kriings *et al.*, 1997) e o estudo de DNA antigo continua a ser uma possibilidade apesar dos reveses de demonstração que as transformações químicas do DNA poderiam mimetizar a evolução (Gilbert, 2006). O impulso foi tão grande que levou à criação do termo Arqueogenética para especificar esta comunicação entre Genética e Arqueologia (Amorim 2000; Renfrew 2000).

3 - DNA mitocondrial

As mitocôndrias são pequenos organelos (1-2µm de comprimento e 0,5-1µm de largura) presentes no citoplasma da maioria das células eucarióticas. Estes organelos são responsáveis, em condições aeróbias, pela obtenção da maior parte da energia necessária às células que as possuem. Delimitadas por duas membranas, uma interna e outra externa, contêm várias cópias de um genoma próprio, o mtDNA. Ao contrário do DNA nuclear que apenas se duplica durante o ciclo celular, o mtDNA sofre duplicação independentemente do ciclo celular (Birky, 2001).

O mtDNA (Figura 1) consiste numa molécula circular de cadeia dupla, sendo uma das cadeias designada por cadeia pesada ou H (de *high*) e a outra por cadeia leve ou L (de *light*). Tais designações reflectem o facto de as cadeias apresentarem diferente densidade num gradiente de céσιο, conferida por uma proporção desigual dos 4 nucleotídeos nas duas cadeias, predominando as adeninas e guaninas na cadeia H e as citosinas e timinas na cadeia L.

A sequência completa do mtDNA foi publicada em 1981 por Anderson (Anderson *et al.*, 1981), ficando conhecida como CRS – *Cambridge Reference Sequence*. Foi revista em 1999 por Andrews *et al.*, tendo-se encontrado algumas discrepâncias, inclusivé de tamanho, mas para não introduzir mais erros, optou-se por manter a numeração inicial.

A molécula de mtDNA humano apresenta 16569 pares de bases e codifica 2 RNAs encontrados nos ribossomas mitocondriais (12s rRNA e 16s rRNA), 22 tRNAs envolvidos na tradução dos 13 genes codificantes de proteínas da cadeia respiratória oxidativa. Todos estes genes foram identificados, sendo três subunidades do complexo citocromo c oxidase (COX1, COX2 e COX3), dois subunidades da ATPase (ATP6 e ATP8), sete subunidades do complexo NADH – CoQ reductase (ND1, ND2, ND3, ND4, ND4L, ND5 e ND6) e o citocromo b é uma subunidade do complexo CoQ – citocromo c reductase. Todos estes 37 genes e

umas poucas bases não codificantes e o controle de replicação da cadeia L (OL) constituem a região codificante do mtDNA, entre as bases 577 e 16023.

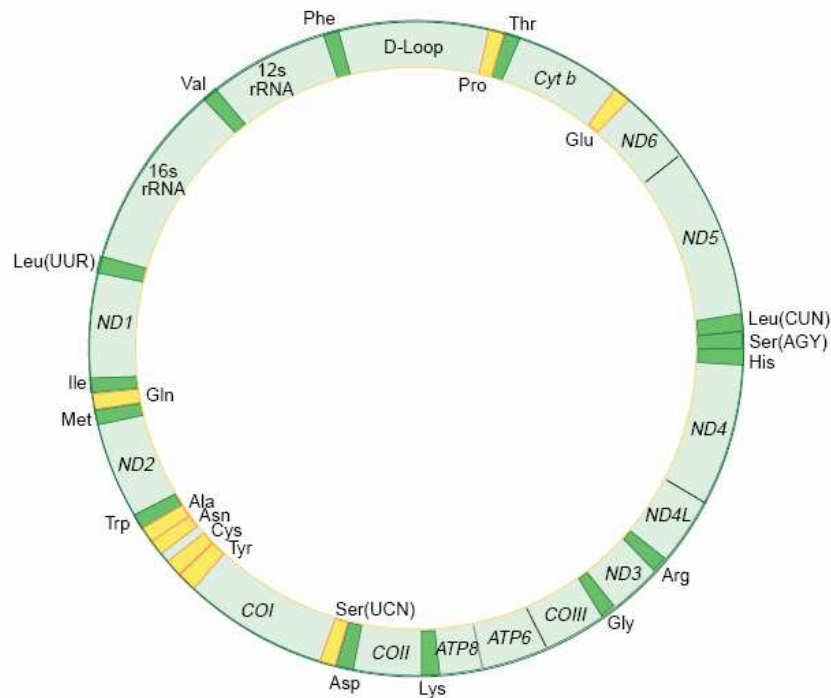


Figura 1 – Representação gráfica da molécula de mtDNA humano, evidenciando as localizações dos diversos genes na região codificante e o D-Loop ou região controle.

As restantes proteínas mitocondriais (pensa-se que sejam por volta de 90%) são codificadas pelo DNA nuclear, sintetizadas no citoplasma e migram posteriormente para as mitocôndrias.

O mtDNA possui ainda uma região não codificante, designada região controle ou D-Loop, entre as posições 16024 e 576, relacionada com a regulação da transcrição e replicação da molécula. No D-Loop existem duas regiões que apresentam uma sequência altamente variável, a nível populacional, quando comparadas com o resto do genoma, sendo por isso designadas regiões hipervariáveis. A região hipervariável I (HVRI), corresponde à região entre as posições 16024 e 16383; a região hipervariável II (HVRII), com 268 bp, está localizada entre as posições 73 e 340. A extensão destas regiões pode ser diferente nos vários estudos, podendo a HVRI estar entre 16024-16365 ou 16090-16365. Em muitos trabalhos são ainda referidas as posições 16390, 16399, 16400 e 72, importantes para a identificação de algumas linhagens.

A porção codificante do mtDNA possui uma taxa de mutação, aproximadamente, dez vezes maior que a taxa de mutação do DNA nuclear (Brown *et al.*, 1979). A região controle tem uma taxa de mutação cinco vezes maior que o resto da molécula na sua totalidade (Ingman *et al.*, 2000) sendo dez vezes maior na região hipervariável I (Vigilant *et al.*, 1991). A taxa de mutação mitocondrial é mais elevada por vários motivos: (1) o facto de o mtDNA não estar envolvido por proteínas – as histonas – tal como o DNA nuclear está; (2) não ter mecanismos de reparação de mutações tão eficientes como o DNA nuclear; (3) durante o processo de produção de energia são produzidos os radicais livres, agentes oxidantes, que podem danificar as moléculas de DNA.

As mutações pontuais podem ser inserções/deleções de bases, sendo muito frequentes nas regiões homopoliméricas poli-C, localizadas entre as posições 302-310 e 16183-16189, e substituições. As substituições, por sua vez, podem ser transições, quando uma pirimidina é substituída por outra ou uma purina por outra (T/C e G/A) ou transversões, quando uma purina é substituída por uma pirimidina ou vice-versa (A/C; G/C; G/T; T/A). As transições são muito mais frequentes que as transversões, sendo os valores para a espécie humana de cerca de 13,75:1 (Belle *et al.*, 2005).

Em mamíferos e na maioria dos outros animais, durante a fecundação, o espermatozóide contribui com muito pouco (ou com nenhum) citoplasma para o zigoto. E há evidências recentes de que as mitocôndrias paternas que penetram no oócito são destruídas por um processo químico, envolvendo provavelmente a ubiquinona, processo este que parece ser específico da espécie, não ocorrendo em cruzamentos experimentais entre espécies próximas (St John, 2002). De modo que, salvo a publicação da presença de mitocôndrias de origem paterna cujo DNA teria adquirido uma deleção de duas bases no gene da proteína ND2, no músculo de um paciente, não há evidência para não considerar a transmissão das mitocôndrias como exclusivamente materna (Schwartz e Vissing, 2002; Kraytsberg *et al.*, 2004).

4 – Contributo do DNA mitocondrial para a Genética Populacional Humana

Como já foi referido, o estudo do mtDNA adquiriu importância crescente em várias áreas, tais como estudos populacionais, clínicos e de genética forense. Começa agora a haver algum intercâmbio entre estas diferentes áreas, o que é essencial para a integração de toda a informação para uma melhor compreensão da dinâmica deste genoma peculiar.

Os genomas actuais derivam de genomas passados por mutações acumuladas ao longo do tempo. Nas porções do genoma, como é o caso do mtDNA, onde não há recombinação que misture os eventos históricos, tornam-se mais fáceis a reconstrução e a datação filogenética. O mtDNA humano funciona como um registo molecular da história genealógica e das migrações de mulheres que transmitiram o seu mtDNA ao longo de várias gerações. Uma vez que o processo de diferenciação molecular ocorreu relativamente rápido e, essencialmente, durante ou após o processo de dispersão para diferentes partes do mundo, conseqüentes variações no mtDNA tendem a ser específicas para determinadas áreas geográficas (Torroni *et al.*, 2006).

Obviamente, esta forma de reconstruir o passado está restringida àqueles que foram reprodutivamente bem sucedidos. As linhagens que não contribuíram para o pool genético actual não podem ser inferidas neste tipo de estudo. Contudo, a evolução das técnicas utilizadas permite, actualmente, estudar DNA fóssil de vestígios humanos, como ossos, permitindo a caracterização de DNA antigo, independentemente de terem contribuído ou não para o pool genético actual.

O primeiro estudo de mtDNA recorreu à digestão de toda a molécula com apenas algumas enzimas de restrição em 21 amostras, com origens geográficas e étnicas diferentes. Este estudo indicou que os padrões obtidos através da digestão com enzimas de restrição, os RFLP (*Restriction Fragment Length Polymorphisms*), poderiam ser usados para traçar a história genética humana (Brown, 1980). O aperfeiçoamento deste método tornou-se promissor para estudos na genética populacional.

Contudo, o estudo que impulsionou a utilização do mtDNA em genética populacional foi publicado por Cann *et al.* (1987). A análise de 147 amostras oriundas de várias regiões geográficas demonstrou que as amostras africanas possuíam maior diversidade do que as restantes e se localizavam na raiz da árvore filogenética. Tal facto foi interpretado como uma clara evidência de que a população africana era a mais antiga de todas, sendo favorável à hipótese uniregional da origem do Homem Moderno. Esta hipótese propõe que as populações humanas resultam de uma diferenciação, relativamente recente, a partir de uma única população ancestral originária de África. O estudo permitiu ainda inferir o tempo decorrido desde a separação das linhagens mitocondriais, cerca de 200.000 anos.

No início dos anos 90, houve um grande aumento dos estudos de genética populacional, recorrendo à análise do mtDNA. Estes estudos basearam-se em RFLPs, distribuídos por toda a molécula de mtDNA, mas rapidamente foram substituídos pela sequenciação de HVRI. Contudo, a variação observada nesta última região é acompanhada

por um nível elevado de mutações recorrentes, o que origina reticulação, logo incerteza, na árvore do mtDNA. A combinação de RFLP de alta-resolução e sequenciação de HVRI permite uma reconstrução filogenética mais segura, passando a ser a estratégia privilegiada (Macaulay *et al.*, 1999).

Com a primeira publicação de 53 genomas mitocondriais humanos de vários locais geográficos (Ingman *et al.*, 2000), a metodologia de aplicação do mtDNA à genética populacional entrou numa nova fase. É agora comum fazer uma caracterização prévia da diversidade do mtDNA numa amostra populacional pela sequenciação de HVRI, o que permite a selecção de algumas linhagens interessantes para uma caracterização mais aprofundada por sequenciação de toda a molécula.

Um conceito fundamental na genética populacional mitocondrial foi introduzido por Torroni *et al.* (1993), num estudo da população de Americanos Nativos, que visava esclarecer a origem, tempo e dimensão das migrações ancestrais para o Novo Mundo a partir da Ásia. O conceito é haplogrupo, sendo este um grupo de haplótipos de mtDNA derivados por descendência da mesma molécula ancestral, possuindo o mesmo padrão basal de mutações. A nomenclatura dos haplogrupos do mtDNA foi então iniciada com o estabelecimento de quatro ramos basais na árvore do mtDNA, denominados alfabeticamente como A, B, C, e D. Mais tarde, os haplogrupos de outras populações de diferentes continentes foram caracterizados, seguindo-se a sua designação por letras. Subgrupos dentro de um haplogrupo, subhaplogrupos, são designados por números, como por exemplo os subhaplogrupos H1 e H2, pertencentes ao haplogrupo H. Ambos estes subhaplogrupos partilham os polimorfismos que caracterizam H, mas todos os indivíduos H1 partilham um ou vários polimorfismos adicionais, que são diferentes dos polimorfismos partilhados pelos indivíduos pertencentes a H2. A nomenclatura para sub-agrupamentos adicionais segue a alternativa letra-número-letra-número, em ordem crescente alfabética e numérica a partir do último grupo publicado. O símbolo asterisco é utilizado para denotar membros de um haplogrupo que não pertencem a nenhum dos subhaplogrupos já denominados, por exemplo, HV* apresenta polimorfismos que não o permitem classificar nem como HV1, R0 ou qualquer um dos subhaplogrupos H.

O debate sobre a origem dos humanos modernos e a forma como estes evoluíram do *Homo erectus* foi, por longo período, controverso, levando à criação de várias hipóteses, sendo a multiregional e o Out-of-Africa as mais debatidas.

De acordo com a hipótese multiregional, os humanos modernos surgiram em diferentes partes do Velho Mundo a partir de populações locais de *Homo erectus*, sendo a continuidade genética entre as populações mantida por fluxo genético. Este modelo implicaria efectivos populacionais grandes e uma significativa proximidade geográfica, de forma a permitir entrecruzamento. Uma consequência deste modelo seria uma diversidade genética e estruturação geográfica maior que a actualmente existente.

A hipótese do Out-of-Africa, contrariamente à hipótese anterior que assume uma evolução paralela, sugere que o *Homo sapiens* surgiu há ~100.000 - 200.000 anos, a partir de um número reduzido de ancestrais (~10.000 adultos) em África. Posteriormente, uma série de migrações permitiram a colonização do resto do Mundo. O estabelecimento dos humanos modernos foi de há ~90.000 anos no Próximo Oriente, 40.000 anos na Europa e Ásia, 50.000 – 60.000 na Austrália e finalmente, 14.000 na América (apesar de alguns autores acreditarem que a colonização do Novo Mundo foi muito mais antiga, há cerca de 35.000 anos). Actualmente, o modelo do Out-of-Africa é de aceitação praticamente unânime.

Os dados de mtDNA vieram apoiar esse ancestral comum à diversidade actual, original do leste de África e com cerca de 200.000 anos. Daqui migrou, colonizando o resto do mundo. A linhagem mitocondrial da qual derivou toda a diversidade observada fora de África é a L3, dando a sua ramificação origem, entre outros, a dois super-haplogrupos, M e N, a partir dos quais se ramificaram vários haplogrupos. O haplogrupo N deu origem ao R. M é mais comum no Leste e Sul Asiático, enquanto N e R predominam no Próximo Oriente e Eurásia. (Torroni *et al.*, 2006).

O novo desenvolvimento tecnológico na aplicação dos marcadores uniparentais na genética populacional está também a contribuir para o desenvolvimento do modelo do “Out of Africa”. O modelo tradicional coloca o ancestral no leste da África Subsariana, seguido por dispersão via Levante há cerca de 45.000 anos atrás (Stringer e Andrews 1988). Contudo, estão a surgir evidências a favor de uma rota de dispersão mais antiga, há cerca de 63.000 anos atrás, mais a sul, através do Corno de África, ao longo da costa tropical do Oceano Índico, em direcção ao sudeste da Ásia e Australásia. De facto, em 2005, estudos de sequenciação completa do mtDNA em duas populações “reliquia”, das ilhas Andaman e Nicobar, no sul da Índia (Thangaraj *et al.*, 2005) e de Orang Asli, na Malásia (Macaulay *et al.*, 2005), ditas descenderem dessa dispersão mais antiga, suportaram a hipótese de uma única rota costeira através do sul da Ásia.

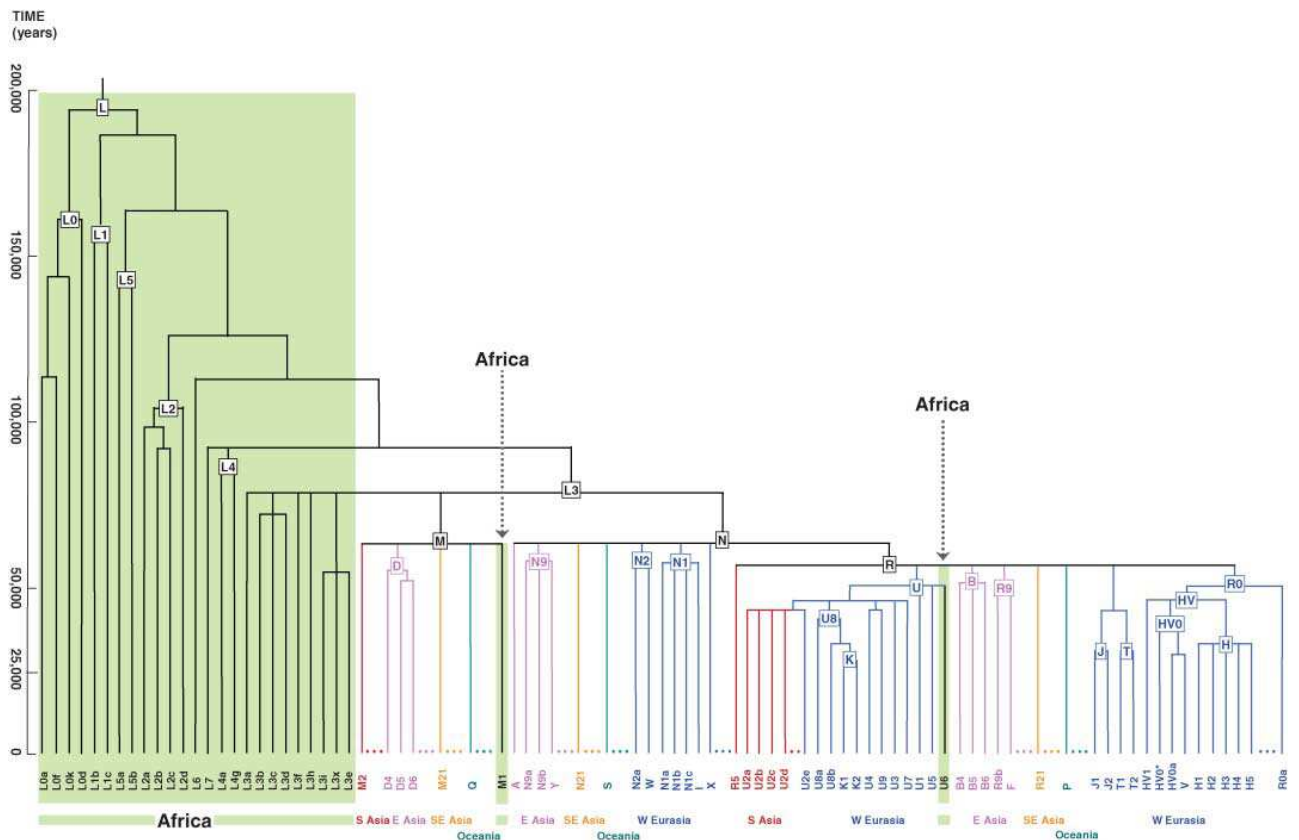


Figura 2 – Representação esquemática da filogenia do mtDNA. Os haplogrupos africanos estão representados a verde e as outras áreas geográficas a diferentes cores. A árvore indica um ancestral comum à diversidade actual, original de África. A diversidade observada fora do continente africano derivou do subhaplogrupo L3, dando a sua ramificação origem aos super-haplogrupos M e N.

Aquela migração mais antiga através do sul da Península Arábica, a ter ocorrido antes de 60.000 anos atrás, não poderia ter ascendido na Península, até ao Próximo Oriente, devido às condições climatéricas da altura. Quase toda a Península seria extremamente árida, o que pode ter contribuído para a migração em direcção a este e sempre através da costa. Os autores (Macaulay *et al.*, 2005), comparando a data obtida para o ancestral comum mais recente no haplogrupo M na Índia em 66.000 anos e 63.000 anos na Australásia, distantes em cerca de 12.000 km, estimaram uma taxa de migração de cerca de 4km/ano. Esta taxa é muito superior à taxa de 0.7km/ano calculada para a re-expansão na Europa após o Último Máximo Glaciar (em Macaulay *et al.*, 2005).

Mas os passos iniciais dessa dispersão permanecem ainda por esclarecer. Um esforço semelhante deve ser aplicado à fonte da migração, ou seja, ao Leste de África, e à

rota de passagem imediata, quer seja Levante, Sul da Arábia ou ambos. Poderão M e N representar, respectivamente, estas diferentes migrações?

Já existem alguns trabalhos publicados para o Leste de África. O mais completo, contendo já sequenciação completa de algumas linhagens, analisou uma amostra da Etiópia, bem como outra do Iémen na Península Arábica (Kivisild *et al.*, 2004). Recentemente, foram publicados dados de toda a região controlada numa amostra do Quênia (Brandstätter *et al.*, 2004). Existem dados já antigos, apenas de HVRI, numa amostra da Núbia, na fronteira entre Egipto e Sudão, bem como de outra amostra do Egipto (Krings *et al.*, 1999). Falta um estudo alargado, com sequenciação de linhagens relevantes, no Sudão e Somália.

Quanto à Península Arábica, têm sido publicados alguns estudos recentes, com dados de HVRI e algumas sequências completas na Arábia Saudita (Abu-Amero *et al.*, 2006) e do Iémen (Kivisild *et al.*, 2004).

5 – Sudão e leste de África

O Sudão localiza-se no leste de África, sendo o maior país em área de África (2.505.810 Km²; o 10º país maior em área do mundo). Tem fronteiras a norte com o Egipto, a nordeste com o Mar Vermelho, a leste com a Eritreia e a Etiópia, a sudeste com o Quênia e o Uganda, a sudoeste com a República Democrática do Congo e a República Centro-Africana, a oeste com o Chad e a noroeste com a Líbia (Figura 3). A estimativa de 2007 para a população no Sudão é de cerca de 39.992.490, o que o torna o 33º país com o maior número de habitantes; a densidade populacional na zona metropolitana de Cartum está a crescer rapidamente, estimando-se em cerca de 5 a 7 milhões, dos quais 2 milhões são refugiados das zonas de conflito a sul e das zonas afectadas pelas secas a oeste e leste.

A costa com o Mar Vermelho alonga-se por 853 Km, e a principal fonte fluvial é o Nilo e seus tributários. Os Nilos azul (um curso de 800 Km ao longo do Sudão) e branco encontram-se em Cartum para formar o rio Nilo, que corre para norte, através do Egipto, em direcção ao Mar Mediterrâneo. O terreno é em geral planície, quebrada por alguns maciços montanhosos: a oeste a cadeia Jebel Marra, no sul a montanha mais alta é o Monte Kinyeti Imatoing e a leste os planaltos do Mar Vermelho. A taxa de queda de chuva aumenta em direcção a sul, sendo o norte dominado pelo Deserto da Núbia e o sul por pântanos e floresta tropical. Nas áreas semi-desérticas do norte e oeste, a subsistência consiste na

agricultura básica, sendo muitas populações nômadas, migrando com os seus rebanhos de ovelhas e camelos. Uma agricultura excedentária é possível perto do rio Nilo e do lago Núbia, na fronteira com o Egito. O Sudão é ainda rico em minerais, incluindo petróleo, gás natural, ouro e prata, mas a sua economia atravessa sérios problemas, sendo o rendimento *per capita* muito baixo.

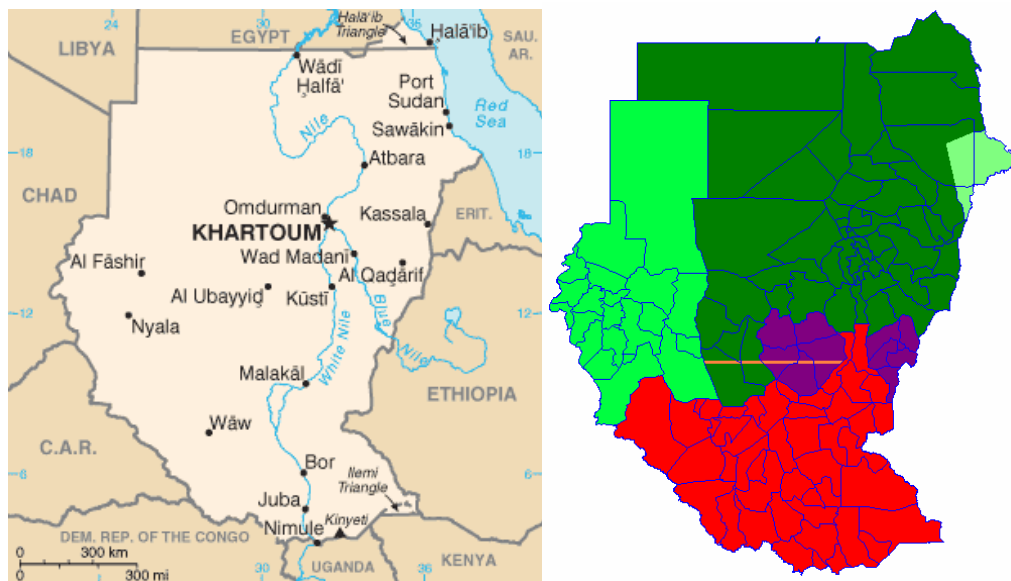


Figura 3 – Mapas do Sudão, mostrando a sua localização em África e a sua divisão em regiões (Norte - verde escuro; Darfur – verde intermédio; Frente leste – verde claro; Sul – vermelho; Montanhas Nuba e Nilo azul – roxo).

O nome Sudão deriva do árabe *Bilad-al-sudan*, que significa o “país dos negros”. Cerca de 597 tribos, com 400 dialectos diferentes, agrupam-se nas duas maiores culturas que constituem o Sudão: árabes com origens na Núbia e africanos não árabes.

Os estados do norte (Figura 3) cobrem a maioria do Sudão e incluem quase todos os centros urbanos. Os seus cerca de 22 milhões de habitantes são muçulmanos que falam árabe, organizados em vários grupos tribais.

A região Sul tem uma população de cerca de 6 milhões e é predominantemente rural, com uma economia de subsistência. Esta região foi afectada pela Guerra desde a independência em 1956, excepto num período de 10 anos. Está negligenciada, tendo mais de 4 milhões de refugiados e uma taxa de mortalidade de 2 milhões de pessoas. A maioria

desta população pratica crenças indígenas tradicionais e alguns são cristãos. A sociedade também está organizada em muitas tribos, com muitos dialectos. A língua franca é uma variante do árabe chamada árabe Juba e o inglês é usado pela elite.

Algumas tribos africanas do oeste, como os Fallata, também conhecidos por Fulani e Hausa, migraram para o Sudão há muito tempo atrás e estabeleceram-se em várias regiões, principalmente no norte, falando a maioria árabe e a sua língua original.

Evidências arqueológicas confirmaram que a região norte do Sudão estava habitada há pelo menos 60.000 anos atrás. Ao longo dos últimos 100.000 anos, o Vale do Nilo e o Sahara passaram por várias fases pluviosas intercaladas por fases de seca, o que inevitavelmente contribuiu para a alteração do habitat humano nesta região (Bubbenzer e Riemer, 2007). Desde há 25.000-8.000 anos atrás, o meio ambiente nesta zona, evoluiu gradualmente para o que é actualmente. Deste período, foram encontrados vários locais de assentamento das populações na zona de fronteira entre o Egipto e o Sudão. Parece terem sido locais de ocupação sazonal de populações semi-nómadas.

Uma cultura sedentária apareceu por volta de 8.000 anos A.C. (Edwards, 2007), com aldeias fortificadas, subsistindo pela caça, pesca e apanha de grão e criação de gado. A região era conhecida para os Egípcios por Cush, partilhando com aqueles fortes laços culturais e religiosos. Contudo, no oitavo século A.C., os Cushitas ficaram sob o domínio de uma linha monárquica agressiva, os quais, governando a partir da capital Napata, estenderam gradualmente a sua influência até ao Egipto. Cerca de 750 A.C., o rei Cushita Kashta conquistou o Egipto Superior, tornando-se governador de Tebes até 740 A.C. O seu sucessor, Piankhy, conquistou o delta, reunindo o Egipto sob a vigésima-quinta dinastia, que durou cerca de 100 anos, quando uma intervenção na área da moderna Síria, seguida de retaliação dos Assírios levou a que a dinastia voltasse a Napata, continuando a dominar o Cush e estendendo os domínios a sul e leste.

Em 590 A.C., um exército egípcio invadiu Napata, levando a que a corte Cushita se movesse para Meroé, perto da sexta catarata, criando-se o reino Meroítico. Este, no seu auge, cerca do III e II séculos A.C., estendia-se da terceira catarata até Sawba, perto da actual Cartum. Apesar de independentes do Egipto, a tradição faraónica persistiu entre os governantes Meroíticos, que tinham estelas para registar as suas conquistas e erigiam pirâmides para conter as suas tumbas. Estes objectos, juntamente com as ruínas de palácios, templos e banhos, atestam um sistema centralizado e político, com estruturação profissional e trabalho forçado. Um sistema de irrigação bem mantido permitiu uma

densidade populacional superior à de períodos mais tardios. Pelo século I A.C., o uso de hieróglifos deu origem a uma escrita Meroítica que adaptou o sistema de escrita egípcio a uma língua indígena, já relacionada com a língua da Núbia falada mais tarde. No século II D.C., o povo conhecido como Nobatae ocupou o banco oeste do Nilo, no norte de Cush, casando e estabelecendo-se entre os Meroíticos como aristocracia militar. Em 350 D.C., o exército abissínio conquistou Meroé, terminando a independência do reino.

No século VI, três estados emergiram como herdeiros políticos e culturais dos Meroítas: Nobatia ou Ballanah no norte; Muqurra no centro; e Alawa na velha Meroé. O governo dos três reinos estava a cargo da aristocracia guerreira. Um missionário enviado em 540 D.C. pela imperatriz bizantina iniciou a conversão dos reis Núbios em cristãos monofisitas.

Após muitas tentativas falhadas de conquistas militares, o comandante árabe do Egito concluiu o primeiro de uma série de tratados renováveis com os Núbios, que estabeleceram as relações entre os dois povos durante mais de 600 anos. O Islão progrediu na área ao longo de um longo período de tempo através de casamentos e contactos com mercadores e colonos árabes. Em 1315, um príncipe muçulmano de sangue real Núbio subiu ao trono. Os dois mais importantes grupos árabes a emergir na Núbia foram os Jaali e os Juhayna. Ambos os grupos descendiam da população local pré-islâmica. Actualmente, a cultura do Sudão do norte combina elementos Núbios e Árabes.

Em 1600, o povo Funj apareceu no sul da Núbia e suplantou os redutos do velho reino cristão de Alawa, estabelecendo o “Sultanado Azul”, com capital no Sinnar. Em 1820, Muhammad Ali do Egito invadiu o Sudão, tendo recebido a rendição do último sultão Funj. Deste modo, o Sudão ficou sob o domínio do Egito de 1821-1885, enquanto este, por sua vez, estava sob domínio Otomano. As autoridades egípcias melhoraram muito as infra estruturas do Sudão, especialmente no norte e no que respeitava à irrigação e produção de algodão. A partir de 1882, o Egito e o Sudão ficaram sob governação de Inglaterra, mas uma revolta do líder religioso Mahdi levou à retirada das forças egípcias e britânicas do Sudão em 1885, ficando sob domínio de um curto estado teocrático, o Mahdiya, que impôs leis islâmicas tradicionais. Mas muitas das tensões tribais mantiveram-se durante este período. Em 1890, os ingleses reconquistaram o Sudão, tendo sido governado como uma colónia britânica, separado administrativamente em 2 territórios, o norte e o sul. A independência ocorreria em 1956 (a do Egito tinha ocorrido em 1936). Um ano antes iniciava-se a primeira guerra civil sudanesa entre o norte e o sul, porque este receava o domínio pelo norte aquando da independência. Historicamente o norte do Sudão tinha laços

mais fortes com o Egito e era predominantemente árabe e muçulmano, enquanto o sul era uma mistura de cristãos e Animismas; esta divisão foi ainda enfatizada pela divisão administrativa dos ingleses. Esta primeira guerra prolongou-se de 1955 a 1972. Em 1983, iniciou-se a segunda guerra civil sudanesa, entre o norte e o sul, tendo durado mais de 20 anos (acordo de paz em 2005, com autonomia do sul por 6 anos, após o que haverá referendo para avaliar desejo de independência) e resultando em imensas mortes, refugiados e fome.

É importante distinguir os árabes sudaneses dos árabes do Médio Oriente. Os árabes sudaneses descendem, em primeiro lugar, dos Núbios antigos, dos quais não se sabe a origem. Durante um período de séculos, a imigração árabe para o Sudão, o entrecruzamento Núbios-árabes e a introdução do islão e da língua árabe, arabizaram os Núbios, tornando-os os árabes sudaneses actuais. Fisicamente, os Núbios são semelhantes a habitantes da Etiópia e Eritreia. Os árabes sudaneses são ainda divididos em muitas tribos de origem Núbia ou Árabe e alguns sudaneses falam uma língua Núbia.

Quando os conflitos entre norte e sul se resolviam, alguns confrontos tribais ocorriam na região oeste do Darfur, nos inícios de 1970s, entre tribos pastoris e tribos agrícolas, após a grande fome africana. Os rebeldes acusaram o governo central de negligenciar economicamente o Darfur. Milícias árabes armadas pela administração para parar as disputas entre as tribos têm sido acusadas de praticar genocídio. A luta levou à migração de centenas de milhares de pessoas, muitos refugiando-se no Chad. O governo reclamou vitória em 1994, mas em 2003 as lutas reiniciaram-se. O povo do Darfur consiste predominantemente em africanos negros de crença muçulmana, enquanto as milícias são árabes. As Nações Unidas estimam que mais de 3,5 milhões de pessoas no Darfur dependem fortemente da ajuda humanitária para a sua sobrevivência.

A frente leste é uma coligação de rebeldes operando no Sudão leste, na fronteira com a Eritreia. Reclamam mais igualdade no governo e ameaçam boicotar a saída do petróleo. O acordo de paz foi assinado em 2006.

A nível arqueológico, existe uma grande lacuna no registo fóssil de exemplares da transição de *Homo erectus* para *Homo sapiens* em África, ou seja, entre o período de há 300.000 a 100.000 anos atrás.

Em 2003 (White *et al.*, 2003), a caracterização de 3 crânios descobertos em Herto, Awash, Etiópia, veio diminuir esta lacuna, providenciando evidências para a localização, datação e contextualização da emergência do *Homo sapiens*. Os crânios foram datados de

160.000 a 154.000 anos atrás, e apresentam características morfológicas de transição entre as formas fósseis arcaicas africanas e as formas humanas anatomicamente modernas do Pleistoceno tardio, tendo sido classificados como uma subespécie extinta, *Homo sapiens idaltu*.

Também na Etiópia, perto do rio Omo, tinham sido descobertos nas décadas de 60 e 70 dois crânios de *Homo sapiens*. Estudos recentes de datação (McDougall *et al.*, 2005) mostraram tratarem-se dos mais velhos vestígios de Homem Moderno, com cerca de 190.000 anos.

Para além destes vestígios no leste de África, só foram encontrados vestígios contemporâneos daquelas datas na África do Sul, Marrocos e Próximo Oriente (revisão em Mellars, 2006). Na África do Sul, os vestígios foram encontrados em grutas na Província do Cabo: do rio Klasies, com cerca de 75.000-125.000 anos; de Blombos, com cerca de 75.000 a 130-140.000 anos; de Diepkloof, com 60.000-80.000 anos; de Bloomplaas entre 70.000 e 150.000 anos. No Norte de África, foram descobertos em 1991 sete fósseis de Homem Moderno em Jebel Irhoud, Marrocos, cuja datação revelou 160.000 anos de idade.

No Levante há vestígios de uma tentativa de migração empreendida por volta de 90.000-100.000 anos, tendo-se descoberto algumas ossadas de Homem Moderno em Skhul e Qafzeh, Israel. Estas ossadas estavam revestidas de ocre, logo estas populações já possuíam ritualidade e simbolismo, enterrando os seus mortos, mas estavam associadas a uma tecnologia lítica menos desenvolvida do que a observada noutros locais em África. Esta migração deve ter sido muito curta e localizada, tendo estas populações de homens modernos sido substituídas por Neandertais, pelo menos a partir de há 70.000 anos.

Na Península Arábica estão a ser iniciados estudos integrativos dos sítios paleolíticos para se construir uma visão geográfica e temporal da sua distribuição. Provavelmente irão ser necessárias pesquisas arqueológicas aquáticas porque o nível do mar actual está muito mais elevado do que há cerca de 65.000 anos atrás.

6- Objectivos

Este trabalho tem como objectivo a caracterização da diversidade mitocondrial numa amostra de 102 indivíduos naturais do Sudão, para aplicações de genética forense e populacional.

A nível forense, pretende-se obter uma base de dados, obedecendo aos critérios de qualidade forenses, da diversidade mitocondrial para as regiões hipervariáveis I e II (HVRI e HVRII). As bases de dados mitocondriais no Leste de África que poderiam ser usadas a nível forense são escassas, limitando-se actualmente ao Quénia e este trabalho vem pois diminuir essa lacuna. Pretende-se efectuar análises de vários parâmetros de diversidade mitocondrial, como a probabilidade de *match*, para as regiões HVRI e II, nestas populações do leste de África.

A nível da genética populacional, pretende-se caracterizar o *pool* genético mitocondrial da população do Sudão, explorando a informação filogenética que pode ser retirada da afiliação dos haplótipos em haplogrupos e da sua contextualização no panorama da África de leste e África Subsariana. Esta caracterização geral da diversidade mitocondrial sudanesa permitirá, particularmente, a identificação de linhagens L3(xM,N), cujos diversos haplótipos serão estudados aprofundadamente por sequenciação completa. A total caracterização desta diversidade dos haplótipos L3(xM,N) do Sudão, quando comparada com outros genomas completos de L3(xM,N) publicados na literatura, permitirão novas achegas para uma melhor resolução da diversidade de L3(xM,N) no Leste de África e seu contributo para a migração Out-of-Africa, bem como uma melhor datação das linhagens L3.

II - MATERIAL E MÉTODOS

1 - Amostras estudadas

Foram analisadas 102 amostras de indivíduos originários do Sudão, emigrados no Dubai, Estados Árabes Unidos. A origem dos indivíduos, bem como dos pais e avós foi confirmada por inquérito quando as amostras foram colhidas. A colheita de amostras obedeceu aos critérios básicos de consentimento informado.

O material biológico utilizado foi sangue, contido em papel (FTA cards-Whatman).

2 - Extração de DNA

A extração do DNA foi efectuada através do método de resina Chelex 100 (100-200 mesh, sodium form, biotechnology grade, Bio-Rad), que compreende vários passos (Lareu *et al.*, 1994). Primeiro efectua-se a lise dos glóbulos vermelhos (ruptura das membranas celulares e nucleares), a partir de um pequeno quadrado (com cerca de 1cm de lado) da mancha de sangue contida no papel FTA, pela adição de 1 mL de água desionizada esterilizada e a sua incubação, durante 30 minutos, à temperatura ambiente. Seguidamente, e após centrifugação a 17500g durante 4 minutos, o sobrenadante contendo os fragmentos celulares é rejeitado, de modo a recuperar a fracção dos glóbulos brancos contidos no sedimento. Ao sobrenadante, adiciona-se 200µL de uma solução de resina Chelex a 5% e faz-se a sua incubação a 56°C durante 30 minutos, com agitação posterior no vórtex. Segue-se uma nova incubação a 100°C durante 8 minutos para desnaturação das proteínas, com vista à eliminação da sua interferência nas reacções posteriores. Após nova agitação no vórtex, procede-se à centrifugação a 17500g durante 4 minutos.

O DNA resultante da extração, foi conservado a -20°C, para posterior amplificação.

3 - Amplificação do DNA por PCR

Em todas as 102 amostras foram analisadas as regiões HVRI e HVRII do DNA mitocondrial. Para a amplificação destas regiões foram usados dois pares de primers: para o HVRI, o L15997 (5'-CAC CAT TAG CAC CCA AAG CT-3') e H16401 (5'-TGA TTT CAC GGA

GGA TGG TG-3'); para o HVRII o L48 (5'-CTC ACG GGA GCT CTC CAT GC-3') e H408 (5'-CTG TTA AAA GTG CAT ACC GCC A-3').

Destas 102 amostras foram seleccionadas 22 amostras para sequenciação completa do mtDNA. Os primers (Tabela 1) foram desenhados de modo a obter 32 fragmentos de cerca de 600bp, com sobreposição para garantir que todas as posições são sequenciadas, segundo protocolo modificado de Maca-Meyer *et al.* (2001). O par de primers 32 corresponde aos do HVRI.

Na mistura da reacção para a amplificação das regiões HVRI, HVRII e todos os 32 fragmentos foi usada a seguinte composição: 12,9µL de água desionizada esterilizada, 0,5µL de desoxinucleótidos trifosfatados (dNTPs a 200mM), 2,5µL de tampão de PCR (10X), 2,5µL de cada primer (a 2,5mM), 0,1µL de Taq DNA polimerase (DNA polimerase do *Thermus aquaticus*). Para cada amostra a amplificar foram usados 4µL de DNA, num volume final de reacção de 25 µL.

As reacções de amplificação foram efectuadas num termociclador 2700 (AB Applied Biosystems) de acordo com as seguintes condições: desnaturação inicial a 95°C durante 2 min; 35 ciclos de desnaturação a 95°C durante 30 segundos, *annealing* a 60°C durante 30 segundos e extensão a 72°C durante 30 segundos; extensão final a 72°C durante 10 minutos.

Foi sempre usado, em todas as reacções de amplificação, um controlo negativo, em que a amostra foi substituída por água desionizada esterilizada.

Tabela 1 – Sequência (5'–3') dos *primers* utilizados para a sequenciação completa do mtDNA.

<i>Primer</i>	<i>Forward</i>	<i>Reverse</i>
H1	AGC CAT TTA CCG TAC ATA GCA CA	CTG TTA AAA GTG CAT ACC GCC A
H2	CAA AGA ACC CTA ACA CCA GCC	GGG AGG GGG TGA TCT AAA AC
H3	GTC ACA CGA TTA ACC CAA GTC A	GTA TAC TTG AGG AGG GTG ACG G
H4	GAG TGC TTA GTT GAA CAG GGC C	TTA GAG GGT TCT GTG GGC AAA
H5	GCC TGG TGA TAG CTG GTT GTC C	GGA ACA AGT GAT TAT GCT ACC T
H6	CAC CGC CTG CCC AGT GAC ACA T	TCG TAC AGG GAG GAA TTT GAA
H7	AAA GTC CTA CGT GAT CTG AGT TC	GGC GTA GTT TGA GTT TGA TGC
H8	TCT AGC CAC CTC TAG CCT AG	ATG CTG GAG ATT GTA ATG GGT
H9	CCA CTC ACC CTA GCA TTA CTT A	ACT CAG AAG TGA AAG GGG GCT A
H10	CCA ATA CTA CCA ATC AAT ACT C	GGT GAT GGT GGC TAT GAT GGT G
H11	TGG GCC ATT ATC GAA GAA TT	GAC AGG GGT TAG GCC TCT TT
H12	AAA GGC GGG AGA AGC CCC G	TGG CCC CTA AGA TAG AGG AGA
H13	CCT GGA GCC TCC GTA GAC CT	GCA CTG CAG CAG ATC ATT TC
H14	CCG GCG TCA AAG TAT TTA GC	GGG TTC TTC GAA TGT GTG GTA G
H15	AGA AGA ACC CTC CAT AAA CCT G	AGA TTA GTC CGC CGT AGT CG
H16	TCC CTC CCT TAC CAT CAA ATC A	TTT CAC TGT AAA GAG GTG TTG G
H17	ACC CCC TCT AGA GCC CAC TG	GAG CGA AAG CCT ATA ATC ACT G
H18	CTC GGA CTC CTG CCT CAC TCA	GTG GCC TTG GTA TGT GCT TT
H19	GGC CTA CTA ACC AAC ACA CTA	AAC CAC ATC TAC AAA ATG CCA GT
H20	TCC GCC AAC TAA TAT TTC ACT T	AAT GAG GGG CAT TTG GTA AA
H21	AAA GGA TTA GAC TGA ACC GAA	CCA TGA TTG TGA GGG GTA GG
H22	CTC CGA CCC CCT AAC AAC CC	CAA GGA AGG GGT AGG CTA TG
H23	AAA ACT AGG CGG CTA TGG TA	GGA GAA TGG GGG ATA GGT GT
H24	GGC TCA CTC ACC CAC CAC ATT	ACG AAC AAT GCT ACA GGG ATG AA
H25	ACA ACC CAG CTC TCC CTA AG	ATT TTC TGC TAG GGG GTG GA
H26	AGC CCT ACT CCA CTC AAG CAC	AGG GTG GGG TTA TTT TCG TT
H27	AAG CGC CTA TAG CAC TCG AA	TGG TTG AAC ATT GTT TGT TGG
H28	TCT TTC TTC TTC CCA CTC ATC C	CAT TGG TCG TGG TTG TAG TCC
H29	CCC CAT TAC TAA ACC CAC ACT C	TTG AAC TAG GTC TGT CCC AAT G
H30	CTC CCG TGA GGC CAA ATA TC	GTC TGC GGC TAG GAG TCA AT
H31	TCC CCA TCC TCC ATA TAT CC	TGA TGT GGA TTG GGT TTT TAT GTA
H32	CAC CAT TAG CAC CCA AAG CT	TGA TTT CAC GGA GGA TGG TG

4 – Electroforese nativa em gel de poliacrilamida e visualização por coloração com nitrato de prata

A qualidade do mtDNA amplificado foi confirmada em electroforese nativa num gel de poliacrilamida, na concentração T9C5 (9% de acrilamida e 5% desses 9% de crosslinker bisacrilamida). O tampão do gel é 0,375 M Tris/HCl, pH8,8 e o tampão das pontes é 0,125M Tris/Glicina, pH8,8.

O gel tem cerca de 3mm e para permitir o seu manuseamento é colocado a polimerizar sobre uma película hidrofílica (Gel-bond). O tampão das pontes é usado para embeber tiras (10cm x 2 cm) de papel Whatman que são colocadas directamente sobre o gel, no ânodo e no cátodo. O controlo da corrida é efectuado pela chegada ao ânodo do corante azul de bromofenol. O sistema electroforético é horizontal, em placa Multiphor (Pharmacia), com temperatura controlada a 4°C. As condições eléctricas foram controladas por uma voltagem fixa de 200V.

A etapa de coloração do gel com nitrato de prata compreende vários passos. Primeiramente, o DNA é fixado mergulhando o gel em etanol a 10% durante 10 minutos e depois em ácido nítrico a 1% durante 5 minutos. Em seguida, o gel é lavado com água desionizada (duas lavagens de cerca de 20 segundos cada) e imerso numa solução de nitrato de prata a 0,2%, durante 20 minutos, no escuro e em constante agitação. Os excessos de prata devem ser retirados lavando novamente o gel em água desionizada (duas lavagens de cerca de 20 segundos cada). Por último procede-se à revelação dos fragmentos de DNA, aos quais o nitrato de prata se ligou, pela adição de uma solução de carbonato de sódio a 0,28M e formaldeído a 0,02%. A reacção deve ser parada pela colocação do gel numa solução de ácido acético a 10% durante 30 segundos. Finalmente o gel é lavado com água e pode ser conservado depois de seco à temperatura ambiente.

5 – Sequenciação automática

As amostras com boa qualidade de amplificação podem seguir para sequenciação automática.

O primeiro passo consiste na purificação das amostras amplificadas, recorrendo a colunas Microspin S-300HR (Amersham), para remoção de primers e nucleótidos livres. As

especificações do comerciante foram seguidas: inserir as colunas num tubo eppendorf e centrifugar a 740g durante 1 minuto; descartar o eppendorf contendo o tampão que embebia a coluna, inserir a coluna em novo eppendorf, pipetar a amostra amplificada e centrifugar a 740g durante 2 min. A amostra amplificada e purificada é recolhida no eppendorf.

Após esta etapa de purificação das amostras, foram preparadas as reacções de sequenciação.

Para cada reacção de sequenciação foram utilizados até 2,5µL de DNA amplificado, 1µL do kit Big Dye Terminator Cycle Sequencing Ready Reaction (AB Applied Biosystems), 1µL de tampão de diluição (AB Applied Biosystems) e 0,5µL de um primer a 2,5mM, num volume final de 5µL. Tanto para o HVRI como para o HVRII foram realizadas reacções de sequenciação para o sentido forward e reverse; para a sequenciação completa só foi realizada a reacção de sequenciação no sentido forward, recorrendo-se apenas a confirmação no reverse em casos de interrupção da sequência pela presença de poli-C extensos.

As condições da reacção de sequenciação foram: desnaturação inicial a 96°C durante 2 min; 35 ciclos de desnaturação a 96°C durante 15 segundos, *annealing* a 50°C durante 9 segundos e extensão a 60°C durante 2 minutos; extensão final a 60°C durante 10 minutos. As reacções foram efectuadas num termociclador 2700 (AB Applied Biosystems)

No final da reacção de sequenciação, as amostras foram sujeitas a uma nova etapa de purificação, para remover nucleótidos e primers não incorporados.

Esta nova purificação, foi feita recorrendo a colunas de purificação com Sephadex. As colunas foram inseridas em tubos eppendorf e em cada uma delas foram colocados 750µL de Sephadex que passa a constituir o filtro para a passagem das amostras e sua purificação. Seguidamente foram centrifugadas a 1700g durante 4 minutos. O líquido que permanece no eppendorf é depois rejeitado e a coluna conservada.

Para cada coluna assim obtida, foram debitados 5µL de amostra e centrifugadas a 1700g durante 4 minutos. A amostra purificada permanece no eppendorf.

Antes de serem aplicadas no sequenciador automático, é necessário adicionar cerca de 8µL de formamida ultra-pura (AB Applied Biosystems), sendo então os cerca de 12µL aplicados nas placas apropriadas ao sequenciador.

O sequenciador automático utilizado foi o ABI 3100 (AB Applied Biosystems), tendo-se utilizado as matrizes providenciadas pelo fornecedor e adequadas ao kit comercial

utilizado. Os perfis electroforéticos foram analisados nos programas de análise da AB Applied Biosystems. A sua edição foi efectuada por dois investigadores, manualmente e por leitura automática no programa BioEdit (Hall, 1999) contra a sequência referência CRS revista por Andrews *et al.* (1999).

6 – Afiliação em haplogrupos

A classificação dos haplótipos de HVRI e HVRII em haplogrupos seguiu as publicações mais recentes, atendendo às alterações gerais recentemente sugeridas por Torroni *et al.* (2006), como a mudança de nome de (pre-HV)1 para R0a. Para haplogrupos de afiliação subsariana, foram seguidas as alterações fulcrais de nomenclatura sugeridas por Salas *et al.* (2004) e atendeu-se às actualizações da sequenciação completa de Torroni *et al.* (2006). Para os haplogrupos de afiliação euroasiática seguiu-se Richards *et al.* (2000) e as actualizações da sequenciação completa de Achilli *et al.* (2004).

O haplogrupo classificado previamente como L3g (Salas *et al.*, 2002) sofreu alteração para L4g em Kivisild *et al.* (2004) e L4b em Doron *et al.* (submetido), devido a informação da sequência completa. Seguiu-se a nomenclatura de L4b.

Tabela 2 – Critérios usados na classificação das amostras do Sudão em haplogrupos. Posições variáveis, por comparação com CRS, para HVRI (menos 16000) e HVRII. As substituições indicadas são transições, excepto se a mudança de base é uma transversão, estando nesse caso indicadas a seguir à posição onde ocorrem.

Haplogrupo	HVRI	HVRII
R0a	126 362	64 263
HV1	067	263
H*	CRS	263
R*	CRS	73 263
N*	223	73 263
U5	270	73 263
U6a	172 219 278	73 263
T	126 294	73 263
J1	069 126	73 295 263
K	224 311	73 263
M1	129 189 223 249 311	73 195 263
M1a	129 189 223 249 311 359	73 195 263
M7a	209 223	73 263
L0a1	129 148 168 172 187 188G 189 223 230 311 320	93 146 152 185 189 236 247
L2	223 278 390	73 146 150 152 182 263
L2a	223 278 294 390	73 146 152 182 195 263
L2a1	223 278 294 309 390	73 143 146 152 195 263
L2a2	189 223 229 278 291 294 311 390	73 152 182 195 263
L2b	114A 129 213 223 278 390	73 146 150 152 182 195 198 204 263
L2c	223 278 390	73 93 146 150 152 182 195 198 263 325
L2c1	223 278 318 390	73 93 146 150 152 182 195 198 263 325
L2d1	129 189 278 300 311 354 390 399	73 146 150 195 263
L2d2	111A 145 223 239 278 292 355 390 399	73 146 150 195 263
L3*	223	73 263
L3b	124 223 278 362	73 263
L3d	124 223	73 152 263
L3e5	041 223	73 150 263 398
L3f	209 223 311	73 263
L3f1	209 223 292 311	73 263
L3f2	209 223 234 311	73 263
L3h1	223 311	73 263
L3h1b	223 256A 284 311	73 189C 195 263
L3h1b1	179 223 256A 284 311	73 189C 195 263
L3h2	111 184 223 304 311	73 150 263 318
L3x	169 223 311	73 150 263
L4b2	223 293T 311 355 362 399	73 146 263
L5b	129 148 166 187 189 223 278 311 355 362	73 152 182 195 247 263

7 – Populações usadas para comparação

Foi efectuada uma pesquisa cuidada da literatura para selecção de populações para comparação de resultados (Tabela 3). Os critérios de selecção foram: proximidade geográfica do Sudão; número razoável de amostras estudadas, idealmente a rondar pelo menos 100 indivíduos; e, sempre que possível, com informação para HVRI e HVRII. Para além de populações africanas, foram incluídas as populações arábicas da Arábia Saudita e Iémen, e a europeia de Portugal.

Tabela 3 - Populações usadas para comparação.

Grupo populacional	População	n	Regiões mitocondriais estudadas	Referência
Norte África	<i>Norte de África</i>			
	Tunísia – Zriba	50	HVRI	Cherni <i>et al.</i> (2005)
	Marrocos	50	HVRI	Brakez <i>et al.</i> (2001)
	Egipto	68	HVRI	Krings <i>et al.</i> (1999)
	Núbia	80	HVRI	Krings <i>et al.</i> (1999)
África Subsariana	<i>África de Leste</i>			
	Sudão	102	HVRI+HVRII	Presente estudo
	Etiópia	270	HVRI	Kivisild <i>et al.</i> (2004)
	Quênia	100	HVRI+HVRII	Brandstätter <i>et al.</i> (2004)
	<i>África Sudeste</i>			
	Moçambique	109	HVRI+HVRII	Pereira <i>et al.</i> (2001)
	<i>África Central</i>			
	Corredor Sael-Sudão	448	HVRI	Cerny <i>et al.</i> (2007)
<i>África Oeste</i>				
	Angola - Cabinda	109	HVRI+HVRII	Beleza <i>et al.</i> (2005)
Península Arábica	Iémen	115	HVRI	Kivisild <i>et al.</i> (2004)
	Arábia Saudita	120	HVRI+HVRII	Abu-Amero <i>et al.</i> (2007)
Europa	Portugal	549	HVRI+HVRII	Pereira <i>et al.</i> (2004)

Para a comparação das sequências completas do Sudão, foram usadas 94 e 4 sequências completas de L3(xM,N) e L4, respectivamente, já publicadas e referidas em detalhe no Anexo 1. Quase todas são sequências completas, excepto as de Kivisild *et al.* (2006), que não possuem informação para a região controlo.

8 – Análise estatística

Para a análise dos dados populacionais foram usados vários softwares informáticos.

O programa Arlequin versão 3.1 (Excoffier *et al.*, 2005), foi usado para estimar medidas *standard* de diversidade genética, distribuições *mismatch*, diversidade genética F_{ST} entre pares de populações e AMOVA, cuja descrição sucinta se passa a efectuar.

a) *Diversidade genética (haplotípica) da população*: probabilidade de dois haplótipos escolhidos ao acaso na população serem diferentes. É equivalente à heterozigotia esperada para bases de dados diplóides (Nei, 1987).

A diversidade genética e a sua variância são estimadas pelas fórmulas:

$$\hat{H} = n/n - 1 \left(1 - \sum_{i=1}^k p_i^2 \right)$$

$$V(\hat{H}) = 2/n(n-1) \left\{ 2(n-2) \left[\sum_{i=1}^k p_i^3 - \left(\sum_{i=1}^k p_i^2 \right)^2 \right] + \sum_{i=1}^k p_i^2 - \left(\sum_{i=1}^k p_i^2 \right)^2 \right\}$$

Em que n é o número de cópias do gene na amostra, k é o número de haplótipos e p_i é a frequência da amostra de determinado haplótipo i .

b) *Número médio de pares de diferenças*: número médio de diferenças entre todos os pares de haplótipos da amostra populacional, através do qual podem ser estimados parâmetros de expansão populacional (Tajima, 1993).

Este índice π é definido pela fórmula:

$$\hat{\pi} = \sum_{i=1}^k \sum_{j<i} p_i p_j \hat{d}_{ij},$$

em que \hat{d}_{ij} é uma estimativa do número de mutações que

ocorreram desde a divergência dos haplótipos i e j , k representa o número de haplótipos e p_i é a frequência do haplótipo i .

A variância total, assume a não recombinação entre locais, a neutralidade selectiva e é obtida pela fórmula:

$$V(\hat{\pi}) = \frac{3n(n+1)\hat{\pi} + 2(n^2 + n + 3)\hat{\pi}^2}{11(n^2 - 7n + 6)} \quad (\text{Tajima, 1993})$$

c) *Mismatch distribution*: distribuição do número observado de diferenças entre todos os pares de haplótipos da amostra populacional. A distribuição é geralmente multimodal para populações em equilíbrio demográfico e unimodal para populações que passaram por uma expansão populacional recente. A distribuição unimodal observada é comparada com uma distribuição estimada através do *stepwise expansion model*.

d) *Diversidade nucleotídica*: probabilidade de dois nucleotídeos homólogos, retirados ao acaso da população serem diferentes. É equivalente à diversidade genética mas ao nível dos nucleotídeos (Tajima, 1983; Nei, 1987).

O índice de diversidade nucleotídica é definido pela seguinte fórmula:

$$\hat{\pi}_n = \frac{\sum_{i=1}^k \sum_{j<i} p_i p_j \hat{d}_{ij}}{L}$$

e a variância da diversidade nucleotídica é calculada pela fórmula:

$$V(\hat{\pi}_n) = \frac{n+1}{3(n-1)L} \hat{\pi}_n + \frac{2(n^2+n+3)}{9n(n-1)} \hat{\pi}_n^2 \quad (\text{Tajima, 1983; Nei, 1987})$$

e) *Teste de neutralidade Tajima's D*: este teste é baseado no *infinite-site model* sem recombinação. Compara duas estimativas do parâmetro de mutação θ , em que $\theta=2Mu$, sendo $M=N$ em populações haplóides de efectivo populacional N .

O teste estatístico D é definido pela fórmula:

$$D = \frac{\hat{\theta}_\pi - \hat{\theta}_s}{\sqrt{\text{Var}(\hat{\theta}_\pi - \hat{\theta}_s)}}, \text{ em que } \hat{\theta}_\pi = \hat{\pi} \text{ e } \hat{\theta}_s = S / \sum_{i=0}^{n-1} (1/i), \text{ S é o número de locais de}$$

segregação na amostra.

A significância do teste estatístico D é testada pela geração de amostras aleatórias sob a hipótese de neutralidade selectiva e de população em equilíbrio, usando um algoritmo de simulação de coalescência adaptado de Hudson (1990). O valor de P do parâmetro D é então obtido como a proporção de valores F_s aleatórios menores ou iguais ao observado.

f) *Teste de neutralidade Fu's (F_s)*: este teste é também baseado no *infinite-site model* sem recombinação. Calcula a probabilidade de observar uma amostra neutra aleatória com um número de alelos semelhante ou inferior ao valor observado, dado o número observado de pares de diferença que é usado como estimador do θ .

O valor F_s é calculado por:

$$F_s = \ln \left(\frac{S'}{1-S'} \right) \quad (\text{Fu, 1997})$$

Em que $S' = \Pr(K \geq K_{obs} \mid \theta = \theta_\pi)$

Fu descreveu este teste F_s como sendo bastante sensível à expansão demográfica das populações o que geralmente corresponde a valores de F_s bastante negativos.

g) *Estrutura genética das populações inferida pela análise de variância (AMOVA)*: este tipo de análise baseia-se numa abordagem da variância das frequências genéticas, mas tendo em conta o número de mutações entre os diferentes haplótipos. O utilizador define a priori grupos populacionais e testa a sua estrutura genética, pela análise hierárquica de variância total em partições dos componentes de covariância devidas a (1) diferenças intra-individuais; (2) diferenças inter-individuais; (3) e, no caso de vários grupos de populações, diferenças inter-populacionais.

No caso de uma estrutura genética hierárquica simples, tal como a de populações de indivíduos haplóides, a implementação do algoritmo conduz ao índice de fixação F_{ST} que é idêntico à média ponderada de F_s sobre os loci, $\hat{\theta}_w$, definido por Weir e Cockerham (1984).

Este F_{ST} pode ser expresso pela fórmula:

$$F_{ST} = \frac{f_0 - f_1}{1 - f_1} = \frac{\bar{t}_1 - \bar{t}_0}{\bar{t}}, \text{ em que } f_0 \text{ é a probabilidade de identidade por descendência de dois}$$

genes diferentes tirados da mesma população, f_1 será a probabilidade de identidade por descendência de dois genes diferentes tirados de duas populações diferentes, \bar{t}_1 é o tempo médio de coalescência de dois genes provenientes de duas populações diferentes e \bar{t}_0 será o tempo médio de coalescência de dois genes provenientes da mesma população.

Quando se testa vários grupos de populações, são calculados os índices F_{ST} , F_{SC} e F_{CT} , o primeiro correspondente à variância entre haplótipos intra-populações; o segundo entre populações dentro de grupos e o terceiro entre grupos. As fórmulas para estes cálculos estão descritas em pormenor no manual do Arlequin.

A significância dos índices de fixação é testada através de uma estratégia de permutação não-paramétrica, descrita por Excoffier *et al.* (1992).

h) *Distância genética entre populações por comparação duas a duas*: os F_{ST} entre duas populações podem ser usados como distâncias genéticas entre populações, pela aplicação de uma transformação que lineariza a distância com o tempo de divergência entre as populações. Estes valores de F_{ST} são apresentados sob a forma de uma matriz de distâncias.

A distribuição nula destes valores F_{ST} , sob a hipótese de inexistência de diferenças entre as populações, é obtida por permutações de haplótipos entre populações. O valor de P do teste é a proporção de permutações que levam a valores de F_{ST} maiores ou iguais ao observado. Os valores de P são também fornecidos sob a forma de uma matriz.

A probabilidade de *match* para haplótipos das regiões hipervariáveis foi calculada como a soma dos quadrados das frequências haplotípicas (Brandstätter *et al.*, 2004). As posições 16182, 16183 e as variações de tamanho nos poli-C em 309 e 315 não foram incluídas para distinguir os haplótipos por serem muito recorrentes e poderem variar mesmo entre tecidos do mesmo indivíduo.

A distribuição de *inter-match* entre todos os pares de haplótipos de duas populações foi calculada no Arlequin, colocando uma população fictícia resultante da soma de haplótipos de duas populações, obtendo a *mismatch distribution* desta, e subtraindo-lhe as *intra mismatch distributions* das duas populações.

Os Networks foram obtidos no Programa Network versão 4.201. Inicialmente usou-se o *Reduced Median Network* (Bandelt *et al.*, 1995) para inferir quais as mutações com recorrência mais elevada, testou-se um *reduced threshold* igual a 2, mas como os networks apresentavam ainda uma elevada reticulação, incompatível com uma fácil visualização, optou-se por um *reduced threshold* igual a 1. Essa nova *file* transformada foi então usada para construir um Median-Joining Network (Bandelt *et al.*, 1999). A estatística *rho* (Forster *et al.*, 1996) foi usada para estimar a idade de expansão desde o *Most Recent Common*

Ancestor (MRCA) de um dado haplogrupo, tendo o cálculo sido efectuado no programa Network. Para as estimativas de idade com base na diversidade em HVRI, apenas o fragmento entre as posições 16090-16365 foi usado, porque a taxa de mutação foi calculada para apenas este segmento, sendo de 1 mutação em cada 20180 anos (Forster *et al.*, 1996). Para as estimativas de idade com base na diversidade da região codificante, foi usada uma taxa de mutação de 1 mutação em cada 5138 anos (Mishmar *et al.*, 2003), não tendo sido contabilizadas as inserções/deleções. Todas as idades são apresentadas como valor \pm desvio padrão, em vez de se apresentar o intervalo de 95% de confiança (multiplicando o desvio padrão por 2) por ser esta a forma mais comum de apresentação nos trabalhos mais recentes (ex: Macaulay *et al.*, 2005; Olivieri *et al.*, 2006).

A árvore filogenética para as sequências completas L3(xM,N) e L4 foi construída segundo o método de máxima parcimónia, tendo sido inicialmente construída à mão, confirmando-se a ramificação por networks obtidos no programa Network.

III – RESULTADOS e DISCUSSÃO

1 – Diversidade haplotípica em HVRI e HVRII

No anexo 2 apresenta-se a tabela com os haplótipos observados no Sudão para as regiões HVRI e HVRII.

Como seria de esperar, o Sudão apresenta uma elevada diversidade haplotípica para ambas as regiões hipervariáveis, HVRI e HVRII (Tabelas 4-6), que se reflecte nos diversos parâmetros de diversidade, como diversidade haplotípica, número de haplótipos, diversidade nucleotídica e número médio de pares de diferenças.

Relativamente ao HVRI, estes elevados valores de diversidade são compartilhados pelas restantes populações de África. Os valores são elevadíssimos no leste de África, com os máximos na Etiópia e Quênia, baixando tanto para norte como para sul. A amostra do Egipto pode constituir uma excepção porque possui um número bastante inferior de indivíduos estudados.

A amostra constituída por 12 populações do Corredor Sael-Sudão (Cerny *et al.* 2007), incluindo Camarões, Chad, Nigéria e Níger, apresenta também uma elevada diversidade, mas pode ser resultado de provir de uma tão alargada região demográfica. A população de Cabinda, um enclave de Angola, na costa oeste da África Subsariana, apresenta ainda uma elevada diversidade, superior à observada em Moçambique, na mesma latitude mas na costa leste de África.

As populações de Marrocos e Tunísia apresentam valores de diversidade muito baixos, especialmente quando comparados com os da população do Egipto, que pode ser classificado geograficamente no mesmo grupo de populações do Norte de África. Como foi já descrito na literatura (Pereira *et al.*, 2005a), tal diminuição da diversidade pode reflectir a elevada estrutura populacional observada nas populações islâmicas do Norte de África ou ser um efeito de a amostragem no Norte de África ser efectuada em pequenos grupos populacionais, provenientes de regiões rurais. Claramente, tal nível de estrutura populacional, quer real quer derivado de enviesamento amostral, não é partilhado pelas populações islâmicas do leste de África e da Península Arábica.

Na população de Portugal, os valores de diversidade são inferiores aos de África Subsariana, e da Península Arábica, como seria de esperar para uma população europeia.

Tabela 4 - Diversidade molecular das populações analisadas em HVRI.

População	n	Diversidade haplotípica	Nº haplótipos (%)	Nº médio de pares de diferenças	Diversidade nucleotídica	Tajima's D ¹	Fu's Fs ²
Marrocos	50	0.959 +/- 0.018	33 (66)	4.300 +/- 2.165	0.012 +/- 0.006	-1.457	-25.665***
Tunísia	50	0.903 +/- 0.022	16 (32)	3.948 +/- 2.010	0.011 +/- 0.006	-1.433	-3.219
Egipto	68	0.993 +/- 0.005	59 (86)	7.045 +/- 3.349	0.019 +/- 0.010	-1.750*	-25.020***
Núbia	80	0.977 +/- 0.008	53 (66)	8.202 +/- 3.843	0.022 +/- 0.011	-1.420	-24.778***
Sudão	102	0.989 +/- 0.004	75 (73)	7.073 +/- 3.347	0.019 +/- 0.011	-1.740*	-24.913***
Etiópia	270	0.993 +/- 0.001	161 (59)	8.543 +/- 3.961	0.023 +/- 0.012	-1.634*	-24.251**
Quénia	100	0.995 +/- 0.002	82 (82)	9.377 +/- 4.342	0.026 +/- 0.013	-1.356	-24.545***
Moçambique	109	0.960 +/- 0.008	49 (44)	7.780 +/- 3.650	0.021 +/- 0.011	-0.894	-23.619***
Corredor Sael-Sudão	448	0.990 +/- 0.001	206 (46)	7.613 +/- 3.558	0.021 +/- 0.011	-1.578*	-24.179**
Angola - Cabinda	110	0.988 +/- 0.003	71 (64)	9.139 +/- 4.236	0.025 +/- 0.013	-1.032	-24.540***
Iémen	115	0.980 +/- 0.005	65 (56)	7.542 +/- 3.547	0.021 +/- 0.011	-1.604*	-24.778***
Arábia Saudita	120	0.986 +/- 0.003	77 (64)	6.815 +/- 3.232	0.019 +/- 0.009	-1.815**	-24.912***
Portugal	549	0.956 +/- 0.007	260 (47)	4.555 +/- 2.243	0.013 +/- 0.007	-2.149***	-24.885***

¹ Valor de Tajima's (p : *=0,01< p ≤0,05; **=0,001< p ≤0,01; ***= p ≤0,001).

² Valor de Fu's Fs statistic (p : *=0,01< p ≤0,05; **=0,001< p ≤0,01; ***= p ≤0,001).

Para as poucas populações para as quais se dispunha da diversidade em HVRII, foram efectuados os mesmos cálculos (Tabela 5). Verificou-se que a diversidade em HVRII é inferior à de HVRI, para todos os parâmetros analisados, em todas as populações.

Mas quando se junta a informação de HVRI e HVRII (Tabela 6), a taxa de diversidade é bastante superior, permitindo um maior poder de distinção entre os haplótipos. De facto, sabe-se que a heterogeneidade na taxa de mutação entre as posições de HVRII é muito elevada, havendo posições *hotspot*, extremamente mutáveis, e outras que quase não variam. Daí que, se a informação de HVRII para a filogenia é bastante reduzida, a informação para a distinção entre haplótipos, importante para as ciências forenses, é importante.

Tabela 5 - Diversidade molecular das populações analisadas em HVRII.

População	n	Diversidade haplotípica	Nº haplótipos (%)	Nº médio de pares de diferenças	Diversidade nucleotídica	Tajima's D ¹	Fu's Fs ²
Sudão	102	0.963 +/- 0.009	49 (48)	3.959 +/- 1.998	0.014 +/- 0.008	-0.953	-25.895***
Quênia	100	0.978 +/- 0.006	64 (64)	5.721 +/- 2.763	0.021 +/- 0.011	-0.618	-25.251***
Moçambique	109	0.846 +/- 0.032	35 (32)	5.180 +/- 2.527	0.019 +/- 0.010	-0.179	-14.351***
Angola - Cabinda	110	0.971 +/- 0.007	55 (50)	7.371 +/- 3.474	0.027 +/- 0.014	-0.010	-24.824***
Arábia Saudita	120	0.962 +/- 0.008	58 (48)	3.355 +/- 1.733	0.012 +/- 0.007	-1.857**	-26.188***
Portugal	549	0.905 +/- 0.010	138 (25)	2.905 +/- 1.528	0.011 +/- 0.006	-1.863***	-25.802***

¹ Valor de Tajima's (p : *=0,01< p ≤0,05; **=0,001< p ≤0,01; ***= p ≤0,001).

² Valor de Fu's Fs statistic (p : *=0,01< p ≤0,05; **=0,001< p ≤0,01; ***= p ≤0,001).

O número médio de pares de diferenças reportado nas tabelas 4 a 6 representa um índice que se obtém a partir das *mismatch distributions*.

Tabela 6 - Diversidade molecular das populações analisadas em HVRI+HVRII.

População	n	Diversidade haplotípica	Nº haplótipos (%)	Nº médio de pares de diferenças	Diversidade nucleotídica	Tajima's D ¹	Fu's Fs ²
Sudão	102	0.991 +/- 0.004	82 (80)	11.032 +/- 5.054	0.017 +/- 0.009	-1.561*	-24.358***
Quênia	100	0.997 +/- 0.002	90 (90)	15.099 +/- 6.806	0.024 +/- 0.012	-1.154	-24.104***
Moçambique	109	0.971 +/- 0.007	63 (57)	12.960 +/- 5.881	0.020 +/- 0.010	-0.670	-24.182***
Angola - Cabinda	110	0.994 +/- 0.002	87 (79)	16.510 +/- 7.408	0.026 +/- 0.013	-0.682	-24.019**
Arábia Saudita	120	0.992 +/- 0.003	92 (76)	10.170 +/- 4.677	0.016 +/- 0.008	-1.883**	-24.383***
Portugal	549	0.986 +/- 0.003	358 (65)	7.445 +/- 3.485	0.012 +/- 0.006	-2.111***	-24.136**

¹ Valor de Tajima's (p : *=0,01< p ≤0,05; **=0,001< p ≤0,01; ***= p ≤0,001).

² Valor de Fu's Fs statistic (p : *=0,01< p ≤0,05; **=0,001< p ≤0,01; ***= p ≤0,001).

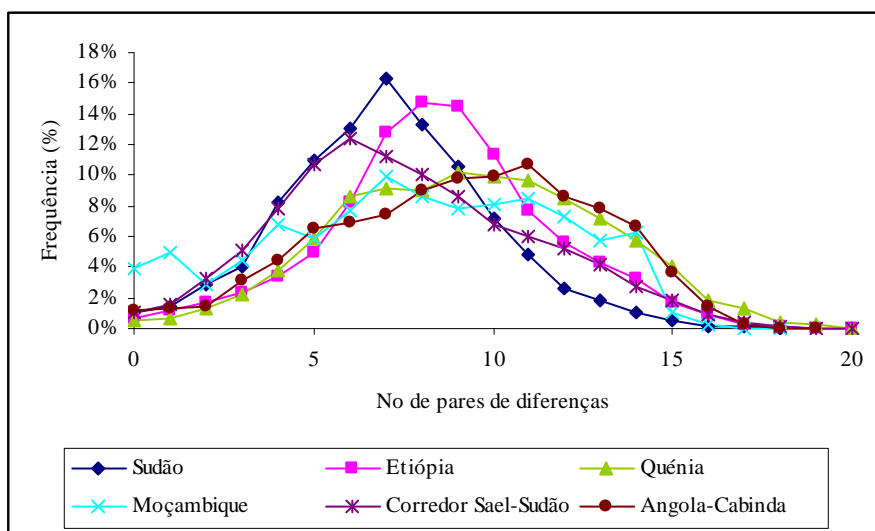
Estas são as distribuições do número de diferenças nucleotídicas entre todos os pares de haplótipos presentes na amostra e podem dar informações importantes quanto a parâmetros demográficos das populações. A distribuição é geralmente multimodal para populações em equilíbrio demográfico e unimodal para populações que passaram por uma expansão populacional recente (Rogers e Harpending, 1992). Mas estas inferências quanto à expansão demográfica devem ser interpretadas com cuidado porque outros factores

podem levar a resultados semelhantes. Por exemplo, se uma população resulta da mistura de dois *pools* genéticos muito diferentes entre si, e ainda não passou tempo suficiente para esbater as diferenças entre eles, uma distribuição bimodal pode ser obtida.

As figuras 4-6 representam as *mismatch distributions* para as diversas populações e regiões estudadas.

Todas as populações apresentam distribuições unimodais para a diversidade em HVRI, excepto as populações de Marrocos e Tunísia. Tal facto pode resultar de o *pool* genético destas populações ser uma mistura de linhagens mitocondriais de origem euroasiática com linhagens de origem subsariana, em proporções bastante elevadas.

A população europeia possui uma média mais baixa do que as restantes populações. As populações da Península Arábica possuem distribuições muito semelhantes às do Sudão e Egipto. E as médias mais elevadas são atingidas para as populações do Quénia e Angola.



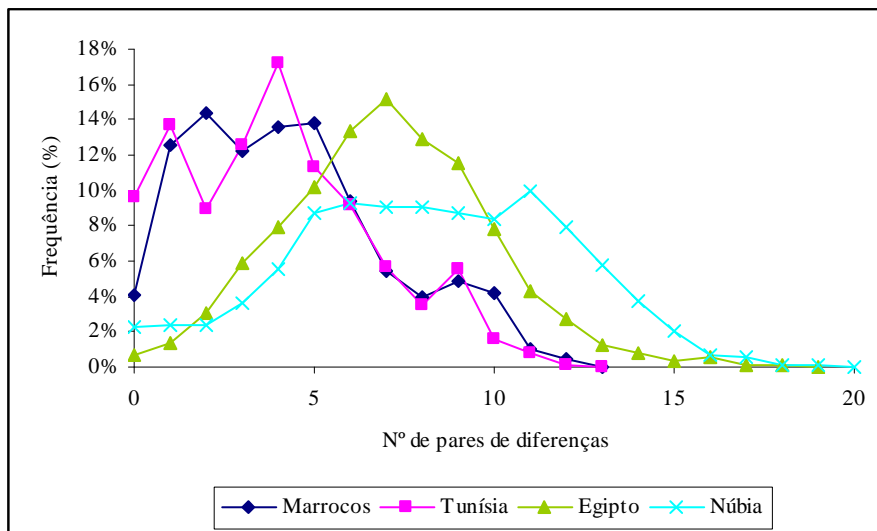
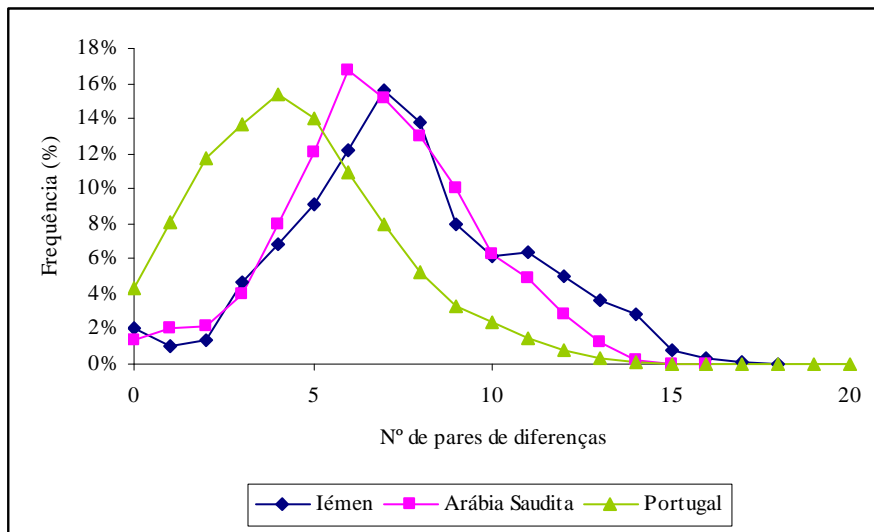


Figura 4 – Distribuição do número de pares de diferenças (*Mismatch Distributions*) para HVRI nas diversas populações estudadas.

As distribuições para HVRII e HVRI+HVRII seguem um padrão semelhante, com valores médios comparativamente mais baixos em HVRII e mais elevados em HVRI+HVRII.

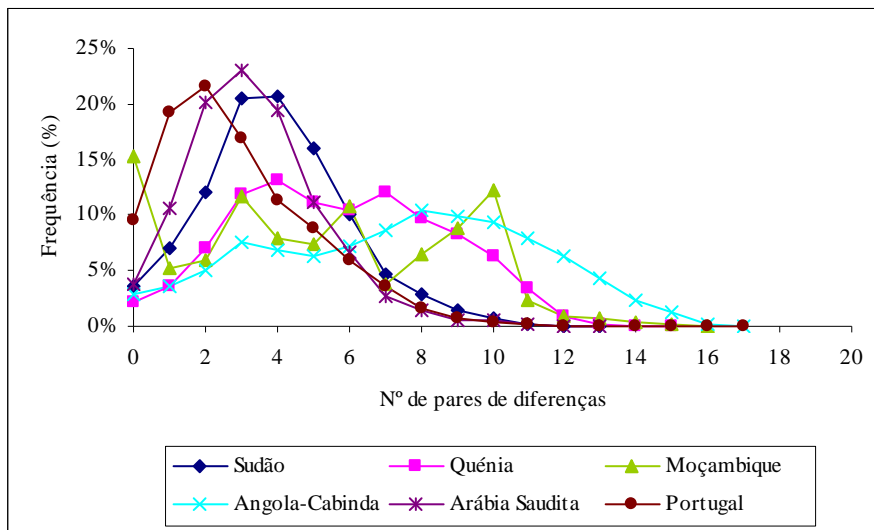


Figura 5 – Distribuição do número de pares de diferenças (*Mismatch Distributions*) para HVRII nas diversas populações estudadas.

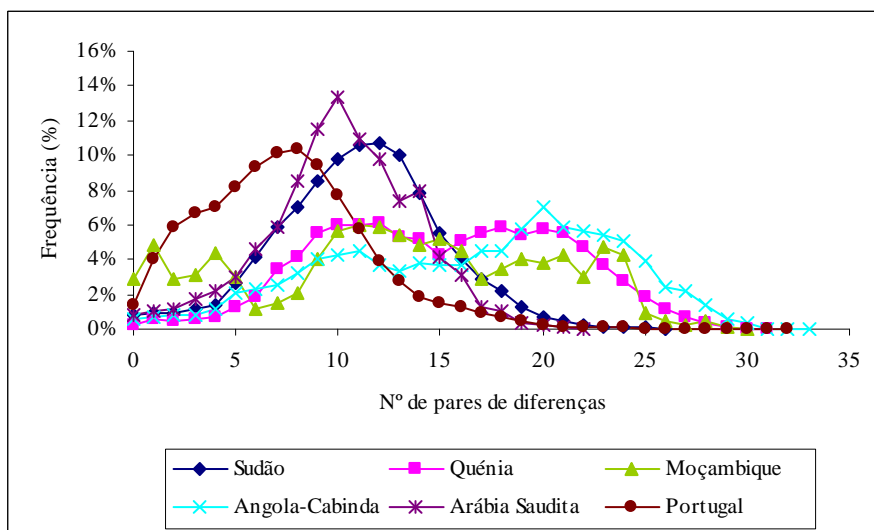


Figura 6 – Distribuição do número de pares de diferenças (*Mismatch Distributions*) para HVRI+HVRII nas diversas populações estudadas.

Outra forma de avaliar se as populações poderão ter passado por um episódio de expansão demográfica é pela aplicação de testes de neutralidade. Se a população está em expansão, os seus parâmetros de diversidade divergem significativamente do esperado segundo o modelo de selecção neutral, sendo os valores negativos. A população também pode divergir significativamente do esperado, mas com valores positivos, se passou por períodos de *bottleneck*, ou seja, de perda de diversidade. Estes valores negativos

estatisticamente significativos também podem ser obtidos se há selecção negativa sobre haplótipos raros; enquanto os valores positivos estatisticamente significativos podem ser obtidos se há selecção positiva de alguns haplótipos mais frequentes. Actualmente, não é possível distinguir entre o efeito demográfico e o efeito de selecção quando em presença de valores estatisticamente significativos.

De qualquer modo, para todas as populações, excepto a Tunísia, obtiveram-se valores negativos e estatisticamente significativos para pelo menos um dos testes usados para verificação da neutralidade (Tabelas 4 a 6). O que é geralmente interpretado como sendo indicador de que estas populações sofreram uma expansão demográfica recente.

2 – Probabilidade de *match a priori*

Um parâmetro importante para as ciências forenses é a capacidade de distinção *a priori* entre haplótipos, ou seja, para uma dada amostra populacional, qual a probabilidade de *match* aleatório. Na tabela 7 representa-se a probabilidade aleatória de *match* para estas populações, nas diversas regiões mitocondriais analisadas. Este parâmetro reflecte o quadro descrito anteriormente para os outros parâmetros de diversidade. Particularmente para o Sudão, a probabilidade de identidade ao estudar apenas a região HVRI é de 2%, duplicando quando se analisa apenas HVRII e reduzindo ligeiramente para 1,8% se estudar as duas regiões simultaneamente.

Tabela 7 - Probabilidade (%) aleatória de *match* para as regiões HVRI, HVRII e HVRI+HVRII.

População	n	HVRI	HVRII	HVRI+HVRII
Marrocos	50	6,0	-	-
Tunísia	50	11,4	-	-
Egipto	68	2,1	-	-
Núbia	80	3,5	-	-
Sudão	102	2,0	4,6	1,8
Etiópia	270	1,1	-	-
Quénia	100	1,4	3,1	1,2
Moçambique	109	4,8	16,1	3,7
Corredor Sael-Sudão	448	1,2	-	-
Angola-Cabinda	110	2,0	3,7	1,4
Iémen	115	2,8	-	-
Arábia Saudita	120	2,1	4,6	1,6
Portugal	549	4,5	9,7	1,6

A *mismatch distribution* também pode ser usada para estimar o número de diferenças nucleotídicas entre todos os pares de comparações haplotípicas entre duas populações, o que serve de informação para a probabilidade de *inter-match* entre populações. Usou-se este método para calcular o *inter-match* do Sudão versus as outras populações. A distribuição está representada na Figura 7 e as probabilidades de *inter-match* (correspondente a 0 diferenças entre os pares de haplótipos) estão reportadas na Tabela 8.

Tabela 8 - Probabilidade (%) aleatória de *inter-match* para as regiões HVRI, HVRII e HVRI+HVRII entre a amostra do Sudão com todas as outras.

População	HVRI	HVRII	HVRI+HVRII
Marrocos	0,88	-	-
Tunísia	0,92	-	-
Egipto	0,48	-	-
Núbia	0,64	-	-
Etiópia	0,33	-	-
Quênia	0,24	1,59	0,11
Moçambique	0,00	0,74	0
Corredor Sael-Sudão	0,52	-	-
Angola-Cabinda	0,07	0,65	0,09
Iémen	0,25	-	-
Arábia Saudita	0,22	2,12	0,06
Portugal	0,90	3,48	0,31

Como se pode verificar, a probabilidade de *inter-match* (0 diferenças) dos haplótipos do Sudão com os haplótipos das outras populações é menor que o *match* no Sudão, sendo sempre inferior a 1%. Curiosamente, os valores de *inter-match* são maiores com as populações de Portugal, Tunísia e Marrocos, porque estas populações têm frequências mais elevadas do haplótipo CRS, que também está presente no Sudão.

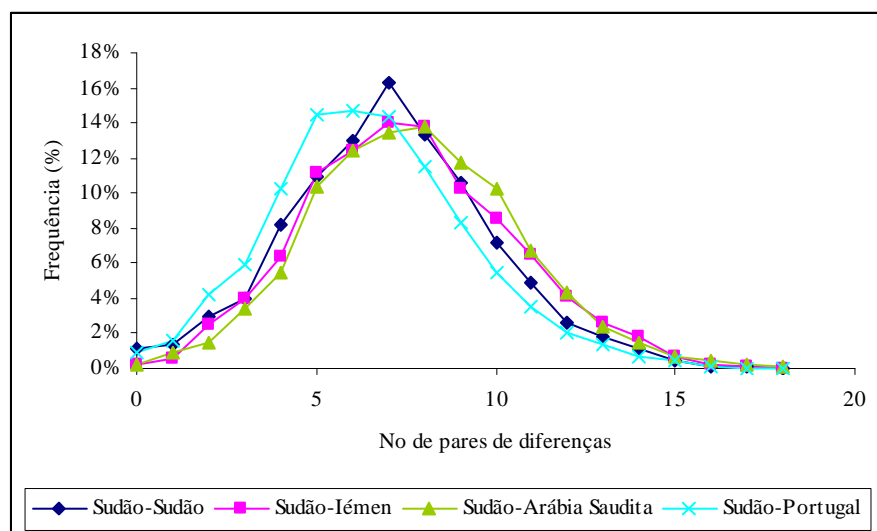
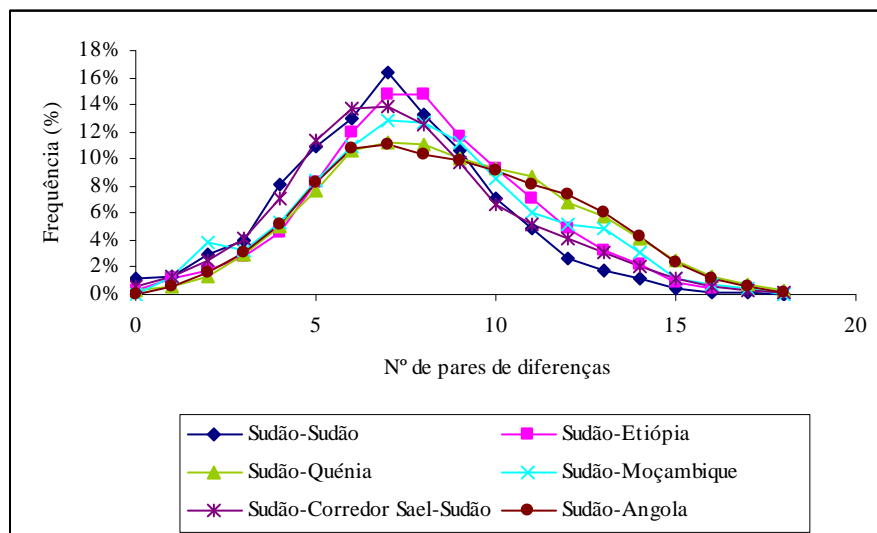
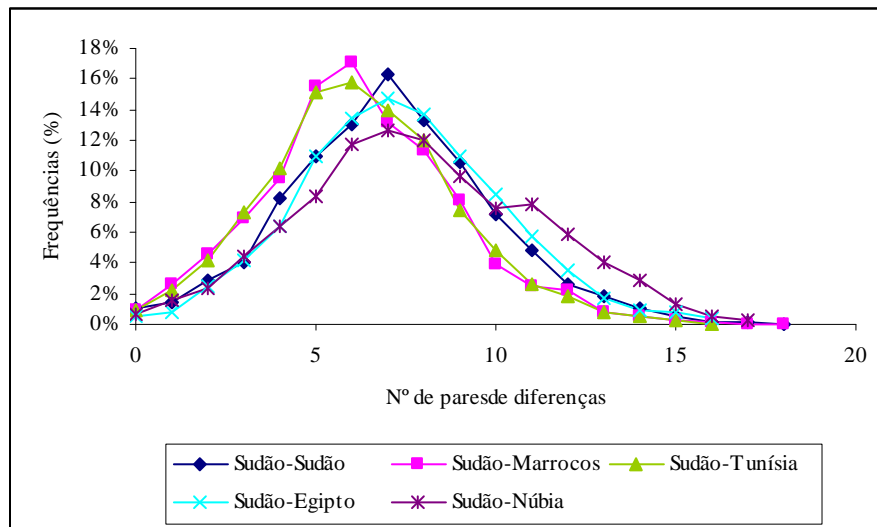


Figura 7 – Distribuições de *inter-match* do Sudão versus as outras populações, bem como a *mismatch distribution* do Sudão, para HVRI.

Mas quanto às distribuições de *inter-match* verifica-se que a extensa maioria dos haplótipos diverge em várias posições nucleotídicas, com média entre 5 e 8 diferenças para o HVRI. A distribuição de *inter-match* com todas as outras populações não diverge da *mismatch distribution* para os haplótipos do Sudão.

O mesmo se verifica para HVRII e HVRI+HVRII, com média menor para o primeiro e maior para o segundo, relativamente a HVRI isolado (Figuras 8 e 9).

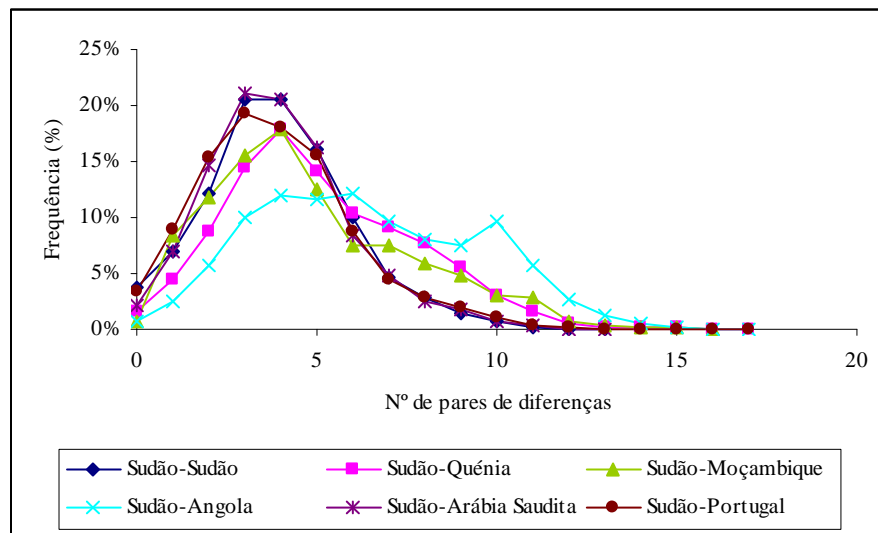


Figura 8 – Distribuições de *inter-match* do Sudão versus as outras populações, bem como a *mismatch distribution* do Sudão, para HVRII.

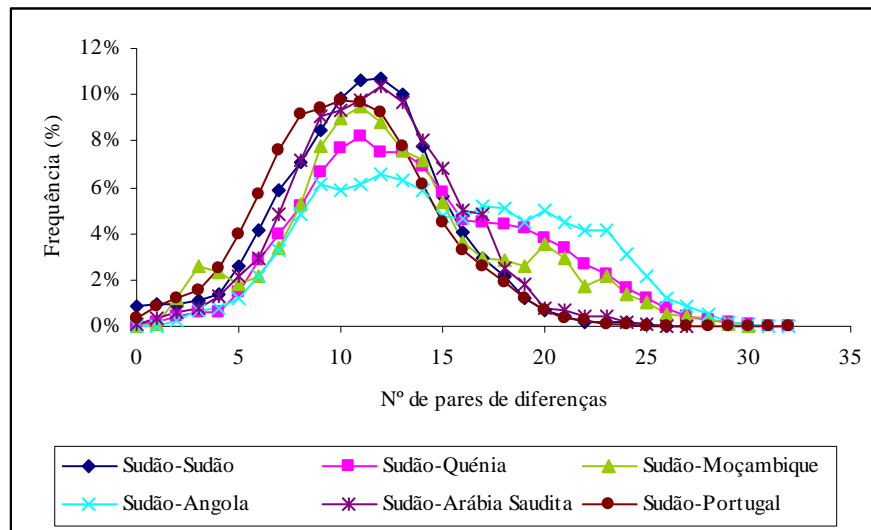


Figura 9 – Distribuições de *inter-match* do Sudão versus as outras populações, bem como a *mismatch distribution* do Sudão, para HVRI+HVRII.

3 – Estruturação populacional baseada na diversidade haplotípica

Quando a diversidade em HVRI é utilizada para estimar as distâncias genéticas F_{ST} entre todos os pares de populações (Tabela 9), verifica-se que todas as populações são significativamente diferenciadas. O que indica, à partida, que a maior percentagem de variação genética estará presente a nível intra-populacional.

A partição da diversidade genética foi testada através da análise da variância molecular, AMOVA (Tabela 10). A extensa percentagem de variação ocorre intra-populacionalmente, com valores mínimos de variação entre populações do mesmo grupo e entre grupos, incluindo ou não a população mais divergente de Portugal.

Tabela 9 – Distâncias genéticas F_{ST} entre todos os pares de populações para HVRI. O valor de P para todas as comparações é igual a $0,00000\pm 0,00000$.

	Marrocos	Tunísia	Egipto	Núbia	Sudão	Etiópia	Quênia	Moçambique	Corredor Sael-Sudão	Angola-Cabinda	Iémen	Arábia Saudita	Portugal
Marrocos													
Tunísia	0.06857												
Egipto	0.02377	0.06814											
Núbia	0.03183	0.05895	0.01509										
Sudão	0.02566	0.05251	0.00906	0.01705									
Etiópia	0.02332	0.04940	0.00697	0.01487	0.00897								
Quênia	0.02252	0.04939	0.00593	0.01394	0.00796	0.00589							
Moçambique	0.04027	0.06704	0.02357	0.03150	0.02543	0.02308	0.02233						
Corredor Sael-Sudão	0.02475	0.05060	0.00846	0.01633	0.01044	0.00838	0.00737	0.02447					
Angola-Cabinda	0.02586	0.05263	0.00928	0.01727	0.01129	0.00919	0.00818	0.02563	0.01066				
Iémen	0.03016	0.05688	0.01357	0.02153	0.01553	0.01338	0.01243	0.02986	0.01483	0.01575			
Arábia Saudita	0.02689	0.05357	0.01033	0.01831	0.01233	0.01022	0.00923	0.02664	0.01169	0.01255	0.01678		
Portugal	0.04240	0.06814	0.02603	0.03383	0.02789	0.02557	0.02483	0.04186	0.02689	0.02809	0.03224	0.02909	

Tabela 10 – Resultados de AMOVA para HVRI. *=0,01<p≤0,05; **=0,001<p≤0,01; ***=p≤0,001.

	Porcentagem de variação			Índices de fixação		
	Entre grupos	Entre populações do grupo	Intra-populações	F_{SC}	F_{ST}	F_{CT}
“Portugal” vs “Marrocos/Tunísia/Egipto” vs “Núbia/Sudão/Etiópia/Quénia/ CorredorSaelSudão/AngolaCabinda/ Moçambique” vs “Iémen/ArábiaSaudita”	0,89	1,66	97,46	0.01672***	0.02542***	0.00885*
“Portugal” vs “Marrocos/Tunísia/Egipto/ Núbia/Sudão/Etiópia/Quénia/ CorredorSaelSudão/AngolaCabinda/ Moçambique” vs “Iémen/ArábiaSaudita”	0.91	1.73	97.36	0.01743***	0.02641***	0.00913
“Portugal” vs “Marrocos/Tunísia/Egipto/ Núbia/Sudão/Etiópia/Quénia/ CorredorSaelSudão/AngolaCabinda/ Moçambique/Iémen/ArábiaSaudita”	0	2.32	97.81	0.02313***	0.02186***	0.00000
“Marrocos/Tunísia/Egipto” vs “Núbia/Sudão/Etiópia/Quénia/ CorredorSaelSudão/AngolaCabinda/ Moçambique” vs “Iémen/ArábiaSaudita”	0.11	1.65	98.23	0.01655***	0.01766***	0.00112
“Marrocos/Tunísia/Egipto/ Núbia/Sudão/Etiópia/Quénia/ CorredorSaelSudão/AngolaCabinda/ Moçambique” vs “Iémen/ArábiaSaudita”	0	1.73	98.34	0.01726***	0.01660***	0.00000
“Marrocos/Tunísia/Egipto/ Núbia/Sudão/Etiópia/Quénia/ CorredorSaelSudão/AngolaCabinda/ Moçambique/Iémen/ArábiaSaudita”	-	1.71	98.29	-	0.01707***	-

4 – Diversidade de haplogrupos

Para além da informação que pode ser retirada da análise global da diversidade mitocondrial, outra fonte importante é o estudo filogenético dessa diversidade. Procedeu-se, assim, à afiliação dos haplótipos de HVRI e HVRII em haplogrupos, grupos de sequências relacionadas por partilha de uma ancestralidade comum, segundo os critérios descritos no Material e Métodos. A classificação de cada um dos haplótipos está descrita no Anexo 2 e a distribuição dos haplogrupos está representada na Figura 10.

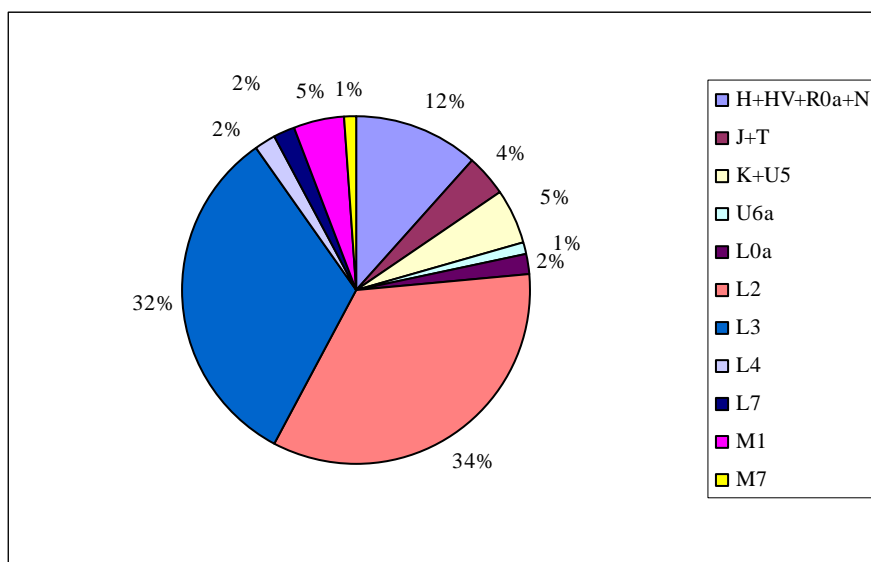


Figura 10 – Distribuição dos haplogrupos na população do Sudão.

Alguns dos haplogrupos presentes no Sudão são típicos de populações Euroasiáticas, como os H, HV, R0a, N*, J, T, K e U5, atingindo uma proporção de 21%. Estes haplogrupos (Figura 11) são mais frequentes no Norte de África, constituindo por volta de 2/3 do pool genético destas populações, e diminuem em frequência em direcção ao sul do continente africano, rondando 1/3 do pool genético dos povos do Leste de África e estando virtualmente ausentes em Angola e Moçambique. Contrariamente ao observado para populações europeias, caracterizadas por uma maior proporção de haplótipos derivados do haplogrupo H (como H1 e H3) do que dos basais R0a, HV e N* (Pereira *et al.*, 2005b), a população do Sudão assemelha-se a populações do Levante (Roostalu *et al.*, 2007) e da Península Arábica (Abu-Amero *et al.*, 2007, Kivisild *et al.*, 2004), contendo uma frequência mais elevada destes haplogrupos basais. Curiosamente, excepto o indivíduo que

não foi possível classificar para além de N*, não foram detectadas outras linhagens prováveis ancestrais do macrohaplogrupo N.

O haplogrupo U6, presente em populações do Norte de África e Ibéria, aparece em apenas um indivíduo no Sudão (1%). Esta linhagem, para além das substituições basais que a afiliam em U6a, apresenta uma transição na posição 103 em HVRII, que parece ser típica do leste de África (Olivieri *et al.*, 2006 e dados próprios não publicados).

O haplogrupo M7 é frequente em populações do sudeste asiático e o indivíduo observado no Sudão pertencente a este haplogrupo pode representar migração recente daquela região. O haplogrupo M1 apresenta-se em frequência elevada no Corno de África, especialmente na Etiópia, atingindo uma frequência de 17%. Durante certo tempo houve discussão se este haplogrupo seria original de África ou representaria uma migração back-to-Africa (ver Quintana-Murci *et al.*, 1999 versus Kivisild *et al.*, 2004). Mais recentemente, novos dados de sequenciação completa apontam para que este haplogrupo ou um seu ancestral tenha sido introduzido em África, numa migração back-to-Africa, através do Levante, há cerca de 40.000 a 45.000 anos atrás (Olivieri *et al.*, 2006). Este haplogrupo está presente numa percentagem de 5% no Sudão. Assim como para o macrohaplogrupo N não foram detectadas no Sudão linhagens ancestrais, o mesmo pode ser afirmado para o macrohaplogrupo M. Esta ausência é partilhada pelas outras populações do Leste de África estudadas até ao momento.

Os restantes haplogrupos, L0, L2, L3(xM,N), L4 e L5 são típicos de populações subsarianas e constituem 72% do pool genético do Sudão. Os haplogrupos L0, L4 e L5 são raros no Sudão, rondando os 2% cada. L0 é um haplogrupo mais frequente nos povos Khoisan do sul de África, podendo ser um sinal da diversidade primordial dos povos subsarianos, posteriormente substituída pela extensa expansão Bantu (Pereira *et al.*, 2001; Salas *et al.*, 2002); apresenta-se em frequências mais elevadas, no presente, em populações do sul de África, como Moçambique e Angola. L4 e L5 são haplogrupos raros nas restantes populações africanas.

Os haplogrupos L2 e L3(xM,N) partilham uma frequência de 32-34% no Sudão, sendo também os haplogrupos mais frequentes nos outros povos do leste e sul de África. Mas estes haplogrupos parecem apresentar, em geral, uma diversidade mais elevada no leste de África do que no resto do continente.

A diversidade relativa do haplogrupo L2 está apresentada na Figura 12 e o network da diversidade HVRI das linhagens L2 observadas nas diversas populações aqui estudadas está representado na Figura 13. A reticulação é muito elevada e mesmo usando um drástico *reduced threshold* igual a 1, existem ainda algumas ramificações não resolvidas. A idade total obtida para o haplogrupo L2, com base nos dados representados é de 80.200 ± 21.300 , comparável com a estimativa obtido por Salas *et al.* (2002) de 70.100 ± 15.300 .

O subhaplogrupo L2a é o mais frequente em praticamente todas as populações, chegando a atingir uma proporção de 100% na Tunísia e Núbia. Este haplogrupo, identificado por uma transição na posição 16294, para além das basais de L2, é um haplogrupo diverso na África oeste e leste, mas raro em populações Khoisan. Chen *et al.* (2000) determinaram uma idade rondando os 39.000-51.400 anos para L2a; a estimativa obtida neste trabalho aponta para 59.600 ± 19.200 anos, a qual é muito próxima da apresentada por Salas *et al.* (2002) de 55.150 ± 19.350 . Algumas das linhagens L2a, contendo as substituições adicionais 16189-16290-16309 e 16189-16192-16290-16309 pela sua considerável frequência mas baixa diversidade em Moçambique, foram implicadas na expansão Bantu na costa leste de África (Pereira *et al.*, 2001). Nenhuma destas linhagens de provável expansão Bantu para o sudeste de África foi observada no Sudão.

Segundo Salas *et al.* (2004) L2b, L2c e L2d não existem no leste de África e a sua presença na América seria devida à migração de escravos principalmente originais da África Ocidental. Algumas destas linhagens foram observadas na amostra do Sudão, em concordância com algumas linhagens L2b terem já sido observadas no Egipto e Etiópia, L2c na Arábia Saudita e L2d no Líbano e no Quênia. Estas linhagens parecem apresentar uma baixa frequência nas diversas regiões de África e a confirmar-se a sua origem ocidental, por sequenciação completa, a sua presença nas outras regiões poderá ser um sinal de introdução recente pelo comércio de escravos. De realçar que o L2c não pode ser distinguido de L2* com base apenas na informação de HVRI, sendo necessária a informação de HVRII, logo não se distinguindo no network representado na Figura 13.

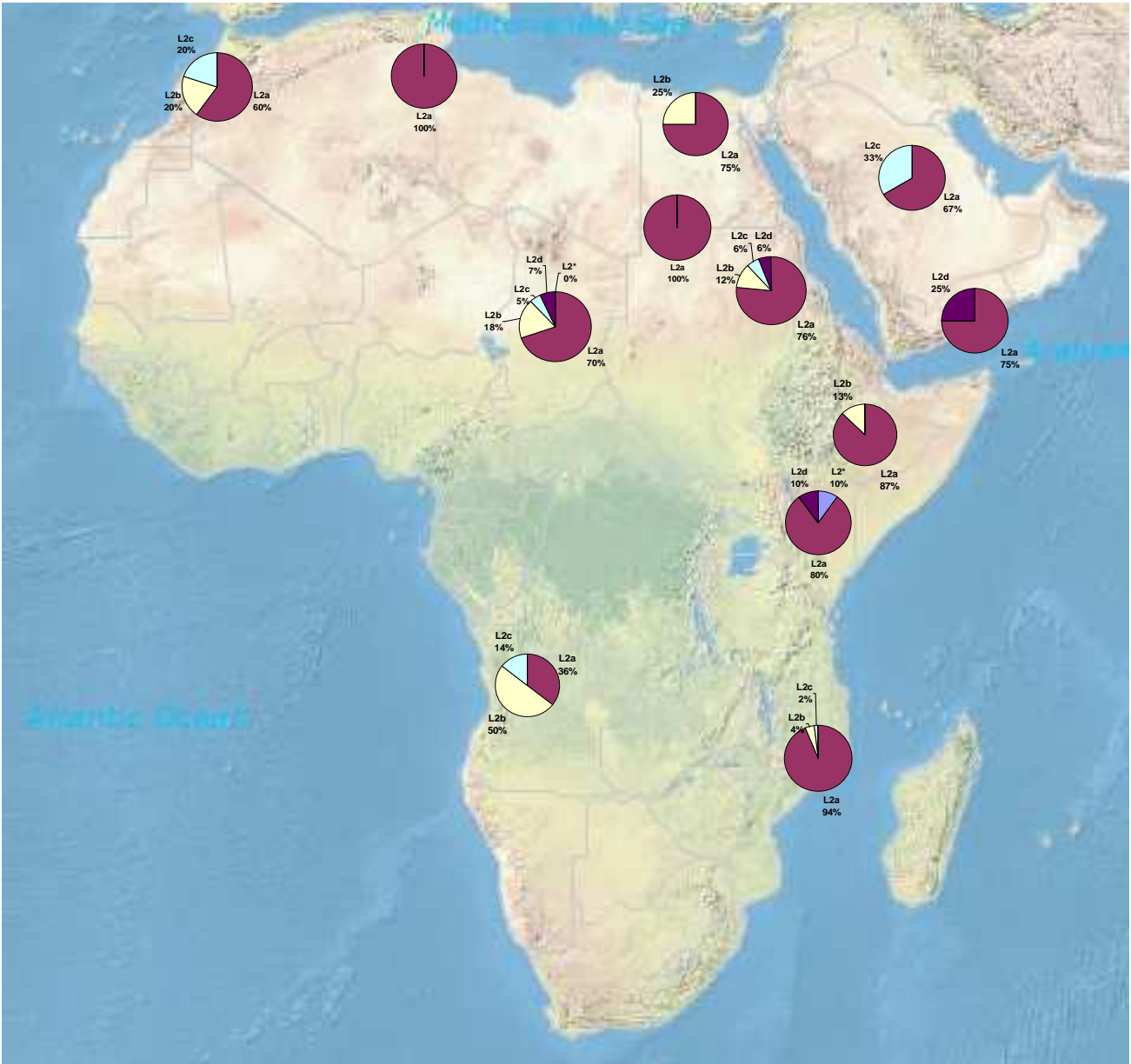


Figura 12 - Distribuição dos subhaplogrupos L2 nas populações de África e Península Arábica.

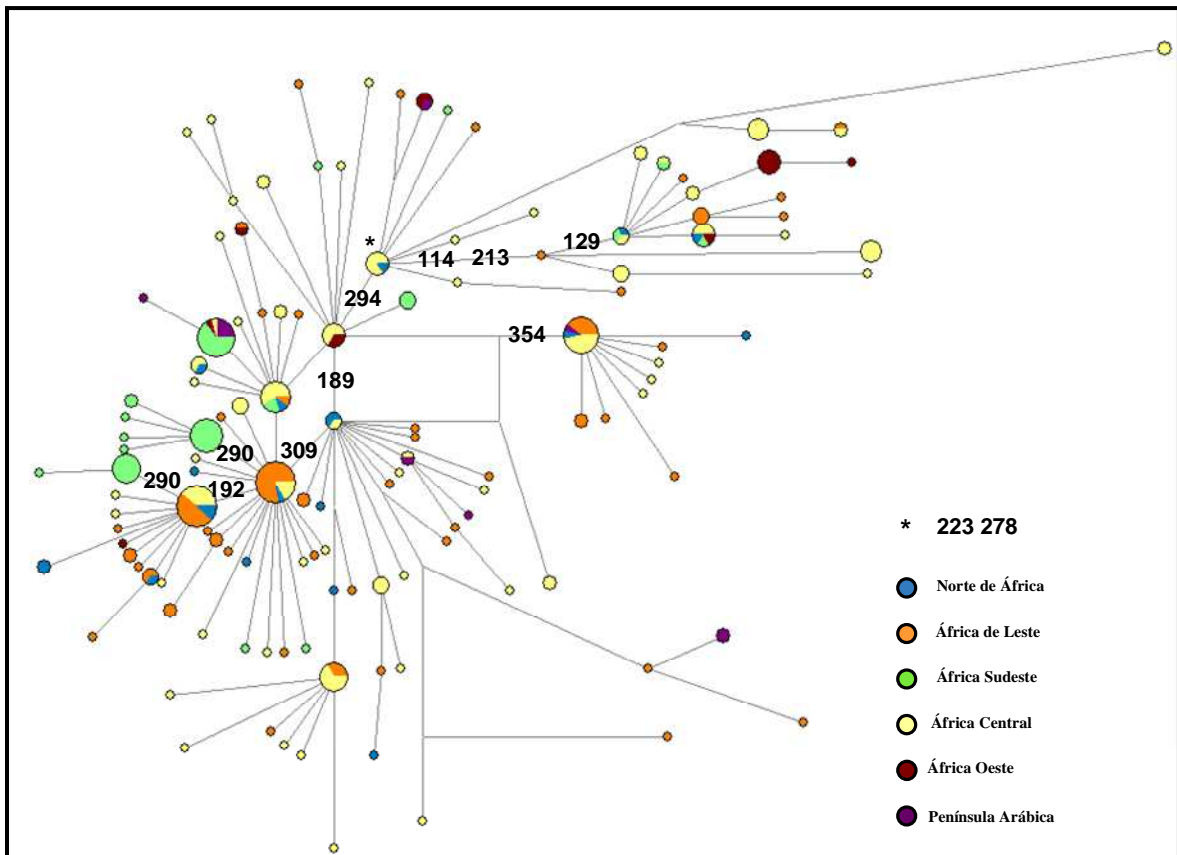


Figura 13 - Reduced network da diversidade em HVRI (entre as posições 16090-16365) para o haplogrupo L2 nas populações de África e Península Arábica. O tamanho dos ramos é proporcional ao número de mutações, estando algumas representadas (sem o prefixo 16000) e a área dos círculos proporcional ao número de amostras. O asterisco localiza a raiz do network.

O haplogrupo L3 é o haplogrupo com mais sucesso, tendo dois dos seus ramos, os macro-haplogrupos M e N colonizado todo o mundo fora de África. Neste trabalho analisamos o haplogrupo L3(xM,N), ou seja, os ramos de L3 que são característicos de África.

Salas *et al.* (2002) mostraram que estas linhagens L3(xM,N), embora dispersas por toda a África, são mais frequentes no leste de África, o que aponta para a sua origem nesta região. Contudo, na Figura 11, pode-se observar que a frequência de L3(xM,N) em Angola e no Corredor Sael-Sudão é bastante elevada. Mas quando se atende à diversidade em termos de haplogrupos (Figura 14), as populações do leste de África são mais diversas, sendo esta uma evidência a favor desta região ser o provável local de origem do haplogrupo L3(xM,N). A estimativa da idade de L3, com base na diversidade em HVRI (Figura 15) é de

56.300 ± 9.600 anos, da mesma ordem de grandeza da estimativa de 61.300 ± 11.650 obtida por Salas *et al.* (2002). De realçar que o subhaplogrupo designado como L3g por Salas *et al.* (2002), definido pelas substituições nas posições 16223-16293T-16311-16355-16362 foi entretanto demonstrado como sendo um haplogrupo mais ancestral que L3, com base na informação completa do mtDNA, tendo sido designado L4g por Kivisild *et al.* (2004) e por L4b por Doron *et al.* (submetido). Deste modo, seria de esperar que a idade do L3 fosse mais recente do que a obtida por Salas *et al.* (2002).

A comprovar uma menor diversidade de L3(xM,N) fora do leste de África, observou-se que grande proporção destas linhagens observadas no Corredor Sael-Sudão, Angola e Moçambique são pertencentes ao subhaplogrupo L3e. Dentro deste subhaplogrupo verificou-se que L3e1 e L3e3 são elevadamente dispersos provavelmente devido à expansão Bantu (Salas *et al.*, 2004); enquanto L3e2 é mais frequente a norte, e dispersou-se para oeste, onde o subgrupo L3e2b predomina; o menos frequente L3e4 tem uma dispersão semelhante.

No leste de África, o subhaplogrupo mais frequente é o L3f, apresentando elevada diversidade na região. Contudo, o seu derivado L3f1 é mais frequente no oeste de África (Salas *et al.*, 2004). A idade estimada para o subhaplogrupo L3f foi de 39.600 ± 11.900 anos, considerando não apenas o ramo com 16209-16311 no network da Figura 15, mas também o ramo definido por 16209, aceitando que ocorreu uma recorrência na posição 16311. Salas *et al.* (2002) estimaram uma idade inferior para L3f, de cerca de 36.400 ± 12.800.

Os subhaplogrupos L3b e L3d partilham uma origem comum, apresentando uma substituição na posição 16124, mas existe um ramo que partilha as substituições em 16278 e 16362 com L3b, mas não apresenta a 16124. As estimativas para o MRCA de L3b e L3d são de 42.400 ± 12.400, considerando todos aqueles ramos e de 31.900 ± 12.200 anos no ramo partilhando a 16124. Ambos, L3b e L3d, são quase exclusivos do oeste, havendo alguns tipos partilhados com o leste, sudeste e sul, e, no caso de L3b, com oeste-centro. L3b foi identificado como sendo um marcador da expansão Bantu (Salas *et al.*, 2004), daí apresentar algumas linhagens com um claro sinal de expansão, como se pode comprovar no network (Figura 15).

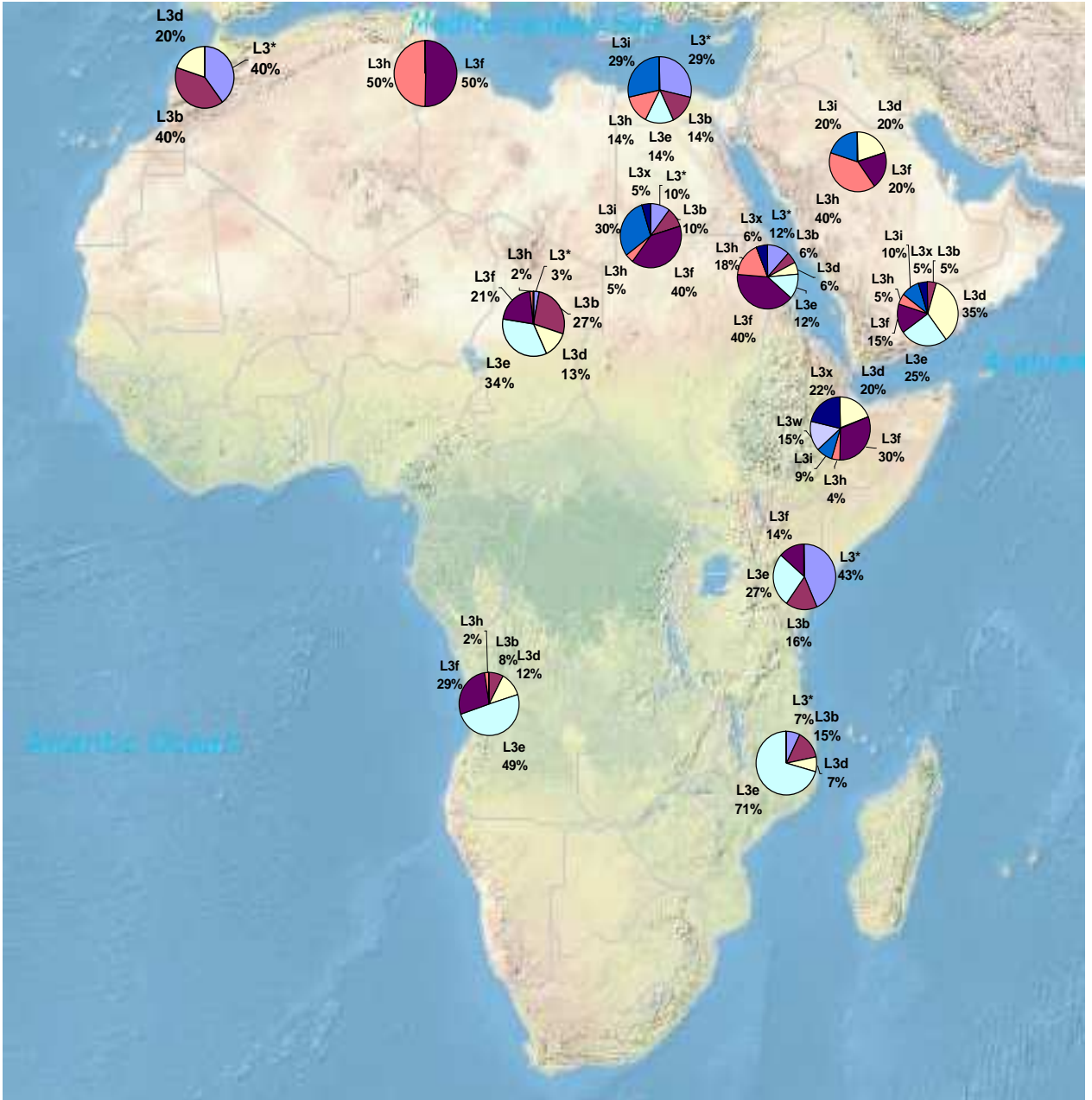


Figura 14 - Distribuição dos subhaplogrupos L3(xM,N) nas populações de África e Península Arábica.

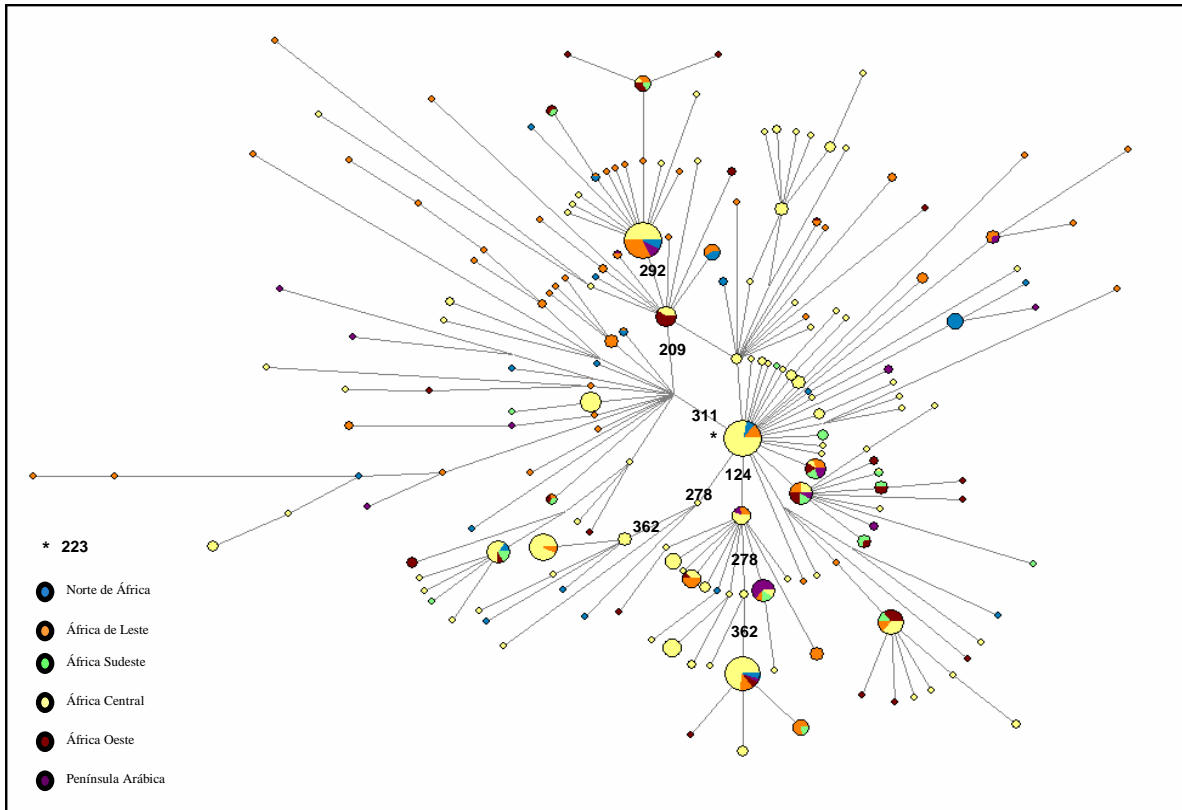


Figura 15 – Reduced network da diversidade em HVRI (entre as posições 16090-16365) para o haplogrupo L3(xM,N) nas populações de África e Península Arábica. O tamanho dos ramos é proporcional ao número de mutações, estando algumas representadas (sem o prefixo 16000) e a área dos círculos proporcional ao número de amostras. O asterisco localiza a raiz do network.

5 – Filogenia das linhagens L3(xM,N) com base na sequência completa de mtDNA

Um total de 22 amostras com provável haplótipo L3(xM,N) quando analisada a diversidade em HVRI foi selecionado para posterior sequenciação completa da molécula, de modo a que pelo menos um dos haplótipos diferentes em HVRI estivesse representado. Destas, 21 pertenciam ao haplogrupo L3(xM,N) e uma era L4 (antes classificada como L3h1?). Os dados obtidos estão reportados no Anexo 3. Estas sequências completas foram usadas para construir duas árvores filogenéticas, juntamente com outras 94 sequências L3(xM,N) e 4 sequências L4 completas já publicadas (Figuras 16 e 17).

A árvore filogenética de L3 está representada na Figura 16. Em geral, as idades estimadas através da diversidade na região codificante são mais antigas e com menor desvio padrão do que as estimadas através da diversidade em HVRI, de acordo com o esperado dada a maior redundância dos polimorfismos em HVRI. Assim sendo, a idade para o haplogrupo L3 é de 69.600 ± 5.900 anos. Da raiz de L3 saem 5 ramos característicos da África Subsariana, L3bcd, L3eix, L3f, L3h e L3a, comprovando a elevada multifurcação, particular deste haplogrupo em África. Os restantes dois ramos a divergirem da raiz de L3 são os macro-haplogrupos M e N.

O ramo L3bcd tem uma idade muito aproximada do haplogrupo L3(xM,N) geral, de 69.500 ± 10.100 anos e é definido pelo polimorfismo na posição 13105 e apenas 4 motivos CA na região 514-523. Os haplogrupos L3b e L3d são muito bem sucedidos, tendo muitas ramificações. O L3d é mais antigo, com uma idade de 43.700 ± 7.200 anos, e encontrando-se disperso por toda a África Subsariana, apesar de a maioria dos indivíduos sequenciados ser do oeste de África ou descendentes destes naturais do Novo Mundo. O haplogrupo L3b é mais recente, com uma idade de 28.000 ± 6.300 anos e com aparente menor dispersão fora do oeste de África. O único representante L3c é proveniente da Etiópia.

O ramo L3eix é mais recente, com 46.200 ± 7.200 anos e definido pelos polimorfismos nas posições 10819 e na altamente recorrente 150. O L3e é um haplogrupo bem sucedido, com cerca de 35.600 ± 4.900 anos. Logo após a emergência deste haplogrupo, deve-se ter originado o sub-ramo L3e3'4'5, seguido próximo de L3e3'4; o L3e5 tem uma idade muito mais recente, 11.600 ± 3.900 anos e parece ser mais abundante no leste de África do que no oeste como acontece para L3e3'4 e restantes subhaplogrupos L3e1 e L3e2. Estes últimos são também recentes, com idades de 17.600 ± 4.800 e $22.900 \pm$

4.600 anos, respectivamente. Os haplogrupos L3i e L3x parecem ser mais frequentes no leste de África e têm idades de 34.700 ± 9.200 e 29.500 ± 7.400 anos, respectivamente.

L3f tem uma idade de 61.100 ± 9.800 anos e o seu subhaplogrupo mais antigo é o L3f1, com 51.000 ± 11.200 anos, que parece ser mais frequente no oeste de África. Pelo contrário, os haplogrupos mais recentes L3f2 e L3f3, com idades de 34.700 ± 10.500 e 20.600 ± 7.300 anos respectivamente, são mais frequentes no leste de África.

O haplogrupo L3h apresenta uma idade muito elevada, de 91.200 ± 12.700 anos porque o subhaplogrupo L3h2 contém muitos polimorfismos. Tal pode ser explicado como um artefacto resultante de este haplogrupo ser raro e ainda estar por descrever muita da sua diversidade ou por ser de facto uma linhagem que por algum motivo acumulou maior diversidade, como foi descrito para a linhagem H4 (Herrnstadt *et al.*, 2002) e L2 (Howell *et al.*, 2004). O subhaplogrupo L3h1 data de 64.200 ± 9.100 anos. Todo o haplogrupo L3h é mais frequente no leste de África.

A única sequência completa L3(xM,N) do Próximo Oriente, pertence a um haplogrupo de origem recente em África, L3d3, sendo pois de provável introdução pós-migração(ões) Out-of-Africa. Todas as linhagens do Novo Mundo são também afiliáveis em subhaplogrupos recentes, nunca constituindo os ramos basais e mais antigos dos diversos subhaplogrupos.

Figura 16 (na próxima página) – Árvore filogenética do haplogrupo L3 para sequências completas. A proveniência das amostras está indicada nas caixas identificadas por diferentes cores: vermelho para o sul de África; verde para o Próximo Oriente; azul para a África leste; amarelo para a África oeste (incluindo as sequências do Novo Mundo); e cinzento para as amostras do Sudão caracterizadas neste trabalho. Nas caixas brancas indicam-se os haplogrupos. Nas caixas creme indicam-se as idades e desvios padrões. As posições recorrentes estão sublinhadas; as que recorrem para o estado ancestral estão indicadas com “!”; a base mutada está indicada, em maiúscula se for transição e em minúscula se for transversão; as inserções/deleções estão indicadas.

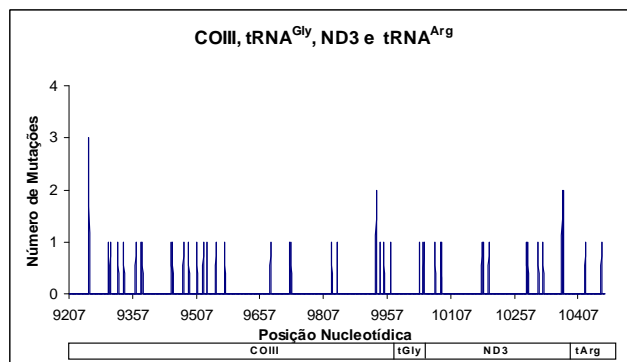
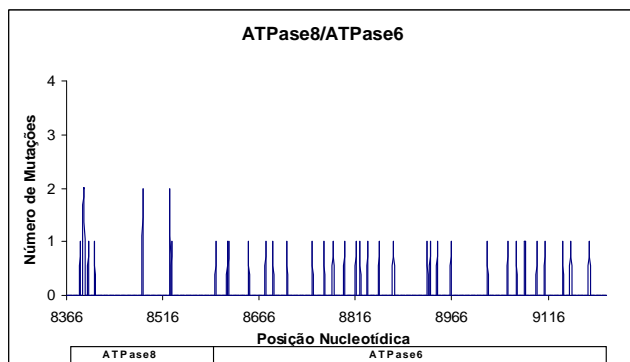
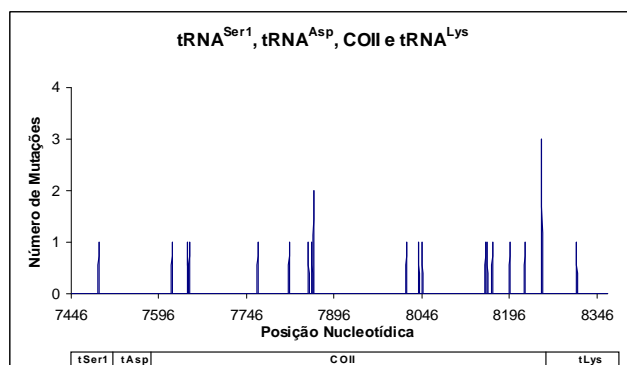
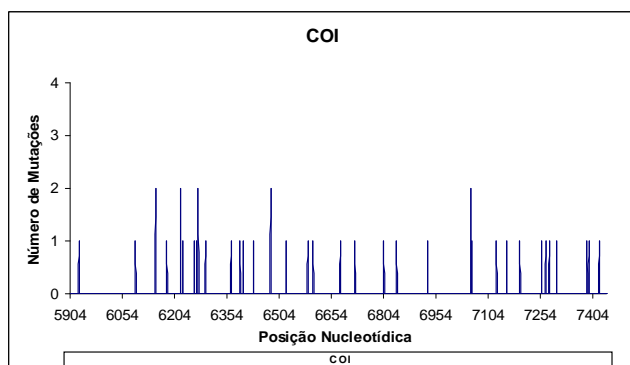
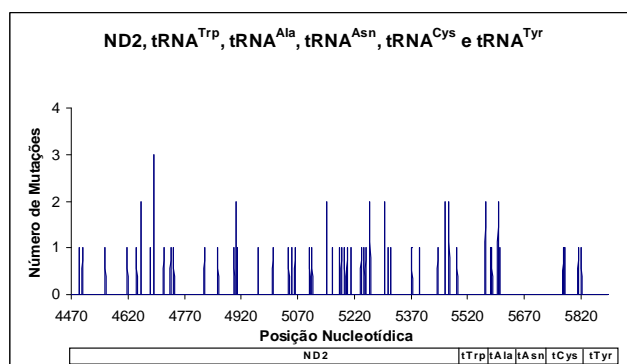
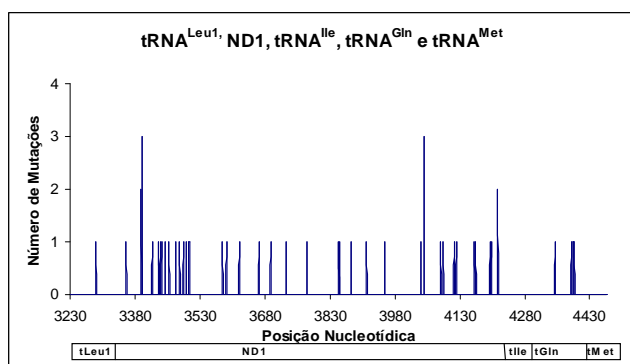
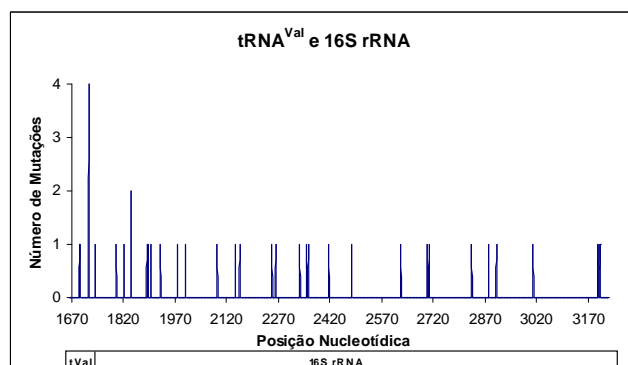
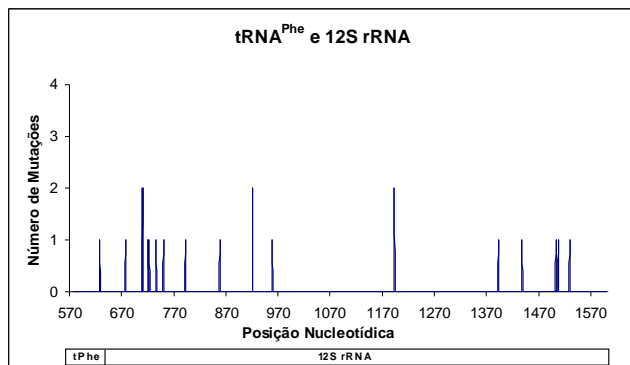
6 – A recorrência na região codificante do mtDNA no haplogrupo L3(xM,N) e implicações para a sua aplicação forense

Nos últimos anos de aplicação de mtDNA aos estudos forenses, tem-se vindo a investir no desenvolvimento de kits de tipagem de SNPs na região codificante, que permitam uma afiliação mais segura dos haplogrupos, baseando-se em polimorfismos definidores e pouco recorrentes. Estes kits usam-se como complemento da caracterização da linhagem mitocondrial por sequenciação de HVRI e HVRII. Mas alguns estudos revelaram que, ao contrário do esperado, existe ainda uma grande heterogeneidade entre as taxas de mutação das posições da região codificante (Bandelt *et al.*, 2006), sendo aconselhável fazer um estudo prévio antes do desenho de um novo kit.

Com vista a um eventual futuro desenvolvimento de um kit de SNPs para sub-caracterização do haplogrupo L3(xM,N), efectuamos o estudo da taxa de recorrência dos polimorfismos da região codificante. Na Figura 18 ilustra-se a taxa de mutação relativa das diversas posições na região codificante e, para comparação, na região controlo, apesar de não serem directamente comparáveis, uma vez que a informação para a região codificante se baseia nos 115 indivíduos, enquanto para a região controlo só em 74 indivíduos. E no anexo 4 reporta-se a listagem do espectro dessas taxas de mutação relativas.

Existe uma certa heterogeneidade nas taxas de mutação relativas para as várias regiões da região codificante. Por exemplo, os genes COXII e 12s rRNA possuem uma baixa diversidade. Em geral, os tRNAs são pouco variáveis, mas existem excepções como o da Prolina e o da Leucina 2. Em HVRII há poucas posições variáveis, mas são altamente recorrentes. Em HVRI há muitas posições polimórficas, sendo umas poucas muito recorrentes (principalmente a 16311). Na região central da região controlo, a posição 16519 é altamente recorrente.

As posições recorrentes dentro do haplogrupo L3(xM,N) diferem das que são mais recorrentes (Freitas e Pereira, 2007) na amostra representativa da diversidade filogenética mitocondrial global, publicada por Ingman *et al.* (2000) (Figura 19), mostrando que os padrões de recorrência podem ser muito variáveis entre os vários haplogrupos. O que mais uma vez justifica um estudo prévio para cada haplogrupo.



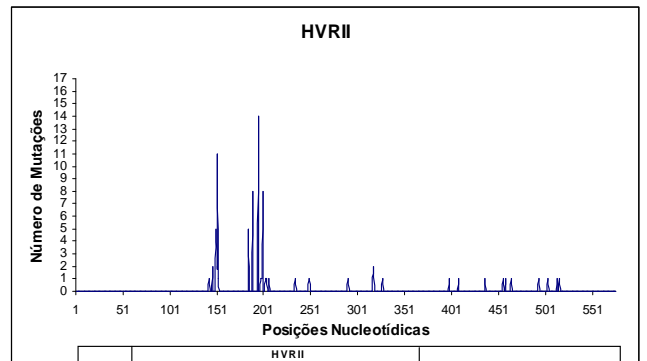
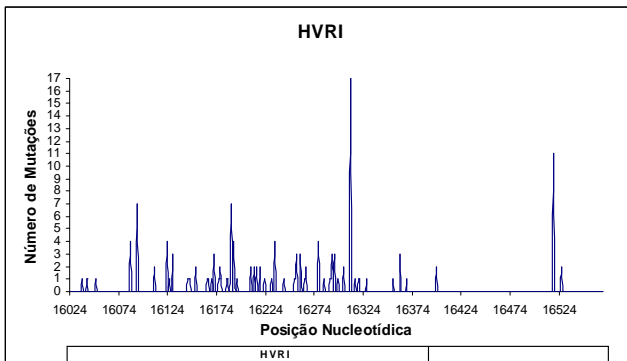
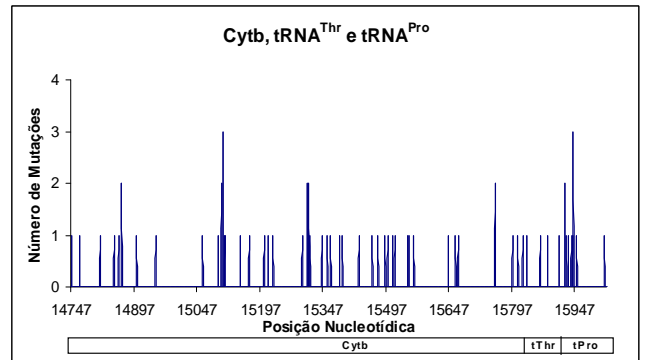
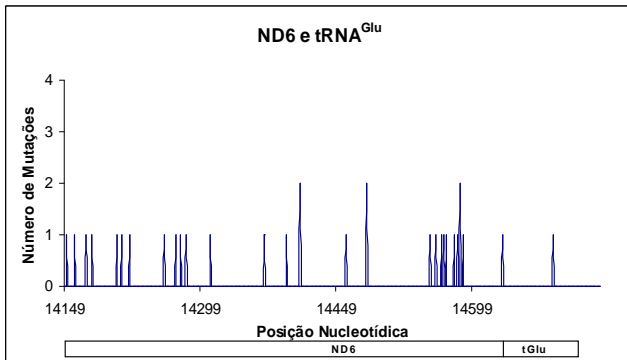
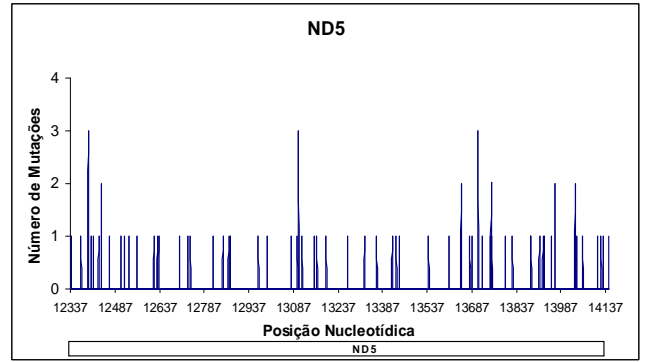
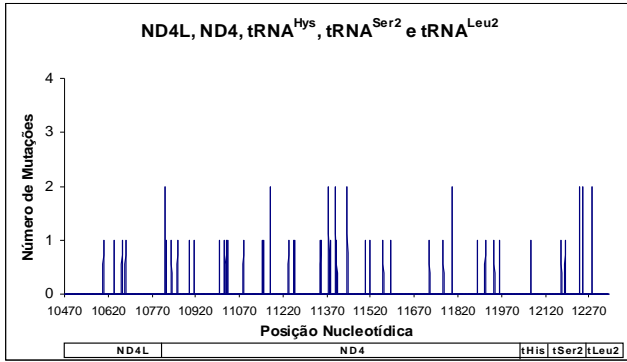
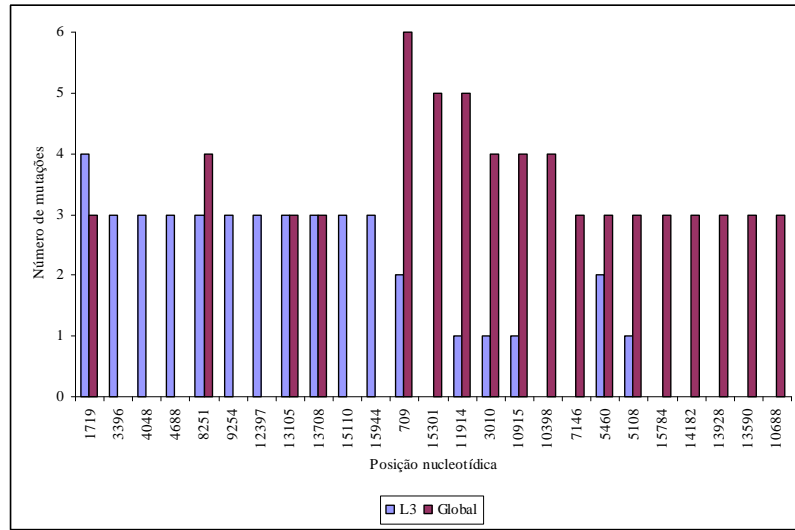
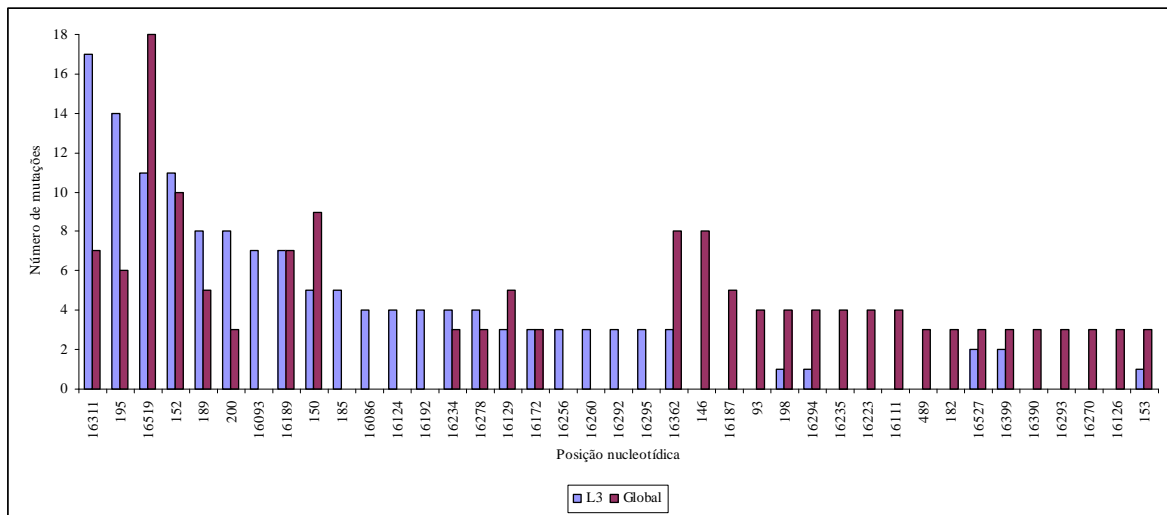


Figura 18 – Taxas de mutação relativas para o haplogrupo L3(xM,N), a partir de 115 indivíduos para a região codificante e de 74 indivíduos para a região controle (HVRI e HVRII).



A



B

Figura 19 – Comparação das taxas de mutação relativas para as posições mais recorrentes da região codificante (A) e da região controlo (B) entre o haplogrupo L3(xM,N) (a azul) e uma amostra global publicada por Ingman *et al.* (2000) (a vermelho).

IV – CONCLUSÕES

1 - Aplicações à genética forense – a diversidade mitocondrial em populações do leste de África

As potencialidades do estudo do mtDNA na resolução de casos forenses estão bem demonstradas. As características peculiares deste genoma implicam contudo um estudo aprofundado dos dados antes da sua aplicação a casos concretos.

Ficou demonstrado nos últimos anos que, para que as bases de dados de mtDNA sejam informativas nas populações euroasiáticas, são necessárias amostragens consideráveis (Pereira *et al.*, 2004). De facto, amostragens de cerca de 500 indivíduos implicam ainda uma considerável partilha de haplótipos e, conseqüentemente, baixa resolução do HVRI e HVRII.

Ademais, uma cuidada análise filogenética dos haplótipos possibilita a detecção de possíveis erros, tais como a troca de amostra em análises separadas de HVRI e HVRII e uma má edição da sequência, contribuindo para um controlo de qualidade essencial na aplicação forense (Salas *et al.*, 2007).

Para o leste de África, só existiam dados obedecendo a estes critérios de qualidade, que têm vindo a ser definidos na aplicação do mtDNA a casos forenses, para o Quênia (Brandstätter *et al.*, 2004). A saber, tipagem simultânea de HVRI e HVRII, uma amostragem razoável, rondando uma centena de indivíduos, e avaliação filogenética dos haplótipos. Neste trabalho diminuámos esse défice ao caracterizarmos a diversidade filogenética em HVRI e HVRII em 102 indivíduos naturais do Sudão.

Os dados de diversidade genética standard obtidos neste trabalho na amostra de 102 indivíduos naturais do Sudão e sua comparação com os dados publicados atestam a elevada diversidade destas populações do leste de África. E, conseqüentemente, a elevada resolução quando se estuda HVRI e HVRII para os haplótipos presentes no leste de África. Para o Sudão, a probabilidade de identidade de haplótipos ao estudar apenas a região HVRI é de 2%, de 4% para HVRII e 1,8% quando se caracterizam as duas regiões simultaneamente. As probabilidades de 2 haplótipos se distinguirem em apenas uma ou duas posições nucleotídicas, o que tem implicações na avaliação da evidência genética forense, especialmente quando as amostras provêm de tecidos diferentes (como por

exemplo, vários talos de cabelo), são ainda muito baixas, rondando os 1,4% e 2,9%, respectivamente. Curiosamente, a probabilidade de identidade de haplótipos entre a amostra do Sudão e as outras amostras revela que o maior valor surge para a comparação com a população Portuguesa e as do Norte de África, o que é devido à elevada proporção do haplótipo CRS nestas populações e à sua presença no Sudão. A partilha de haplótipos entre o Sudão e as populações vizinhas do leste de África é mínima, em geral, e mesmo dentro dos haplogrupos mais frequentes, L2 e L3, como pode ser constatado nos networks apresentados neste trabalho (mesmo tendo estes sido inferidos com base na diversidade no segmento entre as posições 16090 e 16365 em HVRI, uma limitação imposta pela determinação da taxa de mutação ter sido estimada nesta região).

A sequenciação completa de 21 haplótipos L3(xM,N) e sua comparação com 94 amostras previamente publicadas possibilitou constatar a elevada heterogeneidade das taxas de mutação para as diferentes posições nucleotídicas da molécula de mtDNA. Algumas posições localizadas na região codificante apresentam uma considerável taxa de recorrência dentro do haplogrupo L3(xM,N), pelo que devem ser consideradas no desenho de kits de SNaPshot. Se os kits tiverem como objectivo permitir uma afiliação das amostras nos diversos subhaplogrupos L3(xM,N), aquelas posições recorrentes devem ser evitadas; se os kits tiverem como objectivo potenciar a distinção entre haplótipos, aquelas devem ser incluídas. Contudo, para os haplogrupos subsarianos, como ficou demonstrado em L3(xM,N), este segundo tipo de kit não parece ser uma ferramenta tão informativa e necessária como acontece para os haplogrupos eurasiáticos (Coble *et al.*, 2004).

Como seria de esperar, a taxa de recorrência em HVRI e HVRII é muito mais elevada do que na região codificante, sendo mais heterogénea em HVRII do que em HVRI.

A taxa de recorrência em L3(xM,N) difere da taxa de recorrência na diversidade global do mtDNA, o que comprova a necessidade de fazer um estudo particular de cada caso.

2 - Aplicações à genética populacional – inferências para o Out-of-Africa

A confirmar a predominância de indivíduos subsarianos na génese da população Sudanesa verificou-se uma frequência de apenas 21% de linhagens euroasiáticas. Mesmo a influência árabe deve ter tido um impacto genético mínimo, tendo ocorrido

predominantemente uma islamização e arabização da população local. Comparando esta frequência na amostra estudada neste trabalho, com a frequência de 31% reportada para as linhagens euroasiáticas na amostra da Núbia (Klings *et al.*, 1999), parece haver uma estruturação populacional, dentro do Sudão, para estas linhagens, decrescendo a sua frequência em direcção ao sul.

Quanto aos haplogrupos ditos marcadores da migração back-to-Africa (Olivieri *et al.*, 2006), através do Levante, só se observou no Sudão um indivíduo afiliável no haplogrupo U6, numa linhagem que parece ter tido expansão já no leste de África (Olivieri *et al.*, 2006 e dados próprios não publicados), e uma proporção de 5% de M1, cujo centro de expansão parece ter sido a Etiópia (Quintana-Murci *et al.*, 1999).

Todas estas linhagens pertencentes aos macro-haplogrupos M e N são já derivadas, tendo-se confirmado nesta amostra aumentada do Sudão que não existem linhagens basais M e N neste país, como acontece para o restante leste de África. Esta ausência parece ser uma evidência forte de que terão sido linhagens L3 a efectuar a(s) migração(ões) para fora de África, originando só então os supra-haplogrupos M e N. Curiosamente, publicações recentes de dados mais alargados de populações da Península Arábica, nomeadamente, Arábia Saudita (Abu-Amero *et al.*, 2007) e Iémen (Kivisild *et al.*, 2004), têm mostrado também a ausência de linhagens basais de M e N, bem como de haplogrupos L3 particulares. O que parece indicar que a ter ocorrido a onda de migração mais antiga através do Sul da Península Arábica, esta região deve ter funcionado como corredor de passagem e não como centro de expansão. Este dado está de acordo com a elevada taxa de migração estimada por Macaulay *et al.* (2005), de cerca de 4km/ano, muito superior à taxa de 0,7km/ano calculada para a re-expansão na Europa após o Último Máximo Glaciar.

Do pool genético subsariano sudanês, estão presentes em baixa frequência os haplogrupos L0, L4 e L5, os quais também são raros na extensa maioria de África (L0 é apenas mais frequente nos povos Khoisan do sul de África; Pereira *et al.*, 2001; Salas *et al.*, 2002), predominando os haplogrupos L2 e L3(xM,N), com uma frequência de 32-34% cada.

A análise do haplogrupo L2 revelou uma elevada reticulação para algumas das posições HVRI, tendo uma idade de 80.200 ± 21.300 anos. O subhaplogrupo mais frequente no Sudão é o L2a, como é verificado para o resto do continente, e a idade estimada é de 59.600 ± 19.200 anos, próxima da apresentada por Salas *et al.* (2002) de 55.150 ± 19.350 ; as linhagens deste haplogrupo geralmente associadas à expansão Bantu (Pereira *et al.*, 2001) não foram observadas no Sudão. Os poucos exemplos de linhagens L2b, L2c e L2d

observados no Sudão podem ser introduções recentes devidas à migração de escravos principalmente originais da África Ocidental. Estas linhagens também são observadas na Península Arábica, o que pode indicar de igual modo a migração recente de escravos da África ocidental ou através do leste de África.

Para o haplogrupo L3, a caracterização da sequência completa de mtDNA em 21 haplótipos do Sudão e a sua comparação com 94 sequências completas permitiu contribuir para uma afinação da datação destas linhagens L3. Comprovou-se que as idades estimadas com base na diversidade presente na região codificante são mais antigas que as estimadas com base na diversidade em HVRI, o que pode ser explicado pela saturação de mutação em HVRI em linhagens já antigas como são todas as subsarianas. Portanto, é de esperar que as idades estimadas para L2 neste trabalho sejam estimativas por deficiência, devendo ser interpretadas como limites mínimos.

A idade estimada para o haplogrupo L3 foi de 69.600 ± 5.900 anos, estabelecendo o limite máximo para o Out-of-Africa. Identificaram-se cinco ramos subsarianos a partir da raiz, para além dos dois ramos que constituem os macro-haplogrupos M e N. As idades estimadas para três desses cinco ramos subsarianos são muito próximas do evento que originou o próprio haplogrupo L3, testemunhando a elevada expansão demográfica ocorrida neste período, no leste de África; 69.500 ± 10.100 anos para L3bcd, 61.100 ± 9.800 anos para L3f e 64.200 ± 9.100 anos para L3h1 (a estimativa para L3h em geral pode estar ainda enviesada por um efeito de selecção ou de amostragem em L3h2). A localização geográfica destes haplogrupos é também importante: L3bcd parece estar distribuído por toda a África, predominando na África ocidental, especialmente o L3b; L3f está também bastante disperso, predominando L3f1 no oeste, enquanto L3f2 e L3f3 se localizam no leste; L3h parece ser bastante localizado no leste. O ramo L3eix apresenta uma idade mais recente de 46.200 ± 7.200 anos e enquanto L3i e L3x predominam no leste, L3e parece ter-se originado a oeste, havendo no entanto o subgrupo L3e5 que parece ter-se expandido mais a leste. O ramo L3a é muito raro, não tendo sido ainda possível efectuar uma estimativa fidedigna da sua idade e localização predominante.

Curiosamente, existem muitos subhaplogrupos predominantes no leste de África cuja idade de expansão aponta para cerca de 30.000 anos: 34.700 ± 9.200 anos para L3i, 29.500 ± 7.400 anos para L3x e 34.700 ± 10.500 anos para L3f2. Esta data é comparável com a estimativa de expansão do típico haplogrupo do leste de África M1a de 28.800 ± 4.900 anos estimada por Olivieri *et al.* (2006). O que parece indicar uma outra acentuada expansão demográfica no leste de África por volta desta data. Não há indícios arqueológicos

particulares para esta época, tirando o facto de o período designado por *Middle Stone Age* ter acabado por volta dos 35.000 anos. Este período, caracterizado por produções líticas em que os núcleos são preparados antes das pontas serem removidas e retocadas, foi iniciado há 100.000 anos, mas pode ser mais antigo na África do Sul, indo até 200.000 anos. Outra característica importante foi a existência de diversos períodos alternados de períodos pluviosos e períodos de seca, associados a expansões e contracções demográficas.

O declínio máximo deve ter ocorrido há 50.000-40.000 anos atrás, quando uma profunda aridez se tornou extensa (Newman, 1955). Esse sinal de expansão demográfica detectado no mtDNA do leste de África por volta dos 30.000 anos pode reflectir uma recuperação demográfica após aquele período de aridez.

O facto da sequenciação completa de uma amostra que se classificaria como “L3h1?”, com base em HVRI e HVRII, ter revelado um ramo que partilha o polimorfismo em 5460 com L4a, para além dos polimorfismos basais de L4, mostra que muita da diversidade deste haplogrupo pode estar ainda por revelar, sendo mal afiliado em L3(xM,N) com base na diversidade das regiões hipervariáveis.

A caracterização da sequência completa de haplótipos L3(xM,N) na Península Arábica e Próximo Oriente e a sua comparação com estas e outras sequências completas do leste de África será essencial num futuro próximo. Ela fornecerá os meios de avaliar qual a proporção destes haplótipos que pode ser explicada por migração antiga, há volta dos 60.000 anos e 40.000 anos atrás aquando as duas prováveis ondas de migração Out-of-Africa, distinguindo-a da proporção de haplótipos que teve que ser necessariamente introduzida após esse período, por se tratar de diversidade recentemente gerada em África.

V - REFERÊNCIAS BIBLIOGRÁFICAS

- Abu-Amero, K.K., Gonzalez, A.M., Larruga, J.M., Bosley, T.M e Cabrera, V.M. (2007). Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. *BMC Evol Biol.* **7**:32.
- Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., Cruciani, F., Zeviani, M., Briem, E., Carelli, V., Moral, P., Dugoujon, J.M., Roostalu, U., Loogvali, E.L., Kivisild, T., Bandelt, H.-J., Richards, M., Villems, R., Santachiara-Benerecetti, A.S., Semino, O. e Torroni, A. (2004). The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet.* **75**:910-918.
- Amorim, A. (1999). Archaeogenetics. *J. Iberian Archaeol.* **1**:15-25.
- Amorim, A. e Pereira, L. (2005). Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. *Forensic Sci Int.* **150**:17-21.
- Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R. e Young, I.G. (1981). Sequence and organization of the human mitochondrial genome. *Nature.* **290**:457-465.
- Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. e Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet.*: **23**:147.
- Bandelt, H.-J., Forster, P., Sykes, B.C. e Richards, M.B. (1995). Mitochondrial portraits of human populations. *Genetics* .**141**:743-753.
- Bandelt, H.-J., Forster, P. e Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**:37-48.

- Bandelt, H.-J., Kong, Q.-P., Yao, Y.-G., Richards, M. e Macaulay, V. (2006). Estimation of mutation rates and coalescence times: some caveats. In: Bandelt, H.-J., Macaulay, V. e Richards, M. (eds). *Mitochondrial DNA and the Evolution of Homo sapiens*. Springer-Verlag, Berlin.
- Behar, D.M., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E., Scozzarri, R., Makkan, H., Tzur, S., Comas, D., Bertranpetit, J., Quintana-Murci, L., Tyler-Smith, C., Spencer Wells, R. e Rosset, S. e The Genographic Consortium. The dawn of human matrilineal diversity: Insights from complete mitochondrial genomes. (submetido).
- Beleza, S., Gusmao, L., Amorim, A., Carracedo, A. e Salas, A. (2005). The genetic legacy of western Bantu migrations. *Hum Genet.* **117**:366-375.
- Belle, E.M., Piganeau, G., Gardner, M. e Eyre-Walker, A. (2005). An investigation of the variation in the transition bias among various animal mitochondrial DNA. *Gene.* **355**:58-66.
- Belledi, M., Simoni, L., Casalotti, R. e Destro-Bisol, G. (2000). Male and female differential patterns of genetic variation in human populations. In: Renfrew, C. e Boyle, K. (eds) *Archaeogenetics: DNA and the population prehistory of Europe*. McDonald Institute Monographs. Oxbow Books. Oxford. UK. pp 295-300.
- Birky, C.W. Jr. (2001). The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annu Rev Genet.* **35**:125-148.
- Brandstatter, A., Peterson, C.T., Irwin, J.A., Mpoke, S., Koech, D.K., Parson, W. e Parsons, T.J. (2004). Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database. *Int J Legal Med.* **118**:294-306.
- Brakez, Z., Bosch, E., Izaabel, H., Akhayat, O., Comas, D., Bertranpetit, J. e Calafell, F. (2001). Human mitochondrial DNA sequence variation in the Moroccan population of the Souss area. *Ann Hum Biol.* **28**:295-307.

- Brinkmann, B., Klintschar, M., Neuhuber, F., Huhne, J. e Rolf, B. (1998). Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet.* **62**:1408-1415.
- Brown, W.M., George, M. Jr. e Wilson, A.C. (1979). Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci USA.* **76**:1967-1971.
- Brown, W.M. (1980). Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc Natl Acad Sci USA.* **77**:3605-3609.
- Bubbenzer, O. e Riemer, H. (2007). Holocene climatic change and human settlement between the Central Sahara and the Nile Valley: archaeological and geomorphological results. *Geoarchaeology.* **22**:607-620.
- Cann, R.L., Stoneking, M. e Wilson, A.C. (1987). Mitochondrial DNA and human evolution. *Nature.* **325**:31-36.
- Cerny, V., Salas, A., Hajek, M., Zaloudkova, M. e Brdicka, R. (2007). A bidirectional corridor in the Sahel-Sudan belt and the distinctive features of the Chad Basin populations: a history revealed by the mitochondrial DNA genome. *Ann Hum Genet.* **71**:433-452.
- Chen, Y.-S., Olckers, A., Schurr, T.G., Kogelnik, A.M., Huoponen, K. e Wallace, D.C. (2000). mtDNA variation in the South African Kung and Khwe and their genetic relationships to other African populations. *Am J Hum Genet* **66**:1362-1383.
- Cherni, L., Loueslati, B.Y., Pereira, L., Ennafaa, H., Amorim, A. e El Gaaied, A.B. (2005). Female gene pools of Berber and Arab neighboring communities in central Tunisia: microstructure of mtDNA variation in North Africa. *Hum Biol.* **77**:61-70.
- Coble, M.C., Just, R.S., O'Callaghan, J.E., Letmanyi, I.H., Peterson, C.T., Irwin, J.A. e Parsons, T.J. (2004). Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in the Caucasians. *Int J Legal Med* **118**:137-146.

- Diez-Sanchez, C., Ruiz-Pesini, E., Lapena, A.C., Montoya, J., Perez-Martos, A., Enriquez, J.A. e Lopez-Perez, M.J. (2003). Mitochondrial DNA content of human spermatozoa. *Biol Reprod.* **68**:180-185.
- Edwards, D.N. (2007). The archaeology of Sudan and Nubia. *Annu Rev Anthropol.* **36**:211-228.
- Excoffier, L., Laval, G. e Schneider, S. (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**:47-50.
- Excoffier, L., Laval, G. e Schneider, S. (2005). Arlequin ver. 3.1 *User Manual. Computational and Molecular Population Genetics Lab. University of Berne. Switzerland.* Pp145.
- Excoffier, L., Smouse, P. e Quattro, J. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131**:479-491.
- Forster, P., Harding, R., Torroni, A. e Bandelt, H.-J. (1996). Origin and Evolution of Native American mtDNA Variation: A reappraisal. *American Journal of Human Genetics* **59**:935-945.
- Freitas, F. e Pereira, L. (2007). Heterogeneity in coding mtDNA mutation rates: implications in forensic genetics. *Forensic Sci Int:Genetics* (em publicação).
- Fu, Y.-X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**:915-925.
- Gilbert, M.T. (2006). Postmortem Damage of mitochondrial DNA. In: Bandelt, H.-J., Macaulay, V. e Richards, M. (eds). *Human Mitochondrial DNA and the evolution of Homo sapiens.* Springer-Verlag. Berlin, Heidelberg. Pp 91-115.
- Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* **41**:95–98.

- Herrnstadt, C., Elson, J.L., Fahy, E., Preston, G., Turnbull, D.M., Anderson, C., Ghosh, S.S., Olefsky, J.M., Beal, M.F., Davis, R.E. e Howell, N. (2002). Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet.* **70**:1152-1171.
- Howell, N., Elson, J.L., Turnbull, D.M. e Herrnstadt, C. (2004). African Haplogroup L mtDNA sequences show violations of clock-like evolution. *Mol Biol Evol.* **21**:1843-1854.
- Hudson, R.R. (1990). Gene genealogies and the coalescent process. In Futuyama D, Antonovics J (eds). *Oxford Surveys in Evolutionary Biology*. Oxford University Press, New York, pp 1-44.
- Ingman, M., Kaessmann, H., Paabo, S. e Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature.* **408**:708-713.
- Jeffreys, A.J., Wilson, V. e Thein, S.L. (1985). Individual-specific 'fingerprints' of human DNA. *Nature.* **316**:76-79.
- Jenuth, J.P., Peterson, A.C., Fu, K. e Shoubridge, E.A. (1996). Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. *Nat Genet.* **14**:146-151.
- Kivisild, T., Reidla, M., Metspalu, E., Rosa, A., Brehm, A., Pennarun, E., Parik, J., Geberhiwot, T., Usanga, E. e Villems, R. (2004). Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *Am J Hum Genet.* **75**:752-770.
- Kivisild, T., Shen, P., Wall, D.P., Do, B., Sung, R., Davis, K., Passarino, G., Underhill, P.A., Scharfe, C., Torroni, A., Scozzari, R., Modiano, D., Coppa, A., de Knijff, P., Feldman, M., Cavalli-Sforza, L.L. e Oefner, P.J. (2006). The role of selection in the evolution of human mitochondrial genomes. *Genetics.* **172(1)**:373-87.
- Kraytsberg, Y., Schwartz, M., Brown, T.A., Ebralidse, K., Kunz, W.S., Clayton, D.A., Vissing, J. e Khrapko, K. (2004). Recombination of human mitochondrial DNA. *Science.* **304**:981.

- Krings, M., Stone, A., Schmitz, R.W., Krainitzki, H., Stoneking, M. e Paabo, S. (1997). Neandertal DNA sequences and the origin of modern humans. *Cell*. **90**:19-30.
- Krings, M., Salem, A.E., Bauer, K., Geisert, H., Malek, A.K., Chaix, L., Simon, C., Welsby, D., Di Rienzo, A., Utermann, G., Sajantila, A., Paabo, S. e Stoneking, M. (1999). mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration? *Am J Hum Genet*. **64**:1166-1176.
- Lareu, M.V., Phillips, C.P., Carracedo, A., Lincoln, P.J., Syndercombe-Court, D. e Thomson, J.A. (1994). Investigation of the STR locus HUMTH01 using PCR and two electrophoresis formats: UK and Galician Caucasian population surveys and usefulness in paternity investigation. *Forensic Sci. Int.* **25**;66(1):41-52.
- Maca-Meyer, N., Gonzalez, A.M., Larruga, J.M., Flores, C. e Cabrera, V.M. (2001). Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet*. **2**:13.
- Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonne-Tamir, B., Sykes, B. e Torroni, A. (1999). The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet*. **64**:232-249.
- Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., Taha, A., Shaari, N.K., Raja, J.M., Ismail, P., Zainuddin, Z., Goodwin, W., Bulbeck, D., Bandelt, H.J., Oppenheimer, S., Torroni, A. e Richards, M. (2005). Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science*. **308**:1034-1036.
- McDougall, I., Brown, F.H. e Fleagle, J.G. (2005). Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature*. **433**:733-736.
- Mellars, P. (2006). Why did modern human populations disperse from Africa 60,000 years ago? A new model. *Proc Natl Acad Sci USA*. **103**:9381-9386.

- Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M.D., Sukernik, R.I., Olckers, A. e Wallace, D.C. (2003). Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA*. **100**:171-176.
- Monson, K.L., Miller, K.W.P., Wilson, M.R., DiZinno, J.A. e Budowle, B. (2002). The mtDNA population database: an integrated software and database resource for forensic comparison. *Forensic Sci Commun* **4** (no. 2).
- Nachman, M.W. e Crowell, S.L. (2000). Estimate of the mutation rate per nucleotide in humans. *Genetics*. **156**:297-304.
- Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY, USA.
- Newman, J.L. (1995). *The peopling of Africa: a geographic interpretation*. Yale University Press.
- Olivieri, A., Achilli, A., Pala, M., Battaglia, V., Fornarino, S., Al-Zahery, N., Scozzari, R., Cruciani, F., Behar, D.M., Dugoujon, J.M., Coudray, C., Santachiara-Benerecetti, A.S., Semino, O., Bandelt, H.J. e Torroni, A. (2006). The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. *Science*. **314**:1767-1770.
- Pereira, L., Macaulay, V., Torroni, A., Scozzari, R., Prata, M.J. e Amorim, A. (2001). Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann Hum Genet*. **65**:439-458.
- Pereira, L., Cunha, C. e Amorim, A. (2004). Predicting sampling saturation of mtDNA haplotypes: an application to an enlarged Portuguese database. *Int J Legal Med*. **118**:132-136.

- Pereira, L., Cunha, C., Alves, C. e Amorim, A. (2005a). African female heritage in Iberia: a reassessment of mtDNA lineage distribution in present times. *Hum Biol.* **77**:213-229.
- Pereira, L., Richards, M., Goios, A., Alonso, A., Albarran, C., Garcia, O., Behar, D.M., Golge, M., Hatina, J., Al-Gazali, L., Bradley, D.G., Macaulay, V. e Amorim, A. (2005b). High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res.* **15**:19-24.
- Quintana-Murci, L., Semino, O., Bandelt, H.-J., Passarino, G., McElreavey, K. e Santachiara-Benerecetti, A.S. (1999) Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nat Genet.* **23**:437-441.
- Renfrew, C. (2000). *Archaeogenetics: towards a population prehistory of Europe*. In: Renfrew C, Boyle K (eds) *Archaeogenetics: DNA and the population prehistory of Europe*. McDonald Institute Monographs. Oxbow Books. Oxford. UK. pp 3-11.
- Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo C, Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Rienzo, A., Novelletto, A., Oppenheim, A., Norby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Scozari, R., Torroni, A. e Bandelt, H.-J. (2000). Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet.* **67**:1251-76.
- Roewer, L., Krawczak, M., Willuweit, S., Nagy, M., Alves, C., Amorim, A., Anslinger, K., Augustin, C., Betz, A., Bosch, E., Caglia, A., Carracedo, A., Corach, D., Dekairelle, A.F., Dobosz, T., Dupuy, B.M., Furedi, S., Gehrig, C., Gusmao, L., Henke, J., Henke, L., Hidding, M., Hohoff, C., Hoste, B., Jobling, M.A., Kargel, H.J., de Knijff, P., Lessig, R., Liebeherr, E., Lorente, M., Martinez-Jarreta, B., Nievas, P., Nowak, M., Parson, W., Pascali, V.L., Penacino, G., Ploski, R., Rolf, B., Sala, A., Schmidt, U., Schmitt, C., Schneider, P.M., Szibor, R., Teifel-Greding, J. e Kayser, M. (2001). Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. *Forensic Sci Int.* **118**:106-13.

- Rogers, A.R. e Harpending, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* **9**:552-569.
- Roostalu, U., Kutuev, I., Loogväli, E.L., Metspalu, E., Tambets, K., Reidla, M., Khusnutdinova, E.K., Usanga, E., Kivisild, T. e Villems, R. (2007). Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. *Mol Biol Evol.* **24**:436-448.
- Salas, A., Richards, M., De la Fe, T., Lareu, M.V., Sobrino, B., Sanchez-Diz, P., Macaulay, V. e Carracedo, A. (2002). The making of the African mtDNA landscape. *Am J Hum Genet.* **71**:1082-111.
- Salas, A., Richards, M., Lareu, M.V., Scozzari, R., Coppa, A., Torroni, A., Macaulay, V. e Carracedo, A. (2004). The African diaspora: mitochondrial DNA and the Atlantic slave trade. *Am J Hum Genet.* **74**:454-465.
- Salas, A., Bandelt, H.J., Macaulay, V. e Richards, M.B. (2007). Phylogeographic investigations: the role of trees in forensic genetics. *Forensic Sci Int.* **168**:1-13.
- Schwartz, M. e Vissing, J. (2002). Paternal inheritance of mitochondrial DNA. *N Engl J Med.* **347**:576-580.
- Sobrino, B., Brión, M. e Carracedo, A. (2005). SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci Int.* **154**:181-194.
- St John, J.C. (2002). The transmission of mitochondrial DNA following assisted reproductive techniques. *Theriogenology.* **57**:109-123.
- Stringer, C.B. e Andrews, P. (1988). Genetic and fossil evidence for the origin of modern humans. *Science.* **239**:1263-1268.
- Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**: 437-460.

- Tajima, F. (1993). Measurement of DNA polymorphism. In Takahata, N. (author) Clark, A.G. (editor) *Mechanisms of Molecular Evolution: Introduction to Molecular Paleopopulation Biology*. Tokyo, Sunderland, MA: Japan Scientific Societies Press, Sinauer Associates, Inc., p. 37-59.
- Thangaraj, K., Chaubey, G., Kivisild, T., Reddy, A.G., Singh, V.K., Rasalkar, A.A. e Singh, L. (2005). Reconstructing the origin of Andaman Islanders. *Science*. 308:996.
- Torroni, A., Schurr, T.G., Cabell, M.F., Brown, M.D., Neel, J.V., Larsen, M., Smith, D.G., Vullo, C.M. e Wallace, D.C. (1993). Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet*. **53**:563-590.
- Torroni, A., Achilli, A., Macaulay, V., Richards, M. e Bandelt, H.J. (2006). Harvesting the fruit of the human mtDNA tree. *Trends Genet*. **22**:339-345.
- Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. e Wilson, A.C. (1991). African populations and the evolution of human mitochondrial DNA. *Science*. **253**:1503-1507.
- Weir, B.S. e Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**:1358-1370.
- White, T.D., Asfaw, B., DeGusta, D., Gilbert, H., Richards, G.D., Suwa, G. e Howell, F.C. (2003). Pleistocene Homo sapiens from Middle Awash, Ethiopia. *Nature*. **423**:742-747.
- Willuweit, S. e Roewer, L. (2007). On behalf of the International Forensic Y Chromosome User Group Y chromosome haplotype reference database (YHRD): Update. Forensic Science International: *Genetics* **1**:83-87.

VI-ANEXOS

Anexo 1 – Sequências completas ou região codificante de L3 e L4 para comparação com as obtidas neste trabalho.

Designação	Designação Original	Referência	Região	País	Haplogrupo
L3-1	AF346967	Ingman et al. 2000	Oeste de África	sem informação	L3e3b
L3-2	AF346980	Ingman et al. 2000	Oeste de África	sem informação	L3e1f
L3-3	AF346994	Ingman et al. 2000	Oeste de África	sem informação	L3e2b
L3-4	AF347000	Ingman et al. 2000	Este de África	sem informação	L3h1a2a
L3-5	AF347014	Ingman et al. 2000	Oeste de África	sem informação	L3d1a
L3-6	AF347015	Ingman et al. 2000	Oeste de África	sem informação	L3e2b1
L3-7	AF381991	Maca-Meyer et al. 2001	Oeste de África	Mauritânia	L3b1a
L3-8	AF381998	Maca-Meyer et al. 2001	Próximo Oriente	Jordânia	L3d3
L3-9	AY195782	Mishmar et al. 2003	África do Sul	África do Sul	L3d1d
L3-10	AY195784	Mishmar et al. 2003	África do Sul	África do Sul	L3b1a
L3-11	104	Howell et al. 2004	Oeste de África	EUA	L3b1a
L3-12	108	Howell et al. 2004	Oeste de África	EUA	L3b1a
L3-13	136	Howell et al. 2004	Oeste de África	EUA	L3e2b
L3-14	140	Howell et al. 2004	Oeste de África	EUA	L3d1c
L3-15	155	Howell et al. 2004	Oeste de África	EUA	L3b1
L3-16	159	Howell et al. 2004	Oeste de África	EUA	L3e1a
L3-17	163	Howell et al. 2004	Oeste de África	EUA	L3e4
L3-18	164	Howell et al. 2004	Oeste de África	EUA	L3b1
L3-19	166	Howell et al. 2004	Oeste de África	EUA	L3e2b
L3-20	180	Howell et al. 2004	Oeste de África	EUA	L3e3a
L3-21	189	Howell et al. 2004	Oeste de África	EUA	L3b1
L3-22	196	Howell et al. 2004	Oeste de África	EUA	L3e2a
L3-23	211	Howell et al. 2004	Oeste de África	EUA	L3e1a
L3-24	216	Howell et al. 2004	Oeste de África	EUA	L3e1
L3-25	217	Howell et al. 2004	Oeste de África	EUA	L3e1a
L3-26	242	Howell et al. 2004	Oeste de África	EUA	L3e3b
L3-27	309	Howell et al. 2004	Oeste de África	EUA	L3b1a
L3-28	381	Howell et al. 2004	Oeste de África	EUA	L3d1b
L3-29	387	Howell et al. 2004	Oeste de África	EUA	L3e1a
L3-30	413	Howell et al. 2004	Oeste de África	EUA	L3d1'2'3
L3-31	578	Howell et al. 2004	Oeste de África	EUA	L3f1b1
L3-32	579	Howell et al. 2004	Oeste de África	EUA	L3f1b
L3-33	580	Howell et al. 2004	Oeste de África	EUA	L3f1b1
L3-34	581	Howell et al. 2004	Oeste de África	EUA	L3f1b3
L3-35	582	Howell et al. 2004	Oeste de África	EUA	L3e2a

L3-36	583	Howell et al. 2004	Oeste de África	EUA	L3e2b
L3-37	584	Howell et al. 2004	Oeste de África	EUA	L3f1b
L3-38	DQ112710	Kivisild et al. 2006	Oeste de África	República Dominicana	L3b2
L3-39	DQ112711	Kivisild et al. 2006	Oeste de África	República Dominicana	L3b1a
L3-40	DQ112712	Kivisild et al. 2006	Oeste de África	República Dominicana	L3b1a
L3-41	DQ112713	Kivisild et al. 2006	Oeste de África	República Dominicana	L3d2
L3-42	DQ112714	Kivisild et al. 2006	Oeste de África	República Dominicana	L3d1d
L3-43	DQ112715	Kivisild et al. 2006	Oeste de África	República Dominicana	L3d1b
L3-44	DQ112716	Kivisild et al. 2006	Oeste de África	República Dominicana	L3d3
L3-45	DQ112717	Kivisild et al. 2006	Oeste de África	República Dominicana	L3f1b4
L3-46	DQ112718	Kivisild et al. 2006	Oeste de África	República Dominicana	L3h1b1
L3-47	DQ112719	Kivisild et al. 2006	Oeste de África	República Dominicana	L3e1b
L3-48	DQ112720	Kivisild et al. 2006	Oeste de África	República Dominicana	L3e2b2
L3-49	DQ112721	Kivisild et al. 2006	Oeste de África	República Dominicana	L3e2b
L3-50	DQ112722	Kivisild et al. 2006	Oeste de África	República Dominicana	L3e4
L3-51	DQ112726	Kivisild et al. 2006	Oeste de África	Burkina	L3e2a
L3-52	DQ112728	Kivisild et al. 2006	Oeste de África	Burkina	L3e2a
L3-53	DQ112729	Kivisild et al. 2006	Oeste de África	Burkina	L3e2a
L3-54	DQ112734	Kivisild et al. 2006	Oeste de África	Burkina	L3e2b1
L3-55	DQ112736	Kivisild et al. 2006	Oeste de África	Burkina	L3f1b
L3-56	DQ112738	Kivisild et al. 2006	Oeste de África	Burkina	L3b1b
L3-57	DQ112739	Kivisild et al. 2006	Oeste de África	Burkina	L3d1b
L3-58	DQ112740	Kivisild et al. 2006	Oeste de África	Burkina	L3e2a
L3-59	DQ112742	Kivisild et al. 2006	Oeste de África	Burkina	L3b1a
L3-60	DQ112743	Kivisild et al. 2006	Oeste de África	Burkina	L3b1a
L3-61	DQ112744	Kivisild et al. 2006	Oeste de África	Burkina	L3d1a
L3-62	DQ112745	Kivisild et al. 2006	Oeste de África	Burkina	L3b1b

L3-63	DQ112746	Kivisild et al. 2006	Oeste de África	Burkina	L3b2
L3-64	DQ112747	Kivisild et al. 2006	Oeste de África	Burkina	L3e3b
L3-65	DQ112748	Kivisild et al. 2006	Oeste de África	Burkina	L3b1a
L3-66	DQ112756	Kivisild et al. 2006	Oeste de África	Burkina	L3d2
L3-67	DQ112757	Kivisild et al. 2006	Oeste de África	Burkina	L3e5
L3-68	DQ112847	Kivisild et al. 2006	África do Sul	sem informação	L3d3a
L3-69	DQ112851	Kivisild et al. 2006	África do Sul	sem informação	L3d3a
L3-70	DQ112853	Kivisild et al. 2006	África do Sul	sem informação	L3d3a
L3-71	DQ112857	Kivisild et al. 2006	África do Sul	sem informação	L3d3a
L3-72	DQ112884	Kivisild et al. 2006	Oeste de África	sem informação	L3d
L3-73	DQ112949	Kivisild et al. 2006	Este de África	Sudão	L3d1'2'3
L3-74	DQ112956	Kivisild et al. 2006	Este de África	Sudão	L3f2
L3-75	DQ112957	Kivisild et al. 2006	Este de África	Sudão	L3h1a1
L3-76	DQ112958	Kivisild et al. 2006	Este de África	Sudão	L3i1a
L3-77	DQ112960	Kivisild et al. 2006	Este de África	Sudão	L3i1a
L3-78	DQ112961	Kivisild et al. 2006	Este de África	Sudão	L3h1a1
L3-79	lcl_10_L3x(Tor82)	Torrioni et al. 2006	Este de África	Etiópia	L3x1
L3-80	lcl_11_L3i(Tor70)	Torrioni et al. 2006	Este de África	Etiópia	L3i2
L3-81	lcl_12_L3i(Tor69)	Torrioni et al. 2006	Este de África	Etiópia	L3i1b
L3-82	lcl_13_L3e5(Tor79)	Torrioni et al. 2006	Este de África	Etiópia	L3e5
L3-83	lcl_14_L3e(Tor5)	Torrioni et al. 2006	Este de África	Etiópia	L3e2b
L3-84	lcl_15_L3d1(Tor4)	Torrioni et al. 2006	Este de África	Etiópia	L3d1b
L3-85	lcl_16_L3b(Tor6)	Torrioni et al. 2006	Este de África	Etiópia	L3b1a
L3-86	lcl_17_L3c(Tor84)	Torrioni et al. 2006	Este de África	Etiópia	L3c
L3-87	lcl_18_L3f(Tor80)	Torrioni et al. 2006	Este de África	Etiópia	L3f3
L3-88	lcl_19_L3f(Tor83)	Torrioni et al. 2006	Este de África	Etiópia	L3f2a
L3-89	lcl_20_L3f1(Tor75)	Torrioni et al. 2006	Este de África	Etiópia	L3f1b
L3-90	lcl_21_L3f1(Tor81)	Torrioni et al. 2006	Este de África	Etiópia	L3f1a
L3-91	lcl_22_L3h(Tor31)	Torrioni et al. 2006	Este de África	Etiópia	L3h1b1
L3-92	lcl_23_L3h(Tor67)	Torrioni et al. 2006	Este de África	Etiópia	L3h2
L3-93	lcl_24_L3a(Tor73)	Torrioni et al. 2006	Este de África	Etiópia	L3a1
L3-94	lcl_9_L3x(Tor72)	Torrioni et al. 2006	Este de África	Etiópia	L3x2a
L4-1	lcl_7_L7(Tor66)	Torrioni et al. 2006	Este de África	Etiópia	L4a1
L4-2	lcl_8_L4(Tor71)	Torrioni et al. 2006	Este de África	Etiópia	L4b2a1
L4-3	DQ112845	Kivisild et al. 2006	África do Sul	sem informação	L4b2a2
L4-4	DQ112796	Kivisild et al. 2006	Oeste de África	sem informação	L4b2b

Anexo 2: Posições variáveis, por comparação com CRS, para HVRI (menos 16000) e HVRII. As substituições indicadas são transições, excepto se a mudança de base é uma transversão ou deleção/inserção, estando nesse caso indicadas a seguir à posição onde ocorrem. Inserções de uma ou duas citosinas estão indicadas por “.1” e “.2”. As substituições em *itálico* apresentavam heteroplasmia, *n* indica o número de indivíduos.

Haplótipo	n	HVRI	HVRII	Haplogrupo
1	2	CRS	152 263 315.1	H
2	1	CRS	263 309.1 315.1	H
3	1	CRS	263 309.2 315.1	H
4	1	189	263 315.1	H
5	1	129	263 315.1	HV*
6	1	067 192	263 309.1 315.1	HV1
7	1	069 126 145 261	073 152 263 295 309.1 315.1	J1
8	2	224 311	073 152 263 309.1 315.1	K
9	1	224 311	073 263 315.1	K
10	1	129 183A/C 189 215	073 195 263 309.1 315.1	N*
11	1	189 261 270 291 311	073 150 263 309.2 315.1	U5
12	1	270 296	073 075 150 263 315.1	U5
13	1	172 179 183A/C 189 219 278	073 103 263 309.1 315.1	U6a
14	2	126 186 189 222 294	073 263 315.1	T
15	1	126 292 294 296 304 311	073 152 263 315.1	T
16	1	093 126 189 362	043insA 064 152 263 309.1 315.1	R0a
17	1	126 145 362	060insT 064 200 263 315.1	R0a
18	1	126 189 362	064 263 309.2 315.1	R0a
19	1	126 362	064 263 309.1 315.1	R0a
20	2	129 148 168 172 187 188C/G 189 223 230 311 320	093 152 189 200 236 247 263 309.1 315.1	L0a1
21	1	129 209 223 230 278 294 301 354 390	073 143 146 152 189 195 263 309.1 315.1	L2a
22	1	044 189 223 229 278 291 294 311 390	073 152 182 257 263 315.1	L2a
23	1	183A/C 189 223 229 278 291 294 311 390	073 152 182 257 263 315.1	L2a
24	1	183A/C 189 223 229 278 291 294 311 390	073 152 182 262 263 315.1	L2a

25	1	189 192 223 266 278 294 355 390	073 143 146 152 195 263 315.1	L2a
26	1	189 192 223 278 294 311 390	073 152 182 195 247 263 315.1	L2a
27	1	189 223 278 294 362 390	073 152 195 263 309.1 315.1	L2a
28	2	209 223 271 278 294 301 354 390	073 143 146 152 182 195 263 315.1	L2a
29	1	209 223 278 294 301 354 390	073 143 146 152 189 195 263 315.1	L2a
30	1	092 189 223 278 294 309 390	073 143 146 152 195 263 315.1	L2a1
31	1	093 223 278 294 309 390	073 146 152 195 263 315.1	L2a1
32	1	182A/C 183A/C 189 223 278 294 309 352 390	073 143 146 152 195 263 309.1 315.1	L2a1
33	3	183A/C 189 223 278 294 309 390	073 143 146 152 195 263 309.1 315.1	L2a1
34	1	183A/C 189 223 278 294 309 390	073 143 146 152 195 263 315.1	L2a1
35	2	189 192 223 278 289 294 309 390	073 143 146 152 195 263 309.1 315.1	L2a1
36	1	189 192 223 278 292 294 309 390	073 143 146 152 195 263 315.1	L2a1
37	1	189 192 223 278 294 309 390	073 143 146 152 195 263 309.1 315.1	L2a1
38	4	189 223 278 294 309 390	073 143 146 152 195 263 315.1	L2a1
39	1	223 224 278 309 390	073 146 152 182 183 186 263 309.1 315.1	L2a1
40	1	129 183A/C 189 223 229 278 291 294 311 390	073 152 182 195 263 309.2 315.1	L2a2
41	1	114C/A 129 213 223 234 278 390	073 146 150 152 182 183 195 198 203 263 315.1	L2b
42	1	114C/A 129 213 223 278 390	073 146 150 152 182 183 195 198 204 263 315.1	L2b
43	1	114C/A 129 145 213 221 223 278 390	073 150 152 182 183 195 204 263 315.1	L2b
44	1	114C/A 213 223 278 390	073 146 152 182 183 195 198 204 263 315.1	L2b
45	1	223 274 278 311 390	073 083 093 146 150 152 182 195 198 263 315.1 325	L2c
46	1	170 223 233 278 318 390	073 093 146 150 152 182 195 198 263 315.1 325	L2c1
47	1	129 183A/C 189 278 300 354 390	073 146 150 195 263 315.1	L2d1
48	1	111C/A 145 184 223 239 278 292 355 390 399 400	073 146 150 152 182 263 315.1	L2d2
49	1	223	073 152 189 195 263 309.1 315.1	L3*
50	1	172 182A/C 183A/C 189 223	073 150 195 263 309.1 315.1	L3*
51	2	165 177 178 192 223 311 399	073 146 263 315.1	L3*
52	1	086 169 193 195	073 150 200 249delA 263 315.1 375insGAAA	L3x?

53	1	169 223 278 311	073 150 204 263 315.1	L3x
54	1	124 223 278 362	073 263 315.1	L3b
55	1	093 223 278 362	073 263 315.1	L3b
56	1	124 223	073 151 152 195 263 315.1	L3d
57	1	124 223	073 189 195 200 263 309.1 315.1	L3d
58	1	041 186 223 266	073 150 195 263 315.1 398	L3e5
59	1	041 223	073 150 263 309.1 315.1 398	L3e5
60	2	041 223	073 150 263 315.1 398	L3e5
61	1	209 213 223 234 311	073 185 189 263 309.1 315.1	L3f
62	3	209 213 223 311	073 195 263 315.1 328	L3f
63	1	209 223 234 311	073 189 263 315.1	L3f
64	1	192 209 223 292 311	073 152 185 189 200 263 315.1	L3f1
65	1	209 218 223 292 311	073 189 200 263 309.1 315.1	L3f1
66	1	209 223 292 311	073 189 195 200 263 309.1 315.1	L3f1
67	1	209 223 292 311	073 189 200 263 309.1 315.1	L3f1
68	2	209 223 292 311	073 189 200 263 315.1	L3f1
69	2	176 209 223 234 292	073 152 189 263 309.1 315.1 318	L3f2
70	1	093 172 179 192 215 223 256C/A 278 284 311	073 189A/C 195 263 315.1	L3h1b
71	1	093 172 179 215 223 256C/A 278 284 311	073 189A/C 195 263 315.1	L3h1b
72	1	189 223 311 399	073 146 153 263 315.1	L3h1
73	1	223 311 362	073 150 263 309.1 315.1	L3h1?
74	1	129 192 223 255 298 311	073 263 309.1 315.1	L3h1?
75	1	111 184 192 223 234 304	073 150 263 315.1 318	L3h2
76	1	093 129 148 166 183delA 187 189 223 278 294 311 355 362	073 152 182 195 247 263 315.1	L5b
77	1	093 129 148 166 183A/C 186 189 223 278 311 355 362	073 152 182 195 247 263 315.1	L5b
78	1	093 223 293A/T 311 355 362 399	073 146 257 263 309.1 315.1	L4b2
79	1	223 293A/T 311 355 362 399	073 146 257 263 315.1	L4b2
80	1	129 182A/C 183A/C 189 223 249 311	073 152 195 263 315.1	M1

81	1	129 183A/C 189 223 249 311	073 195 263 315.1	M1
82	1	129 183A/C 189 212 223 249 304 311 359	073 195 263 315.1	M1a
83	1	129 183A/C 189 223 249 256 311 359	073 146 189 195 263 309.2 315.1	M1a
84	1	129 183A/C 189 223 249 311 359	073 195 263 280 309.1 315.1	M1a
85	1	140 209 223	073 152 263 315.1	M7a

Anexo 3: Posições variáveis, por comparação com CRS, para a molécula inteira do mtDNA. As substituições indicadas são transições, excepto se a mudança de base ou deleção/inserção estão indicadas.

Haplótipo	mtDNA	Haplogrupo
SUD1	073 150 263 309.1 315.1 398 521delCA 711 1438 2352 2706 4769 7028 7280 8392 8701 8860 9540 10398 10819 10873 11719 12705 14212 14766 15301 15326 16041 16223 16519	L3e5
SUD2	073 151 152 195 263 315.1 521delCA 750 921 1438 2706 3360 3957 4769 5046 5147 5238 5567 6272 6680 7028 7424 8618 8701 8860 9540 10398 10873 11719 12705 13105 13886 14284 14766 15301 15326 15924 16124 16223 16519	L3d1b
SUD5	073 146 263 315.1 723 750 1438 1719 2706 3290 4388 4688 4742 4769 5108 5300 5492 7028 7861 8701 8781C/A 8860 8943 9509 9540 9575 10398 10873 11590 11719 12175 12236 12519 12705 14587 14766 14862 15301 15326 15646 16165 16177 16178 16192 16223 16311 16399	L3h1a2a
SUD8	073 150 263 315.1 318 521delCA 628 750 859 1438 1530 2706 3879 4769 5460 5813 5930 7028 7861 8020 8701 8860 9098 9380 9540 9575 9965 10325 10398 10858 10873 11440 11719 12469 12705 13080 13753 13755 13824 14766 15301 15326 15910 16111 16184 16192 16223 16234 16304 16519	L3h2
SUD13	073 263 315.1 521delCA 750 1438 2706 3450 4769 5773 6221 7028 8701 8860 9449 9540 10086 10373 10398 10873 11002 11719 12705 13105 13914C/A 13932 14766 15301 15311 15326 15824 15944delT 16093 16223 16278 16362 16519	L3b1a
SUD17	073 263 315.1 521delCA 750 1438 2706 3450 4769 5773 6221 7028 8288insACCCCTCT 8701 8860 9300 9449 9540 10086 10373 10398 10873 11002 11719 12705 13105 13419 13914C/A 14766 15301 15311 15326 15824 15944delT 16124 16223 16278 16362 16519	L3b1a
SUD18	073 150 195 263 315.1 398 521delCA 1438 2352 2706 4769 7028 8392 8701 8860 9540 10398 10819 10873 11719 12397 12705 12969C/A 14212 14766 15301 15326 16041 16186 16223 16266 16519	L3e5
SUD33	073 150 200 249delA 263 315.1 375insGAAA 494 750 1438 1842 2706 3435 3483 4769 5900A/C 6227 6392 6401 7028 7129 8311 8701 8817 8860 9180 9540 9941 10398 10819 10873 11719 12705 13681 13708 13722 14766 15301 15326 16086 16169 16193 16195	L3x2a
SUD39	263 309.1 315.1 750 1438 1719 2706 4048 4388 4769 5300 5492 5899insC 7028 7861 8047 8485 8701 8860 9509 9540 9575 10398 10873 11084 11152 11590 11719 12432 12705 12741 13708 13926 14110 14278 14766 15301 15326 16129 16192 16223 16255 16298 16311	L3h1a
SUD47	073 152 189 263 309.1 315.1 318 521delCA 750 959 1438 2706 3396 4218 4643 4769 5181 6602 7028 7258 8158 8251 8701	L3f3

	8860 9540 9932 10398 10604 10873 11719 11770 12705 14766 15301 15326 15502C/G 15514 15940 15944delT 16176 16209 16223 16234 16292	
SUD72	073 152 189 195 263 309.1 315.1 521delCA 750 921 1393 1438 2706 4769 5147 7028 7424 8618 8701 8860 9540 10398 10873 11719 12280 12705 13105 13886 14284 14766 15301 15326 16223	L3d*
SUD81	073 189 200 263 309.1 315.1 750 1438 1822 2706 3396 3483 4218 4769 7028 7819C/A 8527 8701 8860 8932 9540 9950 10398 10873 11440 11719 12705 14122 14766 14769 15301 15326 15514 15944delT 16209 16223 16292 16311 16519	L3f1b
SUD83	073 150 263 309.1 315.1 750 769 1018 1438 1914 2706 3987 4137 4769 4977 5318 5460 7028 7389 7849 8392 8701 8860 8928 9168 9540 10398 10400 10813 10873 11719 12705 12988 14766 14793 15301 15326 16223 16311 16362	L4
SUD100	073 195 263 315.1 328 736 750 1438 2706 3197 3396 3591 3693 4218 4350 4769 5194 5460 5601 7028 8701 8860 9540 9950 10398 10873 11719 12441 12507 12705 13967 14148 14554 14766 15106 15151 15301 15317 15326 15514 15944delT 16209 16213 16223 16311 16519	L3f1a
SUD82	073 150 204 263 315.1 750 1438 2706 3483 4769 5899insC 6401 7028 8311 8701 8817 8860 9540 10398 10819 10873 11719 12705 13708 14766 15172 15301 15326 15346 16169 16223 16278 16311	L3x1
SUD91	073 189A/C 195 263 315.1 522delAC 750 1438 1719 2706 3777 4388 4769 5300 7028 7861 8701 8860 9323 9509 9524 9540 9575 10044 10398 10873 11590 11719 12705 14410 14766 14950 15301 15326 16093 16172 16179 16192 16215 16223 16256C/A 16278 16284 16311	L3h1b1
SUD95	073 189A/C 195 263 315.1 437 522delAC 750 1438 1719 2706 3777 4388 4769 5300 7028 7861 8701 8860 9323 9509 9524 9540 9575 10044 10398 10873 11590 11719 12705 14410 14766 14950 15301 15326 16093 16172 16179 16215 16223 16256C/A 16278 16284 16311	L3h1b1
SUD96	073 146 153 263 315.1 750 1438 1508 1719 2706 4388 4688 4742 4769 4965 5300 5492 7028 7055 7861 8701 8781C/A 8860 8943 9509 9540 9575 9839C/A 10398 10873 11590 11719 12175 12519 12705 13158 14587 14766 15229 15301 15326 15646 16189het 16223 16311 16399	L3h1a2a
SUD58	16519 073 185 189 263 309.1 315.1 521delCA 750 1438 2706 3396 4218 4769 5821 6182 6722 7028 8676 8701 8860 9365 9540 9731C/A 10398 10873 11719 12280 12534 12705 13105 14260 14766 14862 15301 15314 15326 15479 15514 15944delT 16209 16213 16223 16234 16311	L3f2
SUD68	073 189 200 263 309.1 315.1 750 1438 1822 1891 2251 2706 4218 4769 5601 6150 7028 7819C/A 7852 8527 8701 8860 8932 9540 9950 10398 10873 11440 11719 12705 14766 14769 15301 15326 15514 15944delT 16209 16218 16223 16292 16311 16519	L3f1b
SUD76	073 152 185 189 200 263 315.1 750 1438 1822 2626 2706 3396 4218 4769 5601 7028 7819C/A 8527 8701 8860 8932 9540 9950	L3f1b

	10398 10873 11440 11719 12705 14766 14769 15301 15326 15514 15944delT 16018 16192 16209 16223 16292 16311 16519	
SUD92	16519 073 189 263 315.1 521delCA 750 1438 2706 3396 4218 4769 5821 6182 7028 8676 8701 8860 9365 9540 9731C/A 10398 10873 11719 12280 12534 12705 13105 14766 15301 15314 15326 15479 15514 15944delT 16209 16223 16234 16311	L3f2

Anexo 4 – Espectro de taxas de mutação para a região codificante em 155 sequências L3 e para a região controle em apenas 74 dessas sequências L3. As bases encontram-se ordenadas por ordem decrescente de recorrência. As posições não variáveis não estão apresentadas.

Região codificante

Posição	Recorrência						
		11800	2	2158	1	4040	1
1719	4	12236	2	2251	1	4086	1
3396	3	12248	2	2262	1	4092	1
4048	3	12280	2	2332	1	4117	1
4688	3	12441	2	2352	1	4122	1
8251	3	13651	2	2357	1	4164	1
9254	3	13752	2	2416	1	4167	1
12397	3	13967	2	2483	1	4200	1
13105	3	14034	2	2626	1	4203	1
13708	3	14410	2	2702	1	4350	1
15110	3	14484	2	2707	1	4388	1
15944	3	14587	2	2831	1	4392	1
709	2	14869	2	2880	1	4393	1
711	2	15106	2	2903	1	4395	1
921	2	15311	2	3010	1	4491	1
1193	2	15314	2	3197	1	4500	1
1842	2	15758	2	3200	1	4560	1
3394	2	15924	2	3203	1	4619	1
4218	2	628	1	3204	1	4643	1
4655	2	678	1	3290	1	4679	1
4907	2	721	1	3360	1	4715	1
5147	2	723	1	3420	1	4733	1
5262	2	736	1	3435	1	4742	1
5300	2	750	1	3438	1	4823	1
5460	2	792	1	3441	1	4859	1
5471	2	859	1	3450	1	4902	1
5567	2	959	1	3459	1	4910	1
5601	2	1393	1	3474	1	4965	1
6150	2	1438	1	3483	1	5004	1
6221	2	1503	1	3492	1	5046	1
6272	2	1508	1	3498	1	5054	1
6480	2	1530	1	3504	1	5063	1
7055	2	1692	1	3505	1	5102	1
7861	2	1694	1	3582	1	5108	1
8393	2	1737	1	3591	1	5162	1
8485	2	1799	1	3621	1	5181	1
8527	2	1822	1	3666	1	5186	1
9932	2	1888	1	3693	1	5194	1
10370	2	1891	1	3729	1	5201	1
10373	2	1900	1	3777	1	5211	1
10816	2	1927	1	3850	1	5238	1
11176	2	1977	1	3852	1	5246	1
11377	2	2000	1	3879	1	5250	1
11401	2	2093	1	3915	1	5310	1
11440	2	2145	1	3957	1	5316	1

5371	1	7858	1	9380	1	11260	1
5372	1	8020	1	9449	1	11347	1
5393	1	8041	1	9452	1	11350	1
5441	1	8047	1	9477	1	11383	1
5492	1	8155	1	9490	1	11404	1
5580	1	8158	1	9509	1	11503	1
5581	1	8167	1	9524	1	11518	1
5584	1	8197	1	9533	1	11563	1
5605	1	8222	1	9554	1	11590	1
5773	1	8311	1	9575	1	11722	1
5774	1	8387	1	9682	1	11770	1
5777	1	8392	1	9728	1	11887	1
5813	1	8400	1	9731	1	11914	1
5821	1	8410	1	9827	1	11944	1
5930	1	8530	1	9839	1	11963	1
6092	1	8598	1	9941	1	12070	1
6182	1	8616	1	9950	1	12175	1
6227	1	8618	1	9965	1	12188	1
6261	1	8650	1	10034	1	12340	1
6267	1	8676	1	10042	1	12373	1
6293	1	8688	1	10044	1	12396	1
6366	1	8709	1	10070	1	12406	1
6392	1	8749	1	10084	1	12414	1
6401	1	8767	1	10086	1	12432	1
6431	1	8781	1	10181	1	12469	1
6524	1	8799	1	10184	1	12507	1
6587	1	8817	1	10197	1	12519	1
6602	1	8823	1	10286	1	12534	1
6680	1	8835	1	10289	1	12561	1
6722	1	8853	1	10314	1	12618	1
6806	1	8875	1	10325	1	12630	1
6842	1	8928	1	10424	1	12634	1
6932	1	8932	1	10463	1	12705	1
7058	1	8943	1	10604	1	12732	1
7129	1	8965	1	10640	1	12741	1
7158	1	9022	1	10667	1	12816	1
7196	1	9053	1	10679	1	12850	1
7258	1	9067	1	10819	1	12870	1
7270	1	9079	1	10837	1	12873	1
7280	1	9080	1	10858	1	12969	1
7301	1	9098	1	10899	1	12999	1
7389	1	9111	1	10915	1	13080	1
7394	1	9139	1	11002	1	13101	1
7424	1	9151	1	11017	1	13116	1
7493	1	9180	1	11024	1	13158	1
7618	1	9300	1	11029	1	13167	1
7645	1	9305	1	11084	1	13197	1
7648	1	9323	1	11149	1	13269	1
7765	1	9337	1	11152	1	13326	1
7819	1	9365	1	11239	1	13368	1
7852	1	9377	1	11257	1	13419	1

13434	1	14769	1	Região controle		16179	1
13443	1	14818	1	Posição	Recorrência	16184	1
13542	1	14851	1	16311	17	16185	1
13611	1	14862	1	195	14	16186	1
13681	1	14905	1	16519	11	16188	1
13687	1	14950	1	152	11	16193	1
13722	1	15061	1	189	8	16195	1
13749	1	15099	1	200	8	16223	1
13753	1	15113	1	16093	7	16230	1
13755	1	15115	1	16189	7	16243	1
13800	1	15151	1	150	5	16254	1
13824	1	15172	1	185	5	16255	1
13884	1	15208	1	16086	4	16261	1
13886	1	15217	1	16124	4	16264	1
13914	1	15229	1	16192	4	16265	1
13926	1	15299	1	16234	4	16284	1
13932	1	15317	1	16278	4	16290	1
13954	1	15346	1	16129	3	16291	1
14040	1	15358	1	16172	3	16294	1
14060	1	15367	1	16256	3	16298	1
14110	1	15388	1	16260	3	16316	1
14122	1	15394	1	16292	3	16319	1
14128	1	15434	1	16295	3	16320	1
14148	1	15466	1	16362	3	16327	1
14152	1	15479	1	16111	2	16355	1
14161	1	15496	1	16153	2	16368	1
14173	1	15502	1	16178	2	143	1
14180	1	15514	1	16209	2	153	1
14207	1	15519	1	16213	2	198	1
14212	1	15550	1	16215	2	199	1
14221	1	15553	1	16218	2	203	1
14260	1	15565	1	16266	2	204	1
14272	1	15646	1	16304	2	207	1
14278	1	15664	1	16399	2	235	1
14284	1	15670	1	16527	2	249	1
14311	1	15799	1	146	2	291	1
14370	1	15812	1	151	2	328	1
14371	1	15824	1	318	2	398	1
14395	1	15833	1	16037	1	408	1
14461	1	15865	1	16041	1	437	1
14554	1	15883	1	16042	1	456	1
14560	1	15910	1	16051	1	458	1
14566	1	15928	1	16126	1	465	1
14569	1	15932	1	16145	1	494	1
14571	1	15940	1	16146	1	504	1
14580	1	15942	1	16147	1	513	1
14584	1	15952	1	16165	1	516	1
14590	1	16018	1	16166	1		
14634	1			16169	1		
14690	1			16176	1		
14750	1			16177	1		