



## DEPARTAMENTO DE CIÊNCIAS DA VIDA

FACULDADE DE CIÊNCIAS E TECNOLOGIA  
UNIVERSIDADE DE COIMBRA

Diversidade haplotípica associada a variantes no *locus* APOL1 que conferem resistência à doença do sono numa amostra populacional da Ilha do Príncipe.

---

Sérgio Amaro Antunes Teixeira

2012





## DEPARTAMENTO DE CIÊNCIAS DA VIDA

FACULDADE DE CIÊNCIAS E TECNOLOGIA  
UNIVERSIDADE DE COIMBRA

Diversidade haplotípica associada a variantes no *locus* APOL1 que conferem resistência à doença do sono numa amostra populacional da Ilha do Príncipe.

Dissertação apresentada à Universidade de Coimbra para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Evolução e Biologia Humanas, realizada sob a orientação científica do Professor Doutor Jorge Macedo Rocha (Universidade do Porto), Prof.<sup>a</sup> Doutora Manuela Alvarez (Universidade de Coimbra) e do Doutor Licínio Manco (Universidade de Coimbra).

---

Sérgio Amaro Antunes Teixeira

2012



## Resumo

A Apolipoproteína L1 (APOL1) é o factor tripanolítico do soro humano que impede que os parasitas do género *Trypanosoma* provoquem doença do sono quando infetam o Homem. No entanto, há pelo menos duas sub-espécies (*Trypanosoma brucei gambiense* e *Trypanosoma brucei rhodesiense*) que são capazes de anular a atividade tripanolítica de APOL1. Recentemente foram identificadas duas variantes africanas (G1 e G2) do gene da APOL1 que são capazes de restaurar a atividade tripanolítica da proteína e têm um papel importante na resistência à doença do sono. Estes alelos aumentam, porém, significativamente o risco de algumas formas graves de doença renal.

Com o objetivo de reconstruir a história evolutiva das variantes G1 e G2 foi desenvolvido um sistema de genotipagem de polimorfismos localizados em regiões vizinhas das mutações que originaram estas variantes, baseado na ressequenciação direta de um fragmento do gene da APOL1 e na genotipagem simultânea de cinco microssatélites flanqueantes. O sistema foi testado num estudo piloto em que se utilizaram amostras de habitantes da ilha do Príncipe, onde se registou no princípio do século XX uma epidemia de doença do sono com grande mortalidade.

As frequências das variantes G1 (0,20) e G2 (0,10) na ilha do Príncipe são das mais altas até agora observadas no continente africano, com exceção de algumas regiões do Gana e da Nigéria. A variante G1, em particular, é especialmente frequente (0,30) nos indivíduos com os quatro avós nascidos na ilha e que provavelmente descendem de sobreviventes da epidemia. No entanto, não é possível saber para já até que ponto esta frequência mais elevada resultou da maior resistência à infecção, ou reflete a composição étnica dos povoadores originais que ocuparam a ilha antes das levas de imigração de origem cabo-verdiana que se sucederam à vaga de doença do sono.

A análise da diversidade intra-alélica mostrou que a variante G1 tem uma idade aproximada compreendida entre 5600 a 1624 anos, compatível com uma origem durante a difusão da agricultura tropical em África e consistente com a hipótese de que esta variante foi favorecida devido à resistência que confere contra a doença do sono. A variante G2, apesar de menos frequente que G1, tem níveis mais elevados de diversidade intra-alélica compatíveis com uma idade mais antiga de

14784 a 4648 anos, sendo menos provável que a seleção natural tenha tido um papel relevante na sua difusão, se se assumir que esta mutação resultou de um único evento mutacional. No entanto, a multimodalidade dos perfis de diversidade haplotípica de G2 associada ao facto de esta variante resultar de uma deleção, sugerem que é possível que o alelo G2 pode ter tido origens múltiplas, relativamente recentes, que podem estar a ser confundidas com uma origem única antiga.

O sistema de análise da diversidade haplotípica associada às variantes de APOL1 agora desenvolvido poderá ser facilmente utilizado em futuras análises de amostras representativas de todo o continente africano com vista ao esclarecimento da história evolutiva de G1 e G2, incluindo a origem e principais rotas de difusão destas variantes.

***Palavras-Chave:*** APOL1 • Resistência à doença do sono • Ilha do Príncipe • Diversidade haplotípica.

## **Abstract**

Apolipoprotein L1 (APOL1) is the human serum's trypanolytic factor, which impairs parasites from the genus *Trypanosoma* to cause human sleeping sickness upon infection. However, there are at least two subspecies (*Trypanosoma brucei gambiense* and *Trypanosoma brucei rhodesiense*) that are able to overcome APOL1's trypanolytic activity. Recently, two APOL1 African variants (G1 and G2) have been identified that are able to restore Apol1 trypanolytic activity and that play a key role in sleeping sickness disease resistance. These variants, however, increase significantly the risk of some severe forms of kidney disease.

With the objective of reconstructing G1 and G2's evolutionary history we developed a genotyping system for neighboring polymorphisms, based on direct resequencing of a APOL1 gene fragment and on simultaneous genotyping of five flanking microsatellites. The system was tested in a pilot study involving samples of Island of Principe where, at the beginning of the 20<sup>th</sup> century, a high mortality sleeping sickness epidemic was recorded.

G1 and G2's frequencies, 0, 2 and 0, 1 respectively, found on Principe's island are among the highest observed so far in the African continent, with the exception of some Ghana and Nigeria regions. The G1 variant, in particular, is especially frequent (0,3) in individuals whose four grandparents were born on the island and probably survived the epidemic. However, it is not possible to know, for now, to which extent this higher frequency results from an increased resistance to the disease, or reflects the ethnic composition of the original settlers that occupied the island before the waves of Cape Verdean immigration that followed the sleeping sickness onslaught.

The intra-allelic diversity analysis showed that the G1 variant has an approximate age between 5600 and 1624 years, compatible with an origin during the diffusion of tropical agriculture in Africa, and consistent with the hypotheses that this variant was favored due to its role in the resistance to sleeping sickness. The G2 variant, although less frequent than G1, displayed higher levels of intra-allelic diversity compatible with an older age between 14784 and 4648 years, but less compatible with selection if it is assumed that this variant resulted from a single mutational event. However, the multimodality of G2 haplotype diversity profiles associated with the fact that this variant results from a deletion, suggests that it is

possible that the G2 allele might have had relatively recent multiple origins, that might be wrongly interpreted as a single ancient origin.

The system for screening the haplotype diversity associated with the APOL1 variants here developed can be easily used in future analysis of samples that are representative of the whole African continent with the objective of clarifying G1 and G2's evolutionary history, including their origin and main diffusion routes.

**Keywords:** APOL1 · Sleeping sickness resistance · Principe's island · Haplotype diversity.

## **Agradecimentos**

*Este trabalho não teria sido uma possibilidade sem o apoio e ajuda incontornáveis das seguintes pessoas às quais não podia deixar de agradecer:*

O meu especial agradecimento ao orientador desta tese o Prof. Jorge Rocha por me ter aberto à possibilidade de estudar em Genética Evolutiva Humana. A sua contribuição sob a forma dos seus ensinamentos, conhecimento, interesse, rigor e ajuda na execução do trabalho, foi decisiva na escolha do tema e em todos aspetos desta etapa fulcral do mestrado.

Um agradecimento à Prof.<sup>a</sup> Manuela Alvarez pelo seu interesse neste trabalho e ter aceitado ser minha coorientadora.

Um especial agradecimento vai também ao Prof. Licínio Manco que abraçou imediatamente participar neste projeto. As suas sugestões e os seus contributos, sempre oportunos e pertinentes foram muito importantes para este trabalho.

Gostaria de agradecer ao Prof. Nuno Ferrand que prontamente disponibilizou o excelente aparato laboratorial do CTM (Centro de Testagem Molecular) bem como o seu altamente qualificado pessoal técnico.

Um muito obrigado à Susana Lopes do CTM por toda a ajuda e entusiasmo com que endereçou este meu trabalho. Também no CTM, a minha segunda casa neste período, gostaria de agradecer à Sandra Afonso, Sara João, Patrícia Henriques, Sofia Silva, Sandra Reis, Diana Castro, Raquel Godinho, e todos que de alguma forma foram solícitos quando mais precisei.

Uma palavra especial aos meus pais e irmã. Vocês são um perene abraço.

À Ana, a tua paz e compreensão são dádivas que carrego próximo.



## Índice geral

1. Introdução.....	1
1.1. Variação genética e suscetibilidade a doenças infecciosas .....	1
1.2. Variação genética no <i>locus</i> da apolipoproteína L1 (APOL1).....	2
1.2.1 Implicações para a doença renal .....	2
1.2.2 Função Tripanolítica das variantes do APOL1 .....	5
2. Material e Métodos.....	9
2.1. Amostras.....	9
2.2. Detecção de variação genética no gene APOL1 e regiões flanqueantes.....	10
2.2.1 Sequenciação direta do <i>locus</i> APOL1 .....	10
2.2.2 Variação genética em STRs flanqueantes .....	12
2.3. Inferência de haplótipos e outras análises estatísticas.....	13
3. Resultados e Discussão.....	14
3.1. Genotipagem da variação no <i>locus</i> APOL1 e nas regiões adjacentes.....	14
3.2. Variação genética no <i>locus</i> APOL1 e nas regiões adjacentes .....	17
3.2.1 Frequência da variante G1 .....	17
3.2.2 Frequência da variante G2.....	21
3.2.3 Variação haplotípica associada às variantes G1 e G2 .....	23
3.2.4 Idade das variantes variante G1 e G2 .....	26
3.3. Conclusões.....	31
4. Referências bibliográficas .....	33

## Índice de Figuras

Figura 1. Representação esquemática do gene APOL1 (incluindo intrões) com a posição das variantes G1 (amarelo) e G2 (verde). Distância física em kb (kilobases). .....	3
Figura 2. a) Distribuição das frequências do alelo G1. b) Distribuição das frequências relativas do alelo G2. c) Distribuição da frequência dos 3 genótipos de risco G1G1, G2G2 e G1G2 combinados. As cores mais escuras indicam frequências mais elevadas. Retirado de Rosset <i>et al.</i> (2011). .....	4
Figura 3. Ilustração esquemática descrevendo a associação entre as variantes do APOL1, que constituem o alelo G1, e a atividade tripanolítica da apolipoproteína L1 por um lado, e aumento do risco para insuficiência renal não-diabética por outro. Adaptado de Kronenberg (2011). .....	6
Figura 4. Estrutura dos domínios da apolipoproteína L1 em que se destaca a localização das variantes G1 e G2 no domínio de interação com a SRA. Adaptado de Zenker e Mertens (2010). .....	6
Figura 5. Ilha do Príncipe e sua localização geográfica na costa africana. In Maurer (2009). .....	9
Figura 6. Representação esquemática do gene APOL1 (incluindo intrões) com a posição das variantes G1 (vermelho) e G2 (verde), os STRs identificados e região sequenciada. .....	11
Figura 7. Resultados da sequenciação para os <i>loci</i> ( <i>rs73885319</i> e <i>rs60910145</i> ) que definem a G1. Setas indicam posições de interesse. a,b) Indivíduo homocigótico normal (A;T). c,d) Indivíduo heterocigótico (A/G; T/G.) e,f) Indivíduo homocigótico para as variantes (G;G). .....	14
Figura 8. Resultados da sequenciação para a G2. Setas verticais indicam as posições de interesse na sequência. A variante G2 que consiste na deleção de 6 pares de bases numa região delimitada por 2 sequências repetitivas ATAA, identificadas na figura. A deleção de 6 nucleótidos entre estes dois elementos repetitivos conduz à perda na sequência proteica do Apol1 dos 2 aminoácidos Asn388-Tyr389. g) Indivíduo homocigótico normal sem deleção, em que são mostradas as 2 sequências repetitivas ATAA (chavetas); h) indivíduo heterocigótico para a deleção (sequenciação com <i>primer</i> reverso) em que é mostrada a sequência	

(complementar-reversa) com desfasamento até ao início da segunda sequência repetitiva ATAA i) Indivíduo homozigótico com a deleção em que é mostrada a sequência repetitiva ATAA que permanece após a deleção. ....	15
Figura 9. Resultados da reação de multiplex de um único indivíduo para os cinco STRs genotipados.....	15
Figura 10. Distribuição das frequências dos haplótipos associados aos cromossomas W (A), G1 (B) e G2 (C). ....	24
Figura 11. Homogeneidade haplotípica associada aos cromossomas W, G1 e G2. Em abcissas indica-se o número do polimorfismo flanqueante de acordo com a Tabela 2. Em ordenadas indica-se, para cada polimorfismo, a percentagem de haplótipos que mantêm a configuração Em ordenadas indica-se, para cada polimorfismo, a percentagem de haplótipos que mantêm a configuração haplotípica mais comum de cada tipo de cromossoma (W, G1, G2) desde o <i>locus</i> de referência até esse polimorfismo. A linha tracejada destaca o polimorfismo de referência pelo <i>locus</i> rs60910145. ....	27
Figura 12. Curvas de verosimilhança da idade das variantes G1 (A e B) e G2 (C e D) assumindo taxas de mutação dos microssatélites adjacentes de $\mu=0,0001$ /geração (A e C) e $\mu=0,001$ /geração (B e D). ....	28

## Índice de tabelas

Tabela 1. Características dos STRs utilizados para analisar a variação haplotípica associada às variantes G1 e G2 do gene da APOL1.....	12
Tabela 2. Frequências alélicas e níveis de heterozigotia calculados em 154 habitantes da ilha do Príncipe para os 19 polimorfismos utilizados para caracterizar a diversidade haplotípica no <i>locus</i> APOL1.....	18
Tabela 3. Distribuição dos genótipos associados à variante G1 e valores esperados de acordo com o formalismo de <i>Hardy-Weinberg</i> em 154 indivíduos residentes na ilha do Príncipe.....	19
Tabela 4. Distribuição dos genótipos associados à variante G1 e valores esperados de acordo com o formalismo de <i>Hardy-Weinberg</i> em 33 indivíduos residentes no Príncipe com os quatro avós nascidos na ilha. ....	20
Tabela 5. Distribuição dos genótipos associados à variante G2 e valores esperados de acordo com o formalismo de <i>Hardy-Weinberg</i> em 154 indivíduos residentes na ilha do Príncipe.....	21
Tabela 6. Distribuição dos genótipos associados à variante G2 e valores esperados de acordo com o formalismo de <i>Hardy-Weinberg</i> em 33 indivíduos residentes no Príncipe com os quatro avós nascidos na ilha. ....	22
Tabela 7. Estatísticas sumário da diversidade haplotípica no gene APOL1.....	23
Tabela 8. Estimativas da idade das variantes G1 e G2 do <i>locus</i> APOL1.....	29

# 1. Introdução

## 1.1. Variação genética e suscetibilidade a doenças infecciosas

A Genética Evolutiva Humana é o estudo de como um genoma humano difere de um outro, e as implicações deste facto para a compreensão da nossa espécie no que concerne ao seu passado e à sua atualidade (Jobling *et al.*, 2004). Sendo inexorável a íntima ligação do nosso passado evolutivo com os padrões de diversidade genética atual, pois esta poderá ser reminescente de variadíssimos processos demográficos, mutacionais e/ou seletivos (Goldstein e Chikhi, 2002; Barnes *et al.*, 2011). Um exemplo fascinante e contundente do impacto da nossa história evolutiva pode ser encontrado no estabelecimento de uma base genética explicativa da variação interindividual na suscetibilidade a doenças infecciosas humanas. Sendo incontornável neste aspecto o caso providenciado pela malária.

A manutenção de hemoglobinopatias, como anemia falciforme e a beta-talassemia, em taxas relativamente altas em populações humanas com sobreposição geográfica ao chamado cinturão da malária levou, há cerca de 60 anos, J.B.S. Haldane a postular que estas doenças hemáticas quando decorrentes de um genótipo heterozigótico poderiam proteger um indivíduo de uma infecção malárica que, de outra forma seria mortal (Cooke e Hill, 2001). Este postulado foi subsequentemente validado de uma forma empírica como sendo um dos poucos exemplos onde um agente infeccioso corporiza um veículo de seleção modelando, assim, um dos aspectos da variabilidade genética exibida pelas populações humanas atuais (Baker e Antonovics, 2012).

É nesta moldura conceptual e empírica que o presente trabalho é proposto, desta feita com um intrigante e recente avanço na área das doenças renais onde variantes genéticas no *locus* APOL1 aumentam o risco de insuficiência renal crónica não-diabética e insuficiência renal terminal, especialmente em populações com ancestralidade africana, ao mesmo tempo que protegem contra a doença do sono.

## 1.2. Variação genética no *locus* da apolipoproteína L1 (APOL1)

### 1.2.1 Implicações para a doença renal

A observação de que indivíduos afro-americanos têm maior susceptibilidade à insuficiência renal crónica (IRC) não-diabética e insuficiência renal terminal (IRT) do que Euro-americanos (Rosset *et al.*, 2011), motivou vários investigadores a investigar os fatores causais destas diferenças. Entre estes incluem-se não só fatores socioeconómicos e de estilo de vida, mas também, como recentemente demonstrado (Kronenberg, 2011), fatores de índole genética. Recorrendo a técnicas de mapeamento por miscigenação (MALD)<sup>1</sup>, foi inicialmente detetada uma forte associação entre variantes do gene da cadeia 9 da miosina não muscular (*nonmuscle myosin heavy chain 9*, MYH9) no cromossoma 22q12 e a nefropatia não-diabética em afro-americanos (Kao *et al.*, 2008; Kopp *et al.*, 2008). No entanto, apesar da evidência de expressão renal deste gene e de um papel em doenças autossómicas dominantes associadas com glomeruloesclerose progressiva, não foram encontrados polimorfismos que pudessem estar relacionados com a nefropatia não-diabética (Freedman *et al.*, 2010). Este facto conduziu a que dois grupos de investigação, de forma independente, concluíssem que o gene MYH9 não era responsável pelo risco observado, tendo-se antes focado no *locus* adjacente APOL1, no qual foram encontradas duas variantes associadas quer à IRC quer à IRT (Genovese *et al.*, 2010; Tzur *et al.*, 2010). Genovese *et al.* (2010) e Tzur *et al.* (2010) usaram dados do projeto 1000 Genomes (<http://www.1000genomes.org/>) para procurar polimorfismos nucleotídicos (SNPs, *Single Nucleotide Polymorphisms*) biologicamente relevantes que pudessem explicar a diferença de risco de insuficiência renal entre americanos de origem europeia e africana. Esta investigação levou à descoberta de duas variantes derivadas não-sinónimas, rs73885319 e rs60910145, conducentes às substituições

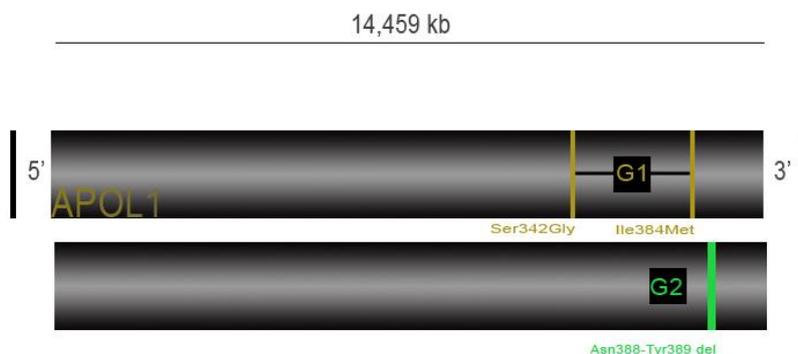
---

<sup>1</sup>MALD - *Mapping by Admixture Linkage Disequilibrium*.

aminoacídicas Ser342Gly e Ile384Met, respetivamente, no último exão do gene APOL1. Estas variantes encontram-se em perfeito desequilíbrio gamético (LD)<sup>2</sup> e formam um haplótipo denominado G1 pelo grupo de Genovese (Figura1). Além disto, Genovese *et al.* (2010) também observaram uma associação com outra variante do gene APOL1, rs7178513, caracterizada por uma deleção de seis pares de bases (*base pairs*), designada G2 (alteração proteica Asn388-Tyr389 del), de menor frequência do que G1, que nunca coocorre com as mutações definidoras do alelo G1 no mesmo cromossoma parental (Figura 1).

Os dois estudos propuseram que ambos os alelos, G1 e G2, tinham um efeito recessivo no risco de IRC e IRT.

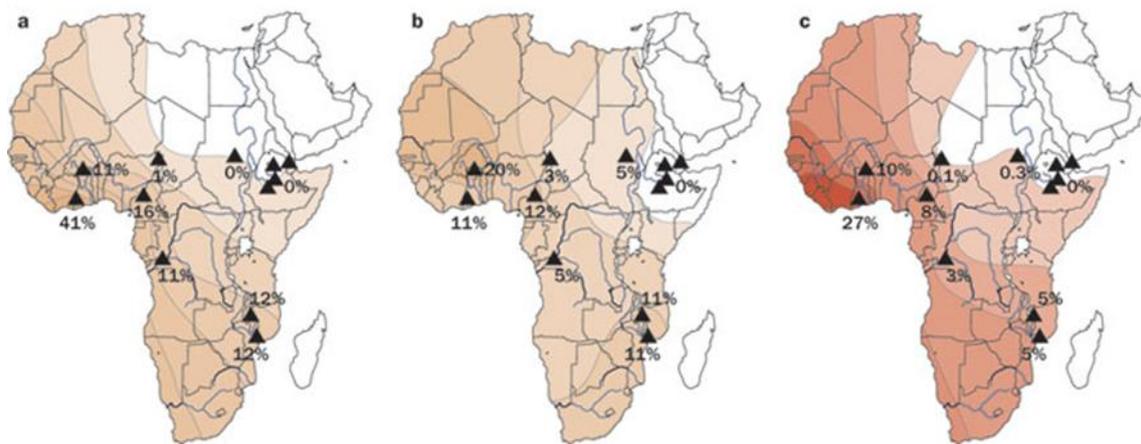
As variantes G1 e G2 foram encontradas a uma frequência conjunta de 46 a 52% nos Yoruba, um grupo étnico da África Ocidental (Genovese *et al.*, 2010; Rosset *et al.*, 2011; Friedman and Pollak, 2011), mas não em populações Europeias, Japonesas ou Chinesas amostradas no Projeto HapMap (<http://hapmap.ncbi.nlm.nih.gov/>) (Figura 2).



**Figura 1.** Representação esquemática do gene APOL1 (incluindo intrões) com a posição das variantes G1 (amarelo) e G2 (verde). Distância física em kb (kilobases).

<sup>2</sup>LD - *Linkage disequilibrium*.

Este facto levantou a possibilidade de estas variantes terem sido alvo de seleção exclusivamente no continente africano (Genovese *et al.*, 2010). Esta observação acabou por ficar intimamente relacionada com uma intrigante função adicional do gene APOL1.



**Figura 2.** a) Distribuição das frequências do alelo G1. b) Distribuição das frequências relativas do alelo G2. c) Distribuição da frequência dos 3 genótipos de risco G1G1, G2G2 e G1G2 combinados. As cores mais escuras indicam frequências mais elevadas. Retirado de Rosset *et al.* (2011).

### 1.2.2 Função Tripanolítica das variantes do APOL1

A apolipoproteína L1 é normalmente encontrada nas partículas lipoproteicas de alta densidade (HDL)<sup>3</sup> e é-lhe atribuído um papel na lise de tripanossomas (Pays *et al.*, 2006). O complexo HDL-APOL1 é conhecido como factor lítico de tripanossomas (TLF)<sup>4</sup>, um mecanismo de imunidade inata que evoluiu no Homem e em outros primatas (Pays *et al.*, 2006; Namangala, 2011).

Os tripanossomas Africanos, dos quais a forma arquetípica é o *Trypanosoma brucei brucei* (*T. b. brucei*), são transmitidos pela mosca *tsetse* (*Glossina* sp.), sendo a espécie humana naturalmente resistente a várias espécies de tripanossomas (Pays *et al.*, 2006; Namangala, 2011). Não obstante, duas subespécies de *T. b. brucei*, *Trypanosoma brucei gambiense* (*T. b. gambiense*) e *Trypanosoma brucei rhodesiense* (*T. b. rhodesiense*) desenvolveram a capacidade para infectar a nossa espécie por mecanismos diferentes, e tornaram-se os agentes causadores da doença do sono ou Tripanossomíase Humana Africana (THA) (Kioy *et al.*, 2004; Pays *et al.*, 2006; Cortes, 2008; Namangala, 2011).

O *T. b. gambiense* causa infeção crónica, enquanto *T. b. rhodesiense*, geralmente implica uma infeção mais aguda (Kioy *et al.*, 2004; Cortes, 2008). O *T. b. gambiense* é encontrado no oeste, centro e algumas partes do leste de África, enquanto *T. b. rhodesiense* predomina no sul e no leste. Em ambos os casos, os parasitas desenvolvem-se primeiramente na linfa e órgãos periféricos (estádio 1), atingindo depois o sistema nervoso central (estádio 2), onde causam graves complicações neurológicas; sem tratamento a doença é fatal (Kioy *et al.*, 2004).

Embora se saiba que *T. b. rhodesiense* inibe a atividade lítica do alelo normal ancestral da APOL1 produzindo uma proteína associada à resistência ao soro humano (SRA)<sup>5</sup>, o mecanismo utilizado pelo *T. b. gambiense* para suprimir a resistência humana continua desconhecido (Namangala, 2011).

Genovese *et al.* (2010) mostraram que as variantes G1 e G2 no *locus* Apol1 são capazes de restaurar a resistência humana a *T. b. rhodesiense*, provavelmente

---

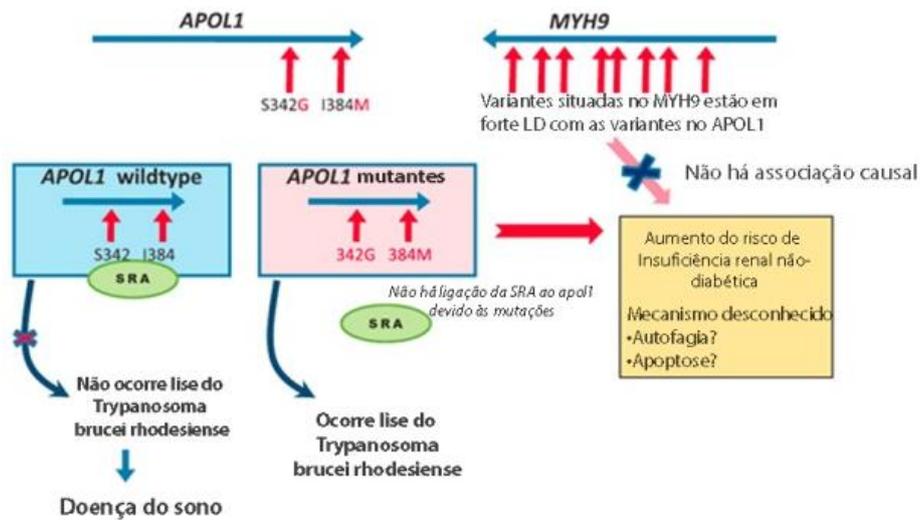
<sup>3</sup>HDL - *High Density Lipoproteins*.

<sup>4</sup>TLF- *Trypanosome Lytic Factor*.

<sup>5</sup>SRA -*Human Serum Resistant-associated Protein*.

por interferirem com a ligação da SRA à apolipoproteína L1 (Figuras 3 e 4). No entanto, estas mutações são incapazes de restaurar a atividade lítica contra *T. b. gambiense*, que tem um mecanismo de evasão ao sistema imunológico do hospedeiro que não envolve a SRA (Pays *et al.*, 2006; Namangala, 2011).

Uma vez que a capacidade de restaurar a atividade lítica contra *T. b. rhodesiense* é, possivelmente, uma vantagem adaptativa em áreas onde a doença do sono é endémica



**Figura 3.** Ilustração esquemática descrevendo a associação entre as variantes do APOL1, que constituem o alelo G1, e a atividade tripanolítica da apolipoproteína L1 por um lado, e aumento do risco para insuficiência renal não-diabética por outro. Adaptado de Kronenberg (2011).



**Figura 4.** Estrutura dos domínios da apolipoproteína L1 em que se destaca a localização das variantes G1 e G2 no domínio de interação com a SRA. Adaptado de Zenker e Mertens (2010).

(ou ao longo de surtos epidémicos), é provável que a alta frequência dos alelos de risco do APOL1 para a doença renal em África seja o preço a pagar pela proteção contra a doença do sono (Rosset *et al.*, 2011).

De facto, Genovese *et al.* (2010) demonstraram que as variantes G1 e G2 poderiam ter sido alvos de seleção positiva há relativamente pouco tempo.

Neste contexto, a associação de variantes do gene MYH9 anteriormente detetada terá resultado da sua ligação aos variantes de G1 e G2 que foram selecionados.

Apesar do notável corpo de dados acumulado até à data, a distribuição geográfica dos alelos G1 e G2 ainda não está suficientemente caracterizada e há questões relacionadas com a origem, evolução e dispersão destas variantes vantajosas do APOL1 que continuam por esclarecer. Por exemplo, existem várias áreas em África que continuam por amostrar (Figura 2) e a distribuição detalhada dos alelos G1 e G2 no continente africano não está ainda estabelecida. É também difícil compreender porque é que a alta frequência de variantes tripanolíticas é sobretudo encontrada em populações da África Ocidental, como os Yoruba, enquanto que o *T. b. rhodesiense*, em relação ao qual as mutações discutidas conferem resistência, é essencialmente encontrado a oriente.

Além disto, a origem geográfica dos alelos G1 e G2, e os contributos da seleção e da migração para a sua difusão em África ainda continuam por clarificar.

Uma das principais lacunas relacionadas com o estudo da história evolutiva destes alelos é a falta de informação sobre os níveis de diversidade que lhes estão associados nas regiões cromossómicas adjacentes. Esta diversidade intra-alélica é, mais do que a frequência das variantes, uma das características essenciais para inferir o local de origem das mutações e a sua idade (Slatkin e Rannala, 2000).

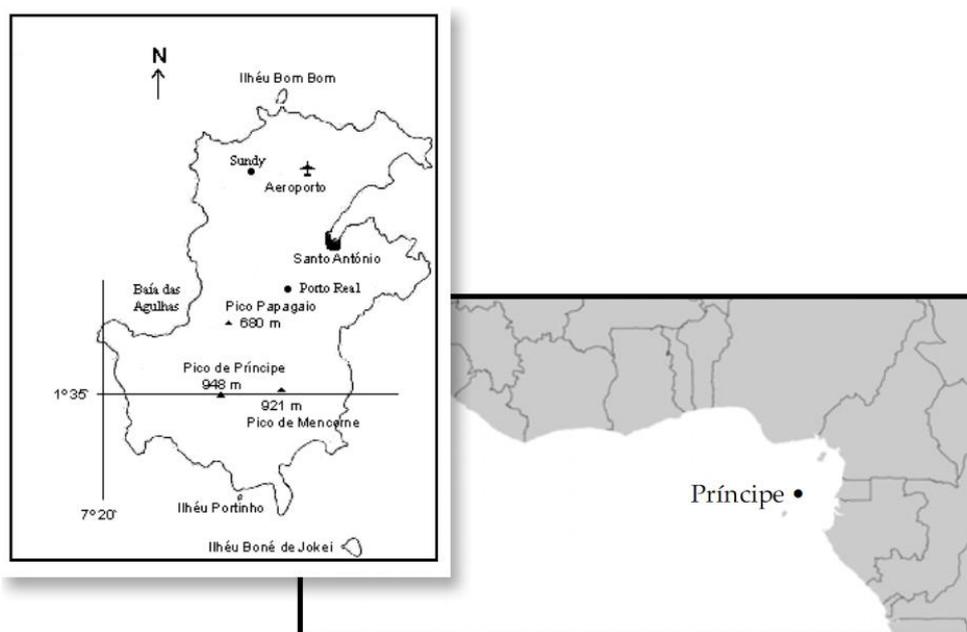
O principal objetivo deste trabalho é contribuir para o estudo da história evolutiva das variantes G1 e G2 do *locus* APOL1 através do desenvolvimento de um sistema de genotipagem que permita documentar eficazmente a variação haplotípica associada a estas variantes em populações africanas. A validade deste sistema foi verificada com um estudo piloto com amostras da ilha do Príncipe, cuja colonização com escravos e trabalhadores contratados originários de vários pontos da costa africana permite avaliar numa única população a diversidade

intra-alélica que se acumulou em diferentes áreas do continente. Uma vantagem adicional do estudo da ilha do Príncipe é a ocorrência de uma epidemia documentada de doença de sono entre 1900 e 1901, com elevada mortalidade (Costa, 1954; Günther, 1973 *in* Maurer, 2009), cujos sobreviventes são atualmente uma escassa minoria da população da ilha mas nos quais a frequência dos variantes de resistência G1 e G2 poderá ser invulgarmente elevada.

## 2. Material e Métodos

### 2.1. Amostras

A amostra analisada inclui 154 indivíduos não aparentados, residentes na ilha do Príncipe (Figura 5). O DNA foi extraído com o *kit QIAmp® DNA Micro* (Qiagen®) a partir de esfregaços bucais conservados em etanol (96%), com ligeiras modificações às instruções do fabricante de forma a rentabilizar o processo de extração. Para garantir quantidades suficientes de material biológico em futuras utilizações, o DNA extraído em cada dador foi submetido a uma amplificação genómica completa (*Whole Genome Amplification, WGA*) com o *kit GenomiPhi™ v2 DNA Amplification Kit* (GE Healthcare Illustra™).



**Figura 5.** Ilha do Príncipe e sua localização geográfica na costa africana. *In* Maurer (2009).

## 2.2. Detecção de variação genética no gene APOL1 e regiões flanqueantes

Para identificar as variantes G1 e G2 e documentar a variação genética a elas associada optou-se por sequenciar diretamente um fragmento que englobasse o exão 6 do gene APOL1 - onde se localizam os SNPs característicos dessas variantes e pudesse captar variação de sequência nas regiões vizinhas. Adicionalmente, foram ainda genotipados 5 microssatélites (*short tandem repeats*; STRs) localizados em posições flanqueantes do fragmento sequenciado.

### 2.2.1 Sequenciação direta do *locus* APOL1

O fragmento-alvo usado na sequenciação direta (Figura 6), com um tamanho de 920 bp, foi amplificado com *primers* selecionados com o programa *Primer3Plus* (<http://www.bioinformatics.nl/cgi-bin/Primer3plus/Primer3plus.cgi>; Untergasser *et al.*, (2007)), usando a sequência do transcrito de referência NG\_023228.1 ([http://www.ncbi.nlm.nih.gov/nuccore/NG\\_023228.1](http://www.ncbi.nlm.nih.gov/nuccore/NG_023228.1)).

Em cada PCR usou-se um volume final de reação de 10 µL, com 5 µL de *Taq PCR Master Mix Kit* (QIAGEN®, 5 u/µl), 0,4 µL de *primer* 5' (5' GGC AGA TGA GCT CCG TAA AG 3'; 10pM), 0,4 µL de *Primer* 3' (5' CCC TGC CAG GCA TAT CTC 3'; 10pM), 3,2 µL de água de desionizada e 1 µL de DNA (10-15 ng/ µL). O protocolo de PCR (*polymerase chain reaction*) consistiu num passo inicial de desnaturação a 95 °C (15 min), ao qual se sucederam 9 ciclos de desnaturação a 95 °C (30 seg.), temperaturas de hibridização com passos decrescentes de 0,5 °C entre 63 e 59 °C, por um período de 45 seg. e, finalmente, uma extensão a 72 °C por 1 min.. O resto da amplificação consistiu em 21 ciclos com desnaturação a 95 °C (30 seg.), hibridização a 59 °C (45 seg.) e extensão a 72 °C (1 min.). No fim foi feita ainda uma extensão a 60 °C durante 30 min..

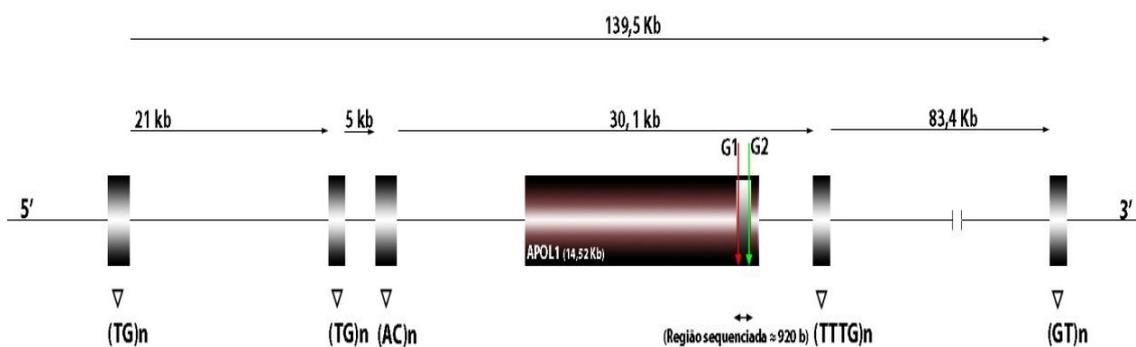
Após PCR, foi adicionado a cada tubo 1 µL de EXO-SAP (96 µL (1u/µL) Exonuclease I (EXO I) + 24 µL (20u/µL) *FastAP™ Thermosensitive Alkaline*

*Phosphatase* (Thermo Scientific Fermentas) a 37°C durante 15 min., ao fim dos quais a reação enzimática foi parada por inativação a 85°C durante 15 min..

A sequenciação do DNA foi feita com o método de Sanger (Sanger *et al.*, 1977) usando o kit *BigDye® Terminator v3.1 Cycle Sequencing* (Applied Biosystems®) nas duas cadeias a fim de assegurar que cada posição tivesse duas leituras. Para cada amostra foi preparado um volume de sequenciação de 10 µL contendo 7 µL de água desionizada, 1 µL de 5x *Sequencing Buffer* (Applied Biosystems®), 0,5 µL de *BigDye® Terminator v3.1* (Applied Biosystems®), 0,5 µL de *primer* 5' (ou 3') (a 10pM) e 1 µL de DNA. Nas reações de sequenciação fez-se uma primeira desnaturação a 94 °C seguida de 25 ciclos, cada um a 96 °C (10 seg.), 55 °C (5 seg.) e 60 °C (4 min.).

No fim das reações o excesso de reagentes foi removido por filtragem em colunas de *Sephadex* e o DNA foi analisado por eletroforese capilar no sequenciador automático *ABI 3130xl Genetic Analyzer* (Applied Biosystems®).

As seqüências obtidas foram analisadas com a aplicação *Seqscape v. 2.5* (Applied Biosystems®), e comparadas com a seqüência consenso NG\_023228.1.



**Figura 6.** Representação esquemática do gene APOL1 (incluindo intrões) com a posição das variantes G1 (vermelho) e G2 (verde), os STRs identificados e região sequenciada.

## 2.2.2 Variação genética em STRs flanqueantes

Com o objectivo de aumentar a resolução da variação haplotípica associada às variantes G1 e G2, foi feita uma pesquisa de sequências repetitivas potencialmente polimórficas nas regiões flanqueantes do gene APOL1, recorrendo à interface gráfica

UCSC Genome Browser (<http://genome.ucsc.edu/>). Com esta pesquisa foi possível identificar cinco STRs situados a diferentes distâncias de APOL1 (Figura 6).

Na tabela 1 apresentam-se as características dos cinco STRs flanqueantes e as sequências dos *primers* das reações de PCR usadas na sua genotipagem.

**Tabela 1.** Características dos STRs utilizados para analisar a variação haplotípica associada às variantes G1 e G2 do gene da APOL1.

	Repetição <sup>1</sup>	Distância a rs73885319 (kb) <sup>2</sup>	<i>primers</i> <sup>3</sup>	Tamanho médio do amplificação (bp) <sup>4</sup>
<b>STR1</b>	(TG) <sub>23</sub>	-51,05	5': 5'TGG AGC CCT TTA AGG AA3' 3':5'GAG GGT GCT GGC TGT T3'	225
<b>STR2</b>	(TG) <sub>17</sub>	-30,06	5': 5'CAC CGG GAG GAA GAA AG'3' 3': 5'CTT GGT TAT CCC CAC CTG'3'	103
<b>STR3</b>	(AC) <sub>13</sub>	-25,17	5': 5' CAT CAG CCA ACA AGT GG3' 3':5'TGG TGT CCA ACT TCA TCC3'	158
<b>STR4</b>	(TTTG) <sub>7</sub>	+4,21	5':5'TGC ACT GAT GTT TGT TGG'3' 3':5'TTC AGG AGA CCC ATC TCA 3'	204
<b>STR5</b>	(GT) <sub>15</sub>	+87,57	5': 5'CAG AGC TGC TGT TTC CAC 3' 3': 5'TTT AAC CCA CAT GCA TCC 3'	128

<sup>1</sup>Número de repetições da sequência de referência APOL1-001 ENST00000397278

([http://www.ensembl.org/Homo\\_sapiens/Search/Details?species=Homo\\_sapiens;idx=Gene;end=1;q=apol1](http://www.ensembl.org/Homo_sapiens/Search/Details?species=Homo_sapiens;idx=Gene;end=1;q=apol1)).

<sup>2</sup>Distância a um dos SNPs que define a variante G1 e que é aqui tomado como referência.

<sup>3</sup> Primers das reações de PCR usadas para genotipar os STRs.

<sup>4</sup> Tamanho do fragmento amplificado com o número de repetições indicado na segunda coluna da tabela.

Os cinco STRs foram amplificados por PCR numa única reação multiplex em que aos *primers* 5' foram adicionadas oligonucleotídeos marcados com diferentes fluorescências (*tails*) que permitiram identificar os fragmentos associados a cada STR por electroforese capilar.

Em cada reação multiplex usou-se um volume final de 10  $\mu\text{L}$  com 5  $\mu\text{L}$  de *Taq PCR Master Mix Kit* (5 u/ $\mu\text{L}$  QIAGEN®), 1  $\mu\text{L}$  de uma solução com os 10 *primers* numa concentração de 10 pM para os *primers* diretos e 100 pM para os *primers* reversos, 3  $\mu\text{L}$  água de desionizada e 1  $\mu\text{L}$  de DNA (10-15 ng/  $\mu\text{L}$ ). O protocolo de PCR envolveu um passo inicial de desnaturação a 95 °C por 15 min., ao qual se sucederam 11 ciclos compostos por 30 seg. de desnaturação a 95 °C, uma temperatura de hibridização 0,5 °C mais baixa a cada ciclo (*Touch down*), de 60 para 55 °C, por um período de 1min. e, finalmente, uma extensão a 72 °C por 30 seg. O restante programa consistiu em mais 2 ciclos. O primeiro, composto por desnaturação 95 °C /30 seg.; hibridização 55 °C /1 min. e extensão 72 °C /30 seg., num acumulado de 21 ciclos. O segundo, que visava a incorporação das caudas fluorescentes, composto por desnaturação 95 °C /30 seg.; *annealing* 53 °C /45 seg. e extensão 72 °C /30seg., num total de 8 ciclos. No fim foi ainda feita uma extensão final de 30 min. a 60 °C.

Os produtos de amplificação foram separados por eletroforese capilar no sequenciador automático *ABI™ 3130xl Genetic Analyzer* da Applied Biosystems®. Os fragmentos foram analisados recorrendo à aplicação *GeneMapper® v. 4.1* (Applied Biosystems®).

### **2.3. Inferência de haplótipos e outras análises estatísticas**

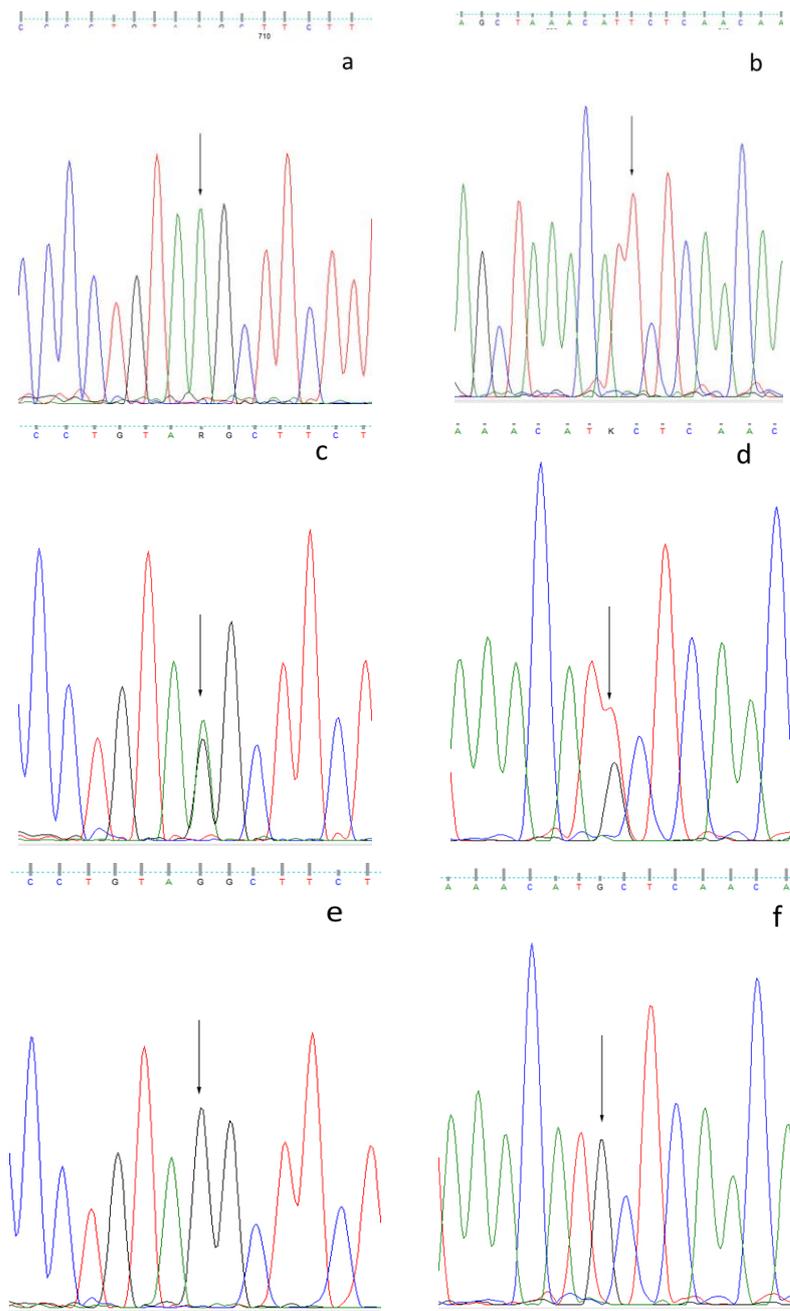
Os haplótipos associados às variantes de APOL1 foram inferidos a partir dos dados genotípicos de acordo com o método ELB (*Excoffier-Laval-Balding*) (Excoffier *et al.*, 2003) disponibilizado no programa informático *ARLEQUIN v. 3.5.1.3*. (Excoffier e Lischer, 2010)

O programa *ARLEQUIN* foi também usado para o cálculo de frequências alélicas, heterozigotias e análise de adequação ao formalismo de *Hardy-Winberg*.

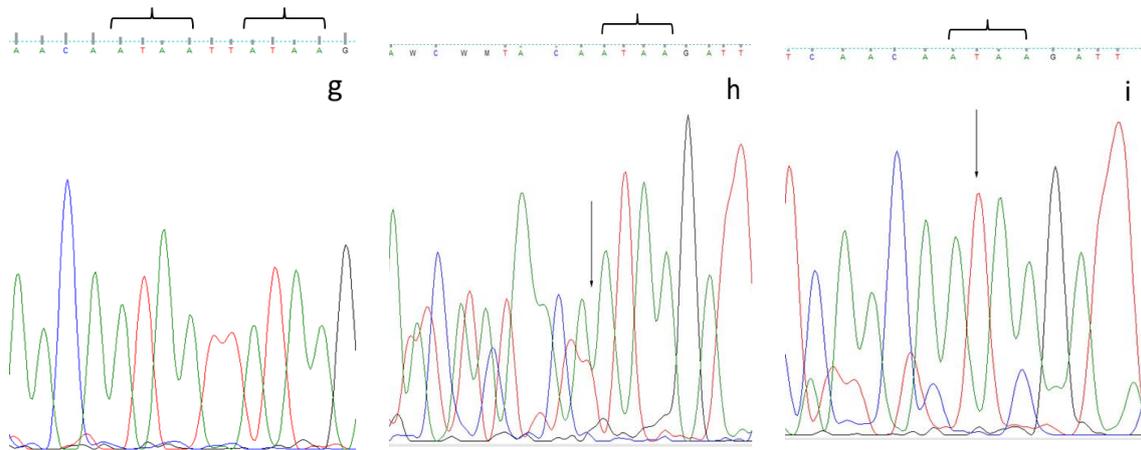
### 3. Resultados e Discussão

#### 3.1. Genotipagem da variação no *locus* APOL1 e nas regiões adjacentes

As Figuras 7 e 8 mostram os resultados das reações de sequenciação em indivíduos com diferentes genótipos com presença ou ausência das variantes G1 e G2 no *locus* APOL1.

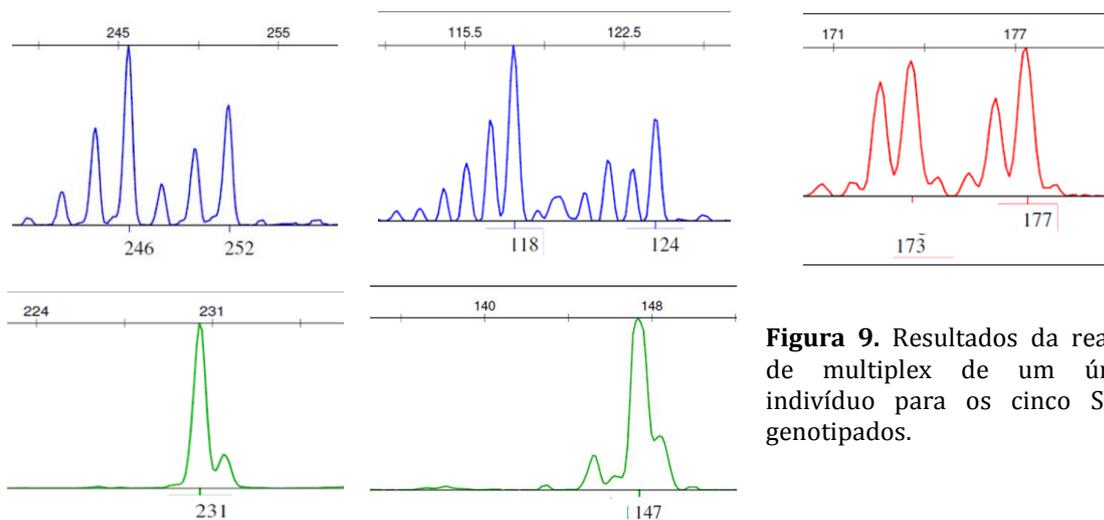


**Figura 7.** Resultados da sequenciação para os *loci* (*rs73885319* e *rs60910145*) que definem a G1. Setas indicam posições de interesse. a,b) Indivíduo homocigótico normal (A;T). c,d) Indivíduo heterocigótico (A/G; T/G.) e,f) Indivíduo homocigótico para as variantes (G;G).



**Figura 8.** Resultados da sequenciação para a G2. Setas verticais indicam as posições de interesse na sequência. A variante G2 que consiste na deleção de 6 pares de bases numa região delimitada por 2 sequências repetitivas ATAA, identificadas na figura. A deleção de 6 nucleótidos entre estes dois elementos repetitivos conduz à perda na sequência proteica do Apol1 dos 2 aminoácidos Asn388-Tyr389. g) Indivíduo homocigótico normal sem deleção, em que são mostradas as 2 sequências repetitivas ATAA (chavetas); h) indivíduo heterocigótico para a deleção (sequenciação com *primer* reverso) em que é mostrada a sequência (complementar-reversa) com desfaseamento até ao início da segunda sequência repetitiva ATAA i) Indivíduo homocigótico com a deleção em que é mostrada a sequência repetitiva ATAA que permanece após a deleção.

Na figura 9 apresentam-se exemplos da resolução alcançada na reação de multiplex para os cinco STRs utilizados para caracterizar a diversidade haplotípica associada aos variantes da APOL1.



**Figura 9.** Resultados da reação de multiplex de um único indivíduo para os cinco STRs genotipados.

As metodologias de genotipagem aqui desenvolvidas podem considerar-se robustas, tendo permitido a obtenção de informação sobre a variação genética de 13 SNPs, um polimorfismo de inserção/deleção e 5 STRs em apenas dois passos experimentais. Nos 13 SNPs observados por sequenciação de DNA incluem-se os dois polimorfismos que definem a variante G1 (rs60910145 e rs73885319), um SNP adjacente não documentado nas bases de dados e 10 SNPs previamente identificados. A inserção/deleção é o polimorfismo associado à variante G2 que, como se referiu, consiste na deleção de um fragmento com seis pares de bases.

## 3.2. Variação genética no *locus* APOL1 e nas regiões adjacentes

Na tabela 2 apresentam-se as frequências alélicas e os níveis de heterozigotia esperada, calculados em 154 habitantes da ilha do Príncipe para os 19 polimorfismos analisados com a metodologia apresentada na secção *Material e métodos*.

### 3.2.1 Frequência da variante G1

A frequência da variante G1 estimada na amostra total é de  $0,20 \pm 0,023$  e situa-se entre os valores mais altos de 0,41-0,46 registados em povos da região Centro-Oeste de África, como os *Asante* do Gana (Figura 2 a) ou os Yoruba da Nigéria (Behar *et al.*, 2011) e valores mais baixos de 0,10 observados nas regiões de língua *Bantu* para as quais há dados disponíveis (Figura 2a)

Uma explicação possível para os resultados agora obtidos no Príncipe é que eles refletem simplesmente a contribuição proporcional de trabalhadores forçados provenientes das regiões da Costa dos Escravos (Centro-Oeste de África) e do Congo-Angola (onde se falam línguas *Bantu*) para o povoamento da ilha. Por exemplo, se essa contribuição tiver sido equitativa, como alguns resultados relativos à ilha de São Tomé parecem indicar (Tomas *et al.*, 2002), a frequência esperada no Príncipe seria dada aproximadamente por:

$$0,4 \times 0,5 + 0,10 \times 0,5 = 0,25$$

o que estaria próximo do que foi encontrado.

**Tabela 2.** Frequências alélicas e níveis de heterozigotia calculados em 154 habitantes da ilha do Príncipe para os 19 polimorfismos utilizados para caracterizar a diversidade haplotípica no locus APOL1.

Locus 1		Locus 2 (20989) <sup>1</sup>		Locus 3 (4895)		Locus 4 (24607)		Locus 5 (78)		Locus 6 (1)		Locus 7 (127)		Locus 8 (30)		Locus 9 (80)		Locus 10 (28)		Locus 11 (5)		Locus 12 (12)		Locus 13 (151)		Locus 14 (49)		Locus 15 (15) <sup>2</sup>		Locus 16 (128)		Locus 17 (12)		Locus 18 (4069)		Locus 19 (83357)		
STR 1		STR 2		STR 3 (D22S1173)		SNP 1 (rs2239785)		P 2 (não documentad)		SNP 3 (rs116136671)		SNP 4 (rs136174)		SNP 5 (rs136175)		SNP 6 (rs136176)		SNP 7 (rs73885316)		SNP 8 (rs142955744)		SNP 9 (rs73403889)		SNP 10 (rs136177)		SNP 11 (rs16996616)		SNP 12 (rs73885319)		SNP 13 (rs60910145)		Inserção/Deleção (rs71785313)		STR 4		STR 5		
Alelo	Frequência	Alelo	Frequência	Alelo	Frequência	Alelo	Frequência	Alelo	Frequência	Alelo	Frequência	Alelo	Frequência	Alelo	Frequência	Alelo	Frequência	Alelo	Frequência	Alelo	Frequência	Alelo	Frequência	Alelo	Frequência	Alelo	Frequência	Alelo	Frequência	Alelo	Frequência	Alelo	Frequência	Alelo	Frequência	Alelo	Frequência	
226	0,0032468	106	0,064935	173	0,0032468	A	0,38312	A	0,97078	A	0,92208	A	0,93831	A	0,93831	A	0,94481	A	0,016234	G	0,0097403	A	0,0064935	A	0,92857	A	0,090909	A	0,7987	T	0,7987	W <sup>3</sup>	0,89935	223	0,016234	133	0,097403	
228	0,0032468	108	0,0064935	175	0,029221	G	0,61688	G	0,029221	G	0,077922	C	0,061688	G	0,061688	G	0,055195	C	0,98377	T	0,99026	G	0,99351	G	0,071429	G	0,90909	G	0,2013	G	0,2013	Del <sup>3</sup>	0,10065	227	0,11039	143	0,042208	
230	0,0032468	110	0,022727	177	0,72727																																	
240	0,042208	112	0,048701	179	0,094156																																	
242	0,085714	114	0,0032468	181	0,13636																																	
244	0,27922	116	0,019481	183	0,0097403																																	
246	0,077922	118	0,27922																																			
248	0,090909	120	0,15584																																			
250	0,10065	122	0,10714																																			
252	0,25974	124	0,16558																																			
254	0,081169	126	0,055195																																			
256	0,016234	128	0,068182																																			
258	0,0064935	130	0,0032468																																			
Heterozigotia	0,82012187	0,84361148	0,44265966	0,47267813	0,056732325	0,14369664	0,11576893	0,11576893	0,10428758	0,03193304	0,01929026	0,01289571	0,13265565	0,165290926	0,32155662	0,32155662	0,989869578	0,57265244	0,47052523																			

<sup>1</sup>Entre parêntesis indica-se a distância em bp entre cada marcador e o marcador anterior  
<sup>2</sup>Os loci que definem as variantes G1 e G2 estão assinalados a vermelho e amarelo, respectivamente.  
<sup>3</sup>Del- Deleção; W- *Wildtype* (sem deleção)

No entanto, esta interpretação pode ser demasiado simplista porque, tal como se referiu na introdução, a ilha do Príncipe sofreu várias modificações importantes da sua estrutura demográfica original, que provocaram uma diminuição significativa dos descendentes da população original, formada nas primeiras fases do povoamento. Uma dessas modificações envolveu a imigração maciça de trabalhadores de Cabo Verde para compensar a crise de mortalidade causada pela epidemia de doença do sono do princípio do século XX, e cujos descendentes representam hoje em dia a maioria dos habitantes do Príncipe (Maurer, 2009).

Este tipo de movimentos populacionais mostra que o povoamento da ilha do Príncipe foi um processo complexo, marcado pela heterogeneidade genética dos povos neles envolvidos, cujo amalgamento completo numa população homogénea está longe de estar concluído. A heterogeneidade genética da ilha do Príncipe é aliás refletida na ausência de verificação dos pressupostos do formalismo de *Hardy-Weinberg* quando se analisa a distribuição dos genótipos que envolvem a variante G1 (Tabela 3).

O desacordo ilustrado na Tabela 3, deve-se a um excesso de homocigóticos e é provável que tenha resultado da junção na amostra total de indivíduos provenientes de subpopulações de origens diferentes, entre as quais os cruzamentos não se fazem ao acaso. Assim, para testar esta suposição, procedeu-se à análise das frequências génicas e do equilíbrio de *Hardy-Weinberg* em subconjuntos da amostra, estratificados de acordo com a informação genealógica que foi possível obter durante a recolha das amostras.

**Tabela 3.** Distribuição dos genótipos associados à variante G1 e valores esperados de acordo com o formalismo de *Hardy-Weinberg* em 154 indivíduos residentes na ilha do Príncipe.

G1G1	Genótipos		Teste $\chi^2$	Frequência de G1
	G1/N <sup>1</sup>	N/N		
11 (6,24) <sup>2</sup>	40 (49,52)	103 (98,24)	$\chi^2=5,69$ ; 1 g.l; $p<0,05$	0,20±0,023

<sup>1</sup> N designa todos os cromossomas que não têm o variante G1.

<sup>2</sup> Valores esperados entre parêntesis.

Na tabela 4 mostra-se o resultado obtido para a distribuição dos genótipos da variante G1 numa subamostra de 33 indivíduos com os quatro avós nascidos na ilha do Príncipe. Muitos destes dadores consideram-se descendentes das primeiras famílias de *escravos forros* que emergiram no Príncipe e designam-se a si mesmos por *minu'le*, o que significa literalmente meninos da ilha no crioulo do Príncipe que alguns deles ainda falam, apesar de ser uma língua em extinção (Maurer, 2009).

Neste caso já não há desacordo com os pressupostos do equilíbrio de *Hardy-Weinberg* e a frequência da variante G1 ( $0,32\pm 0,067$ ) é mais elevada do que a obtida na amostra total ( $0,20\pm 0,023$ ). Esta melhoria do acordo com o formalismo de *Hardy-Weinberg* mostra que os habitantes do Príncipe com maior probabilidade de representarem os estratos populacionais que originalmente colonizaram a ilha não estão totalmente amalgamados com outros setores da população, pelo que não estão estabelecidas as condições de panmixia. Por outro lado, a frequência mais elevada de G1 registada na subamostra da tabela 4 aproxima-se dos valores mais altos do continente africano registados nas regiões centrais da África Ocidental (Figura 2a). Saber se estes valores refletem uma predominância de escravos da África Centro-Oriental nos estádios iniciais do povoamento, ou a um enriquecimento seletivo da variante G1 durante a epidemia de doença do sono é, para já, impossível com os dados disponíveis. No futuro, porém, se se esclarecer qual é a contribuição proporcional de várias regiões de África para o povoamento do Príncipe, será possível verificar até que ponto as frequências de G1 foram influenciadas por contribuições demográficas ou seletivas.

**Tabela 4.** Distribuição dos genótipos associados à variante G1 e valores esperados de acordo com o formalismo de *Hardy-Weinberg* em 33 indivíduos residentes no Príncipe com os quatro avós nascidos na ilha.

G1G1	Genótipos		Teste $\chi^2$	Frequência de G1
	G1/N <sup>1</sup>	N/N		
4 (3,34) <sup>2</sup>	13 (14,32)	16 (15,34)	$\chi^2=0,28$ ; 1 g.l; $p>>0,05$	$0,32\pm 0,067$

<sup>1</sup> N designa todos os cromossomas que não têm o variante G1.

<sup>2</sup> Valores esperados entre parêntesis.

### 3.2.2 Frequência da variante G2

A frequência observada da variante G2 na amostra total da ilha do Príncipe ( $0,10 \pm 0,017$ ; Tabela 5) é próxima das frequências registadas em várias regiões africanas com exceção da zona fronteira entre o Gana e o Burkina Faso, onde G2 atinge as frequências mais elevadas do continente (cerca de 0,2; Figura 2a).

Neste caso, e ao contrário do que acontece com G1, a distribuição dos genótipos associados à variante G2 na amostra total está de acordo com o esperado de acordo com o formalismo de *Hardy-Weinberg* (Tabela 5).

**Tabela 5.** Distribuição dos genótipos associados à variante G2 e valores esperados de acordo com o formalismo de *Hardy-Weinberg* em 154 indivíduos residentes na ilha do Príncipe.

Genótipos			Teste $\chi^2$	Frequência de G1
G2G2	G2/I <sup>1</sup>	I/I		
3 (1,56) <sup>2</sup>	25 (27,88)	126 (124,56)	$\chi^2=1,64$ ; 1 g.l; $p >> 0,05$	$0,10 \pm 0,017$

<sup>1</sup> I designa todos os cromossomas que não têm a variante G2.

<sup>2</sup> Valores esperados entre parêntesis.

Na subamostra de dadores com quatro avós nascidos na ilha do Príncipe a frequência de G2 ( $0,03 \pm 0,02$ ; Tabela 6) é mais baixa do que na amostra total, embora os intervalos de confiança das estimativas a 95% acabem por se sobrepor. Também nesta subamostra há acordo entre os valores observados e esperados segundo o formalismo de *Hardy-Weinberg* (Tabela 6).

Estes resultados mostram que a subestruturação populacional detetada com a variante G1 não afecta a variante G2, provavelmente porque a frequência de G2 é menos heterogénea nas populações de origem do que a frequência de G1.

Aceitando que a doença renal associada às variantes de APOL1 tem manifestação recessiva, tal como sugerido por Genovese *et al.* (2010), a frequência esperada de genótipos de risco (G1G1, G1G2 e G2G2) na população total é de 0,09, aproximando-se do registado em outras regiões africanas do Centro-Oeste de África, com exceção do Gana, onde essa frequência é mais elevada (Figura 2b). Na subamostra de dadores com quatro avós nascidos na ilha, a frequência de genótipos de risco não é muito diferente (0,11).

Assumindo, por outro lado, que basta a presença de uma das variantes G1 ou G2 para que haja proteção contra a infecção por *T. b. rhodesiense* (Genovese *et al.* 2010) é possível calcular a frequência esperada de indivíduos protegidos somando as frequências dos genótipos G1G1, G1G2, G2G2, G1W, G2W, em que W representa todos os cromossomas que não tenham a variante G1 nem a variante G2. Essas frequências são de 0,49 e 0,45 na população geral e na subamostra com quatro avós nascidos no Príncipe.

**Tabela 6.** Distribuição dos genótipos associados à variante G2 e valores esperados de acordo com o formalismo de *Hardy-Weinberg* em 33 indivíduos residentes no Príncipe com os quatro avós nascidos na ilha.

G2G2	Genótipos		Teste $\chi^2$	Frequência de G1
	G2/I <sup>1</sup>	I/I		
0 (0,03) <sup>2</sup>	2 (1,94)	126 (33,03)	$\chi^2=1,64$ ; 1 g.l; $p >> 0,05$	0,03±0,02

<sup>1</sup> I designa todos os cromossomas que não têm a variante G2.

<sup>2</sup> Valores esperados entre parêntesis.

### 3.2.3 Variação haplotípica associada às variantes G1 e G2

Com o objetivo de analisar a variação haplotípica associada às variantes G1 e G2 (também designada por variação intra-alélica), a fase cromossômica (ou haplotípica) dos genótipos obtidos para os 19 polimorfismos estudados neste trabalho (Tabela 2) foi determinada estatisticamente com o programa *ARLEQUIN* v. 3.5.1.3. (Excoffier e Lischer 2010).

Na amostra total de 154 indivíduos (308 cromossomas) observaram-se 171 haplótipos diferentes ( $k=171/308=0,56$ ) dos quais 14 estão associados à variante G1, 13 à variante G2 e 144 a nenhuma das duas variantes.

Na tabela 7 apresenta-se uma caracterização mais detalhada da diversidade haplotípica associada a cada um destes três tipos de cromossomas.

**Tabela 7.** Estatísticas sumário da diversidade haplotípica no gene APOL1.

Tipo de cromossoma	$n_c^2$	$n_h^3$	$K=(n_h/n_c)$	$H^4$	$P^5$
<b>G1</b> <sup>1</sup>	62	14	0,23	0,53	0,68
<b>G2</b> <sup>1</sup>	31	13	0,42	0,86	0,26
<b>W</b> <sup>1</sup>	215	144	0,67	0,99	0,04

<sup>1</sup>Cromossomas associados às variantes G1 (G1), G2 (G2) ou a nenhuma das variantes (W).

<sup>2</sup>Número total de cromossomas associados a G1, G2 ou W.

<sup>3</sup>Número de haplótipos diferentes associados a G1, G2 ou W.

<sup>4</sup>Heterozigotia haplotípica: probabilidade de retirar ao acaso dois haplótipos diferentes numa amostra de cromossomas G1, G2 ou W.

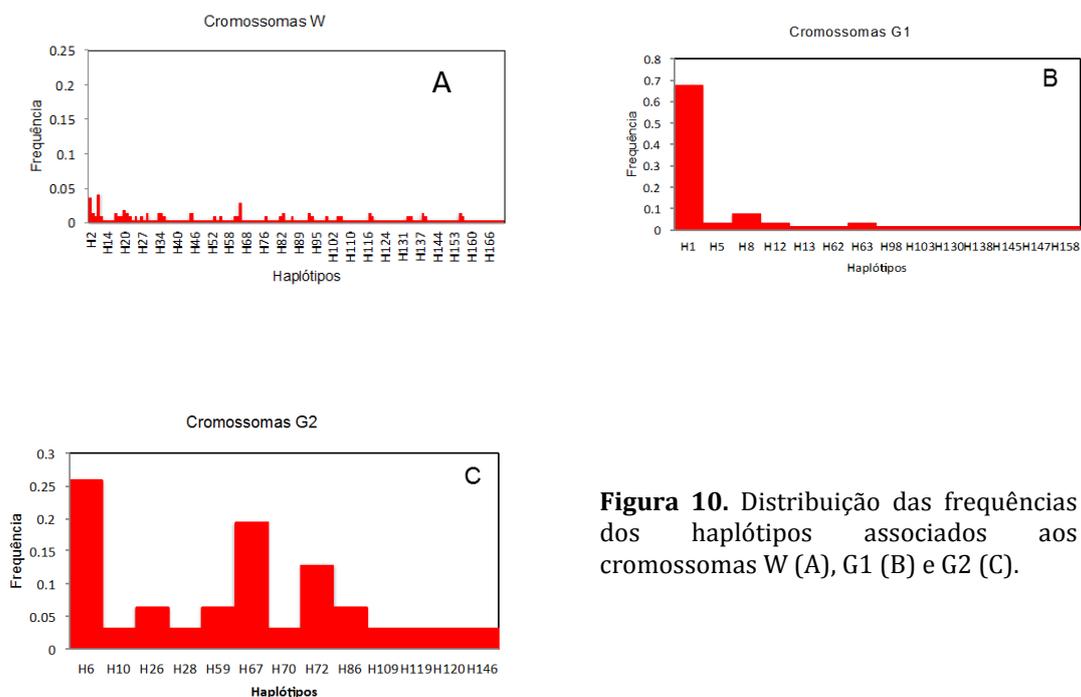
<sup>5</sup>Frequência do haplótipo mais comum associado aos cromossomas G1, G2 ou W.

Os cromossomas que não transportam as variantes G1 e G2 (aqui designados por W) são os cromossomas mais antigos e apresentam a maior diversidade haplotípica (Tabela 7). Por exemplo, 67% destes cromossomas estão distribuídos por haplótipos diferentes ( $k=0,67$ ) e a probabilidade de amostrar dois cromossomas W com haplótipos diferentes é de 99% ( $H=0,99$ ). Por outro lado, a frequência do haplótipo mais comum nos cromossomas W é muito baixa ( $P=0,04$ ),

o que indica que estes haplótipos estão distribuídos de modo tendencialmente uniforme (Figura 10A).

Pelo contrário, os cromossomas associados à variante G1 estão representados por um único haplótipo predominante (H1) com uma frequência relativamente elevada ( $P=0,68$ ) (Tabela 7; Figura 10B). Este haplótipo predominante é muito provavelmente o haplótipo original onde se acumularam as mutações nos SNPs rs73885319 e rs60910145 que definem a variante G1. A sua predominância indica que a variante G1 é suficientemente recente para que a recombinação ou mutação nos STRs adjacentes não tenham tido tempo para promover a acumulação de heterogeneidade intra-alélica. A homogeneidade intra-alélica de G1 é também reflectida noutros indicadores. Por exemplo, em contraste com os cromossomas W, a fracção de haplótipos diferentes nos cromossomas com a variante G1 é de apenas 23% ( $k=0,23$ ), e a probabilidade de amostrar dois haplótipos G1 diferentes é de apenas 53% ( $H=0,53$ ; Tabela 7).

O grau de diversidade intra-alélica da variante G2 situa-se entre os níveis observados para os cromossomas G1 e W. Ao contrário de G1, a variante G2 não tem apenas um haplótipo predominante que se destaque pela sua frequência dos restantes haplótipos (Figura 10B e C).



O haplótipo G2 mais comum (H6) tem uma frequência de 0,26 (P=0,26), mas o segundo haplótipo mais frequente tem uma frequência muito próxima de cerca de 0,19 (H67) e há mesmo um terceiro haplótipo (H72) cuja frequência chega a atingir 0,13 (Figura 10C; Tabela 7). Este tipo de distribuição multimodal dificulta a identificação do cromossoma original onde surgiu inicialmente a deleção de seis pares de bases que define a variante G2 (rs7178513).

Uma alternativa é definir o haplótipo ancestral com base na combinação dos alelos mais frequentes de cada um dos 19 polimorfismos analisados no seio dos cromossomas associados a G2 (Coelho *et al.*, 2005), o que neste caso confirmaria o haplótipo mais comum (H6) como haplótipo ancestral. No entanto, não se pode excluir a possibilidade de G2 ter resultado de pelo menos três mutações independentes, sobretudo se se tiver em conta que a probabilidade de recorrência de inserções/deleções é muito mais elevada do que a das mutações pontuais (Cooper *et al.*, 1995).

### 3.2.4 Idade das variantes variante G1 e G2

A diversidade intra-alélica das variantes G1 e G2 dá-nos informações relevantes sobre a sua antiguidade relativa. Quando mutações que definem cada uma das variantes surgiram num cromossoma ancestral, ficaram inicialmente associadas a um único haplótipo. Com o passar do tempo, a recombinação entre cromossomas G1, G2 e W produz novos arranjos haplotípicos e a fracção de cromossomas associados a cada haplótipo original ( $P$ ; Tabela 7), que originalmente era de 100%, vai decaindo.

Esta queda dos valores de  $P$ , e conseqüente aumento da diversidade intra-alélica, é também facilitada pela taxa de mutação relativamente elevada dos STR que acrescenta novos alelos às associações haplotípicas iniciais. Nestas condições, tal como notaram Stephens *et al.* (1998), é possível estabelecer a seguinte relação analítica entre os valores de  $P$ , as fracções de recombinação entre *loci* adjacentes e o tempo decorrido desde a origem de uma variante:

$$P=(1-r)^t \cong e^{-rt} \quad (1)$$

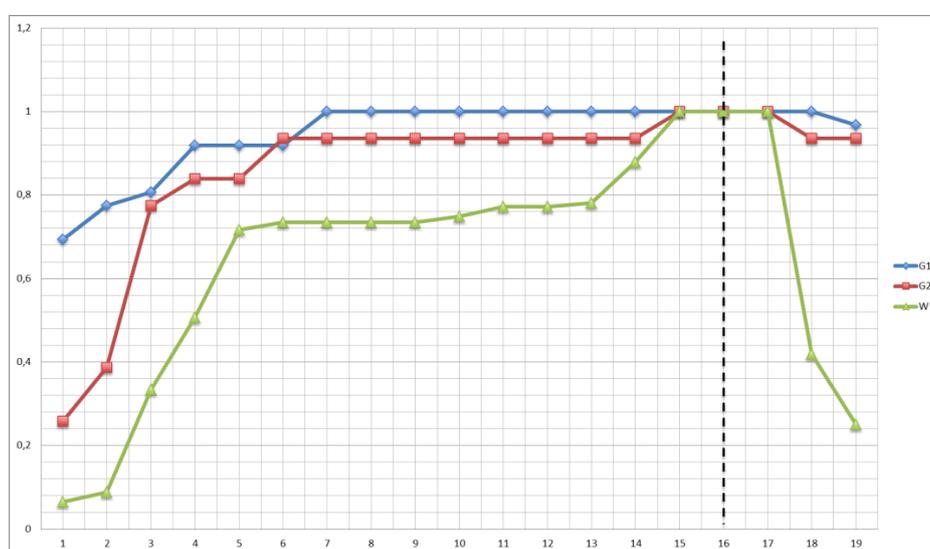
em que  $r$  é a taxa combinada de mutação nos STRs e recombinação entre *loci* adjacentes e  $t$  é o número de gerações decorrido desde a origem de uma variante. Se se souber os valores de  $r$  é possível calcular a idade de uma variante em gerações  $t$ , resolvendo

$$t = -\ln (P)/r \quad (2)$$

Em caso de não se saber os valores de  $r$ , é pelo menos possível estimar quantas vezes um determinado variante é mais antigo que outro calculando a razão entre os respectivos valores de  $-\ln (P)$ .

A figura 11 ilustra a diferença entre a diversidade haplotípica associada a cromossomas W, G1 e G2 em termos dos respectivos valores de  $P$  calculados à medida que nos afastamos, para a direita e para a esquerda, de uma posição de referência centrada no *locus* rs60910145, onde ocorreu uma das mutações que

define a variante G1. Como se pode verificar, os valores de  $P$  vão sendo cada vez mais baixos à medida que aumenta a distância ao *locus* de referência, devido ao aumento da probabilidade de recombinação. No entanto, a erosão da homogeneidade haplotípica é muito mais marcada nos cromossomas W do que nos cromossomas G1 e G2, para a mesma distância ao marcador de referência. Como estas distâncias implicam a mesma fracção de recombinação, a diferença entre os perfis de heterogeneidade haplotípica da figura 11 só pode dever-se a diferenças na idade dos diferentes tipos de cromossomas.

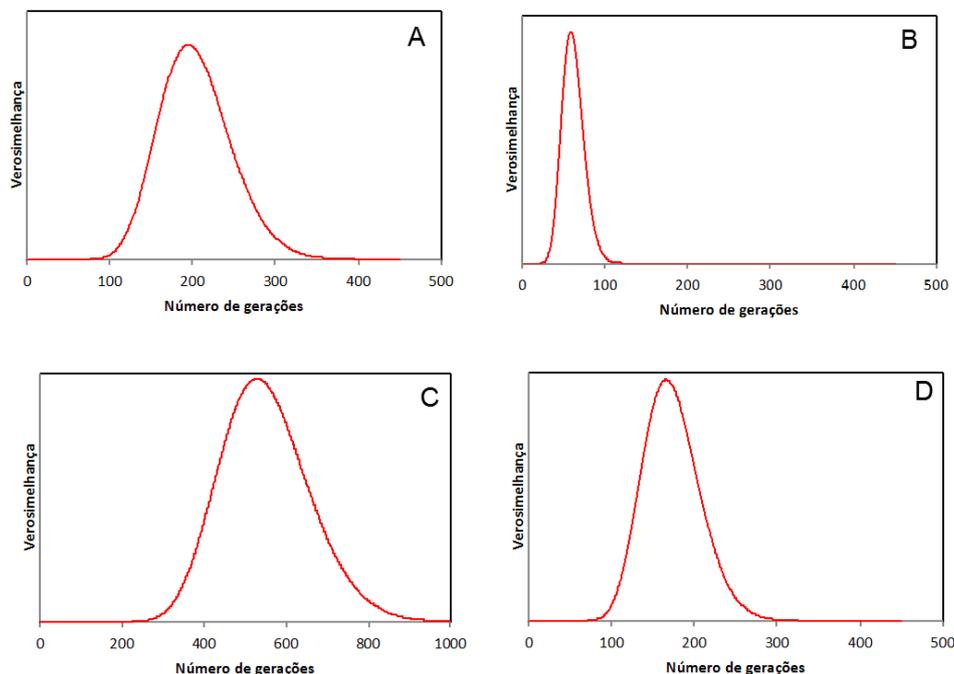


**Figura 11.** Homogeneidade haplotípica associada aos cromossomas W, G1 e G2. Em abcissas indica-se o número do polimorfismo flanqueante de acordo com a Tabela 2. Em ordenadas indica-se, para cada polimorfismo, a percentagem de haplótipos que mantém a configuração Em ordenadas indica-se, para cada polimorfismo, a percentagem de haplótipos que mantém a configuração haplotípica mais comum de cada tipo de cromossoma (W, G1, G2) desde o *locus* de referência até esse polimorfismo. A linha tracejada destaca o polimorfismo de referência pelo *locus* rs60910145.

Se se aceitar a relação  $1\text{cM}=1\text{Mb}$  (Kong *et al.*, 2002) e uma taxa de mutação média para STRs de  $\mu=0,001/\text{geração}$  (Weber e Wong, 1993) o valor de  $r$  à esquerda do marcador de referência será de  $0,000512 + 0,003=0,003512$ , em que a primeira parcela corresponde à fracção de recombinação entre o *locus* 1 e o *locus* de referência (*locus* 16; rs60910145; cf. Tabela 2) e a segunda parcela corresponde à taxa de mutação acumulada dos três STRs situados à esquerda desse *locus* (cf.

Tabela 2). Para a direita do marcador de referência,  $r$  será dado por  $0,0008743+0,002=0,0038743$ , em que a primeira parcela corresponde à fracção de recombinação entre o *locus* 19 e o *locus* de referência e a segunda parcela corresponde à taxa de mutação acumulada dos dois STRs situados à direita desse *locus* (cf. Tabela 2). Se, por outro lado, aceitarmos que a taxa de mutação dos STRs é uma ordem de grandeza inferior ( $\mu=0,0001$ /geração) como por vezes é calculado (Alves *et al.*, 2011), os valores de  $r$  à esquerda e à direita do *locus* de referência serão 0,000812 e 0,00011, respetivamente. Usando estes valores e a lógica da equação (2) é possível tentar estimar mínimos e máximos da idade absoluta das variantes G1 e G2, recorrendo a uma função de verosimilhança baseada na distribuição binomial.

A figura 12 mostra as curvas de verosimilhança da idade das variantes G1 e G2 para dois conjuntos diferentes de valores de  $r$ , obtidos com o pressuposto de que a taxa média de mutação em STRs é 0,001/geração ou 0,0001/geração. Na tabela 8 apresentam-se as estimativas da idade em anos, calculadas a partir das modas das curvas de verosimilhança, bem como os respectivos intervalos de confiança a 95%.



**Figura 12.** Curvas de verosimilhança da idade das variantes G1 (A e B) e G2 (C e D) assumindo taxas de mutação dos microssatélites adjacentes de  $\mu=0,0001$ /geração (A e C) e  $\mu=0,001$ /geração (B e D).

Como seria de esperar das diferenças entre os níveis de variação intra-alélica das duas variantes (Figura 9B e C; figura 10; Tabela 7), G1 é significativamente mais recente do que G2. Esta diferença está de acordo com os resultados obtidos por (Genovese *et al.*, 2010) que também mostraram que a diversidade haplotípica de G1 é menor do que a de G2, com base num número muito mais elevado de marcadores adjacentes (Figura 10; Tabela 7). Estes autores verificaram ainda, usando testes de seleção adequados, que os níveis de diversidade haplotípica sugerem que G1 é uma variante demasiado recente para ter atingido as frequências alélicas observadas na África Ocidental sem que tenha havido qualquer vantagem seletiva, embora não tenha sido feito uma estimativa formal da sua idade. As idades de 5600 anos ou 1624 anos agora calculadas (Tabela 8) são compatíveis com o aparecimento da variante G1 durante a difusão da agricultura tropical em África (Ehret, 2002) e reforçam a ideia de que o principal fator seletivo responsável pela sua disseminação foram os agentes da doença do sono, cuja capacidade de infectar a nossa espécie está intimamente associada à criação de gado. No caso da idade mais antiga de G1 (5600 anos), é possível estimar com base em equações determinísticas (Hartl e Clark, 2007), que seria necessário um coeficiente de seleção de 4 a 5% para a variante atingir uma frequência semelhante à observada no Príncipe (0,20), para um tamanho efetivo populacional de 10000. Se se aceitar para G1 a idade mais recente (1624 anos), o coeficiente de seleção será de cerca de 15%, mantendo o mesmo pressuposto quanto ao tamanho da população.

**Tabela 8.** Estimativas da idade das variantes G1 e G2 do *locus* APOL1.

<b>Variante</b>	<b>Idade em anos (<math>\mu=0,0001</math>)<sup>1</sup></b>	<b>Idade em anos (<math>\mu=0,001</math>)<sup>2</sup></b>
<b>G1</b>	5600 <sup>3</sup> (3248-8428) <sup>4</sup>	1624 (980-2520)
<b>G2</b>	14784 (9268-14784)	4648 (2912-7000)

<sup>1</sup>Assumindo uma taxa de mutação por geração de 0,0001

<sup>2</sup>Assumindo uma taxa de mutação por geração de 0,001

<sup>3</sup>Assumindo 28 anos para a duração média de uma geração (Fenner, 2005)

<sup>4</sup>Intervalo de confiança a 95%

Em relação a G2, como Genovese *et al.* (2010) reconhecem, a sua maior diversidade haplotípica (maior idade) combinada com frequências alélicas mais baixas (Tabela 2) são menos compatíveis com modelos de seletivos, ou pelo menos com modelos seletivos com elevado coeficiente de seleção. Por exemplo, no caso da estimativa de idade mais elevada (Tabela 8), seria necessário um coeficiente de seleção entre 1 a 2% para obter uma frequência de 0,10, tal como observado, numa população com um tamanho efetivo de 10000 indivíduos ao fim de 14784 anos. Para a datação mais recente o coeficiente de seleção seria entre 4 e 5%. Em qualquer caso, estes coeficientes só fariam sentido se se tivesse demonstrado que não era possível explicar a distribuição de G2 com modelos neutros, o que ainda não foi feito inequivocamente. Há contudo que ter em conta que, se se assumir que G2 não resultou de uma única mutação, mas sim de pelo menos três mutações independentes, tal como se discutiu na secção anterior, é possível que a maior diversidade haplotípica observada nesta variante não seja devida à sua antiguidade mas antes à mistura de haplótipos com origens diferentes na população amostrada. Esta possibilidade corresponde à situação normalmente designada por “*soft sweep*” (favorecimento seletivo de sinal fraco) e é uma causa conhecida de diminuição da capacidade detectar seleção (Hermisson e Pennings, 2005; Pennings e Hermisson, 2006a, b; Pritchard *et al.*, 2010) Para avaliar melhor a plausibilidade desta hipótese, será contudo necessário aumentar o número de regiões africanas caracterizadas para a diversidade intra-alélica de G2, de forma a identificar os possíveis locais de origem das mutações independentes.

### 3.3. Conclusões

O estudo piloto agora realizado mostrou que é possível em condições técnicas relativamente simples e reprodutíveis obter uma descrição da diversidade intra-alélica dos variantes G1 e G2 do gene APOL1 que capta propriedades relevantes para estudar a história evolutiva destes alelos em populações africanas. Para tal será necessário estender substancialmente o número de indivíduos e áreas geográficas amostradas, a fim de que possam ser identificadas regiões de origem plausíveis e possíveis rotas de dispersão. Em todo o caso, com o trabalho agora realizado foi possível obter uma prova de princípio da utilidade do sistema desenvolvido e obter informação sobre aspectos desconhecidos da variação genética de APOL1, de entre os quais se destacam os seguintes:

1. A conjugação da sequenciação direta de uma região do gene da APOL1 com uma técnica de genotipagem simultânea por PCR-multiplex de STRs adjacentes permite caracterizar, de forma rápida e reprodutível, a variação genética em 19 polimorfismos informativos, que podem ser usados para estudar a variação haplotípica das variantes G1 e G2 em várias populações.
2. A utilização das técnicas de genotipagem agora desenvolvidas numa amostra piloto da população do Príncipe permitiu verificar que, nesta ilha, as frequências das variantes G1 (0,20) e G2 (0,10) são das mais elevadas do continente Africano, onde só se encontram valores mais elevados nas regiões da Nigéria e do Gana.
3. No caso da variante G1 foi detectada uma heterogeneidade intra-populacional que se reflete na ausência de uma distribuição conforme ao esperado de acordo com o formalismo de *Hardy-Weinberg*. Quando se restringe a análise a uma subpopulação constituída apenas por indivíduos com os quatro avós nascidos na ilha do Príncipe, a heterogeneidade é eliminada, verifica-se acordo

com o equilíbrio de *Hardy-Weiberg* e a frequência do alelo G1 é consideravelmente mais elevada do que na população geral (0,30). Esta heterogeneidade provavelmente reflete uma diferença entre a frequência de G1 nos estratos populacionais mais próximos da população original, e segmentos populacionais que imigraram para o Príncipe, sobretudo a partir de Cabo Verde, em períodos mais recentes.

4. A variante G1, apesar de mais frequente, tem níveis de diversidade haplotípica marcadamente mais baixos do que G2, caracterizados pelo predomínio de um único haplótipo que se estende por 139,5 kb e representa, 68% dos cromossomas G1 amostrados na ilha do Príncipe. Estes níveis de diversidade haplotípica são compatíveis com uma idade recente da variante G1, situada entre valores modais de 5600 a 1624 anos, compatível com uma origem durante a difusão da agricultura tropical em África, e consistente com o modelo de que houve favorecimento seletivo de G1 devido à resistência que esta variante confere aos agentes da doença do sono.
  
5. A variante G2, é mais antiga do que G1, apesar de ter frequências mais baixas. Em contraste com G1, o perfil de diversidade haplotípica de G2 é multimodal, com três haplótipos de frequências moderadas. Se se aceitar que a mutação associada a G2 teve origem única, este perfil é compatível com idades que variam entre 14784 e 4648 anos e é mais difícil de explicar pela ação da seleção natural. No entanto, a multimodalidade dos perfis haplotípicos de G2 e o facto de a variante resultar de uma deleção, cuja probabilidade de recorrência é maior do que a de mutações pontuais, sugerem que esta variante pode ter tido origem em pelo menos três mutações recorrentes. Nestas condições a maior diversidade haplotípica de G2 pode não refletir uma maior antiguidade mas antes a inclusão na mesma amostra de sequências haplotípicas de origem diferente.

## 4. Referências bibliográficas

- Alves, I.; Coelho, M.; Gignoux, C.; Damasceno, A.; Prista, A.; Rocha, J. 2011. Genetic homogeneity across Bantu-speaking groups from Mozambique and Angola challenges early split scenarios between East and West Bantu populations. *Human biology*, 83(1):13-38.
- Baker, C.; Antonovics, J. 2012. Evolutionary Determinants of Genetic Variation in Susceptibility to Infectious Diseases in Humans. *PLoS ONE*, 7(1):e29089.
- Barnes, I.; Duda, A.; Pybus, O. G.; Thomas, M. G. 2011. Ancient urbanization predicts genetic resistance to tuberculosis. *Evolution*, 65(3):842-848.
- Behar DM, Kedem E, Rosset S, Haileselassie Y, Tzur S, Kra-Oz Z, Wasser WG, Shenhar Y, Shahar E, Hassoun G *et al.* . 2011. Absence of APOL1 risk variants protects against HIV-associated nephropathy in the Ethiopian population. *American journal of nephrology*,34(5):452-459.
- Coelho, M.; Luiselli, D.; Bertorelle, G.; Lopes, A. I.; Seixas, S.; Destro-Bisol, G.; Rocha, J. 2005. Microsatellite variation and evolution of human lactase persistence. *Human genetics*, 117(4):329-339.
- Cooke, G. S.; Hill, A. V. 2001. Genetics of susceptibility to human infectious disease. *Nature Reviews Genetics*, 2(12):967-977.
- Cooper, D. N.; Krawczak, M.; Antonarakis, S. E. 1995. The Nature and Mechanisms of Human Gene Mutation. In: Sriver, C. R.; Beaudet, A. L.; Sly, W.; Valle, D., (eds.). *The Metabolic and Molecular Bases of Inherited Disease*. New York, McGraw-Hill: 259-291.
- Cortes, S. 2008. *Controlo da Tripanossomose Humana Africana*. Tese de Doutoramento. Universidade Nova de Lisboa.
- Costa V. 1954. *A Ilha do Príncipe e a Doença do Sono (Nota Prévia)*. I Congresso Nacional de Medicina Tropical. Lisboa.
- Ehret, C. 2002. *The Civilizations of Africa: A History to 1800*. University Press of Virginia.
- Excoffier L, Laval G, and Balding D. 2003. Gametic phase estimation over large genomic regions using an adaptive window approach. *Human genomics*, 1(1):7-19.
- Excoffier L, and Lischer HEL. 2010, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3):564-567.

- Fenner JN. 2005. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *American Journal of Physical Anthropology*, 128(2):415-423.
- Freedman, B. I.; Kopp, J. B.; Langefeld, C. D.; Genovese, G.; Friedman, D. J.; Nelson, G. W.; Winkler, C. A.; Bowden, D. W.; Pollak, M. R. 2010, The apolipoprotein L1 (APOL1) gene and nondiabetic nephropathy in African Americans. *Journal of the American Society of Nephrology*, 21(9):1422-1426.
- Friedman, D. J.; Pollak, M. R. 2011. Genetics of kidney failure and the evolving story of APOL1. *Journal of Clinical Investigation*, 121(9):3367-3374.
- Genovese, G.; Friedman, D. J.; Ross, M. D.; Lecordier, L.; Uzureau, P.; Freedman, B. I.; Bowden, D. W.; Langefeld, C. D.; Oleksyk, T. K.; Uscinski Knob, A. L.; Bernhardt, A. J.; Hicks, P. J.; Nelson, G. W.; Vanhollebeke, B.; Winkler, C. A.; Kopp, J. B.; Pays, E.; Pollak, M. R. 2010, Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science*, 329(5993):841-845.
- Goldstein, D. B.; Chikhi, L. 2002. Human migrations and population structure: what we know and why it matters. *Annual Review of Genomics and Human Genetics*, 3(1):129-152.
- Hartl, D. L.; Clark, A. G. 2007. *Principles of Population Genetics*. Sinauer Associates.
- Hermisson, J.; Pennings, P. S. 2005. Soft Sweeps. *Genetics*, 169(4):2335-2352.
- Jobling, M. A.; Hurles, M.; Tyler-Smith, C. 2004. *Human evolutionary genetics: origins, peoples & disease*. New York, Garland Science.
- Kao, W. H. L.; Klag, M. J.; Meoni, L. A.; Reich, D.; Berthier-Schaad, Y.; Li, M.; Coresh, J.; Patterson, N.; Tandon, A.; Powe, N. R.; Fink, N. E.; Sadler, J. H.; Weir, M. R.; Abboud, H. E.; Adler, S. G.; Divers, J.; Iyengar, S. K.; Freedman, B. I.; Kimmel, P. L.; Knowler, W. C.; Kohn, O. F.; Kramp, K.; Leehey, D. J.; Nicholas, S. B.; Pahl, M. V.; Schelling, J. R.; Sedor, J. R.; Thornley-Brown, D.; Winkler, C. A.; Smith, M. W.; Parekh, R. S. 2008. MYH9 is associated with nondiabetic end-stage renal disease in African Americans. *Nature Genetics*, 40(10):1185-1192.
- Kioy, D.; Jannin, J.; Mattock, N. 2004. Focus: Human African trypanosomiasis. *Nature Reviews Microbiology*, 2(3):186-187.
- Kong, A.; Gudbjartsson, D. F.; Sainz, J.; Jonsdottir, G. M.; Gudjonsson, S. A.; Richardsson, B.; Sigurdardottir, S.; Barnard, J.; Hallbeck, B.; Masson, G.; Shlien, A.; Palsson, S. T.; Frigge, M. L.; Thorgeirsson, T. E.; Gulcher, J. R.; Stefansson, K. 2002. A high-resolution recombination map of the human genome. *Nature Genetics*, 31(3):241-247.

- Kopp, J. B.; Smith, M. W.; Nelson, G. W.; Johnson, R. C.; Freedman, B. I.; Bowden, D. W.; Oleksyk, T.; McKenzie, L. M.; Kajiyama, H.; Ahuja, T. S.; Berns, J. S.; Briggs, W.; Cho, M. E.; Dart, R. A.; Kimmel, P. L.; Korbet, S. M.; Michel, D. M.; Mokrzycki, M. H.; Schelling, J. R.; Simon, E.; Trachtman, H.; Vlahov, D.; Winkler, C. A. 2008. MYH9 is a major-effect risk gene for focal segmental glomerulosclerosis. *Nature Genetics*, 40(10):1175-1184.
- Kronenberg, F. 2011. APOL1 variants and kidney disease. There is no such thing as a free lunch. *Nephrology Dialysis Transplantation*, 26(3):775-778.
- Maurer P. 2009. *Principense (Lung'le) : grammar, texts, and vocabulary of the Afro-Portuguese Creole of the island of Príncipe, Gulf of Guinea*. London/Colombo, Battlebridge.
- Namangala, B. 2011. How the African trypanosomes evade host immune killing. *Parasite Immunology*, 33(8):430-437.
- Pays, E.; Vanhollebeke, B.; Vanhamme, L.; Paturiaux-Hanocq, F.; Nolan, D. P.; Pérez-Morga, D. 2006. The trypanolytic factor of human serum. *Nature Reviews Microbiology*, 4(6):477-486.
- Pennings, P. S.; Hermisson, J. 2006a. Soft Sweeps III: The Signature of Positive Selection from Recurrent Mutation. *PLoS Genetics*, 2(12):e186.
- Pennings, P. S.; Hermisson, J. 2006b. Soft Sweeps II: Molecular Population Genetics of Adaptation from Recurrent Mutation or Migration. *Molecular Biology and Evolution*, 23(5):1076-1084.
- Pritchard JK, Pickrell JK, and Coop G. 2010, The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. *Current Biology* 20(4):208-215.
- Rosset, S.; Tzur, S.; Behar, D. M.; Wasser, W. G.; Skorecki, K. 2011. The population genetics of chronic kidney disease: insights from the MYH9-APOL1 locus. *Nature Reviews Nephrology*, 7(6):313-326.
- Sanger F, Nicklen S, and Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12):5463-5467.
- Slatkin, M.; Rannala, B. 2000. Estimating allele age. *Annual Review of Genomics and Human Genetics*, 1(1):225-249.
- Stephens. J. C.; Reich, D. E.; Goldstein, DB, Shin HD, Smith MW, Carrington M, Winkler C, Huttley GA, Allikmets R, Schriml L *et al.* . 1998. Dating the Origin of the CCR5- $\Delta$ 32 AIDS-Resistance Allele by the Coalescence of Haplotypes. *American journal of human genetics*, 62(6):1507-1515.

- Tomás, G.; Seco, L.; Seixas, S.; Faustino, P.; Lavinha, J.; Rocha, J. 2002. The peopling of Sao Tome (Gulf of Guinea): origins of slave settlers and admixture with the Portuguese. *Human biology*, 74(3):397-411.
- Tzur, S.; Rosset, S.; Shemer, R.; Yudkovsky, G.; Selig, S.; Tarekegn, A.; Bekele, E.; Bradman, N.; Wasser, W. G.; Behar, D. M.; Skorecki, K. 2010, Missense mutations in the APOL1 gene are highly associated with end stage kidney disease risk previously attributed to the MYH9 gene. *Human Genetics*, 128(3):345-350.
- Untergasser, A.; Nijveen, H.; Rao, X; Bisseling, T.; Geurts, R.; Leunissen J.A.M. 2007. *Primer3Plus*, an enhanced web interface to *Primer3*. *Nucleic Acids Research*,35(suppl 2):71-74.
- Weber, J. L.; Wong, C. 1993. Mutation of human short tandem repeats. *Human Molecular Genetics*, 2(8):1123-1128.
- Zenker, M.; Mertens, P. R. 2010, Arrest of the true culprit and acquittal of the innocent? Genetic revelations charge APOL1 variants with kidney disease susceptibility. *International Urology and Nephrology*, 42(4):1131-1134.

