

TESE DE MESTRADO

Maria de Lurdes Pontes Rebelo

Licenciatura em Bioquímica

Aplicação do Estudo de Polimorfismos
do cromossoma y em Genética Forense
y-Strs e Marcadores Bialélicos

Orientadora: Prof. Doutora Maria de Fátima Terra Pinheiro

2003

ICBAS

UNIVERSIDADE DO PORTO

TESE DE MESTRADO

Maria de Lurdes Pontes Rebelo
Licenciatura em Bioquímica

**APLICAÇÃO DO ESTUDO DE POLIMORFISMOS
DO CROMOSSOMA Y EM GENÉTICA FORENSE
Y-STRS E MARCADORES BIALÉLICOS**

Orientadora: Prof. Doutora Maria de Fátima Terra Pinheiro

Dezembro de 2003

**ICBAS
UNIVERSIDADE DO PORTO**

Resumo

O estudo de STRs do cromossoma Y, devido à sua especificidade masculina é, actualmente, muito usado no âmbito da Genética e Biologia Forense, fundamentalmente na resolução de casos do âmbito criminal e de investigação de filiação, para além de apresentar interesse noutros campos, nomeadamente em estudos de evolução humana.

Apesar destas potenciais utilizações, também existem sérias limitações que decorrem das propriedades específicas do cromossoma Y, nomeadamente o facto de todos os indivíduos do sexo masculino familiares de um suposto pai/suspeito partilharem a mesma informação relativamente a este cromossoma. Por isso, na conclusão das perícias médico-legais os marcadores deste cromossoma não devem ser utilizados isoladamente, mas em conjunto com outros polimorfismos autossómicos.

Este trabalho consistiu em efectuar o estudo populacional (Norte de Portugal) de *loci* STR e marcadores bialélicos do cromossoma Y, tendo em vista a sua aplicação à resolução de perícias médico-legais do âmbito da Genética e Biologia Forense, particularmente nos casos de filiação, casos criminais e identificação genética individual. Os polimorfismos analisados foram oito “Short Tandem Repeats” do cromossoma Y (Y-STRs): DYS19, DYS389I/II, DYS390, DYS391, DYS392, DYS393 e DYS385I/II e sete marcadores bialélicos, seis “Single Nucleotide Polimorphisms” (SNPs): sY81, SRY-8299, M9, DYS257, SRY-2627, SRY-1358 e uma inserção *Alu*, o YAP, na população do Norte de Portugal. Os primeiros demonstraram ser os mais úteis em Genética e Biologia Forense, devido aos seus elevados Índices de Diversidade e Poder de Discriminação, à sua relativamente baixa taxa de mutação e à simplicidade técnica de caracterização dos respectivos alelos.

As comparações populacionais realizadas, permitiram inferir que partilhamos vários haplótipos (Y-STRs) com as restantes populações da Península Ibérica e que, por sua vez, a referida partilha decresce progressivamente em relação a outras populações europeias. Não há haplótipos comuns entre a nossa população e as populações da Índia, China, Japão e Coreia. Relativamente aos haplogrupos aqui encontrados (marcadores bialélicos), as comparações populacionais efectuadas permite-nos agrupar a nossa população com as restantes populações Ibéricas e algumas populações europeias, confirmando a existência de uma origem comum das linhagens das populações do Ocidente da Europa.

Para futuro deve-se investir na análise de um maior número de marcadores em simultâneo e, se possível, acrescentar novos sistemas, de modo que numa única análise se obtenha a maior informação possível, obviando-se tempo despendido na análise e quantidade de amostra consumida. Seria conveniente também aliar o estudo dos marcadores bialélicos ao dos Y-STRs pois deste modo discriminava-se melhor as linhagens de cromossomas Y da nossa população.

Abstract

The interest in Y chromosome polymorphisms has continuously increased, not only in evolutionary and genealogical studies, but also in forensics. Having in mind this multidisciplinary view, we have studied the *loci* DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 and DYS385I/II in a population sample of 122 individuals, from the North of Portugal, involved in paternity testing and some biallelic markers in 99 individuals from the same population: sY81, SRY-8299, M9, DYS257, SRY-2627, SRY-1358 and the *Alu* insertion.

Haplotypes and haplogroups were determined and several statistic parameters were calculated using Arlequin software ver. 2000: gene diversity, F_{ST} values and genetic differentiation. Genetic distances were also calculated and the correspondent dendrograms and the MDS graphics debugged.

The analysis of these Y-STR markers showed 111 unique haplotypes. The two most frequent haplotypes were repeated three times and seven others only two times. Consequently, the value of the accumulated haplotypic diversity was $0,998 \pm 0,001$, meaning that the probability that two individuals taken randomly from this population share the same haplotype is 0,002. The Y-STRs are the best markers for discriminate individuals.

Comparisons of our population with others from Iberia, Europe and Asia, showed that we share many haplotypes with Iberia and the number of shared haplotypes decreases progressively with other European populations. None of the haplotypes is shared by North of Portugal and India, China or Japan. AMOVA analysis shows that the biggest percentage of variance occurs within populations.

We can conclude from the data obtained with the Y-STRs and Biallelic markers studied that our population can be phylogenetically included in the Y chromosome lineages of Europe. Additionally we can reinforce the forensic utility of these markers, namely in sexual assaults.

In the future we pretend to combine the information obtained with the study of Y-STRs and biallelic markers and perhaps use some high throughput technologies for typing SNPs more efficiently like Real Time PCR, in order to achieve a more effective discrimination between male individuals.

Résumé

L'intérêt sur le chromosome Y a augmenté dans les dernières années, à cause de ces caractéristiques de transmission, dans les études anthropologiques et dans la médecine légale. En regardant ça, on a étudié les *loci* DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 et DYS385I/II et les marqueurs biallelics sY81, SRY-8299, M9, DYS257, SRY-2627, SRY-1358 et la insertion *Alu*, sur deux groups distincts de personnes du Nord du Portugal, impliqués en cas de paternité.

L'analyse des haplotypes et des haplogroups obtenus, a permis de déterminer quelques paramètres statistiques qu'on a calculés avec le Arlequin software ver. 2000: « gene diversity », valeurs de F_{ST} et la différenciation génétique. Les distances Génétiques ont été calculées pour dessiner le dendrograms correspondant et les « MDS » graphiques.

Avec l'analyse de ces Y-STR marqueurs dans 122 individus on a trouvés 111 unique haplotypes. Le plus fréquent haplotype a apparu trois fois et les autres sept n'ont apparu que deux fois. Comme, la valeur de la diversité haplotypique est de $0,998 \pm 0,001$; la probabilité que deux hommes de la population du Nord du Portugal aient le même haplotype est de 0,002. Les Y-STRs sont les meilleurs marqueurs pour distinguer les individus.

En comparant notre population avec des autres de la Péninsule Ibérique, de l'Europe et d'autres populations on conclut que, nous partageons les haplotypes et les haplogroups avec les populations européennes. Nous pouvons alors conclure qu'on appartient aux lignages des chromosomes Y du reste de l'Europe. En plus, nous pouvons observer

l'utilité de ces marqueurs dans le monde criminel, surtout dans le cas des agressions sexuelles.

Pour l'avenir on prétend combiner l'information obtenue avec l'étude de Y-STRs et les biallelic marqueurs, pour permettre de discriminer plus efficacement les individus.

Agradecimentos

À Comissão Coordenadora do Mestrado em Medicina Legal, em especial ao Sr. Professor Doutor J. Pinto da Costa, por me ter permitido e incentivado à frequência naquele que foi o primeiro Mestrado na área da Medicina-Legal.

À Prof. Doutora Maria de Fátima Pinheiro, que aceitou ser minha orientadora, por me ter apoiado pessoal e profissionalmente e por todos os seus conselhos que me ajudaram a elaborar da melhor forma possível este trabalho.

À Prof. Doutora Marian de Pancorbo e à Doutora Azucena Castro, pelo apoio na parte Estatística e pela amizade, carinho e paciência que sempre me dispensaram; bem hajam!

Ao Professor Doutor José Luis Blázquez Caeiro e ao Doutor Juan Carlos Carril Juanatey, por me terem aturado e pelos seus ensinamentos na área antropológica.

A todos os meus colegas do Laboratório de Genética e Biologia Forense: Maria João, David, Paula, Gabriela e Maria José, pela ajuda no trabalho propriamente dito e pela compreensão nos momentos de maior ansiedade.

Aos meus: Miguel Ângelo e Cláudia Beatriz, os que mais sentiram as minhas falhas, à minha mãe que dentro das suas possibilidades me apoiou incondicionalmente e ao Mário pela sua paciência e apoio. Espero poder compensá-los por tudo o que sofreram!

A todos os que por ventura me ajudaram de outras formas e que não mencionei, o meu agradecimento sincero.

ÍNDICE

1. OBJECTIVOS	1
2. INTRODUÇÃO	5
2.1. <i>Estrutura</i>	7
2.2. <i>Genes e Funções biológicas</i>	9
2.3. <i>Características especiais do cromossoma Y</i>	10
2.4. <i>Primeiras utilizações forenses</i>	12
2.5. <i>Polimorfismos do cromossoma Y: Y-STRs</i>	13
2.5.1. <i>Seleção dos STRs para este estudo</i>	14
2.5.1.1. <i>DYS19</i>	15
2.5.1.2. <i>DYS389I/II</i>	16
2.5.1.3. <i>DYS390</i>	16
2.5.1.4. <i>DYS391</i>	17
2.5.1.5. <i>DYS392</i>	17
2.5.1.6. <i>DYS393</i>	18
2.5.1.7. <i>DYS385</i>	18
2.5.2. <i>Aplicações dos STRs do cromossoma Y em Genética e Biologia Forense</i>	19
2.5.2.1. <i>Testes de paternidade e criminalística</i>	19
2.5.2.2. <i>Recomendações da ISFG relativas à utilização do cromossoma Y</i>	22
2.5.2.3. <i>Legislação Portuguesa</i>	23
2.5.3. <i>Aplicações em estudos de evolução humana</i>	24
2.6. <i>Polimorfismos do cromossoma Y humano: Marcadores bialélicos</i>	27
2.6.1. <i>Abundância</i>	28
2.6.2. <i>Recorrência</i>	29
2.6.3. <i>Nomenclatura</i>	29
2.6.4. <i>Especificidade populacional dos haplogrupos</i>	32
2.6.5. <i>Aplicação na Genética e Biologia Forense</i>	34
2.6.6. <i>Estratégias de tipagem</i>	34
2.6.7. <i>Polimorfismos bialélicos utilizados neste estudo</i>	40
2.6.7.1. <i>YAP (DYS287)</i>	40
2.6.7.2. <i>sY81 (DYS271 ou M2)</i>	41

2.6.7.3. <i>SRY-8299 (SRY-4064)</i>	42
2.6.7.4. <i>M9</i>	42
2.6.7.5. <i>DYS257</i>	43
2.6.7.6. <i>SRY-1532 (SRY-10831)</i>	44
2.6.7.7. <i>SRY-2627 (SRY-2628 ou M167)</i>	45
3. MATERIAL E MÉTODOS	47
3.1. <i>Colheita de amostras</i>	48
3.2. <i>Extracção do DNA</i>	48
3.2.1. <i>Extracção com fenol-clorofórmio-álcool isoamílico</i>	49
3.2.2. <i>Extracção com fenol-clorofórmio-álcool isoamílico e lise diferencial (mistura de células)</i>	51
3.2.3. <i>Extracção com Chelex</i>	54
3.2.4. <i>Extracção de DNA em esfregaços bucais com “Chelex”</i>	55
3.3. <i>Quantificação do DNA</i>	56
3.4. <i>Amplificação do DNA por PCR</i>	57
3.4.1. <i>Y-STRs</i>	57
3.4.2. <i>Marcadores bialélicos</i>	60
3.5. <i>Separação dos fragmentos</i>	63
3.5.1. <i>Y-STRs</i>	63
3.5.2. <i>Y-SNPs e YAP</i>	65
2.6. <i>Cálculo Estatístico</i>	68
2.6.1. <i>Análise intrapopulacional</i>	68
2.6.1.1. <i>Frequências alélicas e haplotípicas</i>	68
2.6.1.2. <i>Diversidade Génica/Haplotípica</i>	69
2.6.2. <i>Comparação de populações</i>	69
2.6.2.1. <i>Número médio de diferenças entre pares (π)</i>	70
2.6.2.2. <i>Estrutura da população: Análise molecular da Variância (AMOVA)</i>	70
2.6.2.3. <i>FSTs linearizados de Slatkin’s</i>	71
2.6.2.3.1. <i>Y-STRs</i>	72
2.6.2.3.2. <i>Marcadores bialélicos</i>	72

4. RESULTADOS	73
4.1. STRs	74
4.1.1. Análise por locus	77
4.1.2. Análise haplotípica e parâmetros com interesse forense	78
4.1.3. Comparações populacionais	78
4.1.4. Aplicação dos Y-STRs na conclusão de perícias médico-legais	84
4.2. Marcadores bialélicos	90
4.2.1. Distribuição das frequências alélicas dos loci SNP e YAP do cromossoma Y analisados	90
4.2.2. Haplogrupos construídos a partir dos loci bialélicos estudados	91
4.2.3. Distâncias genéticas-árvore filogenética	93
5. DISCUSSÃO	98
5.1. Amostragem	99
5.2. Caracterização dos sistemas estudados no cromossoma Y	99
5.3. STRs	100
5.3.1. Diversidade Génica/ Haplotípica	100
5.3.2. Haplótipos encontrados	101
5.3.3. Comparação entre várias populações	102
5.4. Polimorfismos bialélicos	104
5.4.1. Diversidade Génica e frequências dos Haplogrupos	104
5.4.2. Diferenciação populacional e agrupamento das populações	106
5.4.3. História evolutiva baseada nos polimorfismos do cromossoma Y	108
6. CONCLUSÕES	112
7. BIBLIOGRAFIA	116

1. OBJETIVOS

Neste trabalho analisaram-se as linhagens paternas do cromossoma Y e, posteriormente, compararam-se com outras populações para determinação de ancestrais comuns. As linhagens paternas foram estudadas mediante análise dos polimorfismos bialélicos e microsatélite (STRs) da região não recombinante do cromossoma Y, de herança masculina. Este estudo permitiu sobretudo calcular o poder de discriminação dos referidos polimorfismos, com vista á sua aplicação na área da Genética e Biologia Forense.

1. Objectivos do trabalho:

O principal objectivo deste trabalho consistiu em efectuar o estudo populacional (Norte de Portugal) de *loci* STR e marcadores bialélicos do cromossoma Y, tendo em vista a sua aplicação à resolução de perícias médico-legais do âmbito da Genética e Biologia Forense, particularmente nos casos de filiação, casos criminais e identificação genética individual. Os polimorfismos analisados foram oito “Short Tandem Repeats” do cromossoma Y (Y-STRs): DYS19, DYS389I/II, DYS390, DYS391, DYS392, DYS393 e DYS385I/II e sete marcadores bialélicos, seis “Single Nucleotide Polimorphisms” (SNPs): sY81, SRY-8299, M9, DYS257, SRY-2627, SRY-1358 e uma inserção *Alu*, o YAP, na população do Norte de Portugal.

O estudo de STRs do cromossoma Y, devido à sua especificidade masculina, tem-se revelado de grande importância nesta área, para além de apresentar interesse noutros campos, nomeadamente em estudos de evolução humana.

No âmbito da investigação de paternidade, nos casos em que o filho é do sexo masculino, permite excluir a paternidade de um indivíduo falsamente acusado de ser o pai biológico, mesmo nos casos em que este está ausente. Nesta situação, pode-se analisar indivíduos do sexo masculino, com ele aparentados pela linha paterna como, por exemplo, um irmão, um tio ou mesmo o avô (Kayser *e col.*, 1997; Jobling *e col.*, 1997; Roewer *e col.*, 1996).

Relativamente à criminalística biológica, o estudo dos STRs deste cromossoma reveste-se de grande importância, porque os autores de crimes violentos, na grande maioria dos casos, são indivíduos do sexo masculino (Roewer, 1998). Nos crimes sexuais, quando

há mistura de células masculinas e femininas, a análise dos polimorfismos deste cromossoma proporciona informação acrescida em relação aos polimorfismos dos autossomas, porque fornece dados específicos do agressor, sem ser necessário proceder à separação das fases vaginal e espermática (método de lise diferencial); para além de que, quando a agressão sexual tiver sido praticada por vários indivíduos, esta análise permite determinar o número de contribuidores na mistura de células analisada. Outra aplicação, é a possibilidade de se detectar células epiteliais masculinas em ejaculados de indivíduos vasectomizados.

No que respeita à identificação biológica individual, a caracterização dos Y-STRs tem um interesse considerável, podendo-se determinar de imediato o sexo e, nos casos em que o investigado não esteja disponível e possua ascendência e/ou descendência masculina, seguir-se a sua linhagem, sendo possível, coadjuvado com outros dados, fazer-se uma identificação positiva.

Para alcançar o objectivo proposto, seguiu-se a seguinte estratégia:

- 1- Estudaram-se amostras da população do Norte de Portugal, na qual se efectuaram testes de filiação, tendo-se incluído apenas indivíduos fenotipicamente sãos e não aparentados.
- 2- Caracterizaram-se as linhagens do cromossoma Y, tanto Y-STRs como marcadores bialélicos, Y-SNPs e a inserção YAP, tendo-se usado diferentes metodologias para os distintos marcadores analisados. Assim, a metodologia utilizada para o estudo dos Y-STRs, envolveu o uso de várias técnicas de extracção de DNA, seleccionadas de acordo com a natureza das amostras analisadas, bem como a amplificação do DNA, por PCR, e a separação dos produtos amplificados por electroforese capilar. Em relação aos marcadores bialélicos, depois de se ter realizado a extracção do DNA, escolheu-se a técnica, também de acordo com as características das amostras usadas, tendo-se procedido à digestão enzimática, seguida de electroforese em géis de agarose com brometo de etídio.
- 3- Determinaram-se os haplótipos (Y-STRs) da população em estudo e relacionaram-se com os de outras populações da Península Ibérica, Europa e Ásia. Não se fizeram

comparações com populações africanas, por não existirem dados disponíveis. Efectuou-se o mesmo relativamente aos haplogrupos (marcadores bialélicos), desta feita comparando populações das mais distintas origens.

- 4- Calculou-se o poder de discriminação dos diversos polimorfismos, tendo em vista a identificação genética na resolução de perícias médico-legais, depois de se ter determinado as frequências dos haplótipos e dos haplogrupos e as respectivas diversidades.

2. INTRODUÇÃO

2. Introdução

Os indivíduos normais, da espécie humana e do sexo masculino, possuem um cromossoma Y e os do sexo feminino não. O cromossoma Y determina a masculinidade e o desenvolvimento dos testículos, numa fase precoce da embriogénese. Este papel especializado confere-lhe propriedades únicas que chamaram a atenção de especialistas na história da genética humana.

O estudo do cromossoma Y é, actualmente, muito usado no âmbito da Genética e Biologia Forense, fundamentalmente na resolução de casos do âmbito criminal e de investigação de filiação. Relativamente à primeira aplicação, como já foi referido nos objectivos do trabalho, a grande maioria dos crimes violentos é cometida por homens (mais de 90%), por este motivo a análise de amostras deixadas pelo agressor na cena do crime, podem ser informativas quando caracterizadas para os marcadores do cromossoma Y (Jobling *e col.*, 1997). Quando existe mistura de DNA de células femininas com células masculinas, como na maioria dos crimes sexuais, a análise destes marcadores proporciona informação específica acerca do agressor.

Quanto aos casos de filiação, concretamente as perícias relativas à investigação de paternidade, quando o filho é do sexo masculino, comparando a informação genética dos cromossomas Y do suposto pai e do filho, poder-se-á excluir aquele da paternidade que lhe é atribuída, com alguma facilidade.

Apesar destas potenciais utilizações, também existem sérias limitações que decorrem das propriedades específicas do cromossoma Y (Jobling *e col.*, 1997), nomeadamente o facto de todos os indivíduos do sexo masculino familiares de um suposto pai/suspeito partilharem a mesma informação relativamente a este cromossoma.

Embora este estudo se limite ao cromossoma Y, na conclusão das perícias médico-legais os marcadores deste cromossoma não devem ser utilizados isoladamente, mas em conjunto com outros polimorfismos autossómicos que obviam muitos dos problemas que se descreverão posteriormente.

2.1. Estrutura

O cromossoma Y humano representa apenas 2% do genoma. Através de análise citológica verifica-se que este é um cromossoma de pequeno tamanho – o segundo mais pequeno (aproximadamente 60 Mb de comprimento), com uma parte do braço longo composto por heterocromatina polimórfica, cujo tamanho varia de indivíduo para indivíduo. É nesta região que têm sido descobertos vários polimorfismos de tamanho do tipo microsatélite (STRs), com interesse médico-legal. A parte que resta (eucromatina) tem cerca de 30 megabases, é de tamanho constante em homens normais, compreendendo o braço curto (Yp), centrómero e o braço longo proximal, sendo esta a de maior interesse do ponto de vista genético (Figura 1).

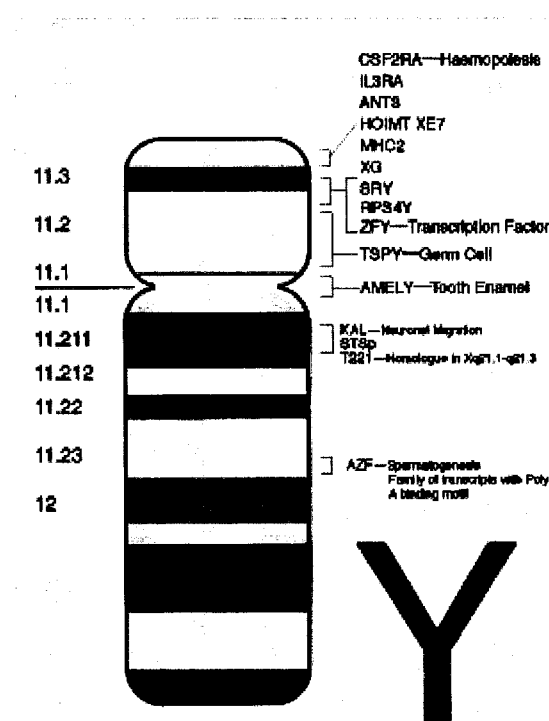


Figura 1. Representação esquemática do cromossoma Y humano, com alguns dos seus genes assinalados.

A região eucromática consiste em quatro classes de sequências: blocos de sequências homólogas ao cromossoma X, sequências repetidas no centrómero e na sua proximidade, “famílias” de sequências repetitivas específicas do cromossoma Y e sequências de cópia única, também específicas deste cromossoma (Lutz Roewer, 1998).

A maior parte deste cromossoma (95%) não recombina (região não recombinante ou NRY) encontrando-se sempre no estado haplóide, de tal modo que é herdada em bloco, ou seja, como um único grupo de ligamento. Esta região constitui o maior bloco não recombinante de todo o genoma humano (YCC, 2002). Apenas umas pequenas porções distais (PAR1 e PAR2), designadas por regiões pseudoautossómicas (2,7 Mb), apresentam homologia com o cromossoma X, sendo responsáveis pelo correcto emparelhamento entre os dois cromossomas sexuais, dando lugar, pelo menos, a uma recombinação em cada meiose (Figura 2.). Portanto, a transmissão destas regiões é semelhante à dos autossomas e, por este facto, não será tratada aqui.

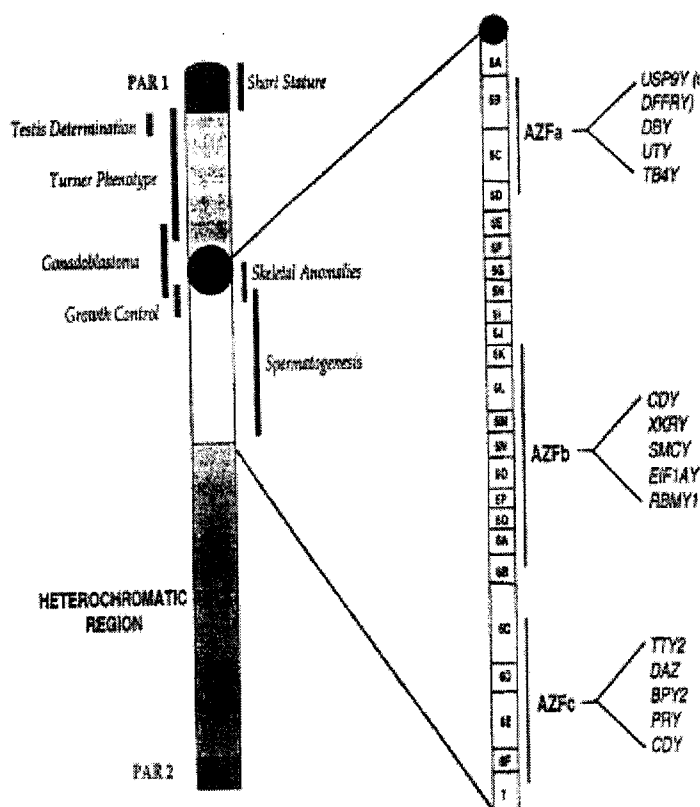


Figura 2. Representação esquemática das várias regiões de interesse do cromossoma Y: região heterocromática, eucromática e regiões PAR 1 e 2.

Tem sido defendida a hipótese dos cromossomas X e Y terem tido uma ancestralidade comum, pois uma grande parte das sequências dos dois cromossomas é idêntica, de tal modo que o cromossoma Y pode ser considerado como um cromossoma X degradado, depois de ter sofrido reordenamento cromossómico durante a evolução, incluindo translocações, tendo sido eliminada a recombinação de uma grande parte do cromossoma e protegidos os genes determinantes do sexo (Graves, 1995).

Após a sequenciação completa do cromossoma Y humano, cujos resultados foram publicados, em Junho do ano corrente, na revista *Nature*, algumas das teorias sobre a origem do cromossoma Y, entre elas a anteriormente referida, foram postas em causa. A hipótese que, actualmente, é considerada mais consistente é que os cromossomas X e Y evoluíram a partir de um mesmo par autossómico. A diferenciação terá ocorrido há cerca de 300 milhões de anos; no entanto, através da sequenciação do cromossoma Y, constatarem existir uma apreciável semelhança entre este cromossoma e o X. Para além disso, descobriram ainda que 95% do cromossoma Y está dividido em três regiões eucromáticas e uma heterocromática. Dessas regiões eucromáticas, a primeira é idêntica à região equivalente do X, a segunda é degenerada em relação ao ancestral autossómico comum ao X e ao Y e a terceira é uma região palindrómica (sequências iguais, mas invertidas). As duas últimas regiões são as maiores e mais ricas em genes, a maioria dos quais codifica para proteínas, que se expressam sobretudo nos testículos.

2.2. Genes e funções biológicas

Na espécie humana os heterocromossomas ou cromossomas sexuais têm estruturas e conteúdos bem diferenciados. O cromossoma Y humano tem importantes funções biológicas específicas do sexo masculino, com consequências directas na aptidão (“fitness”), que incluem determinação testicular e fertilidade (Quintana-Murci e *col.*, 2001). A simples presença de um cromossoma Y determina, por si mesmo, o desenvolvimento de um indivíduo varão, independentemente do número de cromossomas X que possua.

Inicialmente, pensou-se que o sexo nos humanos se devia simplesmente à presença de um ou mais genes que confeririam a masculinidade. Neste sentido, foi descoberto o SRY (Sex Reversal Y), o gene mais bem conhecido, que se localiza no braço curto, na região mais próxima do limite pseudoautossómico (PAR1), sendo o factor que determina o desenvolvimento testicular (Su e Lau, 1993). É um gene de cópia única, sem homólogo no cromossoma X, pequeno e com um único exão. Pensou-se que, na sua ausência, se desenvolveria um indivíduo do sexo feminino. No entanto, cedo se verificou que a situação era mais complexa. As mutações no gene SRY estão associadas com um fenótipo feminino estéril. É o gene que regula a cascata da determinação testicular, codificando para uma proteína de 233 amino-ácidos, que se expressa nos

testículos dos adultos, entre outros tecidos, e está representada na figura 3, não se conhecendo, com precisão, o mecanismo do seu funcionamento.

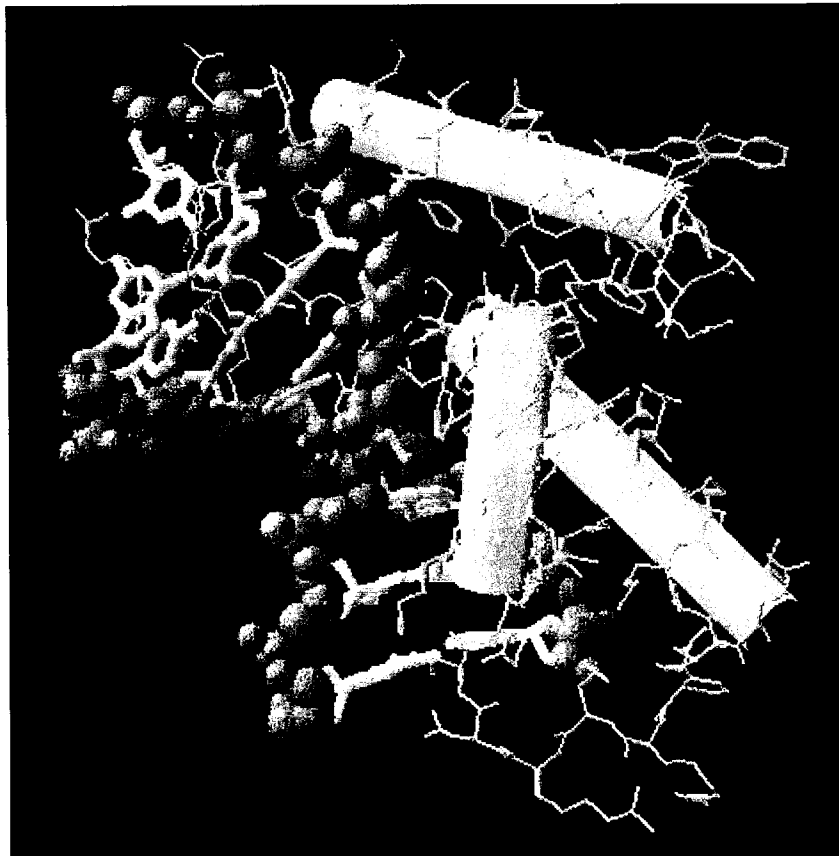


Figura 3. Proteína codificada pelo gene SRY.

2.3. Características especiais do cromossoma Y

O cromossoma Y tendo uma função determinante do sexo, é herdado de uma forma haplóide uniparental paterna. Devido ao seu estado haplóide, a maioria do cromossoma não recombina com nenhum outro, durante a meiose, como já foi referido.

Estas propriedades têm consequências importantes (construção de haplótipos individuais, ancestralidade paterna, fácil estimativa da separação temporal dos acontecimentos que levam ao aparecimento de ramos nas árvores filogenéticas, entre outros) na compreensão da genética populacional, uma vez que o cromossoma Y passa de pais para filhos sem alterações, exceptuando a gradual acumulação de mutações. Nesta perspectiva, o estudo dos polimorfismos deste cromossoma tem maior interesse do que o dos sistemas autossómicos.

Isto significa que todos os cromossomas Y contêm em si uma “gravação” dos acontecimentos mutacionais que ocorreram em todos os cromossomas ancestrais anteriores. Por isso, em princípio, é possível reconstruir a história das linhagens paternas por comparação dos cromossomas Y modernos, usando a análise de polimorfismos de DNA (STRs e sistemas bialélicos), sendo viável a construção de árvores filogenéticas e, através delas, deduzir o passado evolutivo das populações humanas actuais.

Tem sido descrito um número crescente de marcadores polimórficos informativos do cromossoma Y, que incluem substituições de bases, inserções/deleções, microsátélites e minisátélites, úteis em estudos populacionais e na resolução de perícias médico-legais. Destes, muitos podem ser estudados por PCR (Jobling *e col.*, 1997).

Os microsátélites do cromossoma Y (Y-STRs) definem a variabilidade (haplótipos) dentro dos haplogrupos, permitindo calcular as diferenças entre distintas populações humanas (distâncias genéticas), com maior precisão. Acresce ainda a possibilidade de, através delas, se poder estimar os tempos de coalescência de haplogrupos bem definidos.

As variantes bialélicas dos SNPs e dos polimorfismos de inserção/delecção, representam, em geral, eventos moleculares únicos na evolução humana. E, embora em alguns casos possam ser características de pequenos grupos ou indivíduos, podem definir haplogrupos relacionados entre si pela mesma raiz filogenética com grande distribuição geográfica ou, em alguns casos, associados a uma região particular.

2.4. Primeiras utilizações forenses

Uma das primeiras aplicações do cromossoma Y no âmbito médico-legal, foi a determinação do sexo para fins de identificação individual (Witt e Erickson, 1989), o que constituiu um elemento importante na análise de vestígios biológicos. De facto, obter informação acerca da existência de células masculinas, a partir de DNA extraído de exsudatos vaginais, fornecia um resultado preliminar positivo, ao qual se pode, actualmente, seguir a amplificação de vários *loci* informativos. Neste sentido, foram

desenvolvidos vários métodos rápidos e precisos para a determinação do sexo em vestígios biológicos de origem humana.

Uma das metodologias usadas no primeiro teste para diagnóstico do sexo (Lau *e col.*, 1984), empregava sondas específicas para sequências repetidas da região não codificante (heterocromatina) do braço longo do cromossoma Y (Yq) (Cooke, 1976). Esta técnica tinha algumas limitações, nomeadamente a de não detectar cromossomas Y que não possuíam esta região do cromossoma (apresentavam uma deleção), o que pode acontecer em indivíduos normais e férteis. Por outro lado, esta região pode também estar presente em mulheres, fenotipicamente normais por translocação. Além disso, podem também estar ausentes fragmentos do Yq em indivíduos com o Síndrome de Turner. Por conseguinte, utilizando-se estas sondas, pode-se incorrer num elevado risco de diagnóstico incorrecto do sexo.

O único conjunto de sequências repetidas que parece exibir propriedades de especificidade cromossómica significativa é a “Alphoid Satellite Family”, localizada nas regiões peri-centroméricas de todos os cromossomas humanos (Willard e Wayne, 1987). Estas sequências, no cromossoma Y, têm uma organização repetitiva superior, detectável com enzimas de restrição, gerando um fragmento de tamanho característico. Sabe-se que existem cerca de 100 cópias desta sequência neste cromossoma, tendo sido descrita uma técnica que usa uma sonda específica para detecção destas repetições alfoídes (Stalvey e Erickson, 1987; Gaensslen *e col.*, 1992).

Depois da descoberta da PCR, esta técnica começou a ser utilizada para a determinação do sexo, em amostras forenses (Witt e Erickson, 1989; Nørby S e Eriksen B, 1992). Foi usada em primeiro lugar para estudar as sequências alfa ou alfoíde (Pascal *e col.*, 1993; Pfitzinger *e col.*, 1993) que existem em diferentes números de cópias no cromossoma X (cerca de 5000) e no cromossoma Y (cerca de 100). Posteriormente, a PCR foi usada para a detecção do *locus* DYZ1 (milhares de cópias: 500-8000 no cromossoma Y) (Ludes *e col.*, 1991) ou para a detecção do *locus* da amelogenina (a proteína do esmalte mais importante no desenvolvimento da matriz extracelular dos dentes), ou parte dele (Mannuci *e col.* 1994). Trata-se de uma região de cópia única, homóloga para os cromossomas X e Y, mas com tamanhos diferentes.

A análise de um *locus* homólogo no cromossoma X e no cromossoma Y é mais fiável e hoje largamente usada, uma vez que a detecção da banda específica do X confirma a validade do método (Akane *e col.*, 1991). A detecção deste gene homólogo para os dois cromossomas, com tamanhos distintos, tem demonstrado ser de grande utilidade na confirmação da presença de esperma ou de células masculinas (em indivíduos azoospermicos) em exsudatos vaginais colhidos a vítimas de agressão sexual (Mannuci *e col.*, 1994).

O desenvolvimento das técnicas de determinação do sexo prosseguiu no sentido de se obter essa informação, a partir de quantidades reduzidas de DNA (aumento da sensibilidade) extraído de vários tipos de amostras, tais como, sangue, manchas de sangue e de esperma, exsudatos e raízes de cabelos (Pfitzinger *e col.*, 1993; Tanaka, 1992), aliadas à detecção fluorescente dos produtos de PCR.

A partir da segunda metade da década de 90, a caracterização de vários STRs, todos eles na parte não recombinante do cromossoma Y, no mesmo indivíduo, permite reconstruir o seu haplótipo (diversos polimorfismos herdados em bloco), dado que só existe um destes cromossomas, que não sofre recombinação, em cada indivíduo.

2.5. Polimorfismos do cromossoma Y: Y-STRs

Têm sido descritos múltiplos polimorfismos na parte não recombinante do cromossoma Y (Kayser *e col.* 1997, de Knijff *e col.*, 1997). A estrutura e a relação que existe entre repetição/alelo, são comparáveis com outros sistemas equivalentes autossómicos. Excepcionalmente, a análise por PCR amplifica, em alguns sistemas, dois produtos alélicos como por exemplo, o DYS385 e o DYS389I/II. Este elevado número de sequências duplicadas, reflecte uma peculiaridade morfológica do cromossoma Y, uma vez que não foram descritos em autossomas *loci* com estas características, até ao momento. Estes *loci* que geram duas bandas estão localizados no Yq. Outra observação interessante foi que a duplicação não é o nível mais elevado de multiplicidade dos alelos atribuídos aos Y-STRs, uma vez que parecem ter sido encontrados padrões triplicados nos *loci* DYS385 e DYS19.

Relativamente ao DYS389I/II, a estrutura foi satisfatoriamente clarificada (Cooper *e col.*, 1996), porquanto um local de hibridação do “primer” (iniciador) está duplicado, de tal modo que o primer 5’ (“forward”) hibrida duas vezes na mesma cadeia e, por esse motivo, surgem dois produtos de amplificação, ambos contendo repetições em número variável (de Kniff *e col.*, 1997), sendo possível atribuir os respectivos alelos a cada *locus*, pois os produtos da PCR aparecem em intervalos de tamanho diferentes (239-263bp e 353-385bp), diferindo em pelo menos 100bp.

Em contraste, as repetições e as sequências flanqueadoras do DYS385, terão sido, provavelmente, duplicadas na totalidade como inserções de blocos de sequências longas das quais resultou a existência de mais de um *locus*. Neste caso, os alelos não podem ser inequivocamente atribuídos a um ou outro *locus*, devido à sobreposição de tamanhos. Por este motivo, o resultado da sua caracterização aparece como se fossem haplótipos constituídos por dois alelos.

A análise das sequências dos Y-STRs mono-localizados revelou unidades de repetição tetraméricas (DYS19, DYS389I/II, DYS390, DYS391, DYS393), triméricas (DYS388, DYS392) e diméricas (DYS288) (de Knijff *e col.*, 1997). Os sistemas bi-localizados, demonstraram ser diméricos (YCAI-III, Mathias *e col.*, 1994) e tetraméricos (DYS385, Schneider P *e col.*, 1997).

Os *loci* DYS389II e DYS390 são compostos, possuindo duas unidades de repetição diferentes. Os restantes são sistemas simples, ou seja, são constituídos por uma única unidade de repetição.

2.5.1. Seleção dos STRs para este estudo

Aquando do início deste trabalho, tinham sido descritos um conjunto de STRs por vários autores (Kayser *e col.*, 1997; de Knijff *e col.*, 1997) que realizaram um estudo em colaboração, os quais poderiam ser úteis na resolução de perícias médico-legais proporcionando, no seu conjunto, um elevado poder discriminativo. Deste modo, depois de seleccionados alguns deles, efectuou-se o estudo da nossa população (Norte de Portugal), tendo-se acrescentado o DYS385 (Schneider *e col.*, 1998) que, por ser um

dos *loci* “duplicados”, aumenta o poder informativo. Os marcadores escolhidos foram: DYS19, DYS389I/II, DYS390, DYS391, DYS392, DYS393 e DYS385I/II (Tabela 1).

Este conjunto de polimorfismos constitui o chamado haplótipo mínimo de uma base de dados de referência criada na Europa (Y-STR Haplotype Reference Database), graças ao esforço de um grupo de trabalho denominado “The International Forensic User Group”, depois de participarem, com sucesso, num exercício de qualidade obrigatório (Roewer *et al.*, 2001).

Tabela 1. STRs estudados no cromossoma Y

STR	Nº alelos	Intervalo de repetições	Intervalo de tamanhos (bp)
DYS19	10	10-19	174-210
DYS389I	7	7-13	235-263
DYS389II	9	23-31	355-387
DYS390	8	20-27	191-227
DYS391	6	8-13	271-299
DYS392	11	8-18	233-263
DYS393	6	12-17	108-136
DYS385	14	8-21	352-412

Nota: Estes intervalos de tamanho dos alelos são relativos aos “primers” utilizados neste estudo. Quanto ao *locus* DYS389I/II, foi incluído um bloco (GATA)3 na contagem do número de repetições (+3 do que a nomenclatura sugerida por Kayser *et al.*, 1997).

2.5.1.1. DYS19

GDB Access: X77751

Primer 1: *5'CTA CTG AGT TTC TGT TTA TAG T3'

Primer 2: 5'ATG GCA TGT AGT TGA GGA CA3'

Unidade de repetição consenso: GATA.

Foi o primeiro STR descrito no cromossoma Y humano, tendo sido, inicialmente, denominado de diferentes formas: Y27H39, DYS394 ou CHLC.GATA 73F02.P18616. Trata-se de uma repetição tetranucleotídica com um motivo de repetição GATA, com duas regiões distintas: 3 repetições do “core” (GATA)₃ invariável, seguidas de uma estrutura variável (GATA)_n. Esta variação pode ocorrer entre 10 e 19 repetições (Roewer *et al.*, 1992).

A estrutura repetitiva pode estar deslocada uma base da extremidade 5', originando um motivo repetitivo AGAT, ou de duas bases da referida extremidade, dando lugar a um motivo TAGA, sem alterar o número de repetições.

2.5.1.2. *DYS389I/II*

GDB Access: G09600

Primer 1: *5'CCA ACT CTC ATC TGT ATT ATC TAT 3'

Primer 2: 5'TCT TAT CTC CAC CCA CCA GA 3'

Unidade de repetição consenso: (GATA) n (GACA) n.

É um STR tetramérico composto, situado no *locus* DYS389, também denominado CHLC.GATA30F10.P14912 (Murray, 1993). No caso do DYS389I, compõe-se por dois motivos de repetição adjacentes do tipo (TCTG) n (TCTA) n, variando de 7 a 13 repetições; no do DYS389II, é formado por quatro motivos de repetição, do tipo (TCTG) n (TCTA) n (TCTG) n (TATA) n, variando de 23 a 31 repetições.

2.5.1.3. *DYS390*

GDB Access: G09611

Primer 1: *5'TAT TAA TTA CAC ATT TTT GGG CC 3'

Primer 2: 5'TGA CAG TAA AAT GAA CAC ATT GC 3'

Unidade de repetição consenso: (GATA) n (GACA) n.

É um STR tetramérico composto (Murray, 1993), constituído por sete motivos de repetição adjacentes, do tipo (TCTA)₂(TCTG)_n(TCTA)_n(TCTG)_n(TCTA)_n(TCAT)₁

(CTAT)2, dos quais o primeiro e os dois últimos são invariáveis; os quatro restantes, variam no número de repetições, sendo a soma dos quatro motivos variáveis a que se utiliza para identificar as diferentes variantes alélicas.

2.5.1.4. *DYS391*

GDB Access: G09613

Primer 1: *5'CTA TTC ATT CAA TCA TAC ACC CA 3'

Primer 2: *5'GAT TCT TTG TGG TGG GTC TG 3'

Unidade de repetição consenso: (GATA).

Trata-se de uma repetição tetranucleotídica com um motivo TCTA, flanqueado por dois motivos TCTG que apresentam um número de repetições constante (Murray, 1993). Também se denomina por CHL.GATA32C10.P14988 e tem uma estrutura que é representada por (TCTG)3(TCTA)n(TCTG)1, sendo os alelos designados de acordo com o número de repetições do “core” (TCTA)n.

2.5.1.5. *DYS392*

GDB Access: G09867

Primer 1: *5'TCA TTA ATC TAG CTT TTA AAA ACA A3'

Primer 2: 5'AGA CCC AGT TGA TGC AAT GT 3'

Unidade de repetição consenso: (ATT).

É o único STR trinucleotídico simples (Murray J C, 1993) utilizado neste estudo e também é denominado por CHLC.ATA25F04.P16979, cujo motivo de repetição é (TAT)n.

2.5.1.6. *DYS393*

GDB Access: G09601

Primer 1: *5'GTG GTC TTC TAC TTG TGT CAA TAC 3'

Primer 2: *5'AAC TCA AGT CCA AAA AAT GAG G 3'

Unidade de repetição consenso: (GATA).

É um STR tetranucleotídico simples (Murray, 1993), também designado por CHLC.GATA43GO6.P14915, DYS395 e CHLC.GATA 81DO5.P18481, cuja unidade de repetição é (GATA) n, embora a ISFG (International Society of Forensic Genetics) recomende que a leitura se efectue uma base antes, relativamente à extremidade 5' (Bär W e col., 1997). Ou seja, a sequência repetitiva passa a ser (AGAT)n; no entanto, o número de repetições mantém-se o mesmo usando uma ou outra unidade de repetição.

2.5.1.7. *DYS385*

GDB access: Z93950 ("Chromosome Y repeat regions")

Primer 1: *5'AGC ATG GGT GAC AGA GCT A 3'

Primer 2: *5'GGG ATG CTA GGT AAA GCT C 3'

Unidade de repetição consenso: (GAAA).

Trata-se de um *locus* "duplicado", à semelhança do que ocorre com o DYS389I/II, mas, neste caso, não se conseguem, com os primers descritos, distinguir os alelos de cada um deles, pelo que são designados como se se tratasse de um par alélico.

Na tabela 2 resumem-se as características referidas para os Y-STRs.

Tabela 2. Características dos Y-STRs

STR	Nº de acesso GDB	Unidade de repetição consenso	Nº de repetições	Primers
DYS19	X77751	GATA	10 – 19	*5'CTA CTG AGT TTC TGT TTA TAG T3' 5'ATG GCA TGT AGT TGA GGA CA3'
DYS389I/II	G09600	(GATA)n(GACA)n	7-13/23-31	*5'CCA ACT CTC ATC TGT ATT ATC TAT 3' 5'TCT TAT CTC CAC CCA CCA GA 3'
DYS390	G09611	(GATA)n(GACA)n	20-27	*5'TAT TAA TTA CAC ATT TTT GGG CC 3' 5'TGA CAG TAA AAT GAA CAC ATT GC 3'
DYS391	G09613	GATA	8-13	*5'CTA TTC ATT CAA TCA TAC ACC CA 3' 5'GAT TCT TTG TGG TGG GTC TG 3'
DYS392	G09867	ATT	8-18	*5'TCA TTA ATC TAG CTT TTA AAA ACA A3' 5'AGA CCC AGT TGA TGC AAT GT 3'
DYS393	G09601	GATA	12-17	*5'GTG GTC TTC TAC TTG TGT CAA TAC 3' 5'AAC TCA AGT CCA AAA AAT GAG G 3'
DYS385	Z93950	GAAA	8-21	*5'AGC ATG GGT GAC AGA GCT A 3' 5'GGG ATG CTA GGT AAA GCT C 3'

2.5.2. Aplicações dos STRs do cromossoma Y em Genética e Biologia Forense

2.5.2.1. Testes de paternidade e criminalística

O primeiro polimorfismo do cromossoma Y foi descrito por Casanova *e col.* e por Lucotte e Ngo, em 1985.

As técnicas de identificação na área da Genética e Biologia Forense, a partir da década de 1990, foram dominadas pela aplicação de sistemas STR ligados aos autossomas. Relativamente aos Y-STRs, apenas tinham sido descritos três *loci* diméricos, um tetramérico e um pentamérico e, por este facto, a sua aplicação à resolução de casos forenses foi muito restrita. Destes STRs, apenas o DYS19 (tetramérico) tinha sido utilizado no âmbito forense e introduzido em vários laboratórios, como marcador usado por rotina. Uma das hipotéticas razões para o uso limitado dos polimorfismos do cromossoma Y é a ocorrência em menor frequência do que nos autossomas, cerca de 1:4 (Kayser, 1997).

A análise dos haplótipos do cromossoma Y, particularmente os que se definem através da variabilidade microsatélite, tem sido largamente utilizada na resolução de casos do âmbito criminal, como na identificação de DNA masculino em casos que envolvem

manchas com mistura de células femininas e masculinas e em testes de paternidade, particularmente quando o suposto pai é falecido e o pretenso filho é do sexo masculino. Esta circunstância é devida ao facto dos STRs apresentarem um elevado grau de polimorfismo e, conseqüentemente, oferecem um significativo poder de discriminação entre indivíduos (Jobling *e cols.*, 1997; Kayser *e cols.*, 1997).

No âmbito da criminalística, uma vez que a maior parte dos agressores são indivíduos do sexo masculino, a capacidade de se detectar e discriminar o DNA masculino e a informação que se obtém a partir dos Y-STRs, torna estes sistemas um complemento vantajoso ao painel de sistemas autossómicos já utilizados.

Nos casos de agressões sexuais, a análise dos Y-STRs oferece uma nova perspectiva. Pois, como já tem sido referido ao longo deste trabalho, estes marcadores aumentam a taxa de sucesso na identificação da fracção masculina, nas misturas de células presentes em fluidos biológicos humanos, onde a lise diferencial não foi bem sucedida ou seria demasiado arriscada, como por exemplo, no caso de amostras muito degradadas ou em que o número de espermatozóides é baixo.

A maior vantagem do uso dos Y-STRs é precisamente na interpretação dos resultados em amostras com misturas de células, no âmbito das agressões sexuais, se estiver envolvido mais de um contribuidor (Prinz *e col.*, 1997).

A caracterização dos Y-STRs tem um interesse considerável em testes de paternidade, especialmente em casos em que o suposto pai não está disponível, em que haja ascendência e/ou descendência masculina. Pois, qualquer indivíduo do sexo masculino da mesma linhagem paterna do suposto pai, possui a mesma informação relativamente ao cromossoma Y (o mesmo cromossoma Y), de modo que, a análise de apenas dois indivíduos – o filho (geralmente criança) e o adulto com ele relacionado – pode resolver o caso.

Teoricamente, a existência de apenas uma diferença num *locus* do haplótipo, demonstra a não paternidade. No entanto, deve ser identificada mais do que uma para ser possível excluir uma provável mutação, confirmando com segurança, deste modo, a exclusão de paternidade. No caso de não exclusão, para se efectuar uma análise estatística rigorosa, é necessária uma base de dados com um elevado número de perfis (Kayser *e col.*, 1997).

Acresce ainda que, têm sido sugeridos vários protocolos “multiplex”, que melhoram a capacidade de resposta na rotina da análise destes sistemas e na criação de bases de dados (Kayser *e col.*, 1997; Roewer, 2001), disponíveis através da Internet (<http://ystr.charite.de>). Com a utilização laboratorial destas técnicas que amplificam simultaneamente um elevado número de *loci*, o poder de discriminação será bastante elevado, perspectivando-se a possibilidade da substituição dos sistemas autossómicos, numa base de dados de violadores, por exemplo. Recentemente, foi descrito um sistema “multiplex” que inclui 20 STRs distintos, o “20-Plex” (Butler *e col.*, 2002).

A natureza não recombinante do cromossoma Y é uma espada de dois gumes. Por um lado, como foi exposto, pode ser usado com segurança para despistar genealogias paternas. Por outro lado, a subestruturação populacional pode levantar problemas significativos neste tipo de análise. Por exemplo, em processos de imigração especificamente masculinos, o chamado efeito do fundador e a deriva genética (“genetic drift”), podem levar a uma representação excessiva de determinados haplótipos numa dada população.

Por isso, embora se possa usar os microsátélites, deste cromossoma, com segurança para fins de exclusão de um suposto pai e/ou suspeito, quando não se verificar exclusão a interpretação dos resultados pode ser problemática, porque todos os indivíduos do sexo masculino familiares do suposto pai/suspeito partilham o mesmo haplótipo. Por outro lado, dependendo da história demográfica de determinada população, alguns haplótipos podem estar demasiadamente representados, não proporcionando considerável poder informativo e a coincidência de haplótipos ser devida ao acaso.

Outra aplicação do ponto de vista genealógico, é a possibilidade da existência da associação entre sobrenomes e determinados haplótipos ou pequenos grupos com haplótipos específicos. Situações paradigmáticas das referidas associações são o caso do apelido inglês “Sykes” e da sucessão dos sacerdotes Judeus a partir de Aarão, segundo a tradição bíblica (Skoreki *e col.*, 1997). Em estudos futuros, a confirmação da existência destas associações, conjuntamente com a especificidade populacional dos haplótipos do cromossoma Y, podem ser muito úteis na determinação da origem populacional em amostras de DNA, que surgem no quotidiano laboratorial. Recentemente, em comunicação oral, Jobling referiu que, provavelmente, não existem tais associações (XIXth Congress of the ISFG, Arcachon, França, 2003).

A utilidade do estudo do cromossoma Y tem ainda sido demonstrada com sucesso em alguns casos famosos e controversos, como é o caso da paternidade de uma criança, último filho de uma escrava (Sally Hemings) do presidente dos Estados Unidos da América, Thomas Jefferson, que partilhava o mesmo haplótipo Y com o referido presidente. O estudo envolveu a análise de 19 amostras, colhidas a indivíduos representativos das possíveis linhagens envolvidas: a do próprio presidente Thomas Jefferson e a de Hemings. Depois de uma morosa pesquisa dos possíveis descendentes vivos, apenas conseguiram amostras de cinco dos sete descendentes de um tio paterno do presidente que colaboraram no estudo, uma vez que o seu irmão e o único filho conhecido, faleceram sem deixar descendentes. Pela parte do possível filho do presidente e da sua escrava, recorreram ao único tetraneto vivo conhecido.

Existia também um terceiro indivíduo que afirmava ser o filho primogénito, fruto da relação entre Jefferson e a sua escrava, que também foi incluído no estudo, utilizando amostras de cinco dos seus descendentes vivos. Incluíram, também, três descendentes de dois sobrinhos de Jefferson, filhos de uma sua irmã, que teriam contactado com a escrava e cinco descendentes de uma outra família da mesma localidade, como controlo. Foram estudados 19 marcadores do cromossoma Y, 11 STRs, 7 SNPs e um minisatélite MSY1, por três laboratórios que colaboraram no trabalho. Todos os polimorfismos caracterizados eram idênticos nos familiares de Jefferson e nos de Hemings, o que viria a confirmar a paternidade de Jefferson relativamente ao último filho da sua escrava Sally Hemings. Esta confirmação só foi viável face à existência de outros dados disponíveis, que não apenas os genéticos. Relativamente aos resultados do estudo do cromossoma Y, ficou demonstrado, deste modo que, apesar de haver um lapso de 15 gerações, recorrendo ao estudo dos polimorfismos do cromossoma Y, foi possível averiguar a paternidade atribuída a Thomas Jefferson, seguindo a sua genealogia por linha paterna (Butler, 2001).

2.5.2.2. Recomendações da ISFG relativas à utilização do cromossoma Y

À semelhança do que tem sido feito relativamente a outros polimorfismos do DNA a Sociedade Internacional de Genética Forense elaborou um documento com algumas recomendações respeitantes ao uso de STRs do cromossoma Y; porquanto, nos últimos anos, a utilização destes polimorfismos tornou-se comum nos laboratórios forenses. As

aplicações incluem, geralmente, os casos de paternidade em que o pretendo pai não está disponível e na investigação criminal, quando estão envolvidos vários indivíduos do sexo masculino, como já foi amplamente mencionado neste trabalho.

Como, também, se tem vindo a referir, tem sido descrito um número elevado de novos marcadores do tipo Y-STRs e muitos se esperam vir a conhecer num futuro próximo. Muitas das recomendações genéricas relativas aos STRs autossómicos são aplicadas a este cromossoma, destacando-se as seguintes:

- a) A sequência lê-se sempre na direcção de 5'-3' e, sempre que possível, a cadeia de DNA utilizada é a que foi originalmente descrita na literatura disponível, de preferência retirada do GenBank;
- b) Deve ser usada uma nomenclatura standardizada, de preferência o sistema D#S#, desde que não exista outra nomenclatura já divulgada e utilizada.

Relativamente à classificação dos alelos, também se aplicam as regras gerais já descritas para os STRs autossómicos, designando-os pelo número de unidades de repetição, constituindo um procedimento fácil para repetições simples como é o caso do DYS392 ou o DYS388.

Se estiver presente uma repetição parcial, designa-se pelo número de repetições completas, seguidas de um ponto e do número de bases da sequência incompleta.

Se já se adoptou uma nomenclatura para um determinado sistema, deve-se manter o seu uso, como é o caso do DYS391, pois este contempla apenas o número de sequências variáveis; mas se se tratar de um alelo descrito pela primeira vez, este deve ser designado pelo número de repetições total, quer estas sejam ou não variáveis.

Os resultados dos sistemas “duplicados”, como é o caso do DYS385, em que os fragmentos não se podem associar inequivocamente a determinado *locus*, devem ser tratados como genótipos e os respectivos alelos devem estar separados por um hífen, como por exemplo, “DYS385*11-14” (Gill *e col.*, 2001).

2.5.2.3. Legislação Portuguesa

Em Portugal não existe legislação específica que regule a aplicação da análise do DNA na resolução de casos relativos à Genética e Biologia Forense, quer do âmbito civil quer do âmbito penal. A conclusão destes casos pressupõe, na generalidade das situações, a identificação genética individual, sendo as perícias efectuadas nos Serviços de Genética e Biologia Forense do Instituto Nacional de Medicina Legal (INML), com delegações no Porto, Coimbra e Lisboa. Há outros laboratórios no nosso país que realizam este tipo de perícias, nomeadamente o Laboratório de Polícia Científica da Polícia Judiciária (criminalística) e a unidade de Genética Populacional do Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP) (testes de filiação).

Nos últimos anos tem havido um esforço conjunto para a padronização de todos os processos que envolvem a análise do DNA (incluindo os marcadores do cromossoma Y a usar), bem como a realização de exercícios colaborativos, como por exemplo, testes anuais de proficiência externa, sob o patrocínio do GEP-ISFG (Grupo Espanhol e Português da ISFG). Os Serviços atrás mencionados do INML, bem como os laboratórios portugueses associados, incluindo os atrás referidos efectuam estes testes anuais. Para tal, as amostras são enviadas por uma organização externa, o Instituto Nacional de Toxicologia sediado em Madrid, e o objectivo principal é a avaliação da capacidade de execução dos procedimentos na análise do DNA, por parte dos laboratórios participantes. Para além disso, realizam-se reuniões anuais, com vista à discussão dos resultados obtidos nos aludidos exercícios, bem como a obtenção de uma maior uniformidade nos marcadores escolhidos e na forma como são apresentados aos tribunais, nomeadamente no que diz respeito a vários aspectos da análise estatística dos resultados (Robino *e col.*, 2001).

2.5.3. Aplicações em estudos de evolução humana

Para além da utilização da análise do cromossoma Y na área da Genética e Biologia Forense, este cromossoma tem dado um importante contributo na compreensão da história das populações, ou seja na área antropológica.

O número e a variedade dos diferentes sistemas polimórficos do cromossoma Y, constituem a principal razão da sua escolha em estudos de genética populacional, onde a estratégia genealógica necessária se pode basear nas diferenças das taxas de mutação dos marcadores utilizados.

Nos últimos anos, principalmente a partir de 1996, o número crescente de polimorfismos que têm sido identificados, em especial os que são susceptíveis de análise por PCR, corresponde ao aumento do número de antropólogos moleculares que utilizam a variabilidade do cromossoma Y, para verificar a ancestralidade das populações do homem moderno (*Homo sapiens sapiens*).

Existem, fundamentalmente, duas teorias ou modelos para explicar a origem do homem moderno: a teoria multi-regional e a teoria da origem única.

O modelo multi-regional sugere que o homem moderno terá surgido a partir de formas humanas anteriores (*Homo erectus*) com uma certa continuidade regional; ou seja, a partir de várias regiões em simultâneo (Cavalli-Sforza, 1998).

O modelo da origem única ou “out-of-Africa”, propõe que o homem moderno pode ter tido origem num ancestral único que vivia em África e que se dispersou pelos restantes continentes, substituindo formas humanas arcaicas. Esta teoria, corroborada, em primeiro lugar, por estudos que envolveram a análise do DNA mitocondrial (mtDNA) e, depois, pelos autossomas, foi, também, apoiada por estudos filogenéticos do cromossoma Y (Quintana-Murci *e col.*, 2001). Pois, os polimorfismos deste cromossoma mostram uma maior diversidade nas populações africanas do que nas restantes, situando-se o primeiro ramo, nas árvores filogenéticas, precisamente na divisão entre populações africanas e populações não africanas.

É possível determinar a estrutura de populações locais e continentais, confirmando ou reformulando a hipótese atrás descrita sob a origem e migrações humanas originais (out-of-Africa and back again), usando uma combinação de microsátélites do cromossoma Y, sistemas que evoluem rapidamente, e marcadores binários, mais estáveis.

É de destacar, no entanto, que o cromossoma Y contém apenas a história paterna de uma população. Por isso, tem de se ter algum cuidado na interpretação dos resultados, que podem divergir um pouco daqueles que se obtêm através do estudo do mtDNA.

Pois, os resultados de ambos os estudos podem ser influenciados por distintos factores que interferem na dispersão de indivíduos numa população, tanto do ponto de vista biológico como social e cultural (consequentemente, do cromossoma), nas sociedades humanas.

Existe também alguma discrepância no que se refere à estimativa do tempo que dista do ancestral comum mais recente (MRCA-More Recent Common Ancestral), usando os diferentes sistemas na análise populacional (mtDNA, DNA autossómico e cromossoma Y). Os valores encontrados, quer usando o estudo de marcadores bialélicos, quer os Y-STRs, rondam os 59 000 anos, em contraste com valores de 100 000 a 200 000, utilizando o mtDNA. Existem uma série de factores que podem explicar este facto, tais como: o número distinto de marcadores e de cromossomas utilizados nos diferentes estudos; o tamanho efectivo das populações estudadas usando o cromossoma Y, o mtDNA e os autossomas, pois a população do cromossoma Y é um quarto da dos autossomas e, por esse facto, as linhagens autossómicas demorarão quatro vezes mais tempo a ser extintas; diferentes métodos e hipóteses utilizadas nos cálculos estatísticos das estimativas (diferentes modelos); diferentes efeitos da selecção sobre os vários *loci* (Quintana -Murci *e cols.*, 2001).

Mesmo tendo em conta todos estes factores, a idade do MRCA estimada permanece mais recente do que seria de esperar, invocando-se os efeitos da selecção natural sobre os genes do cromossoma Y, para explicar as diferenças observadas. Em *loci* que não sofrem recombinação, a selecção natural, quer positiva quer negativa, actuando num *locus* único, influencia os restantes *loci* alterando a história da molécula na sua globalidade.

Por outro lado, existem diversas forças selectivas que modelam a evolução da parte não recombinante do cromossoma Y (NRY). Têm sido descritos diversos mecanismos que envolvem tanto a selecção positiva como negativa. Uma característica de todos estes modelos é, a longo prazo, uma redução na diversidade de sequências levando em última instância à perda de funções biológicas na região NRY (Quintana-Murci *e col.*, 2001).

2.6. Polimorfismos do cromossoma Y humano: Marcadores bialélicos

A diversidade do cromossoma Y permite distinguir diferentes linhagens, criando haplótipos especificamente masculinos que se distribuem pelas populações humanas de uma forma não aleatória. Estes haplótipos, quando se definem através do estudo de polimorfismos que constituem acontecimentos únicos, como os SNPs (Single Nucleotide Polymorphisms), são denominados haplogrupos (HG).

O número destes haplogrupos é relativamente pequeno mas, quando se estuda uma bateria que inclui um número elevado de marcadores bialélicos (cerca de 50, segundo vários autores; Lareu *M e col.*, 2001), esta constitui um meio simples e poderoso de exclusão no âmbito da investigação de paternidade ou criminalística e a sua especificidade populacional pode permitir fazer inferências acerca da população de origem de determinado indivíduo, como por exemplo, na situação em que este deixe uma amostra de material biológico numa cena de crime.

Os SNPs são os polimorfismos bialélicos mais frequentes, ocorrendo com uma frequência de cerca de 1% na espécie humana, correspondendo a posições nas quais pode surgir duas bases alternativas. A maioria destes polimorfismos surge em regiões que não alteram a funcionalidade génica, mas, também, podem situar-se em regiões codificantes (cSNPs) e em regiões reguladoras (rSNPs), afectando a expressão de determinados genes. Neste momento, são conhecidos centenas de SNPs do cromossoma Y (Underhill *e col.*, 2001).

Os SNPs que correspondem a mutações de um único par de bases (Underhill *e col.* 1997) e o YAP, um polimorfismo de inserção/delecção do elemento *Alu* (Hammer, 1994) ou uma inserção LINE (Santos *e col.*, 2000), têm baixas taxas de mutação (cerca de 2×10^{-8} por base e por geração) e são normalmente considerados como acontecimentos únicos, na evolução humana (UMEs - Unic Mutational Events) (Jobling, 2001).

Ao apresentarem uma baixa taxa de mutação, o estudo dos SNPs pode ser importante nos casos de paternidade em que só se dispõe do pretenso pai e do filho e em que este é do sexo masculino; bem como, podem ser interessantes tendo em vista a automatização

do processo de caracterização das suas variantes alélicas, pelo facto de serem polimorfismos de elevada simplicidade (binários) (Lareu M *e col.*, 2001).

Tendo em consideração as características dos marcadores bialélicos, com base no Princípio da Parsimónia e de acordo com um modelo de mutações sucessivas, é possível construir uma árvore filogenética que permite relacionar diferentes indivíduos de uma amostra populacional, como pertencentes a um determinado haplogrupo (Jobling e Tyler-Smith, 2000). Estes polimorfismos são, por todas as razões apontadas, os melhores marcadores para a construção das referidas árvores filogenéticas dos cromossomas Y. Os nódulos destas árvores constituem os chamados grupos haplotípicos ou haplogrupos, definidos pelos estados alélicos para determinado número de marcadores binários estudados. Os haplogrupos permitem analisar a diversidade e o conhecimento das taxas de mutação possibilita estimar uma idade do ancestral comum mais recente, numa amostragem de cromossomas.

Sempre que possível, são analisadas regiões homólogas da região NRY humana de espécies geneticamente relacionadas, como sejam chimpanzés, gorilas, orangotangos, que são sequenciadas, para determinar os estados ancestrais dos polimorfismos humanos (Underhill *e col.*, 2000, Hammer *e col.*, 2001); porquanto, pensa-se que a partir daquelas formas ancestrais teriam surgido as formas humanas actuais, por acumulação de várias mutações.

2.6.1. Abundância

Uma vez que a eucromatina do cromossoma Y tem 30 Mb, é de esperar que existam milhares de polimorfismos deste tipo (SNPs). Até há pouco tempo, era possível a análise de cerca de 40 marcadores bialélicos, por PCR, dos quais todos, exceptuando dois, são SNPs. Neste momento, são conhecidos mais de oito centenas de SNPs, muitos deles ainda não caracterizados e cuja utilidade, nesta área, ainda não foi confirmada (Underhill *e col.*, 2000).

2.6.2. Recorrência

Apesar da sua baixa taxa de mutação, conhecem-se mutações recorrentes neste cromossoma. Dado que existem cerca de 3×10^9 de cromossomas Y no mundo, poderá acontecer ocorrer uma mesma substituição na mesma localização nalguns indivíduos de todo o mundo; no entanto, será um acontecimento muito pouco frequente, a não ser que a mutação inicial tenha ocorrido há muito tempo atrás.

Um dos primeiros SNPs a ser descoberto foi a transição A-G no SRY-1532 (Whitfield *et al.*, 1995) que reverteu de tal modo que encontramos alelos A tanto em haplogrupos ancestrais como em derivados. Descreveram-se recentemente mais dois marcadores recorrentes (M64 e M108) (Underhill *et al.*, 2000).

Por isso, nenhum SNP deve, em rigor, considerar-se único, uma vez que embora a taxa de mutação seja baixa, o elevado número de cromossomas Y humanos existentes no planeta ($\sim 3 \times 10^9$), excede aquele valor (cerca de 2×10^{-8} por base por geração, como já foi referido) (Jobling, 2001). No entanto, como o tamanho das amostragens populacionais que se usam é reduzido, não é frequente encontrar fenómenos de recorrência, conseguindo-se construir árvores filogenéticas, quando se caracterizam múltiplos SNPs, apesar da possibilidade sua existência.

2.6.3. Nomenclatura

A nomenclatura relativamente aos haplogrupos definidos pelos marcadores bialélicos, na grande maioria por Y-SNPs, tem sido um dos problemas com que os estudiosos do cromossoma Y se têm deparado, uma vez que as que existiam não estavam padronizadas e eram confusas. Há, pelo menos, sete distintos sistemas de nomenclatura dos haplogrupos (Tyler Smith e Jobling, 2000; Underhill, 2000; Hammer, 2001; Karafet, 2001; Semino, 2000; Su, 1999 e Capelli, 2001), tendo-se já conseguido fazer algumas equivalências entre elas. Por isso, foi necessário o estabelecimento de uma nomenclatura universalmente aceite, que permitisse a classificação sistemática e não ambígua de novos haplogrupos que, inevitavelmente, possam vir a ser descobertos.

Para obviar este problema, o “Y Chromosome Consortium”, coordenado por Michael Hammer, publicou recentemente um trabalho, sugerindo uma nova nomenclatura (YCC,

2002). Esta permite a construção de uma árvore filogenética (Figura 4), partindo da caracterização da maior parte dos polimorfismos, por técnicas de PCR, numa amostragem comum, sendo suficientemente flexível para permitir alterações, ocasionadas pela descoberta de possíveis novas mutações e linhagens na região NRY. Esta árvore inclui 153 haplogrupos e faz a correspondência com as nomenclaturas anteriormente referidas, sendo a raiz indicada por uma seta. Esta raiz da árvore filogenética, obteve-se a partir da determinação do estado ancestral de cada um dos locais polimórficos humanos, tendo sido sequenciadas regiões homólogas à NRY, de espécies animais próximas, como por exemplo, chimpanzés, gorilas e orangotangos. A letra Y refere-se à totalidade do cladograma.

Os haplogrupos foram agrupados em “clades”, denominados por letras maiúsculas de A-R, que se dividem em subgrupos, os “subclades”. Existem linhagens que correspondem a nódulos internos dos grupos principais, que são denominados de “paragrupos”. Estes distinguem-se dos haplogrupos, porque são assinalados com um *(asterisco). Os “subclades”, dentro de cada haplogrupo, denominam-se acrescentando um sistema alternativo alfanumérico, a seguir à letra maiúscula, correspondente ao haplogrupo principal, por exemplo E1. Os “nested clades”, dentro destes “subclades” principais, são classificados usando letras minúsculas. Por isso, a nomenclatura alterna números e letras até aos ramos mais terminais da árvore, os “tip haplogroups”. Por exemplo, o haplogrupo 22, segundo Jobling e Tyler-Smith (2000), passa a denominar-se R1b8, na nova nomenclatura.

Como alternativa a esta nomenclatura, o mesmo grupo de trabalho, sugere que os haplogrupos possam ser denominados pelas mutações que definem as linhagens, em vez da própria linhagem. Deste modo, surgiria a informação do haplogrupo, designado pelas letras A-R, seguida da mutação que o define. Como exemplo das duas nomenclaturas alternativas, refere-se o haplogrupo H1a, designação adoptada pela primeira, que passará a ser denominado por H-M36, de acordo com a segunda. Esta última metodologia é mais informativa, no que diz respeito a alterações da topologia da árvore.

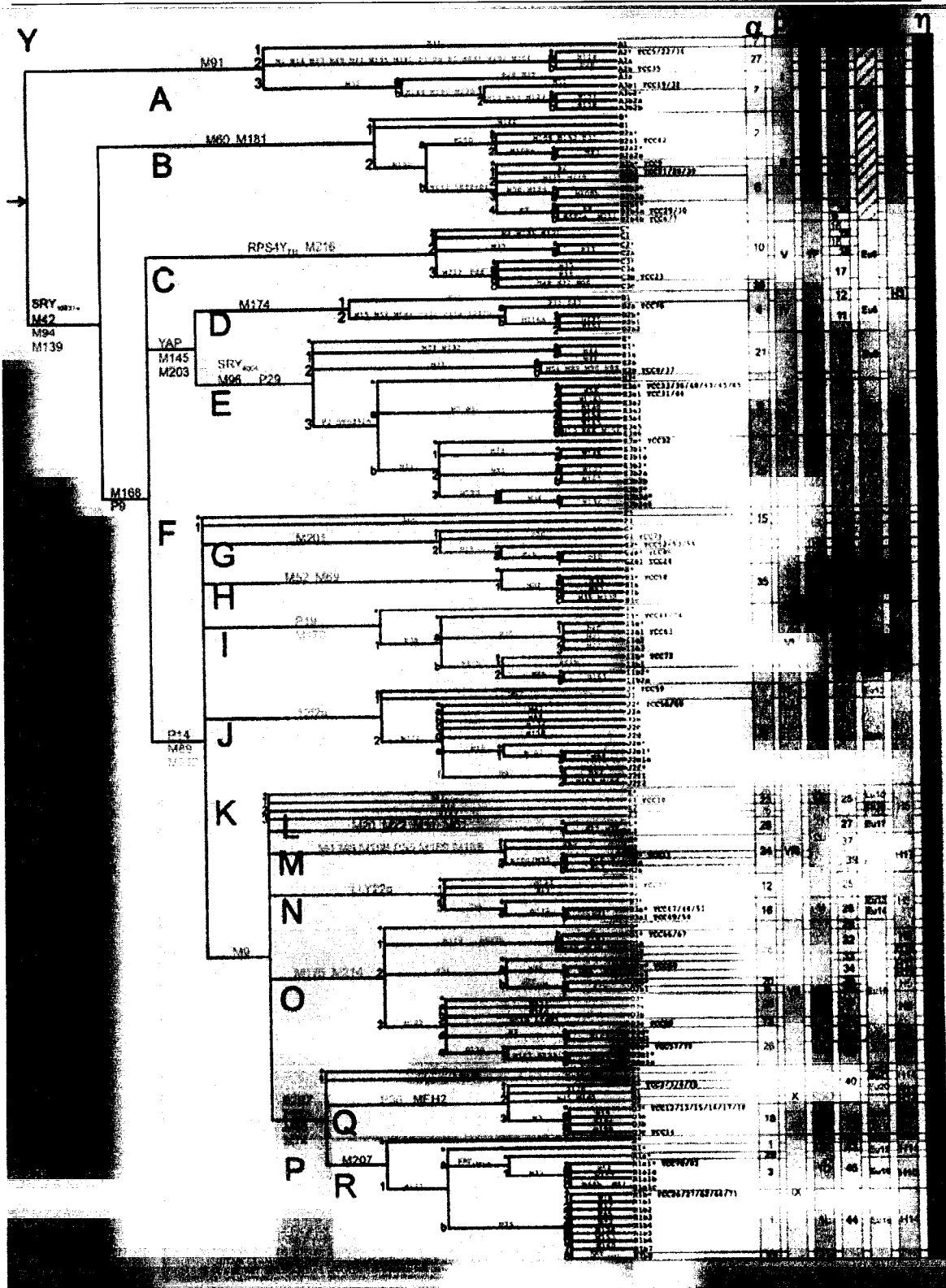


Figura 4. Árvore construída pelo YCC.

2.6.4. Especificidade populacional dos haplogrupos

Desde a descoberta dos primeiros marcadores polimórficos do cromossoma Y, que se tornou visível que os seus haplogrupos se distribuem de forma não aleatória nas populações. Comparando-os com os polimorfismos do DNA autossómico e do DNA mitocondrial, ficou demonstrado que este cromossoma apresenta a maior diferenciação genética entre populações geograficamente distintas. Esta constatação pode ser explicada pelo facto da taxa de migração feminina ser oito vezes maior que a masculina, como resultado do chamado fenómeno da patrilocalidade, em que as mulheres tendem a deslocar-se para as terras dos futuros maridos e não o inverso (Seielstad MT *e col.*, 1998; Jobling, 2001). Por outro lado, também têm sido descritos casos em que as mulheres têm uma taxa de migração elevada para populações vizinhas, especialmente naquelas em que existe a tradição de ser o filho mais velho a herdar o património familiar. Noutros casos as mulheres mantêm-se na sua zona geográfica, sendo os homens a deslocar-se; é o caso dos soldados, marinheiros e monges. No entanto, há vários autores que chegaram à conclusão que a taxa de migração feminina não é muito distinta da masculina, sendo o maior factor que influencia a distribuição das linhagens masculinas, o da deriva genética (Bortolini MC *e col.*, 2002).

Esta diferenciação geográfica dos haplogrupos, sugere que o cromossoma Y pode ser um marcador útil para se poder predizer ou deduzir a população de origem.

A data de ocupação da Europa pelo homem moderno correntemente aceite é de ~40 000 anos (YBP-Years Before the Present). A variação dos padrões de diversidade genética autossómica na Europa tem sido interpretada através do modelo da difusão démica Neolítica, de acordo com o desenvolvimento agrícola. Pelo contrário, estudos que utilizaram o DNA mitocondrial permitiram inferir acerca de muitas linhagens fundadoras até ao Paleolítico, sem mostrar grandes variações na referida diversidade.

Vários estudos que analisaram amostras de populações europeias e próximas da Europa, usando múltiplos polimorfismos bialélicos que, teoricamente definiam 10 haplogrupos, demonstraram existir padrões de diferenciação geográfica não aleatórios, mostrando diferenças significativas de distribuição geográfica para cinco dos seis haplogrupos encontrados. O HG 2 (segundo a nomenclatura de Jobling e Tyler Smith) é a linhagem mais ancestral encontrada na Europa e, por isso, existem cromossomas deste haplogrupo

distribuídos por todo o território europeu, constituindo a única linhagem com frequência elevada e que não apresenta variação considerável.

O gradiente de distribuição populacional para dois dos haplogrupos, o HG1 e o HG9, que representam 45% dos cromossomas Y nestas populações, é consistente com a hipótese da difusão démica. Uma interpretação possível é de que os cromossomas do HG9 se deslocaram na grande expansão dos migrantes agricultores do Leste próximo e o HG1, seria uma linhagem predominante pré-existente na Europa. As estimativas das idades destas linhagens não são inconsistentes com a teoria já referida.

As declinações relativas de outros haplogrupos, ou seja, o facto das suas frequências serem reduzidas, como por exemplo o HG3, restringem-se a regiões específicas e, provavelmente, reflectem movimentos populacionais distintos, incluindo um do Mar do Norte (Rosser *et al.*, 2000).

Uma análise estatística dos dados deste estudo, permite ainda concluir que as populações se relacionam sobretudo com a proximidade geográfica e não tanto com fundamento em afinidades linguísticas.

Existe ainda um aspecto importante a referir, os estudos efectuados para verificar as distribuições dos haplogrupos através das regiões geográficas, que permitem obter mapas, partem de amostras colhidas deliberadamente a pessoas autóctones dessas regiões, excluindo outro tipo de amostras, designadamente de indivíduos residentes, que não são representativos de determinada população, por poderem ter origem geográfica diferente.

Na prática Médico-Legal, as amostras a analisar são, muitas vezes, de populações de origem indeterminada e com composição genética desconhecida. Por isso, seria mais preciso, para efectuar os cálculos estatísticos, haver dados de amostras populacionais mais representativos. No entanto, como actualmente são usados marcadores genéticos altamente polimórficos, especialmente os STRs autossómicos, como complemento destes marcadores, que proporcionam elevado poder de discriminação, mesmo não se tendo em conta aquele pressuposto, os resultados obtidos não serão muito diferentes da realidade biológica.

2.6.5. Aplicação na Genética e Biologia Forense

Tem havido um crescente interesse em utilizar os marcadores bialélicos na área da Genética e Biologia Forense. No entanto, a análise de um pequeno número de marcadores deste tipo tem pouco valor informativo, do ponto de vista da identificação genética; porquanto, tem sido referido que serão necessários aproximadamente 50 *loci* bialélicos para proporcionar um poder informativo idêntico ao dos STRs em “multiplexes”, actualmente usados. Para a sua caracterização têm sido utilizadas várias estratégias de tipagem, baseadas na PCR, envolvendo desde géis de agarose com brometo de etídio, dot-blot reverso, até ao uso recente de “microarrays” em vidro, em meios que não recorrem à electroforese. Todavia, constatou-se que para se conseguir desenvolver “arrays” desta natureza é muito complicado, não tanto do ponto de vista de análise estatística, mas mais na perspectiva bioquímica.

No âmbito da Criminalística Biológica, um dos principais problemas que se levanta é o da interpretação de misturas; pois, nesta situação, é por vezes difícil distinguir a presença de um alelo relativamente ao “background”, quando a quantidade de DNA dos contribuidores, na referida mistura, é escassa. Nestas circunstâncias, torna-se imprescindível o uso de métodos de quantificação e caracterização de DNA eficazes (Peter Gill, 2001).

Pelo que foi exposto, os marcadores bialélicos do cromossoma Y são hoje usados como complemento de outros polimorfismos de DNA, designadamente, STRs autossómicos e Y-STRs. Futuramente, com o desenvolvimento de novas tecnologias, já em fase de investigação, estes marcadores podem ter um interesse relevante nesta área.

2.6.6. Estratégias de tipagem

O facto dos SNPs poderem ser usados como marcadores para o mapeamento de doenças genéticas multifactoriais, estimulou a descoberta de milhões destes polimorfismos, bem como o desenvolvimento de tecnologias de elevado rendimento. Por isso, têm sido desenvolvidos vários métodos de tipagem dos marcadores bialélicos, ao contrário do que acontece com os Y-STRs, pois a técnica usada para a caracterização destes últimos marcadores, amplificação simultânea de vários sistemas por PCR, demonstrou ser

eficiente. Assim, na tentativa de se poder estudar o maior número deste tipo de marcadores em simultâneo, têm sido descritas várias metodologias, algumas das quais serão resumidamente referidas. Acresce que a escolha do método depende daquilo que se pretende investigar, nomeadamente o número de marcadores e o tamanho da amostra populacional.

Metodologia convencional

♦ PCR-RFLP

Como será posteriormente referido, esta foi a técnica utilizada neste estudo, que consiste na amplificação de determinados *loci*, nos quais se detectam marcadores bialélicos, através da utilização de enzimas de restrição, que reconhecem ou não locais de corte, conforme a sequência. Deste modo, são obtidos fragmentos de tamanhos diferentes, detectáveis, por exemplo, em simples géis de agarose com brometo de etídio. Esta técnica permite o estudo de um número escasso de polimorfismos, pois de outro modo torna-se uma técnica fastidiosa.

Metodologias mais recentes

Têm sido propostas novas metodologias para o estudo de SNPs, tais como “DHPLC”, a “PCR em Tempo Real”, a mini-sequenciação, os “microarrays” e a “MALDI-TOF MS”.

♦ Cromatografia líquida de elevada performance desnaturante (DHPLC)

(Underhill et al., 1997)

Explora a retenção diferencial de moléculas de dupla cadeia “homo” e “heteroduplexes”, permitindo a detecção automática de variações dos produtos de PCR. Esta técnica detecta rapidamente emparelhamentos deficientes, apenas devidos a um único nucleótido ou a pequenas inserções/delecções (identificação de novos polimorfismos e genotipagem). No entanto, trata-se de uma metodologia não suficientemente robusta para ser aplicada em Medicina Legal, devido às dificuldades na standardização dos padrões “heteroduplex” (Barros *e col.*, 1991; Barros *e col.*, 1992). No entanto, estão a ser explorados métodos que usam gradientes desnaturantes, por serem mais seguros (Lareu *e col.*, 2001).

♦ PCR em tempo real

Tecnologia que utiliza a PCR em tempo real e que se aplica, quando se testa um número reduzido de SNPs numa amostragem grande (Ranade K *e col.*, 2001; Laréu M *e col.*, 2001).

Das metodologias citadas é uma das mais promissoras e consiste na monitorização da reacção de PCR em tempo real, usando fluorescência. Existem diferentes sistemas que usam a fluorescência e que relacionam os sinais da mesma com as quantidades dos produtos de PCR.

Entre estes sistemas destaca-se o LightCycler e o sistema comercializado pela Applied Biosystems que utiliza a actividade 5' nuclease da Taq Polimerase, o ABI PRISM 7000:

- LightCycler

É um termociclador ultra-rápido, que utiliza o ar para promover o aquecimento e arrefecimento (PCR) e capilares, onde se processa a reacção de PCR. Estes dois factores permitem a redução do tempo de cada ciclo da PCR para cerca de 30 segundos. Por isso, uma PCR completa, de 30-40 ciclos, demora apenas 20-30 minutos.

Este aparelho foi inicialmente desenvolvido para o uso de sondas FRET. Esta metodologia usa duas sondas, uma marcada com fluoresceína na extremidade 3', a dadora do fluorocromo. Esta é excitada pela fonte de energia do aparelho, sendo parte dessa energia de excitação transferida para outra sonda, a receptora que, por sua vez, emite luz num maior comprimento de onda. Esta segunda sonda é marcada na extremidade 5' por um outro composto fluorescente, conforme o SNP a analisar. Só quando as duas sondas estão hibridadas e próximas uma da outra (1-4 bases) é que ocorre a transferência de energia, ou seja a FRET (Fluorescence Resonance Energy Transfer), que é directamente proporcional à quantidade de DNA alvo disponível, isto é, ao produto de PCR.

Esta estratégia de detecção permite monitorizar o processo de amplificação em tempo real, porque a intensidade do sinal de FRET é proporcional à quantidade do produto específico de PCR produzido.

Para além disso, é ainda possível estabelecer o genótipo, usando duas sondas, uma de menor estabilidade a "sensor" que se deve posicionar directamente na mutação a ser detectada e outra, a "anchor", que reconhece uma sequência adjacente.

É sabido que uma sonda que possua apenas uma diferença numa única base, ou seja, uma não complementaridade relativamente à sequência a estudar, pode permitir a

hibridação; no entanto, a temperatura de “melting” é mais baixa do que uma sonda que seja perfeitamente complementar. Por isso, é possível distinguir, através desta temperatura da sonda “sensor”, os diferentes alelos. Assim, se a mutação não está presente a hibridação da sonda é perfeita e a temperatura de “melting” é mais elevada.

No final da reacção faz-se uma análise das curvas de temperaturas de “melting” das sondas, tendo sido determinado que diferenças numa só base (SNPs) pode provocar uma alteração de 3-10°C.

- ABI PRISM 7000

A reacção de PCR usa a actividade 5' da Taq Polimerase, e consiste na utilização de uma sonda, que é cindida graças àquela actividade da Taq e, se a sonda for fluorogénica, é possível que os produtos de PCR sejam detectados sem haver necessidade de uma posterior hibridação ou electroforese. A referida sonda, consiste num oligonucleótido marcado com um corante fluorescente o “reporter” e um “quencher”, situados em extremidades opostas. Quando a sonda está intacta, a proximidade da “quencher” provoca a FRET, ou seja, a transferência da fluorescência da sonda “reporter” para a “quencher”, o que provoca uma redução da fluorescência “reporter”, porque esta é totalmente absorvida pela “quencher”. As cisões da sonda fluorogénica (“repórter”) durante a PCR libertam-na, provocando uma intensidade de fluorescência, por cessação da FRET.

No caso da caracterização de sistemas bialélicos, são incluídas sondas específicas para cada um dos alelos. Estas podem ser facilmente distinguíveis, porque são marcadas com diferentes fluorocromos. A hibridação completa da sonda é necessária para que ao deslocamento inicial desta se siga a sua cisão; por conseguinte, se houver uma não coincidência entre a sonda e o DNA alvo, há redução da eficiência da hibridação da sonda o que, consequentemente, facilita a sua cisão. Um aumento substancial de cada um dos corantes fluorescentes, indica homozigotia para um dos alelos, conforme a cor da fluorescência empregue na marcação do respectivo alelo. O aumento das duas cores será indicativo da presença de uma heterozigotia.

♦ Mini-sequenciação (SnaPshot ddNTP primer Extension – Applied Biosystems)

Para se efectuar a minisequenciação é necessário um primer específico, desenhado de tal modo que inclua todas as bases exactamente até uma base antes da potencial base mutada. Posteriormente, são adicionados ddNTPs (dideoxinucleótidos) marcados com

fluorescência, que serão incorporados de uma forma complementar, de acordo com a sequência. A reacção é interrompida depois da incorporação de um terminador na extremidade 3' do primer não marcado, adjacente à possível base mutada. Esta reacção não inclui a incorporação de dNTPs.

A cor dos picos obtidos, permite a identificação dos alelos presentes no SNP.

◆ DNA Microarrays (Genome Chips)

Têm sido utilizadas várias terminologias para descrever esta tecnologia, tais como: “biochip”, “DNA chip”, “DNA microarray” e “gene array”. Um “array” é uma arranjo ordenado de amostras que fornece um meio de “matching” de amostras conhecidas ou desconhecidas, baseado nas regras de emparelhamento das bases (A-T e G-C, para o DNA e A-U e G-C, para o RNA). Este processo é susceptível de automatização, no sentido de identificar amostras desconhecidas. Uma experiência que envolva “arrays” pode usar microplacas ou membranas para fazer o “blotting”, montadas manualmente ou usando sistemas automáticos (robots), para depositar as amostras. Estes podem-se dividir em “macroarrays” ou “microarrays”, diferindo apenas no tamanho do local onde são aplicadas as amostras (“spots”). Nas macroarrays as “spots” podem ter 300 microns ou mais, sendo facilmente visualizadas através de “scanners”. Nas “microarrays” as “spots” são inferiores a 200 microns de diâmetro e, normalmente, contêm milhares de “spots”, sendo necessária robótica especial e aparelhos de imagem, para se efectuar a leitura.

A *Affimetrix, Inc.* possui uma marca registada, a GeneChip®, que se refere aos seus arrays de DNA de alta densidade, baseados em oligonucleótidos. As sondas utilizadas (cDNA ou oligonucleótidos) são imobilizadas numa superfície sólida, como o vidro ou membranas de nylon, usando um sistema automático para a sua distribuição, seguida da junção do DNA a identificar. A leitura pode-se efectuar de várias formas, geralmente automática, dependendo do processo previamente utilizado.

Esta empresa usa oligonucleótidos sintetizados ou *in situ* no próprio chip, ou por síntese convencional, seguida de imobilização no chip. Promove-se depois a hibridação do DNA marcado com o “array”, permitindo deste modo a identificação dos diferentes alelos.

Os “microarrays” de DNA provaram ser uma metodologia poderosa e promissora em análises de alta escala em que se podem estudar centenas ou milhares de SNPs em

poucos indivíduos, mas têm a desvantagem de não poderem ser construídos ou modificadas em pequenos laboratórios, de modo a incluir novos marcadores que possam interessar. Pode-se proceder ainda à minisequenciação em microarrays (Raitio *M e col.*, 2001).

♦ MALDI-TOF-MS (“Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry”)

Esta metodologia inclui a volatilização/ionização, na qual a amostra e a matriz são vaporizadas e os produtos da extensão do “primer” são enviados para o “flight tube” e são submetidos a aceleração por vácuo. Nesta fase, os produtos da extensão do “primer”, carregados negativamente, quando é aplicado um campo eléctrico, são deslocados desde o “flight tube” ao detector.

O tempo que medeia entre a aplicação do campo eléctrico e a colisão dos referidos produtos com o detector, o “time of flight”, constitui uma medida muito precisa do peso molecular destes, sendo certo que as moléculas mais pequenas, chegam mais rapidamente ao detector.

Ao efectuar a análise dos resultados tem de se converter o “time of flight” na massa exacta de cada molécula.

Esta metodologia foi recentemente desenvolvida com a finalidade de analisar SNPs e, embora requeira um equipamento especial, têm sido descritas muitas técnicas, por exemplo o “PinPoint assay” (Paracchini e col., 2002), particularmente interessante porque permite um número elevado de análise simultânea de vários SNPs (“multiplexing”).

Alguns destes métodos têm a desvantagem de consumir demasiado tempo e serem métodos seriados, impedindo a tipagem dos marcadores disponíveis em todas as amostras. Por isso, se se pensar em utilizar os SNPs na área médico-legal, é essencial a utilização de métodos de elevado rendimento como, por exemplo “DNA microarrays” e a Espectrometria de Massa.

2.6.7. Polimorfismos bialélicos utilizados neste estudo

Como já foi referido, neste momento foram descritos múltiplos marcadores bialélicos do cromossoma Y.

Para este estudo foram escolhidos apenas seis SNPs e uma inserção de um elemento *Alu*: sY81, SRY-8299, M9, DYS257, SRY-2627, SRY-1532 e o YAP (DYS287), pelas seguintes razões:

- Por serem os mais conhecidos, quando este trabalho foi iniciado;
- Por serem os mais informativos em populações europeias previamente estudadas;
- Pela possibilidade da sua detecção usando a técnica de PCR, seguida de digestão enzimática, metodologia simples;
- Por existirem as duas variantes alélicas, encontradas em estudos já efectuados em populações europeias próximas;
- Por ser possível definir uma filogenia de haplogrupos já descrita (Jobling e Tyler-Smith, 1995 e Jobling *et al.*, 1997).

A combinação das diferentes formas alélicas obtidas, permitiu definir os haplogrupos a que cada amostra pertencia, na população estudada.

2.6.7.1. YAP (DYS287)

Trata-se de uma inserção (300bp) de um “retrotransposon” que faz parte da família do DNA repetitivo *Alu*, presente no *locus* DYS287, na região proximal do braço longo do cromossoma Y (Yq11) (“Y *Alu* Polymorphism”), em alguns indivíduos do sexo masculino (Hammer e Horai, 1995). É a única inserção *Alu* localizada no cromossoma Y humano.

A referida inserção, provavelmente ocorreu uma vez na evolução, depois da divergência entre a espécie humana e outras espécies de primatas não humanos, uma vez que está ausente nestes últimos. Nos cromossomas Y em que está presente, a sua estabilidade tem sido demonstrada através de estudos de sequenciação.

A sua detecção faz-se mediante amplificação do *locus* DYS287, verificando-se a presença ou ausência da inserção. A variante alélica YAP+ apresenta a dita inserção *Alu*

na sua sequência, dando origem a um produto de PCR com 455 bp; enquanto o alelo YAP-, sem a inserção, tem um tamanho de 150 bp.

A sequenciação da região homóloga deste *locus* em primatas, permite inferir a ancestralidade do alelo YAP-, que não apresenta a inserção na sua sequência.

A inserção apresenta uma elevada frequência em populações africanas, baixa em populações europeias e está ausente em populações asiáticas. A sua elevada frequência na população japonesa constitui uma exceção (Hammer e Horai, 1995). No Noroeste da Península Ibérica foi descrito um gradiente de frequência no sentido Norte-Sul (Pereira *et al.*, 2000).

2.6.7.2. sY81 (DYS271 ou M2)

Descrito pela primeira vez por Seielstad *et al.* (1994), detectado por sequenciação de fragmentos de DNA provenientes de indivíduos de diversas origens populacionais.

Define-se por uma transição de A-G, localizada no *locus* DYS271, posição 168 (Vollrath *et al.* 1992).

Há estudos demonstrativos da existência de desequilíbrio de ligamento entre o alelo G deste SNP e a inserção YAP; ou seja, o alelo G só aparece em indivíduos que apresentam a inserção (Spurdle *et al.*, 1994). Ao longo de toda a história evolutiva não ocorreram deleções deste elemento *Alu*, nem substituições G-A.

Os primatas não humanos apresentam a forma alélica A (ancestral) para o sY81 e ausência da inserção YAP, o que parece indicar que a inserção terá ocorrido num cromossoma Y com o alelo A para o sY81 e que, depois, terá surgido a transição A-G.

A presença da mutação provoca a perda de um local de corte para a enzima de restrição *Nla III* (*Neisseria lactamica*), permitindo a determinação das formas alélicas (sequência de reconhecimento: 5'-CATG-3').

É nas populações subsarianas que têm sido descritas as frequências mais elevadas para o alelo mutado do sY81 (G) (Karafet *et al.*, 1999), mas este alelo também está presente

em populações do Norte de África (Bosch *e col.*, 1999). O alelo G está ausente em populações asiáticas, europeias e de americanos nativos (Rosser *e col.*, 2000).

2.6.7.3. *SRY-8299 (SRY-4064)*

Caracteriza-se por uma transição G-A na posição -8299 do gene determinante dos testículos (a jusante da extremidade 5'), no braço curto do cromossoma Y (Yp). Também é denominado de SRY4064, já que esta é a posição que ocupava no fragmento polimórfico sequenciado por Whitfield *e col.* (1995), utilizando 5 indivíduos de diferentes origens e um chimpanzé. O estado ancestral consiste num G na posição 4064 deste fragmento, que é substituído por um A no alelo mutado.

A presença da mutação provoca a perda de um local de reconhecimento para enzima de restrição *BsrB I* (*Bacillus stearothermophilus*), permitindo distinguir as variantes alélicas para este SNP (local de reconhecimento: 5'-CAG ▴CGG-3').

Esta mutação só está presente em indivíduos com a inserção YAP.

Este marcador bialélico existe com maior frequência nas populações africanas, à semelhança do que acontece com a inserção YAP. Os indivíduos do Centro e Este asiático podem não apresentar o alelo mutado para este sistema (alelo A).

2.6.7.4. *M9*

Trata-se de uma substituição nucleotídica de uma C-G, localizada no intervalo da deleção 50 do braço longo do cromossoma Y (Yq), descrita no mapa elaborado por Vollrath *e col.* (1992).

É um polimorfismo que foi detectado por Underhill *e col.* (1997), utilizando uma técnica para comparar sequências de DNA, denominada de cromatografia líquida desnaturante de alta resolução, a DHPLC ("Denaturing High-performance Liquid Chromatography").

A presença da mutação (G) provoca a perda de um local de reconhecimento para a enzima de restrição *Hinf I* (*Haemophilus influenzae* Rf) (local de reconhecimento: 5'-G⁺ANTC). A variante ancestral foi determinada por comparação de sequências de DNA humano com DNA de 6 Chimpanzés, 4 Gorilas e 2 Orangotangos (Underhill *e col.*, 1997).

Esta mutação está representada por uma linhagem masculina largamente expandida do ponto de vista geográfico. Não se encontra no continente africano, salvo raras exceções, o que parece indicar que tenha sido originada noutro continente. Estudos populacionais têm revelado que as frequências dos dois alelos são muito semelhantes. Na população europeia parece haver um domínio da variante mutada, o alelo G (Rosser *e col.*, 2000).

2.6.7.5. *DYS257*

Trata-se de uma substituição nucleotídica de G-A, localizada dentro de um clone YAC no intervalo 3C2 do braço curto do cromossoma Y humano (Foote *e col.*, 1992) e no intervalo de deleção 3C do mapa elaborado por Vollrath *e col.* (1992), numa posição intermédia entre a região SRY e a região centromérica. Hammer *e col.* (1998) encontram duas cópias adicionais, não polimórficas, deste *locus*.

A variante ancestral *DYS257**G, foi determinada por comparação de dados obtidos por sequenciação de DNA humano com DNA de quatro espécies de primatas (*Pan troglodytes*, *Pan paniscus*, *Gorilla gorilla* e *Pongo pygmaeus*).

A variante com a base alterada (A) provoca o aparecimento de um local de corte para a enzima de restrição *BanI* (*Bacillus aneurinolyticus*), originando um padrão de 5 bandas correspondentes ao alelo *DYS257**A, enquanto a variante *DYS257**G apresenta apenas quatro bandas (local de reconhecimento: 5'-...G⁺G(T/C)(A/G)CC...-3').

A presença do alelo mutado atinge elevadas frequências em populações do Centro-Leste da Europa, compreendendo cerca de metade do número total de cromossomas de populações como a Rússia, Polónia e Eslovénia; as frequências no sudeste e sudoeste europeu são baixas (Rosser *e col.*, 2000). Este gradiente de distribuição pode ter tido

origem num movimento migratório de um grupo de indivíduos do Norte do Mar Cáspio datado de ~7 000 YBP (Cavalli-Sforza *e col.*, 1994).

2.6.7.6. *SRY-1532 (SRY-10831)*

Consiste numa substituição nucleotídica de A-G, na posição -1532 do gene SRY (a jusante da extremidade 5'), no braço curto do cromossoma Y (Yp).

Este polimorfismo, descrito pela primeira vez por Whitfield *e col.* (1995), que identificou o alelo ancestral a partir de um chimpanzé e outros primatas (gorila e orangotango). A sequenciação do fragmento de 18.3 Kb, permitiu inferir que na posição 10831 aparecia a base A, considerada a variante ancestral, enquanto a forma mutada apresentava a base G. Por este motivo também foi conhecida por SRY10831.

Este SNP apresenta mutações recorrentes, isto é, originalmente existiria um A na posição anteriormente descrita que sofreu uma mutação para G que, por sua vez, reverteu para A no decorrer da evolução, voltando à forma original.

A presença do alelo G cria um local de reconhecimento para a enzima *Dra III* (*Deinococcus radiophilus*), permitindo a determinação das variantes alélicas deste SNP por restrição ao originar dois fragmentos (local de reconhecimento: 5'-...CACNNN•CTG...-3').

A presença do alelo A no SRY-1532 em cromossomas que apresentam o estado ancestral para os restantes polimorfismos, é típica de populações africanas sub-sarianas (Karafet *e col.*, 2000). Pelo contrário, aparecem cromossomas com o alelo A deste polimorfismo e o alelo mutado no M9 em populações europeias (Centro da Europa) e asiáticas, estando praticamente ausentes em populações africanas e de americanos nativos. O alelo G está amplamente representado em todas as populações mundiais, que têm sido estudadas.

2.6.7.7. *SRY-2627 (SRY-2628 ou M167)*

Este SNP caracteriza-se por apresentar uma transição C-T, localizada a -2627bp do codão de iniciação do gene SRY (a jusante da extremidade 5'), descrita por Veitia *e col.* (1997), mediante amplificação e posterior sequenciação. Inicialmente foi designado por SRY-2628. Está inserido na região ORF ("Open Reading Frame" do gene SRY), dentro da qual se encontra a chamada "HMG box", um domínio conservado em muitas proteínas que se ligam ao DNA, incluindo um grupo de proteínas nucleares de elevada mobilidade denominadas de "High Mobility Group", entre outras.

A transição foi detectada por uma técnica denominada por FAMA ("Fluorescent Assisted Mismatch Analysis") no contexto de uma duplicação em "tandem" da sequência CCCCACAGGGTG situada, como já foi descrito, a -2627 pb (extremidade 5') do tripleto de iniciação (ATG) do gene SRY. Esta região repetitiva, após a divergência entre humanos e chimpanzés, sofreu uma expansão. Este facto torna difícil definir o estado ancestral, mesmo quando se estudam primatas não humanos (Hurles *e col.*, 1999), uma vez que existem várias sequências possíveis. De qualquer modo, como o polimorfismo se localiza na terceira base de uma das repetições e, todas as cópias desta repetição apresentam um C (tanto em primatas não humanos como em humanos), este alelo foi designado como o alelo ancestral.

Este SNP, quando apresenta a mutação, pode ser reconhecido pela enzima de restrição *BsiHKA I* (*Bacillus stearothermophilus*), permitindo a determinação das suas variantes alélicas por uma simples digestão (local de reconhecimento: 5'-...G(A/T)GC(A/T)•C...-3').

A mutação neste sistema só ocorre em cromossomas Y que apresentam a forma mutada no M9. Está fracamente representada na população mundial, constatando-se que a maior parte dos indivíduos que a apresentam são originários da Península Ibérica (Hurles *e col.* 1999). Este facto parece indicar uma origem Ibérica para este polimorfismo.

Na Tabela 3 resumiram-se as características referidas dos SNPs.

Tabela 3.Características dos polimorfismos bialélicos

Polimorfismo	alelo ancestral/ moderno	Frequência
YAP (DYS287)	YAP– /YAP+	Frequência elevada em África, Baixa frequência na Europa ausente na Ásia
sY81 (DYS271)	A/G	Frequência elevada em populações sub- sarianas; Frequente no Norte de África; Ausente em populações asiáticas europeias e de americanos nativos.
SRY-8299	G/A	Maior frequência nas populações africanas; Praticamente ausente no Centro e Este asiático.
M9	C/G	Largamente expandida geograficamente; Ausente no continente africano; Maior frequência da variante mutada na Europa
DYS257	A/G	Elevada frequência em populações do Centro-leste da Europa.
SRY-1532	G/A	Típica de populações africanas sub- sarianas; A forma recorrente aparece na Europa e Ásia e está ausente em África e América.
SRY-2627	C/T	Apenas na Península Ibérica e França

3. MATERIAL E MÉTODOS

3. MATERIAL E MÉTODOS

3.1. *Colheita de amostras*

As amostras utilizadas neste trabalho foram, na sua maioria, provenientes de indivíduos fenotipicamente saudáveis, não relacionados, envolvidos em casos de filiação, tendo-se efectuado colheita de sangue venoso em tubos com anticoagulante EDTA K3. As ditas amostras, foram armazenadas a -20°C até se proceder à extracção do DNA, a partir do sangue total.

No que se refere aos STRs, estudaram-se 122 indivíduos do sexo masculino não relacionados e 26 pares pai-filho, para verificar a existência de possíveis mutações, embora se considere este número manifestamente insuficiente para o cálculo de taxas de mutação.

Quanto aos marcadores bialélicos, o ideal seria estudar o mesmo número de amostras, mas como a análise foi desfasada no tempo, estudaram-se apenas 99 amostras.

3.2. *Extracção do DNA*

Procedeu-se à extracção do DNA recorrendo, fundamentalmente, a dois métodos:

- 1) A partir de sangue líquido, por um método descrito por Giusti *e col.* (1986), que utiliza a extracção orgânica com fenol-clorofórmio-álcool isoamílico (casos de filiação); quanto às amostras relacionadas com crimes sexuais, foi usado este processo de extracção do DNA e o método de lise diferencial, cujo objectivo é a separação das duas fracções, feminina e masculina.
- 2) A partir de manchas de sangue, pelo método descrito por Singer-Sam *e col.* (1989) que utiliza uma resina quelante, o Chelex (casos de filiação e criminalística biológica) ou utilizando uma variante deste método para extrair DNA de zaragatoas bucais recolhidas em alguns casos, por exemplo, a suspeitos de agressões sexuais.

3.2.1. Extracção com fenol-clorofórmio-álcool isoamílico

Lise/digestão celular

A 200µl de sangue periférico:

- Juntou-se 500µl de Tampão de Lise.
- Incubou-se em banho-maria com agitação, 2h a 56°C (ou O.N. a 37°C).

Extracção

Após incubação, procedeu-se à extracção:

- Juntou-se 200µl de fenol-clorofórmio-álcool isoamílico (25:24:1). Agitou-se no “vortex” para homogeneizar as fases.
- Centrifugou-se a 12 000 rpm, 5 min.
- Recuperarou-se a fase aquosa (~400µl) para outro tubo, devidamente etiquetado.

Precipitação e lavagem

Ao tubo com a fase aquosa após a extracção:

- Adicionou-se 1/10 do seu volume (40µl) de Acetato de Sódio a 2M, 1µl de Glicogénio (20mg/ml) e dois volumes (800µl) de Etanol Absoluto frio (-20°C). Inverteram-se os tubos suavemente para misturar todos os reagentes.
- Os tubos que continham DNA visível, centrifugaram-se 5 min a 13 000 rpm.
- Se o DNA não fosse visível (a maior parte das vezes), colocavam-se os tubos 1h a -20°C e centrifugavam-se 20 min a 13 000 rpm.
- Em qualquer dos casos anteriores, decantou-se o sobrenadante de cada um dos tubos.
- Lavaram-se os “pellets” com 1ml de Etanol a 80% frio (1000µl).
- Centrifugaram-se 1 min a 12 000 rpm e retiraram-se novamente os sobrenadantes.
- Deixou-se secar o “pellet” de DNA no concentrador de DNA (DNA SpeedVac, DNA 110-Savant).
- Ressuspendeu-se em volume adequado de H₂O MiliQ ou tampão TE (~20-40µl).
- Quantificou-se o DNA presente em cada um dos tubos.

Reagentes:

Cloreto de Sódio (NaCl) Merk (6404)

Acetato de Sódio (Na₂CH₃CO₂) Merk (6267)

Proteínase K Sigma (P0390)

SDS Sigma (L-4390)

Fenol-clorofórmio-álcool isoamílico USB (20084)

Etanol Absoluto Merk (0983)

Tris Sigma (T-8524)

EDTA-dissódico Sigma (E-5134)

Ditiotreitol Sigma (D-0632)

Glicogénio Sigma (G-8876)

Soluções:

1. Tampão de Lise: Tris 10mM pH 8, EDTA 10mM, NaCl 0,1M, SDS a 2% (w/v), DTT 40mM, Proteinase K a 0,2 mg/ml.

Nota: O DTT e a Proteínase K só se juntaram no momento em que se adicionaram às amostras.

2. Acetato de Sódio 2M.

3. Glicogénio a 20 mg/ml.

4. Etanol a 80% (v/v).

5. Tampão TE –Tris 10mM, EDTA 1 mM, pH 7,5.

3.2.2. *Extracção com fenol-clorofórmio-álcool isoamílico e lise diferencial (mistura de células)*

Lise/digestão celular

Fase vaginal:

Cortou-se cerca de 2mm² de tecido/papel com a mancha e colocou-se num microtubo de 1,5ml.

Juntou-se 500µl de Tampão de Lise, 15µl Proteínase K (10mg/ml) e 35µl de SDS a 10% (w/v), a cada tubo.

Incubaram-se os tubos 2h a 56°C (em banho-maria).

Centrifugaram-se 10 min a 9 000 rpm, colocando cada tubo perfurado (com uma agulha ao rubro) dentro de outro tubo. No tubo não perfurado, obteve-se um sobrenadante proveniente da lise de células vaginais e um “pellet” espermático.

Transferiu-se o sobrenadante para outro tubo (fase vaginal) e guardou-se a 4°C até se proceder à extracção do DNA respectivo.

Fase espermática:

Ressuspendeu-se o “pellet” espermático em 400µl de Acetato de Sódio a 0.2M. Agitou-se suavemente em “vórtex”.

Juntou-se a cada tubo, 20µl de Proteínase K (10mg/ml), 20µl de DTT (1M) e 30µl de SDS a 10%. Agitou-se esta mistura suavemente e incubou-se toda a noite a 37°C.

Extracção

Extraíu-se o DNA com 200µl de fenol-clorofórmio-álcool isoamílico, a partir dos tubos que continham, respectivamente, a fase vaginal e a espermática. Agitou-se vigorosamente em “vórtex” e centrifugou-se 5 min a 13 000 rpm.

Recolheu-se a fase aquosa, utilizando pontas amarelas cortadas.

Precipitação e lavagem

Juntou-se 1µl de Glicogénio (20mg/ml) e 2 volumes de Etanol absoluto (900µl) a cada tubo, invertendo-os suavemente para precipitar o DNA.

Quando o DNA não era visível colocaram-se os tubos 2h a -20°C.

Centrifugaram-se os tubos 20 min a 12 000 rpm e decantou-se o sobrenadante.

Juntou-se 1ml de Etanol a 80% frio (-20°C) para lavar o “pellet”. Agitou-se suavemente em “vórtex”.

Centrifugou-se 2 min a 12 000 rpm e retirou-se o sobrenadante (~950µl), deixando secar o “pellet” no SpeedVac.

Ressuspendeu-se em volume adequado de H₂O MiliQ ou tampão TE (20-40µl).

Quantificou-se o DNA presente em cada tubo.

Microcon 100

Trata-se de uma alternativa ao uso de Etanol, após extracção do DNA com fenol-clorofórmio-álcool isoamílico.

A fase aquosa, depois da extracção, é colocada no tubo Microcon juntando-lhe 100 µl de TE. Centrifugou-se 10 min a 3 200 rpm e rejeitou-se o tampão.

Juntou-se mais 200 µl de TE ao Microcon, centrifugou-se a 3 200 rpm, tendo-se rejeitado o tubo colector com o tampão.

Adicionou-se 20-200µl de TE, tendo-se invertido o filtro para um tubo colector limpo e centrifugou-se 5 min a 3 200 rpm.

Rejeitou-se o filtro e acondicionou-se o tubo com o DNA a 4°C, ou a temperaturas inferiores, conforme o tempo que mediou até à sua utilização.

Reagentes:

Cloreto de Sódio (NaCl) Merk (6404)

Acetato de Sódio (Na₂CH₃CO₂) Merk (6267)

Proteínase K Sigma (P0390)

SDS Sigma (L-4390)

Fenol-clorofórmio-álcool isoamílico USB (20084)

Etanol Absoluto Merk (0983)

Tris Sigma (T-8524)

EDTA-dissódico Sigma (E-5134)

Ditiotreitol Sigma (D-0632)

Glicogénio Sigma (G-8876)

Soluções:

1. Tampão de Lise: Tris 10mM (pH=8), EDTA 10mM, NaCl 10mM.
2. SDS a 2% (w/v).
3. Acetato de Sódio 0,2M.
4. Proteínase K a 10 mg/ml.
5. DTT 1M.
6. Glicogénio a 20 mg/ml.
7. Etanol a 80%(v/v).
8. Tampão TE –Tris 10mM, EDTA 1 mM, pH 7,5.

3.2.3. *Extracção com Chelex*

1. Pipetou-se 1ml de H₂O MiliQ para um microtubo estéril de 1,5ml.
 2. Juntou-se 3mm² da mancha ou 4µl de sangue total e agitou-se suavemente.
 3. Incubou-se à TA de 15 a 30 min.
 4. Agitou-se em “vortex” 5-10 seg (cuidadosamente ou inverteu-se periodicamente).
 5. Centrifugou-se 2 a 3 min a 10 000-15 000 rpm.
 6. Com micropipeta, eliminou-se todo o sobrenadante, à excepção de 20-30µl, deixando o tecido ou papel no tubo, quando se tratava de uma mancha.
 7. Juntou-se Chelex a 5% (w/v) (170-180µl) até atingir um volume final de 200µl, com uma ponta de 1ml, ou uma ponta amarela com a extremidade cortada, para ter maior diâmetro.
 8. Incubou-se a 56°C durante 15 a 30 min em banho-maria.
 9. Agitou-se vigorosamente em “vortex” 5-10 seg.
 10. Mergulhou-se em água a ferver durante 8 min, em flutuadores.
- Nota: com uma agulha estéril, perfuraram-se as tampas dos tubos, antes de os levar à ebulição.
11. Agitou-se vigorosamente em “vortex” 5-10 seg.
 12. Centrifugou-se 2 a 3 min a 10 000-15 000 rpm.
 13. Recolheu-se o sobrenadante e colocou-se noutro tubo, tendo o cuidado de não arrastar a resina.
 14. Juntou-se uma parte do sobrenadante aos tubos que já continham os reagentes necessários para efectuar a PCR (até 20% do volume total da mistura).

Reagentes:

Chelex Sigma (7901)

Soluções:

Solução stock de Chelex 100 a 5% (w/v) em H₂O MiliQ (ponta de pipeta de 1ml).

3.2.4. Extracção de DNA em esfregaços bucais com “Chelex”

- Numerou-se cada um dos microtubos de 1,5ml, previamente esterilizados, consoante o número de amostras tratadas.
- Pipetou-se 1ml de água MiliQ para cada tubo numerado.
- Adicionou-se ½ de zaragatoa a cada tubo, tendo o cuidado de fazer um controlo negativo (sem zaragatoa).
- Incubaram-se as amostras a 4°C, durante a noite.
- Centrifugaram-se as amostras 2-3 min à TA a 14 000 rpm.
- Rejeitou-se o sobrenadante, utilizando uma pipeta, à excepção de 50µl.
- Adicionou-se água MiliQ até perfazer 200µl e 2µl de Proteínase K (10mg/ml). Misturou-se devagar e com cuidado.
- Incubaram-se as amostras a 37°C (banho-maria), durante 2 horas.
- Adicionou-se 50µl de Chelex a 20% (w/v) a cada tubo.
- Incubaram-se novamente as amostras a 37°C (banho maria), durante 1 hora.
- Agitaram-se vigorosamente os tubos em “vórtex” 20 seg.
- Centrifugaram-se as amostras à TA, a 14 000 rpm, 10-20 seg.
- Levaram-se os tubos à ebulição durante 8 min, tendo o cuidado de os perfurar previamente.
- Agitaram-se as amostras em “vortex” 10 Seg e centrifugaram-se 5 min à TA, a 14 000 rpm.
- Armazenaram-se as amostras a cerca de – 10°C até posterior utilização.

Reagentes

Chelex Sigma (7901)

Proteínase K Sigma (P0390)

Soluções

Solução stock de Chelex 100 a 20% (w/v) em H₂O MiliQ.

3.3. *Quantificação do DNA*

É possível valorizar por Fluorimetria amostras que contêm DNA, utilizando uma substância que se denomina de Bisbenzimidaz, vulgarmente conhecida por Corante de Hoechst 33258 que apresenta alteração de fluorescência na presença de DNA. Este método utiliza-se, essencialmente, quando as amostras contêm quantidades reduzidas de DNA ($<10\text{ng}/\mu\text{l}$), como é o caso de vestígios biológicos analisados no âmbito de agressões sexuais.

O espectro de excitação do Hoechst 33258 é a 356nm e o de emissão a 492nm. Quando este se liga ao DNA absorve a 365nm e emite a 458nm, ou seja, a exposição de uma amostra de DNA, com a solução de Hoechst 33258, a um comprimento de onda de 365nm, através da utilização de uma lâmpada de mercúrio, causa uma excitação da mesma e posterior emissão de luz a 458nm.

Um aparelho (DyNAQuant) regista a emissão a este comprimento de onda e, a medida de fluorescência permite a quantificação directa do DNA.

Antes de se efectuar a leitura da concentração de uma amostra, tem de se utilizar um “branco”, que não contenha DNA e, posteriormente, calibrar o aparelho com uma solução padrão de DNA (DNA de Timo de Boi), de concentração conhecida. Uma vez executada esta calibração, pode-se obter directamente a concentração de DNA da amostra que se pretende quantificar.

3.4. Amplificação do DNA por PCR

A metodologia experimental utilizada para a caracterização dos fenótipos correspondentes aos sistemas polimórficos utilizados neste estudo, baseou-se na utilização da técnica de amplificação de DNA mediante a reacção em cadeia da polimerase (PCR) (Saiki e tal., 1985; Mullis e tal., 1986), seguida de electroforese capilar, no caso dos Y-STRs, utilizando o sequenciador automático ABI 310 e o Software de análise Genescan Analysis v.2.1., que detecta produtos de amplificação marcados com fluorocromos específicos, permitindo a leitura inequívoca dos diferentes tamanhos dos fragmentos.

No caso dos Y-SNPs, após amplificação de cada *locus* em separado, efectuou-se a detecção dos produtos resultantes da restrição enzimática, por electroforese horizontal em géis de Agarose com Brometo de Etídio.

Para amplificação das amostras de concentração conhecida (fluorimetria), utilizaram-se aproximadamente 5-50ng de DNA, relativamente aos STRs e SNPs. Quanto às amostras extraídas com Chelex, utilizou-se entre 10-20% de amostra em relação ao volume total da mistura de PCR.

3.4.1. Y-STRs

No caso do estudo dos Y-STRs foi seguida uma estratégia de amplificação usando vários “multiplexes”, uma vez que estes têm a vantagem de, com a mesma quantidade de DNA, proporcionarem a análise de vários *loci* em simultâneo, sem prejuízo da individualização dos alelos, à excepção do DYS385.

Multiplex 1 (Pentaplex) – amplificação simultânea dos *loci* DYS19, DYS389I/II, DYS391 e DYS393 (Kayser e col., 1997; de Knijff e col., 1997).

Mistura de reacção: Tampão 1X (Perkin Elmer), 200µM de dNTPs, MgCl₂ 2mM, 0,2µM de “primer” do DYS393, 0,4µM de “primer” do DYS19, 0,12µM de “primer” do DYS390 e 0,12µM de “primer” do DYS389 e 2U de polimerase (Taq Gold) num volume final de 25µl.

Multiplex 2 (Triplex) – Amplificação simultânea dos *loci* DYS390, DYS392 e DYS393 (Prinz *et al.*, 1997; Kloosterman *et al.*, 1998).

Mistura de reacção: Tampão 1X, 200µM de dNTPs, MgCl₂ 1.25mM, 0.2µM de “primer” do DYS391 e do DYS393 e 0.25µM de “primer” do DYS392, 2U Taq Gold (polimerase) num volume final de 50µl.

Monoplex (DYS385) – Amplificação do *locus* duplicado DYS385 (Schneider *et al.*, 1998).

Mistura de reacção: Tampão 1X, 200µM de dNTPs, MgCl₂ 3mM, 0,25µM de “primer” e 1U de Taq Gold (polimerase) num volume final de 50µl.

A sequência dos iniciadores ou “primers” já foi descrita no capítulo da Introdução. As condições de amplificação estão resumidas na Tabela 4.

A identificação dos alelos foi levada a cabo recorrendo à electroforese capilar num sequenciador automático (ABI 310). Esta permite uma separação de alta resolução dos STRs, de forma semi-automática.

Tabela 4 – Condições de amplificação dos Y-STRs

	Fluorocromo	Condições de amplificação
Multiplex 1 (pentaplex)	6 - FAM TM	95°C (10 min)
	5 - FAM TM	94°C (60s)/55°C (60s) /72°C (120s)-30 ciclos
	HEX TM	
Multiplex 2 (triplex)	5 - FAM TM	95°C (10 min.)
		94°C (60s)/55°C (60s)/72°C (120s)-30 ciclos Extensão final: 72°C 10 min.
Monoplex (DYS385)	6 – FAM TM	94°C (3 min.) 94°C (30s) /59°C-57°C (30s) (3x2 ciclos a cada temperatura) /72°C (60s)-30 ciclos Extensão final: 72°C 7 min.

A designação alélica dos sistemas correspondentes aos “multiplex” 1 e 2, efectuou-se de acordo com Kayser *e col.* (1997) e de Knijff *e col.* (1997), com excepção do DYS389, ao qual foi adicionado um bloco (GATA)₃ na contagem das repetições dos alelos (+ 3 repetições). Quanto ao DYS385, utilizou-se um “ladder” sequenciado como referência e a designação foi realizada de acordo com Schneider *e col.* (1998).

3.4.2. Marcadores bialélicos

No caso do estudo dos marcadores bialélicos, procedeu-se à amplificação individual de cada um dos SNPs e da inserção *ALU* (YAP) e à respectiva digestão enzimática, para detecção das alterações na sequência dos alelos (variantes alélicas), com excepção do YAP. Neste marcador apenas se detectou a presença ou ausência da inserção, tendo-se, a seguir à digestão enzimática, procedido à electroforese em gel de Agarose com Brometo de Etídio.

Os *loci* bialélicos analisados mediante PCR foram os seguintes: YAP, sY81, SRY-1532, SRY-8299, SRY-2627, M9 e DYS257.

Parâmetros de Amplificação:

Os “primers” (Tabela 5) foram sintetizados comercialmente e foram diluídos a uma concentração de 12,5µM.

Tabela 5. “Primers” utilizados

<i>Locus</i>	“Primers”(5'- 3')	Haplogrupo (J/T-S)*
YAP	5'-CAG GGG AAG ATA AAG AAA TA-3' 5'-ACT GCT AAA AGG GGA TGG AT-3'	4, 21, 8
sY81	5'-AGG CAC TGG TCA GAA TGA AG-3' 5'-AAT GGA AAA TAC AGC TCC CC-3'	8
SRY-1532	5'-TCC TTA GCA ACC ATT AAT CTG G-3' 5'-AAA TAG CAA AAA ATG ACA CAA GGC-3'	3, 7, 19
SRY-8299	5'-ACA GCA CAT TAG CTG GTA TGA C-3' 5'-TCT CTT TAT GGC AAG ACT TAC G-3'	21
SRY-2627	5'-AGG TCT TTT TTG CCT TCT TA-3' 5'-ATG CAC GGT TTC TTT TGA-3'	22
M9	5'-AAA ACC TAA CTT TGC TCA AGC-3' 5'-GCA GCA TAT AAA ACT TTC AGG-3'	26
DYS257	5'-GAA CTT GTC GGG AGG CAA T-3' 5'-TGA TAC ACT TCC TCC TTT AGT GG-3'	1, 3, 18, 22

* Ver correspondência no capítulo *Resultados* (Jobling e Tyler-Smith, 2000).

Mistura de reacção: Tampão 10X concentrado sem $MgCl_2$ (Perkin Elmer), $MgCl_2$ 50mM (Perkin Elmer), dNTPs a 100mM (Perkin Elmer) e Taq Polimerase 5U/ μ l.

Os iniciadores ou “primers” foram usados de modo a terem a concentração final de 1 μ M e utilizaram-se cerca de 50ng de DNA, extraído com fenol-clorofórmio, num volume total de 25 μ l.

Os parâmetros de amplificação estão resumidos na Tabela 4, tendo-se utilizado os termocicladores 9600 da Perkin Elmer e 9700 da Applied Biosystems. Após amplificação, as amostras foram mantidas a 4°C até à digestão (excepto o YAP).

Antes de se efectuar a digestão, procedeu-se à comprovação da amplificação, por electroforese, utilizando um mini-gel de Agarose a 2% numa solução de TBE 1X com Brometo de Etídio a 0,01%.

Carregaram-se 5 μ l de produto amplificado e 1 μ l de solução de carga, usando TBE 1X como tampão. Neste gel incluíram-se as amostras, um controlo positivo e um controlo negativo provenientes da amplificação.

A visualização da migração, efectuou-se após iluminação do gel por UV, num transiluminador (Macrovue LKB 2011), verificando-se se houve amplificação e com intensidade suficiente, de modo a permitir a posterior digestão dos produtos de PCR.

Digestão com enzimas de restrição:

No caso dos Y-SNPs, como se tratam de mutações pontuais, a estratégia escolhida para a sua análise consistiu em PCR-RFLPs. Uma vez amplificada a região correspondente à mutação, é necessário submeter o fragmento amplificado à digestão com a endonuclease adequada (Tabela 6), ou seja, a que permite distinguir as diferentes variantes alélicas e, posteriormente, efectuar a electroforese dos produtos da digestão, interpretando depois o padrão de bandas obtido.

No YAP apenas se detecta a presença ou ausência da inserção.

Tabela 6. – Condições de amplificação e de restrição

Locus	Condições de amplificação	Endonuclease
SRY-1532	34 ciclos 94°C (30s) / 59°C (30s) / 72°C (30s)	<i>DraIII</i>
M9	33 ciclos 94°C (30s) / 58°C (30s) / 72°C (30s)	<i>HinfI</i>
SRY-8299	33 ciclos 94°C (30s) / 62°C (30s) / 72°C (60s)	<i>BsrBI</i>
sY81	32 ciclos 94°C (30s) / 60°C (60s) / 72°C (60s)	<i>NlaIII</i>
DYS257	33 ciclos 94°C (30s) / 60°C 830s) / 72°C (30s)	<i>BanI</i>
SRY-2627	33 ciclos 94°C (30s) / 52°C (30s) / 72°C (120s)	<i>BsiHKAI</i>
YAP	33 ciclos 94°C (30s) / 54°C (30s) / 72°C (30s)	---

3.5. Separação dos fragmentos

3.5.1. Y-STRs

Recorreu-se à electroforese capilar (CE) no sequenciador automático ABI 310 (Perkin Elmer), que foi o primeiro sistema capilar introduzido no mercado, em 1995. Este sistema é capaz de detectar, de forma semi-automática, fragmentos de DNA marcados com fluorocromos que, depois de excitados por um laser, emitem fluorescência captada por uma câmara.

Depois da introdução dos novos instrumentos (aparelhos) de CE em meados dos 1990s (como é o caso do ABI 310), a técnica começou a ser largamente difundida. Esta tem vantagens relativamente aos géis tradicionais, uma vez que os passos de injeção, separação e detecção podem ser totalmente automatizados; são necessárias quantidades mínimas de amostra para o processo de injeção, de modo que as mesmas amostras podem ser novamente usadas se necessário (este é um factor importante quando se utilizam amostras vestigiais); a separação é muito mais rápida, demorando alguns minutos em vez de horas, devido à possibilidade de utilização de elevadas “voltagens”, uma vez que ocorre uma melhor dissipação de calor nos capilares; como se analisa uma amostra por corrida, não existe o problema de contaminação através dos poços adjacentes.

A todas estas vantagens junta-se o facto de se poder obter uma informação quantitativa, para além da qualitativa.

Todas estas características resultam, portanto, numa maior eficiência, resolução, sensibilidade e precisão na determinação do tamanho dos fragmentos, permitindo menor dispêndio de tempo e trabalho, designadamente na preparação dos géis. Além disso, permite uma análise multicolor, resultando uma rápida obtenção de perfis, baseados no estudo de vários *loci* em simultâneo (“multiplex”) (Moretti TR *e col.*, 2001; Buel E *e col.* 2001).

Quanto às desvantagens, a maior de todas diz respeito ao investimento inicial da aquisição do aparelho, que é dispendioso, mas que com o elevado rendimento que proporciona com a possibilidade de se poder fazer a electroforese em vários capilares

em simultâneo, no ABI 3100 (16 capilares), no final de algumas centenas de análises, é compensado.

Neste trabalho, seguiu-se o protocolo de detecção dos fragmentos de DNA marcados com fluorescência, de acordo com o manual do fabricante. Usou-se um filtro ou matriz capaz de detectar os fluorocromos utilizados, o 6-FAM™ (azul), o HEX™ (verde) e o ROX™ (vermelho). Para isso, antes de correr qualquer amostra, construiu-se a “matrix D” (Perkin Elmer P/N 401546).

Preparação das amostras

Fez-se uma mistura igual para todos os tubos que consistiu em:

- 20µl de Formamida desionizada (n+2 amostras)
- 0,5µl de um marcador interno de pesos moleculares, o GS500 ROX™ size standard (P/N 401734) (n+1 amostras).

Desta mistura, distribuiu-se 20,5µl por cada tubo próprio para a electroforese capilar e juntou-se a cada um deles, respectivamente, 1µl de uma amostra já caracterizada (controlo positivo), 1µl de controlo negativo e 1µl de amostra a caracterizar ou tipar.

Agitou-se e centrifugou-se cada um dos tubos, depois de lhes colocar os respectivos “septa” (pequenas válvulas de borracha que permitem a entrada do capilar no tubo).

Desnaturou-se cada uma das amostras, aquecendo os tubos a 95°C cerca de 3-5 minutos, num termociclador (PE 9600), arrefecendo-os logo de seguida, durante alguns minutos.

Colocaram-se as amostras no suporte, obedecendo à ordem em que foram injectadas no capilar e colocaram-se no sequenciador automático (ABI 310).

Separação dos fragmentos de DNA

As amostras foram depois injectadas no capilar durante 3seg a 15 000 V. A separação foi efectuada a 15 000 V, a uma temperatura de 60°C, durante 26-28 min, usando o polímero “310 Genetic Analyser POP™-4” (Applied Biosystems, P/N402838), o Tampão com EDTA 1X (Applied Biosystems P/N 402824) e os capilares com 47cm X 50 µm (Applied Biosystems P/N 402839).

Análise dos fragmentos

Depois dos dados da corrida serem compilados, pelo software do aparelho apropriado para o efeito, procedeu-se à sua análise usando o GeneScan® 2.1 (para Macintosh, Applied Biosystems). A designação alélica foi feita por comparação com padrões conhecidos, nomeadamente amostras provenientes de exercícios de qualidade e “ladders” alélicos sequenciados (DYS385).

Nomenclatura

Para a designação alélica, utilizou-se a nomenclatura sugerida por Kayser *e col.* (1997), exceptuando o sistema DYS389, em que se incluíram 3 unidades de repetição, por alelo (ver web-site da Universidade de Leiden).

3.5.2. Y-SNPs e YAP

A separação dos fragmentos obtidos após PCR e digestão enzimática, realizou-se mediante electroforese horizontal, em géis de Agarose a 3%, com Brometo de Etídio, de acordo com os tamanhos dos fragmentos a separar.

Juntou-se uma quantidade de Agarose correspondente ao tamanho do gel que pretendemos obter a 100 ou 250 ml de TBE 1X. Aqueceu-se a mistura até dissolução completa da agarose.

Posteriormente, deixou-se arrefecer a solução até atingir uma temperatura de cerca de 60°C, tendo-se adicionado, rapidamente, 1µl de Brometo de Etídio, por cada 100ml de solução, antes de solidificar.

Verteu-se esta mistura num molde do gel, evitando a formação de bolhas perto dos “pentes” (formação dos poços), deixando-se solidificar até se obter um gel compacto e homogéneo.

Colocou-se o gel no aparelho de electroforese, uma vez solidificado, e cobriu-se completamente com tampão TBE 1X, contendo Brometo de Etídio (0,001%). Após este procedimento, procedeu-se à aplicação das amostras.

Depositou-se, em cada poço, 10 µl de DNA digerido (contendo já 2,5 µl de solução de carga). Incluíram-se também no gel, um controlo negativo da amplificação e um marcador de pesos moleculares (100bp ladder) (2 µl de solução de carga + 1 µl de ladder + 2 µl H₂O mQ), para se poder determinar o tamanho dos fragmentos (Tabela 7).

As migrações decorreram a 100V, durante o tempo que foi necessário, a fim de se separar os fragmentos, dependendo do seu tamanho.

Após a migração, procedeu-se à visualização do gel, colocando-o num transiluminador.

Tabela 7. Tamanhos das bandas obtidas para cada sistema após restrição enzimática.

Sistema	Enzima	Alelo ancestral	Alelo moderno
SRY-1532	<i>DraIII</i>	SRY-1532*A: 167bp (não corta)	SRY-1532*G: 112bp+55bp
M9	<i>HinfI</i>	M9*C: 182bp+ 93bp+66bp	M9*G: 248bp+93bp
SRY-8299	<i>BsrBI</i>	SRY-8299*G: 362bp+147bp	SRY-8299*A: 509bp (não corta)
sY81	<i>NlaIII</i>	sY81*A: 102bp+ 65bp+ 42bp	sY81*G: 144bp+65bp
DYS257	<i>BanI</i>	DYS257*G: 182bp+ 106bp+ 63bp+ 43bp	DYS257*A: 288bp+182bp+106bp+ 63bp+43bp
SRY-2627	<i>BsiHKA1</i>	SRY-2627*C: 1242bp (não corta)	SRY-2627*T: 944bp+298bp
YAP	---	YAP- Fragmento de 150bp	YAP+ Fragmento de 455bp

2.6. Cálculo Estatístico

2.6.1. Análise intrapopulacional

Na análise do cromossoma Y, os parâmetros com interesse médico-legal são: as frequências dos haplótipos e a Diversidade Génica/Haplotípica (DH), que neste caso, é igual ao Poder de Discriminação (PD) (Sensabaugh, 1982) e ao valor de Probabilidade Média de Exclusão (MEC) (Chakravorty e Li, 1983), utilizado nos testes de paternidade.

2.6.1.1. Frequências alélicas e haplotípicas

Tendo em conta que os sistemas analisados são de herança haplóide, o cálculo das frequências alélicas e haplotípicas (p_i) realizou-se por contagem simples dos alelos ou haplótipos (Li, 1976):

$$p_i = \frac{x_i}{n}$$

em que x_i é o número de vezes que se repete o alelo com n cópias.

O erro padrão associado à estimativa das frequências (s.e.) foi calculado do seguinte modo:

$$\text{s.e.} = \sqrt{\frac{p(1-p)}{N}}$$

em que N representa o número total de indivíduos analisados e p o valor da frequência alélica.

2.6.1.2. Diversidade Génica/Haplotípica

A Diversidade Génica (D.G.) / Haplotípica (D.H.), foi calculada usando a seguinte expressão:

$$D.H.=1-\sum_{i=1}^n p_i^2$$

Sendo p as frequências alélicas/haplotípicas e n o número de indivíduos/haplótipos. Esta fórmula equivale à heterozigosidade autossómica (Nei, 1987).

2.6.2. Comparação de populações

A diferenciação populacional foi testada usando o Software Arlequin versão 2000 (Schneider *et al.*, 2000).

Estimaram-se componentes de variância e estatísticos F , para cada *locus* separadamente. Os distintos componentes de variância, a diferentes níveis, foram combinados para obter estimativas de estatísticos F , somando os componentes de variância num determinado nível hierárquico, no numerador e denominador, produzindo outros estatísticos F , como razões de componentes de variância. Deste modo, os estatísticos F globais não são simples médias aritméticas de estatísticos F de cada *locus*.

Os estatísticos F dos pares de *locus* foram usados como medidores de distâncias entre populações a curto prazo, aplicando-se uma transformação ligeira para linearizar a distância com o tempo de divergência entre populações (Reynolds *et al.*, 1983; Slatkin, 1995). Estes valores de F_{ST} são dados na forma de uma matriz.

2.6.2.1. Número médio de diferenças entre pares (π)

Este parâmetro é o número médio de diferenças entre todos os pares de haplótipos na amostra populacional. É muito útil em genética forense para fazer uma estimativa do número médio de diferenças entre os indivíduos, provenientes da mesma população. Calcula-se através da expressão:

$$\pi = \sum \sum p_i p_j d_{ij}$$

onde, d_{ij} é uma estimativa do número de mutações ocorridas desde a divergência dos haplótipos i e j ; n é o número de haplótipos; p_i e p_j são as frequências dos haplótipos i e j , respectivamente.

2.6.2.2. Estrutura da população: Análise molecular da Variância (AMOVA)

A estrutura genética da população é verificada a partir da análise da variância, como foi definido inicialmente por Cockerham (1969, 1973) e completada por outros estudiosos (Weir e Cockerham, 1984; Long, 1986). A análise da variância molecular partindo da informação sobre haplótipos/haplogrupos de DNA, é usada no programa Arlequin (AMOVA, Excoffier *et al.*, 1992).

A variância total, assumindo a ausência de recombinação entre posições e neutralidade selectiva, obtém-se a partir da seguinte fórmula:

$$V(\pi) = \frac{3n(n+1)\pi + 2(n^2 + n + 3)\pi^2}{11(n^2 - 7n + 6)}$$

Definindo grupos de populações, o utilizador define uma determinada estrutura genética que irá ser testada. Esta análise molecular da variância, permite estimar parâmetros análogos aos estatísticos F, os quais reflectem a correlação da diversidade haplotípica nos diferentes níveis hierárquicos de subdivisão: entre grupos, entre

populações de um mesmo grupo e intrapopulacional (Weir, 1996; Excoffier, 2000; Rousset, 2000).

No caso de uma estrutura simples, consistindo em indivíduos haplóides nas populações, o índice F_{ST} , originado pelo algoritmo utilizado no programa, é exactamente igual ao estatístico F médio definido por Cockherham (1984), tendo em consideração os vários *loci* (Slatkin, 1991).

Os dados em que se baseia a dita análise são, por um lado, as frequências dos haplótipos (Y-STRs) ou haplogrupos (Y-SNPs e YAP) analisados e, por outro, uma matriz de distâncias que reflecte o número de eventos evolutivos que separam cada haplótipo/haplogrupo dos restantes; ou seja, o número de mutações que separam os ditos haplótipos/haplogrupos.

2.6.2.3. *FSTs linearizados de Slatkin's*

Slatkin (1995) considera um modelo demográfico simples onde duas populações haplóides de tamanho N divergiram τ gerações, a partir de uma população de tamanho idêntico. Estas duas populações permaneceram isoladas desde então, sem trocar migrantes. Sob tais condições, o F_{ST} pode-se exprimir em termos de tempo de coalescência $\overline{t_1}$, que é o tempo de coalescência médio de dois genes originários de duas populações diferentes, sendo $\overline{t_0}$ o tempo de coalescência médio de dois genes da mesma população, usando a análise da estratégia de variância. Os F_{ST} s exprimem-se do seguinte modo:

$$F_{ST} = \frac{\overline{t_1} - \overline{t_0}}{\overline{t_1}}$$

Uma vez que t_0 é igual a N gerações e t_1 é igual a $\tau + N$ gerações, a expressão anterior fica:

$$F_{ST} = \frac{\tau}{\tau + N}$$

Logo, a razão $D = F_{ST}/1-F_{ST} = \tau/N$. Sendo, por isso, proporcional ao tempo de divergência entre duas populações.

2.6.2.3.1. *Y-STRs*

Obteve-se uma matriz de distâncias genéticas entre 16 populações, incluindo a analisada, segundo Slatkin (1995). A partir desta matriz de distâncias, construíram-se árvores “Neighbor-joining” (Saitou and Nei, 1987) e a robustez dos nós foi estimada pelo método do “Bootstrap” (Efron, 1982). A matriz das distâncias genéticas para as 16 populações consideradas, foi também representada num mapa bi-dimensional, através de um escalonamento multidimensional não métrico (Torgerson, 1952; Lalouel, 1980).

2.6.2.3.2. *Marcadores bialélicos*

Obteve-se uma matriz de distâncias genéticas entre 60 populações, incluindo a analisada, segundo Slatkin (1995). A partir desta matriz de distâncias, construíram-se árvores “Neighbor-joining” (Saitou and Nei, 1987) e a robustez dos nós foi estimada pelo método do “Bootstrap” (Efron, 1982). A matriz das distâncias genéticas para as 60 populações consideradas, foi também representada num mapa bi-dimensional, através de um escalonamento multidimensional não métrico (Torgerson, 1952; Lalouel, 1980).

4. RESULTADOS

4. RESULTADOS

4.1. STRs

A partir do estudo de 122 indivíduos saudáveis e não relacionados do Norte de Portugal, obtiveram-se 111 haplótipos (Tabela 8.) por contagem simples, tendo-se posteriormente determinado as respectivas frequências e outros parâmetros estatísticos, utilizando o programa Arlequin 2000 (Excoffier *e col.*, 1992).

Nos 26 pares pai-filho estudados, não se observou nenhuma incompatibilidade. Seria necessário estudar muitos mais dos referidos pares, para que se pudesse calcular com rigor uma taxa de mutação.

Tabela 8. Haplótipos de Y-STRs na população do Norte de Portugal (n=122)

Haplótipo	total	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385
P001	1	13	12	30	23	10	11	14	16-16
P002	1	13	12	31	24	10	11	14	17-18
P003	1	13	13	30	22	10	15	14	14-16
P004	1	13	13	30	24	9	11	14	13-15
P005	1	13	13	30	24	9	11	14	14-14
P006	1	13	13	30	24	9	12	15	13-14
P007	1	13	13	30	24	10	11	13	13-16
P008	1	13	13	30	24	11	14	14	11-14
P009	1	13	13	31	25	10	11	15	14-17
P010	1	13	13	31	25	11	11	14	17-18
P011	1	13	13	32	23	10	11	13	13-17
P012	1	13	13	32	24	10	13	14	15-16
P013	1	13	14	30	24	9	11	13	13-14
P014	1	13	14	31	24	9	11	14	13-13
P015	1	13	14	31	24	9	11	15	14-15
P016	1	13	14	31	24	11	11	14	13-14
P017	1	14	10	27	24	11	14	14	11-14
P018	1	14	11	28	23	11	13	14	12-14
P019	1	14	12	28	23	10	13	15	10-15
P020	1	14	12	28	24	10	13	15	10-15
P021	1	14	12	29	23	10	11	13	13-17
P022	1	14	12	29	23	10	11	14	13-14
P023	1	14	12	29	23	11	11	13	12-14
P024	2	14	12	29	23	11	13	14	11-14
P025	2	14	12	29	23	11	14	14	11-14
P026	1	14	12	29	23	11	15	14	11-14

P027	1	14	12	29	25	10	13	14	11-14
P028	1	14	12	29	25	11	13	14	11-14
P029	1	14	12	30	22	10	15	14	13-14
P030	1	14	12	30	23	10	11	14	14-15
P031	1	14	12	30	23	11	15	14	11-15
P032	1	14	13	29	22	11	13	13	11-14
P033	1	14	13	29	24	10	13	14	11-13
P034	1	14	13	29	24	10	13	14	11-14
P035	1	14	13	29	24	11	13	13	11-13
P036	1	14	13	29	24	11	13	13	11-14
P037	1	14	13	29	25	10	11	12	13-14
P038	1	14	13	29	25	10	13	15	12-14
P039	1	14	13	30	23	10	11	13	13-17
P040	1	14	13	30	23	10	11	13	14-15
P041	1	14	13	30	23	10	13	13	13-17
P042	2	14	13	30	23	10	13	14	11-14
P043	2	14	13	30	23	11	13	14	11-14
P044	1	14	13	30	23	11	13	14	11-15
P045	1	14	13	30	23	11	13	14	12-14
P046	1	14	13	30	24	10	13	13	12-14
P047	1	14	13	30	24	10	13	14	18-18
P048	1	14	13	30	24	10	13	15	12-14
P049	1	14	13	30	24	11	13	13	11-15
P050	1	14	13	30	24	11	13	13	12-14
P051	1	14	13	30	24	11	13	14	11-11
P052	1	14	13	30	24	11	13	14	11-13
P053	3	14	13	30	24	11	13	14	11-14
P054	1	14	13	30	24	11	13	15	10-15
P055	3	14	13	30	24	11	13	15	11-14
P056	1	14	13	30	24	11	14	14	11-14
P057	1	14	13	30	24	12	12	14	11-12
P058	1	14	13	30	24	12	13	14	11-14
P059	1	14	13	30	25	10	13	14	11-14
P060	1	14	13	30	25	11	13	14	11-14
P061	1	14	13	30	25	11	13	14	11-15
P062	1	14	13	31	22	10	11	13	13-17
P063	1	14	13	31	23	9	11	14	17-19
P064	1	14	13	31	23	10	13	14	11-14
P065	1	14	13	31	23	11	13	14	12-15
P066	1	14	13	31	24	10	13	14	11-15
P067	1	14	13	31	24	11	11	14	16-16
P068	1	14	13	31	24	11	13	13	11-13
P069	1	14	13	31	24	11	13	13	11-14

P070	2	14	13	31	24	11	13	14	11-14
P071	1	14	14	30	24	10	13	14	11-14
P072	1	14	14	30	24	10	13	14	12-14
P073	1	14	14	31	24	11	13	14	11-14
P074	1	14	14	31	24	11	13	14	11-15
P075	1	14	14	31	24	11	15	14	11-14
P076	1	14	14	31	25	11	13	14	11-14
P077	1	14	14	32	24	10	13	14	11-14
P078	1	14	14	32	24	10	13	14	11-15
P079	1	14	14	33	23	10	11	14	16-18
P080	1	15	12	27	23	11	14	14	12-14
P081	1	15	12	29	22	10	11	14	13-14
P082	1	15	12	29	24	10	11	14	13-17
P083	1	15	12	30	22	10	11	15	14-14
P084	1	15	12	31	22	10	11	15	14-15
P085	1	15	12	31	22	10	12	15	14-14
P086	2	15	13	30	22	10	11	14	13-14
P087	1	15	13	30	23	10	11	13	13-18
P088	1	15	13	30	23	10	13	13	13-17
P089	1	15	13	30	23	11	13	13	12-14
P090	1	15	13	30	23	11	13	14	11-14
P091	1	15	13	30	23	11	13	14	11-15
P092	1	15	13	30	24	10	13	14	11-15
P093	1	15	13	30	24	11	11	14	11-14
P094	1	15	13	30	24	11	13	13	11-13
P095	1	15	13	30	24	11	13	13	11-14
P096	2	15	13	30	24	11	13	14	11-14
P097	1	15	13	31	21	10	11	15	16-16
P098	1	15	13	31	23	11	11	13	13-15
P099	1	15	13	31	24	11	13	16	11-14
P100	1	15	14	30	22	10	11	13	13-15
P101	1	15	14	31	24	11	13	14	11-14
P102	1	15	14	32	23	11	13	14	11-14
P103	1	15	14	32	24	11	13	14	11-15
P104	1	16	12	30	21	10	11	15	15-15
P105	1	16	13	28	23	10	11	13	12-12
P106	1	16	13	30	23	9	11	13	13-13
P107	1	16	13	30	25	10	11	14	11-14
P108	1	16	14	31	23	11	13	15	15-15
P109	1	17	13	29	23	10	11	14	12-12
P110	1	17	13	29	23	10	11	14	12-13
P111	1	17	13	31	26	10	13	14	11-13

4.1.1. Análise por locus

As frequências alélicas, de cada um dos STRs analisados, são apresentadas na Tabela 9., bem como a respectiva Diversidade Génica. A maior Diversidade Génica corresponde ao locus DYS385 (0.847), seguida da do locus DYS389II (0.677). O locus DYS389I é o que apresenta o menor valor (0.535). A Diversidade Génica média dos loci estudados é de 0.626.

Tabela 9. Frequências alélicas e Diversidade Génica para cada um dos STRs estudados.

DYS19			DYS390			DYS385I/II		
Alelo	Freq.	s.d.	Alelo	Freq.	s.d.	Alelo	Freq.	s.d.
13	0.131	0.031	21	0.016	0.012	10-15	0.025	0.014
14	0.590	0.045	22	0.090	0.026	11-11	0.008	0.008
15	0.213	0.037	23	0.328	0.043	11-12	0.008	0.008
16	0.041	0.018	24	0.467	0.045	11-13	0.049	0.020
17	0.025	0.014	25	0.090	0.026	11-14	0.361	0.044
Diversidade génica: 0.592			26	0.008	0.008	11-15	0.082	0.025
			Diversidade génica: 0.663			12-12	0.016	0.012
						12-13	0.008	0.008
						12-14	0.082	0.025
						12-15	0.008	0.008
						13-13	0.016	0.012
						13-14	0.074	0.024
						13-15	0.025	0.014
						13-16	0.008	0.008
						13-17	0.057	0.021
						13-18	0.008	0.008
						14-14	0.025	0.014
						14-15	0.033	0.016
						14-16	0.008	0.008
						14-17	0.008	0.008
						15-15	0.016	0.012
						15-16	0.008	0.008
						16-16	0.025	0.014
						16-18	0.008	0.008
						17-18	0.016	0.012
						17-19	0.008	0.008
						18-18	0.008	0.008
						Diversidade génica: 0.847		
DYS389I			DYS391					
Alelo	Freq.	s.d.	Alelo	Freq.	s.d.			
10	0.008	0.008	9	0.066	0.023			
11	0.008	0.008	10	0.434	0.045			
12	0.197	0.036	11	0.484	0.045			
13	0.639	0.044	12	0.016	0.012			
14	0.148	0.032	Diversidade génica: 0.577					
Diversidade génica: 0.535								
DYS389II			DYS392					
Alelo.	Freq.	s.d.	Alelo	Freq.	s.d.			
27	0.016	0.012	11	0.328	0.043			
29	0.172	0.034	12	0.025	0.014			
30	0.492	0.045	13	0.557	0.045			
31	0.230	0.038	14	0.049	0.020			
32	0.049	0.020	15	0.041	0.018			
33	0.008	0.008	Diversidade génica: 0.582					
Diversidade génica: 0.677								
			DYS393					
			Alelo	Freq.	s.d.			
			12	0.008	0.008			
			13	0.213	0.037			
			14	0.631	0.044			
			15	0.139	0.031			
			16	0.008	0.008			
			Diversidade génica: 0.542					

4.1.2. *Análise haplotípica e parâmetros com interesse forense*

A análise dos *loci* mencionados nos 122 indivíduos, permitiu identificar 111 haplótipos diferentes. Os dois haplótipos mais frequentes, aparecem repetidos três vezes e correspondem ao conjunto dos alelos mais frequentes em cada um dos sistemas, quando considerados de forma isolada, com exceção do sistema DYS393 em que o alelo 15 está presente sendo, no entanto, o terceiro alelo mais frequente neste sistema. Os referidos haplótipos são: DYS19*14, DYS389I*13, DYS389II*30, DYS390*24, DYS391*11, DYS392*13, DYS393*14, DYS385*11-14 e DYS19*14, DYS389I*13, DYS389II*30, DYS390*24, DYS391*11, DYS392*13, DYS393*15, DYS385*11-14.

O conjunto dos marcadores STRs do cromossoma Y estudados na população do Norte de Portugal, permitiu calcular um parâmetro com interesse médico-legal, a Diversidade Haplotípica, equivalendo ao PD (Poder de Discriminação) nos STRs autossómicos, cujo valor é de 0.998 ± 0.001 . Trata-se de um valor bastante elevado, porque a cada indivíduo estudado corresponde, praticamente, um haplótipo. Ou seja, esta elevada diversidade deveu-se à escassa repetição de haplótipos, pois o maior número de vezes que um haplótipo se repete é de três, o que aconteceu apenas duas vezes; seguido de sete haplótipos que se encontram representados duas vezes cada um. Portanto, a probabilidade de que dois indivíduos tomados ao acaso da população pertençam à mesma linhagem paterna, é de 0,002.

4.1.3. *Comparações populacionais*

Recorreu-se ao programa Arlequin 2000 (Excoffier e col., 1992) para analisar dados haplotípicos referentes às seguintes populações: Norte de Portugal, Astúrias (Martínez-Jarreta B e col., 2003), Andaluzia (Prieto V e col., 2003), Barcelona (Borrego N e col., 2003), pertencentes à Península Ibérica; Alemanha (Hidding M e Schmitt, 2000), Bavária (Anslinger K e col., 2000), Budapeste (C. Hungria) e Romenos (residentes em Budapeste) (Füredi e col., 1999), Itália (Cagliá A e col., 1998), Grécia (Parreira KS e col., 2002), Turquia (Henke J e col., 2001), Albânia (Robino C e col., 2003), agrupadas como populações europeias e Índia (Henke J e col., 2001), China (Hidding M e Schmitt, 2000), Japão (Nonaka I e Minaguchi K, 2001) e Coreia (Shin DJ e col., 2001), fazendo

parte do grupo das populações asiáticas, num total de 16 populações. Não se fizeram comparações com populações do continente africano, porque não se dispunha de dados referentes aos haplótipos de todos os marcadores estudados.

Deste estudo comparativo, verificou-se que a nossa população (Norte de Portugal) partilha o maior número de haplótipos com as populações da Península Ibérica, sete com a população das Astúrias (Norte de Espanha), seis com Barcelona (zona mediterrânica da Península Ibérica) e cinco com a região da Andaluzia (Sul de Espanha). À medida que nos afastamos da Península Ibérica, a partilha de haplótipos diminui progressivamente, tendo-se verificado apenas a referida partilha de quatro dos haplótipos com a Alemanha, três com a Hungria, dois com a Itália e um com a Turquia, Albânia e Romenos.

O haplótipo 14-13-29-24-11-13-13-11-14, correspondente aos *loci* ordenados: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 e DYS385I/II, é o mais frequente nas populações espanholas e é partilhado pela população do Norte de Portugal e por todas as populações europeias anteriormente referidas, excepto pela Itália. É de destacar que não se observou a presença de nenhum haplótipo comum entre a população estudada e as populações asiáticas, da Índia, China, Japão e Coreia.

A análise da AMOVA assinala que a maior percentagem de variância é intrapopulacional (Tabela 10.). Relativamente à variância interpopulacional, esta análise indica que o *locus* DYS19 acumula a maior variância, no que respeita aos europeus e mongóis, enquanto o DYS392 apresenta a maior variância no grupo dos mongóis e da Índia (Tabela 11.).

Tabela 10. Análise de AMOVA

Origem da variação	Entre grupos populacionais	Interpopulacional	Intrapopulacional	Total
Soma de quadrados	298,86	155,94	5570,77	6025,57
Componentes de variância	0,1395	0,1023	2,5577	2,7995

Tabela 11. Análise de AMOVA nas populações comparadas

AMOVA para europeus e mongóis					AMOVA para Indianos e mongóis				
Locus	SSD	d.f.	Va	% variação	Locus	SSD	d.f.	Va	% variação
DYS19	42.11227	1	0.04944	12.71821	DYS19	0.78802	1	0.00106	0.30781
DYS389I	30.96649	1	0.03691	10.99343	DYS389I	10.08158	1	0.03647	9.59567
DYS389II	4.89981	1	0.00229	0.61044	DYS389II	2.02745	1	0.00117	0.30225
DYS390	8.25997	1	0.00486	1.32221	DYS390	4.10486	1	-0.00439	-1.17797
DYS391	12.11837	1	0.01307	5.08429	DYS391	1.51128	1	0.00178	0.91647
DYS392	29.28607	1	0.02889	8.37636	DYS392	17.59987	1	0.08559	19.75880
DYS393	7.95740	1	-0.00061	-0.20447	DYS393	0.56574	1	-0.01823	-5.83276
DYS385	20.66778	1	0.02309	4.72376	DYS385	5.67157	1	0.02369	4.69860

Calculou-se a distância genética de Slatkin (1995) entre a população do Norte de Portugal e populações de países europeus, entre os quais os da Península Ibérica e países asiáticos (Tabela 12.).

Tabela 12. Distância genética de Slatkin entre as populações (1) Norte de Portugal, (2) Astúrias, (3) Andaluzia, (4) Barcelona, (5) Alemanha, (6) Bavária, (7) Budapeste (C Hungria), (8) Itália, (9) Grécia, (10) Turquia, (11) Albânia, (12) Romenos (residentes em Budapeste), (13) Índia, (14) China, (15) Japão e (16) Coreia.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.07105	0.00000														
3	0.06611	0.00238	0.00000													
4	0.08548	0.00086	0.00155	0.00000												
5	0.07857	0.01384	0.02230	0.03355	0.00000											
6	0.09589	0.01759	0.02287	0.03427	0.00121	0.00000										
7	0.11163	0.04761	0.03987	0.06482	0.01569	0.01691	0.00000									
8	0.08075	0.02929	0.02412	0.04994	0.01290	0.02015	0.01261	0.00000								
9	0.09884	0.04991	0.03809	0.06902	0.02581	0.02396	0.00000	0.00571	0.00000							
10	0.10603	0.07371	0.06161	0.10119	0.04224	0.04628	0.02061	0.00858	0.00807	0.00000						
11	0.17112	0.09453	0.08236	0.10428	0.08263	0.07628	0.05409	0.06749	0.04699	0.07249	0.00000					
12	0.23839	0.20137	0.17410	0.22360	0.13653	0.13902	0.07697	0.10973	0.09113	0.08068	0.19617	0.00000				
13	0.13327	0.10791	0.09337	0.13178	0.05774	0.05382	0.03345	0.04515	0.03149	0.02888	0.11424	0.06060	0.00000			
14	0.14987	0.12336	0.12115	0.14656	0.10234	0.12427	0.10121	0.07868	0.10366	0.07540	0.17841	0.21936	0.10700	0.00000		
15	0.18564	0.12659	0.11487	0.14053	0.09561	0.10961	0.07652	0.07767	0.07884	0.07578	0.16125	0.11962	0.07031	0.08481	0.00000	
16	0.16514	0.11258	0.11472	0.13315	0.10032	0.11724	0.10024	0.07497	0.09772	0.08007	0.19753	0.19643	0.10620	0.02642	0.05072	0.00000

O dendrograma obtido a partir destas distâncias genéticas (Figura 5) mostra que a população estudada se inclui dentro das linhagens do cromossoma Y da Península Ibérica, Europa Central e Itália, diferenciando-se das populações do Sul, Oriente da Europa e das asiáticas.

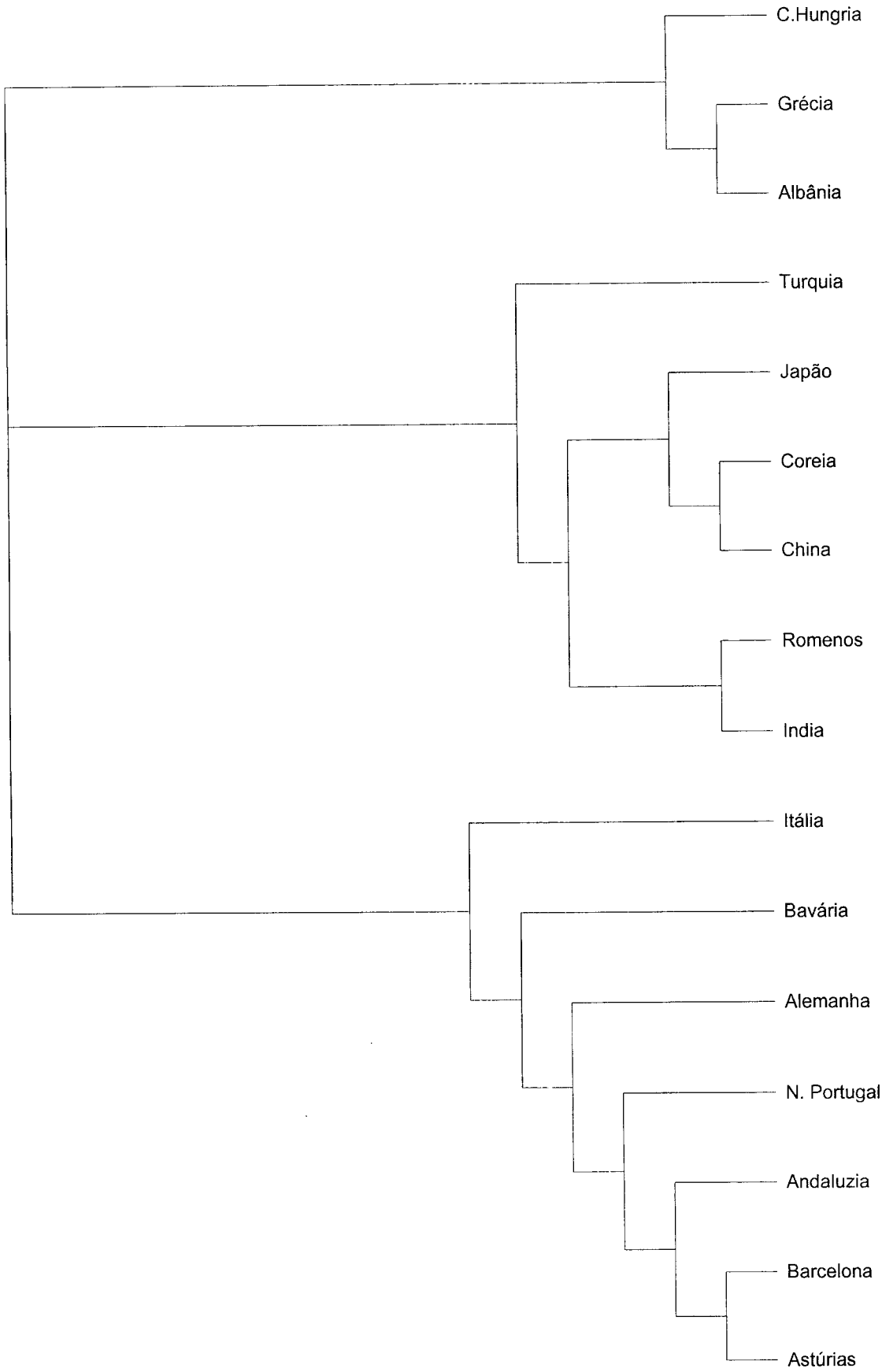


Figura 5. Dendrograma obtido a partir da distância de Slatkin.

Foi também construída uma árvore “neighbor-joining” consenso, para as mesmas 16 populações, usando os haplótipos construídos a partir dos sistemas DYS19, DYS389I/II, DYS390, DYS391, DYS392, DYS393 e DYS385 (Figura 6), baseada no “bootstrapping” das distâncias genéticas segundo Slatkin (1995).

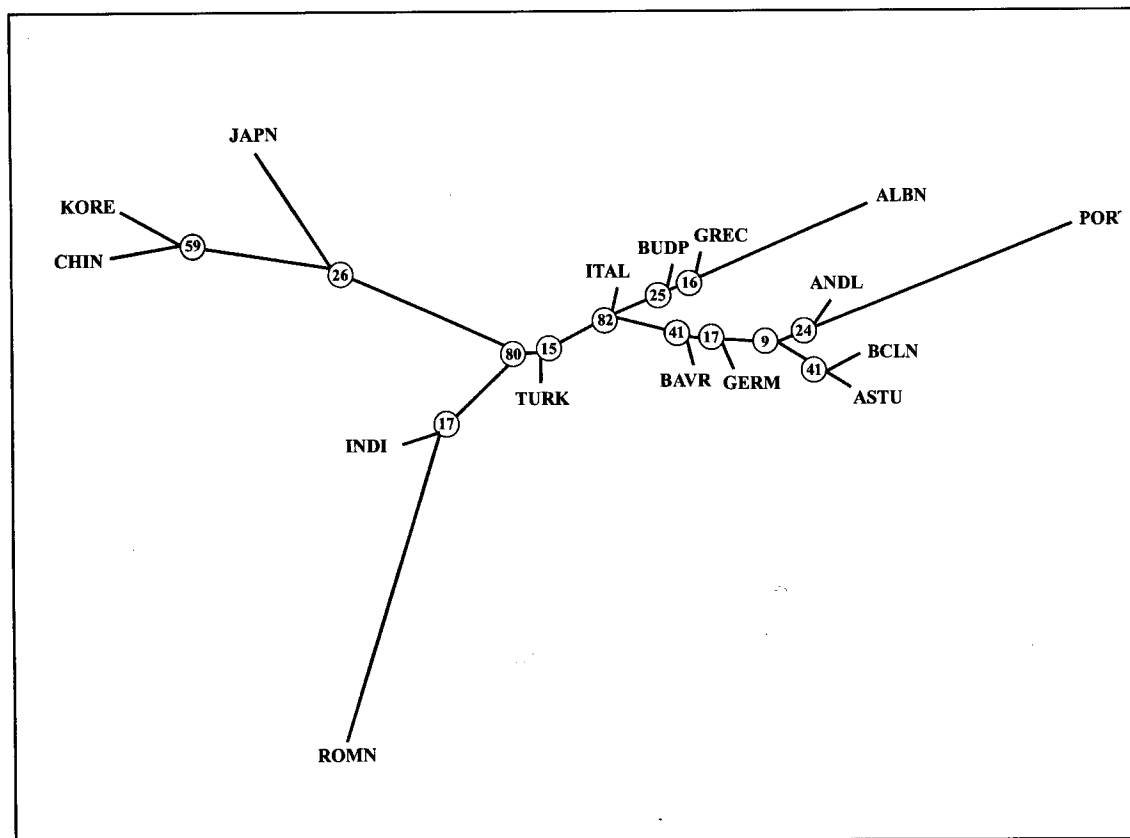


Figura 6. Árvore “neighbor-joining” baseada nas distâncias genéticas de Slatkin (1995), para 16 populações de países europeus entre os quais os da Península Ibérica e países asiáticos.

Procedeu-se também a um escalonamento multidimensional não métrico (MDS), a partir das mesmas distâncias genéticas (Slatkin, 1995), representado na Figura 7. A variância total considerada para a redução dos eixos vectoriais, foi de 96.4 % sendo o coeficiente de stress de 0.088.

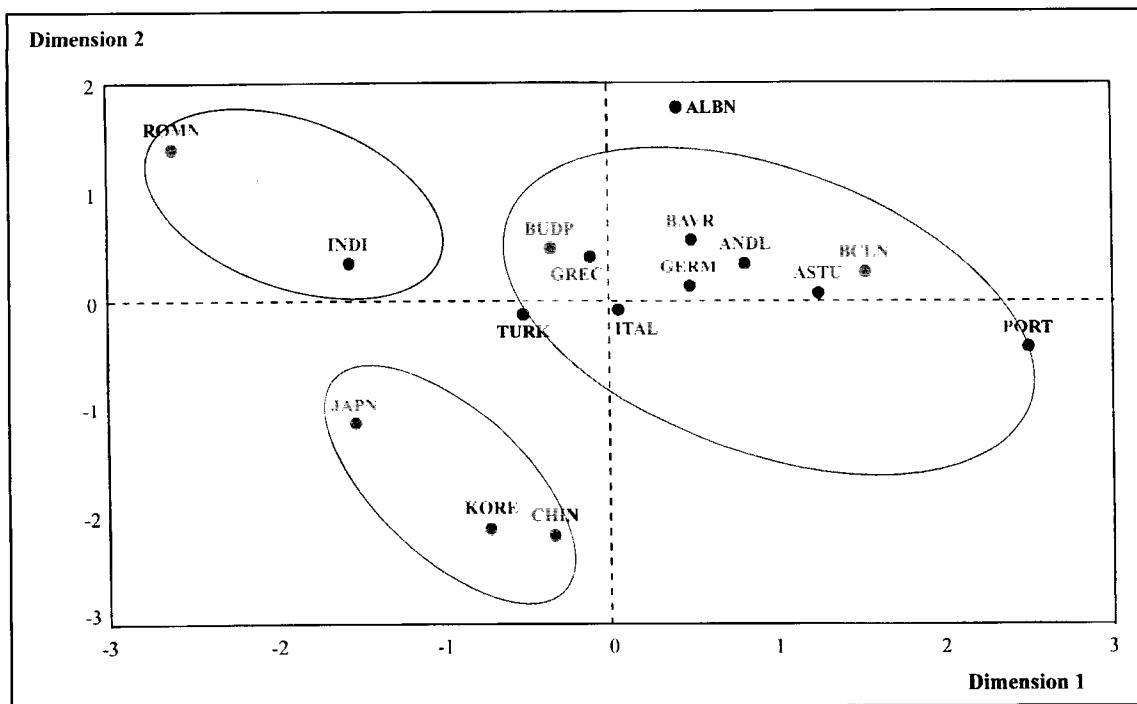


Figura 7. Escalonamento multidimensional (MDS) não métrico aplicado às distâncias genéticas segundo Slatkin (1995) das 16 populações comparadas.

As várias representações gráficas indicam que a nossa população fica incluída no mesmo ramo das restantes da Península Ibérica que, por sua vez, se aproximam das populações do centro da Europa, todas elas situadas acima do eixo positivo da primeira dimensão. Observa-se uma separação das populações mediterrâneas, que se situam sobre o eixo negativo da primeira dimensão, com exceção da Itália que constitui o ponto de união entre estas populações e as do Centro da Europa e da Península Ibérica. Na análise MDS a população albanesa separa-se de todas as demais, mas na árvore da Figura 6 pode-se observar que se situa no ramo das populações mediterrâneas. Por sua vez, a Turquia conflui num nóculo que se separa em dois ramos: por um lado as populações do Norte da Ásia e, por outro, as populações da Índia e os romenos.

4.1.4. Aplicação dos Y-STRs na conclusão de perícias médico-legais

Apresentam-se alguns exemplos dos três tipos de perícias médico-legais do âmbito da Genética e Biologia Forense que são solicitadas, habitualmente, em que o estudo dos Y-STRs é bastante pertinente e tem demonstrado um interesse relevante na sua conclusão. Nalguns casos, que vão ser descritos, não se usaram todos os marcadores do haplótipo mínimo, pois foram concluídos antes deste protocolo ser usado na prática laboratorial quotidiana. Por outro lado, como algumas delas foram concluídas recentemente, para além dos polimorfismos analisados neste trabalho, foram estudados também o DYS437, DYS438, DYS439, GATA A7.1 (DYS460), GATA A7.2 (DYS461), GATA A10, GATA C4 e GATA H4 e que já fazem parte da rotina laboratorial, cujos resultados, eram concordantes com os daqueles.

Em relação à investigação de paternidade, citam-se dois casos em que, depois de se efectuar o estudo de STRs autossómicos, se verificou a existência de incompatibilidade em dois sistemas polimórficos, provavelmente devidas a mutação genética; um caso de identificação genética individual, relativamente a um cadáver e dois casos de agressão sexual, num dos quais havia apenas um suposto agressor e noutro, mais do que um.

Investigação de paternidade

Caso 1 :

Estudou-se um trio “normal” composto por suposto pai, filho (sexo masculino) e mãe, em que se analisou um fragmento do gene homólogo da Amelogenina e 17 STRs autossómicos em dois “multiplex”, utilizando dois “kits” comerciais, o PowerPlex®16 System (Promega) e o AmpFℓSTR® Identifiler™ (Applied Biosystems), cujos resultados se apresentam na Tabela 13. A análise destes resultados permitiu constatar que havia duas incompatibilidades, relativamente aos sistemas HUMTH01 e CSF1PO. Após o cálculo do Índice de Paternidade, em que se considerou todos os sistemas analisados, a relação mãe-filho como certa, as taxas de mutação para os sistemas em que se observaram incompatibilidades e uma probabilidade *a priori* de 50%, para as duas hipóteses pertinentes neste caso, ou seja, a de paternidade e a de não paternidade, obteve-se um valor de $IP = 647,100181$ para a primeira hipótese relativamente à segunda (com $W = 0,99845703$ e $W = 0,00154297$, respectivamente).

Em face deste elevado Índice de Paternidade, do facto da taxa de mutação dos STRs ser relativamente elevada e de nos casos em que se conclui por exclusão de paternidade, ela ocorre num número mais elevado de polimorfismos, a primeira hipótese (a de paternidade) é muito mais consistente do que a segunda. Para além da informação disponível, referente aos STRs autossómicos, considerou-se que o estudo dos Y-STRs, seria uma boa estratégia para complementar a perícia, apesar de se saber que todos os indivíduos relacionados com o pretenso pai, por linha paterna, possuem a mesma informação genética para o cromossoma Y; no entanto, deve-se ter presente que quando se colhem amostras biológicas a um determinado indivíduo, indicado pela mãe como pretenso pai, há indícios favoráveis ao facto deste ser o pai biológico e não outro indivíduo com ele aparentado. Por sua vez, estes indícios são, geralmente, corroborados por outras provas, para além da pericial.

Através do resultado da análise dos Y-STRs verificou-se não haver nenhuma incompatibilidade entre o haplótipo obtido para o suposto pai e o filho, sendo o haplótipo em questão: 14, 13, 29, 24, 10, 13, 13, 11-14, seguindo a ordem dos sistemas já utilizada. É de referir que ao consultar a base de dados do cromossoma Y disponível no site <http://ystr.charite.de>, indicada para estas situações, se encontraram 171 indivíduos com este haplótipo em 13 253 caracterizados para 93 populações europeias.

Em face do exposto, constatou-se a coincidência da existência de duas incompatibilidades em dois polimorfismos, provavelmente devidas a mutação genética prevalecendo a hipótese de paternidade relativamente à de não paternidade. A sequenciação dos alelos dos marcadores em questão, poderá permitir a confirmação destas mutações.

Caso 2 :

Noutra investigação de paternidade, que foi questionada em relação a dois filhos gémeos, sendo um deles do sexo masculino (filho B), após a análise dos sistemas autossómicos já referidos para o caso anterior, não se obteve qualquer incompatibilidade relativamente à filha (filho A) e obtiveram-se duas incompatibilidades em relação ao filho varão, desta feita nos sistemas D21S11 e D19S433 (Tabela 13), obtendo-se um Índice de Paternidade de ($W = 99,8197692$). Neste caso, assumimos como certa a relação mãe-filhos, tendo-se colocado cinco hipóteses: os filhos são irmãos germanos e filhos do suposto pai; apenas o filho A é filho do suposto pai e os gémeos resultaram de fecundação heteroparental; apenas o filho B é filho do suposto pai e os gémeos resultaram de fecundação heteroparental; nenhum dos filhos é filho do suposto pai, sendo irmãos germanos e filhos de um indivíduo tirado ao acaso da população, que não o suposto pai; nenhum dos filhos é filho do suposto pai e resultaram de fecundação heteroparental. Todas as hipóteses referidas são equiprováveis *a priori*.

Depois do estudo dos Y-STRs, verificou-se, novamente, a partilha do mesmo haplótipo, o haplótipo 14-13-30-24-9-12-13-11-14, por parte do suposto pai e do filho do sexo masculino e, mais uma vez, a primeira hipótese é muito mais favorável, isto é, a de paternidade e dos dois filhos serem irmãos germanos. Na base de dados já referida, não encontramos qualquer indivíduo que possuísse o mesmo haplótipo, em 13 085 haplótipos caracterizados em populações da Europa.

Tabela 13. Resultados obtidos em dois casos de investigação de paternidade.

Sistemas estudados	Pretenso pai/caso1	Filho caso1	Mãe caso1	Pretenso pai/caso2	Filho A caso2	Filho B caso2	Mãe caso2
D3S1358	15-16	15-16	15-17	12-15	12-15	12-16	15-16
HUMVWA	16-18	16-18	15-18	16-18	15-16	15-18	15-18
D16S539	12-13	12	11-12	9-11	11-12	11-12	12
D2S1338	16-20	16-17	17-20	16-24	16-19	19-24	17-19
Amelogenina	X-Y	X-Y	X	X-Y	X	X-Y	X
D8S1179	15-16	13-15	13-14	12-13	12-14	13-14	14
D21S11	31.2-34.2	28-34.2	28-31.2	28-29	29	<u>29-30</u>	28-29
D18S51	12-15	12-16	16	14-16	13-14	13-16	13
D19S433	14.2-16	14-16	13-14	13-14.2	11-14.2	<u>11-13.2</u>	11-12
HUMTH01	6-9	<u>9-10</u>	9	7-9.3	7-9.3	9.3	8-9.3
FGA	22-26	22-25	24-25	20-25	23-25	20-24	23-24
D5S818	10-11	11-12	11-12	10-12	10-12	10-12	12
D13S317	11-12	11-14	13-14	11-13	12-13	11	11-12
D7S820	10-11	10	10-12	10-12	10-11	10-11	10-11
HUMTP0X	8	8-12	11-12	8-11	11	8-11	8-11
HUMCSF1PO	10-12	<u>11-13</u>	11	10-12	12	12	11-12
Penta D	11-12	9-12	9-12	9	9	9	9-14
Penta E	12-21	7-12	7-17	5-13	5	5	5-12
DYS19	14	14	--	14	--	14	--
DYS389I	13	13	--	13	--	13	--
DYS389II	29	29	--	30	--	30	--
DYS390	24	24	--	24	--	24	--
DYS391	10	10	--	9	--	9	--
DYS392	13	13	--	12	--	12	--
DYS393	13	13	--	13	--	13	--
DYS385	11-14	11-14	--	11-14	--	11-14	--

Identificação genética individual

Tratou-se de um caso de identificação de cadáver do sexo masculino, em que se dispunha de 5 pretensos irmãos, dos quais três eram do sexo masculino. Após o estudo de vários STRs autossômicos (PowerPlex®16 System), recorreu-se aos Y-STRs para complementar o estudo, uma vez que em relação àqueles se detectaram três incompatibilidades, nos sistemas HUMTH01, D5S818 e FGA (Tabela 14).

Tabela 14. Resultados obtidos relativos à identificação genética de um cadáver.

Sistemas estudados	Cadáver	Suposto Irmão A	Suposto Irmão B	Suposto Irmão C	Suposta Irmã D	Suposta Irmã E
D3S1358	15	15-17	15-17	11-15	11-17	15-17
HUMVWA	15-18	17	16-18	17	16-17	17-18
D16S539	11-13	9-12	9-11	9	9-11	9
Amelogenina	X-Y	X-Y	X-Y	X-Y	X	X
D8S1179	13	12	11-14	12-14	11-12	11-14
D21S11	28-30	31	28-29	31	28-29	31
D18S51	13-15	12-14	14	14-15	14-15	14
HUMTH01	<u>6-9,3</u>	<u>6-9</u>	<u>7-8</u>	<u>8-9</u>	<u>8-9</u>	<u>6-9</u>
FGA	<u>23</u>	<u>21,2-24</u>	<u>21-24</u>	<u>21,2-24</u>	<u>21-24</u>	<u>21-25</u>
D5S818	<u>8-9</u>	<u>12-13</u>	<u>12-13</u>	<u>11-12</u>	<u>12-13</u>	<u>12-13</u>
D13S317	8-14	8	11	8-11	8	11
D7S820	12-13	9-12	8-12	9-12	9-13	8-12
HUMTP0X	8	8	8	8	8	8
HUMCSF1PO	12	9-11	9-11	11-12	11-12	11-12
Penta D	11-12	12	12	9-13	12-13	9-12
Penta E	11-13	5-7	5-7	5-7	5-7	5-11
DYS19	14	14	14	14	--	--
DYS389I	12	14	14	14	--	--
DYS389II	30	30	30	30	--	--
DYS390	22	23	23	23	--	--
DYS393	13	12	12	12	--	--

Neste caso, o haplótipo obtido para o cadáver desconhecido apresentava vários alelos distintos, relativamente aos dos supostos irmãos, ou seja, o alelo 12 no cadáver, para o DYS389I, correspondia ao 14 para os supostos irmãos; o alelo 22 no cadáver, para o DYS390, correspondia ao 23 para os supostos irmãos e o alelo 13 no cadáver, para o DYS393, correspondia ao 12 para os supostos irmãos. Estes resultados estão de acordo com os obtidos com o estudo dos sistemas autossómicos, pelo que se concluiu pela inexistência da relação de parentesco entre o cadáver e os supostos irmãos. Os haplótipos referidos, correspondem a alguns sistemas do haplótipo mínimo, amplificados em “pentaplex” e foram, respectivamente, 14-12-30-22-13, em relação ao indivíduo a identificar e 14-14-30-23-12, para os seus pretensos irmãos. Na base de dados já referida, mesmo com o haplótipo com um número inferior de polimorfismos,

encontrou-se 13 indivíduos em 13 244 europeus, que partilhavam o haplótipo do cadáver e 24 em 13 244 indivíduos europeus, que partilhavam o haplótipo dos supostos irmãos do cadáver.

Agressões sexuais

Caso 1 :

Foi-nos presente material relativo a suposta agressão sexual, designadamente exsudato vaginal colhido à queixosa, do qual se extraiu DNA e uma zaragatoa bucal colhida a um suspeito de ter cometido o crime, da qual também se extraiu DNA. Neste caso, o perfil genético obtido da fase espermática do exsudato vaginal, resultante tanto da análise de STRs autossómicos como de alguns Y-STRs (“pentaplex”), era idêntico ao perfil genético do DNA extraído da zaragatoa bucal colhida ao suspeito, para os sistemas estudados, ou seja, 14-13-30-24-13. Este haplótipo aparece em 244 de 13 244 indivíduos europeus incluídos na base de dados citada.

Caso 2:

Noutra suposta agressão sexual, dispôs-se de uma peça de vestuário da queixosa (cuecas), em que havia suspeita do crime ter sido cometido por mais de um agressor. Efectivamente, após análise de alguns Y-STRs (“pentaplex”), verificou-se a existência de um perfil genético que pode corresponder a uma mistura de células de, pelo menos, dois indivíduos do sexo masculino, uma vez que estão presentes dois picos para os sistemas DYS19 (14 e 16), DYS389I (9 e 10), DYS390 (21 e 25), DYS393 (12 e 13) e apenas um pico para o DYS389II (30). Embora não se possa, com rigor, a partir da área dos picos inferir da quantidade de DNA da amostra, verificou-se que nas mesmas condições de amplificação e separação dos fragmentos obtidos, num dos perfis (14-13-30-25-12) os picos têm uma área superior, podendo corresponder ao maior contribuidor na mistura. A referida mistura pode incluir o haplótipo 14-13-30-25-12 e o 16-12-30-21-13. O primeiro haplótipo aparece em 16 de 13 244 europeus na citada base de dados e, o segundo, em apenas um indivíduo. Neste caso, não nos foi presente qualquer suspeito, com o qual se pudesse fazer comparação.

Em resumo, a análise dos Y-STRs proporcionou uma informação muito valiosa, em todos os casos referidos, demonstrando-se mais uma vez a sua grande utilidade na conclusão de perícias médico legais de resolução mais complexa.

4.2. Marcadores bialélicos

4.2.1. Distribuição das frequências alélicas dos loci SNP e YAP do cromossoma Y analisados

Dos 9 haplogrupos possíveis com os 7 sistemas estudados (SNPs e YAP do cromossoma Y) nesta amostragem (n=99), só encontramos 5. Como podemos constatar o haplogrupo mais frequente é o P*, com 36,4% de frequência (Tabela 15), seguido do BR*, como é comum nos países europeus.

Com o estudo destes polimorfismos obtivemos um valor de Diversidade Génica de $0,6891 \pm 0,0236$.

As correspondências destes haplogrupos na nomenclatura do YCC (Y Chromosome Consortium, 2002) com a usada por Jobling e Tyler-Smith (2000), estão assinaladas na Tabela 15.

Não se encontrou nenhum indivíduo pertencente aos haplogrupos Y*, DE*, R1a e E3a.

Tabela 15. Frequências dos haplogrupos na população do Norte de Portugal.

HAPLOGRUPO		N (frequência absoluta)	Frequência relativa
J/T-S	YCC		
2	BR*	35	0,354
7	Y*	0	0,000
26	K*	13	0,131
4	DE*	0	0,000
1	P*	36	0,364
22	R1b8	4	0,040
21	E*	11	0,111
3	R1a	0	0,000
8	E3a	0	0,000
		N Total=99	

- J/T-S – Nomenclatura segundo Jobling e Tyler-Smith (2000)
- YCC – Nomenclatura sugerida pelo “Consortium” do cromossoma Y (2002)

4.2.2. Haplogrupos construídos a partir dos loci bialélicos estudados

Com o objectivo de comparar os resultados obtidos neste estudo com outros publicados (Karafet *e col.*, 1999; Rosser *e col.*, 2000), realizou-se em primeiro lugar uma equivalência dos haplogrupos segundo a nomenclatura mais actual, sugerida pelo YCC (Y Chromosome Consortium, 2002) e a seguida por Jobling e Tyler-Smith (2000).

Neste trabalho analisaram-se 7 polimorfismos, tendo-se definido 9 haplogrupos, usando-se a nomenclatura binária para fazer referência ao estado ancestral (0) e ao derivado (1), de cada um dos sistemas (Tabela 16).

A mutação mais antiga corresponde a uma alteração no sistema SRY-1532 e constitui a base da árvore filogenética.

O haplogrupo Y* representa o estado ancestral para todos os loci e corresponde ao haplogrupo 7 de Jobling e Tyler-Smith (2000), sendo frequente em populações subsarianas africanas e ausente em populações europeias, incluindo a nossa (Karafet *e col.*, 1999).

O haplogrupo BR* apresenta o alelo moderno para o sistema SRY-1532 e todos os outros ancestrais nos restantes sistemas. Agrupa os haplogrupos 2 e 9 de Jobling e Tyler-Smith. É considerada a linhagem mais antiga da Europa (Rosser *e col.*, 2000), com elevada frequência, que poderá ser constituída por sub-linhagens com provável prevalência regional.

O haplogrupo K* apresenta o alelo moderno para o sistema já referido, bem como para o M9. Localiza-se num dos nódulos internos da árvore filogenética e, provavelmente, tal como o HG BR* poderá ser subdividido noutras linhagens. Ocorre com baixa frequência na nossa população, à semelhança de outros estudos (Rosser *e col.*, 2000).

O haplogrupo P*, apresenta o alelo moderno no sistema DYS257, para além dos dois já referidos. Corresponde ao haplogrupo 1 de Jobling e Tyler-Smith. É um haplogrupo frequente na Europa e apresenta um gradiente de frequência que diminui desde cerca de 98,5% de frequência nos países do Noroeste europeu para Sudeste, no Cáucaso, zona onde terá surgido a agricultura.

O haplogrupo E* corresponde ao haplogrupo 21 segundo Jobling e Tyler-Smith e apresenta o alelo moderno no sistema SRY-8299, relativamente ao DE*, que não foi referido por não se ter encontrado no nosso estudo. É frequente no Sul da Europa, apresentando uma elevada frequência em populações do Norte de África, existindo um

gradiente regional que pode reflectir a barreira genética entre a África e a Europa. Nos países do Sul da Europa têm sido descritas frequências de 10-20%, o que tem sido interpretado com base em factores climatéricos (Rosser *e col.*, 2000).

O haplogrupo R1b8, apresenta o alelo moderno no sistema SRY-2627 relativamente ao P*, tem sido descrito em poucas populações da Europa, como França e Península Ibérica e foi analisado em detalhe por Hurles *e col.* (1999), tendo sido sugerida a sua origem na Península Ibérica. Corresponde ao haplogrupo 22 de Jobling e Tyler-Smith (2000). Aparece na nossa população com frequência de 0,04 (4%).

Quanto aos haplogrupos DE*, R1a e E3a, não aparecem na nossa população e correspondem, respectivamente, aos haplogrupos 4, 3 e 8 de Jobling e Tyler-Smith (2000), restringindo-se o primeiro a países asiáticos (Karafet *e col.*, 1999), o terceiro aparece apenas em populações do Norte de África, com cerca de 5% de frequência. O R1a apresenta frequências mais elevadas em países do centro e Este europeu, apresentando um gradiente que decresce de Nordeste para Sudoeste, sendo as suas frequências para Sueste e Sudoeste muito reduzidas (Rosser *e col.*, 2000).

Tabela 16. Haplogrupos definidos neste trabalho.

Haplogrupo	SRY-1532	YAP	M9	SRY-8299	DYS257	SY81	SRY-1532	SRY-2627
Y* (xBR,A2)	0	0	0	0	0	0	0	0
BR* (xB2b,CE,F1,H,JK)	1	0	0	0	0	0	0	0
K* (xK1,LN,O2B,O3C,	1	0	1	0	0	0	0	0
DE* (xE)	1	1	0	0	0	0	0	0
P* (xR1b,R1a,Q3)	1	0	1	0	1	0	0	0
E* (xE3)	1	1	0	1	0	0	0	0
R1a	1	0	1	0	1	0	1	0
E3a	1	1	0	1	0	1	0	0
R1b8	1	0	1	0	1	0	0	1

Em seguida, construiu-se uma rede filogenética, utilizando o programa Network 1.1 (Bandlet *e cols.*, 1999). A figura 8 mostra a “network” obtida com as frequências encontradas neste trabalho.

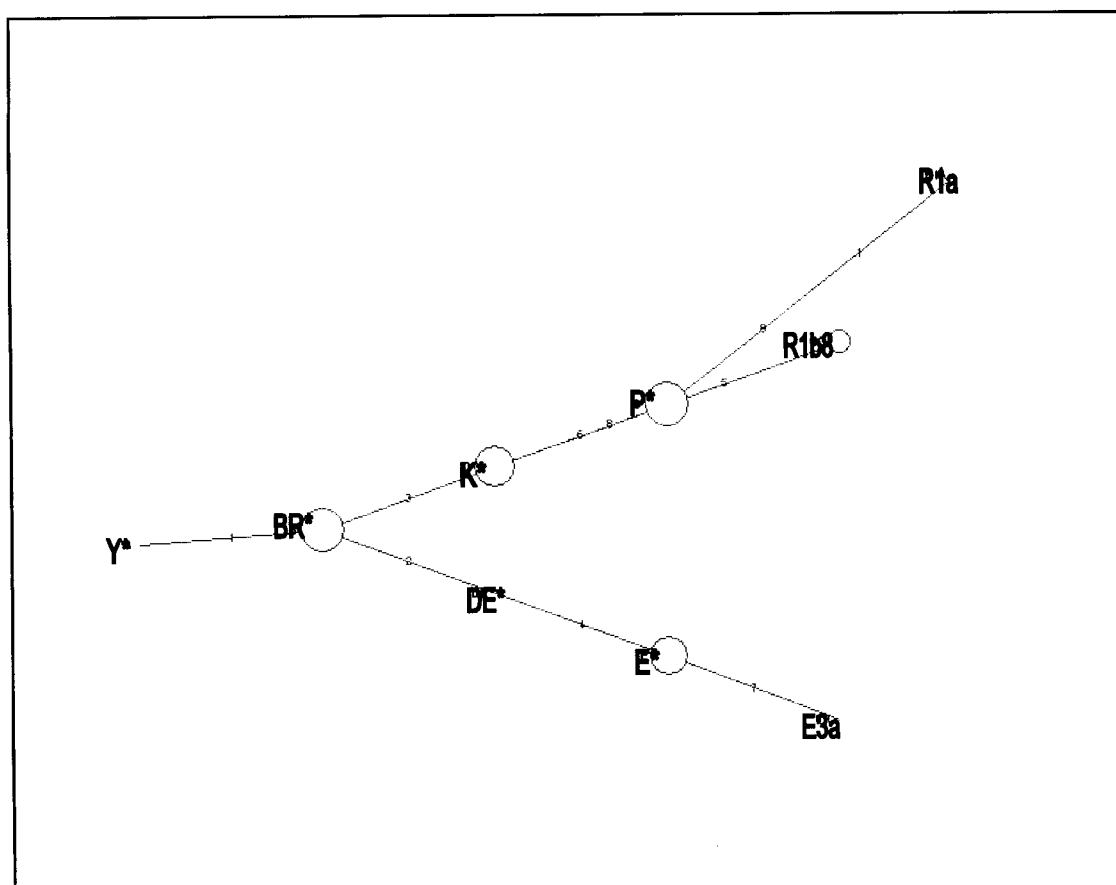


Figura 8. “Network” determinada a partir das frequências dos haplogrupos.

4.2.3. Distâncias genéticas-árvore filogenética

Com o objectivo de se estudar os agrupamentos que se produziam, quando comparamos a nossa população com outras, construiu-se uma árvore “neighbor-joining”, baseada nas distâncias de Slatkin (1995), para 60 grupos populacionais, incluindo os referidos na bibliografia, usando os marcadores bialélicos (SNPs e YAP) estudados (Figura 9). As populações utilizadas nas comparações foram as seguintes:

Norte da Europa: Islândia (Icel), “Saami” (Saam), Norte da Suécia (Nswe), Gotlândia (Gotl), Noruega (Norw), Dinamarca (Dani) e Finlândia (Finn); Europa de Leste: a Estónia (Esto), Letónia e Lituânia (Latv), Rússia (Russ), Bielo-Rússia e Ucrânia (Bela), Geórgia (Geor), Ossétia e Arménia (Osse), Turquia (Turk), Bulgária (Bulg), “Czech” (Czec), Eslováquia (Slov), Roménia e Jugoslávia (Roma), Eslovénia e Hungria (Slve),

Polónia (Poli) e Swanetia (Swan); Centro da Europa: Holanda (Duct), Bavária (Bava), França (Fren), Bélgica (Belg), Escócia ocidental (WSco), Inglaterra de Leste (EAng), Irlanda (Íris) e Inglaterra (Brit); Sul da Europa: a população do Norte de Portugal estudada (Port), Itália (Ital) Sul e Norte de Portugal (SNP), Espanha (Span), Bascos de Arratia e Goierri (Arr); África: Egipto (Egyp), Argélia (Alge), Norte de África (Nafr), Berberes (Bere), “Khoisan” (Khoi), Pigmeus (Pygm), “Bantus” de Leste (Eban), Gambia (Gamb), “Bantus” de Oeste (Wban) e África de Leste (EAfr); Austrália: aborígenes australianos (AAPe), Papua Nova Guiné e Melanésia (Papu); Ásia: “Nentsi” da Tundra (Tun), “Selkups”(Selk), “Buryats”(Bury) e “Evenks” da Sibéria (Even), “Mongóis”(Mong), Japão (Japa), Taiwan e Sul da China (Taiw), Coreia e Indonésia (Kore), Asiáticos de Sudeste (SEAs) e Índia (Indi); Ameríndios: “Inuit” esquimós (Inui), “Navajos” (Nava), “Cheyenne” (Chey) e Pima, Pueblos, Ngöbe e Mixtecs (Pima). Estas populações são as referidas por Karafet *TM e col.* (1999), Rosser *ZH e col.* (2000) e M de Pancorbo *e col.* (2003).

Não se considerou o sistema DYS2627, nas comparações efectuadas, uma vez que não foi usado por Karafet *e cols.* (1999) e, deste modo, o haplogrupo R1b8 (que apresenta uma frequência de 0,04 na nossa população) considerou-se incluído no P*. As distâncias genéticas entre as mesmas populações estão representadas numa escala bidimensional na Figura 10. A variância total considerada para a redução dos eixos vectoriais foi de 97.995 % e o coeficiente de “stress”, de 0.08603.

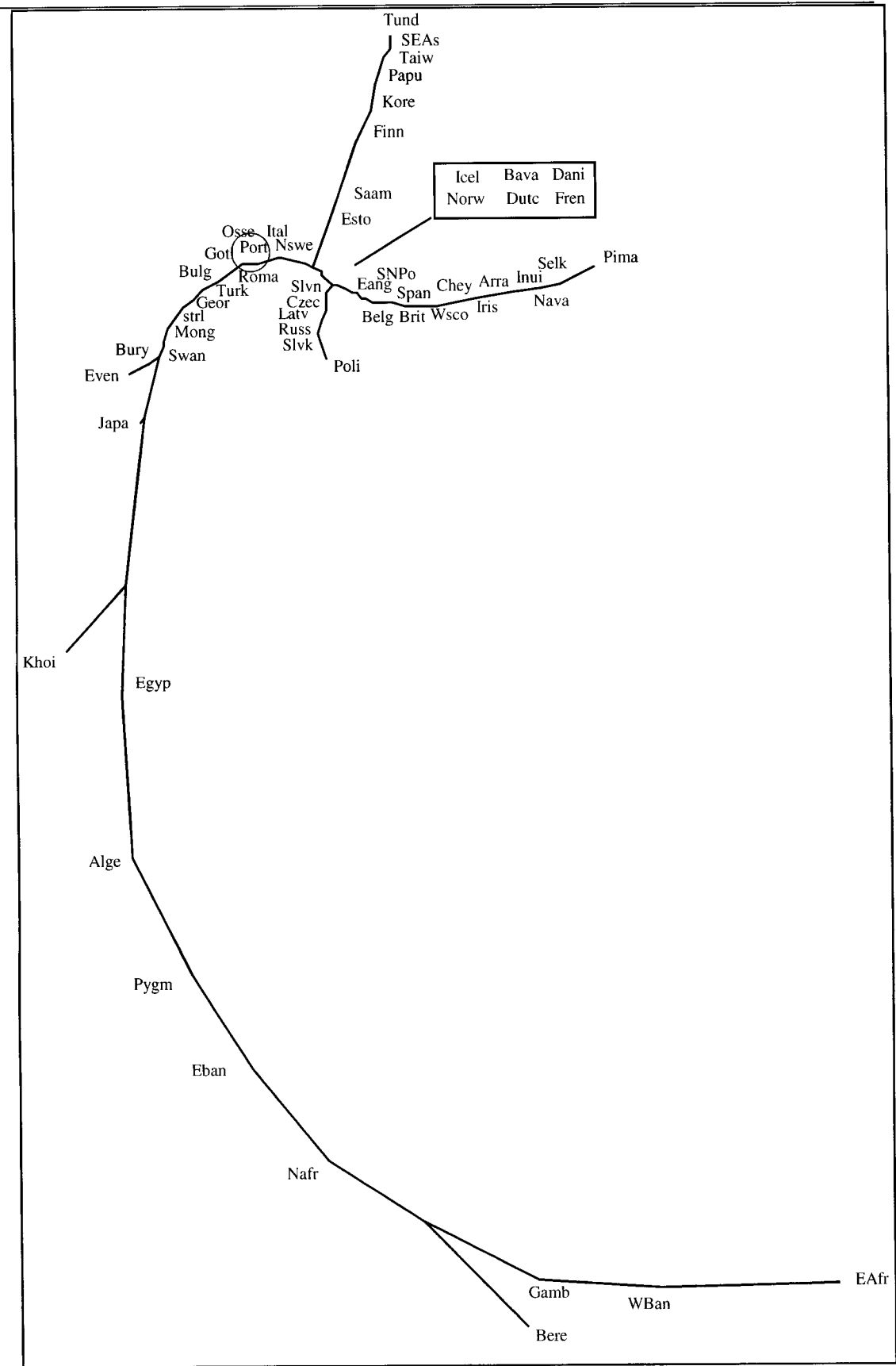


Figura 9. Árvore “neighbor-joining” baseada nas distâncias de Slatkin (1995).

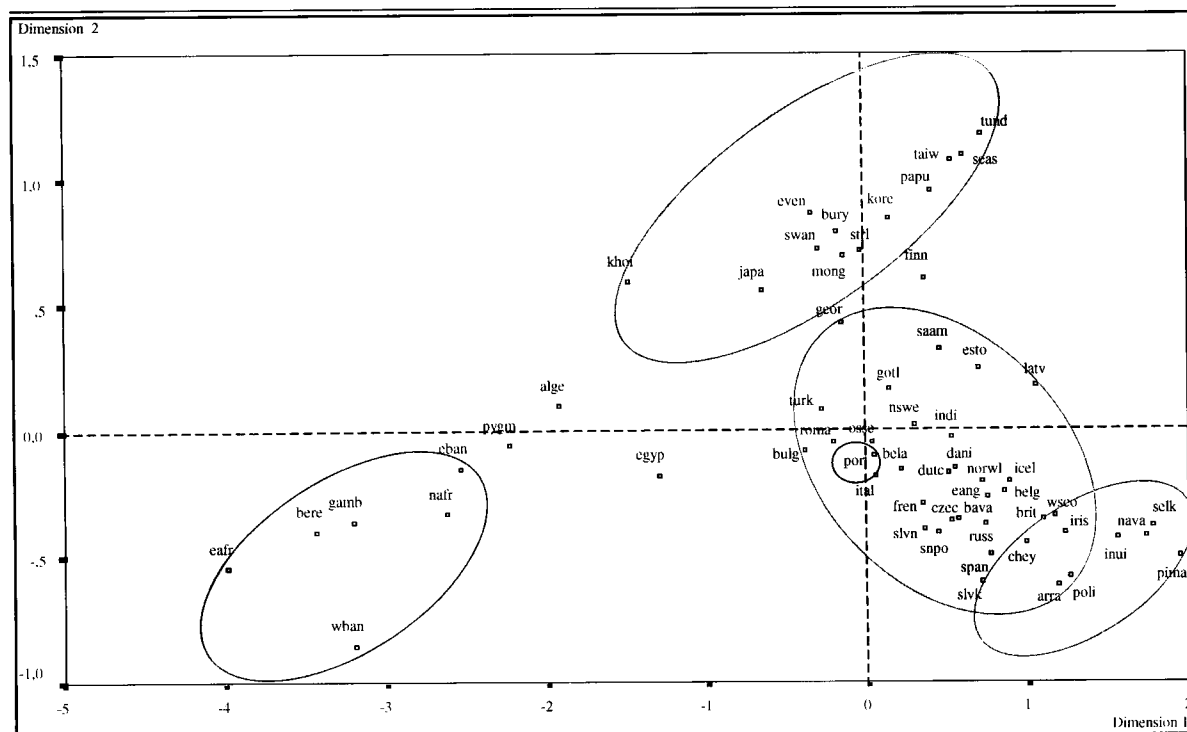


Figura 10. Escalonamento multidimensional (MDS) não métrico aplicado às distâncias genéticas segundo Slatkin (1995) dos 60 grupos populacionais comparados.

No que se refere aos haplogrupos das populações comparadas, verificamos que, no geral, cada população é significativamente diferente das restantes, incluindo a nossa. No que diz respeito às distâncias genéticas (valores de F_{ST}), quase sempre com valores de $p \leq 0,05$ (significando uma elevada subestruturação), constata-se haver, portanto, uma grande especificidade populacional.

Ao realizar o escalonamento multidimensional não métrico, verificamos a existência de grupos populacionais distintos (“clusters”), representados na Figura 10 com diversas cores, correspondendo a diferentes zonas geográficas.

Constatámos que existe um “cluster” (cor verde) que inclui as populações africanas, podendo-se até distinguir o grupo do centro e Sul, do Norte de África.

Outro “cluster” importante (cor de laranja) é constituído pelas populações asiáticas (Mongóis, Japoneses, Tailandeses, Chineses e povos da Tundra).

No “cluster” onde se incluem as populações europeias (cor azul céu), pode-se distinguir pequenos grupos, tais como o Oriente próximo, Cáucaso e Este europeu por um lado, caucasianos e ameríndios por outro e, finalmente, o Noroeste da Europa. A população estudada neste trabalho inclui-se neste grupo (cor azul forte). Este “cluster” abrange as populações nativas americanas ou ameríndios (cor verde), uma vez que não foi

analisado o SNP (DYS199T) que permitiria distingui-las das restantes (Carvajal-Carmona *e col.*, 2000; M de Pancorbo *e col.*, 2003). Este agrupamento pode, no entanto, indicar a influência da colonização europeia na América desde há 500 anos, apesar de se atribuir uma origem asiática às populações nativas americanas (Karafet *e col.*, 1999).

5. DISCUSSÃO

5. Discussão

Os polimorfismos do Cromossoma Y têm sido gradualmente introduzidos nas perícias forenses. O cromossoma Y apresenta baixos níveis de polimorfismo, quando comparado com os cromossomas autossómicos; no entanto, tem utilidade em alguns casos criminais, em particular as agressões sexuais, perícias de filiação mais complexas e identificação genética.

5.1. Amostragem

As amostras seleccionadas obtiveram-se a partir de indivíduos do sexo masculino, nas quais se analisaram Y-STRs e polimorfismos bialélicos, em ordem a determinar haplótipos e haplogrupos representativos de linhagens masculinas. O número de indivíduos estudado foi relativamente baixo, embora permita ter uma ideia muito aproximada do que será a realidade da distribuição das frequências haplotípicas na nossa população. No que se refere aos marcadores bialélicos, poderíamos detectar um número mais elevado de haplogrupos se aumentássemos a amostragem. Por este facto, tencionamos prosseguir o estudo populacional utilizando uma metodologia de tipagem mais eficaz, possivelmente a “PCR em tempo real”.

Por outro lado, seria interessante calcular taxas de mutação, através do estudo de um número mais elevado de pares pai-filho.

5.2. Caracterização dos sistemas estudados no cromossoma Y

Relativamente ao estudo dos Y-STR, o recurso à PCR seguida de análise dos fragmentos no sequenciador automático que utilizam a CE, provou ser uma metodologia eficaz e relativamente simples.

O caminho a seguir de futuro será o de tentar aumentar o número de marcadores analisados em simultâneo e, se possível, acrescentar-lhe mais alguns sistemas informativos, como aliás foi sugerido muito recentemente por Butler (2002). A principal razão para esta necessidade será a de numa única análise obter a maior informação possível, uma vez que muitas vezes temos amostras em quantidades exíguas.

Quanto aos marcadores bialélicos, como já foi referido no capítulo da *Introdução*, a técnica que foi utilizada neste estudo, embora seja simples e útil para analisar um pequeno número de marcadores numa amostragem relativamente pequena, como foi o caso, torna-se fastidiosa, consumindo demasiado tempo e impedindo a análise de todos os marcadores disponíveis em todas as amostras. Um bom exemplo deste problema é o do estudo europeu realizado por Rosser *e col.* (2000), em que marcadores do cromossoma Y que se pensava estarem ausentes da Europa, a partir de estudos anteriores, não foram tipados e em que muitos dos marcadores que foram analisados, foram tipados de forma hierárquica, baseados na informação filogenética.

Para obviar estes e outros problemas, como por exemplo, a interpretação de misturas, no campo da Genética e Biologia Forense, terá de se recorrer a métodos mais eficazes e de elevado rendimento como, por exemplo, a PCR em tempo real, como já foi referido.

5.3. STRs

Embora tenham sido descritos diferentes tipos de polimorfismos, os STRs provaram ser, até ao momento, os marcadores mais indicados em Genética e Biologia Forense, devido aos seus índices de diversidade, à sua relativamente baixa taxa de mutação e à simplicidade técnica de análise.

5.3.1. Diversidade Génica/ Haplotípica

A Diversidade Génica é menor nos países do Ocidente europeu. Este padrão é compatível com a expansão demográfica de Leste para Oeste nestes países. Esta constatação também se pode dever ao facto dos países do Leste Europeu se situarem num local de confluência de migrações entre África, Europa e Ásia. Há autores que sugerem que a expansão da agricultura terá sido um fenómeno mais cultural do que propriamente migratório. Torroni *e col.* (2001) propõem que as populações situadas junto da bacia do Mediterrâneo e no Golfo da Vizcaya, se poderão ter aí refugiado durante a época da última glaciação, há cerca de 15 000 anos, uma vez que são zonas de clima mais temperado. Findo este período, verificou-se o movimento migratório

contrário, ainda na época dos caçadores recolectores. Pode-se pensar, portanto, que após um refúgio em vários locais da Península e no Centro da Europa, terá ocorrido uma dispersão e mistura com aportes genéticos da expansão da agricultura, constituindo a chamada expansão démica pós-glaciar. No entanto, um estudo recente vem argumentar que o fenómeno predominante será o da expansão da agricultura, tradicionalmente defendida por Cavalli-Sforza (Chikhi L *e col.*, 2002). Situação análoga tem sido descrita relativamente ao mtDNA (Rosser *e col.*, 2000).

As Diversidades Génicas de cada *locus*, no estudo efectuado, variam entre 0,356 para o DYS19 e 0,567 para o DYS385. Este último *locus* é o que apresenta maior poder de discriminação, à semelhança do que acontece na grande maioria das populações, sendo o *locus* que mais contribui para o valor de Diversidade Haplotípica (0,998). Este é o parâmetro com mais interesse sob o ponto de vista de identificação genética. Este valor significa que com o estudo destes sistemas consegue-se construir haplótipos, que distinguem diferentes indivíduos de uma população, sendo de destacar que, a maior parte dos haplótipos constituem linhagens únicas de cromossomas Y.

Uma vez que do ponto de vista forense, o principal objectivo consiste na individualização de pessoas ou linhagens, a estratégia analítica escolhida deve permitir diferenciar a maioria dos haplótipos encontrados.

Neste estudo o número mais elevado de haplótipos partilhados é de três, verificando-se para dois deles (P053 e P055), seguido de duas vezes para mais sete haplótipos (P024, P025, P042, P043, P070, P086 e P096).

Como são sistemas herdados de forma haplóide, o valor da Diversidade Haplotípica corresponde ao Poder de Discriminação e à CE (“Exclusion Chance”). Os valores encontrados permitem deduzir que estes sistemas demonstraram ser úteis na resolução de perícias do âmbito médico-legal, tanto na área criminal como em testes de parentesco, quando estão envolvidos indivíduos do sexo masculino, bem como na identificação genética de desconhecidos.

5.3.2. *Haplótipos encontrados*

Os haplótipos encontrados são semelhantes a outros estudos referentes a populações portuguesas, sendo alguns deles partilhados por outras populações europeias (Rosser *et al.*, 2000). Os mais frequentes na nossa população, correspondem ao DYS19*14; DYS389I*30; DYS389II*23; DYS390*14; DYS391*10; DYS392*13; DYS393*14; DYS385*11-14 e ao DYS19*14; DYS389I*30; DYS389II*23; DYS390*14; DYS391*10; DYS392*13; DYS393*15; DYS385*11-14. Estes compreendem os alelos mais frequentes para cada sistema, quando tomados individualmente, com excepção do alelo 15 para o DYS393 que representa o terceiro mais frequente.

Tendo em conta o modelo de mutação “single-step”, podemos inferir que nenhum destes dois haplótipos seria o ancestral, por não ser o mais frequente na Europa, incluindo a Península Ibérica. No entanto, o primeiro dos haplótipos referidos como mais frequentes na nossa população, apenas difere do mais frequente nas populações atrás citadas, no DYS393 que no Norte de Portugal é 14 e naquelas é 13. Acresce ainda que, a partir deste haplótipo mais frequente na nossa população (P053), seguindo aquele modelo de mutação amplamente aceite, podem derivar os haplótipos P096 (com uma mutação no sistema DYS19), o P070 (com uma mutação no sistema DYS389II), o P043 (com uma mutação no sistema DYS390) e o P055 (com uma mutação no sistema DYS393), sendo este último um dos mais frequentes na nossa população. Quanto aos restantes haplótipos na nossa população, o estabelecimento da relação entre eles torna-se mais complexo e sai do âmbito dos objectivos deste trabalho. Estas deduções para ser confirmadas teriam que ser testadas através de programas estatísticos adequados, como tem sido sugerido por vários autores (Holmlund *et al.*, 2003).

5.3.3. *Comparação entre várias populações*

A comparação entre populações torna-se mais robusta quando se utilizam as frequências dos haplótipos (P-Sánchez-Diz *et al.*, 2003). Depois de se proceder a esta comparação, utilizando dados haplotípicos de várias populações europeias, incluindo algumas da Península Ibérica e populações asiáticas, verificou-se que partilhamos muitos haplótipos

com as restantes populações da Península Ibérica. Esta decresce progressivamente relativamente a outras populações europeias, pois partilha-se 4 haplótipos com a Alemanha, 3 com a Hungria, 2 com a Itália e um com a Turquia, Albânia e com um grupo de romenos residentes na Hungria.

O haplótipo 14-13-29-24-11-13-13-11/14 existe na nossa população e em todas as restantes populações da Europa comparadas neste estudo, com excepção da Itália, sendo o mais frequente nas populações espanholas mencionadas e um dos mais frequentes na Europa. Por sua vez, os mais frequentes na nossa população, são mais raros nas restantes. Nenhum dos nossos haplótipos é partilhado com as populações da Índia, China, Japão e Coreia.

Como existe forte associação alélica nestes sistemas, deve-se fazer o cálculo das frequências haplotípicas, tendo em conta uma estimativa de erro, por exemplo por “bootstrapping”, um dos métodos que a incluem e que foi seguido neste estudo.

Para estimativa de frequências e F_{ST} adequados, são necessárias bases de dados destes sistemas, de modo a ser possível corrigir a eventual existência de subestruturação populacional, uma vez que os haplótipos estão confinados às linhagens paternas e, por isso, mais sujeitos à deriva genética.

A análise entre pares de populações efectuada entre a nossa população e as restantes, permitiu constatar a existência de diferenças significativas, com F_{ST} que apresentavam valores de probabilidade inferior a 0,05 ($p \leq 0,05$), como é o caso das populações asiáticas comparadas.

Quando comparámos várias populações da Europa e Ásia, por serem as que tinham os haplótipos completos descritos, verificou-se que somos mais parecidos com alguns europeus, como os da Península Ibérica e os do Centro da Europa, por exemplo, com a Alemanha, e menos parecidos com as populações mais próximas do mediterrâneo como a Grécia, Turquia, Albânia e Romenos. Como seria de esperar somos muito distintos dos povos asiáticos aqui comparados, Índia, China, Japão e Coreia.

A árvore de distâncias genéticas demonstra uma separação entre as várias populações, traduzindo graficamente os valores de F_{ST} obtidos.

Usando uma análise MDS, o gráfico separa claramente as populações em, pelo menos, três grupos principais: europeus, nos quais se inclui a nossa população e dentro destes

ainda se pode individualizar as populações do centro da Europa, asiáticos, as populações da Índia e o grupo dos romenos (Figura 7).

A comparação das distintas populações estudadas no que se refere aos Y-STRs, até ao momento, demonstra que alguns haplótipos são mais frequentes na Europa, como é o caso do haplótipo 14-13-29-24-11-13-13-11/14, já referido, e alguns que apenas se observam nos países asiáticos, por exemplo, aqueles que envolvem alelos de tamanhos mais elevados no DYS385. Portanto, estes haplótipos são de muita utilidade em Genética e Biologia Forense, uma vez que podem permitir calcular um Poder de Discriminação “populacional”, isto é, permitem deduzir se um indivíduo pertence a um determinado grupo populacional relativamente a qualquer outro tomado ao acaso.

5.4. Polimorfismos bialélicos

A capacidade de identificar cromossomas genealogicamente relacionados, é uma característica chave dos marcadores que evoluem lentamente, como é o caso dos marcadores bialélicos. O seu estudo fornece informação muito útil na investigação de eventos demográficos, porque permite comparar grupos de cromossomas que se sabe serem relacionados, evitando problemas devidos a mutações, mais frequentes nos STRs. O aumento do poder de discriminação é o efeito de maior impacto em Genética e Biologia Forense, mas a subestruturação geográfica que se tem constatado em estudos populacionais que usam marcadores bialélicos, é um assunto que requer uma maior reflexão quando se torna necessário o seu uso na conclusão das perícias.

5.4.1. Diversidade Génica e frequências dos Haplogrupos

Os marcadores bialélicos aqui estudados (SNPs e YAP), demonstraram ser insuficientes para discriminar indivíduos, uma vez que proporciona um baixo Poder de Discriminação, apenas $0,6891 \pm 0,0236$, valor esperado, pois tem sido observado noutros estudos. No entanto, a análise destes sistemas na nossa população, permitiu determinar as distâncias genéticas relativamente a outras populações previamente estudadas, para a maioria dos sistemas utilizados neste trabalho, com excepção do Tat, que não

apresentava polimorfismo em populações desta região e, por esse motivo, não foi incluído neste trabalho.

Dos nove haplogrupos possíveis, quando se analisaram os SNPs e a inserção YAP, apenas se obtiveram cinco.

Tal como na Europa e nas populações ameríndias (Rosser *e col.*, 2000), o haplogrupo 1 (J/T-S) ou P* (YCC), é o mais frequente na nossa população, como seria de esperar numa população caucasiana.

A mutação M9 é característica de uma linhagem que não possui a inserção YAP, permitindo distinguir os haplogrupos 26 e 12 do HG1 (J/T-S) e que corresponde ao K* (YCC), haplogrupos estes que correspondem a duas linhagens já antigas que se encontram em todas as populações, excepto em África. Este facto sugere que esta linhagem tenha surgido fora de África e a sua expansão consequência da deriva genética, de um efeito inicial de gargalo e posterior dispersão do Homem moderno. Alternativamente, poderá ter havido um ancestral africano que se perdeu por deriva e/ou selecção.

Outro cenário inclui o fluxo de homens arcaicos e uma origem multiregional (Hammer e Zegura, 1996). Esta teoria tem sido continuamente descartada, uma vez que todos os dados genéticos obtidos, até ao momento, apontam para uma substituição demográfica de formas humanas arcaicas, havendo uma descontinuidade genética entre elas e o homem moderno (Hurles *e col.*, 1999; Caramelli *e col.*, 2003). Este Haplogrupo K* está presente na nossa população, à semelhança de outras não africanas.

O haplogrupo 2 (J/T-S) ou BR* (YCC), é o segundo mais frequente na nossa população, sendo também frequente nas populações europeias (Rosser *e col.*, 2000) e africanas (Karafet *e col.*, 1999).

A origem geográfica da inserção YAP não é certa, embora ela ocorra com elevada frequência em África. Alguns autores sugerem que ela possa ter tido origem na Eurásia e, por isso, não seria de estranhar encontrá-la com uma frequência relativamente elevada na população do Norte de Portugal. A sua ausência ou a sua reduzida frequência pode ser o resultado de um efeito de gargalo, pelo que o número de linhagens com esta inserção terá diminuído.

A baixa frequência do haplogrupo 21 (J/T-S) ou E* (YCC) na nossa população, pode ter resultado do efeito atrás referido. Estes valores de frequência também têm sido obtidos noutras populações portuguesas (Rosser *et al.*, 2000).

A “network” obtida mostra que os haplogrupos 2 e 1 (J/T-S) ou BR* e P* (YCC), são os mais frequentes, indicando semelhança com as populações europeias. O haplogrupo 21 (J/T-S) ou E* (YCC), também com alguma representatividade, reflecte a influência de populações africanas. Situação idêntica foi encontrada por outros autores em populações portuguesas (Rosser *et al.*, 2000).

Os haplogrupos que apresentam uma distribuição geográfica mais restrita, são definidos por acontecimentos mutacionais mais recentes, como o haplogrupo 16 (J/T-S) ou N3 (YCC), não analisado na nossa população, que poderá ser de origem asiática (Karafet *et al.*, 1999). O haplogrupo 22 (J/T-S) ou R1b8 (YCC), que terá uma origem Ibérica e que está presente na nossa população, o que poderá confirmar esta teoria (Hurles *et al.*, 1999; Brión *et al.*, 2000).

Esta teoria tem sido posta em causa, na medida em que, sendo o HG22 (J/T-S) ou R1b8 (YCC) proveniente de linhagens do HG1 ou P* (YCC), por mutação no sistema SRY-2627, seria de esperar que este último haplogrupo, também se encontrasse em populações africanas, o que não acontece. Por isso, poderá suceder que o HG22 (J/T-S), tenha origem a partir do HG26 (J/T-S) ou K* (YCC) e não do HG1 (J/T-S) ou P* (YCC) (Carril JC, 2000). No entanto, para que o haplogrupo R1b8 se originasse a partir do K* deveriam ocorrer dois sucessos mutacionais no DYS257 e no SRY-2627, enquanto a partir de P* só é necessário que ocorra um, o que torna mais provável que R1b8 derive de P* do que de K*.

Haplogrupos mais representados em termos geográficos caracterizam-se por mutações bastante antigas, como é o caso do HG2 (J/T-S) ou BR* (YCC), já referidos.

Outros haplogrupos ausentes na população estudada, correspondem aos HG4, 3 e 8 (J/T-S) ou DE*, R1a e E3a (YCC), que são característicos de determinadas regiões específicas, o primeiro restringe-se a países asiáticos (Karafet *et al.*, 1999), o segundo apresenta frequências elevadas em países do centro e Este europeu e o último aparece apenas em populações do Norte de África, com frequências reduzidas (~5%) (Rosser *et al.*, 2000).

5.4.2. Diferenciação populacional e agrupamento das populações

Efectuou-se a análise dos estatísticos F_{ST} , a partir das características atrás referidas da nossa população, no que diz respeito a semelhanças e diferenças com outras.

O dendrograma construído com os haplogrupos de 60 populações, representativos de todos os grupos raciais, agrupa a nossa população com as populações ibéricas e ainda com populações europeias, confirmando uma origem comum das linhagens paternas de todas as populações do ocidente europeu, para a qual também contribuiu o intercâmbio de indivíduos entre elas. Outra possibilidade é a de que pertençam à mesma onda migratória original.

Dentro deste grupo também encontramos as populações ameríndias, tendo sido demonstrado, através dos estatísticos F_{ST} , não existir diferenças significativas entre estas populações, o que traduz uma relação genética entre elas (incluindo a nossa população). Esta semelhança pode ser devida ao facto de não se ter incluído alguns marcadores (SNPs) que permitiriam distinguir estas populações das europeias, como foi referido por Karafet *e cols.* (1999). Um exemplo dos referidos SNPs, é o DYS199-T que divide o haplogrupo 1 (J/T-S) em dois haplogrupos distintos: o haplogrupo 1C e o 1G. O primeiro corresponderia ao mesmo haplogrupo, ou seja o HG1; o segundo corresponderia a uma linhagem que diferenciaria as populações ameríndias das europeias. Ambos os haplogrupos estão presentes nas populações ameríndias com frequência elevada. Karafet *e cols.* (1999) sugerem que este haplogrupo 1G pudesse ter tido origem no novo mundo, a Este do Estreito de Bering. Este haplogrupo estaria presente em indivíduos (fundadores) da região do Noroeste da Sibéria, correspondendo ou apoiando o fluxo genético ocidental do Alasca para a Sibéria, através do Estreito de Bering (Karafet *e cols.*, 1997. Lel *e cols.* (1997) confirmam a presença do polimorfismo DYS199-T em populações Esquimós da Sibéria e de Chukchi, propondo a hipótese que se teria originado em Bering. A estimativa da idade média mutacional, levada a cabo por Kalafel *e cols.* (1997) favorece, no entanto, a hipótese de uma origem do haplogrupo 1G no Novo Mundo (há 7600 anos).

Para além do “cluster” já descrito, observam-se mais dois que agrupam populações caucasianas. Um deles reúne todas as populações do nordeste da Europa, que, na sua

maioria, falam línguas eslavas. O outro “cluster” agrupa populações do Sudeste da Europa, oriente próximo e Cáucaso (Rosser *e cols.*, 2000).

Estes dois “clusters” distinguem-se perfeitamente do anterior, apesar de serem todos eles populações caucasianas, o que pode ser devido ou ao facto das linhagens paternas terem tido origem em populações fundadoras completamente distintas para cada um dos “clusters”, ou a uma grande divergência após a fundação das referidas linhagens.

Existe ainda um outro “cluster” que agrupa as populações asiáticas e da Oceânia, que partilham as mesmas linhagens paternas, de acordo com os *loci* analisados. A população da Oceânia surgiu por expansão das populações asiáticas, pelo que este agrupamento corresponde esperado. Todavia, se se estudar outros polimorfismos haverá possibilidade de as distinguir, à semelhança do que foi descrito relativamente aos ameríndios.

O “cluster” africano apresenta uma subdivisão entre as populações do centro e Sul do continente e as populações do Norte. Estas últimas têm como haplogrupo maioritário o haplogrupo 21 (J/T-S) ou E* (YCC), sendo este haplogrupo, a razão principal da sua diferenciação com as restantes populações africanas. Esta diferença pode ser hipoteticamente explicada pela expansão da agricultura do Neolítico, a partir do Oriente próximo pelo Norte de África, e que agrupou as populações que teriam uma origem comum, a partir de uma população fundadora que se deslocou desde o Cáucaso ao Norte de África, durante a referida expansão. Portanto, esta população fundadora seria composta por indivíduos com características que os situam dentro do haplogrupo 21 (J/T-S) ou E* (YCC).

5.4.3. História evolutiva baseada nos polimorfismos do cromossoma Y

O resultado da análise dos dados obtidos, a partir da região não recombinante do cromossoma Y, permitem confirmar inexistência de subestruturação genética das nossas linhagens paternas, particularmente no que concerne aos Y-STRs, não existindo diferenças significativas com populações próximas, especialmente com populações espanholas e algumas populações europeias.

Pode-se confirmar a antiguidade de determinada população, pela ausência de determinados polimorfismos, correspondentes a mutações mais recentes, como é o caso das populações africanas.

A recompilação de dados realizada, conjuntamente com o estudo das linhagens paternas da nossa população, parecem apoiar a teoria de que o homem moderno terá sido originário de África. Só nestas populações, se observa o haplogrupo 7 (J/T-S) ou Y* (YCC), do qual surgem todos os restantes haplogrupos por acumulação de mutações. A hipótese da população ancestral ser de origem africana é apoiada por determinação de um ancestral comum, relativamente recente, com cerca de 150 000 anos (ou outros valores dependendo dos estudos realizados com mtDNA, marcadores autossómicos ou do Y). A determinação, sob o ponto de vista arqueológico, de algumas datas de ocupação dos continentes por parte dos colonizadores, pode-se comparar com as obtidos a partir das distâncias genéticas entre as populações analisadas. Quanto mais antiga é a data de ocupação, mais tempo houve para que as diferenças genéticas entre a população dos colonizadores e as populações originais se manifestem, por acumulação de sucessivas mutações (Cavalli-Sforza).

Quando o homem moderno saiu de África, provavelmente terá entrado pela Ásia, ocupando o próximo Oriente. A diversidade encontrada nas populações asiáticas, relativamente às linhagens paternas, revela a existência de mutações ausentes nas populações europeias.

As linhagens encontradas em populações da Oceânia e do continente americano, apresentam semelhanças com as populações asiáticas. Daí que se encontrem distâncias genéticas curtas entre estas populações, quando se analisam os dados relativos aos polimorfismos do Cromossoma Y. As populações ameríndias não se podem comparar neste estudo, porque o polimorfismo que as distingue da estudada, não ter sido analisado para a nossa população, como foi atrás referido; no entanto, o estudo de

Karafet *e cols.* (1999) parece indicar que a origem dos nativos americanos pode relacionar-se com populações asiáticas, uma vez que as suas linhagens são semelhantes a estas.

Todos estes dados parecem confirmar as teorias que indicam ter havido uma primeira vaga que saiu de África e colonizou o Sul da Ásia, chegando à Oceânia, há 60 000-70 000 anos tendo-se, posteriormente, deslocado para as ilhas da Polinésia há 60 000 anos. Por último, chegaram à Austrália há 35 000 anos. Uma parte desta 1ª vaga estendeu-se pelo centro da Ásia, chegando ao nordeste da Europa. Outra vaga de migração que colonizou as ilhas da Polinésia, percorreu a costa asiática que, por sua vez, se dividiu na Sibéria: uma parte passou o Estreito de Bering, formando a primeira população que chegou ao continente americano, há cerca de 10 000-40 000; outra parte, estendeu-se pelo Norte da Europa.

Uma segunda vaga de migração, partiu de África pelo oriente próximo, até chegar ao Sul da Europa, colonizando a Europa há 40 000 anos. Os dados que se obtêm a partir do cromossoma Y apoiam estas diferentes vagas de migração, uma vez que se tem verificado a existência de diferenças nítidas entre as populações do Norte e do Sul da Europa e próximo Oriente.

Os indivíduos da primeira vaga que saiu de África, eram caçadores recolectores, sendo o seu número limitado, mas devido ao facto de serem populações nómadas, a migração foi favorecida. Estas populações sofreram uma redução drástica no seu número, durante a última época glacial, o que apoiou o efeito de gargalo e a diferenciação de acordo com os locais de refúgio. Posteriormente, estas populações foram substituídas ou misturaram-se com indivíduos de migrações posteriores, durante a expansão do Neolítico. Esta última expansão está relacionada com o desenvolvimento da agricultura, a qual ocorreu no Oriente próximo, China, México e Norte dos Andes. A expansão a partir do Oriente próximo ficou favorecida, uma vez que a cultura de trigo e cevada, bem como a criação de animais domésticos, se podia praticar numa extensa área que rodeava o próximo Oriente, grande parte da Europa, do Norte de África não desértico e o Oeste da Ásia. A expansão foi muito lenta e irregular, tendo começado no Oriente próximo e Anatólia, há 9 500 anos, tendo tardado mais 3 500 anos para chegar à Inglaterra (Thorpe, 1996). A descoberta da agricultura fez com que as populações crescessem, o que levou à sua expansão. A apoiar esta hipótese está a presença do haplogrupo 1 (J/T-S) ou P* (YCC) na população europeia e ausência nas restantes

populações, o que sugere ter existido um refúgio europeu, tendo, posteriormente, aquele haplogrupo sido expandido às restantes populações da Europa. A sua presença em populações do Norte de África sugere, também, uma origem europeia recente nestas populações, em vez da possível origem no próximo oriente que seria previsível da expansão da agricultura do Neolítico.

As comparações de dados populacionais levadas a cabo neste trabalho, permitiram deduzir haver uma grande especificidade regional, no que diz respeito à distribuição dos haplogrupos (marcadores bialélicos), sendo a maioria das populações comparadas distintas entre si. Esta observação já tinha sido constatada por outros autores (Rosser *et al.*, 2000), em trabalhos que envolviam amostras de diversas regiões do planeta. Este fenómeno verifica-se mesmo em populações muito próximas, como por exemplo, entre populações do Norte e do Sul da Europa e entre o Este e o Ocidente. Por isso, a partir dos dados dos haplogrupos podemos inferir, com relativa segurança, o grupo étnico a que um indivíduo, envolvido num caso criminal, pertence. Por conseguinte, aquando da valorização da prova no âmbito da criminalística biológica, quando há concordância de haplogrupos (cena de crime e suspeito), tem de se ter em conta a frequência deste na base de dados da população de origem do suspeito. Como foi atrás referido, pode ocorrer que, apesar da frequência desse haplogrupo ser baixa ou, inclusivamente o mesmo estar ausente na população, (onde foi cometido o crime), não se possa descartar a possibilidade do criminoso pertencer a essa população.

A análise dos dois tipos de polimorfismos realizada neste trabalho, foi efectuada em tempos diferentes e com amostragens distintas, por esse motivo não foi possível obter resultados conjuntos. No entanto, o estudo de marcadores bialélicos, aliado à informação obtida a partir de STRs aumenta o poder discriminativo, uma vez que, no seu conjunto, a informação acerca das linhagens do cromossoma Y é mais completa (Roewer L, 2003). Por este motivo, uma vez conhecida a estrutura da população, mediante os resultados obtidos no presente estudo, seria de grande interesse aprofundar este aspecto, investigando as combinações dos haplótipos com os haplogrupos das linhagens paternas existentes na população do Norte de Portugal.

6. CONCLUSÕES

6. Conclusões

1ª- Foram analisados dois tipos de polimorfismo da região não recombinante (NRY) do cromossoma Y, STRs e marcadores bialélicos, tendo os primeiros demonstrado ser os mais úteis em Genética e Biologia Forense, devido aos seus elevados Índices de Diversidade e Poder de Discriminação, à sua relativamente baixa taxa de mutação e à simplicidade técnica de caracterização dos respectivos alelos. Estes dois grupos de marcadores, como são herdados de forma haplóide, são especialmente indicados na resolução de perícias do âmbito criminal, na investigação de paternidade e identificação genética individual, quando estão envolvidos indivíduos do sexo masculino.

2ª- Em relação ao material analisado e às metodologias empregues, destaca-se que as amostras usadas não foram as mesmas para os dois polimorfismos estudados, por terem sido efectuados em tempos diferentes, o que impediu a obtenção de informação mais completa das linhagens do cromossoma Y. Por outro lado, em relação aos Y-STRs, o reduzido número de pares pai-filho analisados (26), obstou ao aparecimento de qualquer incompatibilidade, pois só aumentando este número as detectaríamos e poderíamos determinar, com maior rigor, as respectivas taxas de mutação. Em relação aos marcadores bialélicos, o aumento do número de amostras, proporcionaria a identificação de um número mais elevado de haplogrupos, dentro dos previsíveis, para a nossa população.

O recurso à PCR de vários *loci* em simultâneo, relativamente aos Y-STR, seguida da análise dos fragmentos por electroforese capilar, em sequenciador automático, provou ser uma metodologia eficaz e relativamente simples de executar.

Quanto aos marcadores bialélicos, a técnica que foi utilizada, demonstrou ser simples e útil para analisar um pequeno número de marcadores numa amostragem relativamente reduzida, como foi o caso deste estudo.

3ª - As comparações populacionais realizadas, permitiram inferir que partilhamos vários haplótipos (Y-STRs) com as restantes populações da Península Ibérica e que, por sua vez, a referida partilha decresce progressivamente em relação a outras populações europeias, quando comparámos as frequências haplotípicas do Norte de Portugal com as populações da Europa, incluindo algumas da Península Ibérica. Quando a aludida

comparação é feita com países asiáticos, constatamos não haver haplótipos comuns entre a nossa população e as populações da Índia, China, Japão e Coreia, como era de prever.

4ª- O Poder de Discriminação obtido para os marcadores bialélicos estudados, demonstrou que estes são insuficientes para discriminar indivíduos; no entanto, permitiu-nos confirmar a especificidade geográfica da distribuição dos haplogrupos.

Existem haplogrupos bem representados na nossa população, como sejam, os HG2 e HG1 (J/T-S) ou BR* e P* (YCC), que são os mais frequentes, tendo-se constatado o mesmo na maioria das populações europeias, sendo o primeiro considerado o mais antigo da Europa.

Por outro lado, o haplogrupo 22 (J/T-S) ou R1b8 (YCC), que tem sido descrito em poucas populações da Europa, como França e Península Ibérica, está presente na nossa população, apresentando uma distribuição geográfica restrita. Este facto, parece apoiar a teoria de que este haplogrupo tenha tido uma origem Ibérica.

5ª- A análise dos estatísticos F_{ST} e respectiva construção de dendrogramas, árvores filogenéticas e representações MDS, relativamente aos haplogrupos aqui encontrados (marcadores bialélicos), permite-nos agrupar a nossa população com as restantes populações Ibéricas e algumas populações europeias, confirmando a existência de uma origem comum das linhagens das populações do Ocidente da Europa. Por outro lado, existe uma clara distinção destas com as populações asiáticas e africanas comparadas, podendo-se inferir a existência de uma grande especificidade regional. Este fenómeno verifica-se mesmo entre populações relativamente próximas como, por exemplo, entre o Norte e Sul da Europa e entre o Este e o Ocidente do mesmo continente.

6ª- Dada a importância destes marcadores na área da Medicina Legal, como perspectivas futuras, seria de grande interesse, quanto aos Y-STRs, investir na análise de um maior número de marcadores em simultâneo e, se possível, acrescentar novos sistemas, de modo que numa única análise se obtivesse a maior informação possível, obviando-se tempo despendido na análise e quantidade de amostra consumida.

No que concerne aos marcadores bialélicos, aumentar a amostragem e usar outras metodologias mais eficazes, porquanto a utilizada demonstrou ser fastidiosa, despendendo-se demasiado tempo o que impediu a análise de todos os marcadores

disponíveis em todas as amostras. Para obviar este e outros problemas, *como por exemplo, a interpretação de misturas (incluir?)*, ter-se-á de recorrer a métodos de elevado rendimento, como a PCR em tempo real.

Em face do exposto, tendo em conta a tarefa que nos propomos encetar, com o estudo dos marcadores bialélicos aliado ao dos Y-STRs, estamos cientes de que se conseguirá um assinalável aumento da capacidade informativa das linhagens do cromossoma Y e, conseqüentemente, uma melhor capacidade de alcançar, na maioria dos casos, uma identificação genética positiva em perícias solicitadas ao Serviço de Genética e Biologia Forense.

7. BIBLIOGRAFIA

7. Bibliografia

Akane A, Shiono H, Nakahori Y, Seki S, Nagafuchi S, Yamada M, Nakagome Y (1991). **Sex identification of forensic specimens by polymerase chain reaction (PCR): two alternative methods.** *Forensic Sci Int* **49**: 81-88.

Anslinger K, Keil W, Weichhold G, Eisenmenger W (2000). **Y-chromosomal STR haplotypes in a population sample from Bavaria.** *Int J Legal Med* **113**: 189-192.

Bandlet HJ, Forster P, Röhl A, (1999). **Median-joining networks for inferring intraspecific phylogenies.** *Mol Biol Evol* **16**(1): 37-48.

Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, Mayr W, Olaisen B (1997). **DNA recommendations. Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems.** *Int J Legal Med* **110**: 175-176.

Barros F, Carracedo A, Lareu MV, Rodriguez-Calvo MS (1991). **Electrophoretic human leukocyte antigen HLA-DQA1 DNA typing after polymerase chain reaction amplification.** *Electrophoresis* **12**(12):1041-1045.

Barros F, Lareu MV, Carracedo A (1992). **Detection of polymorphisms of human DNA after polymerase chain reaction by miniaturized SDS-PAGE.** *Forensic Sci Int* **55**(1):27-36.

Bertrantpetit J, Bosch E, Calafell F, Pérez-Lezaun A (1998). **Variación genética en el cromosoma Y: aplicaciones forenses.** In *Livro de resumos das Jornadas de Genética Forense*. Reunião do GEP-ISFH, Ekainak 1998, pág. 68-76.

Borrego N, Piqué E, Xifró A, Brandt-Casadevall C, Ortega M, Corbella J, Moreno P, Huguet E, Gené M (2003). **Haplotype frequencies of DYS19, DYS389 I/II, DYS390, DYS391, DYS393 and DYS385 STR loci in Barcelona (North-East Spain).** *Int Congress Series* **1239**: 399-402. (Barcelona)

Bortolini MC, Salzano FM, Bau CH, Layrisse Z, Petzl-Erler ML, Tsu LT, Hill K, Hurtado AM, Castro-de-Guerra D, Bedoya G, Ruiz-Lina A (2002). **Y-chromosome biallelic polymorphisms and Native American population structure.** *Ann Hum Genet* **66** (4): 255-259.

Bosch E, Calafell F, Santos FR, Pérez-Lezaun A, Comas D, Benchemsi N, Tyler-Smith C, Bertranpetit J (1999) **Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome.** *Am J Hum Genet* **65**: 1623-1638.

Brión M, Lareu MV, Pereira L, González-Neira A, Salas A, Prata MJ, Amorim A, Carracedo A (2000). **Y chromosome lineages in Iberians: construction of highly informative haplotypes using biallelic markers, Stirs and the minisatelite MSY1.** *Prog For Gen* **8**: 263-265.

Buel E, LaFountain M, Schwartz M, Walkinshaw M (2001). **Evaluation of capillary electrophoresis performance through resolution measurements.** *J Forensic Sci* **46**(2): 341-345.

Butler JM (2001). **DNA testing in high-profile cases (Biology, technology and genetics)** in *Forensic DNA typing, Biology & technology behind STR markers*, Academic Press, pag. 253-272.

Butler JM, Schoske R, Vallone PM, Kline MC, Redd AJ, Hammer MF (2002). **A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers.** *Forensic Sci Int* **129**: 10-24.

Cagliá A, Dobosz M, Boschi I, d'Aloja E, Pascali VL (1998). **Increased forensic efficiency of a STR-based Y-specific haplotype by addition of the highly polymorphic DYS385 locus.** *Int J Legal Med* **111**: 142-146. (Itália)

Calafell F, Bertranpetit J (1994). **Principal component analysis of gene frequencies and the origin of Basques.** *Am J Phys Anthropol* **93**(2): 201-215.

Capelli C, Wilson JF, Richards M, Stumpf MP, Gratrix F, Oppenheimer S, Underhill P, Pascali VL, Ko TM and Goldstei DB (2001). **A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania.** *Am J Hum Genet* **68**: 432-443.

Caramelli D, Lalueza-Fox C, Vernesi C, Lari M, Casoli A, Mallegni F, Chiarelli B, Dupanloup I, Bertranpetit J, Barbujani G, Bertorelle G (2003). **Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans.** *PNAS* **100** (11): 6593-6597.

Carracedo A e col. (2001). **Results of a collaborative study of the EDNAP group regarding the reproducibility and robustness of the Y-chromosome STRs DYS19, DYS389 I and II, DYS390 and DYS393 in a PCR pentaplex format.** *Forensic Sci Int* **119**: 28-41.

Carril JC (2000). **Estructura y Perfil Genético de Poblaciones de la Península Ibérica mediante la Aplicación de Marcadores (STRs y SNPs) del Cromosoma Y.** *Tesis doctoral*, Facultad de Biología, Universidad de Santiago de Compostela.

Carvajal-Carmona LG, Soto ID, Pineda N, Ortiz-Barrieta D, Duque C, Ospina-Duque J, McCarthy M e col. (2000). **Strong Amerind/white Sex bias and a possible Sephardic contribution among the founders of a population in Northwest Colombia.** *Am J Hum Genet* **67**: 1287-1295.

Casanova M, Leroy P, Boucekkine C, Weissenbach J, Bishop C, Fellous M, Purrello M, Fiori F, Siniscalco M (1985). **A Y-linked DNA polymorphism and its potencial for estimatin genetic and evolutionary distance.** *Science* **230**: 1403-1406.

Cavalli-Sforza LL, Menozzi P, Piazza A (1994). **The history and geography of human genes.** *Princetown University Press*, Princeton, New Jersey, pag. 263-266.

Cavalli-Sforza LL e Cavalli-Sforza F (1995). **The great human diasporas: the history of diversity and evolution.** Addison-Wesley Publishing Co., Reading, MA.

Cavalli-Sforza LL (1982). **African Pygmies**. Orlando, Florida. Academic Press.

Chikhi L, Nichols RA, Barbujani G, Beaumont MA (2002). **Y genetic data support the Neolithic demic diffusion model**. *Proc Natl Acad Sci USA* **99**(17): 11008-11013.

Cockerham CC (1969). **Variance of gene frequencies**. *Evolution* **23**: 72-83.

Cockerham CC (1973). **Analysis of gene frequencies**. *Genetics* **74**: 679-700.

Cooke H (1976). **Repeated sequences specific to human males**. *Nature* **262**: 182-186.

Cooper G, Amos W, Hoffman D, Rubinsztein DC (1996). **Network analysis of human Y microsatellite haplotypes**. *Hum Mol Genet* **5**: 1759-1766.

Efron B (1982). **The Jackknife, the Bootstrap, and other resampling plans**. Society for Industrial and Applied Mathematics, Philadelphia.

Excoffier L, Smouse PE, Quattro JM (1992). **Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data**. *Genetics* **131**: 479-491.

Excoffier L (2000). **Analysis of population subdivision**, in *Handbook of Statistical Genetics*, D Balding, M Bishop and C Cannings (eds.) Wiley & Sons, Ltd..

Foot S, Vollrath D, Hilton A, Page DC (1992). **The human Y chromosome: overlapping DNA clones spanning the euchromatic region**. *Science* **258**: 60-66.

Foster EA, Jobling MA, Taylor PG, Donnelly P, de Knijff P, Mieremet R Zerjal T, Tyler-Smith C (1998). **Jefferson fathered slave's last child**. *Nature* **396**: 27-28.

Fukshansky N, Bar W (1999). **Biostatistical evaluation of mixed stains with contributors of different ethnic origin**. *Int J Legal Med* **112**: 383-387.

Füredi S, Woller J, Pádár Z, Angyal M (1999). **Y-STR haplotyping in two Hungarian populations.** *Int J Legal Med* **113**: 38-42.(Budapest e Romenos).

Gaensslen RE, Berka KM, Grosso DA, Ruano G, Pagliaro EM, Messina D, Lee HC (1992). **A Polymerase Chain Reaction (PCR) Method for sex and species determination with novel controls for Deoxyribonucleic Acid (DNA) template length.** *Journal of Forensic Sciences, JFSCA*, Vol.37(1): 6-20.

Gill P, Brenner C, Brinkmann B, Carracedo A, Jobling MA, de Knijff P, Kayser M, Krawczak M, Mayr WR, *e col.* 2001. **DNA comission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs.** *Int J Legal Med* **114**: 305-309.

Gill P, Brenner C, Brinkmann B, Budowle B, Carracedo A, Jobling MA, de Knijff P, Kayser M, Krawczak M, Mayr WR, *e col.* (2001).**DNA comission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs.** *Forensic Sci Int* **124**: 5-10.

Giusti A, Baird M, Pasquale S, Balazs I, Glassberg J (1986). **Application of deoxyribonucleic acid (DNA) polymorphisms to the analysis of DNA recovered from sperm.** *J Forensic Sci*, **31**: 409-417.

Graves JAM (1995). **The evolution of mammalian Sex chromosomes and the origin of sex determining genes.** *Philos Trans R Soc Lond B* **350**: 305-312.

Grignani P, Peloso G, Fattorini P, Previderè C (2000). **Highly informative Y-chromosomal haplotypes by the addition of three new STRs DYS437, DYS438 and DYS439.** *Int J Legal Med* **114**: 125-129.

Griffiths RC, Li WH (1983). **Simulating allele frequencies in a population and the genetic differentiation of populations under mutation pressure** *Theor Popul Biol* **23**(1):19-33.

Hammer MF (1994). **A recent insertion of an *Alu* element on the Y chromosome is a useful marker for human population studies.** *Mol Biol Evol* **11**: 749-761.

Hammer MF, Horai S (1995). **Y chromosomal DNA variation and the peopling of Japan.** *Am J Hum Genet* **56**: 951-962.

Hammer MF e Zegura SL (1996). **The role of the Y chromosome in human evolutionary studies.** *Evol Anthropol* **5**: 116-134.

Hammer MF, Spurdle AB, Karafet T, Bonner MR, Wood ET, Novelletto A, Malaspina P, Mitchell RJ, Horai S, Jenkins T, *e col.* (1997). **The geographic distribution of human Y chromosome variation.** *Genetics* **145**: 787-805.

Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR e Zegura SL (1998). **Out of Africa and back again: Nested cladistic analysis of human Y chromosome variation.** *Mol Biol Evol* **15**: 427-441.

Hammer, MF, Redd AJ, Wood ET, Bonner MR, Jarjanazi H, Karafet T, Sanatchiara-Benerecetti S, Oppenheim A, Jobling MA, Jenkins T, Ostrer H, Bonn  Tamir B (2000). **Jewish and middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic Haplotypes.** *Proc Natl Acad Sci USA* **97**: 6769-6774.

Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H e Zegura SL (2001). **Hierarchical patterns of global human y-chromosome diversity.** *Mol Biol Evol* **18**: 1189-1203.

Henke J, Henke L, Chatthopadhyay P, Kayser M, D lmer M, Cleef S, P che H, Felske-Zech H (2001) **Application of Y-chromosomal STR Haplotypes to Forensic Genetics.** *Croat Med J* **42**: 292-297.

Hidding M Schmitt (2000). **Haplotype frequencies and population data of nine Y-chromosomal STR polymorphisms in a German and a Chinese population.** *Forensic Sci Int* **113**: 47-53.

Holmlund G, Nilsson H, Langö-Warensjö A, Rosén B, Lindblom B (2003). **Y-chromosome STR haplotypes in a Swedish population.** *Int Congress Series* **1239**: 343-348.

Hurles ME, Veitia R, Arroyo E, Armenteros M, Bertranpetit J, Pérez-Lezaun A, Bosch E, Shlumukova M, Cambon-Thomsen A, McElreavey K, López de Munain A, Röhl A, Wilson IJ, Singh L, Pandya A, Santos FR, Tyler-Smith C, Jobling MA (1999). **Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism.** *Am J Hum Genet* **65**: 1437-1448.

Jobling MA e Tyler-Smith C (1995). **Fathers and sons: the Y chromosome and human evolution.** *Trends Genet* **11**: 449-456.

Jobling MA, Pandya A, Tyler-Smith C (1997). **The Y chromosome in forensic analysis and paternity testing.** *Int J Legal Med* **110**: 118-124.

Jobling MA e Tyler Smith C (2000). **New uses for new haplotypes: the human Y chromosome, disease and selection.** *Trends Genet* **16**: 356-362.

Jobling MA (2001). **In the name of the father: Surnames and genetics.** *Trends Genet* **17**: 353-357.

Kayser M, Cagliá A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Hrezog B, Hidding M, Honda K, Jobling M, Krawczak M e col. (1997). **Evaluation of Y-chromosomal STRs: a multicenter study.** *Int J Legal Med* **110**: 125-133.

Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF (1999). **Ancestral Asian source(s) of New World Y-chromosome founder haplotypes.** *Am J Hum Genet* **64**: 817-831.

Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL and Hammer MF (2001). **Paternal population history of East Asia: Sources, patterns, and microevolutionary processes.** *Am J Hum Genet* 69 (3): 615-628.

Kloosterman AD, Pouwels M, Daselaar P, Janssen HJT (1998). **Population Genetic study of Y-chromosome specific STR-loci in Dutch Caucasians.** In *Progress in forensic genetics*, Edit by Bjørnar Olaisen , Bend Brinkmann, Patrick J Lincoln, Vol 7, pg. 491-493.

de Knijff P, Kayser M, Cagliá A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Hrezog B, Hidding M, Honda K, Jobling M, Krawczak M *et al.* (1997). **Chromosome Y microsatellites: population genetic and evolutionary aspects.** *Int J Legal Med* 110: 134-140.

Lareu M, Puente J, Sobrino B, Quintáns B, Brión M, Carracedo A (2001). **The use of the LightCycler for the detection of Y chromosome SNPs.** *Forensic Sci Int* 118: 163-168.

Lalouel JM (1980). **Distance analysis and multidimensional Scaling.** In *Current Developments in Anthropological Genetics, Vol. 1: Theory and Methods.* Mielke J, Crawford M, eds., New York and London: Plenum Press, pg. 209-250.

Lau Y-F, Schonberg S (1984). **A male-specific DNA probe detects heterochromatin sequences in a familial Yq-chromosome.** *Am J Hum Genet* 36: 1394-1396.

Li WH (1977). **Distribution of nucleotide differences between two randomly chosen cistrons in a finite population.** *Genetics* 85: 331-337.

Long JC (1986). **The allelic correlation structure of Gainj and Kalam speaking people.I. The estimation and interpretation of Wright's F-statistics.** *Genetics* 112: 629-647.

Lucotte G *et al.* (1985). **p49f, a highly polymorphic probe, that detects TaqI RFLPs on the human Y chromosome.** *Nucleic Acids Res* 13: 8285.

Ludes B e Mangin P (1991). **Application of PCR data on forensic science.** *European Forensic Science*, PCR Symposium, London, 13-15 Maio 1991.

Mannuci A, Sullivan KM, Ivanov PL, Gill P (1994). **Forensic application of a rapid and quantitative DNA sex test by amplification of the homologous gene amelogenin.** *Int J Legal Med* **106**: 190-193.

Martínez-Jarreta B, Nievas P, Abecia E, Hinojal R (2003). **Population genetics of Y-chromosomal haplotypes in Asturias (Northern Spain).** *International Congress Series* **1239**: 327-330.

Mathias N, Bayes M, Tyler-Smith C (1994). **Highly informative compound haplotypes for the human Y chromosome.** *Hum Mol Genet* **3**: 115-123.

M de Pancorbo M, Alvarez M, Rodríguez JL, Herrera RJ (2003). **Estudio de 8 polimorfismos bialelicos del cromosoma Y no relacionan las poblaciones vascas com las georgianas o berber** (submetido para publicação).

Moretti TR, Baumstrark AL, Defenbaugh DA, Keys KM, Brown AL, Budowle B (2001). **Validation of STR typing by Capillary Electrophoresis.** *J Forensic Sci* **46(3)**: 661-676.

Mullis KA, Faloona F, Scharf S, Saiki R, Horn G, Erlich H (1986). **Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction.** *Cold Spring Harbor SympQuant Biol* **51**, pg.263-273.

Murray JC (1993). *Genome Database*.

Nei M (1987). **Molecular evolutionary Genetics.** Columbia University Press, New York, USA.

Nei M e Saitou N (1987). **The Neighbor-Joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* **4**: 406-425.

Nonaka I, Minaguchi K (2001) **Allele frequencies and haplotypes of ten Y-specific STRs in the Japanese population.** *J Forensic Sci* 179-182(Japão)

Nørby S e Eriksen B (1992). **Sex identification of forensic samples using PCR analysis for the presence of Y-chromosome specific DNA sequences.** *Advances in Forensic Haemogenetics* 4, Springer-Verlag Berlin Heidelberg, pág.38-40.

Paracchini S, Arredi B, Chalk R e Tyler-Smith C (2002). **Hierarchical high-throughput SNP genotyping of the human Y chromosome using MALDI-TOF mass spectrometry.** *Nucleic Acids Research* vol. 30, No.6 e27.

Parreira KS, Lareu MV, Sánchez-Diz P, Skitsa I, Carracedo A (2002). **DNA typing of short tandem repeat loci on Y-chromosome of Greek population.** *Forensic Sci Int* 126: 261-264. (Grécia 69)Pascal O, Aubert D, Gilbert E, Moisan JP (1993). **Sexing of forensic samples using PCR.** *Int J Leg Med* 104: 205-207.

Pereira L, Brión M, Prata MJ, Jobling MA, Carracedo A, Amrím A (2000) **Gradient of Y chromosome haplogroup 21 across the western Iberia.** *Prog For Genet* 8: 281-283.

Pfützinger H, Ludes B, Mangin P (1993). **Sex determination of forensic samples: co-amplification and simultaneous detection of a Y-specific and na X-specific DNA sequence.** *Int J Leg Med* 105: 213-216.

Prieto V, Torres Y, Ferfán MJ, López-Soto M, García-Hirschfeld J, Sanz P (2003). **Population genetics of Y chromosomal STR haplotypes in south Spain (Andalusia).** *International Congress Series* 1239: 403-408. (Andaluzia)

Prinz M, Boll K, Baum H, Shaler B (1997). **Multiplexing of Y chromosome specific STRs and performance for mixed samples.** *Forensic Sci Int* 85: 209-218.

Quintana-Murci L, Krausz C, McElreavey K (2001). **The human Y chromosome: function, evolution and disease.** *Forensic Sci Int* 119: 28-41.

Raitia M, Lindroos K, Laukkanen M, Pastinen T, Sistonen P, Sajantila A, Syvänen A-C (2001). **Y-chromosomal SNPs in Finno-Ugric-Speaking populations analysed by Minisequencing on microarrays.** *Genome Res* 471(11):471-482.

Ranade K, Chang MS, Ting CT, Pei D, Hsiao CF, Olivier M, Pesich R, Hebert J, Chen YD, Dzau VJ, Curb D, Olshen R, Risch N, Cox DR, Botstein D.(2001). **High-throughput genotyping with single nucleotide polymorphisms.** *Genome Res* 11(7): 1262-1268.

Reynolds J, Weir BS, Cockermann CC (1983). **Estimation for coancestry coefficients: basis for a short-term genetic distance.** *Genetics* 105: 767-779.

Robino C, Siragusa R, Vasino V, Gino S (2001). **Identificazione genetica e processo penale.** *Minerva Med Leg* 121: 53-67.

Robino C, Gino S, Torre C (2003). **Y-chromosome haplotypes in na Albanian population sample.** *International Congress Series* 1239: 457-459.

Roewer L, Arnemann J, Spurr NK, Grzeschik K-H, Epplen JT (1992). **Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts.** *Hum Genet* 89: 389-394.

Roewer L (1998). **The Y chromosome: forensic application and evolutionary aspects.** *Progress in forensic genetics* 7: 407-412.

Roewer L, Kayser M, de Knijff P e col. (2000). **A new method for evaluation of matches in non-recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males.** *Forensic Sci Int* 114: 31-43.

Roewer L, Krawczak M, Willuweit S, Nagy M, Alves C, Amorim A, Anslinger K, Augustin C, Betz A, Bosch E, Cagliá A, Carracedo A e col. (2001) **Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes.** *Forensic Sci Int* 118: 106-113.

Roewer L (2003).**The use of the Y chromosome in forensic genetics-current practices and future perspectives.** *Int Congress Series* 1239: 279-280

Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, *e col.* (2000). **Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language.** *Am J Hum Genet* **67**: 1526-1543.

Rousset F (2000). **Inferences from spatial population genetics**, in *Handbook of Statistical Genetics*, D Balding, M Bishop and C Cannings. (eds.) Wiley & Sons, Ltd..

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985). **Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia.** *Science* **230**(4732):1350-1354.

Saitou N, Nei M (1987). **The neighbour Joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* **4**: 406-425.

Sánchez-Diz P, De La Fé T, Quintáns B, Salas A, Lareu M, Carracedo A (2003). **Y-chromosome STRs in populations of Bantu origin from Mozambique: male contribution to the Africa genetic pool and forensic implications.** *Int Congress Series* **1239**: 419-424.

Santos FR, Pandya A, Tyler-Smith C, Pena SDJ, Schanfield M, Leonard WR, Osipova L, Crawford MH, Mitchell RJ (1999) **The central Siberian origin for native american Y chromosomes.** *Am J Hum Genet* **64**: 619-628.

Schneider S, Roessli D, Excoffier L (2000). **Arlequin v 2.0: a software for population genetics data analysis.** *Genetics and Biometry Laboratory*, University of Geneva, Geneva, Switzerland.

Schneider P, Meuser S, Waiyawuth W, Seo Y, Rittner C (1998). **Tandem repeat structure of the duplicated Y-chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations.** *Forensic Sci Int* **97** (1998) 61-70.

Seielstad MT, Minch E, Cavalli-Sforza LL (1998). **Genetic evidence for a higher female migration rate in humans.** *Nat Genet* **20** (3): 278-280.

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S *e col.* (2000). **The genetic legacy of paleolithic homo sapiens in extant Europeans: A Y chromosome perspective.** *Science* **290**: 1155-1159.

Sensabaugh GF (1982). **The utilization of polymorphic enzymes in forensic science.** *Isozymes Curr Top Biol Med Res* **11**:137-54.

Seielstad MT, Herbert JM, Lin AA, Underhill PA, Ibrahim M, Vollrath D, Cavalli-Sforza LL (1994). **Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition.** *Hum Mol Genet* **3**: 2159-2161.

Shin DJ, Jin HJ, Kwak KD, Choi JW, Han MS, Kang PW, Choi SK, Kim W (2001). **Y-Chromosome multiplexes and their potential for the DNA profiling of Koreans.** *Int J Legal Med* **115**: 109-117.

Singer-Sam J, Tanguay RL, Riggs AD (1989). **Use of Chelex to improve the PCR signal from a small number of cells.** *Amplifications* **3**: 11.

Skoreki *e col.*(1997). **Y chromosome of the Jewish Priests.** *Nature*, **385**, 32.

Slatkin M (1991). **Inbreeding coefficients and coalescence times.** *Genet Res Camb* **58**: 167-175.

Slatkin (1995). **A mesure of population subdivision based on microsatellite allele frequencies.** *Genetics* **139**: 457-462.

Spurdle AB, Hammer MF, Jenkins T (1994) **The Y Alu polymorphism in southern African populations and its relationship to other Y specific polymorphisms.** *Am J Hum Genet* **54**: 319-330.

Stalvey JRD e Erickson RP (1987). **Na improved method for detecting Y chromosomal DNA.** *Hum Genet* **76**: 240-243.

Su H, Lau YF (1993). **Identification of the transcriptional unit, structural organization, and promoter sequence of the human Sex-determining region Y (SRY) gene, using a reverse genetic approach.** *Am J Hum Genet*, **52**: 1503-1507.

Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J *et al.* (1999). **Y-chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age.** *Am J Hum Genet* **65**: 1718-1724.

Tanaka K, Fukushima H, Ota M, Ichinose A, Sato K, Honda K, Kiyono (1992). **Sex determination in bloodstains and single hairs.** *Advances in Forensic Haemogenetics* **4**, Springer-Verlag Berlin Heidelberg, pág.38-40.

The Y Chromosome Consortium, 2002. **A nomenclature system for the tree of human Y-chromosomal binary haplogroups.** *Genome Research* **12**: 339-348.

Thomas MG, Skorecki K, Ben-Ami H, Parfitt T, Bradman N, Goldstein DB (1998) **Origin of old testaments priests.** *Nature* **394**: 138-140.

Thorpe KL, Abdulla S, Kaufman J, Trowsdale J, Beck S (1996). **Phylogeny and structure of the RING3 gene** *Immunogenetics*. **44(5)**:391-396.

Torgerson WS 1952. **Multidimensional scaling: I. Yheory and Method.** *Psychometrika* **17**: 401-419.

Torroni A, Bandlet H, Macauley V, Richards M, Cruciani F, Rengo C, Martinez-Cabrera V, Vilems R, Kivisild T, Metspalu E, Parik J, Tolk H, Tambets K, Forster P, Karger B, Francalacci P, Rudan P, Janicijevic B, Rickards O, Savontaus M, Huoponen K, Laitinen V, Koivumäki S, Sykes B, Hickey E, Novelletto A, Moral P, Sellitto D, Coppa A, Al-Zaheri N, Santachiara-Banarecetti A, Semino O, Scozzari R (2001). **A Signal, from Human mtDNA, of Postglacial recolonization in Europe.** *Am J Hum Genet* **69**: 844-852.

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, e Oefner PJ (1997). **Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography.** *Genome Res* **7**: 996-1005.

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P *e col.*, (2000). **Y chromosome sequence variation and the history of human populations.** *Nat Genet* **26**: 358-361.

Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, e Cavalli-Sforza LL (2001). **The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations.** *Ann Hum Genet* **65**: 43-62.

Weir BS (1996). **Genetic Data analysis II: Methods for discrete Population Genetic Data.** Sinauer Assoc., Inc, Sunderland, MA, USA.

Weir BS e Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358-1370.

Whitfield LS, Sulston JE, Goodfellow PN (1995). **Sequence variation of the human Y chromosome.** *Nature* **378**: 379-380.

Willard HF (2003). **Tales of the Y chromosome.** *Nature* **423(6942)**: 810-1, 813.

Willard HF e Wayne JS (1987). **Hierarchical order in chromosome-specific human alpha satellite DNA.** *Trends Genet* **3**: 192-198.

Witt M e Erickson RP (1989). **A rapid method for detection of Y-chromosomal DNA from dried blood specimens by the polymerase chain reaction.** *Hum Genet* **82**: 271-274.

Veitia R, Ion A, Barboux S, Jobling MA, Souleyreau N, Ennis K, Ostrer H, Tosi M, Meo T, Chibani J, Fellous M, McElreavey K (1997). **Mutations and sequence variants in the testis-determining region of the Y chromosome in individuals with a 46,XY female phenotype.** *Hum Genet* **99**: 648-652.

Vollrath D, Foote S, Hilton A, Brown LG, Beer-Romero P, Bogan JS, Page DC (1992). **The human Y-chromosome: a 43-interval map based on naturally occurring deletions.** *Science* **258**: 52-59.