

Estratégias para o desenvolvimento de um InDel PCR multiplex para definição de haplogrupos do cromossoma Y

Joana Daniela Mendes Damas

Dissertação de candidatura ao grau de Mestre em Genética Forense
apresentada à Faculdade de Ciências da Universidade do Porto

Orientador: Doutora Leonor Gusmão
Categoria: Investigadora
Afiliação: IPATIMUP

AGRADECIMENTOS

À Faculdade de Ciências da Universidade do Porto pela possibilidade de realizar este mestrado e pela atribuição da bolsa de estudo.

Ao Prof. Doutor António Amorim e à Doutora Leonor Gusmão pela orientação e ajuda sempre disponível.

Ao Doutor Rui Pereira pela ajuda no momento mais difícil, o início, e pela constante disponibilidade.

Aos meus colegas de mestrado, em especial à Estefânia, à Joana e ao Miguel por todo o apoio e partilha de experiência.

À Mónica, à Cláudia, à Ana e à Sara pelo bom ambiente de trabalho que criam, ajuda sempre disponível e momentos de descontração.

Ao grupo de genética populacional do IPATIMUP pelo acolhimento e ajuda.

A toda a minha família e amigos, especialmente aos meus pais e irmã, pelo apoio, preocupação e interesse constantes.

A todos os que não referi anteriormente e contribuíram para o desenvolvimento deste trabalho, em qualquer uma das suas vertentes, o meu muito obrigado!

ÍNDICE

AGRADECIMENTOS.....	II
ÍNDICE.....	III
ÍNDICE DE TABELAS	V
ÍNDICE DE FIGURAS	VI
RESUMO	VIII
ABSTRACT	IX
INTRODUÇÃO	1
A - Cromossoma Y	1
A1-Evolução.....	1
A2-Estrutura	2
A3-Transmissão	4
A4-Polimorfismos	4
B – Principais haplogrupos e sua classificação	6
C – InDels.....	9
OBJECTIVO.....	10
MATERIAIS E MÉTODOS	11
A – Pesquisa de InDels.....	11
B – Desenho de Primers	12
C - Amplificação do DNA	12
D – Electroforese	13
E – Purificação e Sequenciação	13
F – Localização dos novos marcadores na Árvore Filogenética	14
G – Variação do marcador M2 dentro do sub-haplogrupo R1b1b2	15
H – Determinação da localização do marcador M2 na região AZF do cromossoma Y.....	15
RESULTADOS E DISCUSSÃO	16
A – Pesquisa de InDels.....	16
B - Eliminação de marcadores repetidos (I)	16
C - Eliminação de “falsos” polimorfismos	20
D - Eliminação de marcadores repetidos (II).....	21
E - Análise das sequências flanqueantes.....	21
F – Desenho de Primers	21
G - Amplificação do DNA.....	29

H – Electroforese	29
I – Determinação da especificidade masculina	30
J – Verificação da presença de polimorfismo.....	30
K – Confirmação dos polimorfismos por sequenciação	32
L – Localização dos novos marcadores na Árvore Filogenética	33
L1 – Localização do marcador M2 (rs1611049)	33
L2 – Localização do marcador M14 (rs35470479)	38
M – Variação do marcador M2 dentro do sub-haplogrupo R1b1b2.....	39
N – Determinação da localização do marcador M2 na região AZF do cromossoma Y	39
CONCLUSÃO	41
REFERÊNCIAS BIBLIOGRÁFICAS	42

ÍNDICE DE TABELAS

Tabela 1: Características gerais das diferentes classes de eucromatina do cromossoma Y.....	4
Tabela 2: Exemplo de marcadores bialélicos do cromossoma Y, diagnósticos de alguns dos haplogrupos (HG) principais, e respectiva distribuição populacional, indicando as regiões e/ou populações onde a sua frequência é mais elevada.....	6
Tabela 3: Critérios usados na pesquisa de novos InDels nas bases de dados dbSNP e MGS.	11
Tabela 4: Definições usadas para o desenho de primers.	12
Tabela 5: Condições utilizadas nas reacções de PCR.	13
Tabela 6: Condições utilizadas nas reacções de sequenciação.	13
Tabela 7: Amostras utilizadas com a finalidade de localização da mutação na árvore filogenética do Cromossoma Y.....	14
Tabela 8: Resultados do estudo de marcadores repetidos no conjunto dos InDels obtidos pela pesquisa nas bases de dados dbSNP e MGS.	17
Tabela 9: InDels retirados de Karafet <i>et al.</i> (2008) e respectivas características (* UCSC, <i>Homo sapiens</i> Genome, Build 36 version 1).	18
Tabela 10: Selecções a que foram submetidos os InDels obtidos, número de InDels que não obedecem aos critérios de inclusão e número de InDels que passa à próxima fase de análise.	20
Tabela 11: Principais características dos InDels seleccionados para o desenho de primers.....	22
Tabela 12: Sequência parcial da zona onde se inserem os marcadores M1 a M16 com marcação de primers <i>forward</i> (>>), <i>reverse</i> (<<), SNPs presentes no primer (■), polimorfismo alvo (**), zonas a excluir (X), InDels no fragmento amplificado (□) e zona repetitiva no fragmento amplificado (□).	23

ÍNDICE DE FIGURAS

Figura 1: Sistemas de determinação do sexo em vertebrados.	1
Figura 2: Diagrama esquemático do cromossoma Y.....	2
Figura 3: Genealogia do cromossoma Y (esquerda) e polimorfismos estruturais identificados (direita).	5
Figura 4: Árvore filogenética do cromossoma Y segundo Karafet <i>et al.</i> (2008)[1].	7
Figura 5: Mapa da distribuição mundial dos haplogrupos do cromossoma Y.	8
Figura 6: Árvore Filogenética do cromossoma Y com indicação dos haplogrupos usados para detecção de polimorfismo.	12
Figura 7: Localização dos InDels incluídos em Karafet <i>et al.</i> (2008) (□) na árvore filogenética do cromossoma Y.	19
Figura 8: Comparação do poder de resolução do gel de acrilamida T12C5 (à direita) com o método de electroforese capilar usado (à esquerda).	29
Figura 9: Verificação da amplificação em amostras do sexo feminino dos marcadores estudados.	30
Figura 10: Gel electroforético, T12C5, do marcador 1, com indicação do haplogrupo a que pertence cada amostra.	31
Figura 11: Gel electroforético, T9C5, do marcador 2, com indicação do haplogrupo a que pertence cada amostra.	31
Figura 12: Gel electroforético, T12C5, do marcador 14, com indicação do haplogrupo correspondente a cada amostra.	32
Figura 13: Extracto do alinhamento dos haplogrupos Q (superior) e R (inferior) para o marcador 2.	32

Figura 14: Extracto do alinhamento dos haplogrupos O, Q, R e T para o marcador 14.	33
Figura 15: Gel electroforético (T12C5) de amostras dos haplogrupos R e Q, para o marcador 2.	34
Figura 16: Representação esquemática das hipóteses (A e B) propostas para o momento da duplicação do marcador M2.	35
Figura 17: Esquema representativo das amostras dos haplogrupos Q e R estudadas com indicação do tipo de resultado electroforético obtido.	36
Figura 18: Esquema representativo das amostras do sub-haplogrupo R1b1b2 estudadas com indicação do tipo de resultado electroforético obtido.	37
Figura 19: Gel electroforético de amostras dos haplogrupos Q e R, para o marcador 14.	38
Figura 20: Localização do marcador 14 na árvore filogenética do cromossoma Y (★).	38
Figura 21: Gel electroforético T12C5 de amostras portadoras da mutação S167.	39
Figura 22: Representação esquemática da localização do marcador M2 (rs1611049) no cromossoma Y.	40

RESUMO

Actualmente, a caracterização dos diferentes haplogrupos do cromossoma Y humano é feita recorrendo ao estudo de SNPs e de um pequeno número de InDels. Devido a apresentarem certas características específicas os InDels têm vindo a ganhar protagonismo para uso forense, uma vez que permitem uma tipagem mais simples e rápida. Assim, de forma a simplificar e tornar mais robusta a caracterização de amostras masculinas nos principais haplogrupos do cromossoma Y, o objectivo deste trabalho foi criar um multiplex de InDels que nos permitisse fazer essa caracterização recorrendo apenas a uma reacção de PCR e a métodos de separação electroforética automatizada. Para a concretização do objectivo proposto, numa primeira análise seleccionaram-se os InDels presentes na compilação de polimorfismos de Karafet et al. (2008), uma vez que estes já são aceites pelo Consórcio do Cromossoma Y (Y Chromosome Consortium) para inferência de haplogrupos, pelo que estão bem caracterizados e estudados populacionalmente. No entanto, quando se observou a sua localização na árvore filogenética do cromossoma Y verificou-se que, recorrendo apenas a estes polimorfismos, não é possível distinguir os principais haplogrupos descritos em populações humanas, estando apenas descritos InDels que permitem a classificação de cromossomas pertencentes aos haplogrupos A, B, O, J e T, e ao parágrafo DE. Sendo assim, concluiu-se que para a elaboração de um multiplex de InDels que permitisse a tipagem exaustiva dos 22 ramos principais da árvore filogenética do cromossoma Y, seria necessária a caracterização populacional de novos polimorfismos deste tipo.

Para a obtenção de novos InDels recorreu-se a duas bases de dados: a dbSNP (Single Nucleotide Polymorphism database) e a MGS (Mammalian Genotyping Service), de onde se retiraram todos os InDels descritos e já validados para o cromossoma Y. Os 154 InDels obtidos foram submetidos a várias triagens de forma a assegurar a eliminação dos polimorfismos repetidos, tanto dentro como entre bases de dados, e de falsos polimorfismos. Como exemplos temos a detecção de um falso InDel devido à presença de um STR, ou a detecção de polimorfismos específicos do cromossoma Y devido à presença de homologia no cromossoma X. Após todas as triagens efectuadas, verificou-se existirem nestas bases de dados apenas 10 novos InDels potencialmente polimórficos e específicos do cromossoma Y, os quais foram seleccionados para um estudo mais aprofundado. Para estes 10 marcadores foram desenhados primers para amplificação por PCR e tipagem de amostras de 13 dos 17 haplogrupos do cromossoma Y representados na árvore filogenética. Destes 10 marcadores apenas se confirmaram verdadeiros polimorfismos 2 deles, o marcador M2 e o M14. Através de um estudo mais detalhado destes dois marcadores em amostras pertencentes a diferentes sub-haplogrupos, concluiu-se que o marcador M2 apresenta o alelo mutado em amostras do haplogrupo R e, dentro deste, apresenta uma reversão no sub-haplogrupo R1b1b2. Por isso, o marcador M2 não é indicado para a caracterização de linhagens dentro do haplogrupo R e a sua distinção de outros haplogrupos. Por sua vez, o marcador M14 apresenta variação para todas as amostras dos haplogrupos Q e R estudadas, quando em comparação com os restantes haplogrupos, pelo que pode ser utilizado para classificação de amostras pertencentes ao parágrafo P, o que ainda não era efectuado por nenhum dos InDels presentes em Karafet et al. (2008). Assim, embora os marcadores confirmados não nos tenham dado informação adicional suficiente para a criação do multiplex desejado inicialmente, a análise detalhada de toda a informação contida nas bases de dados permitiu-nos concluir que a criação de um multiplex desta natureza só será possível com a detecção e caracterização filogenética de novos InDels, o que apenas será viável com a utilização de métodos de sequenciação de segunda geração.

ABSTRACT

The characterization of the different human Y chromosome haplogroups is currently being done by genotyping SNPs and a small number of InDels. However, due to their specific characteristics, InDels are increasingly gaining importance for forensics, since they allow a fast and easy typing. Therefore, in order to simplify and give strength to the characterization of samples in the main Y chromosome haplogroups, the aim of this work was to create an InDel PCR multiplex allowing this characterization using only one PCR reaction followed by automated electrophoresis.

To achieve this target we have selected the InDels present in the compilation of polymorphisms made by Karafet et al. in 2008. We have chosen these polymorphisms because they are already accepted by the Y Chromosome Consortium, are well characterized and studied at the population level. However, when we assessed their position in the Y chromosome phylogenetic tree we found that we cannot distinguish the major haplogroups described for human populations and, using just these polymorphisms, it is only possible to classify the chromosomes belonging to the haplogroups A, B, O, J and T, and the paragroup DE. In this way, it was concluded that the setup of an InDel PCR multiplex, allowing the full characterization of the 22 main branches of the phylogenetic tree of the Y chromosome, could only be possible with the characterization of new InDel polymorphisms.

In order to get new InDels we have extended the search to already described and validated InDels in two databases: dbSNP (Single Nucleotide Polymorphisms) and the MGS (Mammalian Genotyping Service). From that search we obtained 154 InDels that were subjected to multiple screenings to ensure the elimination of both repeated, inside or between databases, and false polymorphisms, as the confusion between the existence of an InDel because of the existence of an STR, or between Y specific polymorphisms and X-Y polymorphisms. After all the screenings performed, it was confirmed that from all markers in these two databases just 10 were new, potentially polymorphic and Y chromosome specific, and therefore were kept for further study. For these 10 markers specific primers were designed for their PCR amplification and typing in representative samples for 13 of the 17 different haplogroups represented in the phylogenetic tree. From these, only 2 polymorphisms were confirmed, M2 and M14 markers. With a more detailed study of these two markers in samples from different sub-haplogroups, it was possible to conclude that M2 shows the mutated allele in samples of the R haplogroup but it also shows a reversion within the R1b1b2 sub-haplogroup. Therefore, the M2 marker is not recommended for the characterization of R haplogroup samples and for the distinction between R and the other haplogroups. For M14 marker the deeper analysis shows that all studied samples from R and Q haplogroups are different from all other haplogroups and so it can be used to identify samples belonging to the paragroup P which was not yet achieved by any of the InDels presents in Karafet et al. (2008). Thus, although the additional information given by the new polymorphisms identified and characterized in the present work, it was not possible to select a set of InDel markers enough to allow the development of the initially desired multiplex. However, the detailed analysis of the information present in the databases let us conclude that the creation of such multiplex will be possible only with the detection and phylogenetic characterization of new InDel markers, which is only possible using second generation sequencing methods.

INTRODUÇÃO

A - Cromossoma Y

A1-Evolução

Actualmente aceita-se que os cromossomas sexuais X e Y se terão originado a partir de um par de autossomas que divergiram acerca de 300 milhões de anos (Figura 1), altura em que surgiram os mamíferos [2].

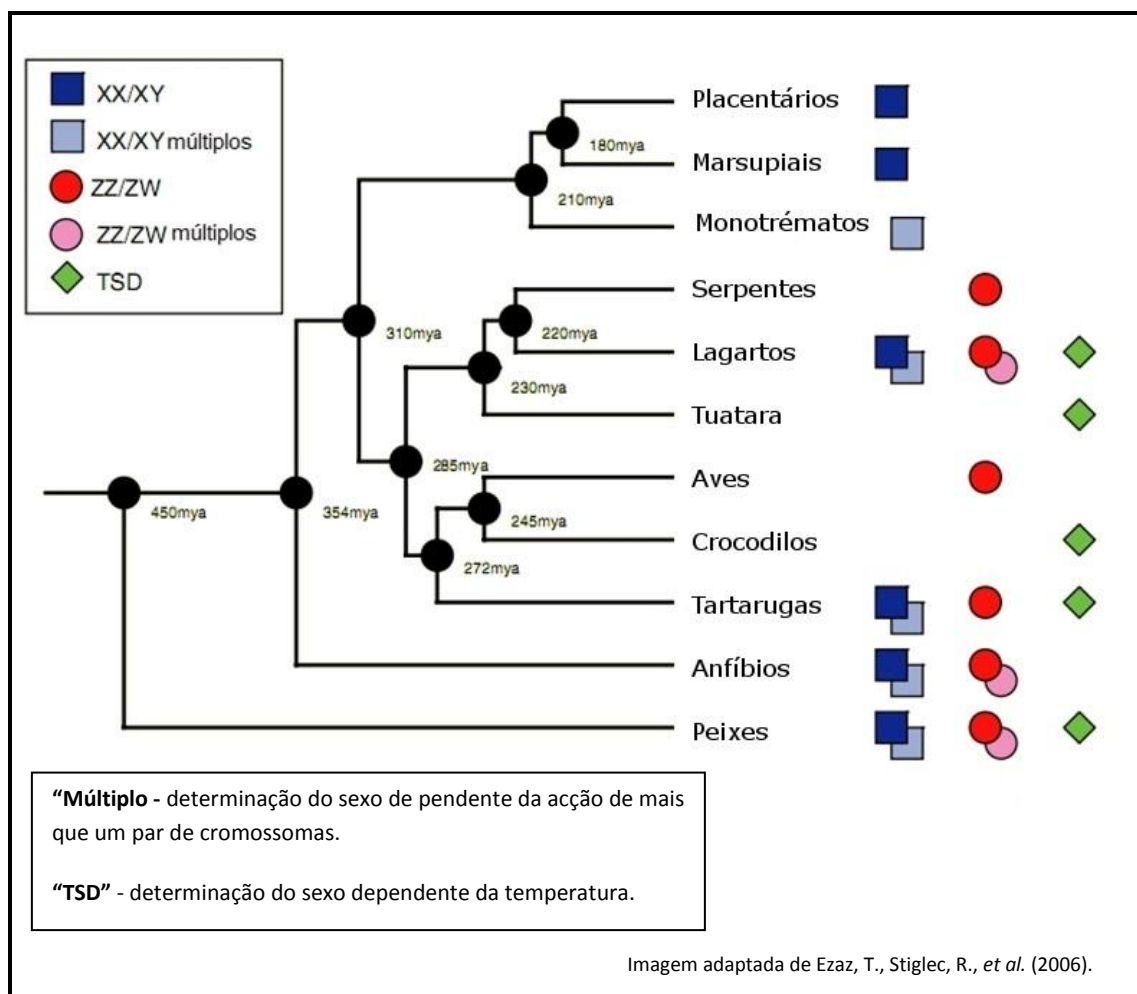


Figura 1: Sistemas de determinação do sexo em vertebrados.

A razão que se pensa ter levado à sua diferenciação foi a perda de recombinação entre os autossomas de origem sendo que esta perda foi favorecida pela aquisição de genes determinantes do sexo (SRY, Sex-determining Region of the Y chromosome). Com a ausência de recombinação houve uma constante acumulação de diferenças entre os cromossomas e, no cromossoma Y, a selecção deixou de actuar num só gene para passar a actuar em toda a região específica do sexo masculino (MSY, Male Specific Region of the Y chromosome) [3]. Com isto, o

conteúdo génico do cromossoma Y foi-se degradando, sendo que dos cerca de 1100 genes do Y ancestral apenas 45 se mantêm e grande parte deles tem um papel importante na reprodução masculina, sofrendo selecção positiva [4].

No entanto, mais recentemente, a sequenciação e análise do genoma do ornitorrinco (*Ornithorhynchus anatinus*) veio sugerir que o sistema de determinação do sexo XY, em mamíferos, não existe há mais de 166 milhões de anos, uma vez que os cromossomas X do ornitorrinco apresentam mais homologias com o cromossoma sexual Z, das aves, do que com os cromossomas X de outros mamíferos [5].

A2-Estrutura

O cromossoma Y (Figura 2) é um dos mais pequenos cromossomas humanos [6]. Representa apenas 2% do genoma humano, e contém aproximadamente 60 Mb de tamanho [7] sendo que, ao contrário dos outros cromossomas, o seu tamanho é muito variável (Figura 3) [8]. Este cromossoma é em grande parte constituído por DNA repetitivo e contém um número muito reduzido de genes. As regiões de eucromatina possuem aproximadamente 23 Mb e retêm 178 unidades de transcrição [9], que codificam, no entanto, apenas 45 proteínas distintas [10], sendo que quase todas elas têm uma função fundamental na determinação do sexo masculino.

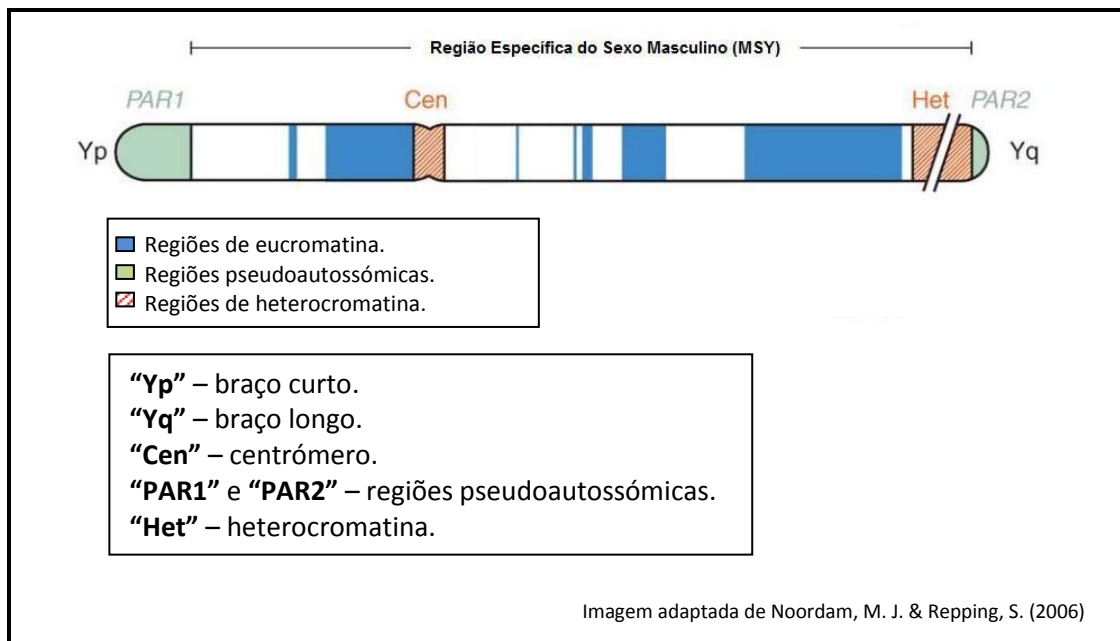


Figura 2: Diagrama esquemático do cromossoma Y.

A região não recombinante do cromossoma Y (NRY, Non-Recombining region of the Y chromosome) compreende aproximadamente 95% do seu DNA pois este cromossoma apenas recombina com o cromossoma X nas regiões pseudoautosómicas (PAR 1 e PAR 2). Mais recentemente a NRY tem sido denominada por região específica do sexo masculino (MSY, Male Specific Y chromosome) devido à verificação de elevada recombinação intracromossómica.

As regiões PAR 1 e PAR2 possuem aproximadamente 2,5 e 1 Mb de tamanho, respectivamente, e localizam-se nas extremidades dos braços curto e longo de ambos os cromossomas. Entre estas regiões existe uma recombinação obrigatória que leva a um correcto emparelhamento de cromossomas durante a meiose no sexo masculino [4]. Para além disso, a existência de PAR 1 e PAR 2 está de acordo com a hipótese de os cromossomas X e Y dos mamíferos se terem originado a partir de um par de autossomas ancestral [4].

A região eucromática do cromossoma Y humano encontra-se dividida pelos seus dois braços, sendo que 8 Mb se situam no braço curto e 14,5 Mb no braço longo [5]. As sequências que compõem esta região podem dividir-se em 3 classes: amplicónica, X-transposta e X-degenerada (Tabela 1) [9].

As sequências X-transpostas (Tabela 1) são compostas por dois blocos no braço curto do cromossoma Y, compreendendo cerca de 3,4 Mb de tamanho. Caracterizam-se por exibirem 99% de identidade com sequências do cromossoma X e resultam de uma mega transposição ocorrida entre os cromossomas X e Y há cerca de 3-4 milhões de anos [11]. Após esta transposição, deu-se uma quebra e inversão no braço curto do cromossoma Y das quais resultaram os dois fragmentos não contíguos que agora constituem estas sequências. Ao contrário do que seria de esperar, tendo em conta a percentagem de semelhança entre estas sequências e o cromossoma X, estas não participam no crossing-over durante a meiose masculina, o que as distingue das regiões PAR. Para além disso, estas são as sequências de eucromatina com a menor densidade de genes e maior densidade de sequências repetitivas [9].

As sequências X-degeneradas (Tabela 1) são formadas por oito blocos situados em ambos os braços do cromossoma Y, apresentando um tamanho total de 8,6 Mb. Estes segmentos apresentam uma identidade de 60 a 96% com as suas regiões homólogas no cromossoma X. Ao contrário do que acontece nas sequências X-transpostas, nestas a densidade de genes é elevada. Dois dos genes codificados nestas sequências são o SRY e o AMELY (Amelogenin, Y-linked). O SRY é responsável pela determinação do sexo nos mamíferos placentários e marsupiais, conduzindo o desenvolvimento embrionário para indivíduos do sexo masculino [9], enquanto o AMELY codifica a amelogenina que vai intervir no desenvolvimento dentário. Este gene apresenta um homólogo no cromossoma X (AMELX), e o facto da amplificação específica do intrão 1 levar à obtenção de fragmentos de tamanho diferente entre os cromossomas X e Y faz com que seja utilizado para a determinação genética do sexo [12].

Por fim, as sequências amplicónicas (Tabela 1) são compostas por sete blocos espalhados por ambos os braços do cromossoma Y, cobrindo cerca de 10,2 Mb, aproximadamente 45% da eucromatina da MSY. Estas sequências, sendo resultado de duplicações intracromossómicas estão organizadas em repetições directas e inversas (palíndromes) e a sua distinção é muito difícil [13]. Dos três tipos de sequências eucromáticas, estas são as que possuem menor densidade de sequências repetitivas e maior densidade génica [9].

Em genética forense, os marcadores do cromossoma Y mais utilizados estão localizados nas regiões X-degeneradas. Isto acontece pois, quando comparadas com as outras regiões, estas possuem menor taxa de homologia, o que torna possível a sua utilização para a detecção de DNA masculino em misturas que contenham DNA feminino.

Tabela 1: Características gerais das diferentes classes de euromatina do cromossoma Y

	Homologias	Distribuição no cromossoma Y	Tamanho Total (Mb)
X-transposta	99% de identidade com o cromossoma X	2 blocos no braço curto	3,4
X-degenerada	60-96% de identidade com o cromossoma X	8 blocos em ambos os braços	8,6
Amplicónica	99.9% de identidade intracromossómica	7 blocos em ambos os braços	10,2

A3-Transmissão

O cromossoma Y transmite-se exclusivamente por via paterna, ou seja, do pai para os descendentes do sexo masculino [14]. Esta característica conjugada com o facto de este cromossoma ser haplóide e não possuir recombinação em quase toda a sua extensão faz com que a transmissão seja a de um cromossoma praticamente inalterado, salvo em caso de ocorrência de mutação. Nestes casos, a mutação ocorrida tem muito pouca probabilidade de ser reparada ou rectificadora [15] pelo que, a sua manutenção ao longo das gerações, faz do cromossoma Y um espelho da história da linhagem paterna [6].

A4-Polimorfismos

A análise da diversidade do cromossoma Y é feita recorrendo tanto a marcadores bi-alélicos como multi-alélicos. No entanto, os polimorfismos encontrados nestes marcadores não podem ser considerados independentes uma vez que, em quase toda a sua extensão, o cromossoma Y não recombina. Por isso, os polimorfismos são estudados de forma haplotípica, sendo denominada por haplótipo a informação correspondente a um conjunto de marcadores multi-alélicos do tipo STR (Short Tandem Repeat), e por haplogrupo, no caso de se tratarem de marcadores bi-alélicos [16].

Os marcadores bi-alélicos incluem os polimorfismos de base única, ou SNPs (Single Nucleotide Polymorphisms), as inserções *Alu* e os InDels. A sua taxa de mutação média é de aproximadamente 10^{-8} pelo que estes polimorfismos são muitas vezes chamados de polimorfismos de evento único. Por sua vez, os marcadores multi-alélicos compreendem os minissatélites e os STRs, ou repetições em tandem, e nestes a taxa de mutação ronda os 10^{-3} .

O elevado número de divisões celulares a que o cromossoma Y está sujeito, a sua estrutura repetitiva, o ambiente oxidativo e a falta de enzimas de reparação no espermatozóide facilitam a ocorrência de mutações, sendo que esta parece ser 4,8 vezes superior à do restante genoma nuclear [10]. No entanto, o seu modo de transmissão e a sua maior susceptibilidade à deriva genética fazem com que as mutações adquiridas nem sempre passem para a geração seguinte, tendendo a perderem-se numa população [10].

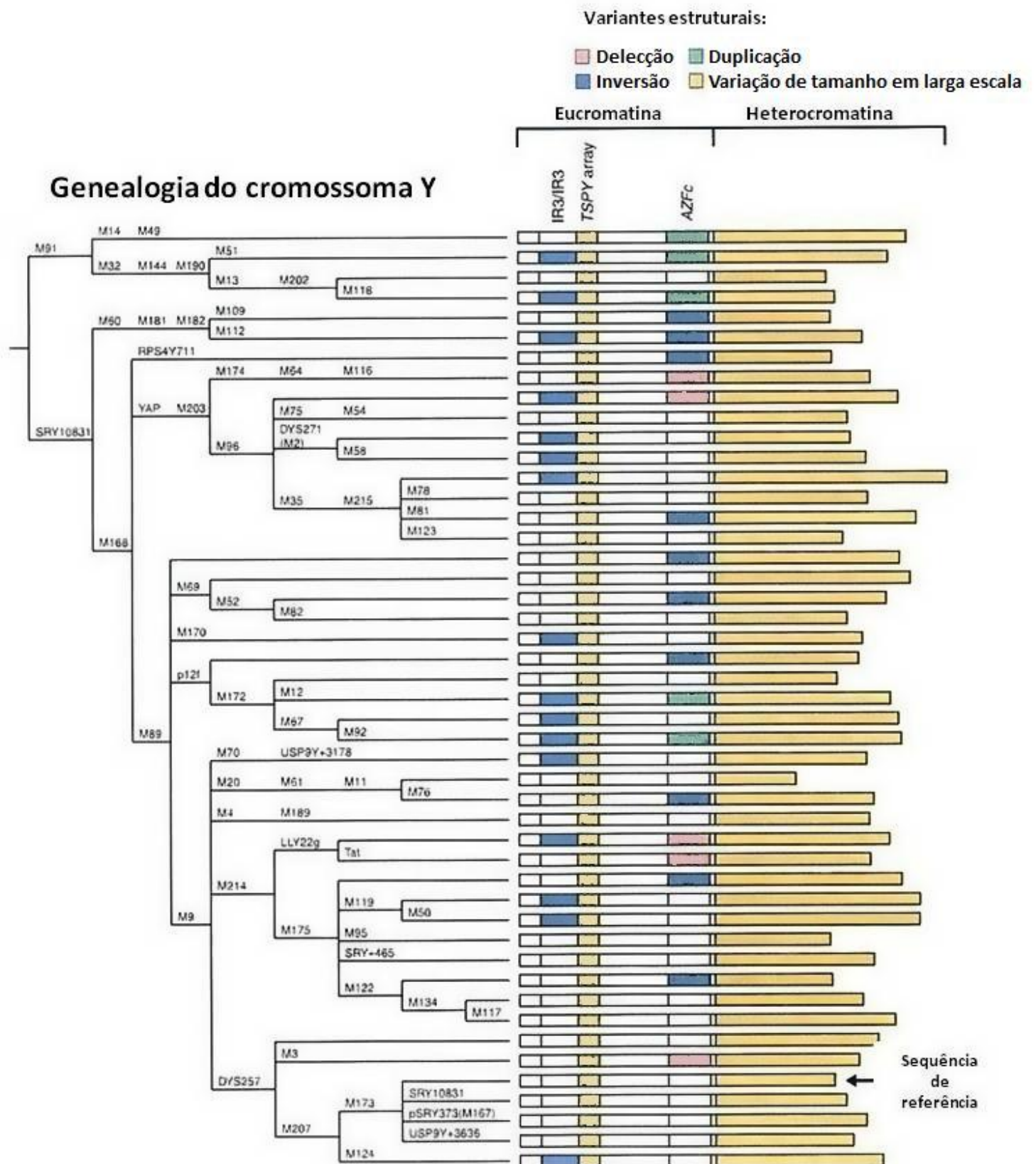


Figura 3: Genealogia do cromossoma Y (esquerda) e polimorfismos estruturais identificados (direita).
Imagem adaptada de Repping, S, *et al.* (2006).

B – Principais haplogrupos e sua classificação

Em 2002, o consórcio do cromossoma Y (Y Chromosome Consortium) publicou a primeira árvore filogenética desse cromossoma, que continha 153 haplogrupos cuja distinção era conseguida através de 245 SNPs [17].

A árvore filogenética do cromossoma Y foi actualizada em 2003 [18] e, mais recentemente, em 2008 [1]. Nesta última actualização foram adicionados aproximadamente 600 marcadores bialélicos que elevaram o número de haplogrupos identificáveis para 311, como se pode observar na Figura 4.

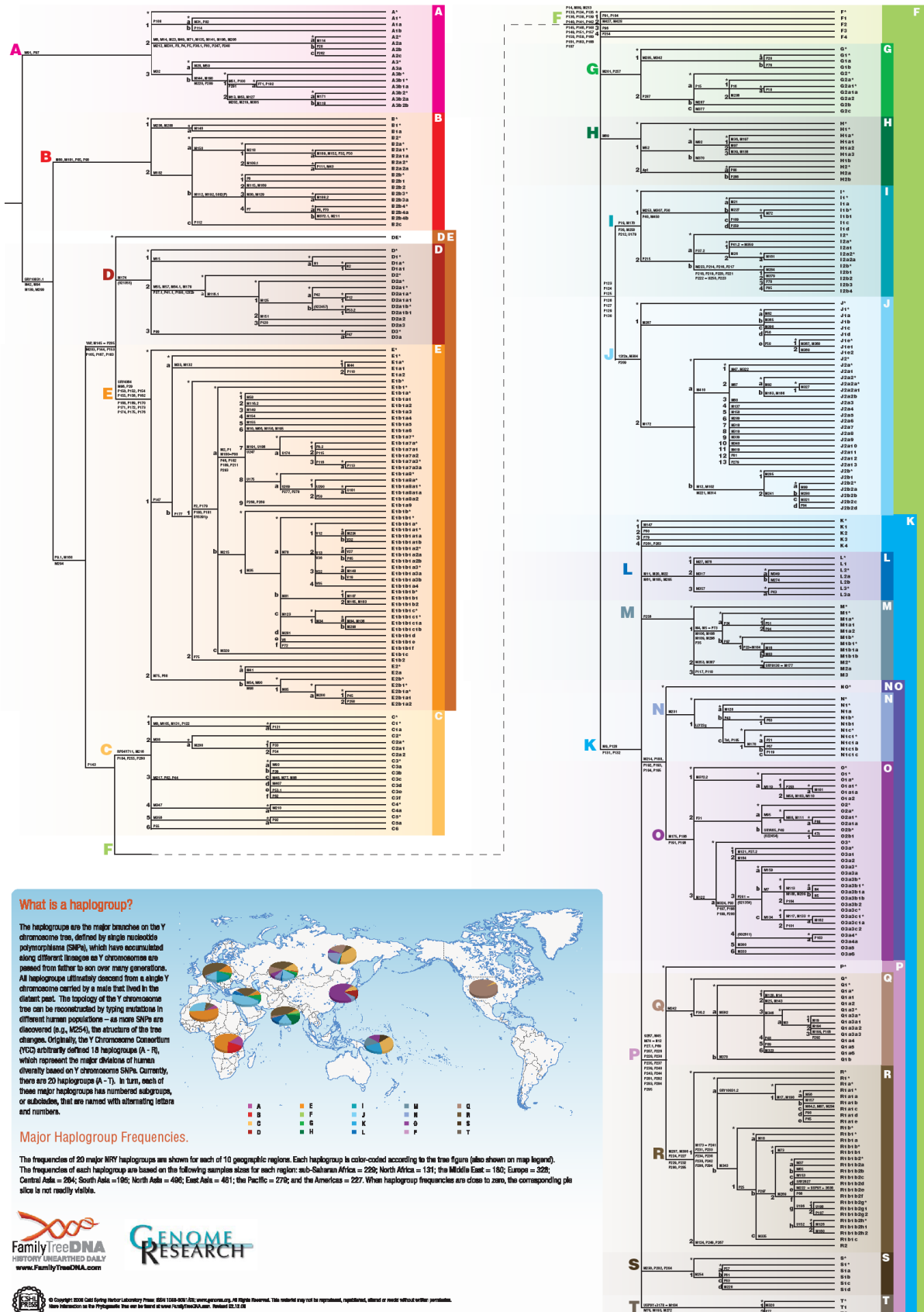
A filogenia do cromossoma Y inclui 20 haplogrupos principais, nomeados com letra maiúscula de A a T. Por sua vez os sub-haplogrupos presentes no interior de cada um dos haplogrupos principais nomeiam-se usando alternadamente números e letras minúsculas, como por exemplo R1a1b1 (Figura 4).

Como consequência das características dos marcadores bialélicos alguns haplogrupos encontram-se confinados a áreas específicas ou continentes, enquanto outros estão amplamente distribuídos no globo (Tabela 2 e Figura 5).

Tabela 2: Exemplo de marcadores bialélicos do cromossoma Y, diagnósticos de alguns dos haplogrupos (HG) principais, e respectiva distribuição populacional, indicando as regiões e/ou populações onde a sua frequência é mais elevada.

HG	SNP	Mutação	Distribuição populacional
A	M91	9T→8T	África (populações Khoisan, Etiópia, Sudão)
B	M60	Ins 1bp	África subsariana (elevada frequência em Pigmeus, África central)
C	M216	C→T	Ásia, Austrália, Oceânia e populações Na-Dene da região norte da América do norte.
D	M174	T→C	Ásia Central (Himalaias, Tibete) e Japão
E	SRY ₄₀₆₄	G→A	África (elevadas frequências em todo o continente), Médio Oriente e Mediterrâneo
G	M201	G→T	Médio Oriente, Mediterrâneo e Cáucaso
H	M69	T→C	Sub-continente Indiano e populações Roma (ciganas) europeias
I	M170	A→C	Europa (frequências elevadas no norte da Europa, Escandinávia)
J	M304	A→C	Médio Oriente, Norte de África, Europa, Ásia Central, Paquistão e Índia
L	M22	A→G	Médio Oriente, Norte de África, Europa, Ásia Central, Mediterrâneo
M	P256	G→A	Oceânia e Indonésia Oriental (Papua Nova Guiné e Melanésia)
N	M231	G→A	Norte da Eurásia
O	M175	Del 5bp	Ásia Oriental, Ásia Central e Oceânia
Q	M242	C→T	Norte da Eurásia (Sibéria), com elevada frequência em populações nativas da América.
R	M207	A→G	Europa (frequências elevadas no oeste Europeu), Ásia central e ocidental
S	P202	T→A	Oceânia e Indonésia
T	M70	A→C	Médio Oriente, África, Europa, Ásia Central, Mediterrâneo

Árvore filogenética do cromossoma Y



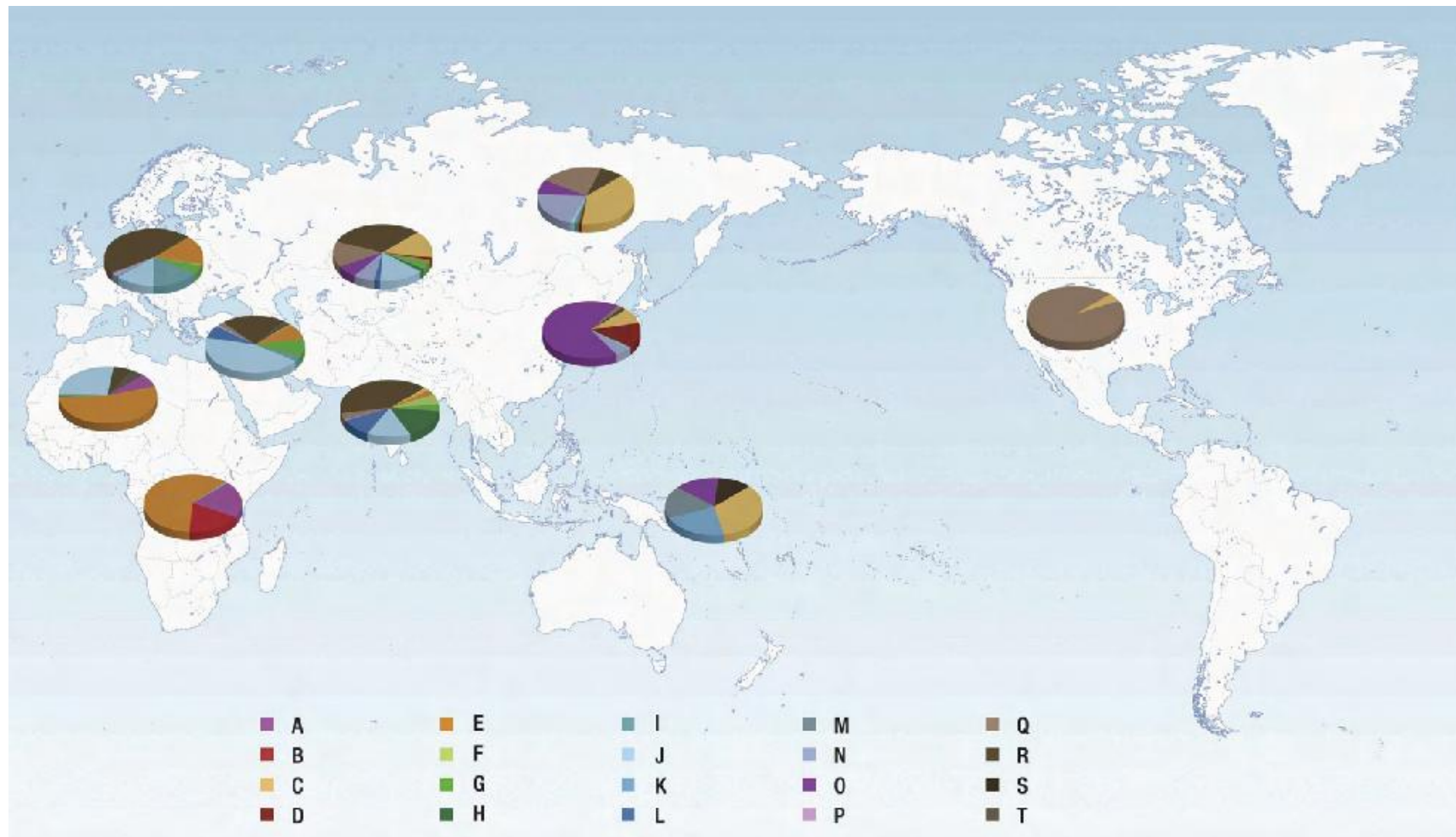


Figura 5: Mapa da distribuição mundial dos haplogrupos do cromossoma Y.

C – InDels

A maioria dos InDels pertencem ao grupo dos polimorfismos bi-alélicos ou di-alélicos e correspondem à inserção ou deleção de uma sequência de DNA de um a centenas de nucleótidos de tamanho [16]. Estes marcadores genéticos, embora não sejam, no genoma humano, tão frequentes como os SNPs, estão amplamente distribuídos e são reconhecidos como uma fonte muito abundante de informação [19-20].

As suas principais características são uma combinação das características dos STRs e dos SNPs [21], sendo elas:

- Ampla distribuição no genoma;
- Baixa probabilidade de recorrência;
- Diferenças significativas nas frequências alélicas entre grupos populacionais;
- Possibilidade de análise recorrendo a amplicões de tamanho reduzido;
- Genotipagem simples, de baixo custo e facilmente automatizada.

As características dos InDels dão a estes marcadores grande importância a nível forense, uma vez que as diferenças nas frequências alélicas os tornam marcadores com potencial para fornecer informação acerca da ancestralidade; a possibilidade de utilização de amplicões de tamanho reduzido abre as portas para a utilização de multiplex e para o sucesso da análise em amostras degradadas; e a simplicidade na análise permite que esta seja feita recorrendo simplesmente a reacções de PCR e posterior análise por métodos electroforéticos, usando assim metodologias já bem estabelecidas nos laboratórios forenses, sem exigências adicionais [21].

OBJECTIVO

Considerando as vantagens da utilização de polimorfismos do tipo InDel, o objectivo deste trabalho foi a criação de um multiplex de InDels específicos do cromossoma Y. Este multiplex teria como finalidade a classificação de amostras do sexo masculino nos principais haplogrupos do cromossoma Y descritos em populações humanas, de uma forma mais simples e robusta que as utilizadas para a tipagem de marcadores do tipo SNPs. Neste sentido, a selecção dos potenciais marcadores a incluir num multiplex deste tipo obedeceu a critérios relacionados com a validade da informação disponível e ao posicionamento e distribuição final dos marcadores na árvore filogenética dos haplogrupos de cromossoma Y. Para isso, usou-se como ponto de partida a lista de InDels compilada em Karafet *et al.* (2008), os quais já se encontram bem caracterizados, estudados populacionalmente e já são utilizados com a finalidade a que propomos o multiplex. Adicionalmente, consultaram-se bases de dados disponíveis online, de forma a efectuar-se um estudo exaustivo da informação relativa a todos os InDels polimórficos específicos do cromossoma Y até agora descritos.

MATERIAIS E MÉTODOS

A – Pesquisa de InDels

Para além dos InDels descritos em Karafet *et al.* (2008) [1] (Figura 7 e Tabela 9) pesquisaram-se outros InDels do mesmo cromossoma, em duas bases de dados: Single Nucleotide Polymorphism Database (dbSNP) do NCBI [22] e base de dados Diallelic Insertion/Deletion Polymorphisms do Mammalian Genotyping Service – National Heart, Lung and Blood Institute (MGS) [23]; os dados usados foram retirados de ambas as bases de dados a 27 de Outubro de 2009.

Os critérios usados para a pesquisa foram os apresentados na Tabela 3.

Tabela 3: Critérios usados na pesquisa de novos InDels nas bases de dados dbSNP e MGS.

dbSNP	MGS
Homo sapiens	Chromosome 25 (Y)
Y chromosome	Marker type : both(biallelic or multiallelic)
Map Weight = 1	Confirmation status: confirmed
InDel/Mixed	
Validated by cluster or by frequency or by submitter or by 2-hit-allele	

Os dados obtidos pela pesquisa foram submetidos a várias triagens com as finalidades seguintes:

- Eliminação de InDels repetidos, recorrendo ao número de referência (rs).
- Eliminação de InDels incluídos em cadeias repetitivas, considerando uma cadeia repetitiva se possuir pelo menos 7 repetições de um mesmo motivo.
- Eliminação de InDels validados por frequência baseada apenas na sequência de Craig Venter.
- Eliminação de InDels com localização em autossomas ou cromossoma X, indicada na dbSNP.
- Eliminação de InDels duplicados, recorrendo à posição física do polimorfismo, obtida a partir de Karafet *et al.* (2008), para os InDels referidos nesse artigo, e a partir do University of California Santa Cruz (UCSC) Genome Browser com NCBI Build 36.1 [24], para os InDels obtidos de MGS e dbSNP.
- Eliminação de InDels duplicados, recorrendo à ferramenta BLAST SNP Sequence do NCBI [25].

Para o último passo indicado, as sequências flanqueantes dos polimorfismos referidos em Karafet *et al.* (2008) foram obtidas através dos primers apresentados pelos autores, através do UCSC In-Silico PCR [26].

Após as triagens referidas, para os InDels restantes, foram estudados os polimorfismos existentes nas sequências flanqueantes. Para esse estudo, a sequência de cada um dos InDels em causa foi retirada do UCSC Genome Browser, com NCBI Build 36.1 [24], usando como critério na procura o rs de cada um deles. Cada uma das sequências retiradas incluí 200 bp a montante e a jusante do respectivo polimorfismo.

B – Desenho de Primers

Nas sequências flanqueantes dos InDels escolhidos foram desenhados primers usando o software Primer3 (v 0.4.0) [27]. As definições usadas para o desenho dos primers encontram-se expostas na Tabela 4.

Tabela 4: Definições usadas para o desenho de primers.

Tamanho do produto	50-200 bp
Tamanho óptimo do primer	20 bp
Temperatura de emparelhamento óptima	60°C
Porcentagem de GCs	20-80%

Para além das definições expostas acima, foi activada a opção “lowercase masking” em que os primers desenhados que possuam um SNP (posições identificadas por letra minúscula na sequência) na extremidade 3’ são automaticamente eliminados.

C - Amplificação do DNA

Usando os primers desenhados anteriormente, foram amplificadas, através da reacção em cadeia da polimerase, amostras pertencentes a 13 diferentes haplogrupos do cromossoma Y e duas amostras de mulheres, como controlo negativo. Foram estudadas um total de 15 amostras pertencendo cada uma delas aos seguintes haplogrupos do cromossoma Y: A, B, C, E (2 amostras, uma com a mutação M2 correspondendo ao sub-haplogrupo E1b1a e outra com a mutação M35 pertencendo ao sub-haplogrupo E1b1b), F, G, H, I, J, L, O, Q, R e T (Figura 6).

As reacções de amplificação foram realizadas num volume final de 10 µL, correspondendo a 5 µL de Multiplex PCR Master Mix (Qiagen GmbH), 1 µL de primer *forward* e 1 µL de primer *reverse* (de um stock a 2 µM) para o marcador em causa, 2,5 µL de água e 0,5 µL de DNA.

Uma vez que as amostras já tinham sido usadas para outros estudos, não foi feita a quantificação do DNA sendo que em trabalhos anteriores a sua concentração foi estimada entre 5 e 20 ng/µL.

As reacções de PCR foram efectuadas em termocicladores GeneAmp PCR System série 2700 (Applied Biosystems). As condições de PCR são descritas na Tabela 5.

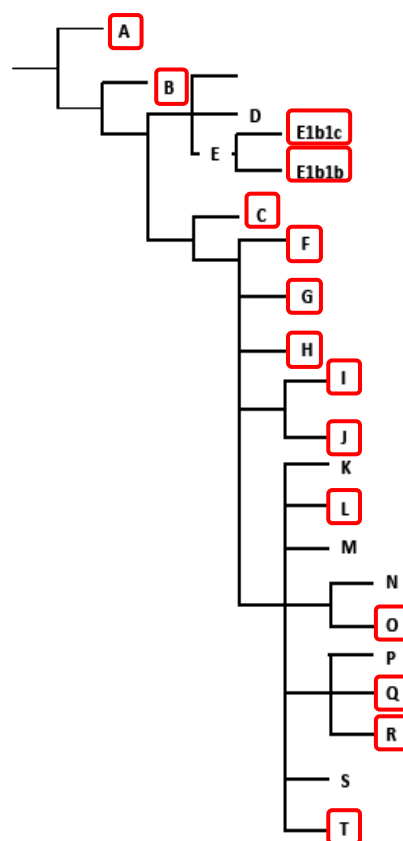


Figura 6: Árvore Filogenética do cromossoma Y com indicação dos haplogrupos usados para detecção de polimorfismo.

Tabela 5: Condições utilizadas nas reacções de PCR.

	Temperatura	Tempo
Desnaturação Inicial	95°C	15'
Desnaturação	94°C	30''
35 ciclos Emparelhamento	60/66°C *	1'30''
Extensão	72°C	1'
Extensão Final	72°C	30'

* a temperatura de emparelhamento foi de 60°C, excepto para M4 cuja temperatura de emparelhamento foi de 66°C.

D – Electroforese

A presença ou ausência de polimorfismo foi detectada através da separação electroforética dos fragmentos obtidos na reacção de PCR. Para isso utilizaram-se duas metodologias: géis não desnaturantes com diferentes concentrações de poliacrilamida, T6C5, T9C5 e T12C5; e electroforese capilar.

No primeiro caso, os géis, com dimensões de 10 cm x 10 cm, foram submetidos a electroforese sobre placa horizontal num Multiphor II (Pharmacia LKB), aplicando-se uma diferença de potencial de 220V. A visualização dos fragmentos de DNA no gel foi feita recorrendo ao método de coloração com nitrato de prata [28].

A electroforese capilar foi levada a cabo num QiAxcel System (Qiagen GmbH), usando o método OM400, com o cartucho QiAxcel High Resolution Kit.

E – Purificação e Sequenciação

Para todos os marcadores nos quais foi detectado ou houve suspeitas de polimorfismo pela separação electroforética, de forma a confirmar a sua presença, foram sequenciadas algumas amostras.

Para isso, os produtos de PCR obtidos foram purificados através da acção da mistura enzimática ExoSAP-IT (USB Corporation), sendo que a 1,5 µL de amplificado foi adicionado 1 µL de ExoSAP. A reacção de purificação foi levada a cabo a 37°C durante 15 minutos, para acção das enzimas, seguida de mais 15 minutos a 85°C, para a sua inactivação, num termociclador GeneAmp PCR System, série 2700 (Applied Biosystems).

A reacção de sequenciação foi realizada num volume final de 5 µL. Este volume resulta de 2,5 µL de produto do PCR purificado, 0,5 µL do primer a usar (*forward* ou *reverse*; de um stock a 2 µM) e 2 µL de kit BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), este último diluído em tampão BigDye Terminator v3.1 5x Sequencing Buffer (Applied Biosystems) na proporção de 1:1. As condições da reacção de sequenciação são indicadas na Tabela 6. Esta reacção foi mais uma vez levada a cabo num termociclador GeneAmp PCR System série 2700 (Applied Biosystems).

Tabela 6: Condições utilizadas nas reacções de sequenciação.

	Temperatura	Tempo
Desnaturação Inicial	96°C	4'
Desnaturação	96°C	20''
35 ciclos Emparelhamento	60°C	10''
Extensão	60°C	1'30''

Os produtos da reacção de sequenciação foram purificados em colunas Sephadex G-50 Fine DNA Grade (GE Healthcare) e ressuspensos em 12 µL de Formamida. A separação electroforética dos produtos foi efectuada num ABI 3130 Genetic Analyzer (Applied Biosystems).

Os resultados foram analisados utilizando o programa Sequencing Analysis v5.2 (Applied Biosystems).

Os alinhamentos foram efectuados usando o programa Geneious v4.8.4 [29].

F – Localização dos novos marcadores na Árvore Filogenética

Após a confirmação da presença/existência do polimorfismo por separação electroforética e sequenciação foram efectuadas novas amplificações para os marcadores em causa, usando, desta vez, um número adicional de amostras de diferentes haplogrupos (n=12 para M2 e n=17 para M14), incluídos nos paragrupos aos quais pertencem os alelos mutados.

As condições de PCR utilizadas foram as descritas na Tabela 5.

Esta análise foi feita com o objectivo de localizar as mutações encontradas na árvore filogenética que descreve os haplogrupos encontrados no cromossoma Y.

As amostras usadas para esta identificação foram as indicadas na Tabela 7.

Tabela 7: Amostras utilizadas com a finalidade de localização da mutação na árvore filogenética do Cromossoma Y.

HAPLOGRUPO	POLIMORFIMO	AMOSTRA	M2	M14
R1a	SRY10831.2	1	✓	✓
R1b1	P25 (+ .2)	2	✓	✓
R1b1	P25 (- .2)	3	✓	✓
R1b1b2a1a1	S21	4	✓	✓
R1b1b2a1a2d	S28	5	✓	✓
R1b1b2a1a2	S116	6	✓	✓
R1b1b2a1a	S127	7		✓
R1b1b2a1a2d3	S139	8	✓	
R1b1b2a1a2f	S145	9	✓	✓
R1b1b2a1*	S167	10	✓	✓
R1b1b2a1a2b	M153	11	✓	✓
R1b1b2	M269	12	✓	✓
R1b1b2a1a2c	SRY2627	13	✓	✓
Q	M242	14		✓
Q	-	15		✓
Q	-	16		✓
Q	-	17		✓
Q	-	18		✓

G – Variação do marcador M2 dentro do sub-haplogrupo R1b1b2

Após a verificação da existência de variação dentro de R1b1b2 para o marcador M2, foram usadas novas amostras pertencentes a este sub-haplogrupo e possuidoras da mutação S167, de forma a confirmar esta variação.

As novas amostras (n=6) foram amplificadas por PCR segundo as condições descritas na Tabela 5, e usando os primers específicos para o marcador M2.

H – Determinação da localização do marcador M2 na região AZF do cromossoma Y

De forma a efectuar a localização do marcador M2 no cromossoma Y foram utilizados como base a sua posição física e os genes mais próximos. Estas informações foram obtidas a partir da dbSNP com NCBI Build 37.1 usando como critério de pesquisa o rs do marcador e os artigos Tessari, A., *et al.* (2004) [30], Park, A. J., *et al.* (2008) [31], Costa, P., *et al.* (2008) [32] e Adams, A. M., *et al.* (2006) [33].

RESULTADOS E DISCUSSÃO

A – Pesquisa de InDels

Como referido anteriormente, o objectivo deste trabalho centra-se na criação de um multiplex de InDels que permita a caracterização de amostras de indivíduos do sexo masculino nos principais haplogrupos do cromossoma Y descritos para populações humanas. Para a concretização deste objectivo é necessário que se conheça o haplogrupo em que cada um dos polimorfismos utilizados se insere. Assim, recorreu-se à compilação de polimorfismos de Karafet *et al.* (2008) uma vez que os polimorfismos aí referidos já se encontram localizados na árvore filogenética do cromossoma Y e estudados a nível populacional. Para os InDels descritos neste artigo (Tabela 9) foi feita a verificação da sua posição na árvore filogenética do cromossoma Y, de forma a averiguar se estes seriam suficientes para a criação do multiplex pretendido. No entanto, veio a apurar-se que, como se mostra na Figura 7, a sua posição na árvore é, nalguns casos, relativamente interna, e em alguns dos haplogrupos o seu número é muito reduzido. Como tal, embora alguns deles pudessem ser utilizados para caracterização de amostras africanas (Figura 7), não são suficientes para a realização do objectivo deste trabalho, que se pretende mais abrangente a nível populacional.

Assim, para reforçar o conjunto dos InDels referidos em Karafet *et al.* (2008), e tornar possível a elaboração do multiplex, partiu-se para a pesquisa de InDels do cromossoma Y, em bases de dados.

As bases de dados escolhidas para a pesquisa foram a Single Nucleotide Polymorphism Database (dbSNP) do NCBI [22] e a Diallelic Insertion/Deletion Polymorphisms do Mammalian Genotyping Service – National Heart, Lung and Blood Institute (MGS) [23].

Ao efectuar-se a pesquisa, e de forma a diminuir a possibilidade de obtenção de falsos polimorfismos, limitou-se o número de InDels reportados apenas para aqueles que já se encontram validados. Esta validação pode, no entanto, ser feita através de frequências (individuais ou populacionais), da observação do alelo mutado em pelo menos dois cromossomas ou, por confirmação do submissor [22].

Como resultado da pesquisa obtiveram-se 154 InDels, 101 a partir da dbSNP e 53 da MGS, sendo possível que alguns se encontrassem já referidos em Karafet *et al.* (2008).

B - Eliminação de marcadores repetidos (I)

De forma a retirar os InDels repetidos, foi efectuada a comparação entre os rs (número de referência) de todos eles. Tendo-se verificado a presença de 13 InDels repetidos (Tabela 8). No entanto, deve ter-se em consideração que 34 dos InDels em Karafet *et al.* (2008) não possuem rs e, por isso, não foram abrangidos nesta análise.

Durante a análise, verificou-se que 23 dos rs obtidos através da dbSNP tinham sofrido fusão. Isto acontece quando pelo menos dois rs indicam um mesmo InDel. Assim, na base de dados, ao detectarem este facto combinaram os rs, passando a manter-se apenas um para cada InDel.

Como resultado final verificou-se que dos InDels obtidos pela pesquisa 13 deles estavam repetidos, 23 foram fundidos e 33 não possuíam rs (Tabela 8).

Nesta fase, foram, então, descartados todos os InDels duplicados, fundidos e, para além destes, de forma a disponibilizar o mesmo tipo de informação para todos InDels e diminuir a possibilidade de repetições, os InDels obtidos pela pesquisa para os quais não foi possível obter o rs.

No entanto, os InDels sem rs descritos em Karafet *et al.* (2008) não foram eliminados, uma vez que já são utilizados para a inferência de haplogrupos.

Assim, transitam para as próximas análises 85 InDels resultantes da pesquisa nas bases de dados destes, 70 foram obtidos pela dbSNP e 15 pela MGS.

Tabela 8: Resultados do estudo de marcadores repetidos no conjunto dos InDels obtidos pela pesquisa nas bases de dados dbSNP e MGS.

		dbSNP	MGS
InDels obtidos pela pesquisa		101	53
InDels retirados	InDels repetidos	8	5
	InDels Fundidos	23	0
	InDels sem rs	0	33
InDels mantidos		70	15

Tabela 9: InDels retirados de Karafet *et al.* (2008) e respectivas características (* UCSC, *Homo sapiens* Genome, Build 36 version 1).

SNP	Haplogrupo (YCC, 2008)	RefSNP ID	Posição no Y *	Mutação
M15	D1	rs3906	20184087-20184088	9bp ins
M17	R1a1	rs3908	20192556	4G->3G
M18	R1b1a	rs3909	20192551-20192550	2bp ins
M39	H1a3		20201636	C del
M57	D2		20211305	1bp ins
M60	B	rs2032623	20337461-20337460	1bp ins
M73	R1b1b1	rs2032634	20348262-20348263	2bp del
M82	H1a	rs2032675	20353562-20353563	2bp del
M84	E1b1b1c1a		20357751	A del
M91	A	rs2032651	20366926	9T->8T
M99	J2b2a		20238682-20238684	1bp del
M111	O2a1		20226693-20226694	TT del
M117	O3a3c1		20224700-20224703	4bp del
M121	O3a1		20366741-20366745	5bp del
M126	R1b1b2h1	rs2032615	20389651-20389654	4bp del
M128	N1a		20227316-20227317	-2bp
M131	C1		20173766-20173774	9bp del
M133	O3a3c1		20175715	T del
M134	O3a3c		20175606	-1bp
M135	A2		20175681	+1bp
M139	BR		20165774	5G->4G
M147	K1		20221470	1bp ins(extraT)
M175	O	rs2032678	14018100-14018104	-5bp
M186	M1	rs2032681	13432855	1bp del
M193	T	rs2032676	13523900-13523899	4bp ins
M199	Q1a3a3	rs2032589	13540505-13540504	1bp ins(extraG)
M254	S1	rs9341297	13532157-13532156	18bp ins
M284	I2b1		21159845-21159848	ACAA del
M317	L2	rs13447360	21164059	GA del
M333	O3a6		13356860-13356909	G ins
M379	I2b2		13536923-13536924	GT del
M390	J1c		2948678	A ins
M419	J2a11		13977300-13977304	AAAAG del
YAP	DE			Alu- ->Alu+
P20	G1a		25029911;23396163	C del
P32	B2a1a		13005252	T-
P48	Q1a4		13006660	T ins
P50	B2a1a		13005250	G del
P62	C3f		20303337	C ins
P63	N1b1		20303215	C ins
P118	M3		13529200	Cdel
P258	E2b1a2		13540552	Ains
P292	Q1a3a3		13540161	G ins
12f2.b	D2			present->absent
12f2a	J			present->absent
MSY2.1	B2b4b			4->3
MSY2.2	O1			4->3

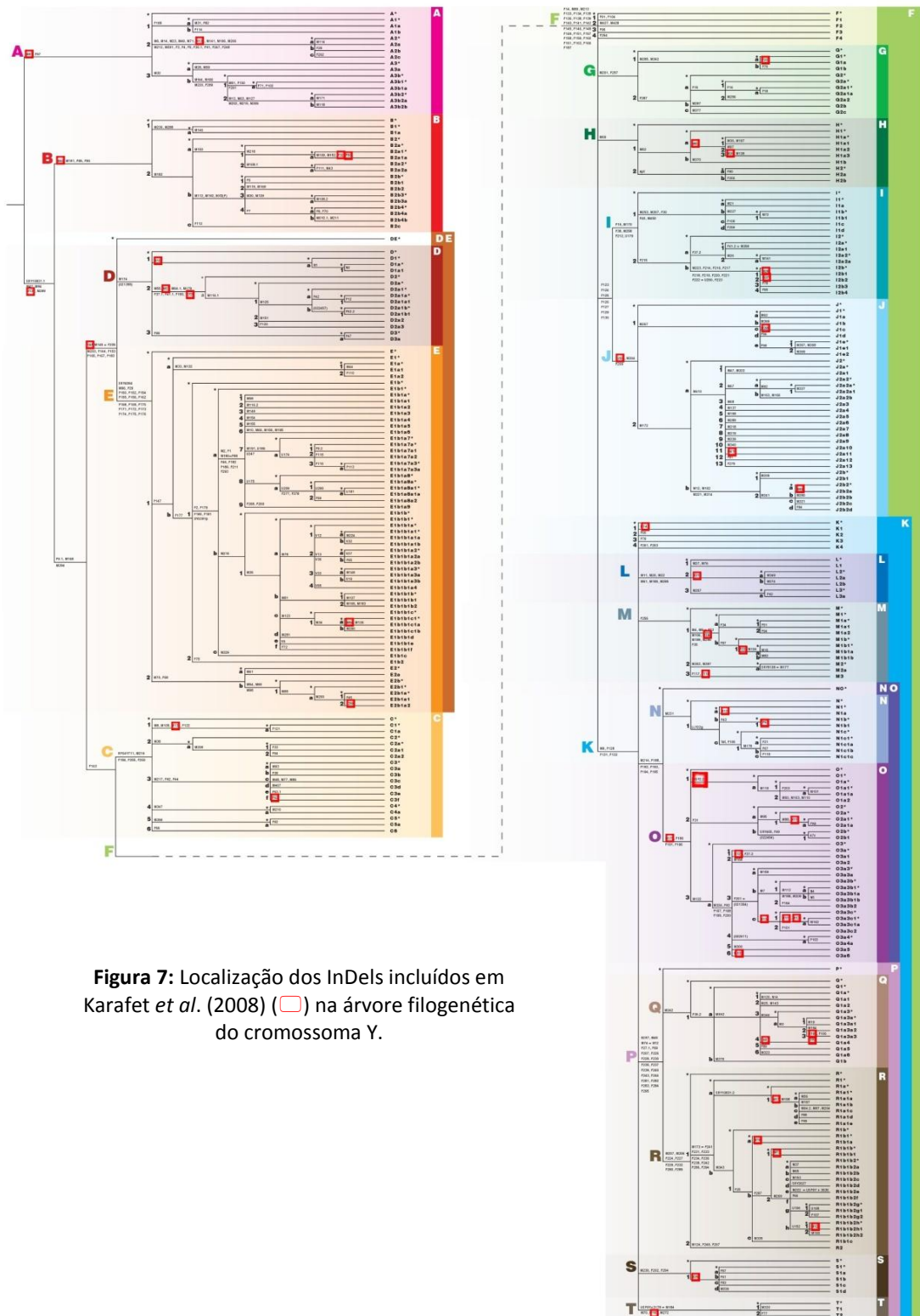


Figura 7: Localização dos InDels incluídos em Karafet *et al.* (2008) (□) na árvore filogenética do cromossoma Y.

C - Eliminação de “falsos” polimorfismos

Mesmo restringindo a pesquisa nas bases de dados, corremos o risco de alguns dos InDels reportados não serem verdadeiros polimorfismos. Por isso, e para aumentar o sucesso na procura de verdadeiros polimorfismos, efectuaram-se várias triagens nos 85 InDels obtidos pela pesquisa.

As triagens efectuadas encontram-se expostas na Tabela 10, assim como, o número de InDels retirados e os mantidos em cada uma delas.

Tabela 10: Selecções a que foram submetidos os InDels obtidos, número de InDels que não obedecem aos critérios de inclusão e número de InDels que passa à próxima fase de análise.

Descrição da selecção	Número de InDels retirados	Números de InDels mantidos para posterior análise
		85
Alelos pertencentes a regiões repetitivas	31	54
Validação efectuada no genoma de Venter	3	51
Amplificação em autossomas ou cr. X	35	16

No que diz respeito às triagens efectuadas, embora todas elas se apliquem na escolha dos verdadeiros polimorfismos, cada uma apresenta um alvo próprio.

A eliminação dos InDels com localização em cadeias repetitivas foi feita para que não houvesse confusão entre a existência do polimorfismo em causa e um STR, uma vez que, os STRs são também InDels mas, atendendo às suas características, não é este o tipo de polimorfismo que procuramos [21].

Os InDels em que a validação era feita pelas frequências baseadas na sequência de Craig Venter também foram retirados uma vez que, a informação relativa a estas frequências nos dava uma contagem de cromossomas Y igual a 2, o que não é possível considerando um cariótipo masculino, tal como o apresentado em Levy *et al.* (2007), e através do qual foram calculadas essas frequências.

Assim, o principal objectivo deste passo foi a eliminação de polimorfismos localizados em regiões não específicas do cromossoma Y ou, para os quais se espera a existência de sequências homólogas noutros cromossomas (mais frequentemente no cromossoma X) responsáveis pela não especificidade do polimorfismo, ao contrário do que é indicado na base de dados.

Por fim, retiraram-se os InDels em que nos era indicada na base de dados a presença adicional do polimorfismo em autossomas ou no cromossoma X.

Esta triagem serviu, tal como a anterior, para eliminar os InDels cuja variação fosse entre o cromossoma Y e um outro cromossoma, de forma a manterem-se apenas os InDels específicos do cromossoma Y.

Após estas três selecções mantêm-se 16 dos InDels obtidos através das bases de dados.

D - Eliminação de marcadores repetidos (II)

Para os 16 InDels que se mantêm até aqui, e de forma a garantir a inexistência de repetições com os referidos em Karafet *et al.* (2008), principalmente com aqueles que presentes neste artigo não possuem rs, foram efectuadas mais duas análises com o objectivo de eliminar estas possíveis repetições.

A primeira consistiu no cruzamento das posições físicas dos polimorfismos. Para isso, foram retiradas de Karafet *et al.* (2008) as posições físicas dos InDels obtidos deste artigo e de UCSC Genome Browser as dos InDels obtidos das bases de dados. Para a obtenção das últimas foi usado NCBI Build 36.1 [24], uma vez que esta foi a versão utilizada no artigo e só assim se torna possível a comparação das posições entre os dois grupos de dados. Através desta análise não se verificou a existência de nenhuma repetição.

Mesmo assim, foi efectuada uma segunda análise recorrendo à ferramenta BLAST SNP Sequence do NCBI [25]. Esta ferramenta permite procurar SNPs numa sequência alvo. Assim sendo, uma vez que os InDels estão incluídos nos polimorfismos que a ferramenta localiza, e os 16 InDels que se mantêm estão todos referenciados na dbSNP, que é a base de dados usada como referência pela ferramenta, estes InDels serão detectados caso existam na sequência alvo introduzida.

Para esta análise usaram-se como sequências alvo as obtidas através do UCSC In-Silico PCR [26] com base nos primers fornecidos em Karafet *et al.* (2008) para cada um dos InDels lá referidos. Mais uma vez, não se verificaram repetições entre os InDels.

Estas análises vieram confirmar a inexistência de repetições entre os 16 InDels obtidos pela pesquisa nas bases de dados dbSNP e MGS e os referidos em Karafet *et al.* (2008), incluindo, desta vez, aqueles que sendo de Karafet *et al.* (2008) não possuem rs.

E - Análise das sequências flanqueantes

De forma a eliminar a existência de limitações na análise dos polimorfismos pretendidos foram analisadas as sequências flanqueantes de cada um deles.

Como resultado desta análise obteve-se a eliminação do marcador M5. Esta eliminação foi efectuada uma vez que este marcador está incluído num mega InDel com 1252 bp de tamanho (M5 - Tabela 12) e, por causa disto, a sua análise estaria sempre dependente da presença deste mega InDel.

F – Desenho de Primers

De todas as triagens efectuadas até ao momento passam para esta fase, e vão ser alvos para o desenho de primers, 15 dos InDels reportados na pesquisa. As características destes InDels são expostas na Tabela 11.

De entre as opções introduzidas no software para o desenho de primers (Tabela 4 em Materiais e Métodos) a opção “lowercase masking” é de grande importância uma vez que, ao eliminar os primers que possuam um SNP na extremidade 3', nos permite aumentar a segurança no emparelhamento dos primers obtidos.

Tabela 11: Principais características dos InDels seleccionados para o desenho de primers.

Marcador	RefSNP ID	Fonte	Alelos observados	Duplicação no Y	Nº de duplicações no Y
M1	rs16664	Marshfield	-/AA	Sim	4
M2	rs1611049	Marshfield	-/ACAACACACTG	Sim	2
M3	rs3840378	dbSNP	-/AA	Não	-
M4	rs4040916	dbSNP	-/A/G/GTAG	Não	-
M6	rs4537491	dbSNP	C/CC/T/TT	Não	-
M7	rs7892881	dbSNP	C/GT/T/TC	Não	-
M8	rs13447439	dbSNP	-/T	Não	-
M9	rs34014582	dbSNP	-/C	Sim	3
M10	rs34415580	dbSNP	-/G	Sim	3
M11	rs34546780	dbSNP	-/T	Não	-
M12	rs34612261	dbSNP	-/TT	Não	-
M13	rs35029086	dbSNP	-/T	Não	-
M14	rs35470479	dbSNP	-/TTAG	Não	-
M15	rs35533781	dbSNP	-/CTCA	Não	-
M16	rs35556857	dbSNP	-/GTAA	Sim	3

Após a introdução dos 15 marcadores no software [27], e considerando aos resultados obtidos, houve cinco casos em que, de acordo com as condições impostas e referidas na Tabela 4 (Materiais e Métodos), os primers obtidos levariam à formação de um produto amplificado com outras alterações de tamanho para além da que se pretendia estudar. Os marcadores em que isto aconteceu foram M7, M8, M10, M11 e M15 sendo que, por este facto, foram excluídos dos passos seguintes deste estudo.

Mais especificamente, o marcador M7 foi retirado uma vez que existem outros InDels (zonas rodeadas a vermelho na Tabela 12) na sua sequência flanqueante que influenciam o tamanho do fragmento amplificado, não sendo possível o desenho de primers evitando estas regiões. O mesmo acontece com os marcadores M11 e M15, o que levou igualmente à sua exclusão.

No caso do marcador M8, embora o marcador em si não pertença a uma região repetitiva, existe, nas suas proximidades, uma zona de repetição (rodeada a verde na Tabela 12) impossível de evitar no desenho de primers e que levaria a alterações extra de tamanho do fragmento amplificado, podendo mascarar o polimorfismo em estudo.

Por fim, o marcador M10 foi retirado uma vez que está incluído num STR de motivo complexo, e como já foi dito anteriormente este não é o tipo de polimorfismo que se pretende utilizar na criação do multiplex.

Assim, nesta fase foram retirados um total de cinco marcadores que não satisfaziam os critérios necessários para a amplificação específica do InDel que se pretendia analisar. Passam, então, para a fase de confirmação do polimorfismo dez marcadores, sendo eles M1, M2, M3, M4, M6, M9, M12, M13, M14 e M16 (Tabela 12).

Tabela 12: Sequência parcial da zona onde se inserem os marcadores M1 a M16 com marcação de primers *forward* (>>), *reverse* (<<), SNPs presentes no primer (■), polimorfismo alvo (**), zonas a excluir (X), InDels no fragmento amplificado (□) e zona repetitiva no fragmento amplificado (□).

[illegible]

M3_rs3840378	<p>AGTTACTATGAATCAAAGCGTAAATACACAAACACAATATACAATCCAAAATAGATGTAC >></p> <p>AGCATTCTGGAATAAAGCAAaaGAGTGTTTCATTCACACACACAGTAGCTTCAAAACgGTT >>>>>>>>>>>>>>>>> ** <<<<<<<<<<<<<<<<<<<<</p> <p>CGATCTGTTTGTTCtCATGTAGTTTTgTAAAGATGGAAAAAAGGACTTTGGTCATCAAG << X X</p>
M4_rs4040916	<p>CACACATTTATGCAGATTTaTGCTTCATtTTTTaAACTCCATAGCAATATTTTGAAAATT >>>>>>>></p> <p>TTGTGCACCATCATTATGAC-AaTAGAGATTTCTATACCTGTTTaTATATTTACCTTTAA >>>>>>>>>> X *</p> <p>TATAGCACTTTCTATTTTCATATGCTGTTATGATGCTGTGCAGCATCATTTCaTTTTTGt <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<</p>
M5_rs4045884	<p>Polimorfismo incluso numa mega inserção-delecção com 1252 bp de tamanho (rs60045165).</p> <p><u>tgaagatttatccctcacataggattccagaacactcctgctgagtctgaagttttattcatcacataggactccagaacactgctaagagaggtctgaa</u> <u>tgtccctcacattgatattccagaacactccttctgagtctgaatgtttgttcctcacataggattcgagaacactcctgctgagtctgaatgtttgtc</u> <u>ccttacctaagattcgagaacattcacactgagatataaatgcttgcccttaacataagatttcagaacactgctcctgaggtctgaaagttttgtccctc</u> <u>Acataggattccagaactcctcctgctgagtctgaaagtttataccgcacataaggattccagaacactgctgctgagttagaaagttttctgtcacata</u></p>

Resultados e Discussão

Resultados e Discussão

[illegible]

[illegible]

G - Amplificação do DNA

A escolha dos haplogrupos usados para amplificação teve como principal objectivo a ampliação da possibilidade de detecção de diferenças entre eles. Para isso, usou-se o maior número de haplogrupos disponíveis, de forma a abranger a quase totalidade dos descritos na árvore filogenética do cromossoma Y [1].

Foram também usadas amostras do sexo feminino, estas para confirmar a presença/ausência dos marcadores em causa noutro cromossoma que não o Y e, assim, demonstrar a especificidade dos polimorfismos para o cromossoma Y.

No que diz respeito às temperaturas de emparelhamento utilizadas, começou por usar-se uma temperatura de emparelhamento de 60°C para todos os pares de primers. No entanto, verificou-se que a esta temperatura ocorria a amplificação de fragmentos inespecíficos para o marcador M4. Assim, e de forma a eliminar estes fragmentos, efectuou-se a reacção de amplificação para este marcador com temperaturas de emparelhamento de 64 e 66° C, tendo-se verificado que estes deixam de se observar à temperatura de 66°C. Por este motivo, passou a utilizar-se esta temperatura de emparelhamento para amplificação do marcador M4.

H - Electroforese

Após a análise dos resultados utilizando diferentes condições de separação electroforética verificou-se que o gel de acrilamida com melhor capacidade de separação dos fragmentos amplificados, de entre os três usados, é o T12C5. Nestas condições é usado um gel contendo uma maior concentração de acrilamida pelo que, apresentando uma malha mais apertada, permite uma melhor separação de fragmentos com pequenas diferenças de tamanho.

Verificou-se também que este gel possui melhor poder de resolução que o método de electroforese capilar utilizado, como se pode verificar na Figura 8. Aí pode ver-se que usando um gel de acrilamida T12C5 somos capazes de diferenciar 4 bandas, o que seria de esperar tendo em conta que o marcador usado está duplicado no cromossoma Y, e as duplicações apresentam alelos diferentes, enquanto usando o método de electroforese capilar apenas se vê uma banda.



Figura 8: Comparação do poder de resolução do gel de acrilamida T12C5 (à direita) com o método de electroforese capilar usado (à esquerda).

I – Determinação da especificidade masculina

Ao mesmo tempo que se comprovava a existência de polimorfismo verificou-se que em grande parte dos marcadores também se obtém amplificação nas amostras do sexo feminino (Figura 9). Isto leva-nos a suspeitar da possibilidade do polimorfismo que estamos a analisar estar localizado numa região não específica do cromossoma Y ou apresentar homologia no cromossoma X [4].

Assim, e tendo em conta que a indicação de presença em autossomas ou no cromossoma X foi um dos factores que nos levou a excluir marcadores da análise ainda antes da amplificação, os marcadores com amplificação nas amostras do sexo feminino foram descartados.

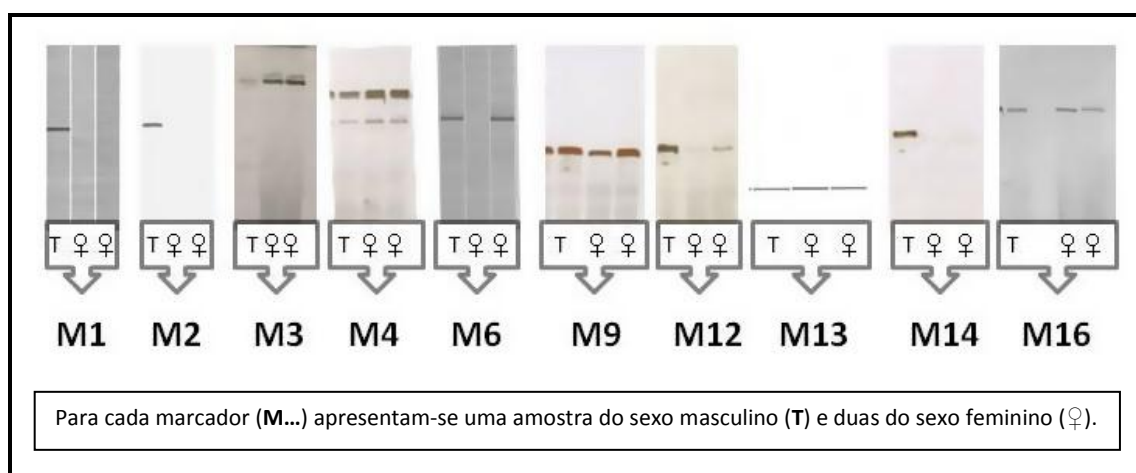


Figura 9: Verificação da amplificação em amostras do sexo feminino dos marcadores estudados.

Passamos então a ter em especial atenção os marcadores M1, M2 e M14, para os quais não se verificou amplificação nas amostras do sexo feminino, podendo-se assim concluir tratar-se de marcadores específicos do cromossoma Y.

J – Verificação da presença de polimorfismo

A Figura 10 mostra-nos o resultado da separação electroforética para o marcador M1. Aqui pode verificar-se que como seria de esperar estão presentes quatro bandas com diferentes mobilidades. No entanto, quando procuramos diferenças entre as amostras dos vários haplogrupos estas não se verificam. Este facto faz com que não se espere que este marcador seja realmente polimórfico, pelo que não nos permite distinguir amostras masculinas dos diferentes haplogrupos do cromossoma Y deixando de ter interesse para o objectivo definido neste estudo.

Também, através dos resultados da separação electroforética, se constatou que os marcadores M2 e M14 apresentam polimorfismo entre os haplogrupos do cromossoma Y.

Para o marcador M2 a amostra pertencente ao haplogrupo R apresenta o alelo longo do InDel, enquanto os outros haplogrupos possuem o alelo curto (Figura 11).

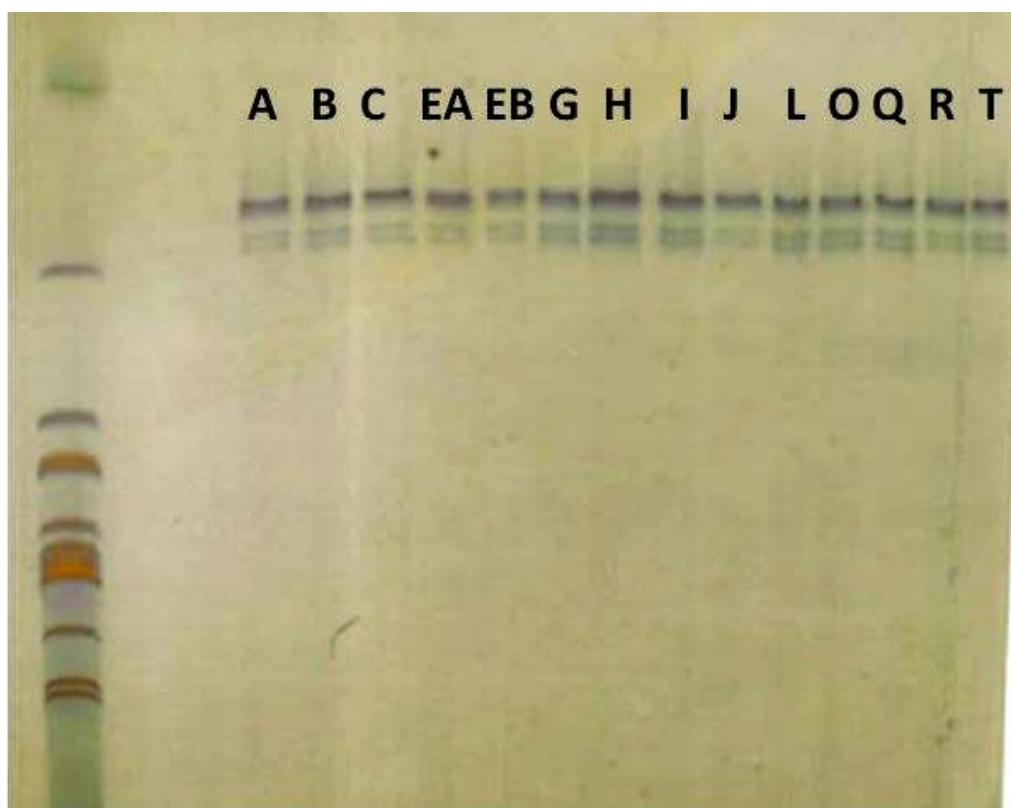


Figura 10: Gel electroforético, T12C5, do marcador 1, com indicação do haplogrupo a que pertence cada amostra.

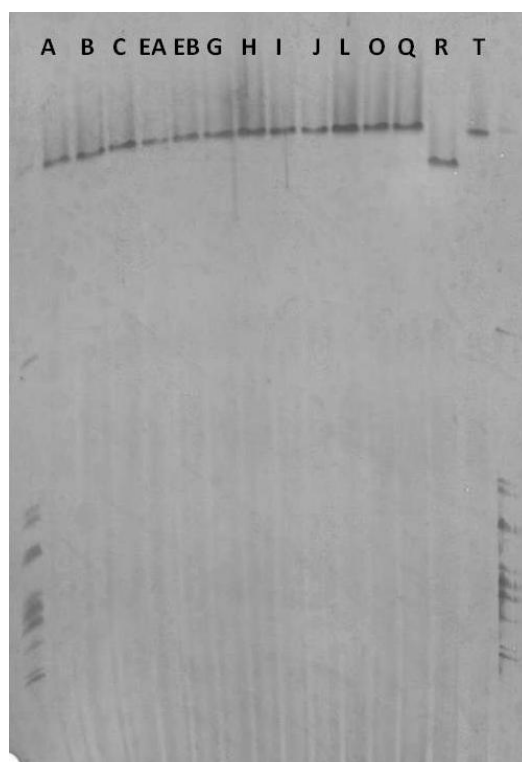


Figura 11: Gel electroforético, T9C5, do marcador 2, com indicação do haplogrupo a que pertence cada amostra.

Por sua vez, no marcador 14 a amostra pertencente ao haplogrupo Q possui um fragmento com maior mobilidade que os dos restantes haplogrupos apresentando, por isso, o alelo curto do InDel em causa (Figura 12).

Como resultado da determinação obtiveram-se diferenças entre haplogrupos em dois marcadores, os marcadores M2 e M14.



Figura 12: Gel electroforético, T12C5, do marcador 14, com indicação do haplogrupo correspondente a cada amostra.

K – Confirmação dos polimorfismos por sequenciação

Para os polimorfismos detectados no passo anterior é necessário, agora, efectuar a confirmação e verificar se a diferença de tamanho identificada pela separação electroforética se deve verdadeiramente ao InDel em causa. Para isso, efectuou-se sequenciação de algumas amostras para ambos os marcadores.

Para o marcador M2 foram sequenciadas as amostras dos haplogrupos Q e R, correspondendo estes aos presumíveis alelos curto e longo, respectivamente. Verificou-se então que a diferença de tamanho observada por electroforese se deve á presença no haplogrupo R e ausência no Q do fragmento [ACAACACACTG], confirmando-se assim que a diferença de tamanho é causada pelo InDel que se pretendia estudar (Figura 13).

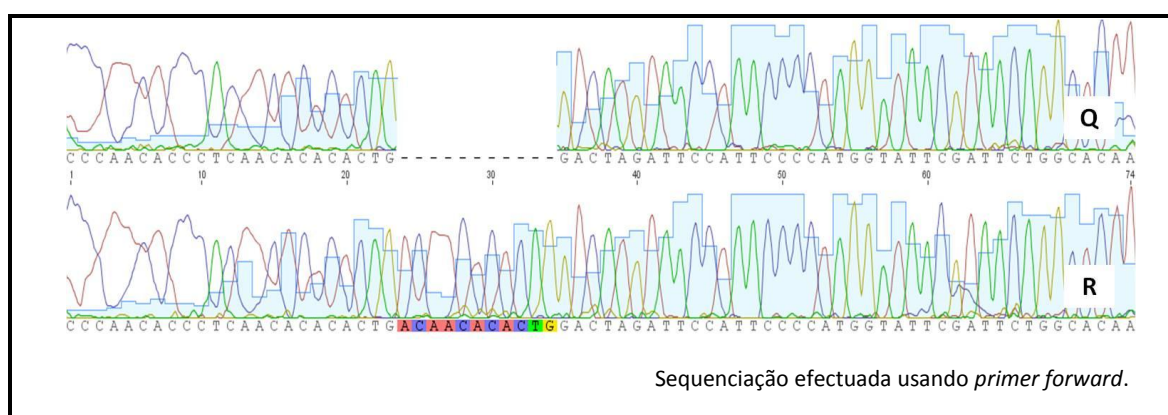


Figura 13: Extracto do alinhamento dos haplogrupos Q (superior) e R (inferior) para o marcador 2.

Por sua vez, para o marcador 14 foram sequenciadas as amostras pertencentes aos haplogrupos O, Q, R e T tendo-se verificado que, os haplogrupos O e T possuem o alelo longo e

os haplogrupos Q e R o alelo curto. Os resultados apresentados na Figura 14 mostram que a diferença de tamanho observada através da separação electroforética resulta da presença ou ausência do fragmento [TTAG], correspondente ao InDel em causa.

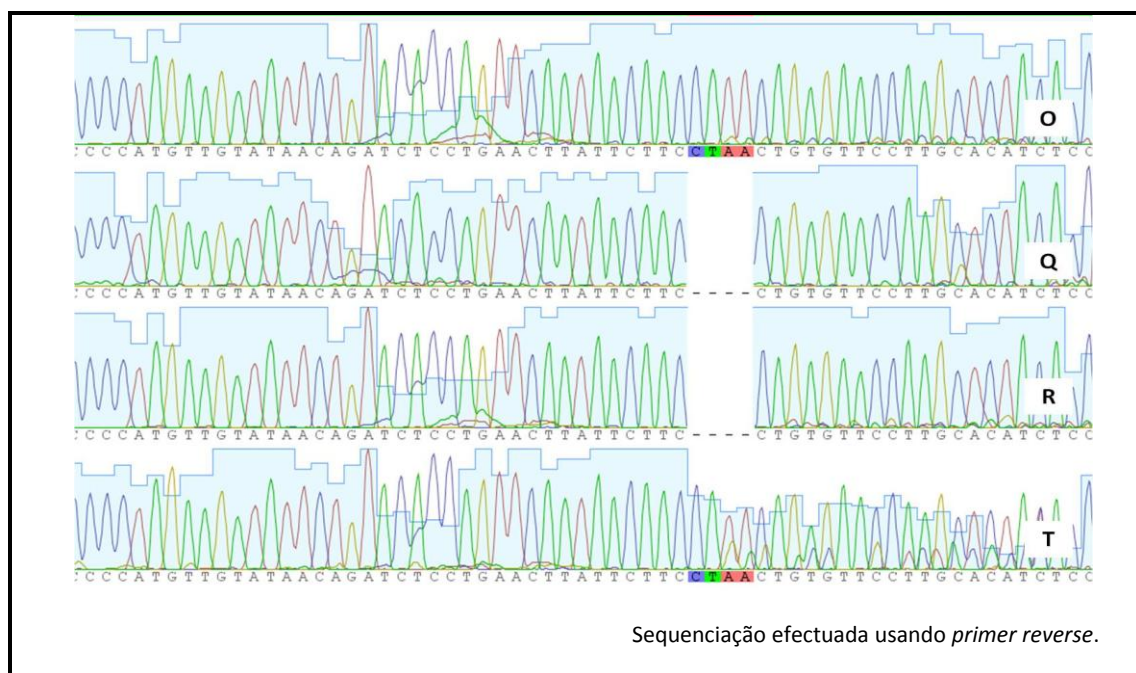


Figura 14: Extracto do alinhamento dos haplogrupos O, Q, R e T para o marcador 14.

L – Localização dos novos marcadores na Árvore Filogenética

De forma a verificar a utilidade dos InDels confirmados é necessário localizá-los na árvore filogenética que descreve os haplogrupos do cromossoma Y [1]. Para isso foram usadas novas amostras pertencentes aos haplogrupos nos quais foi detectado o polimorfismo. As amostras utilizadas foram as apresentadas na Tabela 7, em Materiais e Métodos.

L1 – Localização do marcador M2 (rs1611049)

Para o marcador M2 foram usadas novas amostras pertencentes ao haplogrupo R uma vez que, o polimorfismo foi detectado dentro deste haplogrupo e, desta vez, se pretende verificar a existência de variação dentro deste haplogrupo.

As amostras usadas inicialmente pertencentes aos haplogrupos Q e R foram usadas como referências dos alelos curto e longo, respectivamente.

As variações encontradas estão expostas na Figura 15.

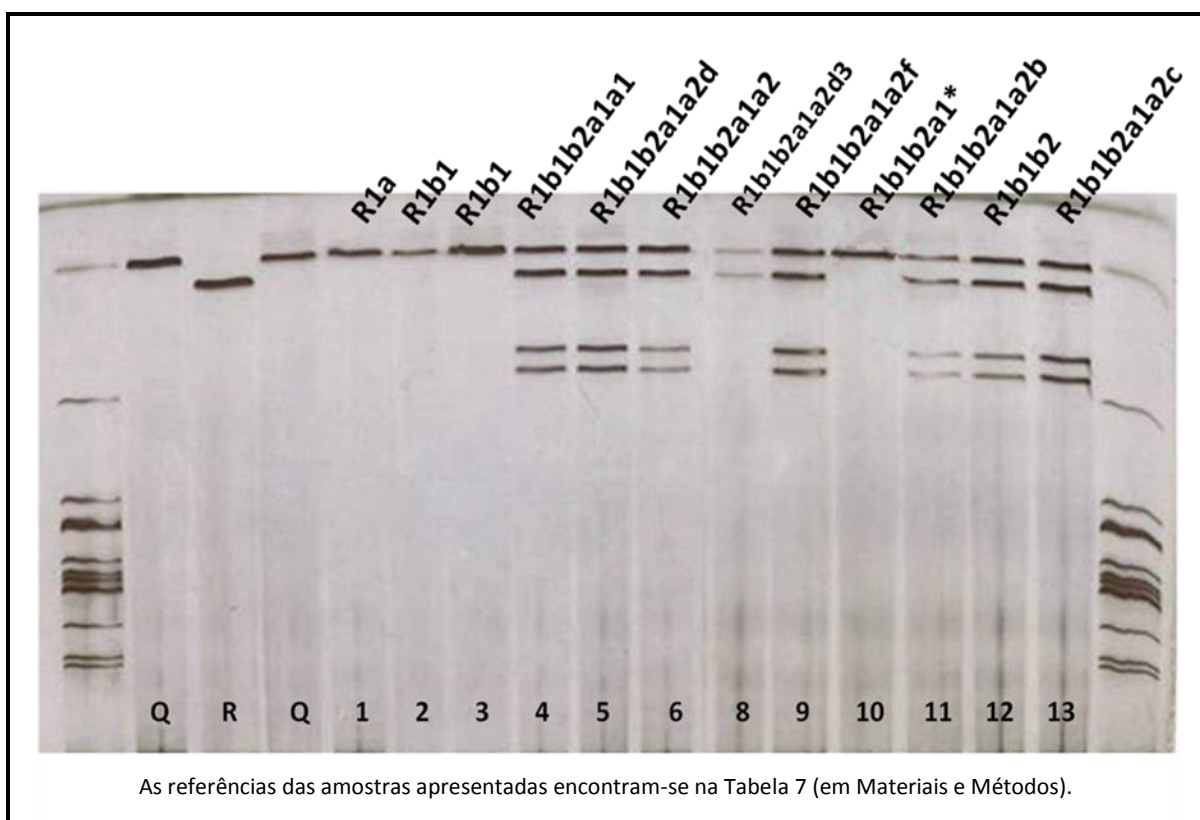


Figura 15: Gel electroforético (T12C5) de amostras dos haplogrupos R e Q, para o marcador 2.

Analisando os resultados obtidos para o marcador M2, as diferenças existentes no haplogrupo R (Figura 17 e Figura 18) conduzem-nos para a hipótese de utilização do InDel na distinção do sub-haplogrupo R1b1b2.

Este marcador é descrito na base de dados como duplicado no cromossoma Y mas, considerando as variações obtidas dentro do haplogrupo R, a localização da ocorrência da duplicação é questionada sendo que, este acontecimento, pode ser explicado de duas formas.

Para a elaboração das hipóteses onde se discute o momento da ocorrência da duplicação, não foi utilizada a amostra identificada como R na Figura 15 uma vez que, embora se saiba que esta amostra pertence ao haplogrupo R, não se sabe em que sub-haplogrupo se situa.

As hipóteses que se põem, para o momento da duplicação são:

- A. A região do cromossoma Y onde se encontra este InDel está duplicada em todos os haplogrupos sendo que, ambos os loci apresentam alelos com o mesmo comprimento/sequência. Assim, os perfis electroforéticos de duas bandas obtidos em amostras de R1b1b2 atribuem-se à ocorrência de mutação no InDel em estudo num dos loci o que leva à existência de alelos de tamanho diferente (Figura 16.A).
- B. A duplicação ocorreu já dentro do haplogrupo R, mais precisamente, distingue o sub-haplogrupo R1b1b2. Assim, as amostras fora de R1b1b2 apresentam apenas uma banda que corresponde ao alelo curto ou longo, caso seja a de maior ou menor mobilidade em gel, respectivamente. Dentro de R1b1b2 atribuem-se à existência de uma banda a presença do mesmo alelo em ambos os loci (o alelo curto para a banda

com maior mobilidade e o alelo longo para a de menor mobilidade), e à existência de duas bandas a ocorrência de mutação no InDel num dos loci (Figura 16.B).

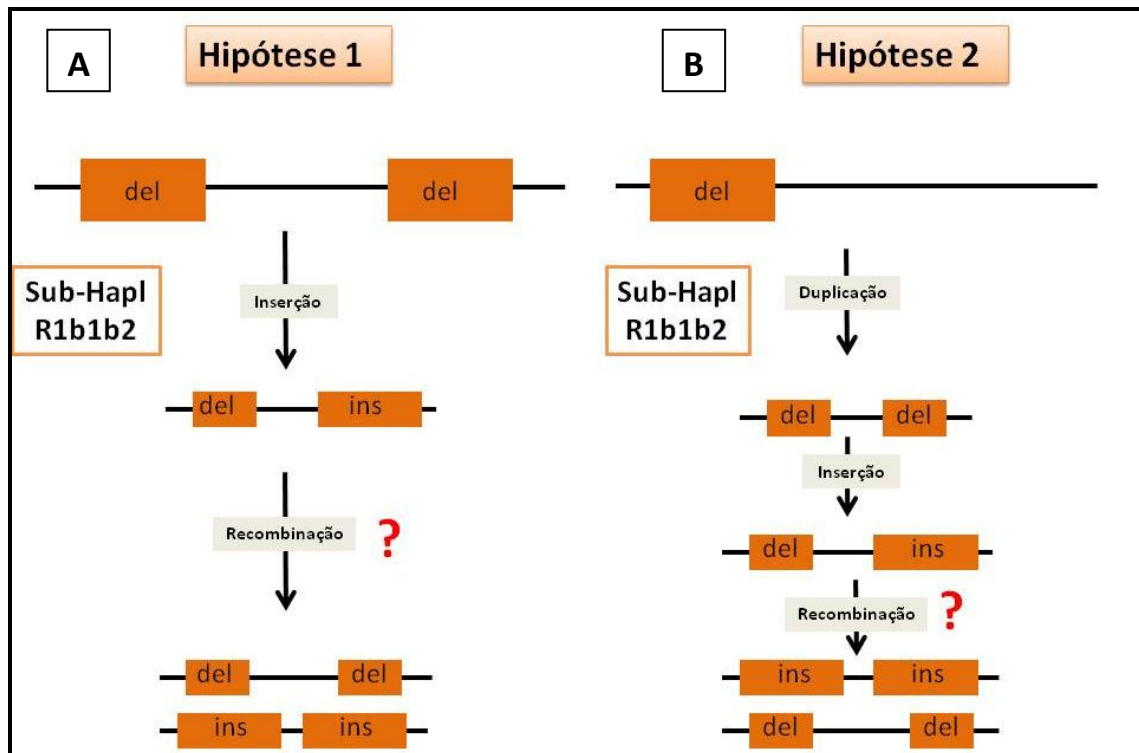


Figura 16: Representação esquemática das hipóteses (A e B) propostas para o momento da duplicação do marcador M2.

Para ambas as hipóteses expostas acima, a existência de amostras em R1b1b2 com perfil electroforético de apenas uma banda pode ser explicada através da existência de um SNP, que embora ainda não descrito a sua real existência não pode ser descartada. Assim sendo, a presença de um SNP no local de ligação do primer num dos locus levaria à ineficiência na amplificação desse locus não nos permite obter as duas bandas características, mas apenas uma única banda.

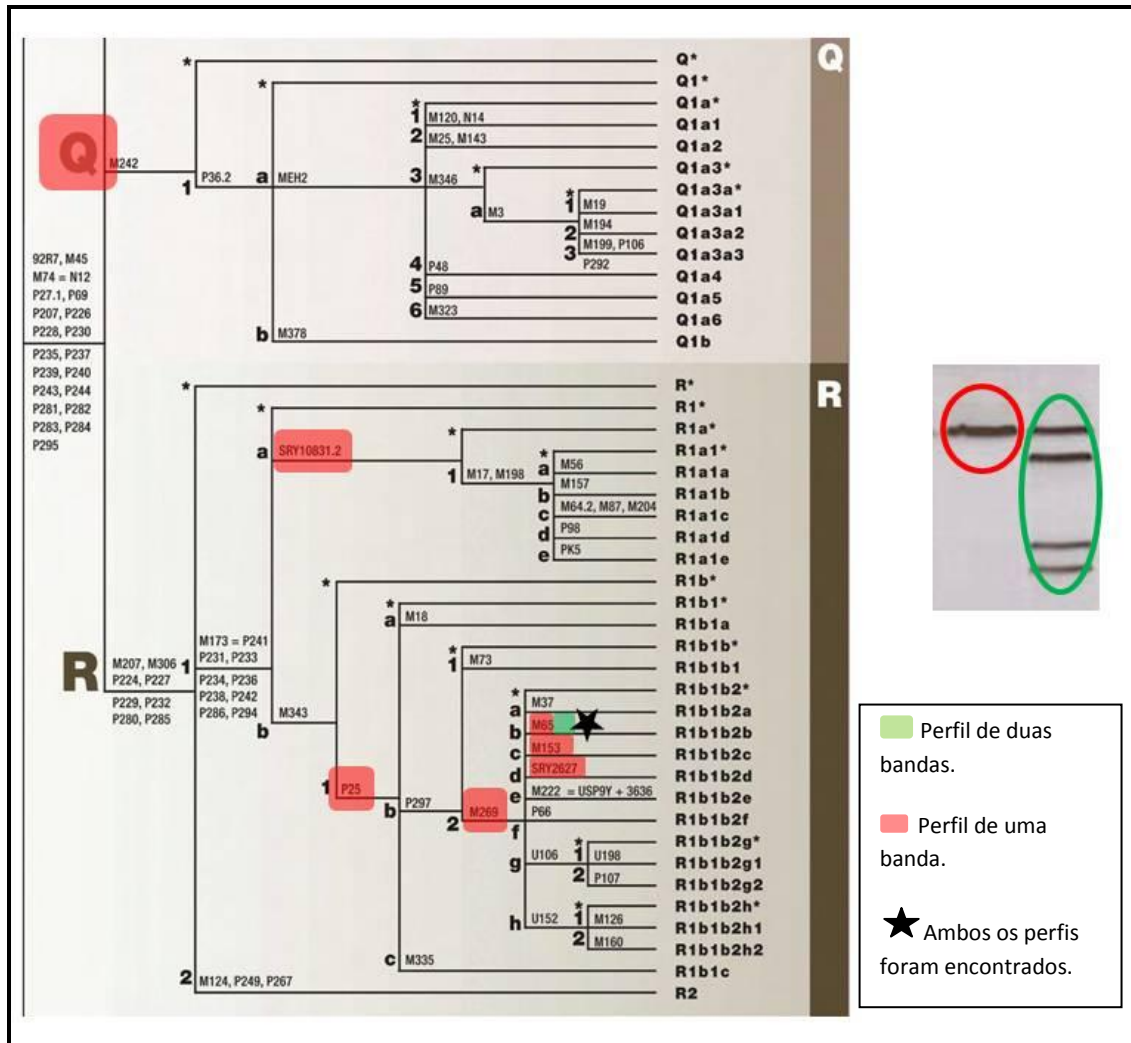


Figura 17: Esquema representativo das amostras dos haplogrupos Q e R estudadas com indicação do tipo de resultado electroforético obtido.

Para além da explicação exposta acima a existência de mais que um padrão de bandas no sub-haplogrupo R1b1b2 faz-nos questionar se este marcador se localizará numa região do cromossoma onde ocorre recombinação intracromossómica, o que poderá ter levado, após a ocorrência da inserção numa das duplicações do InDel, a que as duplicações voltassem a apresentar o mesmo alelo [34].

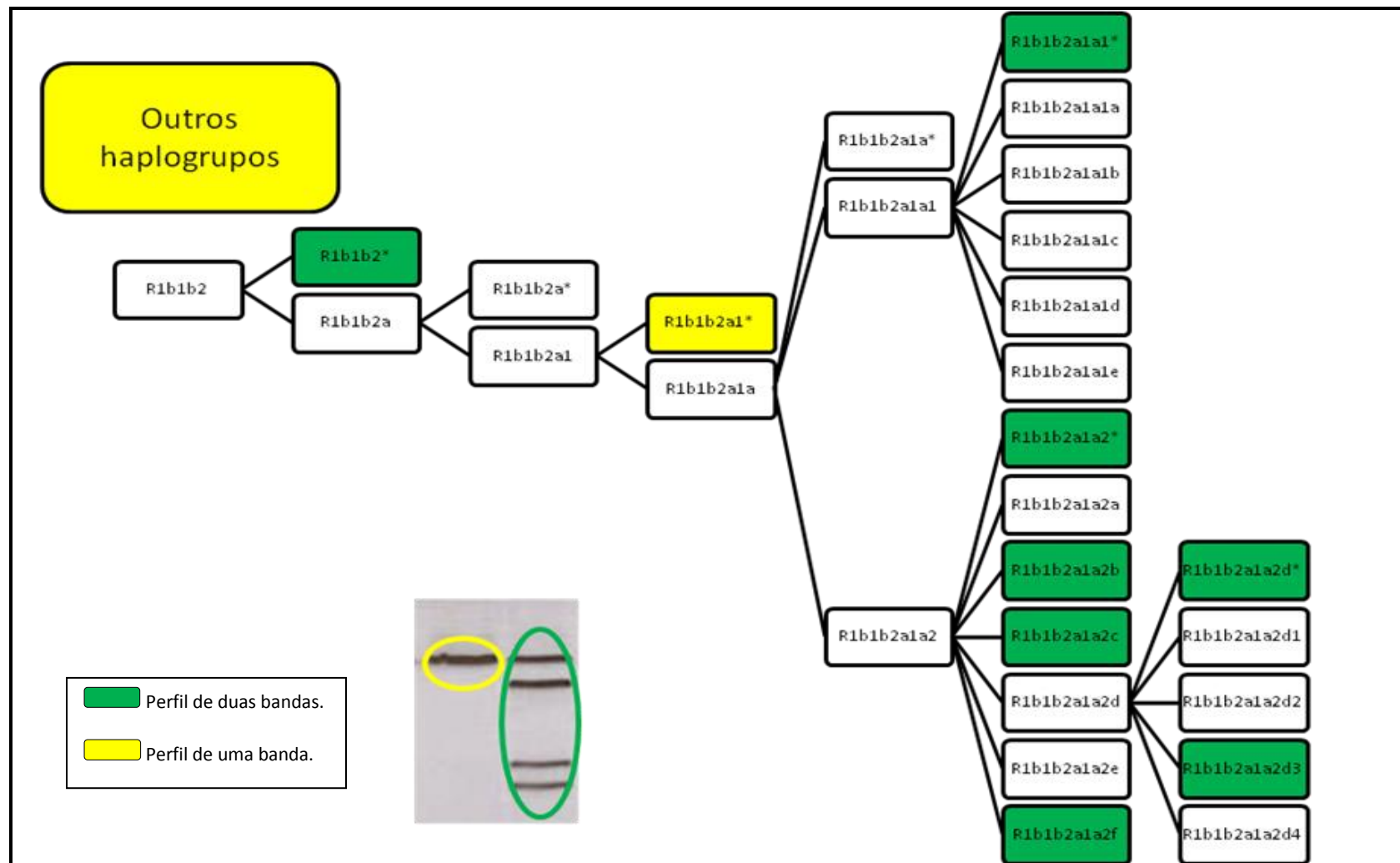


Figura 18: Esquema representativo das amostras do sub-haplogrupo R1b1b2 estudadas com indicação do tipo de resultado electroforético obtido.

L2 - Localização do marcador M14 (rs35470479)

Usando os primers para o marcador 14, foram amplificadas novas amostras dos haplogrupos Q e R de forma a verificar a existência de variação dentro destes haplogrupos. Os resultados obtidos encontram-se expostos na Figura 19.

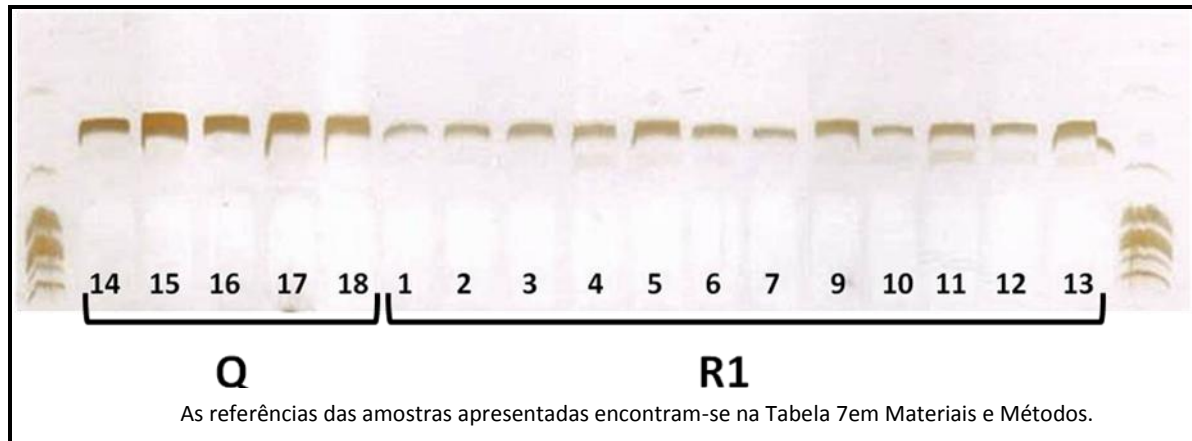


Figura 19: Gel electroforético de amostras dos haplogrupos Q e R, para o marcador 14.

O marcador 14 permite distinguir os haplogrupos Q e R de todos os outros utilizados. Assim, este marcador parece localizar-se no parágrafo P podendo, no entanto, a sua localização, que se indica na Figura 20, sofrer alteração aquando de um estudo mais exaustivo de amostras pertencentes aos diversos haplogrupos que constituem P*.

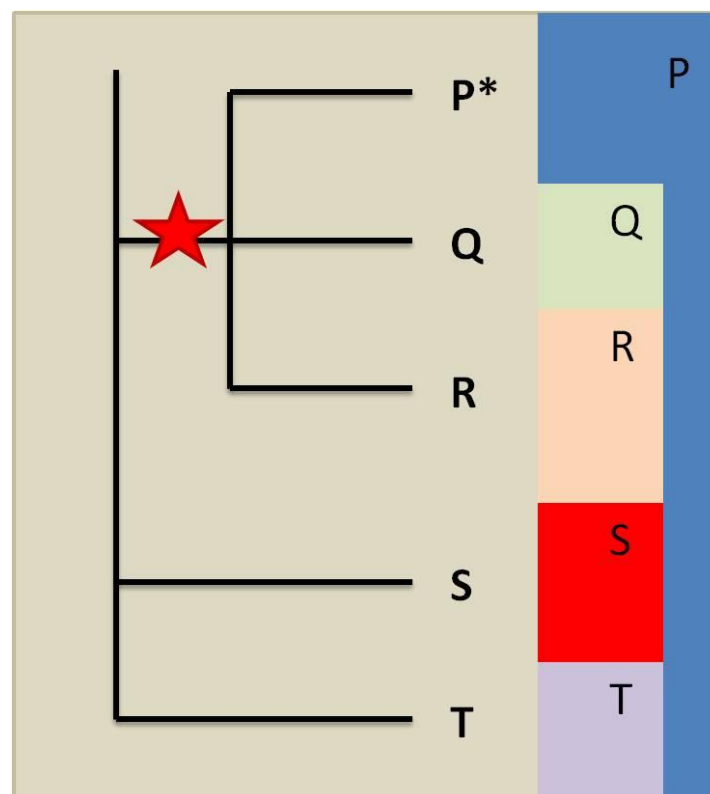


Figura 20: Localização do marcador 14 na árvore filogenética do cromossoma Y (★).

M – Variação do marcador M2 dentro do sub-haplogrupo R1b1b2

De forma a verificar se a variação encontrada dentro do sub-haplogrupo R1b1b2 é verdadeira e não um problema de amostra, uma vez que apenas a amostra 10 (Figura 15) é diferente de todas as outras, foram amplificadas por PCR novas amostras apresentando a mesma mutação daquela onde se detectou diferença, sendo esta mutação a S167 que define R1b1b2a1*. Após amplificação foi efectuada a separação electroforética do produto amplificado destas amostras cujo resultado se pode visualizar na Figura 19.

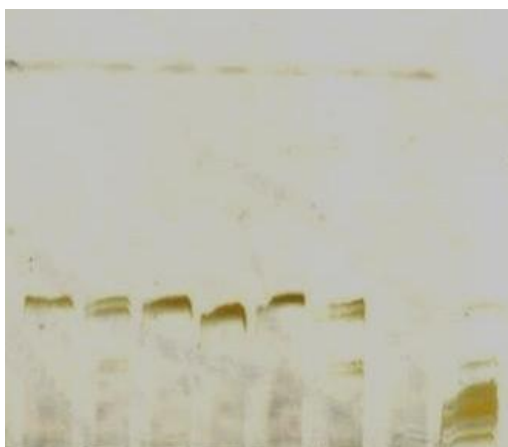


Figura 21: Gel electroforético T12C5 de amostras portadoras da mutação S167.

Analisando os resultados obtidos, confirma-se a existência de variação dentro do sub-haplogrupo R1b1b2 uma vez que, as amostras utilizadas possuem todas a mesma mutação, que define R1b1b2a1*, e observam-se os três padrões de bandas possíveis para este marcador.

Posto isto, coloca-se a possibilidade do InDel em estudo se localizar numa região onde ocorra recombinação intracromossómica com elevada frequência, sendo este fenómeno uma explicação plausível para a variação existente em R1b1b2 [33-34].

N – Determinação da localização do marcador M2 na região AZF do cromossoma Y

Após a detecção e confirmação da variação existente em R1b1b2 para o marcador M2 tornou-se importante a localização deste marcador no cromossoma Y, de forma a clarificar a razão da existência dos diferentes padrões de bandas neste sub-haplogrupo, mais especificamente, verificar se este facto pode dever-se à existência de recombinação intracromossómica na região onde se inclui [35].

Para a realização desta tarefa recorreu-se à base de dados dbSNP onde a partir do rs do marcador em causa se obtiveram a posição física e os genes que flanqueiam o marcador M2. A partir desta informação efectuou-se a representação esquemática da sua localização (Figura 22).

Após a análise da posição do marcador M2 verifica-se que este se localiza num dos vários palíndromes existentes no cromossoma Y, no palíndromo 3 do gene AZF [30, 32]. Estas regiões são conhecidas por serem regiões onde a recombinação intracromossómica é frequente e, por isso, a variação observada para R1b1b2 pode dever-se à ocorrência de recombinação após a mutação no marcador M2 [32].

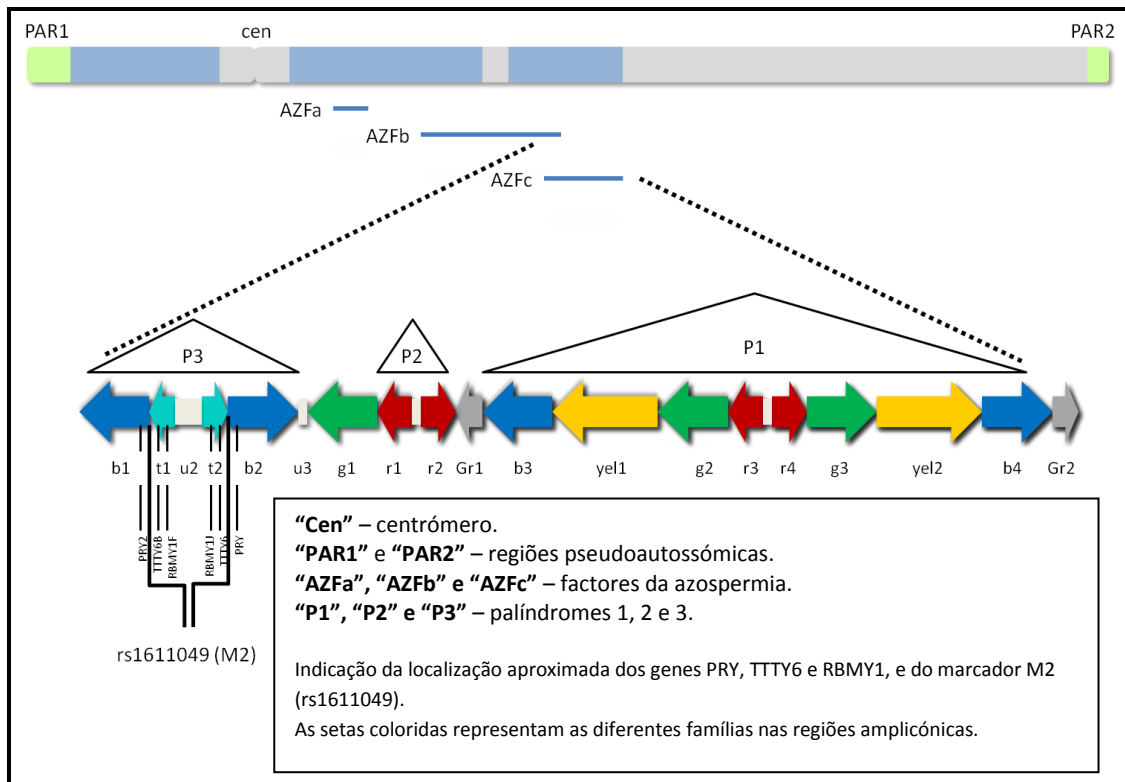


Figura 22: Representação esquemática da localização do marcador M2 (rs1611049) no cromossoma Y.

CONCLUSÃO

Durante a realização deste estudo verificou-se que os InDels presentes na árvore do consórcio do cromossoma Y não são suficientes para a criação de um multiplex de InDels que permita fazer a classificação de amostras de indivíduos do sexo masculino nos principais haplogrupos do cromossoma Y [1]. Isto acontece pois, embora na árvore que os descreve existam em número suficiente para que isso se tornasse possível, grande parte deles caracterizam sub-haplogrupos e não um dos haplogrupos principais, deixando de ser úteis para o objectivo proposto.

De forma a contornar este problema e cumprir o objectivo proposto, partiu-se para a pesquisa em bases de dados de outros InDels ainda não utilizados para distinguir haplogrupos [22-23]. No entanto, veio a verificar-se que os InDels já validados presentes nas bases de dados mais utilizadas não são na sua maioria polimorfismos específicos do cromossoma Y ou então, a sequência flanqueante gera muitos problemas à sua confirmação.

Assim, verificou-se que dos milhares de InDels descritos e presentes em bases de dados para o cromossoma Y, apenas 154 foram já validados, e, destes, apenas 2 se confirmaram ser verdadeiros polimorfismos.

Dos dois polimorfismos confirmados, o marcador M14 permite distinguir o parágrafo P o que se revela de grande utilidade uma vez que ainda nenhum InDel, dos descritos em Karafet *et al.* (2008), tinha sido localizado nesta região da árvore filogenética do cromossoma Y.

A confirmação e o estudo do marcador M2 levantaram questões quanto ao momento de ocorrência da duplicação descrita na região onde se localiza este marcador, a variação observada dentro do sub-haplogrupo R1b1b2 e a existência de um SNP ainda não descrito na zona flanqueante deste polimorfismo [30, 32]. No entanto, a confirmação da existência de variação em R1b1b2 não permitem que este seja um marcador a ser utilizado para caracterização de amostras nos diferentes haplogrupos.

Assim, após todas as triagens levadas a cabo nos InDels obtidos pela pesquisa em bases de dados o marcador M14 vem reforçar os InDels presentes em Karafet *et al.* (2008) não tornando, no entanto, possível a criação do multiplex proposto *a priori*.

Entretanto, perfilam-se como desenvolvimentos futuros: (a) a confirmação da presença/ausência de variação para o marcador M2 nos haplogrupos não estudados; (b) a presença de um alelo nulo, também em M2; (c) a confirmação da localização e estudo populacional de M14; (d) o estudo da diversidade haplotípica dentro dos alelos de M14; e, (e) o teste com amostras de haplogrupo desconhecido.

Após todas as triagens efectuadas nos InDels obtidos a partir das bases de dados podemos concluir que o desenvolvimento de um multiplex que possibilite a distinção de amostras nos principais haplogrupos do cromossoma Y só poderá realizar-se com a detecção e caracterização de novos InDels, sendo que o desenvolvimento e aperfeiçoamento da sequenciação de segunda geração dão um novo impulso à concretização deste objectivo pois, a sua capacidade de tornar a sequenciação mais rápida e económica aumenta as oportunidades de se efectuar a análise do cromossoma Y para indivíduos dos diferentes haplogrupos[36]. Esta análise tornará mais eficaz a detecção de InDels que distingam os vários haplogrupos e que, depois de validados contribuirão para a criação do multiplex pretendido neste trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Karafet, T.M., et al., *New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree*. Genome Research, 2008. **18**(5): p. 830-838.
2. Ali, S. and S.E. Hasnain, *Genomics of the human Y-chromosome: 1. Association with male infertility*. Gene, 2003. **321**: p. 25-37.
3. Li, Z., C.J. Haines, and Y. Han, *"Micro-deletions" of the human Y chromosome and their relationship with male infertility*. Journal of Genetics and Genomics, 2008. **35**(4): p. 193-199.
4. Waters, P.D., M.C. Wallis, and J.A.M. Graves, *Mammalian sex--Origin and evolution of the Y chromosome and SRY*. Seminars in Cell & Developmental Biology, 2007. **18**(3): p. 389-400.
5. Veyrunes, F., et al., *Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes*. Genome Research, 2008. **18**(6): p. 965-973.
6. Gusmão, L., et al., *Y chromosome specific polymorphisms in forensic analysis*. Legal Medicine, 1999. **1**(2): p. 55-60.
7. Quintana-Murci, L., C. Krausz, and K. McElreavey, *The human Y chromosome: function, evolution and disease*. Forensic Science International, 2001. **118**(2-3): p. 169-181.
8. Repping, S., et al., *High mutation rates have driven extensive structural polymorphism among human Y chromosomes*. Nat Genet, 2006. **38**(4): p. 463-467.
9. Skaletsky, H., et al., *The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes*. Nature, 2003. **423**(6942): p. 825-837.
10. Graves, J.A.M., *Sex Chromosome Specialization and Degeneration in Mammals*. Cell, 2006. **124**(5): p. 901-914.
11. Schwartz, A., et al., *Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination*. Hum. Mol. Genet., 1998. **7**(1): p. 1-11.
12. Mannucci, A., et al., *Forensic application of a rapid and quantitative DNA sex test by amplification of the X-Y homologous gene amelogenin*. International Journal of Legal Medicine, 1994. **106**(4): p. 190-193.
13. Noordam, M.J. and S. Repping, *The human Y chromosome: a masculine chromosome*. Current Opinion in Genetics & Development, 2006. **16**(3): p. 225-232.
14. Ruiz-Linares, A., *Y-chromosomes and Evolution*, in *International Encyclopedia of the Social & Behavioral Sciences*, J.S. Neil and B.B. Paul, Editors. 2001, Pergamon: Oxford. p. 16653-16657.
15. Goodwin, W., A. Linacre, and S. Hadi, *An Introduction to Forensic Genetics*. 2007: John Wiley & Sons, Ltd.
16. Butler, J.M., *Forensic DNA Typing - Biology, Technology, and Genetics of Str Markers*. 2005: Elsevier Academic Press.
17. Consortium, T.Y.C., *A Nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups*. Genome Research, 2002. **12**(2): p. 339-348.
18. Jobling, M.A. and C. Tyler-Smith, *The human Y chromosome: an evolutionary marker comes of age*. Nat Rev Genet, 2003. **4**(8): p. 598-612.
19. Vali, U., et al., *Insertion-deletion polymorphisms (indels) as genetic markers in natural populations*. BMC Genetics, 2008. **9**(1): p. 8.
20. Mills, R.E., et al., *An initial map of insertion and deletion (INDEL) variation in the human genome*. Genome Research, 2006. **16**(9): p. 1182-1190.
21. Pereira, R., et al., *A new multiplex for human identification using insertion/deletion polymorphisms*. ELECTROPHORESIS, 2009. **30**(21): p. 3682-3690.

22. NCBI. *Single Nucleotide Polymorphism database (dbSNP)*. Available from: <http://www.ncbi.nlm.nih.gov/snp/>.
23. Mammalian Genotyping Service - National Heart, L.a.B.I. *Diallelic Insertion/Deletion Polymorphisms Database*. Available from: <http://research.marshfieldclinic.org/genetics/home/index.asp>.
24. Cruz, U.o.C.S. *UCSC Genome Browser*. Available from: <http://genome.ucsc.edu/cgi-bin/hgGateway>.
25. NCBI. *BLAST SNP Sequence*. Available from: http://www.ncbi.nlm.nih.gov/SNP/snp_blastByOrg.cgi.
26. Cruz, U.o.C.S. *UCSC In-Silico PCR*. Available from: <http://genome.ucsc.edu/cgi-bin/hgPcr?command=start>.
27. Rozen, S., *Primer3*.
28. Budowle, B., et al., *Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE*. Am J Hum Genet, 1991. **48**: p. 137-144.
29. Drummond AJ, A.B., Buxton S, Cheung M, Cooper A, Heled J, Kearse M, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A *Geneious v4.8.4*. 2010.
30. Tessari, A., et al., *Characterization of HSFY, a novel AZFb gene on the Y chromosome with a possible role in human spermatogenesis*. Mol Hum Reprod, 2004. **10**(4): p. 253-8.
31. Park, M.J., et al., *Characterization of deletions in the DYS385 flanking region and null alleles associated with AZFc microdeletions in Koreans*. J Forensic Sci, 2008. **53**(2): p. 331-4.
32. Costa, P., et al., *Identification of new breakpoints in AZFb and AZFc*. Mol Hum Reprod, 2008. **14** (4): p. 251-258.
33. Adams, S.M., et al., *The case of the unreliable SNP: Recurrent back-mutation of Y-chromosomal marker P25 through gene conversion*. Forensic Science International, 2006. **159**(1): p. 14-20.
34. Rozen, S., et al., *Abundant gene conversion between arms of palindromes in human and ape Y chromosomes*. Nature, 2003. **423**(6942): p. 873-6.
35. Navarro-Costa, P., J. Goncalves, and C.E. Plancha, *The AZFc region of the Y chromosome: at the crossroads between genetic diversity and male infertility*. Hum Reprod Update, 2010. **16**(5): p. 525-42.
36. Morozova, O. and M.A. Marra, *Applications of next-generation sequencing technologies in functional genomics*. Genomics, 2008. **92**(5): p. 255-264.