

SORAIA FIGUEIRINHAS NUNES GUERREIRO

CARACTERIZAÇÃO DOS *LOC*/GLOBALFILER NAS POPULAÇÕES ANGOLANA E GUINEENSE, RESIDENTES NO SUL DE PORTUGAL

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

Orientador – Mestre Paulo Dario

Categoria – Especialista Superior de Medicina Legal

Afiliação – Serviço de Genética e Biologia Forenses da Delegação do Sul do Instituto Nacional de Medicina Legal e Ciências Forenses, I.P.



Ao terminar esta etapa da minha vida não posso deixar de agradecer a todas as pessoas que estiveram a meu lado e me ajudaram na sua concretização, quer pessoal como profissionalmente.

À Doutora Maria Cristina de Mendonça e à Dra. Maria João Porto por permitirem a realização do presente projeto nas instalações do Serviço de Genética e Biologia Forenses da Delegação do Sul do Instituto Nacional de Medicina Legal e Ciências Forenses, I.P.

À Dra. Teresa Ribeiro por me ter recebido e integrado no Serviço de Genética e Biologia Forenses da Delegação do Sul do Instituto Nacional de Medicina Legal e Ciências Forenses, I.P.

Ao meu orientador, Mestre Paulo Dario, por todo o apoio prestado ao longo deste ano, pelos conhecimentos transmitidos, por sempre me incentivar a dar o meu melhor na concretização deste projeto, pelas palavras de apoio nas “horas de desespero” e, por ter tido sempre tempo para me ajudar a resolver os imprevistos e as dúvidas que foram ocorrendo durante toda a realização deste trabalho.

À Professora Doutora Maria José Pinto da Costa, Diretora do Mestrado em Medicina Legal, pela oportunidade de ingressar no mesmo, por todos os conhecimentos transmitidos e apoio prestado e por todas as dúvidas esclarecidas.

À Mestre Heloísa Afonso Costa, à Mestre Cláudia Vieira da Silva, ao Mestre António Amorim e à Dra. Isabel Lucas por todo o apoio, boa disposição, motivação e conhecimentos transmitidos.

A todas as minhas colegas estagiárias no Serviço, nomeadamente à Ana Rita Inácio, à Ana Resende, à Marisa Cruz e à Rita Dario pela orientação prestada no início do estágio e durante o trabalho laboratorial e por todos os momentos de animação partilhados. Em especial, à Marta Proença, por me ter ouvido, aconselhado e ajudado nalguns dias mais complicados, pela sua amizade, que quero preservar, e pelo companheirismo demonstrado.

À Dona Conceição, à Dona Céu e ao Hugo pelos momentos de descontração e risada proporcionados.

A todos os meus amigos, principalmente à Liliana Brito, à Andreia Fernandes, à Catarina Oliveira, à Márcia Vieira e à Catarina Alves que tinham sempre uma palavra de apoio e motivação para me dar e muita paciência para me ouvir e aturar. E, particularmente, à Joana Couceiro e à Carolina Pereira pelo tempo despendido na leitura da dissertação, pelas dicas de escrita e por toda a ajuda prestada.



À Dona Helena e ao Senhor Firmino, pelo carinho, por me escutarem e aconselharem e por me tirarem de casa nos dias em que não conseguia mais olhar para o computador.

Às minhas colegas da Especialização e do Mestrado, Inês Bessa, Mafalda Ferreira, Tânia Moreira, Bárbara Gonçalves e Mariana Plácito por terem sido o meu apoio durante as aulas e durante a minha estadia no Porto, espero que a nossa amizade perdure.

Ao meu namorado, Pedro Mourato, por toda a paciência, compreensão, encorajamento e suporte e por estar sempre presente.

Ao meu pai e ao meu irmão, Armando Guerreiro e Gabriel Guerreiro, por me permitirem seguir os meus sonhos, por estarem sempre do meu lado e por me amarem incondicionalmente assim como eu os amo. À minha tia, Carla Guerreiro, por ser um exemplo de força e determinação e por todas as palavras de motivação e carinho. À minha “madrinha”, Cármen Vaz, por tudo o que fez e continua a fazer por mim e também às minhas avós, Maria Eugénia Guerreiro e Maria Helena Nunes, por me encherem de miminhos e palavras doces. A toda a minha família pela imensa demonstração de orgulho em mim.

À minha mãe, que apesar de não poder partilhar comigo esta “vitória”, sei que, onde quer que esteja, está muito orgulhosa.

A todos, o meu sincero OBRIGADO.



As Repúblicas de Angola e da Guiné-Bissau são países de língua oficial portuguesa, ou PALOP, localizados na costa ocidental africana. Estes países foram alvo de muitas invasões, movimentos migratórios e rotas comerciais, principalmente por parte dos navegadores portugueses, o que contribuiu não só para a diversificação cultural e religiosa das suas populações mas também, para o aumento da variabilidade genética das mesmas.

Após a independência destas colónias do domínio português, alguns dos seus habitantes imigraram para Portugal, para fugirem a situações de mudança e instabilidade política e de guerra estando, atualmente, estas nacionalidades estrangeiras entre as dez mais representadas no nosso país.

Segundo o Relatório de Imigração, Fronteiras e Asilo do Serviço de Estrangeiros e Fronteiras, referente ao ano 2015, residiam em Portugal 388 731 imigrantes, dos quais 18 247 e 17 091 correspondiam a indivíduos angolanos e guineenses, respetivamente, maioritariamente residentes no sul do país.

Devido aos fatos referidos anteriormente, este estudo teve como objetivo a caracterização das populações angolana e guineense, residentes no sul de Portugal, com recurso à análise de 21 STRs autossómicos do *GlobalFiler™ PCR Amplification Kit (Applied Biosystems)*, e a comparação destas populações com outras já publicadas na literatura, nomeadamente a população do Sul de Portugal.

Para tal, analisaram-se 153 e 70 amostras de sangue de indivíduos de origem angolana e guineense, respetivamente, que se apresentaram no Serviço de Genética e Biologia Forenses do INML, I.P. no âmbito de investigações de parentesco biológico.

O estudo populacional foi realizado com recurso ao *software* Arlequin versão 3.5, com o qual foram calculadas as frequências alélicas, o equilíbrio de Hardy-Weinberg e o *linkage disequilibrium*. Os resultados obtidos indicaram que todos os *loci* se encontram em HWE e em LD, para ambas as populações estudadas. A comparação entre as populações foi feita através do cálculo das distâncias genéticas e do teste exato de diferenciação populacional, usando os *softwares* PHYLIP versão 3.695 e Arlequin versão 3.5, respetivamente, os quais demonstraram que as populações de Angola e Guiné-Bissau residentes no sul de Portugal são muito semelhantes à afro-americana e à sul-africana. Contudo, estas populações são geneticamente menos relacionadas com as populações europeias, asiáticas e mexicana. O teste exato de diferenciação populacional apresentou diferenças significativas entre os pares de populações África do Sul/Angola e África do Sul/Guiné-Bissau, para dois *loci* (D22S1045 e D10S1248) e para um *locus*

A heterozigosidade, a homozigosidade e os parâmetros estatísticos de interesse forense foram estimados utilizando o *software* PowerStats versão 1.2. Todos os marcadores genéticos estudados preenchem os requisitos necessários para que possam ser usados em genética forense, porque apesar do D13S317 apresentar valores de heterozigosidade inferiores a 70% para as populações angolana e guineense, todos os STRs analisados exibem resultados superiores a 80% para o poder de discriminação.

O marcador SE33 mostrou-se o mais informativo para a população angolana. No entanto, para a população guineense os *loci* mais informativos são o SE33 e o D21S11. Os resultados obtidos para a população de Angola e da Guiné-Bissau demonstraram que para a primeira existem dois *loci* menos informativos, o D5S818 e o D13S317 e para a segunda existe apenas um, o D13S317.

[illegible]

Os dados referidos permitem concluir que as frequências alélicas determinadas no presente estudo podem constituir uma base de dados de referência, aquando da necessidade do uso de frequências populacionais para indivíduos de origem angolana ou guineense em casos de investigação de parentesco biológico realizados no Serviço de Genética e Biologia Forenses da Delegação Sul do INMLCF, I.P. Este estudo indica, também, que as populações estudadas irão introduzir variabilidade genética na população do Sul de Portugal.

Palavras-Chave: Angola, Guiné-Bissau, Sul de Portugal, STRs autossómicos, *GlobalFiler™ PCR Amplification Kit*, Genética Populacional, Genética Forense.



Angola and Guinea-Bissau are Portuguese-speaking countries (PALOP), located on the west coast of Africa. These countries were subject to several invasions, migratory movements and trade networks, especially by the Portuguese navigators, which contributed not only to the cultural and religious diversity of their populations but also to increase their genetic variation.

After independence of these Portuguese colonies, some of its inhabitants migrated to Portugal, to escape to political instability and war. Actually, these populations are among the ten most represented foreign nationalities in our country.

According to 2015 Immigration, Frontiers and Asylum Report, from Portuguese Foreign Affair Services, the total number of foreign residents was 388 731, which 18 247 and 17 091 corresponded to Angolan and Guinean individuals, respectively, mainly located at the southern region.

Due to facts mentioned above, this study aimed to characterize the Angolan and Guinean populations inhabiting in southern Portugal, through analysis of 21 autosomal STRs of GlobalFiler™ PCR Amplification Kit (Applied Biosystems), and the comparison of these populations with others already published, including South Portugal population.

In these work, were analyzed 153 and 70 blood samples of Angolan and Guinea individuals, respectively, who went to Forensic Genetics and Biology Department of INMLCF, I.P. under biological kinship investigations.

Population study was performed using Arlequin software, version 3.5, which estimated allelic frequencies, Hardy-Weinberg equilibrium and linkage disequilibrium. The results showed that all loci were in HWE and LD for both studied populations. Comparison between populations was made through the calculation of genetic distances and the exact test of population differentiation, using PHYLIP, version 3.695, and Arlequin, version 3.5 softwares, respectively. Their results demonstrated that Angola and Guinea-Bissau populations inhabiting southern Portugal, are very similar to African-American and South-African. However, these populations are genetically less related to the European, Asian and Mexican populations. The exact test of population differentiation exhibited significant differences between South Africa/Angola and South Africa/Guinea-Bissau, for two loci (D22S1045 and D10S1248) and to one locus (D22S1045), respectively. Remaining populations showed significant differences for multiple loci.

Heterozygosity, homozygosity and statistical parameters with forensic interest were inferred using PowerStats software, version 1.2. All genetic markers

studied fulfill the requirements to be used in forensic genetics, because despite the D13S317 present values below 70% for Angolan and Guinean populations, all STRs analyzed showed power of discrimination greater than 80%.

SE33 marker proved to be the most informative for the Angolan population. However, for the Guinean population the most informative loci are the SE33 and D21S11. Results for Angola and Guinea-Bissau populations shown that for the first one there are two less informative loci, D5S818 and D13S317 and for the second one, only one, D13S317.

For all 21 autosomal STRs analyzed, probability of identity, power of discrimination and power of exclusion exhibited values equal to $3,8093 \times 10^{-26}$, 99.999999999999999999999961907% and 99.9999999837329%, respectively, for the Angolan immigrant population and $2,085 \times 10^{-25}$, 99.99999999999999999999997915% and 99.9999997871199%, respectively, for the Guinean immigrant population.

Data referred allow conclude that allelic frequencies calculated in this study can be part of a reference database, when it is necessary to use population frequencies for individuals with Angolan or Guinean origin attempted to Forensic Genetics and Biology Department of INMLCF, I.P. under biological kinship investigations. This work also suggests that populations studied will introduce genetic variability in native South Portugal population.

Keywords: Angola, Guinea-Bissau, South Portugal, autosomal STRs, *GlobalFiler™ PCR Amplification Kit*, Population Genetics, Forensic Genetics.

AGRADECIMENTOS	III
RESUMO	V
ABSTRACT	VII
ÍNDICE DE FIGURAS	XI
ÍNDICE DE TABELAS	XIII
LISTA DE ABREVIATURAS E ACRÓNIMOS	XV
1. INTRODUÇÃO	1
1.1 REPÚBLICA DE ANGOLA	1
1.1.1 GEOGRAFIA E DEMOGRAFIA	1
1.1.2 HISTÓRIA	2
1.2 REPÚBLICA DA GUINÉ-BISSAU	5
1.2.1 GEOGRAFIA E DEMOGRAFIA	5
1.2.2 HISTÓRIA	6
1.3 A MIGRAÇÃO EM PORTUGAL	9
1.4 GENOMA HUMANO	11
1.4.1 ESTRUTURA DO DNA	12
1.4.2 VARIABILIDADE GENÉTICA	13
1.5 GENÉTICA FORENSE	15
1.5.1 MICROSSATÉLITES OU SHORT TANDEM REPEATS (STRs)	17
1.5.1.1 METODOLOGIAS PARA A ANÁLISE DOS STRs	19
1.5.2 BASES DE DADOS DE PERFIS DE DNA	24
1.5.3 GLOBALFILER™ PCR AMPLIFICATION KIT	26
1.6 GENÉTICA POPULACIONAL	29
1.6.1 ESTUDO DAS POPULAÇÕES	29
1.6.1.1 PRINCÍPIO DE HARDY-WEINBERG	30
1.6.1.2 <i>LINKAGE (DIS)EQUILIBRIUM</i>	30
1.6.1.3 DISTÂNCIA GENÉTICA ENTRE POPULAÇÕES	31
1.6.1.4 TESTE EXATO DE DIFERENCIAÇÃO	32
1.6.1.5 HOMOZIGOSIDADE E HETEROZIGOSIDADE	32
1.6.1.6 PARÂMETROS ESTATÍSTICOS DE INTERESSE FORENSE	32
1.7 OBJETIVOS DO ESTUDO	35
2. MATERIAIS E MÉTODOS	37
2.1 AMOSTRAGEM	37
2.2 COLHEITA E ARMAZENAMENTO DAS AMOSTRAS	38



2.3 EXTRAÇÃO DE DNA PELO MÉTODO DE <i>CHELEX</i> ®100.....	38
2.4 AMPLIFICAÇÃO DOS MARCADORES GENÉTICOS.....	38
2.5 ELETROFORESE CAPILAR	40
2.6 ESTUDO DAS POPULAÇÕES	40
2.6.1 ANÁLISE ESTATÍSTICA DOS PARÂMETROS POPULACIONAIS.....	40
2.6.2 HETEROZIGOSIDADE, HOMOZIGOSIDADE E ANÁLISE ESTATÍSTICA DOS PARÂMETROS DE INTERESSE FORENSE.....	41
3. RESULTADOS E DISCUSSÃO	43
3.1 ESTUDO DA POPULAÇÃO ANGOLANA RESIDENTE NO SUL DE PORTUGAL ...	43
3.1.1 ANÁLISE ESTATÍSTICA DOS PARÂMETROS POPULACIONAIS.....	43
3.1.2 HETEROZIGOSIDADE, HOMOZIGOSIDADE E ANÁLISE ESTATÍSTICA DOS PARÂMETROS DE INTERESSE FORENSE.....	48
3.2 ESTUDO DA POPULAÇÃO GUINEENSE RESIDENTE NO SUL DE PORTUGAL.....	53
3.2.1 ANÁLISE ESTATÍSTICA DOS PARÂMETROS POPULACIONAIS.....	53
3.2.2 HETEROZIGOSIDADE, HOMOZIGOSIDADE E ANÁLISE ESTATÍSTICA DOS PARÂMETROS DE INTERESSE FORENSE	57
3.3 DISTÂNCIA GENÉTICA ENTRE AS POPULAÇÕES	61
3.4 TESTE EXATO DE DIFERENCIAÇÃO POPULACIONAL.....	65
4. CONCLUSÃO.....	69
5. REFERÊNCIAS BIBLIOGRÁFICAS.....	73
6. ANEXOS.....	84



Figura 1: Mapa da República de Angola. (Fonte: http://www.africa-turismo.com/mapas/angola.htm consultado a 03/07/2016)	1
Figura 2: Mapa da República da Guiné-Bissau. (Fonte: http://www.africa-turismo.com/mapas/guine-bissau.htm consultado a 03/07/2016)	5
Figura 3: Principais nacionalidades de imigrantes em Portugal. (Adaptado de Dias <i>et al.</i> , 2015).....	10
Figura 4: Organização do genoma humano dentro da célula. (Adaptado de Butler, 2010a)	11
Figura 5: Constituição do DNA e dos genes. (Adaptado de: Liu, G. <i>et al.</i> , 2014)	12
Figura 6: Estrutura do DNA. (Adaptado de Krebs <i>et al.</i> , 2014).....	13
Figura 7: Sequências repetidas em <i>tandem</i>. (Adaptado de Butler, 2010c)	16
Figura 8: Classificação dos alelos. (Adaptado de Butler, 2010d).....	18
Figura 9: Estrutura da microvariante 9.3 do marcador TH01. (Adaptado de Goodwin <i>et al.</i> , 2011c)	18
Figura 10: Tipos de padrões trialélicos. (Adaptado de Butler, 2010e)	19
Figura 11: Extração de DNA com Chelex® 100. (Adaptado de Goodwin <i>et al.</i> , 2011)	21
Figura 12: Amplificação do DNA por PCR. (Adaptado de Butler, 2010g)	22
Figura 13: Representação esquemática da técnica de eletroforese capilar. (Adaptado de Butler, 2010; Li, 2015e)	24
Figura 14: Mapa de Portugal. (Adaptado de Carvalho, 2015)	37
Figura 15: Gráfico MDS.	63
Figura 16: Árvore filogenética.	64



Tabela I: <i>Loci</i> , alelos e fluoróforos do <i>GlobalFiler™ PCR Amplification Kit</i> . (Adaptado de Applied Biosystems, 2015).....	27
Tabela II: Parâmetros estatísticos de interesse forense. (Adaptado de Butler, 2012 e Goodwin <i>et al.</i> , 2011h)	32
Tabela III: Reagentes e respetivos volumes utilizados para a amplificação dos <i>loci</i> STR de cada uma das amostras.	39
Tabela IV: Composição dos controlos positivo e negativo.	39
Tabela V: Condições da amplificação por PCR do <i>GlobalFiler™ PCR Amplification Kit</i> (Applied Biosystems).	40
Tabela VI: Frequências alélicas dos 21 STRs autossómicos na população angolana residente no sul de Portugal.....	44
Tabela VII: Resultados obtidos para os parâmetros estatísticos populacionais relativos à população angolana residente no sul de Portugal.	46
Tabela VIII: Tabela representativa dos valores de significância para o <i>linkage disequilibrium</i>	47
Tabela IX: Parâmetros estatísticos de interesse forense, número de alelos e genótipos, heterozigossidade e homozigossidade para cada um dos 21 STRs autossómicos estudados para a população angolana residente no sul de Portugal	49
Tabela X: Parâmetros estatísticos de interesse forense para todos os STRs autossómicos estudados para a população angolana residente no sul de Portugal	50
Tabela XI: Frequências alélicas dos 21 STRs autossómicos na população guineense residente no sul de Portugal.....	54
Tabela XII: Resultados obtidos para os parâmetros estatísticos populacionais relativos à população guineense residente no sul de Portugal.....	56
Tabela XIII: Tabela representativa dos valores de significância para o <i>linkage disequilibrium</i>	57
Tabela XIV: Parâmetros estatísticos de interesse forense, número de alelos e genótipos, heterozigossidade e homozigossidade para cada um dos 21 STRs autossómicos estudados para a população guineense residente no sul de Portugal	59
Tabela XV: Parâmetros estatísticos de interesse forense para todos os STRs autossómicos estudados para a população guineense residente no sul de Portugal	59



Tabela XVI: Distâncias genéticas entre as populações angolana e guineense residentes no sul de Portugal e as populações do Sul de Portugal, de Espanha, da Holanda, dos EUA (indivíduos caucasianos e africanos), da África do Sul, do Líbano, dos Emirados Árabes Unidos, do Afeganistão, da China, do México e da Coreia	61
Tabela XVII: Valores de ρ para o teste exato de diferenciação populacional entre a população angolana residente no sul de Portugal e as populações guineense residente no sul de Portugal, do Sul de Portugal, de Espanha, da Coreia, do Líbano, dos EUA (indivíduos caucasianos e africanos), da Holanda, da África do Sul, da China, do México, dos Emirados Árabes Unidos e do Afeganistão	66
Tabela XVIII: Valores de ρ para o teste exato de diferenciação populacional entre a população guineense residente no sul de Portugal e as populações angolana residente no sul de Portugal, do Sul de Portugal, de Espanha, da Coreia, do Líbano, dos EUA (indivíduos caucasianos e africanos), da Holanda, da África do Sul, da China, do México, dos Emirados Árabes Unidos e do Afeganistão	67

LISTA DE ABREVIATURAS E ACRÓNIMOS



%	Porcentagem
°C	Grau Celsius
μL	Microlitro
∞	Infinito
n	Multiplicação
A	Adenina
C	Citosina
CODIS	<i>Combined DNA Index System</i>
d.C.	Depois de Cristo
DNA	<i>Deoxyribonucleic Acid</i>
EC	Eletroforese Capilar
EDNAP	<i>European DNA Profiling Group</i>
ENFSI	<i>European Network of Forensic Institutes</i>
ESS	<i>European Standard Set</i>
EUA	Estados Unidos da América
FBI	<i>Federal Bureau of Investigation</i>
Fst	Índice de fixação
G	Guanina
g	Grama
h	Homozigozidade
H	Heterozigozidade
He	Heterozigotia esperada
Ho	Heterozigotia observada
HWE	Equilíbrio de Hardy-Weinberg
INMLCF	Instituto Nacional de Medicina Legal e Ciências Forenses, I.P.
km²	Quilometro quadrado
I	<i>Locus I</i>
L	Número de <i>loci</i>
LD	<i>Linkage Disequilibrium</i>
LE	<i>Linkage Equilibrium</i>
m	Número de comparações efetuadas
MDS	Escala multidimensional
MEGA	<i>Molecular Evolutionary Genetics Analysis</i>
mL	Mililitro
mm	Milímetro

LISTA DE ABREVIATURAS E ACRÓNIMOS



<i>n</i>	Número de amostras analisadas
NDNAD	<i>National DNA Database</i>
NJ	<i>Neighbor joining</i>
PAIGC	Partido Africano para a Independência da Guiné e Cabo Verde
PALOP	País Africano de Língua Oficial Portuguesa
PCR	<i>Polimerase Chain Reaction</i>
PD	Poder de Discriminação
PD_{Comb}	Poder de discriminação combinado
PE	Poder de Exclusão
PE_{Comb}	Poder de exclusão combinado
pH	Potencial hidrogeniónico
PI	Índice de paternidade
PIC	Conteúdo de informação polimórfica
PI_{Comb}	Índice paternal combinado
PIDE	Polícia Internacional e de Defesa do Estado
PM	Probabilidade de Identidade ou <i>Matching</i>
PM_{Comb}	Probabilidade de identidade ou <i>matching</i> combinada
RFLP	<i>Restriction Fragment Length Polymorphism</i>
rpm	Rotação por minuto
SGBF-S	Serviço de Genética e Biologia Forenses da Delegação do Sul
SPSS	<i>Statistical Package for the Social Sciences</i>
STRs	<i>Short Tandem Repeats</i>
T	Timina
VNTRs	<i>Variable Number of Tandem Repeats</i>

1.1 REPÚBLICA DE ANGOLA

1.1.1 GEOGRAFIA E DEMOGRAFIA

A República de Angola é um país africano de língua oficial portuguesa (PALOP) localizado na costa ocidental de África. O seu território ocupa uma área de 1 246 700 km² e é delimitado a norte e a nordeste pela República Democrática do Congo, a leste pela Zâmbia, a sul pela Namíbia e a oeste pelo Oceano Atlântico (Embaixada da República de Angola na Hungria). Está dividida em 18 províncias: Bengo, Benguela, Bié, Cabinda, Kuano-Kubango, Kwanza-Norte, Kwanza-Sul, Cunene, Huambo, Huíla, Luanda, Luanda-Norte, Luanda-Sul, Malanje, Moxico, Namibe, Uíge e Zaire, sendo Luanda a sua capital (Figura 1) (Consulado Geral de Angola no Porto).



Figura 1: Mapa da República de Angola. (Fonte: <http://www.africa-turismo.com/mapas/angola.htm> consultado a 03/07/2016)



Segundo os Resultados Definitivos do Censo de 2014, a população angolana é composta por 25 789 024 habitantes, dos quais 48,5% são do sexo masculino e 51,5% são do sexo feminino. A província de Luanda é a mais povoada com 6 945 386 residentes, representando 27% da população total do país (Ceita *et al.*, 2014).

Angola tem duas estações, a estação das chuvas, que corresponde ao período mais quente e ocorre de outubro a abril, e a estação seca ou do cacimbo, que começa em maio e termina em setembro. A sua temperatura média máxima e mínima corresponde a 27°C e 17°C, respetivamente (Embaixada da República de Angola na Áustria, Croácia e Eslovénia).

Por possuir uma situação geográfica particular, localizar-se numa zona intertropical e subtropical, estar próximo do mar e pelas especificidades do seu relevo, o país divide-se em duas regiões climáticas distintas: a região litoral e a região do interior. A região litoral é relativamente húmida com uma temperatura média acima dos 23°C; a região do interior, por sua vez, divide-se em três zonas: a zona norte, com elevada pluviometria e temperaturas altas, a zona de altitude, que abrange as regiões planálticas centrais, com pouca humidade e temperaturas baixas e a zona sudeste, próxima do deserto do Namibe, com baixa humidade, pouco volume pluviométrico e temperaturas baixas durante todo o ano. Devido a todas estas características Angola é um país riquíssimo em fauna e flora (Embaixada da República de Angola na Áustria, Croácia e Eslovénia).

1.1.2 HISTÓRIA

Os primeiros povos a instalarem-se em Angola, foram os bosquímanos, ótimos caçadores, de baixa estatura e de pele castanho-clara. No entanto, durante o primeiro milénio d.C. povos mais evoluídos, os Bantu, de cor negra, começaram uma das maiores migrações da História. Vindos da região dos Camarões, ao chegarem a Angola, encontraram, não só os bosquímanos, mas também outros povos menos desenvolvidos, impondo-lhes facilmente a sua tecnologia nos domínios da metalúrgica, cerâmica e agricultura (Embaixada da República de Angola em Washington, DC). Estes povos formaram uma série de reinos historicamente importantes. O primeiro surgiu no século XIII e foi designado de Reino do Congo (Warner, 1991).

Em 1482, chegam ao Congo os portugueses, liderados por Diogo Cão. O contacto entre o navegador e o rei do Congo, *Nzinga Nkuwu*, foi cordial, estabelecendo-se relações diplomáticas e trocas comerciais. *Nzinga Nkuwu*



ofereceu escravos, recursos minerais e marfim, em troca de armas de fogo. Os portugueses conseguiram também, influenciar o povo e o seu rei, levando-os a converterem-se ao cristianismo, a aprenderem a língua portuguesa e a adotarem costumes europeus (Brito, 2015; Diogo, 2014; Filomena *et al.*, 2008).

A segunda expedição, realizada por Paulo Dias de Novais, em 1560, pretendia expandir as relações com os outros reinos, inclusive o reino do *Ndongo*, situado a sul do Reino do Congo. Todavia, o rei *Ngola Kilwanji* não se mostrou interessado nas propostas portuguesas e Novais é preso e feito escravo durante seis anos, ao fim dos quais é libertado em troca de auxílio contra um reino vizinho. Em 1576, volta a Angola e funda a cidade de Luanda, dando início ao período colonial. Apesar de toda a resistência oferecida pelos reis do *Ndongo*, as expedições militares prosseguiram, levando ao desmoronamento do reino. Portugal assume então o reino do *Ndongo* em 1664 e passa a denominá-lo de Reino de Angola (Diogo, 2014; Filomena *et al.*, 2008).

Após esta conquista, os portugueses intensificaram o comércio de escravos e o processo de colonização. Angola foi um dos maiores exportadores de escravos dos séculos XVIII e XIX sendo que, a maioria destes indivíduos, teve como destino o Brasil, sendo-lhes atribuída a responsabilidade pelo crescimento da economia deste país, devido à entrada de mão-de-obra barata (Diogo, 2014).

A crescente repressão ao comércio de escravos fez com que, em 1830, a economia angolana desse um grande salto em direção ao desenvolvimento de uma economia de produção. A abolição do tráfico de escravos, iniciada em 1836 e concluída, definitivamente, em 1870, somada à abertura dos portos angolanos às nações estrangeiras, em 1844, fortaleceu o desenvolvimento dos produtos locais (Filomena *et al.*, 2008).

Com a Conferência de Berlim (1884-1885), Portugal efetivou a ocupação territorial das suas colónias. Após uma implantação morosa e complicada, no final do século XIX é estabelecida uma administração colonial baseada diretamente no território e nos povos a governar (Embaixada da República de Angola em Washington, DC).

A proclamação da República Portuguesa em 1910, seguida do aparecimento do Estado Novo em 1926, levou a uma modernização da colónia portuguesa e consequente evolução da economia (Brito, 2015). A situação vigente era aparentemente calma, até à segunda metade do século XX, altura em que movimentos associativos de grupos nacionalistas enfrentaram o poder instituído. Iniciou-se a formação de organizações políticas, mais explícitas a partir da década



de cinquenta, que reivindicavam os seus direitos em Angola. Tais organizações promoveram campanhas diplomáticas de apoio à luta pela independência do país, desencadeando mais tarde conflitos armados diretos contra o poder colonial, uma vez que os seus representantes não cediam às propostas das forças nacionalistas. Depois de muitos anos de conflito, o país ganhou a sua independência a 11 de novembro de 1975 (Embaixada da República de Angola em Washington, DC; Zau, 2002).

1.2 REPÚBLICA DA GUINÉ-BISSAU

1.2.1 GEOGRAFIA E DEMOGRAFIA

Tal como a República de Angola, a República da Guiné-Bissau também faz parte dos PALOP e fica situada na costa ocidental de África. É limitada a norte pela República do Senegal, a leste e sul pela República da Guiné e a oeste pelo Oceano Atlântico (Portal Oficial do Governo da Guiné-Bissau).

Este país é constituído por uma parte continental e outra insular, o arquipélago dos Bijagós, com cerca de noventa ilhas, das quais apenas dezassete estão habitadas. Ocupa uma extensão de aproximadamente 36 125 km², sendo que apenas 27 700 km² constituem a superfície emersa devido à diminuta elevação do país (Benzinho *et al.*, 2015a).



Figura 2: Mapa da República da Guiné-Bissau. (Fonte: <http://www.africa-turismo.com/mapas/guine-bissau.htm> consultado a 03/07/2016)



Administrativamente, o país está dividido em oito regiões: Bafatá, Biombo, Bolama/Bijagós, Cacheu, Gabú, Oio, Quínara, Tombali, e um sector autónomo, Bissau, a capital. As regiões estão por sua vez divididas em sectores e estes em secções, compostas por aldeias (Figura 2) (Portal Oficial do Governo da Guiné-Bissau).

De acordo com o Instituto Nacional de Estatística da Guiné-Bissau em 2014, a população guineense correspondia a 1 514 451 indivíduos, dividindo-se em 48,9% de habitantes do sexo masculino e em 51,1% do sexo feminino (Camará *et al.*, 2014).

A Guiné-Bissau tem um clima predominantemente tropical com características marítimas, sendo muito quente e húmido durante todo o ano. Contudo, tal como em Angola, podem distinguir-se duas estações distintas: a estação seca, de novembro a abril, e a estação das chuvas, de maio a outubro. A temperatura média anual no país é de 26,8°C. Também de modo semelhante, devido às condições ambientais e dos solos, este território, exhibe uma enorme diversidade de animais e plantas (Benzinho *et al.*, 2015a).

1.2.2 HISTÓRIA

O território da atual Guiné-Bissau é habitado há mais de 200 mil anos. No século XIII, os Mandingas, povos de etnia muçulmana, invadiram a Guiné-Bissau e fundaram no século XV o Reino de Gabú, conhecido por Império de Kansalá, que era vassalo do Império do Mali (Assembleia Nacional Popular da República da Guiné-Bissau).

Em 1446, Nuno Tristão, vindo da costa senegalesa, chega à Guiné-Bissau, sendo posteriormente assassinado por tribos locais que se opunham à sua intrusão (Augel, 2007; Benzinho *et al.*, 2015b).

Por muito tempo, Portugal serviu-se da região, como ponto de apoio para o comércio escravagista ao longo da costa africana, vendo-a como um centro de comércio e não como uma colónia própria para a agricultura, como Angola e Moçambique. Com as sucessivas viagens dos navegadores portugueses, foram-se estabelecendo as bases para a expansão mercantil portuguesa na área e, em 1588, nasce a primeira feitoria fortificada, Cacheu (Augel, 2007).

Durante séculos, o território guineense foi administrado juntamente com o arquipélago de Cabo Verde, inicialmente desabitado, que serviu como “armazém



de depósito de mercadorias”, essencial para o tráfico humano e para o abastecimento da navegação transatlântica (Augel, 2007).

Em 1879, procede-se à separação administrativa de Cabo Verde e a colónia recebe o nome de Guiné Portuguesa tendo como primeira capital Bolama (Augel, 2007; Benzinho *et al.*, 2015b).

Aquando da Conferência de Berlim (1884-1885), foram definidas novas fronteiras e Portugal ficou com os territórios de Angola, Moçambique, São Tomé e Príncipe, Cabo Verde e Guiné-Bissau, procedendo à ocupação imediata destas colónias (Augel, 2007; Benzinho *et al.*, 2015b).

Em 1951, face à pressão internacional, o estatuto de colónia da Guiné Portuguesa é substituído pelo de Província Ultramarina, que passados quatro anos, adquire uma constituição própria e autonomia financeira e administrativa (Augel, 2007; Benzinho *et al.*, 2015b).

Durante o processo de colonização sempre foram visíveis a resistência guineense e a luta pela autodeterminação do povo nativo mas, só em 1956, é fundado o PAIGC (Partido Africano para a Independência da Guiné e Cabo Verde) por Amílcar Cabral. A partir desta data o partido deu início a atividades políticas de regulamentação e consciencialização por todo o território (Assembleia Nacional Popular da República da Guiné-Bissau; Augel, 2007; Benzinho *et al.*, 2015b).

Durante três anos a resistência do PAIGC foi pacífica mas, após o massacre dos estivadores em greve em Bissau no porto de Pinjiguiti, a agosto de 1959, onde as forças portuguesas da PIDE (Polícia Internacional e de Defesa do Estado) interromperam a manifestação matando cerca de 50 pessoas e ferindo outras 100, o PAIGC decide optar pela luta armada a qual começou em 1963 (Assembleia Nacional Popular da República da Guiné-Bissau; Benzinho *et al.*, 2015b).

Apesar da resistência árdua dos portugueses e do assassinato de Amílcar Cabral, em Conacri, no início de 1973, o PAIGC declara, unilateralmente, a independência da Guiné-Bissau a 24 de setembro de 1973. Somente depois da queda da ditadura, a 25 de Abril de 1974, Portugal reconhece oficialmente a independência deste país, a 10 de setembro de 1974, aquando da deliberação da Assembleia Geral das Nações Unidas (Assembleia Nacional Popular da República da Guiné-Bissau; Augel, 2007; Benzinho *et al.*, 2015b).



1.3 A MIGRAÇÃO EM PORTUGAL

Até finais do século XX, a estrutura e o crescimento populacional de Portugal, foram influenciados pela emigração de muitos portugueses em busca de melhores condições de vida noutros países. Os movimentos de saída, tendencialmente de indivíduos em idade ativa, tiveram impacto na estrutura etária da população, contribuindo para o seu envelhecimento e para a escassez de mão-de-obra em alguns setores da produção industrial em Portugal. Contudo, durante o Estado Novo sempre se tentou compensar esta situação através do recrutamento de força de trabalho a partir das colónias africanas que, nesta época, eram oficialmente território português, sendo este fenómeno denominado de migração inter-regional de cidadãos portugueses (Mendes *et al.*, 2012; Pinheiro, 2008).

Com a Revolução de 1974, e a posterior queda do Império Colonial Português, cerca de meio milhão de colonos portugueses regressaram a Portugal. Apesar de não serem considerados imigrantes, por causa da sua nacionalidade portuguesa, o fenómeno dos repatriados – designados retornados – originou, num curto espaço de tempo, o maior crescimento da população residente no país desde então. Com os retornados vieram, aproximadamente, 28 000 imigrantes africanos, para escaparem a situações de mudança e instabilidade política e de guerra nas ex-colónias (Machado, 1997; Pinheiro, 2008).

A integração de Portugal na Comunidade Europeia provocou um enorme crescimento económico, sendo que o país passou a atrair um número significativo de imigrantes trabalhadores a partir de 1980. Enquanto, entre 1970 e 1980, a imigração esteve restringida à circulação a partir das antigas colónias, no final de 1990, houve uma mudança significativa nos fluxos de imigração, observando-se um aumento de indivíduos oriundos do Brasil. No final do século XX, começam a chegar a Portugal cidadãos da Europa de Leste, especialmente da Ucrânia (Pinheiro, 2008).

Portugal deixa assim de ser um país de emigração e torna-se no destino para alguns milhares de imigrantes, vindos principalmente da Europa de Leste, África e Brasil (Pinheiro, 2008).

Segundo o Relatório de Imigração, Fronteiras e Asilo do Serviço de Estrangeiros e Fronteiras, referente ao ano 2015, residiam em Portugal 383 759 estrangeiros com título de residência válido, e 4 972 que prorrogaram os seus vistos de longa duração, perfazendo um total de 388 731 indivíduos (Dias *et al.*, 2015).



A Figura 3 mostra quais as nacionalidades de imigrantes mais representativas em Portugal. Há que referir que a população mais expressiva é a brasileira enquanto a angolana e a guineense se encontram em sexto e oitavo lugar, respetivamente (Dias *et al.*, 2015).

Durante muitos anos, tanto os angolanos como os guineenses estiveram sobre a influência de movimentos migratórios, redes de comércio e invasões que conduziram ao aparecimento de indivíduos com características genéticas provenientes de diferentes origens. Com a posterior migração destas populações, estabeleceram-se relações entre indivíduos de Angola e Guiné-Bissau e os habitantes nativos de Portugal, o que pode contribuir para o aumento da variabilidade genética da população portuguesa nativa (Alvarez *et al.*, 2009)

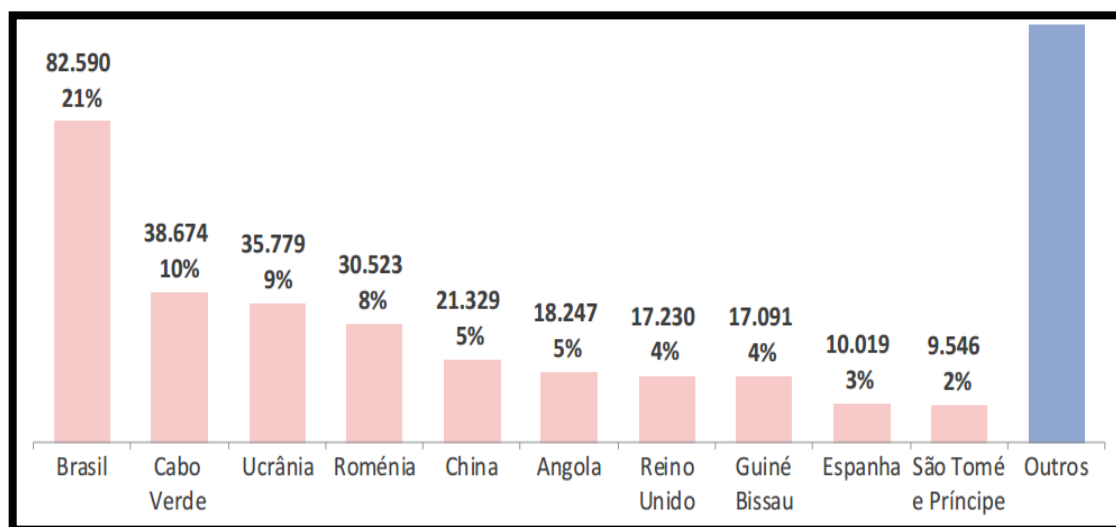


Figura 3: Principais nacionalidades de imigrantes em Portugal. Na parte superior é possível verificar o número de cidadãos de cada nacionalidade bem como a respetiva percentagem no ano de 2015 (Adaptado de Dias *et al.*, 2015).



1.4 GENOMA HUMANO

A célula é a unidade básica da vida e como tal, tem capacidade de produzir os materiais e energia essenciais à sua manutenção e de eliminar os elementos que lhe são nocivos (Butler, 2010a).

Um ser humano é constituído por, aproximadamente, cem mil milhões de células, todas elas originadas a partir de uma única, o zigoto. Este é formado na fecundação do óvulo pelo espermatozoide, onde ocorre a junção da informação genética de ambos os progenitores, sendo posteriormente transmitida a todas as células do descendente. Em todas elas existe um composto orgânico, designado por Ácido Desoxirribonucleico (DNA), que contém a informação genética necessária para o processo de replicação e para a produção de proteínas essenciais ao organismo. A totalidade da informação genética existente numa célula é designada por genoma (Bhatnagar *et al.*, 1999; Butler, 2010a).

O genoma humano é constituído pelo DNA nuclear e pelo DNA mitocondrial (Figura 4). O DNA nuclear está organizado em vinte e dois pares de cromossomas autossómicos, não sexuais, e um par de cromossomas sexuais, importantes para a determinação do sexo do indivíduo, designados por X ou Y. Os cromossomas estão associados a proteínas denominadas de histonas, que têm a função de os proteger. Os indivíduos do sexo feminino são portadores de duas cópias do cromossoma X enquanto os indivíduos do sexo masculino possuem uma única cópia de cada um dos cromossomas sexuais, um X e um Y. (Butler, 2010a).

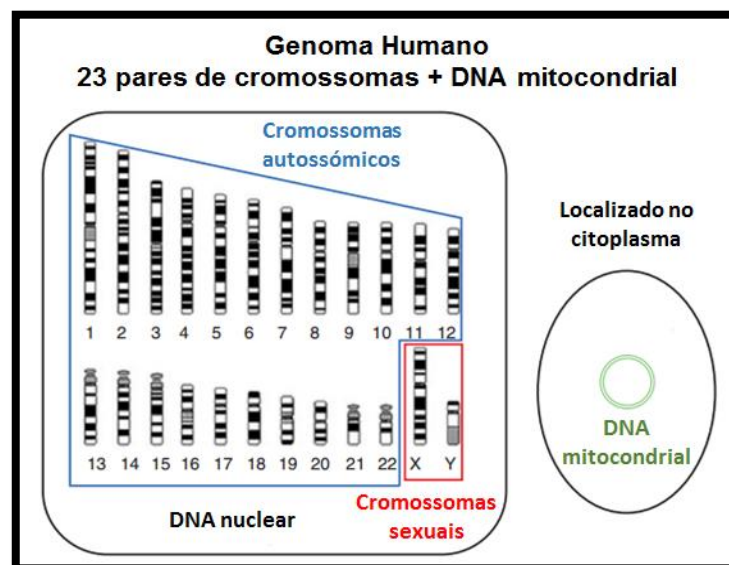


Figura 4: Organização do genoma humano dentro da célula. Dentro do núcleo da célula encontram-se vinte e dois pares de cromossomas autossómicos e um par de cromossomas sexuais; no citoplasma, o DNA mitocondrial. (Adaptado de Butler, 2010a)



O DNA é constituído por regiões codificantes e não codificantes (Figura 5). Os exões, elementos presentes nos genes, correspondem às regiões codificantes, pois contêm a informação necessária para a formação de um produto funcional. No entanto, estes representam apenas cerca de 5% do genoma humano, sendo que o restante corresponde a regiões não codificantes, como os intrões e as regiões intergénicas, onde estão localizados os marcadores genéticos usados para a identificação de indivíduos na genética forense (abordados com mais pormenor na seção 1.5.1) (Butler, 2010a; Regateiro, 2007a).

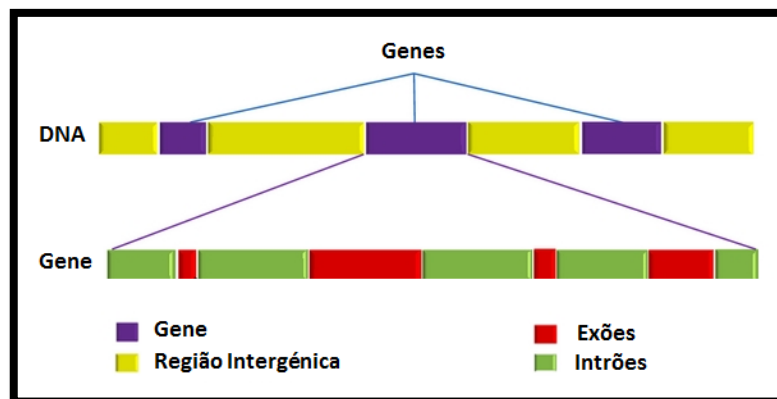


Figura 5: Constituição do DNA e dos genes.

O DNA contém regiões génicas, os genes (a **roxo**), e regiões entre os genes ou intergénicas (a **amarelo**). Os genes são compostos por exões (a **vermelho**) e intrões (a **verde**). (Adaptado de: Liu, G. *et al.*, 2014)

Relativamente à posição ou localização cromossómica de um gene ou de um marcador genético, esta é comumente referida como *locus*, no plural *loci*. Cada uma das formas génicas localizada num determinado *locus* é designada por alelo. Se os dois alelos de determinado *locus* forem diferentes, o indivíduo é denominado de heterozigótico. Por outro lado, se os alelos forem iguais, o indivíduo é homozigótico (Butler, 2010a).

1.4.1 ESTRUTURA DO DNA

O DNA é formado por duas cadeias helicoidais constituídas por nucleótidos. Cada nucleótido é formado por uma base azotada, uma molécula de açúcar, a desoxirribose e um grupo fosfato. As diferenças entre os nucleótidos ocorrem devido à existência de quatro bases distintas, a adenina (A) e a guanina (G), designadas por purinas, e a citosina (C) e a timina (T), designadas por pirimidinas (Butler, 2010a; Regateiro, 2007a).



O esqueleto de cada hélice é composto por ligações longitudinais entre o açúcar de um nucleótido e o fosfato do nucleótido seguinte, ficando as bases dispostas para o seu interior. Estas emparelham de modo complementar, sendo que a adenina se liga à timina por duas pontes de hidrogénio e a citosina à guanina por três destas ligações. A cada conjunto de duas bases azotadas unidas por pontes de hidrogénio dá-se o nome de par de bases (pb) (Martins *et al.*, 2008; Regateiro, 2007a).

As duas cadeias de DNA são antiparalelas, tendo uma delas orientação 5' → 3' e a outra a orientação oposta, 3' → 5'. Esta enumeração tem em conta a estrutura química do DNA e refere-se à posição do átomo de carbono na molécula do açúcar (Figura 6) (Butler, 2010a).

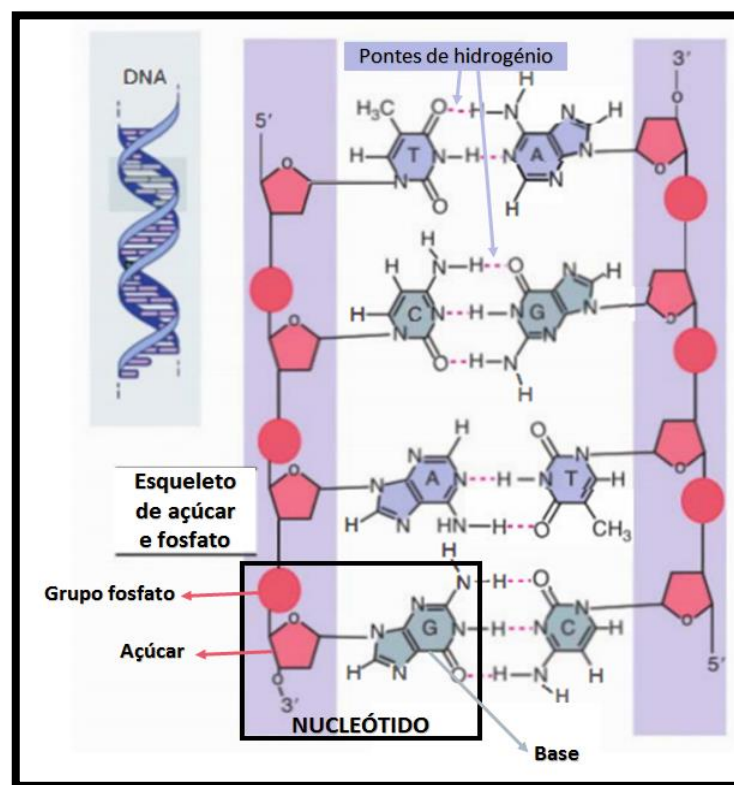


Figura 6: Estrutura do DNA. Representação esquemática do nucleótido, da estrutura helicoidal do DNA e do esqueleto de cada hélice. (Adaptado de Krebs *et al.*, 2014).

1.4.2 VARIABILIDADE GENÉTICA

As células humanas dividem-se continuamente, de tal forma que a probabilidade de ocorrerem erros é elevada. Alguns destes erros são corrigidos por mecanismos de reparação do DNA. Porém, quando estes mecanismos falham, algumas das alterações persistem no genoma dos indivíduos, designando-se por



mutações. Assim, as mutações são todas e quaisquer alterações permanentes na sequência de pares de bases do DNA e podem ocorrer sob a forma de substituições, inserções, deleções, translocações ou duplicações de nucleótidos (Martins *et al.*, 2008).

As mutações por substituição dividem-se em transições, onde ocorre a troca de uma purina ou pirimidina por outra purina ou pirimidina, respetivamente e em transversões, onde ocorre a permutação de uma purina por uma pirimidina, ou vice-versa (Martins *et al.*, 2008).

As alterações na sequência de DNA podem ser vantajosas, desvantajosas ou neutras. As mutações neutras, apesar de provocarem mudanças na sequência, não afetam o fenótipo – características físicas, bioquímicas e/ou fisiológicas observadas num indivíduo ou numa célula, resultantes da interação do meio ambiente com um gene ou mais genes (Martins *et al.*, 2008; Regateiro, 2007a).

Quando as modificações ocorrem nas células somáticas (células não sexuais) podem originar um conjunto de células mutantes idênticas entre si, que vão estar presentes em toda a vida do indivíduo, mas que não irão afetar a da sua descendência. Por outro lado, uma alteração nas células germinativas (células que darão origem aos espermatozoides ou aos óvulos) pode ser transmitida aos descendentes, podendo ficar presente em todas as suas células (Martins *et al.*, 2008).

As mutações podem dar origem a formas génicas alternativas que, quando se encontrem presentes em pelo menos 1% da população passam a constituir polimorfismos desses *loci*, ou variantes alélicas no caso desses *loci* serem genes (Strachan *et al.*, 2015).

A grande variabilidade genética entre indivíduos é, essencialmente, explicada pelos polimorfismos ocorridos, ao longo do tempo, nas regiões não codificantes, que por não sofrerem pressão seletiva, isto é, não darem origem a alterações de fenótipo, são fixados no genoma e passam assim à descendência (Martins *et al.*, 2008; Kiaris, 2012).

Para além das mutações, a recombinação é outro fator responsável pela variabilidade genética na população humana. Este processo permite a produção de novas combinações de genes através da troca aleatória entre segmentos de cromossomas semelhantes durante a formação das células germinativas (Butler, 2010a; Lodish *et al.*, 2013).



1.5 GENÉTICA FORENSE

A genética forense pode ser definida como a área da genética que recorre à análise de amostras biológicas, utilizando técnicas de biologia molecular, com o objetivo de auxiliar a Justiça. Esta ciência permite a identificação biológica dos indivíduos sendo, por isso, usada na resolução de casos de investigação de parentesco biológico, identificação de desconhecidos e criminalística biológica (Dario *et al.*, 2016).

A evolução da genética forense teve início no século passado com a descoberta dos polimorfismos do sistema ABO e das suas características hereditárias, por Karl Landsteiner (Landsteiner *et al.*, 1928), os quais foram usados para a realização de investigações de paternidade, até meados nos anos 80 (Dario *et al.*, 2016).

Contudo, em 1985, Alec Jeffreys descreve pela primeira vez, uma técnica na altura conhecida como DNA *fingerprinting* ou atualmente como DNA *typing*, baseada na existência de polimorfismos em regiões do genoma humano com sequências de DNA repetidas em *tandem* (Jeffreys *et al.*, 1985), isto é, sequências que se repetem sucessivamente na mesma região cromossómica sendo o número de repetições variável de indivíduo para indivíduo (Carracedo *et al.*, 2000; Goodwin *et al.*, 2011a). Estas regiões, com unidades de repetição entre 6 a 100 pares de bases, ficaram conhecidas como minissatélites ou *variable number of tandem repeats* (VNTRs). Jeffreys conseguiu aumentar a performance dos testes de paternidade, utilizando um método de análise de minissatélites designado por *restriction fragment length polymorphism* (RFLP) (Butler, 2010b; Goodwin *et al.*, 2011a).

Foi no Reino Unido que o DNA *fingerprinting* foi utilizado pela primeira vez, em dois casos distintos. O primeiro, um caso de imigração, no qual Andrew, um rapaz nascido no Reino Unido e emigrado no Gana, onde vivia o seu pai, regressou sozinho à sua terra natal sendo impedido de voltar para a sua mãe, Christiana, irmão e irmãs, pelos serviços de migração que suspeitavam das relações de parentesco. Alec Jeffreys, assumiu o caso e procedeu à recolha de sangue de todos os intervenientes. De seguida, determinou o perfil de DNA de cada um e comparou-os entre si. Este procedimento permitiu-lhe concluir que havia cerca de uma hipótese em mil milhões de Christiana não ser mãe de Andrew (Jeffreys *et al.*, 1985). E o segundo, um caso de criminalística, no qual duas raparigas de 15 anos, Lynda Mann e Dawn Ashworth, foram violadas e assassinadas. Nos locais do crime



foi recolhido sêmen de onde foi extraído o DNA. Durante a investigação, a criação de perfis genéticos não só identificou o verdadeiro agressor, como também permitiu excluir um suspeito inocente, que tinha confessado a autoria dos crimes (Butler, 2010b). Contudo, a análise de minissatélites apresentava algumas limitações, tais como, a necessidade de utilizar uma grande quantidade de DNA, muitas vezes indisponível nas cenas de crime, e ser uma técnica extremamente laboriosa e morosa (Goodwin *et al.*, 2011b).

Em meados dos anos 80, Kary Mullis e os seus colegas (Mullis *et al.*, 1987) desenvolveram a técnica de PCR (do inglês, *Polimerase Chain Reaction*), a qual tornou possível a análise do material genético a partir de pequenas quantidades de amostra biológica, impulsionando o desenvolvimento de novos métodos de análise do DNA (Li, 2015a).

No final dos anos 80, foram descobertas outras regiões, semelhantes aos VNTRs mas com unidades de repetição mais pequenas, entre 1 a 6 pares de bases, designadas por microssatélites ou *short tandem repeats* (STRs) (Figura 7) que rapidamente se tornaram os marcadores mais utilizados em genética forense (Butler, 2010b; Goodwin *et al.*, 2011b).

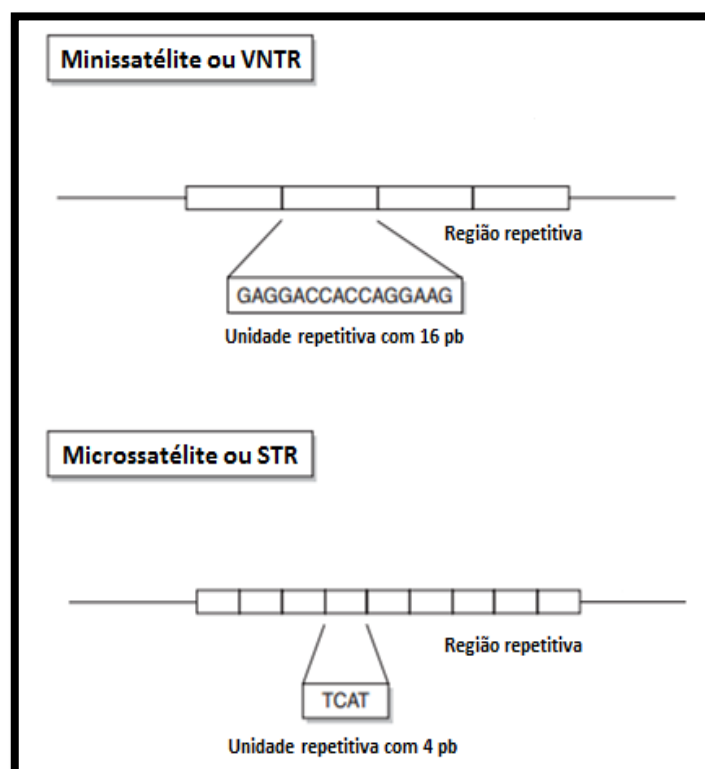


Figura 7: Sequências repetidas em *tandem*. Representação esquemática dos minissatélites ou VNTR e dos microssatélites ou STR. (Adaptado de Butler, 2010c).



1.5.1 MICROSSATÉLITES OU SHORT TANDEM REPEATS (STRs)

Para a resolução das investigações forenses é necessário proceder à identificação genética das amostras biológicas correspondentes. Esta assenta na caracterização de polimorfismos do DNA das regiões não codificantes do genoma humano, também designados por marcadores genéticos (Pinheiro, 2016).

Os STRs constituem a base da identificação genética. Apesar de apresentarem unidades de repetições com tamanho variável, os mais usados em genética forense são os tri, tetra e pentanucleotídicos, com 3, 4 e 5 nucleótidos por unidade de repetição, respetivamente. Estima-se que existam cerca de 500 000 STRs dispersos por todo o genoma, dois quais 6000 a 10000 são tri ou tetraméricos (Pinheiro, 2016).

Estes marcadores genéticos são muito polimórficos e apresentam baixas taxas de mutação na transmissão à descendência; não sofrem pressão seletiva; possuem alelos de tamanhos semelhantes, o que possibilita a amplificação simultânea de vários *loci* levando, também, à redução da ocorrência do fenómeno de *dropout* alélico (amplificação preferencial dos alelos de tamanho reduzido, que pode resultar no desaparecimento dos de maior tamanho); formam poucos produtos *stutter* (subprodutos da amplificação por PCR que diferem do alelo verdadeiro por apresentarem uma, ou mais, unidades de repetição a menos e que levam ao aparecimento de falsos alelos); e permitem a amplificação por PCR a partir de amostras com quantidades diminutas de DNA ou com DNA degradado. Os perfis genéticos obtidos, rápida e economicamente, através da análise de STRs são fáceis de interpretar (Butler, 2011a; Corte-Real, A. *et al.*, 2016; Goodwin *et al.*, 2011c; Pinheiro, 2016).

Existem STRs autossómicos, STRs no cromossoma Y e STRs no cromossoma X. Os primeiros, são usados na resolução de todo o tipo de perícias de genética forense, os segundos e os terceiros, quando se pretende obter informação complementar. Os STRs, principalmente os autossómicos, podem existir na forma de mini-STR, usados quando o material genético se encontra degradado e não é possível a obtenção de resultados (Pinheiro, 2008).

Em 1997, a Comissão Internacional de DNA da Sociedade Internacional de Hemogenética Forense propôs regras para a nomenclatura dos STRs de modo a facilitar a comunicação entre a toda a comunidade forense (Butler, 2010a; Olaisen *et al.*, 1997).



A primeira regra estabelece que a leitura das sequências de DNA seja feita da extremidade 5' para a 3'; a segunda, dita que determinado STR localizado nos genes codificantes de proteínas terá a abreviatura da proteína codificada, como o marcador TPOX (situado no gene que codifica para a peroxidase da tiroide humana) contudo, no caso dos marcadores encontrados nas regiões intergênicas o nome é dado recorrendo à designação D#S## em que, a seguir à letra D, que indica DNA, aparece o número do cromossoma no qual se localiza o STR e, subsequentemente à letra S, que significa sequência única, os números identificadores, como por exemplo o marcador D18S51 que se situa no cromossoma 18 e no 51º *locus* descrito nesse cromossoma; a terceira regra corresponde à classificação da sequência de repetição, em que o número de vezes que esta se repete é o número do alelo em questão (Figura 8) (Butler, 2010a; Olaisen *et al.*, 1997).

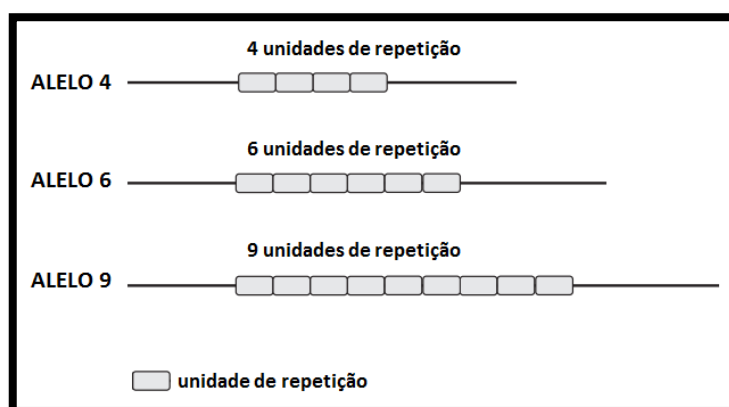


Figura 8: Classificação dos alelos. Os alelos 4, 6 e 9 possuem 4, 6 e 9 unidades de repetição, respetivamente. (Adaptado de Butler, 2010d)

Nem todos os alelos apresentam unidades de repetição completas designando-se de microvariantes, as quais são classificadas com o número de repetições completas seguidas de um ponto decimal e do número de nucleótidos da repetição incompleta (Butler, 2010d). Na figura 9 encontra-se representada a microvariante 9.3 do marcador TH01.

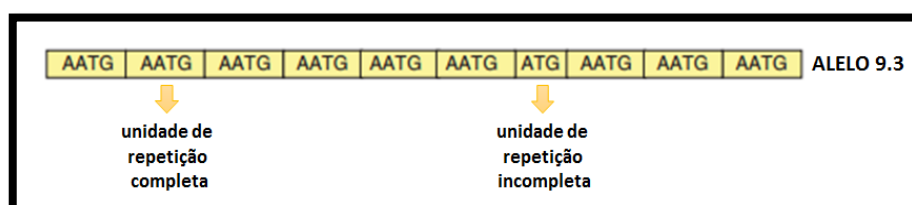


Figura 9: Estrutura da microvariante 9.3 do marcador TH01. O alelo 9.3 deste marcador é constituído por 9 unidades de repetição completas e uma incompleta, constituída por 3 nucleótidos. (Adaptado de Goodwin *et al.*, 2011c)



Nos perfis genéticos é esperado, para determinado *locus* STR, que apareçam um ou dois alelos para indivíduos homozigóticos ou heterozigóticos, respetivamente. Porém, em casos raros, podem observar-se padrões trialélicos, isto é, três alelos num único *locus* (Zamir, A. *et al.*, 2002).

Tim Clayton, propôs a existência de dois tipos de padrões (Figura 10). O tipo I, é explicado por uma mutação somática num *locus* heterozigótico, durante o desenvolvimento do indivíduo, que resulta numa quimera (fenómeno no qual algumas células apresentam o alelo original e outras o alelo mutante). Neste, os 3 alelos apresentam diferentes alturas, sendo a soma das alturas dos dois alelos de menor tamanho equívale à altura do terceiro. No padrão do tipo II, resultante de uma duplicação pontual num *locus* heterozigótico, ao nível das células germinativas, todos os alelos exibem alturas semelhantes (Clayton *et al.*, 2004).

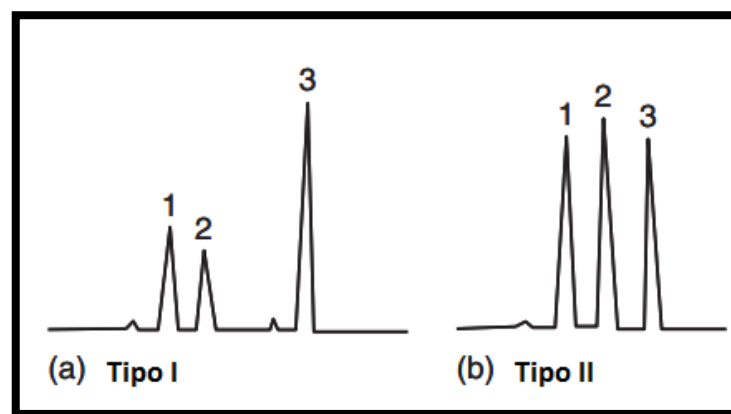


Figura 10: Tipos de padrões trialélicos. (a) padrão trialélico tipo I e (b) tipo II. (Adaptado de Butler, 2010e)

1.5.1.1 METODOLOGIAS PARA A ANÁLISE DOS STRs

Classicamente, a análise dos STRs engloba três etapas: extração do DNA das amostras em estudo, amplificação por PCR e eletroforese capilar.

A primeira, tem como objetivo obter DNA a partir de determinada amostra, em quantidade e pureza suficientes que permitam a aquisição de um perfil genético completo. A escolha do método de extração de DNA depende de vários fatores incluindo o tipo e a quantidade disponível de amostra, a taxa de sucesso com amostras forenses, o tipo de produtos químicos usados, o custo e a rapidez do processo e a experiência do pessoal do laboratório. Independentemente do procedimento escolhido, todos são constituídos por três fases, (1) disrupção das membranas celulares, resultando na lise das células, (2) desnaturação das



proteínas e (3) separação do DNA das proteínas desnaturadas e de outros componentes celulares, sendo que as amostras devem ser manipuladas cuidadosamente para evitar a contaminação amostra-a-amostra ou a introdução de DNA exógeno. O processo de extração é provavelmente aquele em que as amostras estão mais suscetíveis à contaminação nos laboratórios, os quais possuem diferentes locais, conforme a etapa da análise, para o seu manuseamento (Butler, 2010d; Goodwin *et al.*, 2011d; Li, 2015c).

O método de *Chelex*[®] 100, desenvolvido por Walsh (Walsh *et al.*, 1991) e, posteriormente, modificado por Willard (Willard *et al.*, 1998) foi dos primeiros a ser adotado pela comunidade forense, sendo um dos mais usados atualmente por possuir vantagens como a rapidez, simplicidade, reduzida possibilidade de contaminação das amostras, baixo custo, uso de produtos não nocivos para a saúde e aplicabilidade a uma vasta gama de amostras forenses. O *Chelex*[®] 100 é uma resina quelante, que se liga aos íons de magnésio presentes na molécula de DNA, essenciais para o funcionamento das nucleases que a degradam. Estas mantêm-se inativas deixando de interferir com a integridade do material genético e com a, posterior reação de amplificação (Goodwin *et al.*, 2011d; Li, 2015c).

Usando este método, o primeiro passo consiste na lavagem das amostras adicionando água desionizada e incubação à temperatura ambiente, para remover possíveis contaminantes ou inibidores da PCR. De seguida, incubam-se os tubos com as amostras, já com a resina adicionada anteriormente, a 56°C resultando no enfraquecimento das membranas celulares e separação dos aglomerados de células. Continuando o procedimento, as amostras são colocadas em água a 100°C, ocorrendo a destruição das células e libertação do DNA. Após uma centrifugação rápida, a resina e os restos celulares ficam retidos no fundo dos tubos e o sobrenadante, onde se encontra o DNA, é utilizado para a amplificação por PCR (Figura 11). Deve incluir-se, juntamente com os tubos que contêm as amostras a estudar, um sem material genético designado por controlo negativo para verificar se há ou não contaminação dos reagentes utilizados (Butler, 2010f; Goodwin *et al.*, 2011d; Li, 2015c; Walsh *et al.*, 1991; Willard *et al.*, 1998).

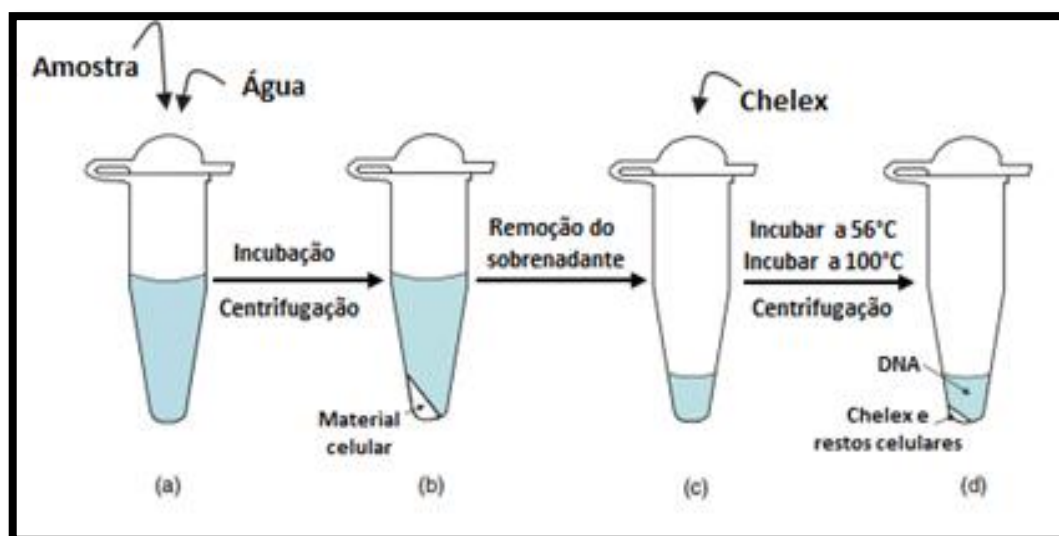


Figura 11: Extração de DNA com *Chelex*[®] 100. (a) Adição de água à amostra, incubação à temperatura ambiente e centrifugação. (b) Remoção do sobrenadante. (c) Adição do *Chelex*, incubação a 56°C e seguidamente, a 100°C e centrifugação. (d) O sobrenadante contém o DNA que será usado na PCR. (Adaptado de Goodwin *et al.*, 2011d)

A segunda etapa da análise dos STRs, é a amplificação por PCR (Mullis *et al.*, 1987) num instrumento designado por termociclador. Esta baseia-se nos processos enzimáticos da replicação de DNA, possibilitando que uma sequência específica seja amplificada exponencialmente, dando origem a múltiplas cópias da mesma. Adequa-se à análise de amostras forenses por ser uma técnica sensível, rápida e permitir a amplificação de várias sequências numa mesma amostra com quantidades diminutas de DNA ou DNA degradado (Butler, 2010g; Goodwin *et al.*, 2011e; Li, 2015d).

Nos dias de hoje, os laboratórios forenses usam, maioritariamente, *kits* comerciais validados para a amplificação e caracterização de amostras biológicas. Estes contêm todos os componentes necessários para o procedimento devendo o material genético ser amplificado de acordo com o descrito no manual do *kit* utilizado para o efeito (Pinheiro, 2008).

Na preparação das amostras é necessário incluir um controlo negativo e outro positivo de modo a monitorizar a eficácia das condições experimentais escolhidas e/ou a técnica do operador. O controlo negativo contém todos os componentes necessários à amplificação, com a exceção do DNA que é substituído por água, avaliando se algum deles foi ou não contaminado com o material genético. O controlo positivo é um indicador da adição ou não dos componentes



às amostras e da eficácia da técnica. Neste é utilizado uma molécula de DNA padronizada e de boa qualidade, disponível no *kit* utilizado (Butler, 2010g).

Classicamente, a amplificação ocorre em três fases, (1) desnaturação, (2) *annealing* e (3) extensão. Primeiro, as amostras preparadas previamente, são expostas à temperatura de 94°C fazendo com que as cadeias helicoidais da molécula de DNA se separem formando duas moléculas de cadeia simples. Isto ocorre devido à rutura das ligações de hidrogénio que as mantêm unidas. Posteriormente, a temperatura é reduzida até aos 50-65°C, permitindo o emparelhamento dos *primers*, que são pequenas sequências de DNA que precedem a região a amplificar de modo a identifica-la, às extremidades 3' das cadeias simples. Por fim, a temperatura é aumentada até à temperatura ótima para o funcionamento da Taq polimerase que catalisa a síntese de novas cadeias de DNA usando a original como molde (Figura 12). Estas fases repetem-se, normalmente, entre 28 a 32 ciclos (Butler, 2010g; Goodwin *et al.*, 2011e).

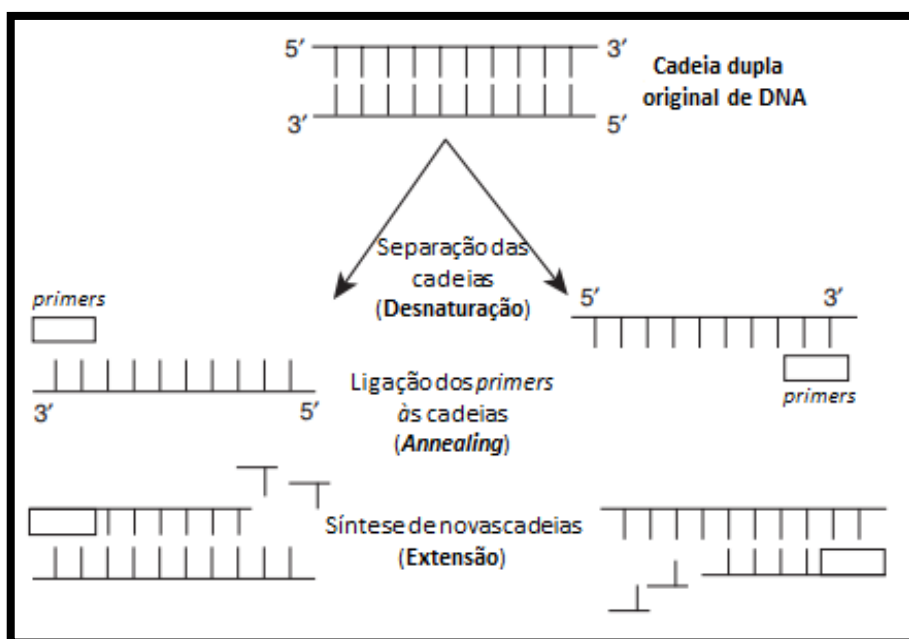


Figura 12: Amplificação do DNA por PCR.

Em cada ciclo, a cadeia dupla de DNA é separada, originando duas cadeias simples, pela exposição a temperaturas elevadas; os *primers* ligam-se às cadeias simples devido à diminuição da temperatura; e, ocorre a síntese de novas cadeias quando as amostras são submetidas à temperatura ótima de atividade da polimerase. (Adaptado de Butler, 2010g)

Após a reação de PCR é necessário separar os STRs amplificados. Para tal procede-se à eletroforese capilar (EC) num sequenciador automático. Antes da EC, as amostras são preparadas numa placa multipoços, nos quais se adiciona o produto amplificado ou o *ladder* alélico, a formamida, reagente que permite que



as cadeias de DNA se mantenham desnaturadas, e o marcador interno. De seguida, a placa é colocada num termociclador para desnaturar as amostras e posteriormente, no sequenciador onde estas são injetadas pelo capilar através da aplicação de um campo elétrico, o qual permite que as moléculas negativas, como o DNA, migrem do polo negativo para o polo positivo. Esta migração é feita consoante o tamanho das moléculas, migrando as mais pequenas mais rapidamente e mais lentamente as de maiores dimensões, até atingirem a câmara de passagem do laser onde, estando marcadas com fluoróforos, são excitadas e emitem fluorescência. Essa fluorescência é detetada no sequenciador, sendo a informação armazenada no software do equipamento. Após a recolha dos dados, o software remove a sobreposição dos espetros e calcula o tamanho dos fragmentos amplificados utilizando o padrão conhecido do marcador interno, o qual contém fragmentos de DNA com tamanhos conhecidos e também marcados com fluoróforos. É então feita uma curva de calibração que compara os picos das amostras desconhecidas com os picos do marcador interno. O último passo é a atribuição de designações aos alelos dos produtos amplificados e para tal é necessário a presença, na corrida de eletroforese, do *ladder* alélico. Assim, os picos desconhecidos são comparados com o *ladder* possibilitando a obtenção de um perfil genético ou eletroforegrama (Figura 13) (Butler, 2010h; Butler, 2011b; Goodwin *et al.*, 2011f).

Esta técnica tem como vantagens, a automatização da injeção, da separação e da deteção; necessitar de pouca quantidade de amostra, o que torna possível repetir o ensaio; a rapidez de separação das moléculas; possibilidade de armazenamento eletrónico do resultado; e a análise de várias amostras simultaneamente quando existem vários capilares no aparelho (Butler, 2010h; Butler, 2011b; Goodwin *et al.*, 2011f).

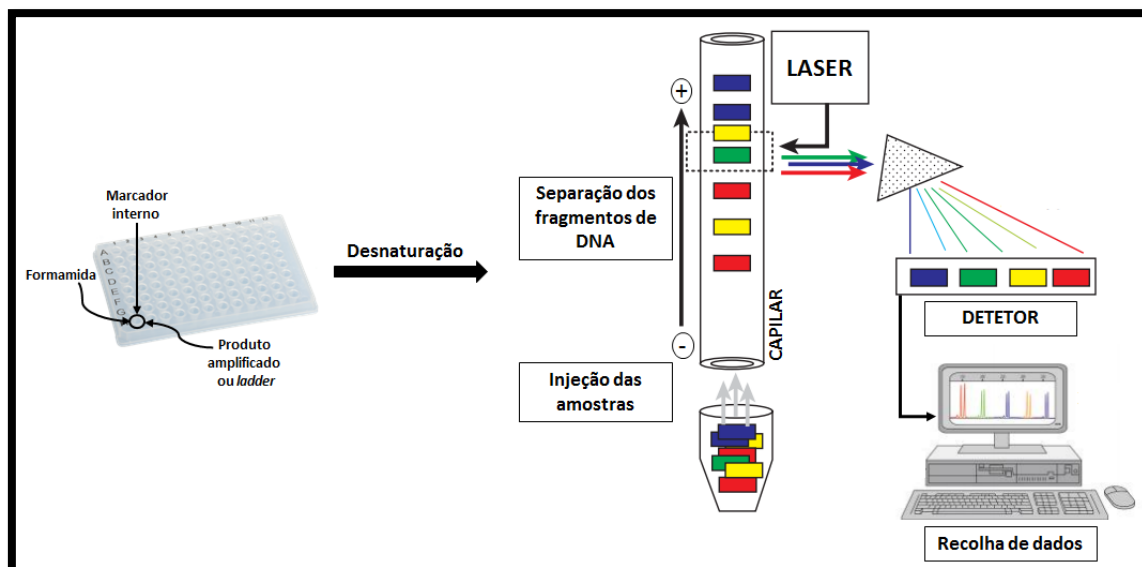


Figura 13: Representação esquemática da técnica de eletroforese capilar. (Adaptado de Butler, 2011b; Li, 2015e)

1.5.2 BASES DE DADOS DE PERFIS DE DNA

Com o objetivo de facilitar as investigações forenses a nível nacional e internacional, tornou-se essencial a criação de metodologias que permitam a comparação de perfis de DNA e a troca de informação entre laboratórios. Para este efeito, foi necessário estabelecer métodos de trabalho idênticos, uniformizar os critérios de nomenclatura e de seleção dos STRs e desenvolver bases de dados de perfis genéticos.

Uma base de dados de perfis de DNA possui um conjunto informatizado e estruturado de registos de perfis genéticos e ficheiros de dados pessoais. Geralmente estes perfis são obtidos de amostras recolhidas no local do crime ou de pessoas. Assim, as bases de dados permitem a comparação entre o perfil genético de uma amostra e os perfis já aí existentes, sendo muito úteis na verificação de ligações entre crimes, de modo a determinar se o mesmo indivíduo está ou não envolvido em diferentes delitos, realizados no país de residência ou noutros; na resolução de casos antigos, arquivados por falta de provas; na identificação de pessoas desaparecidas ou vítimas de grandes catástrofes; e de perpetradores, não suspeitos do crime, quando há correspondência entre os perfis, um originado na análise do vestígio e outro existente na base de dados. (Conselho de Fiscalização da Base de Dados de Perfis de ADN, 2014; Corte-Real, 2016; Pereira, 2008).



A primeira base de dados de perfis de DNA nacional, conhecida como *National DNA Database* (NDNAD), foi criada no Reino Unido a 10 de abril de 1995. Esta incluía o estudo de seis STRs autossómicos (FGA, TH01, vWA, D8S1179, D18S51 e D21S11). A NDNAD tornou-se um modelo para muitos países em todo o mundo levando ao desenvolvimento e crescimento de novas bases de dados de perfis de DNA nacionais (Butler, 2011c).

As bases de dados têm várias estruturas organizacionais, dependendo do enquadramento regulamentar nacional de cada país. Todavia, cada um deles possui um número definido de STRs, que deve ser estudado rotineiramente para todas as amostras, de modo a que estas sejam aceites nas bases. Em 1998, ocorre a seleção destes marcadores baseada nos dados experimentais de exercícios colaborativos entre diversos grupos de trabalho, como o *European DNA Profiling Group* (EDNAP), em pequenos *kits* comerciais de análise de STRs, originalmente concebidos pelo extinto *Forensic Science Service* (Birmingham, Reino Unido), e nas recomendações do grupo de trabalho em DNA da INTERPOL. Um ano mais tarde, a *European Network of Forensic Institutes* (ENFSI) designa um conjunto mínimo de STRs para que os perfis possam ser inseridos nas bases de dados europeias, designado por *European Standard Set* (ESS) (Butler, 2011c; Pereira, 2008; Schneider, 2009).

O ESS, originalmente constituído por sete marcadores genéticos autossómicos, os seis pertencentes à NDNAD mais um (D3S1358), pretendia facilitar a partilha de perfis entre os diversos países, promovendo a cooperação internacional para a resolução de investigações forenses (Butler, 2011c; Pereira, 2008; Schneider, 2009). Em 2001, este conjunto é legislativamente confirmado através da Resolução 2001/C187/01 do Conselho Europeu (Resolução do Conselho de 25 de junho de 2001). Contudo, em 2009, considerando que o intercâmbio de perfis de DNA entre os Estados Membros progrediu rapidamente e que as bases de dados nacionais de DNA aumentaram em dimensão e número, o ENFSI propõe a adição de um novo conjunto de marcadores. Assim, o ESS passa a ser constituído por 12 marcadores autossómicos, os sete anteriormente referidos e mais cinco mini-STRs (D1S1656, D2S441, D10S1248, D12S391 e D22S1045), mantendo-se igual até à atualidade (Resolução do Conselho de 30 de novembro de 2009).

Nos Estados Unidos, a base de dados nacional, designada por *Combined DNA Index System* (CODIS), foi criada em 1998. No entanto, iniciou-se como um projeto piloto em 1990, com 14 laboratórios estatais e locais, tornando-se oficial em 1994, com a lei do “*DNA Identification Act*”, a qual atribuiu responsabilidades ao FBI (*Federal Bureau of Investigation*) para implementar uma base de dados



(Pereira, 2008). O CODIS inclui o estudo de 13 *loci*, sendo que aos sete STRs autossômicos do ESS original foram adicionados mais seis (TPOX, CSF1PO, D16S539, D7S820, D13S317, D5S818) (Butler, 2011a).

Recentemente, o FBI anunciou a expansão do número de *loci* originais do CODIS. Portanto, foram selecionados sete STRs adicionais que serão implementados até 1 de janeiro de 2017, passando o CODIS a ser constituído por 20 marcadores genéticos autossômicos visto que, aos 13 STRs do CODIS inicial será acrescentado um conjunto adicional (D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433 e D22S1045). Para permitir a análise destes novos *loci*, a *Promega Corporation* e a *Applied Biosystems* lançaram novos *kits* (Hares, 2015).

Em Portugal existe, desde 2008, uma Base de Dados de Perfis de DNA, criada e regulamentada pela Lei n.º 5/2008. Esta integra STRs autossômicos de uso obrigatório e complementar. A escolha dos marcadores obrigatórios deve-se à Resolução 2001/C187/01 do Conselho Europeu e à necessidade de assegurar a compatibilidade com os marcadores utilizados nos perfis de outras bases de dados europeias, enquanto os complementares incluem os restantes STRs utilizados pela INTERPOL e pela comunidade científica internacional, permitindo um aumento da capacidade de discriminação e da compatibilização com as bases de dados europeias e evitar situações de falsas coincidências entre perfis (Lei n.º 5/2008; Portaria n.º 270/2009).

1.5.3 GLOBALFILER™ PCR AMPLIFICATION KIT

O *GlobalFiler™ PCR Amplification Kit* da *Applied Biosystems*, permite a amplificação simultânea de 24 marcadores genéticos – 20 STRs autossômicos correspondentes ao novo CODIS mais o *locus* SE33 e 3 marcadores identificativos do sexo do dador da amostra (Amelogenina, Yindel, and DYS391) – utilizando *primers* marcados com 6 fluoróforos (Applied Biosystems, 2015). Este apresenta elevada sensibilidade para amostras com baixa concentração e/ou na presença de inibidores e, por apresentar mini-STRs na sua constituição, permite a análise de amostras degradadas. O seu tempo de amplificação é reduzido e o poder de discriminação superior relativamente a outros *kits* (Applied Biosystems, 2015; Hares, 2015; Thermo Fisher Scientific Inc., 2015).

Este *kit* contém praticamente todos os reagentes necessários para a análise dos STRs, tais como a *GlobalFiler™ Master Mix*, que está tamponizada e contém albumina de soro bovino, cloreto de magnésio, nucleótidos, Taq polimerase e



azida de sódio, o *GlobalFiler™ Primer Set*, constituído pela mistura de *primers* marcados com os diferentes fluoróforos, o *DNA Control 007*, tamponizado e formado por DNA genómico de um indivíduo do sexo masculino, azida de sódio e o *GlobalFiler™ Allelic Ladder*, que possui a maioria dos alelos descritos para cada um dos marcadores estudados (Applied Biosystems, 2015).

Cada um dos constituintes da *master mix* apresenta determinada função: a albumina de soro bovino previne ou reduz a inibição da *Taq* polimerase ligando-se no lugar de alguns inibidores competitivos e podendo assim eliminar ou reduzir a concentração dos mesmos; o cloreto de magnésio estabiliza a interação dos *primers* com a cadeia de DNA; os nucleótidos são necessários para a formação de novas cadeias de DNA; a tamponização mantém o pH ótimo para a reação de amplificação; e a azida de sódio é um agente conservante do tampão e do cloreto de magnésio (Applied Biosystems, 2015; Goodwin *et al.*, 2011e).

No *GlobalFiler™*, os produtos amplificados são marcados com cinco fluoróforos diferentes, designados por 6-FAM™, VIC™, NED™, TAZ™ e SID™. Para o marcador interno, *GeneScan™ 600 LIZ™ Size Standard v2.0*, é utilizado o fluoróforo LIZ™ (Applied Biosystems, 2015). O 6-FAM™ é apresentado a azul, o VIC™ a verde, o NED™ a amarelo, o TAZ™ a vermelho, o SID™ a roxo e o LIZ™ a laranja. A Tabela I mostra os *loci* amplificados no *kit*, quais os fluoróforos que sinalizam cada *loci* e os alelos incluídos no *ladder* alélico e no *DNA Control 007*.

Tabela I: *Loci*, alelos e fluoróforos do *GlobalFiler™ PCR Amplification Kit*.

(Adaptado de Applied Biosystems, 2015)

<i>LOCUS</i>	ALELOS DO <i>LADDER</i> ALÉLICO	FLUORÓFOROS	ALELOS DO <i>DNA CONTROL 007</i>
D3S1358	9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20	6-FAM™	15, 16
vWA	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24		14, 16
D16S539	5, 8, 9, 10, 11, 12, 13, 14, 15		9, 10
CSF1PO	6, 7, 8, 9, 10, 11, 12, 13, 14, 15		11, 12
TPOX	5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15		8, 8
Amelogenina	X, Y	VIC™	X, Y
D8S1179	5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19		12, 13
D21S11	24, 24.2, 25, 26, 27, 28, 28.2, 29, 29.2, 30, 30.2, 31, 31.2, 32, 32.2, 33, 33.2, 34, 34.2, 35, 35.2, 36, 37, 38		28, 31



<i>LOCUS</i>	<i>ALELOS DO LADDER ALÉLICO</i>	<i>FLUORÓFOROS</i>	<i>ALELOS DO DNA CONTROL 007</i>
D18S51	7, 9, 10, 10.2, 11, 12, 13, 13.2, 14, 14.2, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27	VIC™	12, 15
DYS391	7, 8, 9, 10, 11, 12, 13		11
D2S441	8, 9, 10, 11, 11.3, 12, 13, 14, 15, 16, 17	NED™	14, 15
D19S433	6, 7, 8, 9, 10, 11, 12, 12.2, 13, 13.2, 14, 14.2, 15, 15.2, 16, 16.2, 17, 17.2, 18.2, 19.2		14, 15
TH01	4, 5, 6, 7, 8, 9, 9.3, 10, 11, 13.3		7, 9.3
FGA	13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 26.2, 27, 28, 29, 30, 30.2, 31.2, 32.2, 33.2, 42.2, 43.2, 44.2, 45.2, 46.2, 47.2, 48.2, 50.2, 51.2		24, 26
D22S1045	8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19	TAZ™	11, 16
D5S818	7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18		11, 11
D13S317	5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16		11, 11
D7S820	6, 7, 8, 9, 10, 11, 12, 13, 14, 15		7, 12
SE33	4.2, 6.3, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 20.2, 21, 21.2, 22.2, 23.2, 24.2, 25.2, 26.2, 27.2, 28.2, 29.2, 30.2, 31.2, 32.2, 33.2, 34.2, 35, 35.2, 36, 37		17, 25.2
D10S1248	8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19	SID™	12, 15
D1S1656	9, 10, 11, 12, 13, 14, 14.3, 15, 15.3, 16, 16.3, 17, 17.3, 18.3, 19.3, 20.3		13, 16
D12S391	14, 15, 16, 17, 18, 19, 19.3, 20, 21, 22, 23, 24, 25, 26, 27		18, 19
D2S1338	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28		20, 23



1.6 GENÉTICA POPULACIONAL

A genética populacional ocupa-se do estudo das leis da hereditariedade e das várias forças que levam a mudanças evolutivas nas espécies ao longo do tempo. Neste contexto, uma população pode ser descrita como um grupo de organismos, da mesma espécie, que vivem no interior de uma determinada área geográfica suficientemente restrita, pelo que cada membro da população pode cruzar-se (no sentido reprodutivo) com qualquer outro membro do sexo oposto desse mesmo grupo (Abrantes, 2010).

Gregor Mendel, “o pai da genética clássica”, descreveu as leis da hereditariedade (Blumberg, 1997), sendo as mais importantes a lei da segregação e a lei da segregação independente que são a base do *Linkage (Dis)Equilibrium* e do Princípio de Hardy-Weinberg, ambos usados no estudo forense das populações. A primeira afirma que cada característica é determinada por dois alelos que se separam na formação dos gametas e a segunda, refere que os alelos são herdados de forma independente devido à recombinação genética. De acordo com estas leis é possível a existência de inúmeras combinações de pares de alelos (genótipos) para um determinado *locus* (Butler, 2010a; Regateiro, 2007b).

1.6.1 ESTUDO DAS POPULAÇÕES

Para que seja possível estudar as populações é necessário, (1) escolher qual a etnia que se pretende analisar e qual o número de amostras a utilizar, (2) proceder à análise dos STRs, que engloba a extração e amplificação do DNA e, posterior, eletroforese capilar para obtenção de um perfil genético, (3) calcular as frequências alélicas, (4) realizar a análise estatística dos dados e (5) efetuar o estudo comparativo entre populações (Butler, 2012).

Se as amostras selecionadas forem obtidas de forma aleatória e corresponderem a indivíduos representativos da população de referência, não relacionados entre si, não há um número ideal de indivíduos que se possa recomendar para a realização do estudo das populações. Contudo, alguns autores referem o uso de 100 a 200 indivíduos como um valor aceitável (Butler, 2012; Souto, 2016).

Os dados obtidos são avaliados recorrendo à análise estatística de modo a assegurar que as frequências alélicas se encontram de acordo com os princípios de herança genética. Assim, o Princípio de Hardy-Weinberg e o *Linkage*



(Dis)Equilibrium são usados para o estudo da independência dos alelos dentro de um determinado *locus* e entre diferentes *loci*, respetivamente (Butler, 2012).

1.6.1.1 PRINCÍPIO DE HARDY-WEINBERG

O Princípio de Hardy-Weinberg foi descoberto em simultâneo por um matemático inglês, Godfrey Hardy, e um médico alemão, Wilhelm Weinberg, de forma independente, em 1908 (Hardy, 1908; Weinberg, 1908). É uma representação matemática simples da relação das frequências genóticas e alélicas dentro de uma população ideal e é fundamental para a genética forense. Este afirma que dentro de uma população infinita em que os cruzamentos são feitos ao acaso as frequências alélicas e genóticas de determinado *locus* genético permanecem constantes. Quando uma população obedece ao Princípio de Hardy-Weinberg diz-se que está em equilíbrio de Hardy-Weinberg (HWE) e assim, pode-se calcular as frequências genóticas esperadas (ou heterozigotia esperada – H_e) a partir das frequências genóticas observadas (ou heterozigotia observada – H_o), tendo como base as frequências alélicas (Butler, 2012; Goodwin, 2011g; Souto, 2016). A frequência alélica é o quociente entre o número de cópias de um alelo e o número total de alelos observados numa população. Por outro lado, o quociente entre o número de indivíduos com um genótipo particular e o total dos indivíduos analisados é referido por frequência genotípica (Butler, 2012).

Para que a previsibilidade e constância do equilíbrio de Hardy-Weinberg se concretizem há um conjunto de pressupostos a que as populações devem obedecer. A população tem que ser infinitamente grande e pan-mítica, ou seja, os cruzamentos devem ocorrer de forma aleatória. Mas também, não deve ocorrer migração, seleção natural nem mutação. Contudo, as populações reais, geralmente, não obedecem ao HWE e as frequências alélicas e genóticas mudam ao longo do tempo (Butler, 2012; Goodwin, 2011g; Souto, 2016).

1.6.1.2 LINKAGE (DIS)EQUILIBRIUM

A tendência que os alelos de dois ou mais *loci*, que se encontram suficientemente próximos num determinado cromossoma, têm para serem transmitidos à descendência em bloco é denominada por *linkage* (Hartl *et al.*, 2007).



O *linkage equilibrium* (LE) ocorre quando os alelos presentes num determinado *locus* são herdados independente dos alelos existentes noutra *locus* (segregação independente). Por outro lado, o *linkage disequilibrium* (LD) corresponde à associação não aleatória dos alelos de dois ou mais *loci* (segregação dependente). Este pode dever-se à não ocorrência de recombinação, que pode resultar na herança em bloco de dois *loci* localizados no mesmo cromossoma, à seleção natural, que pode conduzir à diminuição ou ao aumento das frequências dos alelos “desfavoráveis” ou “vantajosos”, respetivamente, numa determinada população, à migração (ou fluxo de genes), que facilita a ocorrência de acasalamentos entre nativos de diferentes populações levando ao aparecimento de indivíduos que possuem alelos representativos de ambas as populações, à mutação, a qual pode resultar no aparecimento de novos alelos, e à deriva genética, que ocorre normalmente em pequenas populações e que tal como a seleção natural pode ter como consequência o aumento ou a diminuição da frequência de determinados alelos no entanto, ocorre por acaso (Hartl *et al.*, 2007; Hedrick 2011a; Hedrick, 2011b; Single *et al.*, 2016). Este conceito foi usado pela primeira vez em 1960 por Lewontin e Kojima (Lewontin *et al.*, 1960) continuando até à atualidade. No início, existiam poucos dados acerca de como o estudar, e a sua importância para a biologia evolutiva e genética humana não era reconhecida fora da genética populacional. No entanto, o interesse pelo *linkage disequilibrium* cresceu rapidamente na década de 1980 quando a sua utilidade para o mapeamento genético se tornou evidente e a realização de inúmeras pesquisas na identificação de *loci* intimamente ligados se tornou possível (Slatkin, 2008).

1.6.1.3 DISTÂNCIA GENÉTICA ENTRE POPULAÇÕES

A análise estatística da distância genética mede o grau de diferenciação genética entre populações, por comparação duas a duas. A distância entre elas é tanto maior, quanto maior for o valor estatístico. Assim, se este valor for superior entre duas populações A e B do que entre C e D, podemos dizer que C e D estão mais relacionadas geneticamente do que A e B. Existem vários métodos para a medição da distância genética entre populações, no entanto o mais usado foi descrito por Wright (Wright, 1951) e é designado por índice de fixação (F_{st}). Este mede as diferenças entre as frequências alélicas de duas populações variando entre os valores 0 e 1. Onde 0 corresponde a total similaridade e o 1 a total diferenciação (Jobling *et al.*, 2014; Templeton, 2006).



1.6.1.4 TESTE EXATO DE DIFERENCIAÇÃO POPULACIONAL

Este teste calcula a distribuição dos alelos de cada *loci* entre diferentes populações estudadas, segundo descrito por Raymond e Rousset (1995) e por Goudet *et al.* (1996) (Excoffier *et al.*, 2010a; Rousset, 2015).

1.6.1.5 HOMOZIGOSIDADE E HETEROZIGOSIDADE

A homozigosidade (h) e a heterozigosidade (H) representam a proporção de indivíduos homozigóticos e heterozigóticos para cada *locus*, respetivamente. A primeira é calculada pela fórmula, $h = \sum_{i=1}^n p_i^2$, e a segunda por $H = 1 - h$ (Butler, 2005).

1.6.1.6 PARÂMETROS ESTATÍSTICOS DE INTERESSE FORENSE

Normalmente, nas tabelas com as frequências alélicas encontradas em publicações acerca do estudo forense das populações, aparece uma lista de parâmetros que são usados para a avaliação de cada marcador genético estudado (Butler, 2012). Estes são designados por parâmetros estatísticos de interesse forense e encontram-se descritos na Tabela II.

Tabela II: Parâmetros estatísticos de interesse forense.

(Adaptado de Butler, 2005 e Goodwin *et al.*, 2011h; Huston, 1998)

PARÂMETRO ESTATÍSTICO	DESCRIÇÃO	FÓRMULA MATEMÁTICA
Probabilidade de identidade ou <i>matching</i> (PM)	Probabilidade de dois indivíduos escolhidos aleatoriamente apresentarem genótipos idênticos num determinado <i>locus</i> .	$PM = \sum_{k=1}^m p_k^2$ <p>p_k representa a frequência do genótipo k m corresponde ao número total de genótipos</p>
Poder de discriminação (PD)	Probabilidade de dois indivíduos escolhidos aleatoriamente apresentarem genótipos diferentes num determinado <i>locus</i> .	$PD = 1 - PM$
Conteúdo de informação polimórfica (PIC)	Indica o nível polimórfico de determinado <i>locus</i> .	$PIC = 1 - \sum_{i=1}^n p_i^2 - \left(\sum_{i=1}^n p_i^2 \right)^2 + \sum_{i=1}^n p_i^4$ <p>p_i representa a frequência do alelo i n corresponde ao número total de alelos</p>



PARÂMETRO ESTATÍSTICO	DESCRIÇÃO	FÓRMULA MATEMÁTICA
Poder de exclusão (PE)	Probabilidade de um <i>locus</i> excluir determinado indivíduo de uma paternidade.	$PE = h^2(1 - 2hH^2)$ H e h representam heterozigosidade e homozigosidade, respetivamente
Índice de paternidade (PI)	Probabilidade de um indivíduo, que está a ser estudado, ser o pai biológico por comparação com um indivíduo ao acaso.	$TPI = (h+H)/2h$ H e h representam heterozigosidade e homozigosidade, respetivamente

Para a avaliação do conjunto de marcadores estudados são utilizados, o poder de exclusão combinado (PE_{Comb}), o poder de discriminação combinado (PD_{Comb}), a probabilidade de identidade combinada (PM_{Comb}) e o índice de paternidade combinado (PI_{Comb}), que são calculados de acordo com as fórmulas abaixo referidas, nas quais PE_i , PD_i , PM_i e PI_i representam a probabilidade de exclusão, o poder de discriminação, a probabilidade de identidade e o índice de paternidade para o *locus* i , respetivamente, o L diz respeito ao número de *loci* e o símbolo Π significa multiplicação (Butler, 2005 e Goodwin *et al.*, 2011h; Huston, 1998).

$$PM_{Comb} = \prod_{l=1}^L PM_i$$

$$PD_{Comb} = 1 - \prod_{l=1}^L (1 - PD_i)$$

$$PE_{Comb} = 1 - \prod_{l=1}^L (1 - PE_i)$$

$$PI_{Comb} = \prod_{l=1}^L PI_i$$



1.7 OBJETIVOS DO ESTUDO

O principal objetivo deste projeto é a caracterização genética das populações imigrantes angolana e guineense, com residência no sul de Portugal, visto serem das nacionalidades estrangeiras mais representadas no país, por existirem laços históricos e linguísticos entre Portugal, Angola e Guiné-Bissau e devido à inexistência de meios suficientes para a realização deste tipo de estudos nos países mencionados.

Como objetivos específicos são identificados: o cálculo das frequências alélicas dos 21 STRs autossômicos amplificados pelo *GlobalFiler™ PCR Amplification Kit*; a determinação dos parâmetros estatísticos populacionais e de interesse forense, tais como, o Equilíbrio de Hardy-Weinberg, a heterozigotia esperada, a heterozigotia observada, o *Linkage Disequilibrium*, a probabilidade de identificação ou *matching* (PM), o poder de discriminação, o conteúdo de informação polimórfica, o poder de exclusão e o índice de paternidade; e a comparação entre as frequências alélicas das populações angolana e guineense e as de outras populações já estudadas, inclusive a população portuguesa, para avaliar a distância genética entre as mesmas e entender qual a contribuição genética dos imigrantes angolanos e guineenses na população nativa do Sul de Portugal.



2.1 AMOSTRAGEM

Para a realização deste estudo foram selecionadas, ao acaso, 223 amostras de manchas de sangue, obtidas sob consentimento informado de indivíduos que recorreram ao Serviço de Genética e Biologia Forenses da Delegação do Sul (SGBF-S) do Instituto de Medicina Legal e Ciências Forenses, I.P. (INMLCF) para a realização de perícias de investigação biológica de parentesco. Estas corresponderam a casos estudados no SGBF-S, entre 2000 e 2013, em que os intervenientes, residentes no sul de Portugal, poderiam ou não ser oriundos de Angola e Guiné-Bissau, mas possuíam ascendência materna e paterna proveniente destes países. Os partícipes abrangiam várias faixas etárias e ambos os sexos sendo aparentemente saudáveis e não aparentados. Para impossibilitar a identificação dos mesmos, todas as amostras utilizadas foram anonimizadas.

Das 223 manchas de sangue estudadas, 153 pertenciam a indivíduos angolanos (77 do sexo feminino e 76 do sexo masculino) e 70 a indivíduos guineenses (37 do sexo feminino, 33 do sexo masculino).

Neste caso, o termo Sul de Portugal, refere-se à área de atuação da Delegação do Sul do INMLCF, a qual está delimitada a vermelho na Figura 14.



Figura 14: Mapa de Portugal. Área de atuação da Delegação do Sul do Instituto Nacional de Medicina Legal e Ciências Forenses, I.P. (Adaptado de Carvalho, 2015)



2.2 COLHEITA E ARMAZENAMENTO DAS AMOSTRAS

As manchas de sangue foram colhidas em cartões *Whatman™* (GE Healthcare), secas à temperatura ambiente durante cerca de 24 horas e, posteriormente, armazenadas em envelopes de papel, devidamente identificados.

2.3 EXTRAÇÃO DE DNA PELO MÉTODO *CHELEX®* 100

Procedeu-se ao corte de fragmentos circulares, com cerca de 3 mm de diâmetro, de cada uma das amostras de mancha de sangue, os quais foram colocados em tubos de 1,5 mL devidamente identificados, e incubados em 1000 µL de água *Milli-Q*, à temperatura ambiente durante 15 minutos.

Após a incubação, realizou-se a agitação dos tubos no *vórtex* a baixa rotação, durante 10 segundos, e a sua centrifugação a 14 000 rpm durante 3 minutos, a 20°C, numa centrífuga Eppendorf 5804R. Imediatamente a seguir, retiraram-se cerca de 980 µL de sobrenadante, sem perturbar o *pellet* e adicionaram-se 180 µL da solução *Chelex®* 100 (*Bio-Rad*) a 5%.

Posteriormente, incubaram-se os tubos num banho de água a 56°C, durante 15 minutos, os quais foram depois agitados no *vórtex*, a alta velocidade, durante 15 segundos e de seguida, colocados em água em ebulição, durante 8 minutos. Por fim, realizou-se a agitação das amostras no *vórtex*, a alta velocidade, durante alguns segundos, e a sua centrifugação, a 14 000 rpm, durante 5 minutos, a 20°C.

É de salientar, que para este procedimento foi preparado um controlo negativo nas mesmas condições que as restantes amostras, com a exceção da adição do fragmento circular de mancha de sangue.

A partir deste último passo, procedeu-se à amplificação das amostras ou armazenaram-se os tubos a uma temperatura de -20°C para prevenir a ação das nucleases.

2.4 AMPLIFICAÇÃO DOS MARCADORES GENÉTICOS

Realizou-se a amplificação por PCR dos marcadores genéticos STR, recorrendo ao *GlobalFiler™ PCR Amplification Kit* (*Applied Biosystems*), de acordo as indicações do fabricante descritas no respetivo manual de utilização (*Thermo Fisher Scientific Inc.*, 2015). No entanto, o volume total da reação de amplificação



foi reduzido para metade do indicado no manual, passando de 25 μL para 12,5 μL . Esta redução foi obtida diminuindo para metade o volume utilizado de cada um dos reagentes do respetivo *kit*, mantendo as suas concentrações.

Na Tabela III encontram-se os componentes e respetivos volumes utilizados para a amplificação dos marcadores genéticos para cada uma das amostras.

Tabela III: Reagentes e respetivos volumes utilizados para a amplificação dos *loci* STR de cada uma das amostras.

REAGENTES	VOLUME (μL)
Água	6,5
<i>GlobalFiler™ Master Mix</i>	3,75
<i>GlobalFiler™ Primer Set</i>	1,25
DNA	1
TOTAL	12,5

Primeiramente, preparou-se uma *master mix*, a qual continha a quantidade de cada um dos reagentes, expeto o DNA, necessários para amplificar $n+1$ amostras, sendo n o número de amostras a amplificar. De seguida, adicionaram-se 11,5 μL da *master mix* mais 1 μL de DNA, de cada amostra, a cada um dos tubos de reação de 0,2 mL, perfazendo um volume total, por amostra, de 12,5 μL . Os controlos positivo e negativo foram preparados segundo a Tabela IV. Todos estes passos foram realizados na sala pré-PCR, numa câmara de fluxo laminar previamente descontaminada com radiação ultravioleta, durante 15 minutos.

Tabela IV: Composição dos controlos positivo e negativo.

CONTROLO POSITIVO	CONTROLO NEGATIVO
4,5 μL de Água	7,5 μL de Água
3,75 μL de <i>GlobalFiler™ Master Mix</i>	3,75 μL de <i>GlobalFiler™ Master Mix</i>
1,25 μL de <i>GlobalFiler™ Primer Set</i>	1,25 μL de <i>GlobalFiler™ Primer Set</i>
3 μL de <i>DNA Control 007</i>	-----

Finalmente, colocaram-se os tubos num termociclador *GeneAmp® PCR System 9700* (*Applied Biosystems*), o qual foi programado de acordo com as condições apresentadas na Tabela V. Após a amplificação, as amostras foram



armazenadas entre -30°C e -15°C ou submetidas, logo de seguida, à eletroforese capilar.

Tabela V: Condições da amplificação por PCR do *GlobalFiler™ PCR Amplification Kit* (Applied Biosystems).

INCUBAÇÃO INICIAL	29 CICLOS		EXTENSÃO FINAL	
	DESNATURAÇÃO	ANNEALING / EXTENSÃO		
95°C (1 minuto)	94°C (10 segundos)	59°C (90 segundos)	60°C (10 minutos)	4°C ∞

2.5 ELETROFORESE CAPILAR

Inicialmente, preparou-se uma *mix* que continha 9,6 µL de formamida Hi-Di™ (*Applied Biosystems*) e 0,4 µL do marcador interno *GeneScan™ 600 LIZ™ Size Standard v2.0* por cada *n* amostras analisadas, sendo *n* o número de amostras a analisar. No entanto, com a intenção de compensar eventuais erros de pipetagem, a *mix* foi feita para *n+1* amostras, a qual foi distribuída, em volumes de 10 µL, por cada poço de uma placa *MicroAmp® Optical* (*Applied Biosystems*) de 96 poços. Após a distribuição da *mix*, adicionou-se 1 µL de cada produto amplificado ou 1 µL do *GlobalFiler™ Allelic Ladder* (um *ladder* alélico por cada 15 amostras) aos poços correspondentes. Seguidamente, as amostras foram desnaturadas, num termociclador *GeneAmp PCR System 9700* (*Applied Biosystems*), a 95°C, durante 3 minutos. Decorrido este tempo, colocou-se a placa no sequenciador automático *Applied Biosystems Genetic Analyzer 3130xl*, no qual ocorreu a eletroforese capilar de acordo com as condições descritas pelo fabricante do *GlobalFiler™ PCR Amplification Kit* (*Applied Biosystems*).

Por fim, os resultados da eletroforese foram obtidos sobre a forma de eletroforegrama e analisados com recurso ao *software GeneMapper® ID-X* versão 1.4 (*Applied Biosystems*).

2.6 ESTUDO DAS POPULAÇÕES

2.6.1 ANÁLISE ESTATÍSTICA DOS PARÂMETROS POPULACIONAIS

A análise dos parâmetros populacionais, como o cálculo das frequências alélicas, a heterozigotia esperada e observada, o equilíbrio de Hardy-Weinberg e o



linkage disequilibrium, foi realizada recorrendo ao *software* Arlequin versão 3.5.2.2 (Excoffier *et al.*, 2010b).

Com o auxílio desta ferramenta, também foi possível efetuar o teste exato de diferenciação populacional entre as populações angolana e guineense residentes no sul de Portugal e as populações, já estudadas para os 21 marcadores genéticos, do Sul de Portugal (Almeida *et al.*, 2015), de Espanha (García *et al.*, 2012), da Holanda (Westen *et al.*, 2014), dos EUA (indivíduos caucasianos e africanos) (Hill *et al.*, 2013), da África do Sul (Ristow *et al.*, 2016), do Líbano (Andari *et al.*, 2013), dos Emirados Árabes Unidos (Alhmoudi *et al.*, 2015), do Afeganistão (Älgenäs *et al.*, 2014), da China (Zhang *et al.*, 2016), do México (Ramos-González *et al.*, 2016) e da Coreia (Park *et al.*, 2016).

As distâncias genéticas entre as populações, anteriormente, referidas foram calculadas através do *software* PHYLIP versão 3.695 (Felsenstein, 2005). Para a representação destes dados, recorreu-se ao *software* *Molecular Evolutionary Genetics Analysis* (MEGA) versão 6 (Tamura *et al.*, 2013) e ao *Statistical Package for the Social Sciences* (SPSS) versão 20 (SPSS, 2011), através dos modelos *neighbor joining* (NJ) e escala multidimensional (MDS), respetivamente.

2.6.2 HETEROZIGOSIDADE, HOMOZIGOSIDADE E ANÁLISE ESTATÍSTICA DOS PARÂMETROS DE INTERESSE FORENSE

A heterozigosidade, a homozigosidade e os parâmetros de interesse forense, como a probabilidade de identidade ou *matching*, o poder de discriminação, o poder de exclusão, o conteúdo de informação polimórfica e o índice de paternidade, foram determinados através do *software* PowerStats versão 1.2 (Tereba, 1999).



3.1 ESTUDO DA POPULAÇÃO ANGOLANA RESIDENTE NO SUL DE PORTUGAL

Do total de amostras analisadas para a população angolana residente no sul de Portugal, uma delas, correspondente a um indivíduo do sexo masculino, não foi incluída na análise estatística por apresentar um padrão trialélico do tipo I no marcador genético D8S1179, o qual era constituído pelos alelos 14, 16 e 17. Este padrão encontra-se descrito na *STRbase* pela *International Commission on Missing Persons, Sarajevo, Bosnia and Herzegovina* (Short Tandem Repeat DNA Internet DataBase, 1997a).

Também se pôde observar a presença de alelos raros, os quais não se encontram no *ladder* alélico, em três marcadores distintos e em diferentes indivíduos. Os alelos 26.3, 14.1 e 23.3 foram encontrados nos marcadores SE33, D2S441 e D21S11, respetivamente, sendo que os dois primeiros foram observados em dois indivíduos diferentes do sexo feminino e o terceiro num indivíduo do sexo masculino. Apenas as microvariantes alélicas 14.1 do marcador D2S441 e 23.3 do marcador D21S11 aparecem descritas na *STRbase* pelo *Belgium National Institute of Criminalistics and Criminology* (Short Tandem Repeat DNA Internet DataBase, 1997b) e pela Delegação do Sul do Instituto de Medicina Legal e Ciências Forense (Short Tandem Repeat DNA Internet DataBase, 1997c), respetivamente.

3.1.1 ANÁLISE ESTATÍSTICA DOS PARÂMETROS POPULACIONAIS

As frequências alélicas calculadas para os 21 marcadores genéticos STR estão representadas na Tabela VI enquanto os restantes parâmetros populacionais se encontram na Tabela VII.

3. RESULTADOS E DISCUSSÃO



Tabela VI: Frequências alélicas dos 21 STRs autossômicos na população angolana residente no sul de Portugal.

ALELOS	D3S1358	VWA	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	D2S441	D19S433	TH01	FGA	D22S1045	D5S818	D13S317	D7S820	SE33	D10S1248	D151656	D12S391	D2S1338
5			0,0033																		
6				0,0033	0,0888						0,1086					0,0099	0,0033				
7				0,0263	0,0362						0,3388				0,0329	0,25					
8			0,0132	0,0526	0,3388	0,0033					0,2862			0,0691	0,0296	0,1086					
9	0,0033		0,2171	0,0329	0,1645	0,0033		0,0033			0,1776			0,0329							
9.3											0,0855				0,0263	0,2961		0,0033	0,0132		
10			0,1349	0,2895	0,0888	0,0066		0,0033	0,0954	0,023	0,0033		0,0559	0,0428							
10.2								0,0066							0,3421	0,2105		0,0395	0,0757		
11		0,0329	0,3651	0,25	0,2632	0,0559		0,0099	0,3684	0,0691			0,1776	0,1941			0,0099				
11.2																					
11.3									0,0263						0,352	0,1086		0,2237	0,0822		
12	0,0066		0,1447	0,2829	0,0197	0,1447		0,0362	0,1743	0,1151			0,0658	0,375		0,0033					
12.1																					
12.2										0,0395					0,1546	0,0132	0,0099	0,2434	0,0757		
13	0,0066	0,0066	0,1086	0,0493		0,1941		0,0395	0,0461	0,2993				0,273			0,0033				
13.2										0,0691											
13.3									0,0033						0,0625		0,0461	0,2434	0,2434		
14	0,0691	0,0822	0,0132	0,0132		0,3454		0,0559	0,2467	0,1809			0,023	0,0099							
14.1									0,0033								0,0132				
14.2								0,0066		0,0691									0,023		
14.3																	0,0493	0,148	0,1974	0,1118	0,0066
15	0,3224	0,1842				0,1974		0,2039	0,0362	0,0658			0,2467	0,0033			0,0033				
15.2	0,0033									0,023									0,0099		
15.3																	0,0691	0,0691	0,1151	0,0855	0,0691
16	0,2829	0,2401				0,0428		0,1875		0,0197			0,1743				0,0033				
16.2										0,023									0,0954		
16.3																	0,0592	0,023	0,0033	0,1283	0,0954
17	0,2401	0,2039				0,0066		0,1579		0,0033			0,2368						0,0362	0,0033	
17.3																	0,1415	0,0066		0,2566	0,0691
18	0,0592	0,1579						0,1316				0,0099	0,0197								
18.2												0,0066							0,0263	0,0033	
18.3																	0,1612			0,1184	0,1743
19	0,0066	0,0822						0,0921				0,0822							0,0033	0,0099	
19.3																	0,0724			0,125	0,0789
20		0,0099						0,0263				0,0461								0,0033	
20.1																	0,0263			0,0691	0,1447

3. RESULTADOS E DISCUSSÃO



ALELOS	D3S1358	VWA	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	D2S441	D19S433	TH01	FGA	D22S1045	D5S818	D13S317	D7S820	SE33	D10S1248	D1S1656	D12S391	D2S1338
21								0,0329				0,0987					0,0132				
21.2								0,0033									0,0395			0,0395	0,1283
22												0,1711					0,0033				
22.2												0,0033								0,0296	0,0625
23								0,0033				0,1776					0,0066				0,0033
23.2																					
23.3							0,0066													0,0164	0,0724
24												0,1546					0,0164				
24.2																					
24.3							0,0033														0,0625
25												0,102					0,0362				
25.2																					0,023
26							0,0099					0,0461					0,0559				
26.2																	0,0033				
26.3																					0,0099
27							0,0757					0,0263					0,0395				
27.2																					
28							0,2434					0,0164					0,0559				
28.2																					
29							0,1546					0,0164					0,0197				
29.2																					
30							0,1711					0,0066					0,0197				
30.2							0,023					0,0066									
31							0,0855										0,0099				
31.2							0,0724					0,0132									
32							0,0099										0,0099				
32.2							0,0428														
33							0,0132														
33.2							0,0329														
34							0,0164														
34.2							0,0099														
35							0,0164														
36							0,0066														
37							0,0066														
42.2												0,0033									
43.2												0,0033									
44.2												0,0099									



Tabela VII: Resultados obtidos para os parâmetros estatísticos relativos à população angolana residente no sul de Portugal.

MARCADOR	HETEROZIGOTIA OBSERVADA	HETEROZIGOTIA ESPERADA	HWE (valores de ρ)
D3S1358	0,80263	0,75243	0,9174
vWa	0,78947	0,82986	0,63819
D16S539	0,78289	0,7708	0,64766
CSF1PO	0,81579	0,76904	0,54736
TPOX	0,77632	0,77397	0,50305
D8S1179	0,79605	0,78064	0,48229
D21S11	0,875	0,86757	0,47956
D18S51	0,88816	0,86738	0,50897
D2S441	0,76974	0,76227	0,69744
D19S433	0,86184	0,84499	0,59528
TH01	0,77632	0,75512	0,96524
FGA	0,86184	0,88529	0,23054
D22S1045	0,86184	0,8154	0,01131
D5S818	0,71053	0,74181	0,17013
D13S317	0,67105	0,73102	0,42798
D7S820	0,75658	0,78426	0,83577
SE33	0,94737	0,92572	0,15874
D10S1248	0,78289	0,80528	0,45422
D1S1656	0,86184	0,86123	0,57179
D12S391	0,90132	0,86347	0,01104
D2S1338	0,85526	0,89654	0,13634

Como é possível observar na Tabela VII, apenas os marcadores D22S1045 e D12S391 apresentam valores de ρ inferiores ao nível de significância 0,05, os quais se encontram representados a azul. A estes marcadores foi aplicada a correção de Bonferroni (Weir, 1996), a qual deve ser usada quando se executam simultaneamente vários testes estatísticos independentes sendo calculada pela aplicação da expressão $0,05/m$, onde m é o número de comparações efetuadas. Assim sendo, o nível de significância passou a ser 0,00238 (0,05/21). Visto que os valores de ρ dos dois marcadores são superiores ao nível de significância corrigido pode-se afirmar que esta população se encontra em equilíbrio de Hardy-Weinberg.

Na mesma tabela, também estão disponíveis os valores de heterozigotia observada e de heterozigotia esperada. Os valores de H_o situam-se entre 0,67105 e 0,94737 e os de H_e entre 0,73102 e 0,92572, correspondendo os valores mais baixos ao marcador D13S317 e os mais elevados ao SE33. Através da análise destes parâmetros conclui-se que a população angolana residente no sul de Portugal é maioritariamente heterozigótica para todos os marcadores estudados.

3. RESULTADOS E DISCUSSÃO



Relativamente aos resultados obtidos para o *linkage disequilibrium*, representados na Tabela VIII, observam-se valores de ρ inferiores a 0,05 para 19 pares de *loci*. No entanto, aplicando a correção de Bonferroni (Weir, 1996) para $m=210$, o nível de significância diminui para 0,000238 e os valores de ρ para todos os pares de *loci* passam a ser não significativos, isto é, não se observa nenhum valor de ρ inferior a 0,000238. Posto isto, infere-se que todos os pares de *loci* se encontram em *linkage equilibrium*, ou seja, todos os marcadores analisados são herdados independentemente, tal como esperado pelo fato dos marcadores serem independentes.

Tabela VIII: Tabela representativa dos valores de significância para o *linkage disequilibrium*. O sinal (-) designa “não significativo” para valores de ρ superiores a 0,05, o sinal (+) designa “significativo” para valores de ρ inferiores a 0,05 e o sinal (+*) designa “não significativo” para valores de ρ superiores a 0,000238.

MARCADOR	D3S1358	vWa	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	D2S441	D19S433	TH01	FGA	D22S1045	D5S818	D13S317	D7S820	SE33	D10S1248	D1S1656	D12S391	D2S1338
D3S1358																					
vWa	-																				
D16S539	-	-																			
CSF1PO	-	-	-																		
TPOX	-	-	-	-																	
D8S1179	-	-	-	-	-																
D21S11	-	-	-	+*	-	-															
D18S51	-	-	-	-	-	-	-														
D2S441	-	-	-	-	-	-	-	+*													
D19S433	-	-	-	-	-	-	-	-	-												
TH01	-	-	-	-	-	+*	-	+*	+*	-											
FGA	+*	-	-	-	-	+*	-	-	-	-	-										
D22S1045	-	-	-	-	-	-	-	-	-	-	-	-									
D5S818	-	-	-	-	-	-	-	+*	-	-	-	-	-								
D13S317	-	-	-	-	-	-	-	-	-	-	-	-	-	-							
D7S820	-	+*	-	-	-	-	-	-	-	-	-	-	-	-	+*						
SE33	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-					
D10S1248	-	-	-	-	-	-	-	-	-	-	-	+*	-	-	-	-	+*				
D1S1656	-	-	-	-	-	+*	-	-	-	-	-	-	-	-	+*	+*	+*	-			
D12S391	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+*	+*	-	-		
D2S1338	-	-	-	-	-	-	-	-	+*	-	-	-	-	-	-	-	-	-	-	-	

3.1.2 HETEROZIGOSIDADE, HOMOZIGOSIDADE E ANÁLISE ESTATÍSTICA DOS PARÂMETROS DE INTERESSE FORENSE

Aquando da análise da Tabela IX, é possível observar que a probabilidade de identidade ou *matching* varia entre 0,0177 e 0,1163, correspondendo o valor mais baixo ao marcador SE33 e o mais elevado ao marcador D5S818. Estes valores indicam que é mais provável que dois indivíduos, selecionados ao acaso, apresentem o mesmo genótipo para o marcador D5S818 do que para o SE33. A probabilidade de identidade combinada, representada na Tabela X, foi calculada segundo a fórmula demonstrada na seção 1.6.1.6, pela qual se obteve o resultado $3,8093 \times 10^{-26}$. Assim, pode dizer-se que o fenómeno em que dois indivíduos, pertencentes à população estudada e escolhidos aleatoriamente, possuem o mesmo genótipo para todos os 21 STRs autossómicos analisados é muito raro, o qual pode ocorrer num em cada $3,8093 \times 10^{26}$ indivíduos.

[illegible]

Na análise do conteúdo de informação polimórfica, também presente na Tabela IX, observa-se que o STR autossômico SE33 é o que apresenta maior nível de polimorfismo, sendo CIP igual a 0,9179. Todavia, todos os outros marcadores exibem valores superiores a 0,6846, o qual corresponde ao D13S317.

Os valores mínimos observados para o poder de exclusão e para o índice de paternidade (Tabela IX) correspondem ao marcador D13S317, sendo 0,3849 e 1,52, respectivamente. Por outro lado, os resultados mais elevados para estes dois parâmetros de interesse forense pertencem ao SE33, sendo o poder de exclusão igual a 0,8928 e o índice de paternidade igual a 9,5. O poder de exclusão e o índice de paternidade combinados, os quais estão presentes na Tabela X, indicam que,

3. RESULTADOS E DISCUSSÃO



para o conjunto dos 21 marcadores presentes no *GlobalFiler™ PCR Amplification Kit* (Applied Biosystems), a probabilidade de exclusão de um indivíduo de uma paternidade é de 99,9999999837329% e que é $6,3377 \times 10^9$ vezes mais provável que um indivíduo estudado seja o pai biológico em comparação com uma pessoa ao acaso.

Também foram determinadas as proporções de hetero e homozigossidade (Tabela IX), sendo o marcador SE33 o mais heterozigótico ($H=0,9474$) e o D13S317 o mais homozigótico ($h=0,3289$).

Tabela IX: Parâmetros estatísticos de interesse forense, número de alelos e genótipos, heterozigossidade e homozigossidade para cada um dos 21 STRs autossômicos estudados para a população angolana residente no sul de Portugal.

	NÚMERO DE ALELOS	NÚMERO DE GENÓTIPOS	HOMOZIGOSIDADE	HETEROZIGOSIDADE	PROBABILIDADE DE IDENTIDADE	PODER DE DISCRIMINAÇÃO	CONTEÚDO DE INFORMAÇÃO POLIMÓRFICA	PODER DE EXCLUSÃO	ÍNDICE DE PATERNIDADE
D3S1358	10	21	0,1974	0,8026	0,1144	0,8856	0,708	0,6039	2,5333
vWa	9	30	0,2105	0,7895	0,0547	0,9453	0,8042	0,5797	2,375
D16S539	8	22	0,2171	0,7829	0,0935	0,9065	0,7355	0,5677	2,303
CSF1PO	9	26	0,1842	0,8158	0,1046	0,8954	0,7293	0,6287	2,7143
TPOX	7	23	0,2237	0,7763	0,0916	0,9084	0,738	0,5558	2,2353
D8S1179	10	24	0,2039	0,7961	0,0871	0,9129	0,7464	0,5917	2,4516
D21S11	19	58	0,125	0,875	0,038	0,962	0,8515	0,7447	4
D18S51	17	52	0,1118	0,8882	0,0397	0,9603	0,8501	0,7713	4,4706
D2S441	9	23	0,2303	0,7697	0,0901	0,9099	0,7252	0,5441	2,1714
D19S433	13	48	0,1382	0,8618	0,0444	0,9556	0,8267	0,7183	3,619
TH01	6	15	0,2237	0,7763	0,1035	0,8965	0,7125	0,5558	2,2353
FGA	20	56	0,1382	0,8618	0,0287	0,9713	0,8712	0,7183	3,619
D22S1045	8	27	0,1382	0,8618	0,0807	0,9193	0,7864	0,7183	3,619
D5S818	8	22	0,2895	0,7105	0,1163	0,8837	0,6982	0,4447	1,7273
D13S317	7	14	0,3289	0,6711	0,1055	0,8945	0,6846	0,3849	1,52
D7S820	8	20	0,2434	0,7566	0,0785	0,9215	0,7479	0,5211	2,0541
SE33	29	89	0,0526	0,9474	0,0177	0,9823	0,9179	0,8928	9,5
D10S1248	9	28	0,2171	0,7829	0,0724	0,9276	0,7737	0,5677	2,303
D1S1656	14	48	0,1382	0,8618	0,0395	0,9605	0,8438	0,7183	3,619
D12S391	14	43	0,0987	0,9013	0,0421	0,9579	0,8465	0,7981	5,0667
D2S1338	14	60	0,1447	0,8553	0,0263	0,9737	0,8841	0,7053	3,4545

[illegible]

Conforme os resultados apresentados neste estudo, após a aplicação da correção de Bonferroni (Weir, 1996), nenhum dos loci estudados demonstrou desvios ao equilíbrio de Hardy-Weinberg e o marcador mais informativo para a população angolana imigrante é o SE33 visto que possui maior Ho, PD, CIP, PE e PI e menor PM. Contrariamente, o marcador D5S818 apresenta o valor mais baixo, comparando com os restantes marcadores, para o poder de discriminação, e o mais elevado para a probabilidade de identidade todavia, é o D13S317 que exhibe menor heterozigotia esperada, conteúdo de informação polimórfica, poder de exclusão e índice de paternidade.

Caracterização dos *loci* GlobalFiler nas populações Angolana e Guineense, residentes no sul de Portugal | 50



linkage disequilibrium. Relativamente ao HWE, os dados obtidos no presente estudo são concordantes com os obtidos por Santos *et al.*, Beleza *et al.* e Melo *et al.* Todavia, no trabalho relativo à população angolana residente em Lisboa o TPOX demonstra ser o STR menos informativo. Tal resultado mostra-se discordante com os apresentados neste projeto e nos trabalhos realizados para a população angolana nativa. Apesar de, no presente trabalho, o *locus* D13S317 demonstrar-se como um dos menos informativos não é o marcador que exibe menor poder de discriminação, como referido por Beleza *et al.* e Melo *et al.*, mas sim o *locus* D5S818. Assim sendo, será necessário, no futuro, a realização de mais estudos acerca desta população, nos quais o número de indivíduos deverá ser aumentado e o número de *loci* analisados mantido.



3.2 ESTUDO DA POPULAÇÃO GUINEENSE RESIDENTE NO SUL DE PORTUGAL

Durante a análise dos eletroforegramas, obtidos a partir das amostras estudadas para a população guineense residente no sul de Portugal, foi encontrado o alelo raro 17.1 no marcador D18S51 num indivíduo do sexo masculino, o qual está descrito na *STRbase* pelo *Orchid Cellmark Inc Nashville* e pelo *Swedish National Forensic Centre* (Short Tandem Repeat DNA Internet DataBase, 1997d).

3.2.1 ANÁLISE ESTATÍSTICA DOS PARÂMETROS POPULACIONAIS

As frequências alélicas calculadas para os 21 marcadores genéticos STR estão representadas na Tabela XI enquanto os restantes parâmetros populacionais se encontram na Tabela XII.

3. RESULTADOS E DISCUSSÃO



Tabela XI: Frequências alélicas dos 21 STRs autossômicos na população guineense residente no sul de Portugal.

ALELOS	D3S1358	VWA	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	D2S441	D19S433	TH01	FGA	D22S1045	D5S818	D13S317	D7S820	SE33	D10S1248	D1S1656	D12S391	D2S1338
5			0,0071																		
6					0,1214						0,15										
7				0,0643	0,0357						0,2357					0,0143					
8			0,0357	0,0857	0,1857						0,2286			0,0929	0,0286	0,2214					
9	0,0071		0,1929	0,0357	0,2857			0,0071			0,2571			0,05	0,0071	0,0786					
9.3											0,1071										
10			0,1643	0,2071	0,0786	0,0286			0,1214	0,0143	0,0214		0,0214	0,0571	0,0071	0,3786			0,0071		
10.2																	0,0071				
11			0,2571	0,25	0,2714	0,0286			0,4	0,0786			0,1357	0,1214	0,3571	0,1714		0,0429	0,0214		
11.2																	0,0071				
11.3									0,0214												
12	0,0143		0,1786	0,2643	0,0214	0,0643		0,1	0,1429	0,0929			0,0143	0,5286	0,4357	0,1286		0,1214	0,0786		
12.2										0,0214											
13	0,0357	0,0071	0,1357	0,0929		0,2143		0,0929	0,0929	0,3			0,0071	0,1357	0,1357	0,0071	0,0071	0,2571	0,1286		
13.2										0,1							0,0286				
14	0,1214	0,1	0,0286			0,3214		0,0643	0,2	0,1857			0,0714	0,0071	0,0214		0,0571	0,3143	0,25		
14.2								0,0143		0,1											
15	0,2857	0,1929				0,25		0,1286	0,0214	0,0286			0,1929	0,0071	0,0071		0,0429	0,1286	0,2214	0,0357	
15.2										0,0714											
15.3																			0,0071		
16	0,3357	0,2071				0,0786		0,1786					0,1429				0,0643	0,1071	0,1143	0,05	0,0214
16.2										0,0071							0,0214				
16.3																			0,1		
17	0,1571	0,2429				0,0143		0,1857					0,3929				0,0643	0,0214	0,0071	0,1857	0,0643
17.1								0,0071													
17.2																	0,0071				
17.3																			0,0571	0,0214	
18	0,0286	0,1643						0,0929					0,0214				0,1571	0,0071		0,2714	0,05
18.1																			0,0071		
18.2												0,0071									
18.3																			0,0071	0,0143	
19	0,0143	0,0643						0,0857				0,0571					0,1214			0,1857	0,2
19.3																			0,0071		
19.4																					
20		0,0214						0,0357				0,0714					0,0786			0,0929	0,0571
20.2																	0,0071				

3. RESULTADOS E DISCUSSÃO



ALELOS	D3S1358	VWA	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	D2S441	D19S433	TH01	FGA	D22S1045	D5S818	D13S317	D7S820	SE33	D10S1248	D1S1656	D12S391	D2S1338
21												0,1					0,0286			0,0571	0,0786
21.2																	0,0214				
22												0,1643					0,0214			0,0286	0,1429
23								0,0071				0,2071								0,0286	0,1071
23.2																	0,0214				
23.3												0,0143									
24												0,1929								0,0071	0,1286
24.2							0,0071										0,0286				
25												0,0643								0,0071	0,0929
25.2																	0,0214				
26							0,0143					0,05									0,05
26.2																	0,05				
27							0,05					0,0286									0,0071
27.2																	0,0357				
28							0,1643					0,0143									
28.2																	0,05				
29							0,1857					0,0071					0,0071				
29.2																	0,0214				
30							0,1857					0,0071									
30.2							0,0071														
31							0,0786														
31.2							0,0571					0,0071									
32							0,0429														
32.2							0,0857										0,0214				
33.2							0,0571														
34.2							0,0071														
35							0,0429														
35.1							0,0143														
42.2												0,0071									



Tabela XII: Resultados obtidos para os parâmetros estatísticos populacionais relativos à população guineense residente no sul de Portugal.

MARCADOR	HETEROZIGOSIDADE OBSERVADA	HETEROZIGOSIDADE ESPERADA	HWE (valor de p)
D3S1358	0,78571	0,76917	0,91984
vWa	0,81429	0,82518	0,0037
D16S539	0,82857	0,82312	0,77915
CSF1PO	0,74286	0,80915	0,728
TPOX	0,78571	0,79322	0,99392
D8S1179	0,78571	0,78171	0,66276
D21S11	0,91429	0,88356	0,78271
D18S51	0,75714	0,88304	0,05811
D2S441	0,77143	0,76074	0,99695
D19S433	0,82857	0,84008	0,87163
TH01	0,81429	0,79733	0,28922
FGA	0,81429	0,87266	0,44536
D22S1045	0,75714	0,76886	0,01644
D5S818	0,64286	0,6778	0,53523
D13S317	0,6	0,66752	0,6703
D7S820	0,72857	0,76074	0,269
SE33	0,88571	0,93546	0,58334
D10S1248	0,74286	0,79568	0,41568
D1S1656	0,88571	0,84481	0,75491
D12S391	0,82857	0,84522	0,88998
D2S1338	0,9	0,89024	0,65128

De acordo com os dados apresentados na Tabela XII, apenas os marcadores vWa e D22S1045 apresentam valores de p inferiores ao nível de significância 0,05, os quais se encontram representados a azul. A estes foi aplicada a correção de Bonferroni (Weir, 1996), com a qual o nível de significância passou a ser 0,00238 (0,05/21). Uma vez que os valores de p para os dois marcadores referidos anteriormente são superiores a 0,00238 pode dizer-se que esta população se encontra em equilíbrio de Hardy-Weinberg.

Também foram calculadas a heterozigotia observada e a heterozigotia esperada para cada um dos marcadores estudados, as quais são demonstradas na Tabela XII. A H_o apresenta valores entre os 0,6 e 0,91429 e a H_e entre 0,66752 e 0,93546, correspondendo os valores mais baixos ao marcador D13S317 e os mais elevados ao D21S11 e ao SE33, respetivamente. A análise destes parâmetros permite concluir que, tal como a população angolana residente no sul de Portugal, a população guineense também residente no sul de Portugal, é maioritariamente heterozigótica para todos os 21 marcadores estudados.



Os resultados obtidos para o *linkage disequilibrium* podem ser vistos na Tabela XIII. Existem 11 pares de *loci* com valores de χ^2 inferiores a 0,05, todavia aplicando a correção de Bonferroni (Weir, 1996) para $m=210$, o nível de significância reduz para 0,000238. Com esta correção, todos os valores de χ^2 para estes pares de *loci* passam a ser não significativos, ou seja, todos os χ^2 são superiores a 0,000238. Posto isto, conclui-se que todos os pares de *loci* analisados se encontram em *linkage equilibrium*, isto é, todos os marcadores analisados são herdados independentemente.

Tabela XIII: Tabela representativa dos valores de significância para o *linkage disequilibrium*. O sinal (-) designa “não significativo” para valores de ρ superiores a 0,05, o sinal (+) designa “significativo” para valores de ρ inferiores a 0,05 e o sinal (+*) designa “não significativo” para valores de ρ superiores a 0,000238.

MARCADOR	D3S1358	vWa	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	D2S441	D19S433	TH01	FGA	D22S1045	D5S818	D13S317	D7S820	SE33	D10S1248	D1S1656	D12S391	D2S1338
D3S1358																					
vWa	-																				
D16S539	-	-																			
CSF1PO	+*	-	-																		
TPOX	-	-	-	+*																	
D8S1179	-	-	+*	-	-																
D21S11	-	-	-	-	-	-															
D18S51	-	-	+*	-	+*	-	-														
D2S441	-	-	-	-	-	-	-	-													
D19S433	-	+*	-	-	-	-	-	-	-												
TH01	-	-	-	-	-	-	-	-	-	-											
FGA	-	-	-	-	-	-	-	-	-	-	-										
D22S1045	-	-	-	-	-	-	-	-	-	-	-	-									
D5S818	-	-	-	+*	-	-	-	-	-	-	-	-	-								
D13S317	-	+*	-	+*	-	-	-	-	-	-	-	-	-	-							
D7S820	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-						
SE33	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-					
D10S1248	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-				
D1S1656	+*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
D12S391	-	+*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
D2S1338	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

3.2.2 HETEROZIGOSIDADE, HOMOZIGOSIDADE E ANÁLISE ESTATÍSTICA DOS PARÂMETROS DE INTERESSE FORENSE

Através da análise da Tabela XIV, observa-se que a probabilidade de identidade varia entre 0,0196 e 0,1673, correspondendo o valor mais baixo ao

O poder de discriminação (Tabela XIV), apresenta valores entre 0,8327 e 0,9804, que correspondem aos marcadores D13S317 e SE33, respetivamente. De acordo com estes dados, existe uma maior probabilidade de dois indivíduos, selecionados ao acaso, exibirem diferentes genótipos para o marcador SE33 do que para o D13S317. Para o conjunto dos 21 STRs, foi calculado o valor para o poder de discriminação combinado (Tabela XV), o qual é a igual a 0,9999999999999999999999997915. Assim, pode concluir-se que existe 99,99999999999999999999997915% de probabilidade de dois indivíduos, pertencentes à população guineense residente no sul de Portugal e selecionados aleatoriamente, possuírem diferentes genótipos para todos os marcadores estudados.

Para o poder de exclusão e para o índice de paternidade (Tabela XV) o marcador D13S317 é o que exibe valores mais baixos, 0,2909 e 1,25, respetivamente. Por outro lado, os resultados mais elevados para estes dois parâmetros pertencem ao D21S11, sendo o poder de exclusão igual a 0,8247 e o índice de paternidade igual a 5,8333. O poder de exclusão e o índice de paternidade combinados, os quais estão presentes na Tabela X, indicam que a probabilidade de exclusão de um indivíduo de uma paternidade é de 99,9999997871199% e que é $3,8809 \times 10^8$ vezes mais provável que um indivíduo estudado seja o pai biológico em comparação com uma pessoa ao acaso, analisando os 21 marcadores presentes no *GlobalFiler™ PCR Amplification Kit* (Applied Biosystems).

Caracterização dos *loci* GlobalFiler nas populações Angolana e Guineense, residentes no sul de Portugal | 58

maior número de indivíduos heterozigóticos ($H=0,9143$) e homozigóticos ($h=0,4$), respectivamente.

Tabela XIV: Parâmetros estatísticos de interesse forense, número de alelos e genótipos, heterozigosidade e homozigosidade para cada um dos 21 STRs autossômicos estudados para a população guineense residente no sul de Portugal.

	NÚMERO DE ALELOS	NÚMERO DE GENÓTIPOS	HOMOZIGOSIDADE	HETEROZIGOSIDADE	PROBABILIDADE DE IDENTIDADE	PODER DE DISCRIMINAÇÃO	CONTEÚDO DE INFORMAÇÃO POLIMÓRFICA	PODER DE EXCLUSÃO	ÍNDICE DE PATERNIDADE
D3S1358	9	19	0,2143	0,7857	0,0939	0,9061	0,728	0,5728	2,3333
vWa	8	20	0,1857	0,8143	0,0824	0,9176	0,7942	0,6258	2,6923
D16S539	8	23	0,1714	0,8286	0,0665	0,9335	0,7917	0,6531	2,9167
CSF1PO	7	23	0,2571	0,7429	0,0673	0,9327	0,7755	0,4976	1,9444
TPOX	7	20	0,2143	0,7857	0,0767	0,9233	0,756	0,5728	2,3333
D8S1179	8	19	0,2143	0,7857	0,091	0,909	0,7427	0,5728	2,3333
D21S11	15	41	0,0857	0,9143	0,0388	0,9612	0,8654	0,8247	5,8333
D18S51	13	38	0,2429	0,7571	0,0376	0,9624	0,8643	0,5221	2,0588
D2S441	7	18	0,2286	0,7714	0,0906	0,9094	0,7233	0,5471	2,1875
D19S433	11	31	0,1714	0,8286	0,0518	0,9482	0,8162	0,6531	2,9167
TH01	6	17	0,1857	0,8143	0,089	0,911	0,7590	0,6258	2,6923
FGA	16	38	0,1857	0,8143	0,0396	0,9604	0,8527	0,6258	2,6923
D22S1045	9	18	0,2429	0,7571	0,1016	0,8984	0,7334	0,5221	2,0588
D5S818	8	17	0,3571	0,6429	0,14	0,86	0,6447	0,3455	1,4
D13S317	8	14	0,4	0,6	0,1673	0,8327	0,6017	0,2909	1,25
D7S820	7	18	0,2714	0,7286	0,1049	0,8951	0,7195	0,4738	1,8421
SE33	26	59	0,1143	0,8857	0,0196	0,9804	0,9246	0,7663	4,375
D10S1248	8	22	0,2571	0,7429	0,0776	0,9224	0,7607	0,4976	1,9444
D1S1656	12	29	0,1143	0,8857	0,0567	0,9433	0,8197	0,7663	4,375
D12S391	15	36	0,1714	0,8286	0,049	0,951	0,8212	0,6531	2,9167
D2S1338	12	37	0,1	0,9	0,0339	0,9661	0,873	0,7954	5

Tabela XV: Parâmetros estatísticos de interesse forense para todos os STRs autossômicos estudados para a população guineense residente no sul de Portugal.

PARÂMETRO	VALOR
Probabilidade de identidade combinada	2,085x10 ⁻²⁵
Poder de discriminação combinado	0,99999999999999999999997915
Poder de exclusão combinado	0,999999997871199
Índice de paternidade combinado	3,8809x10 ⁸



Apesar dos marcadores D13S317 e D5S818 apresentarem valores de heterozigosidade inferiores a 70%, exibem poder de discriminação superior a 80%. Por isso, tal como na população angolana residente no sul de Portugal, confirma-se que todos os STRs analisados cumprem o primeiro critério de seleção (Urguhart *et al.*, 1996).

De acordo com os resultados apresentados anteriormente, após a aplicação da correção de Bonferroni (Weir, 1996), nenhum dos *loci* estudados demonstrou desvios ao equilíbrio de Hardy-Weinberg e os marcadores mais informativos, para esta população, são o SE33 e o D21S1. O SE33 apresenta o valor mais elevado para o PD e para o CIP e o mais reduzido para a PM. Já o D21S11, exibe maior heterozigotia observada, poder de exclusão e índice de paternidade. Por outro lado, o marcador D13S317 possui os valores mais baixos para a Ho, o PD, o CIP, o PE e o PI e o mais elevado para a PM, sendo o menos informativo.

Os resultados obtidos foram comparados com os dados referentes à população nativa guineense presentes em dois artigos, um publicado em 2002 (Gonçalves *et al.*, 2002) e outro em 2005 (Pereira *et al.*, 2005). Segundo estes dois estudos, todos os STRs analisados se encontram em HWE sendo o marcador D13S317 aquele que apresenta menor heterozigotia observada, menor poder de discriminação e menor poder de exclusão, tal como verificado no presente trabalho. Porém, será fundamental realizar mais estudos com a esta população para se poder comparar os resultados obtidos. Estes deverão analisar um maior número de indivíduos e manter os STRs estudados. Tanto Gonçalves *et al.* como Pereira *et al.*, não incluem a análise do marcador SE33 nem do *linkage disequilibrium*.



3.3 DISTÂNCIA GENÉTICA ENTRE AS POPULAÇÕES

Os resultados obtidos para o cálculo das distâncias genéticas entre as populações angolana e guineense, residentes no sul de Portugal e as populações do Sul de Portugal, de Espanha, da Holanda, dos EUA (indivíduos caucasianos e africanos), da África do Sul, do Líbano, dos Emirados Árabes Unidos, do Afeganistão, da China, do México e da Coreia estão presentes na Tabela XVI.

Tabela XVI: Distâncias genéticas entre as populações angolana e guineense residentes no sul de Portugal e as populações do Sul de Portugal, de Espanha, da Holanda, dos EUA (indivíduos caucasianos e africanos), da África do Sul, do Líbano, dos Emirados Árabes Unidos, do Afeganistão, da China, do México e da Coreia.

	Angola	Guiné-Bissau	Sul de Portugal	Espanha	Coreia	Líbano	EUA (caucasianos)	Holanda	EUA (africanos)	África do Sul	China	México	Emirados Árabes Unidos
Guiné-Bissau	0,043300												
Sul de Portugal	0,101450	0,133615											
Espanha	0,111145	0,139807	0,006803										
Coreia	0,146409	0,158146	0,094776	0,109137									
Líbano	0,094176	0,121209	0,020412	0,023921	0,087030								
EUA (caucasianos)	0,115739	0,143967	0,009428	0,013691	0,106627	0,036138							
Holanda	0,112929	0,147262	0,009821	0,011980	0,113207	0,033860	0,004732						
EUA (africanos)	0,014584	0,039337	0,081521	0,088541	0,137082	0,075000	0,090937	0,089441					
África do Sul	0,033524	0,059113	0,043186	0,050996	0,092208	0,040075	0,050887	0,051912	0,024017				
China	0,102373	0,124879	0,044688	0,053520	0,036766	0,043556	0,048082	0,050222	0,086575	0,045843			
México	0,181641	0,227332	0,096268	0,101697	0,145555	0,113755	0,103584	0,105789	0,155401	0,121067	0,108865		
Emirados Árabes Unidos	0,088213	0,120263	0,030285	0,033691	0,087220	0,014954	0,039931	0,039263	0,067219	0,032250	0,040614	0,108410	
Afeganistão	0,099368	0,126332	0,036801	0,047380	0,043752	0,033380	0,041564	0,044146	0,083704	0,038557	0,015503	0,102839	0,025873



Relativamente à população de Angola residente no sul de Portugal, verifica-se que existe maior proximidade genética com a população afro-americana ($F_{st} = 0.014584$), seguida das populações sul-africana e guineense, as quais apresentam valores de índice de fixação de 0,033524 e 0,043300, respetivamente. Por outro lado, as populações portuguesa, chinesa, espanhola, holandesa, americana caucasiana, coreana e mexicana são as que exibem maior diferenciação genética em relação à angolana, com valores de F_{st} superiores a 0,1.

A população guineense imigrante demonstra maior semelhança com as populações afro-americana ($F_{st} = 0,039337$), angolana ($F_{st} = 0,043300$) e sul-africana ($F_{st} = 0,059113$). Todavia, encontra-se menos relacionada geneticamente com as populações do Sul de Portugal, de Espanha, caucasiana dos Estados Unidos, da Holanda, do México e da Coreia, as quais apresentam índices de fixação superiores a 0,13.

A população do Sul de Portugal encontra-se mais próxima de outras populações europeias estudadas, Espanha ($F_{st} = 0,006803$) e Holanda ($F_{st} = 0,009821$), e da população caucasiana americana ($F_{st} = 0,009428$). No entanto, apresenta menor semelhança genética com as populações coreana, mexicana e com as populações imigrantes guineense e angolana residentes no sul de Portugal.

As populações chinesa ($F_{st} = 0.036766$), afegã ($F_{st} = 0,043752$), libanesa ($F_{st} = 0,087030$) e emiradense ($F_{st} = 0,087220$) apresentam maior similaridade com a população da Coreia, que por sua vez, se encontra geneticamente menos relacionada com as populações mexicana, africanas e europeias.

Estes dados podem ser observados com maior clareza através do gráfico (Figura 15) e da árvore filogenética (Figura 16) obtidos através dos modelos escala multidimensional (MDS) e *neighbor joining* (NJ), respetivamente.

Os resultados obtidos são concordantes com outros estudos populacionais, onde as populações africanas, incluindo a afro-americana, se mostram mais relacionadas geneticamente entre si do que em relação às populações europeias, asiáticas e do México (Fernandes *et al.*, 2003; Hong *et al.*, 2013; Juárez-Cedillo *et al.*, 2008; Kido *et al.*, 2007; Melo *et al.*, 2010).

Através dos estudos efetuados por Fernandes *et al.* (2003), García *et al.* (2012) e Almeida *et al.* (2015) pode verificar-se a proximidade entre as populações europeias e destas com a população americana caucasiana.

Chouery *et al.* (2010) demonstra que a população libanesa apresenta maior similaridade com outras populações do Médio Oriente e maior diferenciação da população caucasiana europeia. Älgenäs & Tillmar (2014) confirmam que as



frequências alélicas para a população afegã são significativamente diferentes das obtidas para uma população europeia, a sueca, e muito semelhantes quando comparadas com outras populações do Afeganistão e da Ásia. No estudo realizado por Zhang *et al.* (2016), pode observar-se que a população dos Emirados Árabes Unidos é mais relacionada geneticamente com populações asiáticas do que com as populações europeias e africanas. Todos estes dados são concordantes com os obtidos no presente trabalho.

Os resultados obtidos para população da Coreia, são corroborados através de um estudo efetuado por Hong *et al.* (2013) onde esta população mostra maior afinidade genética com populações pertencentes a países do continente asiático, como a China, o Japão, a Tailândia, Índia, entre outras, e menor relação com as populações europeias, nomeadamente a portuguesa e a espanhola, as africanas e a mexicana.

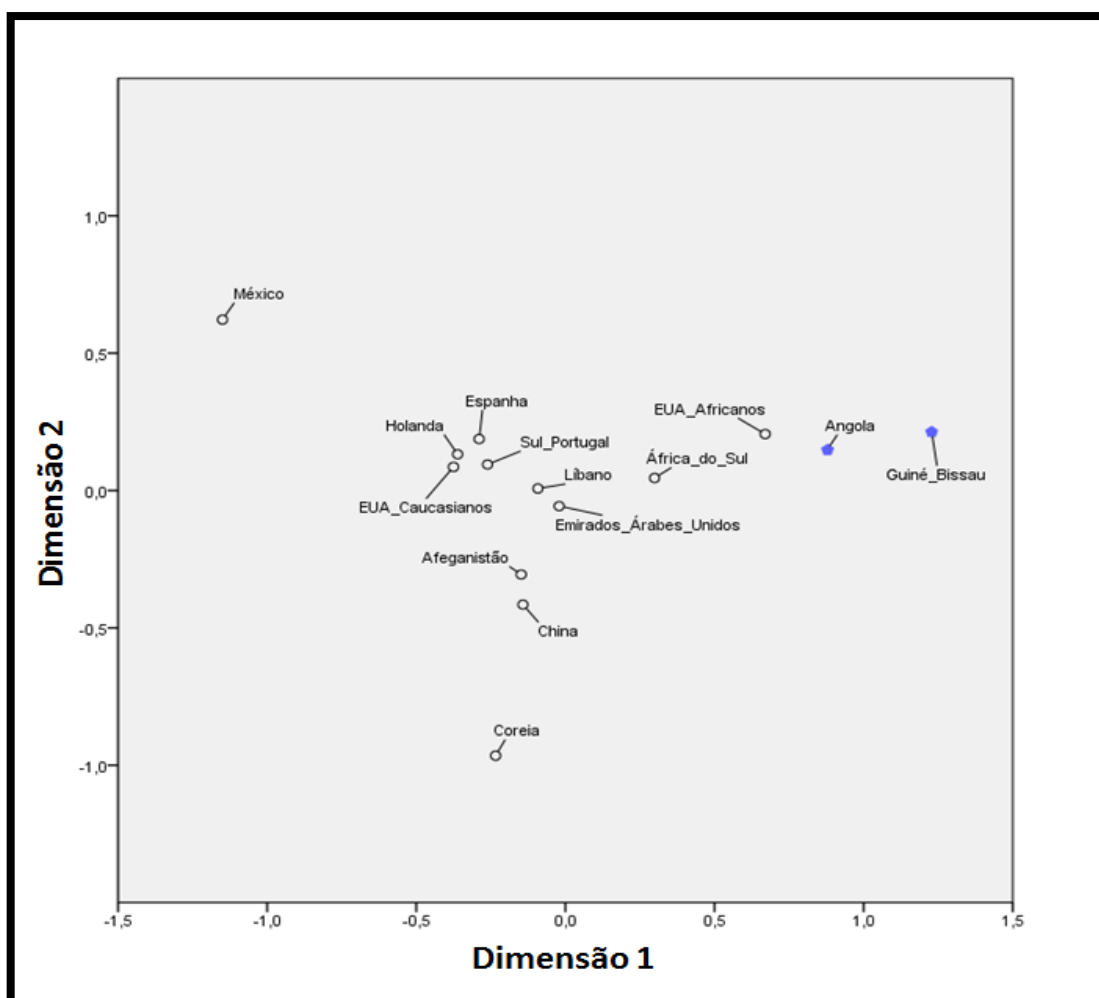


Figura 15: Gráfico MDS. Representação obtida a partir das distâncias genéticas calculadas entre as populações angolana e guineense, residentes no sul de Portugal e as populações do Sul de Portugal, de Espanha, da Holanda, dos EUA (indivíduos caucasianos e africanos), da África do Sul, do Líbano, dos Emirados Árabes Unidos, do Afeganistão, da China, do México e da Coreia.

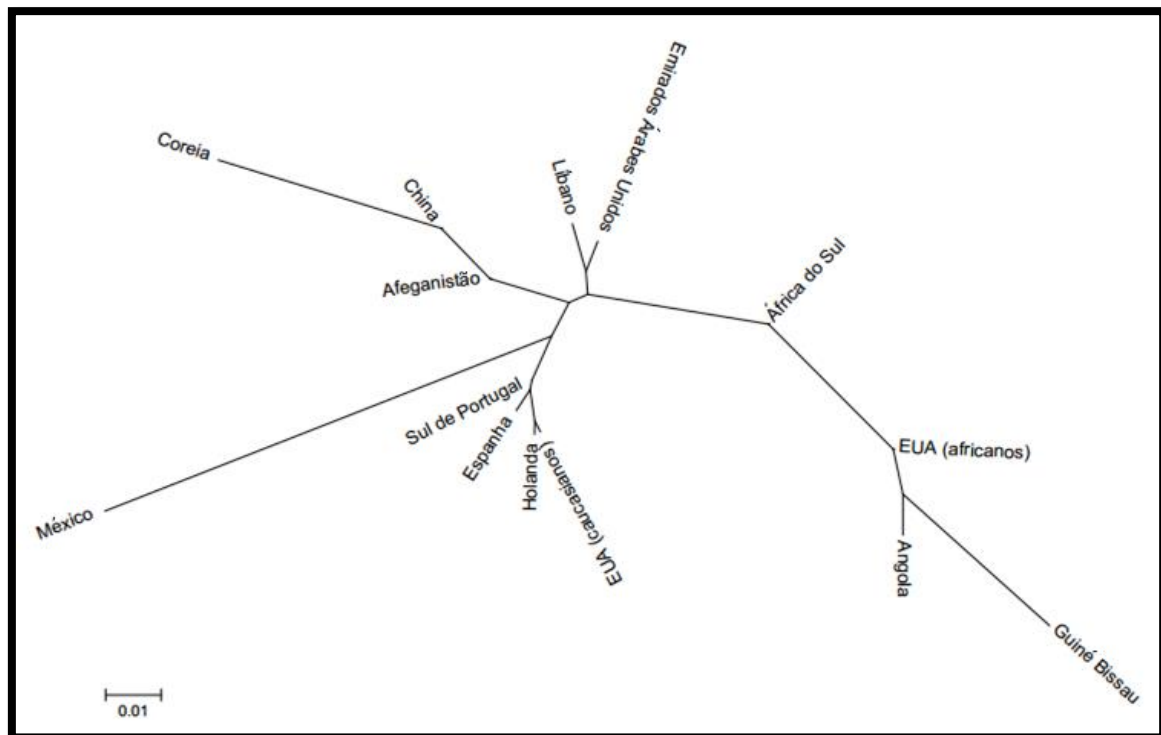


Figura 16: Árvore filogenética. Representação gráfica obtida a partir das distâncias genéticas calculadas entre as populações angolana e guineense, residentes no sul de Portugal e as populações do Sul de Portugal, de Espanha, da Holanda, dos EUA (indivíduos caucasianos e africanos), da África do Sul, do Líbano, dos Emirados Árabes Unidos, do Afeganistão, da China, do México e da Coreia.



3.4 TESTE EXATO DE DIFERENCIAÇÃO POPULACIONAL

Os resultados obtidos para o cálculo do teste exato de diferenciação populacional entre as populações angolana e guineense, residentes no sul de Portugal e as populações do Sul de Portugal, de Espanha, da Holanda, dos EUA (indivíduos caucasianos e africanos), da África do Sul, do Líbano, dos Emirados Árabes Unidos, do Afeganistão, da China, do México e da Coreia estão presentes nas Tabela XVII e XVIII.

Estes dados demonstram que as populações que apresentam diferenças significativas nas frequências alélicas após a aplicação da correção de Bonferroni ($0,05/273=0,00018$), relativamente à população angolana, são a coreana, a mexicana, a chinesa e a holandesa, em 18 de 21 *loci* analisados. Quanto à população guineense, verifica-se que as populações da Coreia, do México e da Holanda exibem diferenças consideráveis em 13, 11 e 9 *loci*, respetivamente, dos 21 marcadores genéticos estudados. Por outro lado, as populações com maior relação genética entre si são a angolana e guineense, residentes no sul de Portugal, a afro-americana e a sul africana. Nos pares Guiné-Bissau vs. Angola, e vice-versa, Angola vs. EUA (africanos) e Guiné-Bissau vs. EUA (africanos) não se observaram quaisquer valores de p significativos. Contudo, os conjuntos Angola vs. África do Sul e Guiné-Bissau vs. África do Sul apresentaram valores inferiores a 0,00018 para dois *loci* (D22S1045 e D10S1248) e um *locus* (D22S1045), respetivamente.

3. RESULTADOS E DISCUSSÃO



Tabela XVII: Valores de p para o teste exato de diferenciação populacional entre a população angolana residente no sul de Portugal e as populações guineense residente no sul de Portugal, do Sul de Portugal, de Espanha, da Coreia, do Líbano, dos EUA (indivíduos caucasianos e africanos), da Holanda, da África do Sul, da China, do México, dos Emirados Árabes Unidos e do Afeganistão. Os resultados sombreados a azul são significativos valores de p inferiores a 0,00018.

	Angola vs. Guiné- Bissau	Angola vs. Sul de Portugal	Angola vs. Espanha	Angola vs. Coreia	Angola vs. Líbano	Angola vs. EUA (caucasianos)	Angola vs. Holanda	Angola vs. EUA (africanos)	Angola vs. África do Sul	Angola vs. China	Angola vs. México	Angola vs. Emirados Árabes Unidos	Angola Vs. Afeganistão
D3S1358	0,49245+- 0,0123	0,03861+- 0,0049	0,00352+- 0,0023	0,00828+- 0,0028	0,01948+- 0,0041	0,01630+- 0,0033	0,00062+- 0,0007	0,83732+- 0,0077	0,04975+- 0,0074	0,00504+- 0,0029	0,01580+- 0,0049	0,04099+- 0,0085	0,30290+- 0,0158
VWA	0,91676+- 0,0053	0,01354+- 0,0028	0,04941+- 0,0057	0,00000+- 0,0000	0,00001+- 0,0000	0,00541+- 0,0013	0,00000+- 0,0000	0,47651+- 0,0130	0,54025+- 0,0223	0,00000+- 0,0000	0,00000+- 0,0000	0,00732+- 0,0016	0,00013+- 0,0000
D16S539	0,66019+- 0,0108	0,00000+- 0,0000	0,00000+- 0,0000	0,02340+- 0,0056	0,00246+- 0,0017	0,00000+- 0,0000	0,00000+- 0,0000	0,44836+- 0,0216	0,02054+- 0,0034	0,02015+- 0,0032	0,00196+- 0,0009	0,06375+- 0,0065	0,22617+- 0,0143
CSF1PO	0,48728+- 0,0052	0,00132+- 0,0004	0,01969+- 0,0045	0,00000+- 0,0000	0,00022+- 0,0002	0,00000+- 0,0000	0,00000+- 0,0000	0,76666+- 0,0117	0,53825+- 0,0229	0,00000+- 0,0000	0,00306+- 0,0021	0,00196+- 0,0011	0,00151+- 0,0010
TPOX	0,21726+- 0,0125	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,74952+- 0,0147	0,01585+- 0,0030	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000
D8S1179	0,44169+- 0,0133	0,00000+- 0,0000	0,00113+- 0,0006	0,00000+- 0,0000	0,00006+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,88709+- 0,0072	0,01669+- 0,0028	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000
D21S11	0,76664+- 0,0107	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,63910+- 0,0124	0,09776+- 0,0113	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000
D18S51	0,62009+- 0,0153	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00025+- 0,0003	0,00000+- 0,0000	0,39900+- 0,0186	0,00360+- 0,0017	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000
D2S441	0,76603+- 0,0095	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00002+- 0,0000	0,00000+- 0,0000	0,77557+- 0,0127	0,33372+- 0,0244	0,00000+- 0,0000	0,00000+- 0,0000	0,02035+- 0,0060	0,00000+- 0,0000
D19S433	0,75802+- 0,0102	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,52827+- 0,0146	0,00628+- 0,0015	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000
TH01	0,28981+- 0,0086	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,33603+- 0,0232	0,07601+- 0,0101	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000
FGA	0,99336+- 0,0012	0,00000+- 0,0000	0,00061+- 0,0005	0,00000+- 0,0000	0,00162+- 0,0010	0,00000+- 0,0000	0,00000+- 0,0000	0,49448+- 0,0197	0,00073+- 0,0008	0,00000+- 0,0000	0,00000+- 0,0000	0,00108+- 0,0006	0,00163+- 0,0004
D22S1045	0,07140+- 0,0053	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,50082+- 0,0116	0,00013+- 0,0001	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000
D5S818	0,07055+- 0,0042	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,71818+- 0,0116	0,07141+- 0,0082	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000
D13S317	0,31911+- 0,0167	0,00076+- 0,0003	0,00229+- 0,0009	0,00000+- 0,0000	0,00023+- 0,0001	0,00466+- 0,0010	0,00176+- 0,0009	0,78188+- 0,0103	0,09135+- 0,0097	0,00000+- 0,0000	0,00000+- 0,0000	0,00017+- 0,0001	0,00000+- 0,0000
D7S820	0,88857+- 0,0039	0,05537+- 0,0073	0,07559+- 0,0089	0,00000+- 0,0000	0,33987+- 0,0274	0,04067+- 0,0046	0,00676+- 0,0017	0,91433+- 0,0080	0,61942+- 0,0191	0,01336+- 0,0028	0,00006+- 0,0001	0,43282+- 0,0217	0,00598+- 0,0015
SE33	0,99928+- 0,0004	0,00716+- 0,0023	0,01210+- 0,0043	0,00000+- 0,0000	0,00300+- 0,0032	0,00276+- 0,0014	0,00000+- 0,0000	0,99313+- 0,0021	0,78914+- 0,0217	0,00000+- 0,0000	0,00000+- 0,0000	0,11715+- 0,0200	0,00100+- 0,0009
D10S1248	0,66177+- 0,0142	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,45001+- 0,0172	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000
D1S1656	0,67372+- 0,0122	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,58480+- 0,0114	0,11429+- 0,0118	0,00000+- 0,0000	0,00000+- 0,0000	0,00004+- 0,0000	0,00000+- 0,0000
D12S391	0,60878+- 0,0107	0,00000+- 0,0000	0,00219+- 0,0008	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,96393+- 0,0028	0,00965+- 0,0028	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000
D2S1338	0,59468+- 0,0095	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,64391+- 0,0099	0,07725+- 0,0109	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000	0,00000+- 0,0000

3. RESULTADOS E DISCUSSÃO



Tabela XVIII: Valores de p para o teste exato de diferenciação populacional entre a população guineense residente no sul de Portugal e as populações angolana residente no sul de Portugal, do Sul de Portugal, de Espanha, da Coreia, do Líbano, dos EUA (indivíduos caucasianos e africanos), da Holanda, da África do Sul, da China, do México, dos Emirados Árabes Unidos e do Afeganistão. Os resultados sombreados a azul são significativos valores de p inferiores a 0,00018.

	Guiné-Bissau vs. Angola	Guiné-Bissau vs. Sul de Portugal	Guiné-Bissau vs. Espanha	Guiné-Bissau vs. Coreia	Guiné-Bissau vs. Líbano	Guiné-Bissau vs. EUA (caucasianos)	Guiné-Bissau vs. Holanda	Guiné-Bissau vs. EUA (africanos)	Guiné-Bissau vs. África do Sul	Guiné-Bissau vs. China	Guiné-Bissau vs. México	Guiné-Bissau vs. Emirados Árabes Unidos	Guiné-Bissau vs. Afeganistão
D3S1358	0,49245+-0,0123	0,02025+-0,0033	0,00279+-0,0007	0,00383+-0,0009	0,00030+-0,0003	0,00773+-0,0016	0,00076+-0,0002	0,29368+-0,0095	0,01993+-0,0018	0,00130+-0,0004	0,01207+-0,0023	0,00229+-0,0007	0,00750+-0,0021
VWA	0,91676+-0,0053	0,90032+-0,0058	0,81036+-0,0082	0,00002+-0,0000	0,33775+-0,0126	0,53176+-0,0159	0,39134+-0,0133	0,98954+-0,0015	0,99504+-0,0006	0,14007+-0,0167	0,01272+-0,0015	0,64566+-0,0099	0,17013+-0,0086
D16S539	0,66019+-0,0108	0,01906+-0,0027	0,00314+-0,0011	0,08757+-0,0096	0,03820+-0,0041	0,00584+-0,0010	0,00934+-0,0017	0,89843+-0,0065	0,20971+-0,0089	0,74116+-0,0140	0,24426+-0,0087	0,24794+-0,0109	0,76914+-0,0148
CSF1PO	0,48728+-0,0052	0,00016+-0,0002	0,00036+-0,0003	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000	0,62295+-0,0080	0,04563+-0,0046	0,00000+-0,0000	0,00004+-0,0000	0,00000+-0,0000	0,00000+-0,0000
TPOX	0,21726+-0,0125	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000	0,10871+-0,0113	0,00042+-0,0001	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000
D8S1179	0,44169+-0,0133	0,00013+-0,0001	0,00273+-0,0012	0,00710+-0,0020	0,14007+-0,0064	0,00000+-0,0000	0,00000+-0,0000	0,81250+-0,0071	0,47120+-0,0142	0,02884+-0,0030	0,00064+-0,0006	0,01922+-0,0023	0,02623+-0,0027
D21S11	0,76664+-0,0107	0,00059+-0,0004	0,00762+-0,0019	0,00000+-0,0000	0,01876+-0,0056	0,00640+-0,0016	0,00021+-0,0002	0,50891+-0,0065	0,54237+-0,0162	0,00000+-0,0000	0,00116+-0,0006	0,01930+-0,0029	0,00002+-0,0000
D18S51	0,62009+-0,0153	0,26646+-0,0128	0,04738+-0,0057	0,00115+-0,0004	0,38127+-0,0148	0,61699+-0,0173	0,02862+-0,0035	0,81137+-0,0080	0,78124+-0,0038	0,00433+-0,0016	0,08021+-0,0061	0,08950+-0,0051	0,00533+-0,0015
D2S441	0,76603+-0,0095	0,00083+-0,0002	0,00000+-0,0000	0,08779+-0,0101	0,00208+-0,0006	0,00651+-0,0022	0,00063+-0,0003	0,63697+-0,0141	0,47449+-0,0169	0,03052+-0,0037	0,00000+-0,0000	0,06326+-0,0073	0,06174+-0,0040
D19S433	0,75802+-0,0102	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000	0,00037+-0,0002	0,00000+-0,0000	0,00000+-0,0000	0,71489+-0,0109	0,05212+-0,0058	0,00000+-0,0000	0,00000+-0,0000	0,00014+-0,0001	0,00001+-0,0000
TH01	0,28981+-0,0086	0,01775+-0,0038	0,00204+-0,0009	0,00000+-0,0000	0,02869+-0,0024	0,00000+-0,0000	0,00000+-0,0000	0,09108+-0,0035	0,75566+-0,0081	0,04366+-0,0044	0,00000+-0,0000	0,10828+-0,0096	0,06000+-0,0032
FGA	0,99336+-0,0012	0,17801+-0,0060	0,17435+-0,0165	0,34914+-0,0185	0,73967+-0,0154	0,06539+-0,0053	0,01187+-0,0023	0,93903+-0,0075	0,71075+-0,0121	0,41083+-0,0279	0,00000+-0,0000	0,75305+-0,0148	0,52457+-0,0123
D22S1045	0,07140+-0,0053	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000	0,20008+-0,0119	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000
D5S818	0,07055+-0,0042	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000	0,00000+-0,0000	0,03035+-0,0021	0,00034+-0,0002	0,00000+-0,0000	0,00000+-0,0000	0,00018+-0,0002	0,00000+-0,0000
D13S317	0,31911+-0,0167	0,00049+-0,0002	0,00055+-0,0004	0,00000+-0,0000	0,00389+-0,0011	0,00206+-0,0014	0,00054+-0,0002	0,54619+-0,0086	0,03429+-0,0032	0,00000+-0,0000	0,00000+-0,0000	0,00237+-0,0008	0,00000+-0,0000
D7S820	0,88857+-0,0039	0,22807+-0,0106	0,29785+-0,0121	0,00014+-0,0002	0,73781+-0,0100	0,15711+-0,0063	0,09233+-0,0066	0,86342+-0,0066	0,67113+-0,0119	0,06845+-0,0071	0,01634+-0,0021	0,64951+-0,0133	0,01364+-0,0018
SE33	0,99928+-0,0004	0,32288+-0,0114	0,50900+-0,0183	0,00000+-0,0000	0,13320+-0,0106	0,14652+-0,0143	0,02828+-0,0056	0,99827+-0,0005	0,93991+-0,0074	0,13645+-0,0114	0,00000+-0,0000	0,77487+-0,0248	0,19074+-0,0195
D10S1248	0,66177+-0,0142	0,01578+-0,0022	0,06776+-0,0090	0,02643+-0,0026	0,01528+-0,0023	0,01281+-0,0018	0,00119+-0,0006	0,93232+-0,0043	0,47116+-0,0180	0,03773+-0,0056	0,01391+-0,0022	0,00761+-0,0016	0,01417+-0,0024
D1S1656	0,67372+-0,0122	0,00085+-0,0008	0,00129+-0,0006	0,00000+-0,0000	0,00284+-0,0009	0,00000+-0,0000	0,00000+-0,0000	0,80118+-0,0075	0,11498+-0,0096	0,00009+-0,0001	0,00115+-0,0007	0,01297+-0,0036	0,00049+-0,0004
D12S391	0,60878+-0,0107	0,06668+-0,0104	0,05449+-0,0072	0,00997+-0,0038	0,03922+-0,0057	0,08435+-0,0106	0,02746+-0,0053	0,97952+-0,0027	0,68925+-0,0141	0,06682+-0,0072	0,00338+-0,0008	0,03065+-0,0035	0,02637+-0,0058
D2S1338	0,59468+-0,0095	0,00000+-0,0000	0,00000+-0,0000	0,00015+-0,0001	0,00000+-0,0000	0,00079+-0,0004	0,00002+-0,0000	0,31553+-0,0096	0,12512+-0,0105	0,00687+-0,0022	0,00000+-0,0000	0,00001+-0,0000	0,00029+-0,0002



Este estudo teve como principal objetivo a caracterização genética das populações angolana e guineense residentes no sul de Portugal através da análise de 21 STRs autossômicos, com recurso ao *GlobalFiler™ PCR Amplification Kit*. Para tal, foram calculados os parâmetros estatísticos populacionais e de interesse forense.

Segundo os resultados obtidos foi possível concluir que:

- ✎ Para ambas as populações, todos os *loci* se encontram em equilíbrio de Hardy-Weinberg e em *linkage disequilibrium*, após a aplicação da correção de Bonferroni.
- ✎ Todos os marcadores genéticos apresentam heterozigosidade superior a 70% com a exceção do D13S317, o qual exibe $H=0,6711$ para a população angolana e $H=0,6$ para a população guineense. No entanto, os STRs autossômicos analisados preenchem os requisitos necessários para que possam ser usados em genética forense visto que, para todos eles, o poder de discriminação é superior a 80%.
- ✎ O estudo da população angolana revelou a presença das variantes alélicas raras 26.3, 14.1 e 23.3, em três indivíduos diferentes, nos marcadores SE33, D2S441 e D21S11, respetivamente. Na população guineense foi detetada a existência do alelo raro 17.1, num indivíduo, no marcador D18S51.
- ✎ O STR SE33 apresenta os valores mais elevados para o poder de discriminação e para o conteúdo de informação polimórfica e o valor mais baixo para a probabilidade de identidade, relativamente às duas populações estudadas. Se por um lado, este marcador também exibe os resultados mais altos para o poder de exclusão e para o índice de paternidade para a população de Angola residente no sul de Portugal, por outro, no que concerne à população da Guiné-Bissau a residir no sul de Portugal, é o D21S11 que apresenta maior PE e PI. Pode assim afirmar-se que para a população angolana o marcador SE33 é o mais informativo e que a população guineense possui dois STRs muito informativos, o SE33 e o D21S11.
- ✎ Para as duas populações estudadas, o *locus* D13S317 apresentou menor CIP, PE e PI. Este também exibiu maior PM e menor PD para a população guineense, no entanto, o D5S818 demonstrou o valor mais elevado para a probabilidade de identificação e o mais reduzido para o poder de

discriminação na análise dos parâmetros de interesse forense obtidos para a população angolana. Pode assim dizer-se, que para a população guineense imigrante no sul de Portugal o marcador D13S317 é o menos informativo e que para a população angolana imigrante no sul de Portugal os *loci* menos informativos são o D5S818 e o D13S317.

- [illegible]

Outro dos objetivos deste projeto era a comparação das populações estudadas com outras populações publicadas na literatura, tais como, a do Sul de Portugal, a de Espanha, a da Holanda, a dos EUA (indivíduos caucasianos e africanos), a da África do Sul, a do Líbano, a dos Emirados Árabes Unidos, a do Afeganistão, a da China, a do México e a da Coreia. Para tal, calcularam-se as distâncias genéticas entre todas elas e o teste exato de diferenciação populacional entre as populações presentemente analisadas e as 12 restantes.

De acordo com os dados obtidos pôde concluir-se que:

- As populações de Angola e Guiné-Bissau residentes no sul de Portugal, apresentam maior semelhança genética entre si e relativamente a outras populações africanas, como a afro-americana e a sul-africana. No entanto, demonstram-se mais afastadas geneticamente das populações europeias, nomeadamente da população do Sul de Portugal, asiáticas e mexicana. Devido à distância genética existente entre as populações estudadas e a



população do Sul de Portugal, pode afirmar-se que as populações angolana e guineense introduzirão variabilidade genética na população portuguesa nativa.

✎ O teste exato de diferenciação populacional demonstrou que não existem diferenças significativas entre as frequências alélicas obtidas para as populações estudadas e para a afro-americana. No entanto, relativamente à população da África do Sul foram observadas diferenças significativas, após a correção de Bonferroni, em dois *loci* (D22S1045 e D10S1248), quando comparada com a população angolana e num *locus* (D22S1045), quando comparada com a população guineense. Para as restantes populações foram encontradas diferenças significativas em múltiplos *loci*.



- Abrantes, D. (2010). Valorização probabilística da prova e respectiva interpretação em genética forense. Em M. F. Pinheiro, *Genética Forense: Perspectivas da Identificação Genética* (pp. 257-325). Porto, Portugal: Universidade Fernando Pessoa
- Älgenäs, C., & Tillmar, A. O. (2014). Population genetics of 29 autosomal STRs and 17 Y-chromosomal STRs in a population sample from Afghanistan. *International Journal of Legal Medicine*, 128: 279-280
- Alhmoudi, O. A., Jones, R. J., Alsafar, H. & Hadi, S. (2015). Population genetics data for 21 autosomal STR loci for United Arab Emirates (UAE) population using next generation multiplex STR kit. *Forensic Science International: Genetics*, 19: 190-191
- Almeida, C., Ribeiro, T., Oliveira, A. R., Porto, M. J., Santos, J. C., Dias, D., & Dario, P. (2015). Population data of the GlobalFiler® Express loci in South Portuguese population. *Forensic Science International: Genetics*, 19: 39-41
- Alvarez, M., Piedade, J., Balseiro, S., Ribas, G., & Regateiro, F. (2009). HLA-G 3'-UTR SNP and 14-bp deletion polymorphisms in Portuguese and Guinea-Bissau populations. *International Journal of Immunogenetics*, 36(6): 361-366.
- Andari, A., Othman, H., Taroni, F., & Mansour, I. (2013). Population genetic data for 23 STR markers from Lebanon. *Forensic Science International: Genetics*, 7: e108-e113
- Applied Biosystems (2015). *GlobalFiler™ PCR Amplification Kit User Guide*. Waltham, Massachusetts, USA: Thermo Fisher Scientific Inc.
- Assembleia Nacional Popular da República da Guiné-Bissau. (2015). Consultado a 3 de julho de 2016. Disponível em: <http://www.anpguinebissau.org/>
- Augel, M. (2007). Panorama Histórico. Em *O desafio do escombros: nação, identidades e pós-colonialismo na literatura da Guiné-Bissau* (pp. 51-70). Rio de Janeiro, Brasil: Garamond Ltda
- Beleza, S., Alves, C., Reis, F., Amorim, A., Carracedo, A. & Gusmão, L. (2004). 17 STR data (AmpF/STR Identifier and Powerplex 16 System) from Cabinda (Angola). *Forensic Science International*, 141: 193-196
- Benzinho, J., & Rosa, M. (2015a). Um Retrato da Guiné-Bissau. Em *Guia turístico: à descoberta da Guiné-Bissau* (pp. 15-23). Escoural, Pombal: Afectos com Letras - ONGD
- Benzinho, J., & Rosa, M. (2015b). Introdução Histórica. Em *Guia turístico: à descoberta da Guiné-Bissau* (pp. 11-13). Escoural, Pombal: Afectos com Letras - ONGD



- Bhatnagar, S. M., Kothari, M. L. & Mehta, L. A. (1999). Molecular Genetics: The Genetic Basis of Inheritance. Em *Essentials Of Human Genetics* (4^a ed., pp. 44-65). Chennai, Índia: Orient Longman
- Blumberg, R. B. (ed.) (1997) *MendelWeb*. Edition 97.1. Disponível em: <http://www.mendelweb.org/>
- Brito, P. (2015). *Estudo de Marcadores Bialélicos (SNPs) do Cromossoma Y numa População Africana (Angola)* (Dissertação de Mestrado em Biologia Celular e Molecular). Universidade de Coimbra, Coimbra.
- Butler, J. M. (2005). STR Population Database Analyses. Em *Forensic DNA Typing* (2^a ed., pp. 473-496). Burlington, USA: Elsevier Academic Press
- Butler, J. M. (2010a). Basics of DNA Biology and Genetics. Em *Fundamentals of Forensic DNA Typing* (pp. 19-41). Burlington, USA: Elsevier Academic Press
- Butler, J. M. (2010b). Overview and History of DNA Typing. Em *Fundamentals of Forensic DNA Typing* (pp. 1-18). Burlington, USA: Elsevier Academic Press
- Butler, J. M. (2010c). Historical Methods. Em *Fundamentals of Forensic DNA Typing* (pp. 43-78). Burlington, USA: Elsevier Academic Press
- Butler, J. M. (2010d). Short Tandem Repeat Markers. Em *Fundamentals of Forensic DNA Typing* (pp. 147-173). Burlington, USA: Elsevier Academic Press
- Butler, J. M. (2010e). STR Genotyping and Data Interpretation. Em *Fundamentals of Forensic DNA Typing* (pp. 205-227). Burlington, USA: Elsevier Academic Press
- Butler, J. M. (2010f). DNA Extraction. Em *Fundamentals of Forensic DNA Typing* (pp. 99-109). Burlington, USA: Elsevier Academic Press
- Butler, J. M. (2010g). DNA Amplification (The Polymerase Chain Reaction). Em *Forensic DNA Typing: Biology, Technology and Genetics of STR Markers* (pp. 125-146). Burlington, USA: Elsevier Academic Press
- Butler, J. M. (2010h). Fundamentals of DNA Separation and Detection. Em *Forensic DNA Typing: Biology, Technology and Genetics of STR Markers* (pp. 175-203). Burlington, USA: Elsevier Academic Press
- Butler, J. M. (2011a). Short Tandem Repeat (STR) Loci and Kits. Em *Advanced Topics in Forensic DNA Typing: Methodology* (pp. 99-139). Waltham, USA: Elsevier Academic Press
- Butler, J. M. (2011b). Capillary Electrophoresis: Principles and Instrumentation. Em *Advanced Topics in Forensic DNA Typing: Methodology* (pp. 99-139). Waltham, USA: Elsevier Academic Press



- Butler, J. M. (2011c). DNA Databases: Uses and Issues. Em *Advanced Topics in Forensic DNA Typing: Methodology* (pp. 213-270). Waltham, USA: Elsevier Academic Press
- Camará, S., Costa, C., Manafá, B., Semedo, S. & Vieira, R. (2014). *Estatísticas Básicas da Guiné-Bissau*. Pinjiguite, Bissau: Instituto Nacional de Estatística.
- Carracedo, A. & Sánchez-Diz, P. (2000). Forensic genetics: from classical serological genetic markers to DNA polymorphisms analyzed by microarray technology. Em M. J. Bogus, *Handbook of analytical separations, Forensic Science* (1ª ed., vol 2, pp. 695-706). Amsterdão, Holanda: Elsevier
- Carvalho, A. (2015). *Nuts 2013 – As Novas Unidades Territoriais para Fins Estatísticos*. Lisboa, Portugal: Instituto Nacional de Estatística, I.P.
- Ceita, C. & Machado, A. (2014). Resultados Definitivos do Recenseamento Geral da População e Habitação. Luanda, Angola: Instituto Nacional de Estatística.
- Chouery, E., Coble, M. D., Strouss, K. M., Saunier, J. L., Jalkh, N., Medlej-Hashim, M., Ayoub, F. & Mégarbané, A. (2010). Population genetic data for 17 STR markers from Lebanon. *Legal Medicine*, 12: 324-326
- Clayton, T. M., Guest, J. L., Urquhart, A. J. & Gill, P. D. (2004). A Basis for Anomalous Band Patterns Encountered During DNA STR Profiling. *Journal of Forensic Science*, 49(6): 1-8
- Conselho de Fiscalização da Base de Dados de Perfis de ADN. (2014). Consultado a 2 de agosto de 2016. Disponível em: <http://www.anpguineebissau.org/>
- Consulado Geral de Angola no Porto. (s.d.). Consultado a 3 de julho de 2016. Disponível em: <http://www.consuladogeralangola-porto.pt/>
- Corte-Real, A. & Porto, M. J. (2016). Identificação genética. Em F. Corte-Real & D. N. Vieira, *Princípios de Genética Forense* (pp. 147-168). Coimbra, Portugal: Imprensa da Universidade de Coimbra
- Corte-Real, F. (2016). Bases de dados de perfis de ADN. Em F. Corte-Real & D. N. Vieira, *Princípios de Genética Forense* (pp. 143-175). Coimbra, Portugal: Imprensa da Universidade de Coimbra
- Dario, P. & Geada, H. S. (2016). Investigação biológica de parentesco. Em F. Corte-Real & D. N. Vieira, *Princípios de Genética Forense* (pp. 93-122). Coimbra, Portugal: Imprensa da Universidade de Coimbra
- Dias, P., Machado, R. & Oliveira, A. (2015). Relatório de Imigração, Fronteiras e Asilo. Oeiras, Portugal: Serviço de Estrangeiros e Fronteiras.
- Diogo, J. (2014). *Parâmetros de Construção Da Democracia em Angola: Perfil Democrático* (Monografia de Licenciatura em Direito). Universidade do



- Extremo Sul Catarinense, Criciúma.
- Embaixada da República de Angola em Washington, DC (2015). Consultado a 3 de julho de 2016. Disponível em: <http://www.angola.org/>
- Embaixada da República de Angola na Áustria, Croácia e Eslovénia. (s.d.). Consultado a 3 de julho de 2016. Disponível em: <http://www.embangola.at/>
- Embaixada da República de Angola na Hungria (s.d.). Consultado a 3 de julho de 2016. Disponível em: <http://angolaembassy.hu/pt/>
- Excoffier, L., & Lischer, H. (2010a). Manual Arlequin ver 3.5.2.1 – An Integrated Software Package for Population Genetics Data Analysis. Bern, Switzerland: Computational and molecular population genetics lab (CMPG) of Institute of Ecology and Evolution University of Berne. Disponível em: <http://cmpg.unibe.ch/software/arlequin35/man/Arlequin35.pdf>
- Excoffier, L., & Lischer, H. (2010b). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3): 564-567
- Felsenstein, J. (2005). PHYLIP (Phylogeny Inference Package) version 3.6. Seattle, USA: Department of Genome Sciences, University of Washington
- Fernandes, A. T., Velosa, R., Jesus, J., Carracedo, A., & Brehm, A. (2003). Genetic differentiation of the Cabo Verde Archipelago population analysed by STR polymorphisms. *Annals of Human Genetics*, 67: 340-347
- Filomena, C. & Cheron, C. (2008). *III Seminário Internacional Organizações e Sociedade: Inovações e Transformações Contemporâneas - Luta pela libertação colonial e guerra civil em Angola*. Porto Alegre, Brasil: Pontifícia Universidade Católica do Rio Grande do Sul.
- Frankham, R., Ballou, J. D., Briscoe, D. A. (2010). Characterizing genetic diversity: single loci. Em *Introduction to Conservation Genetics*. (2ª ed, pp. 66-89). New York, USA: Cambridge University Press
- García, O., Alonso, J., Cano, J. A., García, R., Luque, G. M., Martín, P., Yuso, I. M., Maulini, S., Parra, D. & Yurrebaso, I. (2012). Population genetic data and concordance study for the kits Identifiler, NGM, PowerPlex ESX 17 System and Investigator ESSplex in Spain. *Forensic Science International: Genetics*, 6: e78-e79
- Gonçalves, R., Jesus, J., Fernandes, A. T. & Brehm, A. (2002). Genetic profile of multi-ethnic population from Guiné-Bissau (west African coast) using the new PowerPlex® 16 System kit. *Forensic Science International*, 129: 78-80



- Goodwin, W., Linacre, A. & Hadi, S. (2011a). DNA structure and the genome. Em *An Introduction to Forensic Genetics* (2ª ed., pp. 11-20). West Sussex, United Kingdom: Wiley-Blackwell
- Goodwin, W., Linacre, A. & Hadi, S. (2011b). Introduction to forensic genetics. Em *An Introduction to Forensic Genetics* (2ª ed., pp. 1-9). West Sussex, United Kingdom: Wiley-Blackwell
- Goodwin, W., Linacre, A. & Hadi, S. (2011c). The analysis of short tandem repeats. Em *An Introduction to Forensic Genetics* (2ª ed., pp. 67-79). West Sussex, United Kingdom: Wiley-Blackwell
- Goodwin, W., Linacre, A. & Hadi, S. (2011d). DNA extraction and quantification. Em *An Introduction to Forensic Genetics* (2ª ed., pp. 37-52). West Sussex, United Kingdom: Wiley-Blackwell
- Goodwin, W., Linacre, A. & Hadi, S. (2011e). Polymerase chain reaction. Em *An Introduction to Forensic Genetics* (2ª ed., pp. 53-65). West Sussex, United Kingdom: Wiley-Blackwell
- Goodwin, W., Linacre, A. & Hadi, S. (2011f). The analysis of short tandem repeats. Em *An Introduction to Forensic Genetics* (2ª ed., pp. 67-80). West Sussex, United Kingdom: Wiley-Blackwell
- Goodwin, W., Linacre, A. & Hadi, S. (2011g). Statistical interpretation of STR profiles. Em *An Introduction to Forensic Genetics* (2ª ed., pp. 95-106). West Sussex, United Kingdom: Wiley-Blackwell
- Goodwin, W., Linacre, A. & Hadi, S. (2011h). Appendix A Forensic parameters. Em *An Introduction to Forensic Genetics* (2ª ed., pp. 181-182). West Sussex, United Kingdom: Wiley-Blackwell
- Hardy, G. H. (1908) Mendelian proportions in a mixed population. *Science*, 28: 49-50
- Hares, D. R. (2015). Selection and Implementation of Expanded CODIS Core Loci in the United States. *Forensic Sci. Int. Genet.*, 17: 33-34.
- Hartl, D. L. & Clark, A. G. (2007). Organization of Genetic Variation. Em *Principles of Population Genetics* (4ª ed., pp. 45-92). Sunderland, USA: Sinauer Associate
- Hedrick, P. W. (2011a). Selection. Em *Genetics of Populations*. Massachusetts, USA: Jones and Bartlett Publishers (4ª ed., pp. 111-186)
- Hedrick, P. W. (2011b). Linkage Disequilibrium and Recombination. Em *Genetics of Populations*. (4ª ed., pp. 523-598)
- Hill, C. R., Duewer, D. L., Kline, M. C., Coble, M. D., & Butler, J. M. (2013). US



- population data for 29 autosomal STR loci. *Forensic science international. Genetics*, 7: e82-e83
- Huel, R. L. M., Lara, B., Madacki-Todorovi □, K., Smajlovi □, L., Eminovi □, I., Berbi □, I., Miloš, A., & Parsons, T. J. (2007). Variant alleles, triallelic patterns, and point mutations observed in nuclear short tandem repeat typing of populations in Bosnia and Serbia. *Croat Med J*, 48: 494-502
- Huston, K. A. (1998). Statistical Analysis of STR Data. *Profiles in DNA*, 1(3): 14-15
- Inácio, A., Costa, H. A., Silva, C. V., Ribeiro, T., Porto, M. J., Santos, J. C., Igrejas, G. & Amorim, A. (2015). Study of y-SNPs genetic markers with forensic interest and ancestry informative power in PALOP's immigrant populations in Lisboa. *Forensic Science International: Genetics Supplement Series*, 5: e3-e4.
- Jeffreys, A. J., Brookfield, J. F. & Semeonoff, R. (1985). Positive identification of an immigration test-case using human DNA fingerprints. *Nature*, 317(6040): 818-819
- Jeffreys, A. J., Wilson, V. & Thein, S. L. (1985). Hypervariable 'minisatellite' regions in human DNA. *Nature*, 314(6006): 67-73
- Jobling, M., Hollox, E., Hurles, M., Kivisild, T. & Tyler-Smith, C. (2014). Making Inferences from Diversity. Em *Human Evolutionary Genetics* (2ª ed., pp. 167-224). New York, USA: Garland Science
- Juárez-Cedillo, T., Zuñiga, J., Acuña-Alonzo, V., Pérez-Hernández, N., Rodríguez-Pérez, J. M., Barquera, R., Gallardo, G. J., Sánchez-Arenas, R., García-Penã, M. C., Granados, J. & Vargas-Alarcón, G. (2008). Genetic admixture and diversity estimations in the Mexican Mestizo population from Mexico City using 15 STR polymorphic markers. *Forensic Science International: Genetics*, 2: e37-e39
- Kiaris, H. (2012). Genes, Polymorphisms and Genetic Heterogeneity. Em *Genes, Polymorphisms, and the Making of Societies: How Genetic Behavioral Traits Influence Human* (pp. 19-27). Florida, USA: Universal-Publishers
- Kido, A., Dobashi, Y., Fujitani, N., Hara, M., Susukida, R., Kimura, H., & Oya, M. (2007). Population data on the AmpFISTR Identifiler loci in Africans and Europeans from South Africa. *Forensic science international*, 168(2): 232-235
- Krebes, J. E., Goldstein, E. S. & Kilpatrick, S. T. (2014). Genes Are DNA. Em *Lewin's Genes XI* (pp. 2-25). Burlington, USA: Jones & Bartlett Learning
- Landsteiner, K. & Levine, P. (1928). On individual differences in human blood. *J.*



- Exp. Med.*, 47(5): 757-775
- Lei n.º 5/2008 de 12 de fevereiro de 2008. *Diário da República n.º 30 – 1ª série*. Lisboa: Assembleia da República.
- Lewontin, R. C. & Kojima, K. (1960). The evolutionary dynamics of complex polymorphisms. *Evolution*, 14: 458-472
- Li, R. (2015a). Forensic Biology : A Subdiscipline of Forensic Science. Em *Forensic Biology* (2ª ed., pp. 53-76). Florida, USA: CRC Press
- Li, R. (2015b). Autosomal Short Tandem Repeat Profiling. Em *Forensic Biology* (2ª ed., pp. 369-406). Florida, USA: CRC Press
- Li, R. (2015c). Nucleic Acid Extraction. Em *Forensic Biology* (2ª ed., pp. 111-131). Florida, USA: CRC Press
- Li, R. (2015d). Amplification by Polymerase Chain Reaction. Em *Forensic Biology* (2ª ed., pp. 143-157). Florida, USA: CRC Press
- Li, R. (2015e). DNA Electrophoresis. Em *Forensic Biology* (2ª ed., pp. 143-157). Florida, USA: CRC Press
- Liu, G., & Luan, Y. (2014). Identification of Protein Coding Regions in the Eukaryotic DNA Sequences Based on Marple Algorithm and Wavelet Packets Transform. *Abstract and Applied Analysis*, vol. 2014: 1-14.
- Lodish, H., Berk, A., Kaiser, C. A., Kreiger, M., Bretscher, A., Ploegh, H., Amon, A. & Scott, M. P. (2013). Basic Molecular Genetic Mechanisms. Em *Molecular Cell Biology* (pp. 115-169). New York, USA: W. H. Freeman and Company
- Machado, F. (1997). Contornos e especificidades da imigração em Portugal. *Sociologia - Problemas E Práticas*, 24: 9-44.
- Martins, T., & Carvalho, R. (2008). Conceitos Básicos de Genética com Aplicação à Investigação Criminal. Em M. F. Pinheiro, *CSI Criminal* (1ª ed., pp 83-94). Porto, Portugal: Universidade Fernando Pessoa
- Melo, M. M., Carvalho, M., Lopes, V., Anjos, M. J., Serra, A., Vieira, D. N., Sequeiros, J. & Corte-Real, F. (2010). Genetic study of 15 STRs loci of Identifiler system in Angola population. *Forensic Science International: Genetics*, 4: e153-e157
- Mendel, G. (1866). Versuche über Pflanzenghybriden. *Verhandlungen des naturforschenden Vereines in Brünn*, 4: 3-47
- Mendes, F., Santos, J. & Rego, C. (2012). VII Congresso Português de Sociologia - A Imigração Africana em Portugal nos Últimos Vinte Anos: Oportunidades e Ameaças no Mercado de Trabalho. Porto, Portugal: Universidade do Porto, Faculdade de Letras, Faculdade de Psicologia e Ciências da Educação.



- Mullis, K. B. & Faloona, F. A. (1987). Specific synthesis of DNA in vitro via a polymerasecatalyzed chain reaction. *Methods Enzymol*, 155: 335-350.
- Olaisen, B., Bär, W., Mayr, W., Lincoln, P., Carracedo, A., Brinkmann, B., Budowle, B. & Gill, P. (1997). DNA recommendations – further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. *Forensic Science International*, 87: 179-184.
- Park, H. C., Kim, K., Nam, Y., Park, J., Lee, J., Lee, H., Kwon, H., Jin, H., Kim, W., Kim, W. & Lim, S. (2016). Population genetic study for 24 STR loci and Y indel (GlobalFiler™ PCR Amplification kit and PowerPlex® Fusion system) in 1000 Korean individuals. *Legal Medicine*, 21: 53-57
- Pereira, A. (2008). Bases de dados genéticos. Em M. F. Pinheiro, *CSI Criminal* (1ª ed., pp 11-40). Porto, Portugal: Universidade Fernando Pessoa
- Pereira, J. V., Serra, A., Anjos, M. J., Vide, M. C., Corte-Real, F. & Vieira, D. N. (2005). Allele Frequencies for Fifteen STR Loci in Guinea-Bissau Population. *J Forensic Sci*, 50(1): 243-244
- Pinheiro, M. F. (2008). A Perícia em Genética e Biologia Forense – Criminalística Biológica. Em *CSI Criminal* (1ª ed., pp 11-40). Porto, Portugal: Universidade Fernando Pessoa
- Pinheiro, M. F. (2016). Criminalística biológica. Em F. Corte-Real & D. N. Vieira, *Princípios de Genética Forense* (pp. 41-71). Coimbra, Portugal: Imprensa da Universidade de Coimbra
- Pinheiro, T. (2008). Emigration, Immigration and Interculturality: The Meaning of the European Year of Intercultural Dialogue in Portugal. *Journal of the Institute for Euroregional Studies*, 6: 63-73.
- Portal Oficial do Governo da Guiné-Bissau. (s.d.). Consultado a 3 de julho de 2016. Disponível em: <http://www.gov.gw>
- Portaria n.º 270/2009 de 17 de março de 2009. *Diário da República n.º 53, 1ª série*. Lisboa: Ministério da Justiça e da Saúde.
- Primorac, D., Schanfield, M. & Marjanovic', D. (2014). Forensic DNA Analysis and Statistics. Em *Forensic DNA Applications: An Interdisciplinary Perspective*. (pp. 55-84). Florida, USA: CRC Press
- Ramos-González, B., Aguilar-Velázquez, J. A., de Lourdes Chávez-Briones, M., Delgado-Chavarría, J. R., Alfaro-Lopez, E., & Rangel-Villalobos, H. (2016). Population data of 24 STRs in Mexican-Mestizo population from Monterrey, Nuevo Leon (Northeast, Mexico) based on Powerplex® Fusion and GlobalFiler® kits. *Forensic Science International: Genetics*, 21: e15-e17.



- Regateiro, F. J. (2007a). Bases Celulares e Moleculares da Hereditariedade. Em *Manual de Genética Médica* (1ª ed, pp 9-28). Coimbra, Portugal: Imprensa da Universidade de Coimbra
- Regateiro, F. J. (2007b). Diversidade Humana. Mutações. Reparação do DNA. Em *Manual de Genética Médica* (1ª ed, pp 41-66). Coimbra, Portugal: Imprensa da Universidade de Coimbra
- Resolução do Conselho de 25 de junho de 2001 relativa ao intercâmbio de resultados de análises de ADN. *Jornal Oficial da União Europeia, C 187: 1-4*
- Resolução do Conselho de 30 de novembro de 2009 relativa ao intercâmbio de resultados de análises de ADN. *Jornal Oficial da União Europeia, C 296: 1-3*
- Ristow, P. G., Cloete, K. W., & D'Amato, M. E. (2016). GlobalFiler® Express DNA amplification kit in South Africa: Extracting the past from the present. *Forensic Science International: Genetics*, 24: 194-201
- Rousset, F. (2015). Genepop 4.4 for Windows/Linux/Mac OS X. Disponível em: <http://kimura.univ-montp2.fr/~rousset/Genepop.pdf>
- Santos, R. M., Silva, C. V., Costa, H. A., Sanches, S., Espinheira, R., Santos, J. C. & Amorim, A. (2012). Genetic portrait of na immigrant population from Angola living in Lisboa. *Forensic Science International: Genetics*, 6: e170-e173
- Schneider, P. M. (2009). Expansion of the European Standard Set of DNA Database loci – the current situation. *Profiles in DNA*, 12(1): 6-7.
- Short Tandem Repeat DNA Internet DataBase (1997a). *Tri-Allelic Patterns for D8S1179*. Consultado a 21 de agosto de 2016. Disponível em: http://www.cstl.nist.gov/strbase/var_D8S1179.htm#Tri
- Short Tandem Repeat DNA Internet DataBase (1997b). *Variante Alleles for D2S441*. Consultado a 21 de agosto de 2016. Disponível em: http://www.cstl.nist.gov/strbase/var_D2S441.htm
- Short Tandem Repeat DNA Internet DataBase (1997c). *Variante Alleles for D21S11*. Consultado a 21 de agosto de 2016. Disponível em: http://www.cstl.nist.gov/strbase/var_D21S11.htm
- Short Tandem Repeat DNA Internet DataBase (1997d). *Variante Alleles for D18S51*. Consultado a 21 de agosto de 2016. Disponível em: http://www.cstl.nist.gov/strbase/var_D18S51.htm
- Single, R. M. & Thomson, G. (2016) Linkage Disequilibrium: Population Genetics of



- Multiple Loci. Em R. M. Kliman, *Encyclopedia of Evolutionary Biology* (pp. 400-404). Oxford, United Kingdom: Academic Press
- Slatkin, M. (2008). Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, 9(6): 477-485.
- Souto, L. (2016). Alguns conceitos de genética populacional com relevância em genética forense. Em F. Corte-Real & D. N. Vieira, *Princípios de Genética Forense* (pp. 123-142). Coimbra, Portugal: Imprensa da Universidade de Coimbra
- SPSS, IBM (2011). IBM SPSS statistics for Windows, version 20.0. New York, USA: IBM Corp.
- Strachan, T., Goodship, J. & Chinnery, P. (2015). Principles of Genetic Variation. Em *Genetics and Genomics in Medicine* (pp. 79-116). New York, USA: Garland Science
- Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013) MEGA 6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, 30: 2725-2729
- Templeton, A. R. (2006). Population Structure and History. Em *Population Genetics and Microevolutionary Theory* (pp. 19-246). New Jersey, USA: Wiley-Liss
- Tereba, A. (1999). Tools for analysis of population statistics, *Prof. DNA*, 9: 14-16
- Thermo Fisher Scientific Inc. (2015). *GlobalFiler® PCR Amplification Kit*. Consultado a 5 de agosto de 2016. Disponível em: <https://www.thermofisher.com/order/catalog/product/4476135?ICID=search-product>
- Urquhart, A. J., Oldroyd, N. J., Downes, T., Barber, M., Alliston-Greiner, R., Kimpton, C. P., & Gill, P. D. (1996). Selection of STR loci for forensic identification systems. Em A. Carracedo, B. Brinkmann & W. Bär, *Advances in Forensic Haemogenetics* 6 (pp. 115-117). Heidelberg, Alemanha: Springer Berlin Heidelberg.
- Walsh, P. S., Metzger, D. A., & Higuchi, R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques*, 10(4): 506-13
- Warner, R. (1991). Precolonial Angola and the Arrival of the Portuguese. Em T. Collelo, *Angola: a country study* (3ª ed., pp. 5-14). Washington, D.C., USA: Library of Congress, Federal Research Division
- Weinberg, W. (1908) Über den Nachweis der Vererbung beim Menschen. *Jahresh. Ver. Vaterl. Naturkd. Württemb*, 64: 369-382

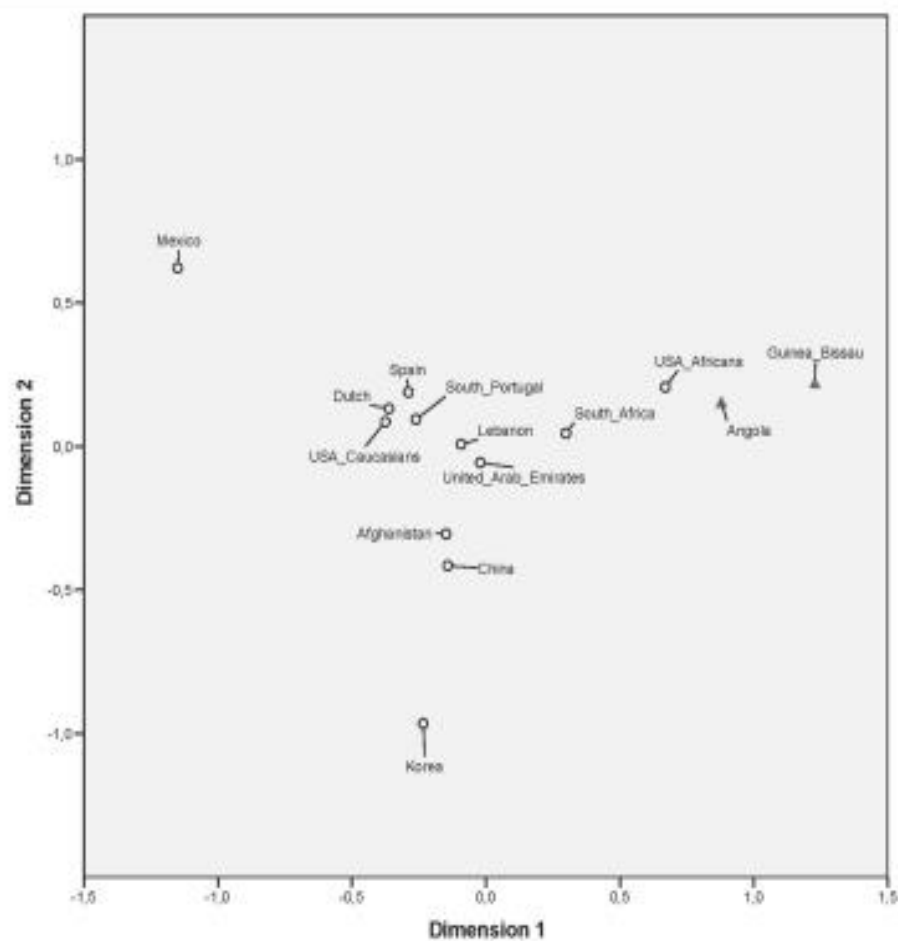


- Weir, B. S. (1996). Disequilibrium. Em *Genetic data analysis II* (pp. 91-140). Sunderland, USA: Sinauer Associates.
- Westen, A. A., Kraaijenbrink, T., de Medina, E. A. R., Harteveld, J., Willemse, P., Zuniga, S. B., Gaag, K. J., Weiler, N. E. C., Warnaar, J., Kayser, M., Sijen, T. & Kniff, P. (2014). Comparing six commercial autosomal STR kits in a large Dutch population sample. *Forensic Science International: Genetics*, 10: 55-63
- Willard, J. M., Lee, D. A. e Holland, M. M. (1998). Recovery of DNA for PCR amplification from blood and forensic samples using a chelating resin. *Methods Mol Biol*, 98: 9-18
- Wright, S. (1951). The genetical structure of populations. *Ann. Euethcs*, 15: 323-54
- Zamir, A., Shpitzen, M., Oz, C., Motro, U., Meiner, V & Gafny, R. (2002). Presentation of a Three-Banded Allele Pattern – Analysis and Interpretation. *Journal of Forensic Science*, 47(4): 1-3
- Zau, F. (2002). Contexto histórico da República de Angola. Em *Angola: Trilhos para o Desenvolvimento* (pp. 35-55). Lisboa, Portugal: Universidade Aberta
- Zhang, H., Xia, M., Qi, L., Dong, L., Song, S., Ma, T., Yang, S. Jin, L., Liming, L. & Li, S. (2016). Forensic and population genetic analysis of Xinjiang Uyghur population on 21 short tandem repeat loci of 6-dye GlobalFiler™ PCR Amplification kit. *Forensic Science International: Genetics*, 22: 22-24



Anexo I: Artigo “*Characterization of GlobalFiler loci in Angolan and Guinean populations inhabiting Southern Portugal*” publicado na revista científica internacional *Internacional Journal of Legal Medicine*

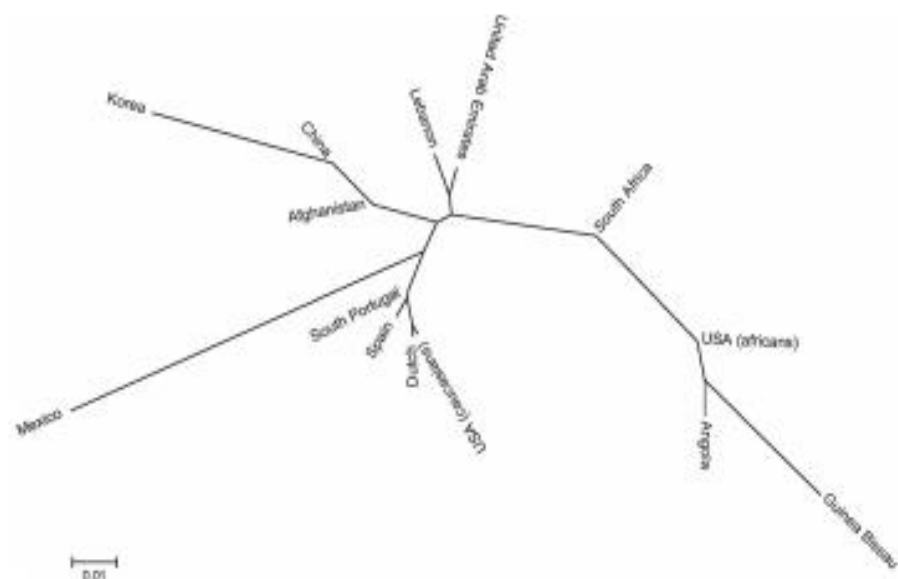
Fig. 1 Multidimensional scaling (MDS) plot based on genetic distance values between Angola and Guinea-Bissau and other populations



The laboratory is ISO/IEC 17025:2005 accredited and participates in the collaborative quality control and proficiency

testing exercises of the Spanish and Portuguese Working Group (GHEP) from ISFG.

Fig. 2 Neighbor-joining tree constructed with genetic distance values between Angola and Guinea-Bissau and other populations



References

- Alvarez M, Piedade J, Balseiro S, Ribas G, Regateiro F (2009) HLA-G 3'-UTR SNP and 14-bp deletion polymorphisms in Portuguese and Guinea-Bissau populations. *Int J Immunogenet* 36(6):361–366. doi:10.1111/j.1744-313X.2009.00875.x
- Santos RM, Silva CV, Costa HA, Gomes PF, Sanches S, Espinheira R, Santos JC, Amorim A (2012) Genetic portrait of an immigrant population from Angola living in Lisboa. *Forensic Sci Int-Gen* 6(6): e170–e173. doi:10.1016/j.fsigen.2012.05.010
- Pinheiro T (2008) Emigration, immigration and interculturality: the meaning of the European Year of Intercultural Dialogue in Portugal. *Eurolimes* 6:63–73 ISSN: 18419259
- Dias P, Machado R, Oliveira A (2015) Relatório de Imigração, Fronteiras e Asilo. Serviço de Estrangeiros e Fronteiras, Oeiras
- Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques* 10(4):506–513
- Hares DR (2015) Selection and implementation of expanded CODIS Core loci in the United States. *Forensic Sci Int-Gen* 17: 33–34. doi:10.1016/j.fsigen.2015.03.006
- Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10(3):564–567. doi:10.1111/j.1755-0998.2010.02847.x
- Tereba A (1999) Tools for analysis of population statistics. *Profiles in DNA* 9:14–16
- Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Seattle, USA: Department of Genome Sciences, University of Washington
- SPSS IBM (2011) IBM SPSS statistics for Windows, version 20.0. New York, EUA: IBM Corp
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA 6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729