U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc

2.º
CICLO

FCUP
i3S
Ipatimup
2017

U.PORTO

Screening of *HBB* gene mutations in population samples from Alentejo and implementation of a SNaPshot® based system for *HBB**S haplotyping

Cátia Sofia Botelho Couto

FC

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

INSTITUTO DE INVESTIGAÇÃO E INOVAÇÃO EM SAÚDE
UNIVERSIDADE DO PORTO

25 ANOS A DESCOBRIR O CANCRO
ipatimup

Screening of *HBB* gene mutations in population samples from Alentejo and implementation of a SNaPshot® based system for *HBB**S haplotyping

Cátia Sofia Botelho Couto

Dissertação de Mestrado apresentada à Faculdade de Ciências da Universidade do Porto

Biologia

2017

MSc

Screening of *HBB* gene mutations in population samples from Alentejo and implementation of a SNaPshot® based system for *HBB*S* haplotyping

Cátia Sofia Botelho Couto

Mestrado em Biologia Celular e Molecular

Departamento de Biologia

2017

Orientador

Maria João Prata, PhD, Faculdade de Ciências da Universidade do Porto (FCUP), Instituto de Patologia e Imunologia Molecular da Universidade do Poto (Ipatimup), Instituto de Investigação e Inovação em Saúde (i3S)

Coorientador

Luísa Azevedo, PhD, Faculdade de Ciências da Universidade do Porto (FCUP), Instituto de Patologia e Imunologia Molecular da Universidade do Poto (Ipatimup), Instituto de Investigação e Inovação em Saúde (i3S)

All corrections determined by the jury, and only those, were incorporated.

The President of the Jury,

Porto, _____/_____/_____

Agradecimentos

À equipa do Grupo de Genética Populacional e Evolução, do Instituto de Investigação e Inovação em Saúde (i3S), pela hospitalidade com a qual fui recebida, com especial agradecimento ao Professor Doutor António Amorim pela oportunidade concedida em realizar o projeto de mestrado neste grupo de investigação.

À Professora Doutora Maria João Prata pela excelente orientação que me proporcionou nesta importante fase, pelos conhecimentos transmitidos, pela dedicação, pelo rigor exigido, pela preocupação e acima de tudo paciência, à qual os agradecimentos nunca serão demais.

À Professora Doutora Luísa Azevedo, pela coorientação do meu projeto, com a qual os primeiros passos do mesmo foram tomados, e que ao longo do ano manteve a sua porta aberta para todas as dúvidas que surgiam.

À Doutora Verónica Gomes pela inteira disponibilidade e paciência nas fases posteriores da investigação realizada, cujo apoio foi indispensável para a concretização do mesmo.

À Doutora Paula Faustino, do Instituto Nacional de Saúde Doutor Ricardo Jorge (INSA), pela cedência das amostras usadas como controlo no presente trabalho.

Às minhas colegas que integraram o mesmo grupo de investigação, pelo apoio constante e boa disposição, em especial à Sofia pela entreajuda e ausência de momentos monótonos.

À Mafalda e Marina pelo companheirismo e bons momentos, com as quais a passagem pela FCUP tornou-se uma experiência ainda mais gratificante.

A todos os meus grandes amigos que tive o prazer de fazer nos últimos 8 anos, demasiados para enumerar, mas com quem partilhei os momentos mais importantes da minha vida, pelo apoio, equilíbrio e companhia.

Aos meus avós que nunca esquecerei, por encherem a minha vida de carinho e aos quais palavras de agradecimento nunca farão justiça ao apoio incondicional que recebi.

Aos meus pais, que me mostraram que com confiança, perseverança e pensamento positivo tudo é possível. Que me apoiaram em todas as decisões que tomei, a quem devo todas as experiências que vivi e o motivo pelo qual posso ter orgulho da pessoa que me tornei. Os maiores e mais fortes pilares da minha vida, a quem estas palavras de agradecimento não são suficientes para expressar toda a gratidão que sinto.

Por todos os sacrifícios que tomaram para que fosse possível concretizar os meus sonhos, o maior obrigada que alguma vez direi.

A todos, o meu sincero, muito obrigada.

Abstract

It is well established that in regions where malaria is endemic the sickle cell allele *HBB*S* is usually found at high frequency due to it being associated with resistance to the disease. When in homozygosity, it is responsible for the Sickle Cell Anemia (SCA), which in Portugal reaches the highest prevalence in south-central regions. This distribution seems to be explained by the pattern of population migrations that marked the past of these regions.

In order to better understand the presence of Sickle Cell Diseases (SCD) in Portugal, this project aimed at screening the *HBB* gene in three population samples from Alentejo - Coruche, Serpa (Pias) and Alcácer do Sal, and further infer the origin of the most common variant, *HBB*S*, through the analysis of the haplotypic background encompassing the entire cluster of β-globin genes.

In this sense, and as an alternative to the conventional haplotyping strategy based on the analysis of Restriction Fragment Length Polymorphism (RFLP), a SNaPshot® Multiplex system was implemented targeting the most informative positions defining *HBB*S* haplotypes, and then applied to examine *HBB*S*-bearing individuals previously identified by Sanger sequencing.

In a total of 266 individuals analysed, 19 harboured the c.20A>T mutation that underlies *HBB*S*. The variant *HBB*C* was not detected in any sample.

The frequencies of *HBB*S* were 2.2, 2.9 and 7.7% in Coruche, Alcácer do Sal and Serpa respectively, with the estimate for Serpa being the highest up to now reported for Portugal, where foci of high prevalence of HBB*S carries have been described in the South of country.

Among the chromosomes of the 19 individuals bearing the *HBB*S* allele, 29% harboured the Benin haplotype, 13.1% the Senegal and 10.5% the Bantu. Remarkably, whereas in Alcáçer do Sal and Coruche only Senegal and Bantu haplotypes were found, in Serpa all haplotypes were Benin.

Taking into account the global distribution of *HBB*S* haplotypes, the findings here obtained reinforce a scenario before proposed positing that the introduction of HBB*S in south Portugal was mediated by gene influx events with distinct sources: one from the region encompassing the Mediterranean basin (captured by the Benin haplotype), and other from sub-Saharan Africa, likely afforded by the transatlantic slave trade (captured by the Senegal and Bantu haplotypes).

Besides *HBB*S*, two β-thalassemia variants, already reported as amongst the most common in Portugal, were also identified: HbVar.827 (two heterozygous from Coruche)

and HbVar.845 (one heterozygous from Alcácer do Sal, one from Serpa and three from Coruche).

Globally, this study provides a refreshed perspective on the epidemiology of hemoglobinopathies in south Portugal.

Keywords: Sickle Cell Disease (SCD); Portugal; Alentejo; Álcacer do Sal; Serpa; Coruche; *HBB* gene; Single Nucleotide Polymorphisms (SNPs); SNaPshot® Multiplex; Haplotypes.

Resumo

Está bem estabelecido que em regiões onde a malária é endémica se regista elevada frequência da variante *HBB*S*, por ser um alelo que confere resistência à malária. Quando em homozigotia, é responsável pela Anemia Falciforme, que em Portugal atinge uma maior prevalência nas regiões centro-sul. Esta destribuição parece resultar dos padrões de migrações populacionais que marcaram o passado dessas regiões.

Tendo em vista uma análise mais fina da distribuição da Anemia Falciforme no sul de Portugal, com este projeto pretendeu-se efetuar o despiste de mutações no gene *HBB* em amostras populacionais do Alentejo - Coruche, Serpa (Pias) e Alcácer do Sal, e inferir a origem das variantes mais comuns, nomeadamente *HBB*S*, através do respetivo estudo haplotípico envolvendo o *cluster* de genes da globina-β.

Nesse sentido, e em alternativa à estratégia convencional de haplotipagem baseada no estudo de *Restriction Fragment Length Polymorphism* (RFLP), uma técnica baseada em SNaPshot® Multiplex foi desenvolvida para a análise dos polimorfismos mais informativos que definem os haplótipos *HBB*S*, e aplicada nas amostras em que esse alelo tinha sido previamente identificado por sequenciação de Sanger.

Em 266 individuos analisados, 19 apresentaram a mutação c.20A>T que define o alelo *HBB*S*. A variante *HBB*C* não foi detetada em nenhuma amostra.

As estimativas de frequência de *HBB*S* foram 2.2, 2.9 e 7.7% em Coruche, Alcácer do Sal e em Serpa respectivamnete, com a última sendo a mais alta alguma vez reportada em Portugal, onde *foci* de alta prevalência de portadores de *HBB*S* foram anteriormente descritos no sul do país.

Entre os cromossomas dos 19 individuos portadores do alelo *HBB*S*, 29% apresentavam o haplótipo Benin, 13.1% o Senegal e 10.5% o Bantu. De notar, que enquanto em Alcácer do Sal e Coruche apenas os haplótipos Senegal e Bantu foram encontrados, em Serpa foram apenas detetados haplótipos Benin.

Atendendo à destribuição global dos haplótipos associados ao alelo *HBB*S*, os dados agora obtidos reinforçam um cenário proposto anteriormente, segundo o qual a introdução *HBB*S* no sul de Portugal foi mediado por fluxos migratórios com origens distintas: um proveniente da região que engloba toda a bacia Mediterrânica (associado ao haplótipo Benin), e outro da África Subsaariana, muito provavelmente decorrente do tráfico de escravos (associado aos haplótipos Senegal e Bantu).

Além do alelo *HBB*S*, duas variantes de β-talassemia, anteriormente descritas entre as mais comuns em Portugal, foram também identificadas: HbVar.827 (duas amostras

heterozigóticas de Coruche) e HbVar.845 (uma amostra heterozigótica de Alcácer do Sal, uma de Serpa e três de Coruche).

Em suma, este estudo fornece uma perpectiva atualizada relativa à epidemiologia das hemoglobinopatias no sul de Portugal.

Palavras-Chave: Anemia Falciforme; Portugal; Alentejo; Álcacer do Sal; Serpa; Coruche; gene *HBB*; *Single Nucleotide Polymorphisms* (SNPs); SNaPshot® Multiplex; Haplótipos.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

A	Adenine
α	*Alpha*
$\alpha_2\beta_2$	Adult Hemoglobin – major form
$\alpha_2\delta_2$	Adult Hemoglobin – minor form
$\alpha_2\varepsilon_2$	Gower 2 Hemoglobin
$\alpha_2\gamma_2$	Fetal Hemoglobin
β	*Beta*
BEN	Benin
bp	Base Pair
C	Cytosine
ºC	Degree Celsius
CAM	Cameroon
CAR	Central African Republic
δ	*Delta*
DNA	Deoxyribonucleic Acid
ε	*Epsilon*
et al.	et alli/et aliae
ExoSAP	Exonuclease I Shrimp Alkaline Phosphatase
FastAP	Thermosensitive Alkaline Phosphatase
FTA	Flinders Technology Associates
FW	Forward
γ	*Gamma*
G	Guanine
Gln	Glutamine
Glu	Glutamic Acid
h	Hour
Hb	Hemoglobin
HbA	Adult Hemoglobin – major form
HbA_2	Adult Hemoglobin – minor form
HBA1 (α_1)	*Alfa* 1 Globin Gene
HBA2 (α_2)	*Alfa* 2 Globin Gene
HBAP1 ($\psi\alpha_1$)	*Alfa* 1 Globin Pseudogene
HbAS	AS Hemoglobin - Sickle Cell Trait
HBB (β)	*Beta* Globin Gene
HBBP1 (ψβ)	*Beta* Globin Pseudogene

*HBB*S*	Sickle Cell Allele
HBD (δ)	*Delta* Globin Gene
HBE1 (ε)	*Epsilon* Globin Gene
HbF	Fetal Hemoglobin
HBG1 ($^{A}\gamma$)	*Gamma* A Globin Gene
HBG2 ($^{G}\gamma$)	*Gamma* G Globin Gene
HBM ($\psi\alpha_2$)	*Alfa* 2 Globin Pseudogene
HBQ1 (θ_1)	*Theta* Globin Gene
HbS	S Hemoglobin – Sickle Cell Disease
HbVar	Database of Human Hemoglobin Variants and Thalassemia Mutations
HBZ (ζ)	*Zeta* Globin Gene
HBZP1 (ψζ)	*Zeta* Globin Pseudogene
∞	Infinite
i.e.	That is
IVS	InterVening Sequencing
L.D.	Linkage Disequilibrium
Lys	Lysine
µL	Microlitre
µM	Micromolar
min	Minute
mL	Milliliter
mRNA	Messenger Ribonucleic Acid
%	Percent
PCR	Polymerase Chain Reaction
Ψ	*Psi*
®	Registered
RBC	Red Blood Cells
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic Acid
RV	Reverse
SBE	Single Base Extension
SCA	Sickle Cell Anemia
SCD	Sickle Cell Disease
s. d.	Standard Deviation
sec	Second
SEN	Senegal

SNP	Single Nucleotide Polymorphism
θ	*Theta*
T	Thymine
™	Trade Mark
v.	Version
Val	Valine
V	Volts
ζ	*Zeta*
ζ2ɣ2	Portland Hemoglobin
ζ2ε2	Gower 1 Hemoglobin

1. State of Art

1.1. Blood Components

In an average adult, 4 to 5 litres is the amount of blood that courses through the body, allowing the distribution of elements necessary for the normal function of each organ. The blood is constituted mainly by four different components, each one with different particular functions: plasma, leukocytes and phagocytic cells (white blood cells), platelets and erythrocytes (red blood cells (RBC)) (Rogers, 2010).

The plasma it's what allows the transport of the mentioned cells, being constituted by about 7 percent of proteins (such as serum albumin, that represent 60 percent of plasma proteins, globulins and fibrinogen) and more than 90 percent water, which gives the fluidity consistent to the solution (Rogers, 2010; Sherwood, 2015).

White blood cells are involved in the process of immunity, protecting the body against infections and diseases, where the exogenous microorganisms and particles can be ingested and destroyed, or by the production of antibodies.

The platelets, that are small cells with 2 to 4 micrometres of diameter, have as main function the controlling of bleedings, by attaching to the injured endothelial surface of blood vessels, and also the blockage of the entering of exogenous microorganisms to the blood flow, avoiding the infection process (Rogers, 2010).

Last but not least, erythrocytes, which account to about 45 percent of the blood volume, existing on average 5 billion RBC in each millilitre of blood. The cells lacks nuclei and mitochondria, and are disc-shaped with a flattened center that provides an increased surface area for its main function, the carrying of oxygen from the lung and its diffusion to all the body tissues. The process is mediated by hemoglobin (Hb), and each erythrocyte can contain approximately 280 million of these molecules (Fox, 2011; Rogers, 2010; Sherwood, 2015). This protein will be deeply dissected in the next section.

1.2. Hemoglobin

As a colored pigment overwhelmingly found in erythrocytes, hemoglobin is the protein that gives the recognised red shade to the blood (Honig and Adams, 1986; Sherwood, 2015). It is a tetrameric protein, having a quaternary structure of four polypeptide chains (also titled subunits). Adult hemoglobin, known as HbA, is composed by two *alpha*-globin (α) and two *beta*-globin (β), commonly represented as $\alpha_2\beta_2$ (Baldwin and Chothia, 1979; Honig and Adams, 1986; Jensen et al., 1998).

For an effective transport of oxygen through the body, oxygen needs to be able to interact with the transporter, but since the amino-acid chains of the globins are not able

to do the binding irreversible, that is assisted by a protein-bound prosthetic group, the heme (Nelson and Cox, 2012).

Each subunit has a reversible oxygen-binding heme group that consists of an organic ring structure called protoporphyrin that has an iron atom linked in its ferrous ($Fe2^{+}$) state. Is to the iron atom that the actual binding of the oxygen occurs (Figure 1) (Nelson and Cox, 2012).

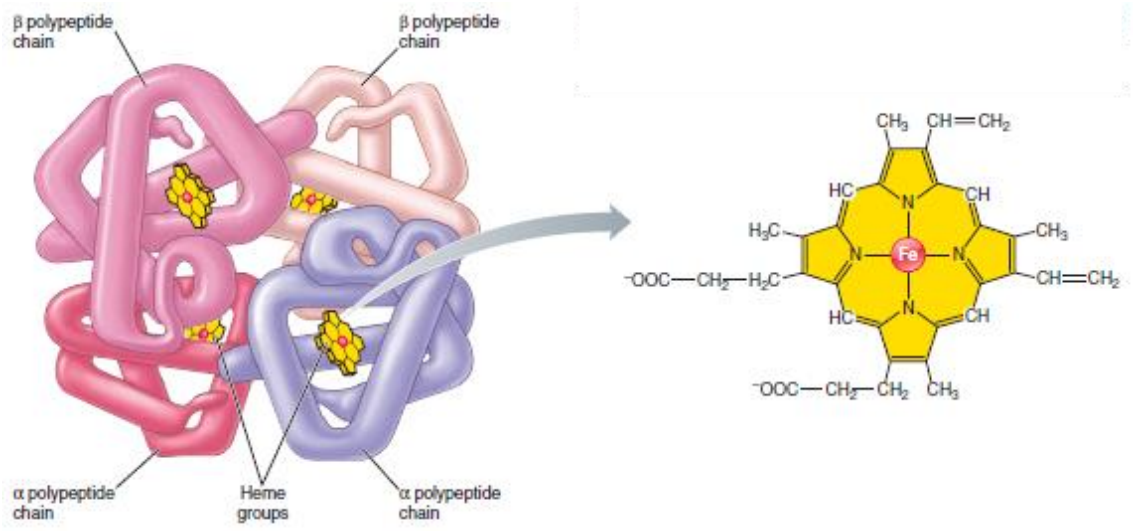

Figure 1 - Hemoglobin molecule and representation of each subunit and heme groups (adapted from Sherwood (2015)).

Although the function of the protein remains the same, there are actually different hemoglobin types, which switch since the time of conception until 48 weeks postnatal age, persisting thereafter through adulthood, probably because of the variation in oxygen requirements during those different times (Weatherall, 2013).

The globins belong to a superfamily that contains two groups, the “α-like” and the “β-like”, encompassing the two kinds of globin chains necessary for the arrangement of different haemoglobin forms. Within each globin group, there is more than one gene that can be translated, allowing the production of more than one α- and β-like globin. The set of genes that code for each globin type, are therefore designated as “α-cluster” and “β-cluster” (Griffiths, 2008).

1.2.1. Hemoglobin Genes

Located in the short arm of chromosome 16 in band p13.3, the α-cluster spans a region of 30kb comprising three genes: *HBZ* (ζ), *HBA2* (α_2) and *HBA1* (α_1); and four pseudogenes (*HBZP1* (ψζ), *HBM* ($\psi\alpha_2$) *HBAP1* ($\psi\alpha_1$) and *HBQ1* (θ_1)). The cluster lies in a GC rich area associated with unmetylated CpG islands (Brittain, 2002; Grosveld et al., 1993; Weatherall, 2013).

On the other hand, the β-cluster is constituted by the genes *HBE1* (ε), *HBG2* (^{G}ɣ), *HBG1* (^{A}ɣ), *HBD* (δ) and *HBB* (β) and the pseudogene *HBBP1* (ψβ), being positioned in an AT rich region of the short arm of chromosome 11, band p15.5, covering approximately 70kb of DNA. The difference between genes ^{G}ɣ and ^{A}ɣ lies on a single substitution that results in products of expression with identical length, though differing at amino-acid residue 136, which in ^{G}ɣ is glycine and in ^{A}ɣ is alanine (Brittain, 2002; Grosveld et al., 1993; Weatherall, 2013).

In spite of the different chromosomal location of the two clusters, they share many common characteristics. In addition of having pseudogenes – duplicate copies of genes that accumulated random mutations making them non-functional – the gene members in both cluster are organized in a sequence that recapitulates the temporal order of appearance of the polypeptides in the course of human development (Figure 2) (Brittain, 2002; Griffiths, 2008; Grosveld et al., 1993; Weatherall, 2013).

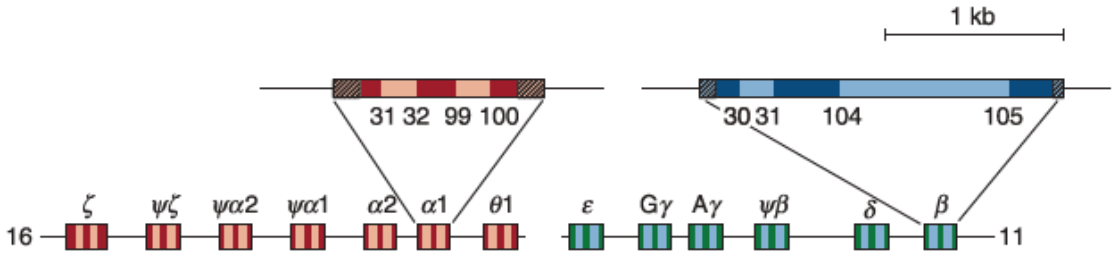

Figure 2 – Representation of α-cluster and β-cluster genes (adapted from Kaushansky et al. (2010)).

1.2.2. Normal Human Hemoglobins

The association of the different globin types, results in the repertoire of hemoglobins that are produced to respond to the variation of oxygen requirements in particular times of human life. During the first two weeks, the embryo is small enough to get the oxygen by simple diffusion processes, however, as long as he growths this process is no longer sufficient to meet the metabolic demands, and therefore, not much before of the fourth week of gestation, three different forms of embryonic hemoglobins start to being produced: Portland (ζ_2ɣ$_2$), Gower 1 ($\zeta_2\varepsilon_2$) and Gower 2 ($\alpha_2\varepsilon_2$) (Brittain, 2002; Grosveld et al., 1993; Weatherall, 2013).

At 10 weeks post-conception, the amount of embryonic hemoglobins drastically decreases, and the production of the fetal hemoglobin (HbF, α_2ɣ$_2$) sets on, remaining the predominant hemoglobin for the rest of gestation. Approaching the 38th week of postconceptional age, a second switch occurs leading to an almost silencing of fetal hemoglobin expression while adult hemoglobin began to be predominantly produced. Therefore, in adults there is the production of HbA ($\alpha_2\beta_2$), HbA$_2$ ($\alpha_2\delta_2$), the major and

minor forms of hemoglobin that account to ~ 97% and ~2% of total hemoglobin, respectively, and a residual proportion of HbF (<1%) (Bard, 1973; Grosveld et al., 1993; Weatherall, 2013).

1.3. Hemoglobin Genetic Disorders

Also designated as hemoglobinopathies, the genetic disorders of Hb are the most common world-wide among the inherited diseases typically associated with an autosomal recessive transmission pattern (Costa et al., 2016). Affecting approximately 7% of the world's population, the conditions are mainly diagnosed through blood tests that include analyses of the quantity and quality of hemoglobin (Hoffbrand et al., 2006).

The two main groups of disorders are thalassemias and structural variants of hemoglobin; each one separately can result in diseases with a large range of severity, although carriers of some genetic mutations can be phenotypically normal. The combination of genetic alterations underlying the two groups of disorders is also possible resulting once again in diverse clinical and pathological manifestations. Together with the large variety of mutations that can affect the globin genes, many other factors can modulate the phenotypic manifestation of the diseases (Bridges and Pearson, 2008; Corrons, 2001)

1.3.1. α and β-thalassemias

Thalassemias, the commonest monogenic diseases in man, are caused by an imbalanced rate of production of normal globin chains of adult hemoglobin. They are divided into α^0-/β^0- and α^+-/β^+- depending on whether the mutation abolishes completely the production of the protein, or diminishes its production, respectively (Flint et al., 1998; Kaushansky et al., 2010).

This abnormalities can be caused by gene deletion or by mutations that affect the transcription of mRNA (by the creation of new splice sites in introns or exons), its translation (by amino-acid substitution encoding a chain terminator codon) or the stability of the molecule (Clarke and Higgins, 2000; Kaushansky et al., 2010). In consequence, the excess of unpaired globin chains can precipitate and damage the membrane of erythrocytes, causing their premature destruction (Steinberg et al., 2009).

Since α-globin is expressed on fetus and adults, α-thalassemia causes pre- and postnatal disease. The range of severity can go from mild anemia, also referred to as α-thalassemia trait, to Hb Bart's hydrops fetalis syndrome, where *HBA1* and *HBA2* genes are absent in both copies of chromosome 16, and therefore the fetus cannot synthesize α-globins to make HbF or HbA (Jobling et al., 2013; Old, 2003b).

β-Thalassemias are a very heterogeneous group of disorders, with more than 200 disease-causing mutations so far identified (Patrinos, 2004), including single nucleotide substitutions, and in a lower frequency than in α-thalassemias, insertions and deletions (Corrons, 2001; Old, 2003a). However, since many mutations are only prevalent in particular population groups, where the predominant ones account for the majority of mutation alleles and the less incidents are considered as rare mutations, at a worldwide scale only about 20 alleles account to the majority of all β-thalassemia (Kaushansky et al., 2010).

1.3.2. Abnormal Hemoglobins

More of 1000 Hb variants have been identified, arising as result of single nucleotide mutations affecting the α-, β-, δ- and ɣ-globin subunits, and therefore, modified HbA, HbA_2 or HbF (Kaushansky et al., 2010; Old, 2003a).

Depending on the hematological and/or clinical consequences, each variant can be sorted into different categories: variants that alter the hemoglobin physical/chemical properties; those from which result unstable proteins; others leading to Hb forms with altered oxygen affinity; M hemoglobins and variants that cause thalassemic phenotypes (Kaushansky et al., 2010).

The modified hemoglobins that have been more frequently studied are HbS, HbC, HbE, HbD_{Punjab} and HbO_{Arab} because of their clinical importance. HbS was the first to be identified at the molecular level by Pauling et al. (1949), who studied patients with Sickle Cell Disease (SCD), an anemia first reported in 1910 that is the most common of the severe structural Hb variants (Hoffbrand et al., 2010; Kaushansky et al., 2010)

1.4. Sickle Cell Disease

The sickle cell allele, *HBB*S*, characterized by the substitution of the nucleotide adenine by a thymine in the position 20 (c.20A>T) of the first exon of the *HBB* gene. This mutation leads to a β-globin chain where the amino-acid glutamic acid at position 6 is replaced by valine (β6Glu>Val) (Saiki et al., 1985).

The most common sickle disorder is Sickle Cell Anemia (SCA), which arises when an individual carries two copies of the mutant gene, and consequently HbS ($\alpha_2\beta^S_2$) is the produced hemoglobin. The disease expression is due to the polymerization of deoxygenated hemoglobin S (Bridges and Pearson, 2008).

In the pulmonary circulation, HbS molecules are in an oxygenated state because of the oxygen tension, presenting function and molecular structure similar to the HbA, and consequently erythrocytes display a typical conformation with a biconcave shape.

However, at the time of the release of the oxygen in the tissues, the protein undergoes a process of aggregation, resulting in filament structures consisting in linear polymers of HbS (Bridges and Pearson, 2008; Honig and Adams, 1986).

The filaments are further organized into fibers that stretch and distort the erythrocytes membrane resulting in the characteristic sickle-shape. This structural modification decreases the flexibility of RBC that allows to pass through the capillaries into larger venules, resulting in microvascular occlusion that interrupts the blood flow (Bridges and Pearson, 2008; Honig and Adams, 1986; Kaushansky et al., 2010).

Although the polymerization of hemoglobin doesn't happen suddenly after deoxygenation, and therefore the erythrocytes can still enter the pulmonary circulation where they become again oxygenated, the repeated sickling cycles contributes to the premature destruction of the cells (Honig and Adams, 1986).

Worldwide, >300 000 infants are born with SCA each year (Piel et al., 2013) and approximately 300 million people are heterozygous carriers of the mutation (HbAS). The heterozygous individuals are clinically described as having the Sickle Cell Trait, which, with rare exceptions, is asymptomatic (Kaushansky et al., 2010).

1.4.1. Other Genotypes Underlying Sickle Cell Disease

The coinheritance of HbS with other hemoglobin variant or β-thalassemias accounts for the remaining SCD, which encompass a number of sickling syndromes with clinical importance. The three most common are HbS/C (β6Glu>Val/β6Glu>Lys), which manifests with a moderate SCD phenotype, HbS/β^+-thalassemia and HbS/β^0-thalassemia, which are two severe SCDs.

HbC was the second variant to be described after HbS, and it is believed to have originated in the broad region encompassing Central West Africa and parts of West Africa, where it's prevalence can reach 28%.(Kaushansky et al., 2010).

Homozygous HBB*C do not suffer from vasoocclusion as HbS, but display a mild to moderate hemolytic anemia, not considered as a life-threatening disease, as a result of the formation of tetragonal crystals in erythrocytes. However the association of HbC and HbS prompts the reduced life expectancy of the double heterouzygous for both variants (Kaushansky et al., 2010; Vekilov, 2003).

Due to the higher prevalence of HbSC individuals contrasting to HbC, it's the reason why HbC is usual considered a sickle variant/disorder, even that those morphological events (polymerization of hemoglobin), are not associated to this variant.

Among other rarer forms are included HbS/D_{Punjab} (β6Glu>Val/β121Glu>Gln) and HbS/O_{Arab} (β6Glu>Val/β121Glu>Lys), both associated with severe phenotypes, or HbS/E

(β6Glu>Val/β26Glu>Lys), a SCD with mild clinical course (Bridges and Pearson, 2008; Kaushansky et al., 2010; Steinberg, 2009).

1.4.2. The Malaria Hypothesis

Although the majority of children born with SCA dies before the age of five due to the severity of the disease, still the frequencies of the *HBB*S* might reach remarkable high values in certain regions, contrary to what would be expected for genes under purifying selection (Piel et al., 2010). Haldane (1949) was the first to purpose an explanation to the paradoxical observations, suggesting that heterozygotes carriers of the *HBB*S* have protection against severe malaria, as soon would be sustained by the epidemiological works performed by Allison (1954).

Since then, growing evidence for the correlation between Sickle Cell Disease and protection against the *Plasmodium falciparum* infection prompted numerous studies in order to analyze the distribution of the SCD in the world that ultimately would lead to formulate “The Malaria Hypothesis” (Figure 3).

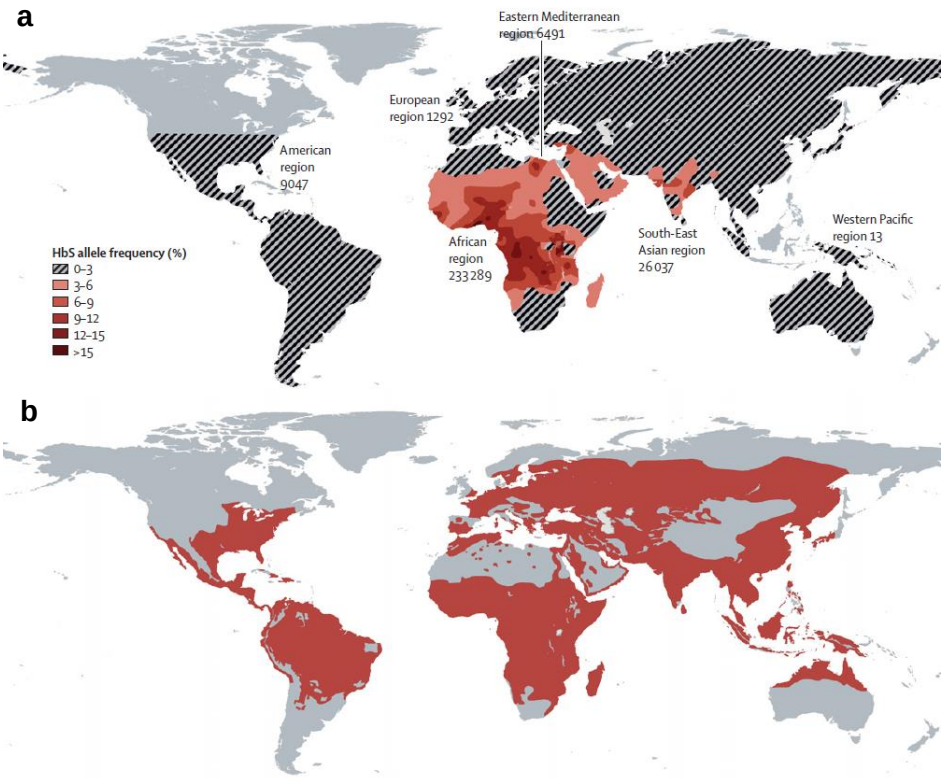

Figure 3 - Global distribution of *HBB*S* **(a)** and Malaria **(b)** (adapted from Rees et al. (2010)).

Evidence supporting such theory came from many studies. Besides the similarity of the distribution of malaria and HbS at local, regional and global scales, other evidence

was provided by clinical studies in endemic malaria regions and by a number of in-vivo and ex-vivo studies (Williams and Weatherall, 2012).

The sickle cell allele is widely distributed throughout sub-Saharan Africa (especially in tropical ecozone), the Middle East and parts of the Indian sub-continent, where the prevalence can range from 5% to 40% (Kaushansky et al., 2010; Weatherall and Clegg, 2001). Such geographical distribution have close resemblance to the prevalence of malaria cases, concerning which Africa accounts for 82% of the current world disease cases, and for 90% of world malaria deaths (Depetris-Chauvin and Weil, 2016).

As a result of migrations to western countries, SDC prevalence also increased in areas where malaria was not endemic (Kaushansky et al., 2010).

In many populations, besides the pressure of balancing selection, the high frequency of consanguineous marriages is also an important factor accounting for the increased frequency of the mutated allele (Williams and Weatherall, 2012).

In parallel with epidemiological studies, many investigations were undertaken in order to identify the exact mechanisms by which *HBB*S* provide protection against malaria. Although those mechanisms are not fully elucidated to the date, it is widely accepted that they probably includes both innate and immune-mediated mechanisms that induce increased phagocytosis of HbAS parasitized red blood cells (Luzzatto, 2012; Rees et al., 2010).

1.4.3. β^S - Globin Haplotypes and Evolution of the Sickle Cell Allele

Through the years, haplotyping of variants from the *beta*-globin gene cluster has revealed to be increasingly useful, allowing to address questions from the field of populations genetics but also to predict the severity of some hemoglobinopathies, like the SCD (Vinson et al., 2004).

It is estimated that human genomes differ one nucleotide every 1000-1500 bases. Those variations are called Single Nucleotide Polymorphisms (SNP's) and can be classified according to i) the region where they are located, ii) their consequences, or iii) the level of evolutionary conservation whereby they can be conservative or nonconservative (Buckingham and Flaws, 2007).

Groups of SNP's in the same chromosome that are in strong linkage disequilibrium (LD) define the so called haplotypes. They are a combination of alleles at different *loci* that are inherited together, unless recombination occurs (Buckingham and Flaws, 2007).

The first and still more conventional strategy used to identify haplotypes in the *beta*-globin gene cluster relies in the study of the Restriction Fragment Length Polymorphisms

(RFLP), and consists in the detection of the number and size of DNA fragments that result after digestion with restriction enzymes (Buckingham and Flaws, 2007). Therefore, variations at different SNP's can enable or prevent the recognition by specific enzymes, and as a consequence different patterns of restriction are associated to distinct haplotypes.

The first DNA haplotype in the β-cluster was described by Kan and Dozy (1978), who found that a *Hpa* I cleavage site was strongly associated with the *HBB*S*. Further works allowed the identification of another SNP's also in strong L.D. with the same allele, although only in some African populations (Pagnier et al., 1984).

As new SNPs were discovered, extending the haplotypic background associated to *HBB*S*, it was possible to identify distinct haplotypes typically found in specific populations. Accordingly, the main haplotypes become commonly known as: Benin (BEN), Bantu/Central African Republic (CAR), Senegal (SEN), Cameroon (CAM) and Arab-Indian, with the last being confined to Eastern Saudi Arabia and the Indian subcontinent (Table 1) (Crawford et al., 2002; Vinson et al., 2004).

When an haplotype doesn't fit in any of the above described, it is considered an atypical haplotype (Leal et al., 2016).

Table 1 - Restriction patterns of *HBB*S* associated haplotypes (adapted from Bitoungui et al. (2015)).

Enzymes / *Haplotypes*	*XmnI (5'Gγ)*	*HindIII (Gγ)*	*HindIII (Aγ)*	*HincII (3?'Ψβ)*	*HinfI (5'β)*
Senegal	+	+	−	+	+
Bantu/Central African Republic	−	+	−	−	−
Cameroon	−	+	+	+	+
Benin	−	−	−	+	−
Arab-Indian	+	+	−	+	−

(+)=Cut by a specific restriction endonuclease; (−)=Is not cut by that specific restriction endonuclease.

A few decades ago, the observation that the main *HBB*S* haplotypes were molecularly quite divergent while presenting a very well-defined geographic distribution (Figure 4), led to hypothesize that multiple *HBB*S* alleles were originated independently by recurrent mutation in the different regions where each haplotype is more prevalent (Kan and Dozy, 1980; Pagnier et al., 1984).

However, the question that still does not have a conclusive answer is whether the sickle mutation had indeed a multicentric origin, having appeared independently in different chromosome backgrounds, or rather its presence in different regions is due to a single origin with subsequent spread of the mutation by population admixture and people migration into regions where malaria was endemic within or outside Africa (Bitoungui et al., 2015; Kulozik et al., 1986).

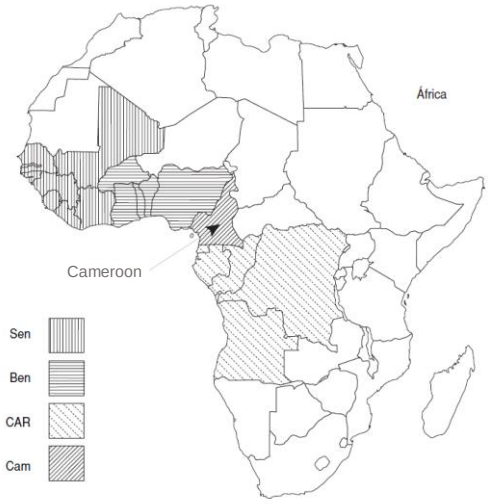

Figure 4 – Haplotypes distribution in African continent (adapted from Rodriguez Romero et al. (1998)).

In a seminal article of Livingstone (1958) a scenario was provided assuming a single origin of *HBB*S* in which balancing selection and gene flow alone could explain the current patterns of distribution of *HBB*S* in Africa, arguing that the spread of agriculture in the last few thousand years was responsible for the spread of the selective advantage of the sickle cell gene, and hence for the spread of the gene itself.

Moreover, the striking features of the entire β-cluster, including the presence of a recombination hotspot and frequent events of interallelic gene conversion that are able to generate highly divergent *HBB*S* haplotypes (Zago et al., 2000), were claimed to refute the multiple origin hypothesis and instead sustain that more likely *HBB*S* arose through a single mutation (Fullerton et al., 1994).

It was also proposed that a combination of recombination, gene mutation and/or conversion events is the best explanation for the origin of *HBB*S* (Antonarakis et al., 1984).

One of the weakest points of the multiple origin hypothesis is that it implies that several identical but independent mutations have occurred in a short period of time after the appearance of malaria in Africa, which seems unlikely given the low mutation rate of nuclear DNA (Currat et al., 2002). Despite that, the multicentric origin of *HBB*S* become indeed the most common view about the subject. Yet, it was not plenty sustained in a very recent comprehensive study addressing the global distribution of *HBB*S* haplotypes, leading the authors to emphasize that a careful reassessment of African *HBB*S* haplotypes is still needed to elucidate the evolutionary dynamics of the sickle allele, not discarding for now a single origin of the sickle mutation (Bitoungui et al., 2015).

Regardless of the evolutionary mechanisms from which arose the *HBB*S* haplotypes, since the referred to as typical haplotypes present a remarkable geographic specificity, they can give important insights to trace many episodes of the demographic history of a population.

In terms of clinical implications, some works emerged indicating that haplotypes of β-globin cluster might modify the severity of the Sickle Cell Disease. As a rule, carriers of *HBB*S* on Senegal and Arab-Indian haplotypes usually have high levels of fetal hemoglobin (> 15%) and concentration of $^{G}\gamma$-globin chains. Because high concentrations of HbF dilute the amount of HbS and disrupts its polymerisation, carriers of Senegal and Arab-Indian haplotypes are associated to reduced severity of phenotypes of the disease (Leal et al., 2016; Steinberg, 2009)

In opposition to the Senegal and Arab-Indian haplotypes, the Bantu/CAR was associated with decreased HbF levels (< 5%) and consequently with a severe SCD phenotype (Leal et al., 2016; Steinberg, 2009).

Reportedly, carriers of Benin and Cameroon have an intermediate clinical course, showing levels of HbF between 5% and 15% (Leal et al., 2016).

The relation between haplotypes and clinical severity, is basically dictated by HbF levels, which have been strongly associated with an specific polymorphism, the rs7482144, that in fact is a variant commonly used in the identification of the haplotypes. Nevertheless, other genetic modulatory elements of the level of fetal hemoglobin also exist, as indicates the variability in those levels among carriers of the same allele at that SNP (Steinberg, 2009).

1.5. Sickle Cell Disease in Portugal

While the high frequency of alleles responsible by SCD is largely explained by the selective advantage that carriers have in environments where falciparum malaria is endemic, due to the migration of peoples, those alleles might also become relatively common in areas where malaria is not endemic (Antonarakis et al., 1984; Kaushansky et al., 2010).

However, in Portugal malaria was endemic until mid-last century, and only in 1973 was considered eradicated by the World Health Organization after extensive campaigns for disease control (Bruce-Chwatt and de Zulueta, 1977; Landeiro, 1933).

Before the 1950s, estimates of 100,000 cases per year were usual, though unevenly distributed across the country (Landeiro, 1933). Then, several malariologic regions were documented, revealing different levels of endemicity. Among the regions with highest

incidence, was Alentejo/Ribatejo and the basins of the Tagus and Sado rivers where rice culture was a traditional activity (Bruce-Chwatt and de Zulueta, 1977).

Is in those regions where malaria used to be a serious health problem, that *HBB*S* reaches the highest values in the country. Since long that those regions were very scarcely populated, and in order to cope with the needs to work in the swampy rice fields on the banks of southern rivers, many African slaves were brought to the region, which partially explain the distribution of the Sickle Cell Trait in the region (Gallup and Sachs, 2001; Martins et al., 1993; Monteiro et al., 1989).

Much before, contacts with Mediterranean people who have settled in different areas across south Portugal due to the better conditions for agriculture, fishing and mining, might have also contributed to the introduction of hemoglobin variants in Portugal (Martins et al., 1993).

2. Objectives

In the present study it was intended the analysis of samples collected in Health Care Centers in Alentejo region in order to access the incidence of β-globin variants *HBB*S* and *HBB*C*, and the haplotypes distribution associated to the variants.

For the purpose, the following objectives were settled:

- Screening of sickle cell variants in samples from Coruche, Serpa and Alcácer do Sal;
- Identification of Single Nucleotide Polymorphisms associated to the main sickle cell haplotypes;
- Design of a new method for analysis of haplotypes associated to chromosomes with the sickle cell variations, based on an Multiplex SNaPshot® system;
- Inference of haplotypes in *HBB*S* bearing individuals;
- Comparison of the results with previous data regarding the incidence of Sickle Cell Disease (SCD) in Portugal and associated haplotypes.

3. Methodology

3.1. Sampling

In the present study, 266 blood samples previous collected in Health Care Centres from the Alentejo region (Portugal) were used to perform the screening of *HBB* gene mutations and subsequent identification of haplotypes associated to the variants. The samples belonged to unrelated individuals who had provided inform consent to participate in the study, allowing research use of their data. Out of them, 35 resided in Alcácer do Sal (Setúbal district), 71 in Serpa (Beja district) and 160 in Coruche (Santarém district).

3.2. Sickle Cell Variants Screening

3.2.1. DNA extraction

Genomic DNA was extracted and purified from dried blood stains collected on FTA® cards (Whatman®), using QIAamp® DNA Investigator Kit by QIAGEN® following the manufacture´s protocol.

3.2.2. PCR

In order to identify sickle cell mutations, the *HBB* genic region encompassing the sequence where the variants are located (chromosome 11:5227110-5226531) was amplified by Polymerase Chain Reaction (PCR), for which a pair of primers was selected flanking the region.

The two primers (Table 2) were designed according to the reference sequence of the *HBB* nucleotide (NM_000518) available in the platform Ensembl (Yates et al., 2015), after which their properties, like melting temperature, GC-content and self-complementarity, were analysed in the platform Primer3 (Untergasser et al., 2012). Then, the platform UCSC Genome Browser – Blat (Kent, 2002) available in http://genome.ucsc.edu/ was used to confirm the specificity of binding to the target genomic region.

Table 2 - Primers for amplification of the target region of *HBB* gene and expected fragment size.

Primer	Sequence (5'>3')	Fragment Length (bp)
HBB_FW	AGCAGGGAGGGCAGGAGCCAG	580
HBB_RV	CCCCTTCCTATGACATGAACTTAACC	

PCR reactions were prepared using 5 µL of HotStarTaq® Master Mix Kit (QIAGEN®), 0.5 µL of each primer (2 µM), 1 µL of Q-Solution (QIAGEN®) (final concentration 10%) and 3 µL of DNA, making a total of 10 µL per reaction. The control reactions were performed by substituting de DNA with distilled water (RNase/DNase/Protease free).

Amplification was conducted in a GeneAmp® PCR System 2700 (Applied Biosystems™) under the conditions presented in Table 3.

Table 3 - PCR protocol for amplification of the target region of *HBB* gene.

Step	Temperature (ºC)	Time	Cycles
Initial Denaturation	95	15 min	1
Denaturation	94	30 sec	10
Annealing	65	1:30 min	
Extension	72	1 min	
Denaturation	94	30 sec	30
Annealing	62	1:30 min	
Extension	72	1 min	
Final Extension	70	10 min	1
Hold	4	∞	

Amplification products were separated by horizontal polyacrylamide gel (8.4%) electrophoresis using a Multiphor II Electrophoresis System (GE Healthcare), with MultiTemp III Thermostatic Circulator (Amersham Biosciences) and Consort EV243 power supply. The running was performed at 180V with 0'GeneRuler 100 bp Plus DNA Ladder (Thermo Scientific™) and the results were visualized by silver staining method, with the following sequential steps: 10 minutes in Ethanol (10%); 5 minutes in Nitric Acid (1%); two washings with distilled water; 20 minutes in Silver Nitrate; two washings with distilled water; DNA visualization with solution made with 3g Sodium Carbonate (0.28M) + 1 mL Formaldehyde (0,02%) + 100 mL water; reaction termination with Acid Acetic (10%) and final washing with distilled water.

3.2.3. Automated DNA Sequencing

The amplified PCR products were submitted to purification by Sephadex™ G-50 Fine DNA Grade (GE Healthcare) gel filtration resin (following manufacture's protocol for columns preparation) and subsequently used in sequencing reactions. In each reaction it was used 2 µL of BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems™) (previously diluted 1:2 with Buffer by Applied Biosystems™), 0.5 µL of HBB_FW primer and 2.5 µL of the purified product.

The PCR reaction was performed in GeneAmp® PCR System 2700 (Applied Biosystems™) under the conditions presented in Table 4.

Fragments were analysed in an ABI PRISM 3130xl analyser (Applied Biosystems™) after purification with Sephadex™ G-50 (GE Healthcare) gel filtration resin and posterior addition of 10 µL of Hi-Di™ Formamide (Applied Biosystems™). Sequences were aligned and analysed using Geneious v.5.5 analysis software (Biomatters, available from http://www.geneious.com/).

Table 4 - PCR protocol for sequencing of the amplified target region of *HBB* gene.

Step	Temperature (ºC)	Time	Cycles
Initial	96	2 min	1
Denaturation	96	15 sec	
Annealing	58	10 sec	35
Extension	60	2 min	
Final Extension	60	10 min	1
Hold	4	∞	

3.3. Genotyping of SNPs Defining Haplotypes

3.3.1. Selection of SNPs

A comprehensive bibliographic research about the haplotypes associated to the sickle cell mutation(s) was performed in order to identify the single nucleotide polymorphisms (SNPs) to be examined with a new approach we sought to develop in this project that consisted in a SNaPshot® Multiplex system.

For that purpose, a reference sequence of the β-cluster was created using the platform Ensembl (Yates et al., 2015), comprising the nucleotide sequences of the genes *HBE1, HBG2, HBG1, HBBP1, HBD* and *HBB* (11:5271844 – 11:5222404). Since the conventional methodology for the identification of *HBB*S* haplotypes relies in the analysis of Restriction Fragment Length Polymorphism (RFLP), for which the primers commonly used were often described, we have begun by identifying the sequences of the fragments that typically would be amplified to be submitted to restriction enzymes digestion. Next, and given that the widely used restriction enzymes were also known, the exact location of the most informative polymorphisms defining the haplotypes were determined, ending up with a total of 6 SNPs: rs7482144, rs113425530, rs2070972, rs10128556, rs968857 and rs16911905 (Figure 5).

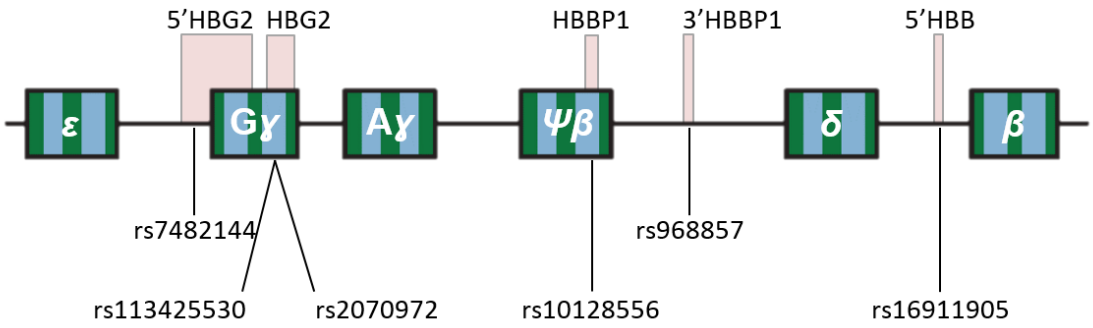

Figure 5 – Representation of the β-globin cluster, the target regions to be amplified (rose squares) encompassing the positions of six polymorphisms that define the haplotypes associated to the sickle cell variant.

3.3.2. Multiplex PCR

Five pairs of new primers were designed in order to amplify the fragments containing the 6 target SNPs (Table 5) . In order to implement a Multiplex PCR, primers for amplification of fragments with different lengths were chosen with the assistance of platforms Primer3 (Untergasser et al., 2012), UCSC Genome Browser – Blat (Kent, 2002) and In-Silico PCR (both available in http://genome.ucsc.edu/), and the AutoDimer software (available from http://www.cstl.nist.gov/div831/strbase/AutoDimerHomepage/AutoDimerProgramHomepage.htm).

Table 5 – Primers for amplification of the target regions of β-globin cluster and expected fragment sizes.

Primer	Sequence (5'>3')	Fragment Length (bp)
5'HBG2_FW	ACAAGAAGGTGAAAAACGG	772
5'HBG2_RV	CTTTATGGCATCTCCCAAG	
HBG2_FW	GCTGCAAGAAGAACAACTACC	400
HBG2_RV	GACAACCATGTGTGATCTCTTA	
HBBP1_FW	CAGGATTCTTTGTTATGAGTGTT	332
HBBP1_RV	CAAGCTGGACTTGCAGTAA	
3'HBBP1_FW	GAGACCTAAACTGAGGAACCTT	208
3'HBBP1_RV	CTTGATGGACCCTAACTGATATA	
5'HBB_FW	GATCACGTTGGGAAGCTATA	166
5'HBB_RV	AGGTCTTCTACTTGGCTCAGA	

PCR reactions were performed in a final volume of 10 µL, containing 5 µL of Qiagen® Multiplex PCR Kit, 1 µL of Primer Mix, 3.5 µL of distilled water (RNase/DNase/Protease free) and 0.5 µL of DNA, under the conditions presented in Table 6. The Primer Mix was

prepared with a final concentration of 2 µM for each primer. Reactions were conducted in GeneAmp® PCR System 2700 (Applied Biosystems™).

Amplification products were separated by horizontal polyacrylamide gel (8.4%) electrophoresis, after which DNA was visualized by the silver staining method previous described in section 3.2.2.

Table 6 – Multiplex PCR protocol for amplification of the target regions of β-globin cluster.

Step	Temperature (ºC)	Time	Cycles
Initial Denaturation	95	15 min	1
Denaturation	94	30 sec	35
Annealing	60	1:30 min	
Extension	72	1 min	
Final Extension	72	10 min	1
Hold	4	∞	

3.3.3. SNaPshot®

For the sake of developing a SNaPshot® Multiplex system to determine simultaneously the genotypes of the 6 target SNPs, Single Base Extension (SBE) primers were designed with the assistance of the platforms already mentioned for the construction of other primers (section 3.3.2). The SBEs hybridize to the sequence immediately adjacent to the target nucleotide base, and for the separation of each primer by capillary electrophoresis, non-annealing tails with different sizes were added to each of the 6 SBE primers. The sequences of the primers are described in Table 7, as well are the possible nucleotides that will be added to the terminal 3' end of each SBE depending on the genotypes at the different SNPs.

Table 7 – Single Base Extension primers for the genotyping of the target polymorphisms.

rs	Polymorphism	Detected Allele	Sequence (5'>3')
7482144	G>A	G/A	GGTGGAGTTTAGCCAGG
113425530	C>A	G/T	GTCGTGAAAGTCTGACAATTGATTCTGGG TGGAA
2070972	A>C	A/C	GACTAAACTAGGTGCCACGTCGTGAAAGT CTGACAACCTCCAGATAACTACACACC
10128556	C>T	G/A	GTCTGACAATGTTGGGGTAGTGAGTTG
968857	T>C	T/C	CAATGCATGACACATGCTTG
16911905	G>C	G/C	TGCCACGTCGTGAAAGTCTGACAACGTTT TAAAATCATTTCCTT

A purification step was performed in order to optimize the samples for the SNaPshot® reactions. It consisted in adding to each 1 µL of amplification product, 0.5 µL of ExoSAP-IT™ (Applied Biosystems™), and then the solution was submitted to the conditions presented in Table 8 - Purification 1.

After that, to each purified sample (1.5 µL), 1 µL of SNaPshot™ Multiplex Kit (Applied Biosystems™), 1 µL of SBE primer Mix, and 1.5 µL of distilled water (RNase/DNase/Protease free) was added. The SBE primer Mix was prepared in order to obtain the following concentrations in the final solution (5 µL): rs7482144 at 0.8 µM, rs113425530 at 1.4 µM, rs2070972 at 0.8 µM, rs10128556 at 0.3 µM, rs968857 at 0.6 µM and rs16911905 at 0.6 µM. The SNaPshot® reactions were performed under the conditions described in Table 8 – SNaPshot®.

A final purification step (Table 8 - Purification 2) was performed by adding up 1 µL of FastAP (Thermo Scientific™) to the final product.

All the described reactions were conducted in GeneAmp® PCR System 2700 (Applied Biosystems™).

Table 8 – Protocols for the purification steps and genotyping of the target polymorphisms.

Purification 1		**SNaPshot®**		**Purification 2**	
Temperature (ºC)	**Time**	**Temperature (ºC)**	**Time**	**Temperature (ºC)**	**Time**
37	15 min	96	10 sec	37	1 h
85	15 min	50	5 sec	85	15 min
		60	30 sec		
		25 cycles			

In order to be analyzed by capillary electrophoresis in an ABI PRISM 3130xl analyser (Applied Biosystems™), 1 µL of the SNaPshot® purified products were mixed with 12 µL of Hi-Di™ Formamide (Applied Biosystems™) and GeneScan™ – 120 LIZ™ Size Standard (Applied Biosystems™). The results were analyzed with the GeneMapper® v.4.0 software.

3.4. Haplotype Determination

Haplotype inference and Hardy-Weinberg equilibrium tests were performed with the Arlequin 3.5v software (Excoffier and Lischer, 2010).

The correspondence here assumed between detected haplotypes and the commonly referred to as Bantu, Benin, Senegal, Arab-Indian and Cameroon haplotypes, is summarized in Table 9.

Both rs113425530 and rs2070972 are positioned in the same restriction enzyme recognition site when RFLP analysis is performed, therefore the genotypes of the two variants should be analyzed as a single result. The presence of nucleotide T in the first SNP **or** C in the second (where the presence of just one of them would be associated to the non-recognition by the restriction enzyme), should be linked to the Benin haplotype and consequently three genotypes can be associated to the same.

Table 9 – Single Nucleotide Polymorphisms alleles associated to the main sickle cell haplotypes.

	rs7482144	**rs113425530**	**rs2070972**	**rs10128556**	**rs968857**	**rs16911905**
Bantu/CAR	G	G	A	G	C	C
Benin	G	T	C	G	T	C
	G	T	A	G	T	C
	G	G	C	G	T	C
Senegal	A	G	A	A	T	G
Arab-Indian	A	G	A	A	T	C
Cameroon	G	G	A	G	T	G

4. Results and Discussion

4.1. Genotyping

4.1.1. Sickle Cell Variants

In this work, three population samples from Alcácer do Sal, Coruche and Serpa were addressed. The demographic characterization of the studied populations is summarized in Table 10, and more detailed information is presented in Table S1.

Table 10 – Overall characteristics of the studied populations.

Location	Gender	n	Birth Year
Alcácer do Sal (Setúbal)	Male	13	1933-2001
	Female	22	1932-1987
Serpa (Beja)	Male	68	1917-2005
	Female	3	1927-1987
Coruche (Santarém)	Male	160	1918-2006
	Female	0	-

Samples had been randomly collected among the users of the Health Care Centers from Alcácer do Sal, Coruche and Serpa. No inclusion criterion regarding health status was applied, meaning, thus, that patients with hemoglobinopathies were not deliberately included nor exclude from the study. However, three of the 35 individuals from Alcácer do Sal and nine of a total of 71 from Serpa were already referenced in the respective centers as having sickle cell trait/anemia.

Sampling had been carried out in the scope of a FCT funded Project - Deciphering the origin of foci of high prevalence of hereditary anemias in Portugal: epidemiological and evolutionary study - POCI/ANT/57037, that involved the genetic characterization of the three populations with lineage markers (i.e. mitochondrial DNA and Y-chromosome). For that reason, sampling of male individuals had been prioritized, a recommendation that was strictly followed in Coruche. For that reason, males are clearly overrepresented on the whole of the three samples here analysed.

From a total of 266 samples screened for *HBB*, 19 were positive for the presence of the c.20A>T mutation, corresponding to the single nucleotide polymorphism rs334 that define allele *HBB*S* (β^S) (Figure 6). Two subjects from Alcácer do Sal, nine from Serpa and seven from Coruche, giving a total of 18 individuals, displayed the allele in heterozygosity, being therefore carriers of Sickle Cell Trait, whereas one subject from Serpa was homozygous, harbouring thus Sickle Cell Anemia. The observed genotypic

distributions and frequency estimates of β^S in the three samples are presented in Table 11.

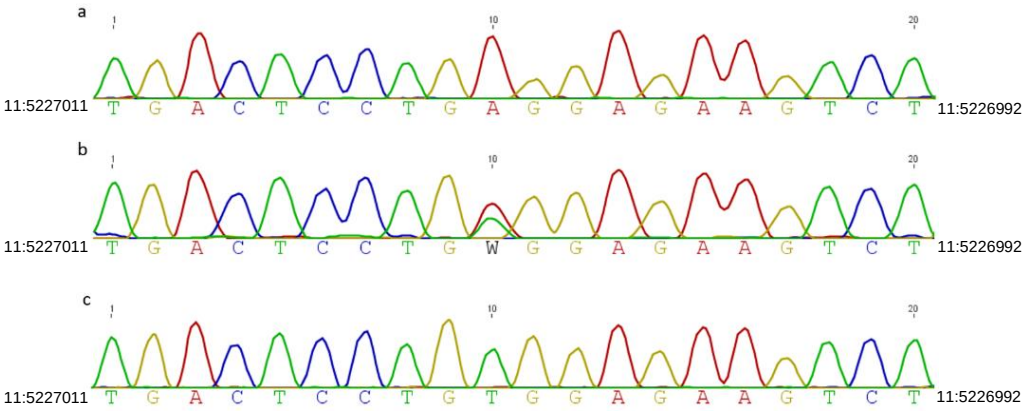

Figure 6 - Example of sequencing results in sickle variants screening; c.20A>T mutation represented on position 10 of the electropherogram. **a –** AA; **b –** AS; **c –** SS.

The two heterozygous from Alcácer do Sal corresponded to the subjects already diagnosed with Sickle Cell Trait. Among the nine subjects from Serpa with information of having Sickle Cell Trait/Anemia, eight were found to be heterozygous HbAS and one homozygous HbS. Additionally in the total sample from Serpa, another HbAS subject was detected who had no indication of being sickle cell carrier.

Table 11 – Globin genotypes of the studied samples and frequencies of each genotype and β^S allele in the different districts of Alentejo region.

Location	Globin Genotype	n	Globin Genotype Frequency (%)	β^S allele Frequency (%)
Alcácer do Sal (Setúbal)	AA	33	94.3	2.9
	AS	2	5.7	
Serpa (Beja)	AA	61	85.9	7.7
	AS	9	12.7	
	SS	1	1.4	
Coruche (Santarém)	AA	153	95.6	2.2
	AS	7	4.4	
Alentejo (all districts considered)	AA	247	92.8	3.8
	AS	18	6.8	
	SS	1	0.4	

The *HBB*C* mutation, which results from the missense variant c.19G>A (rs33930165) was not detected in any of the screened individuals.

One of the most comprehensive studies regarding the prevalence of Hb disorders in Portugal dates back from 1993 (Martins et al.). Then, 15208 samples from the entire

country where analysed and the results were presented considering a geographical distribution by administrative districts. In Portugal as a whole, the frequency of the Sickle Cell Trait was reported to be 0.32%, but its prevalence clearly increased from the north towards the south. In the north, the Sickle Cell Trait was not detected in the districts of Braga, Bragança, Viana do Castelo, Vila Real and Viseu, and only in Porto was found at very low prevalence of 0.05%. Contrarily, the districts of Beja, Setúbal and Santarém presented the highest values of 1.11%, 0.99% and 0.71% respectively, the latter ex-quo with Évora, while in Faro the frequency of *HBB*S* carriers was 0.68%.

Much more recently, a project was conducted involving the screening of hemoglobinopathies only in the centre regions of Portugal, reinforced that in central Portugal the frequency of *HBB*S* was ~0.22%, which is an intermediate value between those common in North and South of the country (Campos, 2007).

In the light of this data, the proportion here estimated of heterozygous carriers of *HBB*S* either in Alcácer do Sal, from the Setúbal district, 5.7%, Coruche from the Santarém district, 4.4%, or Serpa from the Beja district, 12.7%, is remarkably higher than found in Martins et al. (1993) work. Actually, however, the discrepancy was not unexpected . First because samples from this study are not representative of the entire districts, but instead just encompass specific sub-regions within each one, contrarily to the study of Martins et al. (1993), which had a wider coverage in terms of geographic area and so naturally in terms of samples sizes. Second, and more importantly, because the samples here analysed were deliberately collected in sub-regions known to be foci of sickle cell carriers, as already highlighted in Martins et al. (1993) (though not discriminating the data for those foci), or before in Monteiro et al. (1989), who restrictively surveyed Coruche, where 8.2% heterozygous HbAS were found out, leading to estimate at 4.1% the frequency of *HBB*S*.

Thus, the results here obtained are not only in agreement with the described in literature, but also shows that in 2007, the year during which most samples examined in this work were collected, Alcácer do Sal, Coruche and Serpa continued to be hot spots of *HBB*S* frequency, meaning thus that at least between 1989 and 2007, the genetic epidemiology of Sickle Cell Trait/Anemia did not changed substantially in the South of Portugal.

In Serpa, the frequency of *HBB*S* reached the highest frequency, attaining the very unusual value of 7.7%. Besides the elevated proportion of heterozygous (12.7%), it was from Serpa the unique homozygous HbS detected in this study, accounting to 1.4% of the sample. In the absence of further published data for Serpa, it is impossible to evaluate the consistency of the data now obtained. However, we cannot exclude that the high frequency might had been inflated, in a certain extent, by some bias in the sampling

process. In fact, in that sample were only enrolled subjects from Pias, which is a very small parish from Serpa, being likely therefore that many related individuals live there.

Tests of Hardy-Weinberg Equilibrium (HWE) performed for rs334 variant in the three population samples here analysed, did not revealed any significant departures from the HWE expectations (Table 12).

Table 12 - Hardy-Weinberg test results of polymorphism rs334 in Alcácer do Sal, Serpa and Coruche.

Polymorphism	Location	Observed Heterozygosity	Expected Heterozygosity	*p* - value	s. d.
rs334	Alcácer do Sal	0.05714	0.05631	1.00000	0.00000
	Serpa	0.12676	0.14394	0.34121	0.00050
	Coruche	0.04372	0.04293	1.00000	0.00000

4.1.2. Other Polymorphisms

The primers designed flanking the first and second exons of *HBB* gene, allowed to examine the entire amplified region, encompassing exon 1 to the beginning of intron (IVS; InterVening Sequencing) 2, where it was possible to detect other polymorphisms besides the ones responsible for the HbS and HbC variants (Figure 7).

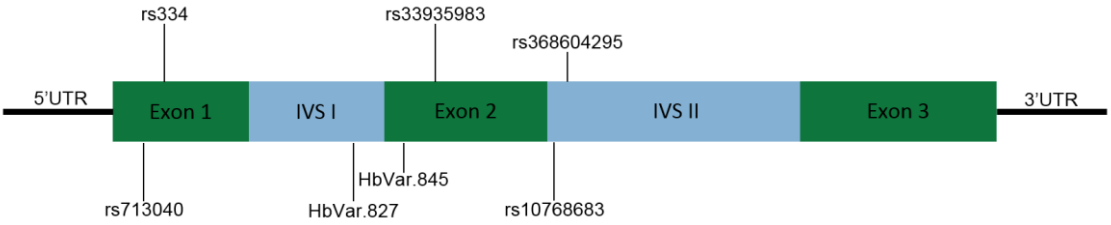

Figure 7 - Representation of *HBB* gene and position of the detected polymorphisms besides rs334.

Two of those polymorphisms were the c.9T>C and c.315+16G>C, corresponding to the SNPs rs713040 and rs10768683 respectively. Both were already contained in the HbVar database (Patrinos, 2004) under the category “rare variants”, without indication of any associated repercussion. The first one is a synonymous mutation, not implying any amino-acid change. The second one is located in the second intron of the *HBB* gene, to wich no clinical consequences are associated according to HbVar, and not leading to any splice alteration as predicted by ESEfinder v.3.0 (Smith et al., 2006).

In the three population samples here analysed, it caught attention the strong relation between the two polymorphisms, since the presence of one allele at one SNP

determined the presence of the allele at the other SNP, in such a way that all homozygous for the derived variant at one SNP were also homozygous for the derived variant at the second, and heterozygous or homozygous to the ancestral alleles displayed identical correlation. The two SNPs were therefore submitted to the analysis of Linkage Disequilibrium (L.D.), which is the non-random association of alleles at different *loci*, having revealed to be in total L.D. (r^2= 100%).

Regarding the genotypic distribution, in the pooled samples, 2.6% were homozygous TT at rs713040 and GG at rs10768683, 20.3% heterozygous at both SNPs and 77.1% homozygous CC also at the two SNPs. Gene frequency of the minor allele at each *loci* was 0.128, which is similar to the values found among the European populations from the 1000 genomes Phase 3 (0.131-0.227) (Yates et al., 2015).

Two other SNPs, rs368604295 and rs33935983, were both found in one heterozygous sample each. None was contained in HbVar (Patrinos, 2004). The first is a variant within the second intron (c.315+26T>G) that not interferes with mRNA splicing, according to the predictions in ESEfinder v.3.0 web resource (Smith et al., 2006) so without any consequences for the normal production of the β-globin subunits. The second is a missense variant in the second exon (c.169G>A), which results in the substitution of amino-acid glycine to serine in translation. Although the final product will be different, the variation does not affect the structure or function of the protein according to PolyPhen (Adzhubei et al., 2010) predictions, that scored the polymorphism as benign and therefore no clinical consequences are expected (Yates et al., 2015).

In addition, two polymorphisms with clinical relevance were found in this study: the rs35004220/HbVar.827 (IVS I-110 (G>A)) and the rs11549407/HbVar.845 (CD39 (C>T)), detected in two and five heterozygous individuals, respectively (Figure 8). The two heterozygous for HbVar.827 were from Coruche, whereas among the heterozygous for HbVar.845, one was from Alcácer (actually the subject was referenced as possible *HBB*S* carrier) one from Serpa and three from Coruche. None of these seven individuals harboured sickle cell mutations.

Both HbVar.827 and HbVar.845 result in a phenotype of thalassemia however with different characteristics.

The variation at SNP rs35004220 is associated with β^+-thalassemia, where the production of the protein is still active but in lower concentrations. Reportedly, the mutation is responsible for the creation of a new splice site that leads to production of about 90% of abnormally spliced mRNA, while only ~10% of the mRNA is normal (Campos, 2007).

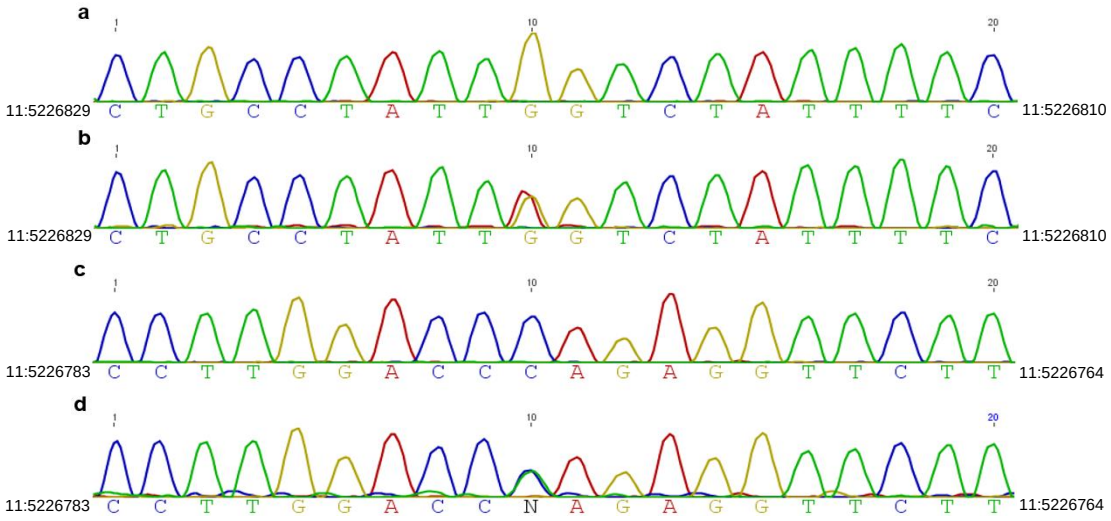

Figure 8 - Examples of β-thalassemias screening results. **a –** Homozygous G/G; **b –** Heterozygous HbVar.827; **c –** Homozygous C/C ; **d –** Heterozygous HbVar.845.

Although this mutation has a wide ethnic distribution, it is one of the most common β-thalassemia variants in the Mediterranean countries, representing 8.57% of β-thalassemia variants in Portugal (Patrinos, 2004).

The HbVar.845 is responsible by a β^0-thalassemia phenotype, where the production of β-globin is completely abolished. This mutation causes the substitution of codon 39 by a premature STOP codon, with subsequent termination of the mRNA translation.

Similarly to the other detected β-thalassemia variant, also HbVar.845 was found to be present in various ethnics backgrounds, but its prevalence is highest in populations from the Mediterranean regions, including Portugal where it accounts to over one-third of β-thalassemia variants (Almeida, 2015; Patrinos, 2004).

Former studies were conducted addressing β-thalassemias in Portugal, although the most recent published data regarding the south districts of the country are from the 90's of the last century (Faustino et al., 1999; Martins et al., 1993; Ribeiro et al., 1997). The panorama for the centre of Portugal was updated in 2007 with the study of Campos (2007).

According to Ribeiro et al. (1997), the reported contribution of five different mutations for the thalassemia phenotype in the south of Portugal was: IVS I-1 (G>A) 52.7%, CD39 (C>T) 30.6%, IVS I-110 (G>A) 8.3%, IVS I-6 (T>C) 2.8% and CD15 (TGG>TGA) 2.8%. The results here obtained meet these data in the sense that the two identified thalassemia mutation are among the most common in South Portugal. However, contrarily to the findings of Ribeiro et al. (1997), the IVS I-1 (G>A) mutation, refereed as the commonest in their work, was not detected in any of the samples here analysed.

Even though it was not possible to analyse the entire *HBB* gene, essentially due to time limitations, the results obtained already afford an important view on the contribution of thalassemias to the burden of hemoglobinopathies in the south of country.

4.2. SNP Genotyping

The implementation of a new method for analysis of haplotypic data associated to the sickle cell *HBB*S* variant was one of the principal aims of the project, since to the date, such studies were performed primarily by Restriction Fragment Length Polymorphism analysis. Although a similar approach had already been applied, the number of polymorphisms analysed in Multiplex system was too low, which highly limited the haplotyping resolution (Crawford et al., 2002).

After a throughout compilation of published studies regarding the SNPs in L.D. with the sickle cell allele, a primarily selection of the most relevant was performed. The criteria for choosing the polymorphisms was in order to permit the analysis of the minimum number of variants without losing the ability to identify unambiguously the associated sickle cell haplotypes.

Among the initially selected SNPs, it was included a SNP located in *HBG1* ($^{A}\gamma$) gene, once it could provide some supplementary information for the haplotyping. However due to the high homology between *HBG1* ($^{A}\gamma$) and *HBG2* ($^{G}\gamma$), a gene that also contained other of the targets SNPs, it was not possible to design an SNaPshot® system that would differentiate the polymorphisms of both genes. For that reason, the less informative of the two SNPs, precisely that within *HBG1*, was not included in the Multiplex system.

The main obstacle to implement the planned strategy allowing the study of the entire range of SNPs that define the *HBB*S* haplotypes, was the homology between the genes in the β-cluster, and consequently, the design of primers specifically to the target variations. Even though, such challenge was finally overcome and the SNaPshot® Multiplex technique was successfully developed for the simultaneous screening of the following 6 SNPs: rs7482144, rs113425530, rs2070972, rs10128556, rs968857 and rs16911905. A representative result of the Multiplex PCR and the SNaPshot® results of two unrelated individuals are illustrated in Figure 9 and Figure 10 respectively.

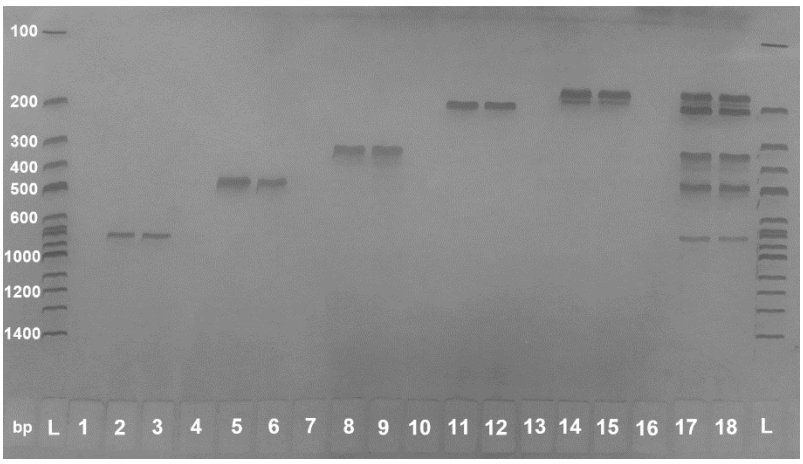

Figure 9 – Multiplex PCR products; **L –** Ladder; **1 –** 5'HBG2 Control; **2,3 –** 5'HBG2; **4 –** HBG2 Control; **5,6 –** HBG2; **7 –** HBBP1 Control; **8,9 –** HBBP1; **10 –** 3'HBBP1 Control; **11,12 –** 3'HBBP1; **13 –** 5'HBB Control; **14,15 –** 5'HBB; **16 –** Multiplex PCR Control; **17,18 –** Multiplex PCR.

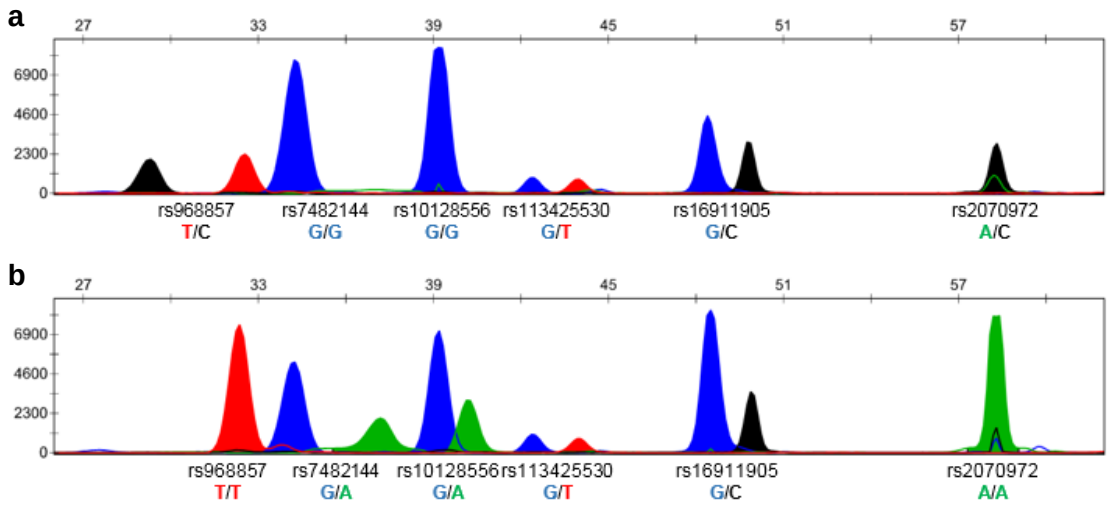

Figure 10 - Electropherograms examples of SnaPshot® reaction results and prediction of genotypes associated to each variants. **a –** sample SH/08; **b –** sample SH/07.

To achieve clear genotyping results as displayed in Figure 10, the steps that precede the SnaPshot® reaction are critical, specifically the amplification of the DNA fragments by Multiplex PCR. An insufficient genome amplification will affect later the efficacy of the SNaPshot® step, and thus particular attention was given to optimize the system of 6-plex multiplex.

All samples with positive result for the presence of the c.20A>T mutation were submitted to the developed approach. The genotyping results are presented in Table 13.

Table 13 – Single Nucleotide Polymorphisms genotypes obtained through SNaPshot® reactions, in samples with the sickle cell variant. **AH** – Alcácer do Sal; **SH/SC** – Serpa; **CC** – Coruche.

	Polymorphisms					
Sample ID	**rs7482144**	**rs113425530**	**rs2070972**	**rs10128556**	**rs968857**	**rs16911905**
AH/02	G/A	G/G	A/A	G/A	T/C	G/C
AH/03	G/A	G/G	A/A	G/A	T/C	G/C
SH/01	G/G	G/T	A/C	G/G	T/C	G/C
SH/02	G/G	T/T	A/A	G/G	T/T	C/C
SH/03	G/G	T/T	A/A	G/G	T/T	G/C
SH/04	G/G	G/T	A/C	G/G	T/C	G/C
SH/05	G/G	G/T	A/A	G/G	T/T	G/C
SH/06	G/G	G/T	A/C	G/G	T/C	G/C
SH/07	G/G	G/T	A/C	G/G	T/C	G/C
SH/08	G/A	G/T	A/A	G/A	T/T	G/C
SH/09	G/G	G/T	A/C	G/G	T/C	G/C
SC/60	G/A	G/T	A/A	G/A	T/T	C/C
CC/04	G/G	G/G	A/C	G/G	C/C	G/C
CC/09	G/A	G/G	A/C	G/A	T/C	G/G
CC/16	G/A	G/G	A/C	G/A	T/C	G/G
CC/24	G/A	G/G	A/C	G/A	T/C	G/G
CC/25	G/A	G/G	A/A	A/A	T/T	G/G
CC/45	G/G	G/G	A/A	G/G	T/C	G/C
CC/151	G/A	G/G	A/A	G/A	T/T	G/G

To evaluate the accuracy of the results provided by the technique, 16 samples previously haplotyped through conventional RFLP analysis in Instituto Nacional de Saúde Doutor Ricardo Jorge (INSA) were used as control. The control samples were from individual with Sickle Cell Disease, except P7, P8, P11, P15 and P16 that were carriers of two distinct hemoglobinopathies: Sickle Cell Trait and β-thalasemia.

The genotypes here obtained by SNaPshot® Multiplex are presented in Table 14.

Table 14 - Single Nucleotide Polymorphisms genotypes obtained through SNaPshot® reactions, in 16 control samples.

	Polymorphisms					
Sample ID	**rs7482144**	**rs113425530**	**rs2070972**	**rs10128556**	**rs968857**	**rs16911905**
P1	G/A	G/G	A/A	G/A	T/C	G/C
P2	G/A	G/T	A/A	G/A	T/T	G/C
P3	G/A	G/T	A/A	G/A	T/T	G/C
P4	G/A	G/G	A/A	G/A	T/C	G/C
P7	G/G	G/G	A/C	G/G	C/C	G/C
P8	G/G	G/T	A/A	G/G	T/T	G/C
P9	G/G	G/G	A/A	G/G	C/C	C/C
P10	G/A	G/G	A/A	G/A	T/C	G/C
P11	G/G	G/G	A/C	G/G	C/C	G/C
P12	G/G	G/G	A/A	G/G	C/C	C/C
P14	G/A	G/G	A/A	G/A	T/C	G/C
P15	G/G	G/G	A/C	G/G	C/C	G/C
P16	G/A	G/G	A/A	G/A	T/T	G/C
P17	G/G	G/T	A/A	G/G	T/C	C/C
P19	G/G	T/T	A/A	G/G	T/T	C/C
P21	G/G	G/T	A/A	G/G	T/C	C/C

In order to evaluate the strength of the association between the studied variants, Linkage Disequilibrium analysis was performed in the platform Ensembl (Yates et al., 2015). As expected, the pairwise L.D. values were statistically significant in the populations from the 1000 genomes project with data available, as is summarized in Table 15.

Table 15 – Linkage Disequilibrium (L.D.) results of the studied SNPs. **(+)** – Significant L.D.; **nd** – No data.

	rs7482144	**rs113425530**	**rs2070972**	**rs10128556**	**rs968857**	**rs16911905**	**rs334**
rs7482144	*	nd	nd	nd	nd	nd	nd
rs113425530	nd	*	+	+	+	+	+
rs2070972	nd	+	*	+	+	+	+
rs10128556	nd	+	+	*	+	+	+
rs968857	nd	+	+	+	*	+	+
rs16911905	nd	+	+	+	+	*	+
rs334	nd	+	+	+	+	+	*

4.3. β^S-Globin Haplotypes

From the raw genotypic data, haplotypes were estimated with the ELB algorithm implemented in Arlequin 3.5v software, which inferred a total of nine haplotypes in the samples from Alcácer, Coruche and Serpa harboring the sickle cell mutation (Table 16). Since those samples were heterozygous for the sickle cell mutation, except one that was homozygous, the genotypes for rs334 were also included in Arlequin input file, and therefore, the inferred haplotypes anchored in the *HBB*S* background were only 3: Haplotypes ID 2, 3 and 8, in Table 16.

Table 16 – Inferred haplotypes associated to the 19 studied samples, by Arlequin 3.5v software, respectively frequency and identification of the main haplotype associated. Haplotypes defined by polymorphisms in the following order: rs7482144, rs113425530, rs2070972, rs10128556, rs968857, rs16911905 and rs334.

Haplotype ID	Haplotype	Frequency	Observations
1	AGAATGA	0.078947	
2	GGAGCC**T**	0.105263	Haplotype Bantu
3	GTAGTC**T**	0.289474	Haplotype Benin
4	GGCGCGA	0.236842	
5	GTAGTGA	0.026316	
6	GGAGTGA	0.078947	
7	AGAATCA	0.026316	
8	AGAATG**T**	0.131579	Haplotype Senegal
9	GGAATGA	0.026316	

According to the assumed correspondence with the conventional haplotype nomenclature (section 3.4 – Table 9), the inferred haplotypes 2, 3 and 8 matched the haplotypes Bantu, Benin and Senegal, respectively. None of the commonly referred as atypical haplotype was detected.

Concerning the samples used as controls, that is, those in which haplotypes had been previously determined by RFLP typing, from the new genotypic data here obtained with the SNaPshot® system, haplotypes were also inferred as above mentioned, and the results are presented in Table 17. For all samples, the haplotypes inferred coincided with those based on RFPL analysis, an observation that clearly sustains the quality of the implemented SNaPshot® methodology.

Table 17 – *HBB*S* haplotypes associated to the control samples. **T**-Thalassemia.

Sample ID	*HBB*S* Haplotype
P1	Senegal/Bantu
P2	Senegal/Benin
P3	Benin/Senegal
P4	Bantu/Senegal
P7	Bantu/T
P8	Benin/T
P9	Bantu/Bantu
P10	Senegal/Bantu
P11	Bantu/T
P12	Bantu/Bantu
P14	Bantu/Senegal
P15	Bantu/T
P16	Senegal/T
P17	Bantu/Benin
P19	Benin/Benin
P21	Benin/Bantu

Among the 19 individuals from Alcácer do Sal, Coruche and Serpa bearing the sickle cell gene, the Benin accounted to 29% of total haplotypes, the Senegal to 13.1%, and the Bantu to 10.5% (Table 18).

In previous studies addressing *HBB*S* haplotypes in Portugal, also the Benin, the Bantu and the Senegal haplotypes were uniquely detected (Lavinha et al., 1992; Monteiro et al., 1989).

However, an interesting finding in this study respects the distribution of *HBB*S* haplotypes in the three population samples: the Benin haplotype was uniquely detected in Serpa where no other haplotype was found among the 11 chromosomes harbouring *HBB*S*; the Senegal haplotype was only found in Coruche, in five out the seven *HBB*S* chromosomes; the remaining two haplotypes from Coruche were Bantu, as also were the two *HBB*S* chromosomes from Alcácer do Sal. This distribution suggest that the introduction of *HBB*S* in Serpa may have been due to an importation event distinct from that (or those) responsible by the presence of *HBB*S* in Alcácer and Coruche.

In order to further explore this question, and given that the samples from Alcácer, Coruche and Serpa were previously characterised for mitochondrial DNA (mtDNA) and Y-chromosome lineages (Pereira et al. (2010)), the results on those markers were recruited for *HBB*S* haplotypes here identified (Table 18).

Table 18 – *HBB*S*, mitochondrial and Y-chromosome haplotypes associated to samples with rs334 variant. **AH** – Alcácer do Sal; **SH/SC** – Serpa; **CC** – Coruche; **A** – HbA; **(-)** – Female; **nd** – No data.

Sample ID	*HBB*S* Haplotype	mtDNA Haplotype	Y-Cr Haplotype
AH/02	Bantu/A	U6b	-
AH/03	Bantu/A	K1a	E1b1b1b*
SH/01	Benin/A	H*	-
SH/02	Benin/Benin	J1b1	R1b1*
SH/03	Benin/A	HV0	-
SH/04	Benin/A	I1a	-
SH/05	Benin/A	T1	J*
SH/06	Benin/A	T1	nd
SH/07	Benin/A	T1	nd
SH/08	Benin/A	J1c1	P*
SH/09	Benin/A	J1c1	nd
SC/60	Benin/A	H2a	R1b1*
CC/04	Bantu/A	K1a	J2*
CC/09	Senegal/A	L1b	nd
CC/16	Senegal/A	U5b	R1b1*
CC/24	Senegal/A	H*	E1b1b1b*
CC/25	Senegal/A	K1a	E1b1b1b*
CC/45	Bantu/A	HV0	P*
CC/151	Senegal/A	U6a	R1b1*

Focusing on the mtDNA results because they are available for all the 19 individuals regardless their sex, (contrarily to Y-chromosome data, which obviously was only obtained for males), it is noteworthy the combined frequency of haplogroups I, J, and T among the *HBB*S* bearing chromosomes, all having the Benin haplotype as background. In the study of Pereira et al. (2010) a thin differentiation of Pias comparatively to Alcácer and Coruche was detected that was mainly attributed precisely to the global frequency of haplogroups I, J, and T in Pias. Since the three mtDNA haplogroups are typical from peri-Mediterranean populations, the authors proposed that Pias had retained a stronger link with Mediterraneans than other Portuguese populations.

Coincidently, the Benin haplotype, which is the most prevalent *HBB*S* haplotypes at the worldwide level, reaches its highest frequency throughout all Mediterranean regions (Bitoungui et al., 2015).

Thus, the Mediterranean influence in Pias seems now reinforced with the observation that the Benin haplotype is restrictively present in the region.

In the *HBB*S* chromosomes from Alcácer (only two) and especially from Coruche (seven in total) the distribution of mtDNA lineages is less unusual in the context of the general Portuguese mtDNA diversity, excepting the presence of a L1b lineage in a sickle cell chromosome from Coruche (CC/09). The presence of mtDNA L lineages outside sub-Saharan Africa, where they are typically found, is mainly attributed to the influx of African lineages that occurred during the Atlantic slave trade, lasted from the mid-15th century until the late-18th century. In Portugal, and especially in the south, L lineages are considerably more common than in the rest of Europe, denoting the demographic impact of the presence of African slaves in the country. The detection that one individual from Coruche carrier of the sickle cell trait in a Senegal haplotype also bears an mtDNA lineage of sub-Saharan extraction, suggests that in Coruche, African slaves had an important role in the importation of the Sickle Cell Trait to the region.

Furthermore, both the Senegal and the Bantu haplotypes found in Coruche, are very common in the regions from Africa that were main sources of slaves brought to Portugal. Whereas the Bantu haplotype attains the highest prevalence in Central, Southern, and East Africa, the Senegal haplotype is more restricted to the far west regions of Africa, where it peaks in frequency (Bitoungui et al., 2015)

Globally, the results here obtained fully meet the picture proposed by Lavinha et al. (1992), positing that the Sickle Cell Trait might have been imported twice into Portugal: first, between the eighth and the thirteenth centuries from the Mediterranean basin (in association with the Benin haplotype) and later, after the fifteenth century from sub-Saharan Africa over an Atlantic route (Senegal and Bantu haplotypes).

This two-step process of introduction of HBB*S was based on the observation in Portugal of a superposition of two Bantu-Senegal foci over a diffuse distribution of the Benin haplotype (Lavinha et al., 1992). In this study, two foci of Bantu-Senegal haplotypes were also detected (in Coruche and Alcácer do Sal), but the distribution of the Benin haplotype was not diffuse, since instead it was only present in Serpa (Pias). This late finding suggests that in Pias a founder effect might have underlain the high frequency of *HBB*S* in the current day population.

Previous studies have provide evidence that *HBB*S* haplotypes are associated to levels of fetal hemoglobin, which is the best known modifier of the clinical features of SCD (Steinberg, 2009). According to those studies, the Senegal and Arab-Indian haplotypes correlate with high percentage of HbF and consequently with milder course of the disease, the Bantu/CAR is related to low levels of HbF and a more severe

phenotype, while the Benin haplotype is thought to confer a relatively favorable clinical outcome (Bitoungui et al., 2015; Steinberg, 2009)

Since no hematological parameters or clinical data were available for the participants in this study, it was not possible to evaluate such correlation in the population samples analysed.

However, the observation that the unique homozygous for *HBB*S* harboured two Benin haplotypes, and was a subject from Pias, a foci of *HBB*S* in the Benin haplotype where the allele reaches the very unusual frequency of 7,7%, seems consistent with the favorable clinical outcome to which the Benin haplotype is associated. Naturally, the issue still deserves future investigation.

5. Final Remarks

The implemented SNaPshot® Multiplex methodology as proven to be an efficient technique to genotype the SNPs difining haplotypes associated to *HBB*S*.

Comparatively to the conventional RFLP technique, the SNaPshot® strategy simplifies substantialy the inference of haplotypes.

In that way it was possible to obtain a fine characterization of *HBB*S* allels in three regions of Alentejo that previously had been already reported as foci of high prevalence of the allele, although for Pias (Serpa) no published data was available to the date.

The data presented here indicates that Pias may be the region where *HBB*S* reaches the highest frequency in Portugal (7.7%), while in Coruche and Alcácer do Sal the frequencies were 2.2% and 2.9% respectively, wich are both high values within the Portuguese context.

Analysis of haplotypes was essential to obtain some insights on the introduction of *HBB*S* in Alentejo. The exclusive detection of Benin haplotype in Pias, whereas in Alcácer and Coruche this haplotype was absent but instead were present the Senegal and Bantu types, seems to be a clear sign that at least two migratory waves were responsible by the presence of the variant in Portugal: one coming from the region encompassing the Mediterranean basin (captured by the Benin haplotype), and other from sub-Saharan Africa, likely afforded by the transatlantic slave trade (captured by the Senegal and Bantu haplotypes). Remarkably, this scenario meets the picture proposed by Lavinha et al. (1992), despite the observed confinement of the Benin haplotype to Pias, wich suggest that a founder effect might be associated to the region.

Along with *HBB*S*, two β-thalassemias variants were detected in seven individuals out of total 266, giving a frequency of 0.013, wich is also a rather high value.

Globaly this study reinforce previous epidemiological data demonstrating that in south of Portugal hemoglobinopathies continue to be a serious concern in terms of health care.

In the future it would be important to complete the screening of *HBB,* wich due to time limitations was not possible to do in this work, and to haplotype the β-thalassemia variants.

More importantly, however, further studies should be done in order to obtain a picture still more detailed on the epidemiology of hemoglobinopathies in Portugal.

6. Bibliographic References

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nature methods *7*, 248-249.

Allison, A.C. (1954). Protection afforded by sickle-cell trait against subtertian malarial infection. British medical journal *1*, 290.

Almeida, L.O. (2015). Human Gene Mutations and Migratory Flows—Portugal and the Mediterranean. Advances in Anthropology *5*, 164.

Antonarakis, S.E., Boehm, C.D., Serjeant, G.R., Theisen, C.E., Dover, G.J., and Kazazian, H.H. (1984). Origin of the beta S-globin gene in blacks: the contribution of recurrent mutation or gene conversion or both. Proceedings of the National Academy of Sciences *81*, 853-856.

Baldwin, J., and Chothia, C. (1979). Haemoglobin: the structural changes related to ligand binding and its allosteric mechanism. Journal of molecular biology *129*, 175-220.

Bard, H. (1973). Postnatal fetal and adult hemoglobin synthesis in early preterm newborn infants. Journal of Clinical Investigation *52*, 1789.

Bitoungui, V.J.N., Pule, G.D., Hanchard, N., Ngogang, J., and Wonkam, A. (2015). Beta-globin gene haplotypes among cameroonians and review of the global distribution: is there a case for a single sickle mutation origin in Africa? Omics: a journal of integrative biology *19*, 171-179.

Bridges, K.R., and Pearson, H.A. (2008). Anemias and other red cell disorders, First edn (New York: McGraw Hill Professional).

Brittain, T. (2002). Molecular aspects of embryonic hemoglobin function. Molecular aspects of medicine *23*, 293-342.

Bruce-Chwatt, L.J., and de Zulueta, J. (1977). Malaria eradication in Portugal. Transactions of the Royal Society of Tropical Medicine and Hygiene *71*, 232-240.

Buckingham, L., and Flaws, M.L. (2007). Molecular Diagnostics: Fundamentals, Methods, & Clinical Applications, First edn (Philadelphia: F.A. Davis).

Campos, J.M. (2007). Rastreio de Hemoglobinopatias na Zona Centro de Portugal. In Departamento de Zoologia (Faculdade de Ciências e Tecnologia da Universidade de Coimbra).

Clarke, G.M., and Higgins, T.N. (2000). Laboratory investigation of hemoglobinopathies and thalassemias: review and update. Clinical chemistry *46*, 1284-1290.

Corrons, J.V. (2001). Anemias por defectos congénitos de la hemoglobina. Hemoglobinopatías estructurales y talasemias. Medicine-Programa de Formación Médica Continuada Acreditado *8*, 2684-2693.

Costa, S.N., Madeira, S., Sobral, M.A., and Delgadinho, G. (2016). Hemoglobinopatias em Portugal e a intervenção do médico de família. Revista Portuguesa de Medicina Geral e Familiar *32*, 416-424.

Crawford, D.C., Caggana, M., Harris, K.B., Lorey, F., Nash, C., Pass, K.A., Tempelis, C., and Olney, R.S. (2002). Characterization of β-globin haplotypes using blood spots from a population-based cohort of newborns with homozygous HbS. Genetics in Medicine *4*, 328-335.

Depetris-Chauvin, E., and Weil, D.N. (2016). Malaria and Early African Development: Evidence from the Sickle Cell Trait. The Economic Journal.

Excoffier, L., and Lischer, H.E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular ecology resources *10*, 564-567.

Faustino, P., Pacheco, P., Loureiro, P., Nogueira, P.J., and Lavinha, J. (1999). The geographic pattern of β-thalassaemia mutations in the Portuguese population. British journal of haematology *107*, 903-904.

Flint, J., Harding, R.M., Boyce, A.J., and Clegg, J.B. (1998). The population genetics of the haemoglobinopathies. Baillière's clinical haematology *11*, 1-51.

Fox, S. (2011). Human Physiology Twelfth edn (New York: McGraw-Hill).

Fullerton, S., Harding, R., Boyce, A., and Clegg, J. (1994). Molecular and population genetic analysis of allelic sequence diversity at the human beta-globin locus. Proceedings of the National Academy of Sciences *91*, 1805-1809.

Gallup, J.L., and Sachs, J.D. (2001). The economic burden of malaria. The American journal of tropical medicine and hygiene *64*, 85-96.

Griffiths, A.J.F. (2008). Introduction to Genetic Analysis, Ninth edn (W. H. Freeman).

Grosveld, F., Dillon, N., and Higgs, D. (1993). The regulation of human globin gene expression. Baillière's clinical haematology *6*, 31-55.

Haldane, J.S. (1949). The rate of mutation of human genes. Hereditas *35*, 267-273.

Hoffbrand, A.V., Pettit, J.E., and Vyas, P. (2010). Color Atlas of Clinical Hematology, Fourth edn (Philadelphia: Mosby/Elsevier).

Hoffbrand, V., Moss, P., and Pettit, J. (2006). Essential Haematology, Fifth edn (Massachusetts: Blackwell).

Honig, G.R., and Adams, J.G. (1986). Human Hemoglobin Genetics, First edn (New York: Springer-Verlag).

Jensen, F.B., Fago, A., and Weber, R.E. (1998). Hemoglobin Structure and Function. Fish physiology *17*, 1-40.

Jobling, M., Hollox, E., Hurles, M., Kivisild, T., and Tyler-Smith, C. (2013). Human Evolutionary Genetics, Second edn (New York: Garland Science,Taylor & Francis Group).

Kan, Y.W., and Dozy, A.M. (1978). Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. Proceedings of the National Academy of Sciences *75*, 5631-5635.

Kan, Y.W., and Dozy, A.M. (1980). Evolution of the hemoglobin S and C genes in world populations. Science *209*, 388-391.

Kaushansky, K., Lichtman, M.A., Beutler, E., Kipps, T.J., Seligsohn, U., and Prchal, J.T. (2010). Williams Hematology, Eighth edn (New York: McGraw-Hill Medical).

Kent, W.J. (2002). BLAT—the BLAST-like alignment tool. Genome research *12*, 656-664.

Kulozik, A., Wainscoat, J., Serjeant, G., Kar, B., Al-Awamy, B., Essan, G., Falusi, A., Haque, S., Hilali, A., and Kate, S. (1986). Geographical survey of βs-globin gene haplotypes: evidence for an independent Asian origin of the sickle-cell mutation. American journal of human genetics *39*, 239.

Landeiro, F. (1933). O sezonismo em Portugal (Agên cia Geral das Colónias, Divisão de Publicações e Biblioteca).

Lavinha, J., Gonçalves, J., Faustino, P., Romão, L., Osório-Almeida, L., Peres, M.J., Picanço, I., Martins, M.C., Ducrocq, R., and Labie, D. (1992). Importation route of the sickle cell trait into Portugal: contribution of molecular epidemiology. Human biology, 891-901.

Leal, A.S., Martins, P.R.J., Balarin, M.A.S., Pereira, G.A., and Resende, G.A.D. (2016). Haplotypes βs-globin and its clinical-haematological correlation in patients with sickle-cell anemia in Triângulo Mineiro, Minas Gerais, Brazil. Jornal Brasileiro de Patologia e Medicina Laboratorial *52*, 6-10.

Livingstone, F.B. (1958). Anthropological implications of sickle cell gene distribution in West Africa. American Anthropologist *60*, 533-562.

Luzzatto, L. (2012). Sickle cell anaemia and malaria. Mediterranean journal of hematology and infectious diseases *4*.

Martins, M., Olim, G., Melo, J., Magalhaes, H., and Rodrigues, M. (1993). Hereditary anaemias in Portugal: epidemiology, public health significance, and control. Journal of medical genetics *30*, 235-239.

Monteiro, C., Rueff, J., Falcao, A., Portugal, S., Weatherall, D., and Kulozik, A. (1989). The frequency and origin of the sickle cell mutation in the district of Coruche/Portugal. Human genetics *82*, 255-258.

Nelson, D.L., and Cox, M.M. (2012). Lehninger Principles of Biochemistry, Sixth edn (New York: W. H. Freeman).

Old, J.M. (2003a). DNA Diagnosis of Hemoglobin Mutations. In Hemoglobin Disorders: Molecular Methods and Protocols, R.L. Nagel, ed. (New Jersey: Humana Press), pp. 101-116.

Old, J.M. (2003b). Methods for Analysis of Prenatal Diagnosis. In Hemoglobin Disorders: Molecular Methods and Protocols, R.L. Nagel, ed. (New Jersey: Humana Press), pp. 117-131.

Pagnier, J., Mears, J.G., Dunda-Belkhodja, O., Schaefer-Rego, K.E., Beldjord, C., Nagel, R.L., and Labie, D. (1984). Evidence for the multicentric origin of the sickle cell hemoglobin gene in Africa. Proceedings of the National Academy of Sciences *81*, 1771-1773.

Patrinos, G.P., B. Giardine, C. Riemer, W. Miller, D.H.K. Chui, N.P. Anagnou, H. Wajcman, and R.C. Hardison (2004). Improvements in the HbVar database of human hemoglobin variants and thalassemia mutations for population and sequence variation studies. Nucleic Acids Research *32*, D537-D541.

Pauling, L., Itano, H.A., Singer, S., and Wells, I.C. (1949). Sickle Cell Anemia, a Molecular Disease. Science *110*, 543-548.

Pereira, V., Gomes, V., Amorim, A., Gusmão, L., and João Prata, M. (2010). Genetic characterization of uniparental lineages in populations from Southwest Iberia with past malaria endemicity. American Journal of Human Biology *22*, 588-595.

Piel, F.B., Patil, A.P., Howes, R.E., Nyangiri, O.A., Gething, P.W., Dewi, M., Temperley, W.H., Williams, T.N., Weatherall, D.J., and Hay, S.I. (2013). Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. The Lancet *381*, 142-151.

Piel, F.B., Patil, A.P., Howes, R.E., Nyangiri, O.A., Gething, P.W., Williams, T.N., Weatherall, D.J., and Hay, S.I. (2010). Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. Nature communications *1*, 104.

Rees, D.C., Williams, T.N., and Gladwin, M.T. (2010). Sickle-cell disease. The Lancet *376*, 2018-2031.

Ribeiro, M., Gonçalves, P., Cunha, E., Bento, C., Almeida, H., Pereira, J., Núñez, G.M., and Tamagnini, G. (1997). Genetic heterogeneity of β-thalassemia in populations of the Iberian peninsula. Hemoglobin *21*, 261-269.

Rodriguez Romero, W., Sáenz Renauld, G.F., and Chaves Villalobos, M.A. (1998). Haplotipos de la hemoglobina S: importancia epidemiológica, antropológica y clínica.

Rogers, K. (2010). Blood: Physiology and Circulation, First edn (New York: Britannica Educational Publishing).

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., and Arnheim, N. (1985). Enzymatic amplification of b-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science *230*, 1350-1354.

Sherwood, L. (2015). Human Physiology: From Cells to Systems, Ninth edn (Boston: Cengage Learning).

Smith, P.J., Zhang, C., Wang, J., Chew, S.L., Zhang, M.Q., and Krainer, A.R. (2006). An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. Human molecular genetics *15*, 2490-2508.

Steinberg, M.H. (2009). Genetic etiologies for phenotypic diversity in sickle cell anemia. The Scientific World Journal *9*, 46-67.

Steinberg, M.H., Forget, B.G., Higgs, D.R., and Weatherall, D.J. (2009). Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management, Second edn (New York: Cambridge University Press).

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3—new capabilities and interfaces. Nucleic acids research *40*, e115-e115.

Vekilov, P.F.-T., Angela; Hirsh, Rhoda Elison (2003). Nucleation and Crystal Growth of Hemoglobins - The Case of HbC. In Hemoglobin disorders: Molecular Methods and Protocols, R.L. Nagel, ed. (New Jersey: Humana Press), pp. 155-176.

Vinson, A.E., Walker, A., Elam, D., Glendenning, M., Kutlar, F., Clair, B., Harbin, J., and Kutlar, A. (2004). A Novel Approach to Rapid Determination of βS-Globin Haplotypes: Sequencing of the Aγ-IVS-II Region. Hemoglobin *28*, 317-323.

Weatherall, D., and Clegg, J. (2001). Inherited haemoglobin disorders: an increasing global health problem. Bulletin of the World Health Organization *79*, 704-712.

Weatherall, D.J. (2013). Globin Genes, Human. In Brenner's Encyclopedia of Genetics, S. Brenner, and J.H. Miller, eds. (Academic Press), pp. 337-339.

Williams, T.N., and Weatherall, D.J. (2012). World distribution, population genetics, and health burden of the hemoglobinopathies. Cold Spring Harbor perspectives in medicine *2*, a011692.

Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., and Gil, L. (2015). Ensembl 2016. Nucleic acids research *44*, D710-D716.

Zago, M., Silva Jr, W., Dalle, B., Gualandro, S., Hutz, M., Lapoumeroulie, C., Tavella, M., Araujo, A., Krieger, J., and Elion, J. (2000). Atypical βS haplotypes are generated by diverse genetic mechanisms. American journal of hematology *63*, 79-84.

Supplementary Data

Table S1 - Demographic characterization and genotypic data of the polymorphisms found in the studied samples.
AH/AC – Alcácer do Sal; **SH/SC** – Serpa; **CC** – Coruche; **F** – Female; **M** – Male.

			Polymorphisms						
Sample ID	Gender	Age	rs334	rs7130040	rs10768683	HbVar.827	HbVar.845	rs368604295	rs33935983
AH/01	F	48	A/A	C/C	C/C	G/G	C/T	T/T	G/G
AH/02	F	60	A/T	C/C	C/C	G/G	C/C	T/T	G/G
AH/03	M	16	A/T	C/C	C/C	G/G	C/C	T/T	G/G
AC/01	F	65	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/02	F	66	A/A	T/C	G/C	G/G	C/C	T/T	G/G
AC/03	F	77	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/04	F	58	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/05	F	31	A/A	T/C	G/C	G/G	C/C	T/T	G/G
AC/06	F	64	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/07	F	59	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/08	M	77	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/09	M	-	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/10	M	72	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/11	M	75	A/A	T/C	G/C	G/G	C/C	T/T	G/G
AC/12	M	65	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/13	M	57	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/14	F	39	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/15	F	46	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/16	F	80	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/17	M	61	A/A	T/C	G/C	G/G	C/C	T/T	G/G
AC/18	F	42	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/19	M	84	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/20	F	46	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/21	F	70	A/A	T/C	G/C	G/G	C/C	T/T	G/G
AC/22	F	69	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/23	F	60	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/24	F	30	A/A	T/C	G/C	G/G	C/C	T/T	G/G
AC/25	M	64	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/26	F	85	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/27	M	67	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/28	M	72	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/29	F	69	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/30	F	60	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/31	F	78	A/A	C/C	C/C	G/G	C/C	T/T	G/G
AC/32	M	82	A/A	T/C	G/C	G/G	C/C	T/T	G/G
SH/01	F	42	A/T	C/C	C/C	G/G	C/C	T/T	G/G
SH/02	M	36	T/T	C/C	C/C	G/G	C/C	T/T	G/G
SH/03	F	30	A/T	C/C	C/C	G/G	C/C	T/T	G/G
SH/04	F	90	A/T	C/C	C/C	G/G	C/C	T/T	G/G
SH/05	M	37	A/T	T/C	G/C	G/G	C/C	T/T	G/G
SH/06	M	99	A/T	C/C	C/C	G/G	C/C	T/T	G/G
SH/07	M	61	A/T	T/C	G/C	G/G	C/C	T/T	G/G
SH/08	M	100	A/T	C/C	C/C	G/G	C/C	T/T	G/G
SH/09	M	79	A/T	C/C	C/C	G/G	C/C	T/T	G/G
SC/01	M	82	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/02	M	85	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/03	M	83	A/A	T/C	G/C	G/G	C/C	T/T	G/G
SC/04	M	86	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/05	M	97	A/A	T/T	G/G	G/G	C/C	T/T	G/G
SC/06	M	79	A/A	T/C	G/C	G/G	C/C	T/T	G/G
SC/07	M	91	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/08	M	78	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/09	M	85	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/10	M	33	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/11	M	40	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/12	M	12	A/A	C/C	C/C	G/G	C/C	T/T	G/G

			Polymorphisms						
Sample ID	Gender	Age	rs334	rs7130040	rs10768683	HbVar.827	HbVar.845	rs368604295	rs33935983
SC/13	M	44	A/A	C/C	C/C	G/G	C/C	T/G	G/G
SC/14	M	83	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/15	M	12	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/16	M	50	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/17	M	94	A/A	T/C	G/C	G/G	C/C	T/T	G/G
SC/18	M	85	A/A	T/C	G/C	G/G	C/C	T/T	G/G
SC/19	M	12	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/20	M	89	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/21	M	78	A/A	T/C	G/C	G/G	C/C	T/T	G/G
SC/22	M	78	A/A	T/C	G/C	G/G	C/C	T/T	G/G
SC/23	M	78	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/24	M	86	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/25	M	83	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/26	M	97	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/27	M	-	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/28	M	78	A/A	T/C	G/C	G/G	C/C	T/T	G/G
SC/29	M	82	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/30	M	81	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/31	M	83	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/32	M	66	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/33	M	86	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/34	M	88	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/35	M	92	A/A	T/C	G/C	G/G	C/C	T/T	G/G
SC/36	M	35	A/A	T/C	G/C	G/G	C/C	T/T	G/G
SC/37	M	82	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/38	M	90	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/39	M	70	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/40	M	54	A/A	T/C	G/C	G/G	C/C	T/T	G/G
SC/41	M	65	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/42	M	67	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/43	M	77	A/A	C/C	C/C	G/G	C/T	T/T	G/G
SC/45	M	84	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/48	M	89	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/49	M	28	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/51	M	80	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/52	M	71	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/53	M	74	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/54	M	84	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/55	M	45	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/56	M	59	A/A	T/T	G/G	G/G	C/C	T/T	G/G
SC/57	M	51	A/A	T/C	G/C	G/G	C/C	T/T	G/G
SC/58	M	79	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/59	M	86	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/60	M	86	A/T	C/C	C/C	G/G	C/C	T/T	G/G
SC/61	M	86	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/62	M	-	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/63	M	57	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/64	M	82	A/A	C/C	C/C	G/G	C/C	T/T	G/G
SC/65	M	86	A/A	T/C	G/C	G/G	C/C	T/T	G/G
SC/66	M	54	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/01	M	78	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/02	M	78	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/03	M	92	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/04	M	88	A/T	C/C	C/C	G/G	C/C	T/T	G/G
CC/05	M	78	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/06	M	51	A/A	C/C	C/C	G/G	C/C	T/T	G/G

			Polymorphisms						
Sample ID	Gender	Age	rs334	rs713040	rs10768683	HbVar.827	HbVar.845	rs368604295	rs33935983
CC/07	M	78	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/08	M	93	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/09	M	99	A/T	C/C	C/C	G/G	C/C	T/T	G/G
CC/10	M	90	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/11	M	75	A/A	C/C	C/C	G/G	C/T	T/T	G/G
CC/12	M	-	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/13	M	87	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/14	M	93	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/15	M	65	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/16	M	39	A/T	C/C	C/C	G/G	C/C	T/T	G/G
CC/17	M	86	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/18	M	74	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/19	M	74	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/20	M	81	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/21	M	-	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/22	M	91	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/23	M	88	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/24	M	87	A/T	C/C	C/C	G/G	C/C	T/T	G/G
CC/25	M	33	A/T	T/C	G/C	G/G	C/C	T/T	G/G
CC/26	M	56	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/27	M	44	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/28	M	86	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/29	M	35	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/30	M	83	A/A	T/C	G/C	G/A	C/C	T/T	G/G
CC/31	M	82	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/32	M	58	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/33	M	74	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/34	M	83	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/35	M	83	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/36	M	91	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/37	M	92	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/38	M	74	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/39	M	78	A/A	C/C	C/C	G/G	C/T	T/T	G/G
CC/40	M	84	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/41	M	84	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/42	M	66	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/43	M	80	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/44	M	88	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/45	M	82	A/T	C/C	C/C	G/G	C/C	T/T	G/G
CC/46	M	68	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/50	M	66	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/51	M	61	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/52	M	74	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/53	M	87	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/54	M	80	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/55	M	88	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/56	M	74	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/61	M	84	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/62	M	88	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/63	M	78	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/64	M	70	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/65	M	46	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/66	M	87	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/67	M	92	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/76	M	-	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/77	M	68	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/78	M	71	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/83	M	84	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/84	M	90	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/85	M	78	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/87	M	82	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/88	M	65	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/89	M	88	A/A	T/T	G/G	G/G	C/C	T/T	G/G

			Polymorphisms						
Sample ID	Gender	Age	rs334	rs713040	rs10768683	HbVar.827	HbVar.845	rs368604295	rs33935983
CC/90	M	81	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/91	M	40	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/99	M	83	A/A	T/T	G/G	G/G	C/C	T/T	G/G
CC/100	M	64	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/101	M	90	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/102	M	71	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/103	M	78	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/104	M	80	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/105	M	70	A/A	T/T	G/G	G/G	C/C	T/T	G/G
CC/111	M	67	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/112	M	72	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/113	M	71	A/A	C/C	C/C	G/G	C/T	T/T	G/G
CC/114	M	82	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/115	M	90	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/116	M	80	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/117	M	42	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/118	M	60	A/A	T/T	G/G	G/G	C/C	T/T	G/G
CC/119	M	46	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/120	M	77	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/121	M	60	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/122	M	75	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/123	M	77	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/124	M	76	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/125	M	93	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/126	M	78	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/127	M	38	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/128	M	77	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/129	M	80	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/130	M	74	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/131	M	98	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/132	M	83	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/133	M	85	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/134	M	61	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/135	M	73	A/A	C/C	C/C	G/A	C/C	T/T	G/G
CC/137	M	93	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/138	M	84	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/139	M	66	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/140	M	89	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/141	M	84	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/142	M	75	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/144	M	81	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/145	M	91	A/A	T/C	G/C	G/G	C/C	T/T	G/A
CC/146	M	68	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/147	M	55	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/148	M	74	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/149	M	96	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/150	M	77	A/A	T/T	G/G	G/G	C/C	T/T	G/G
CC/151	M	76	A/T	T/C	G/C	G/G	C/C	T/T	G/G
CC/152	M	65	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/153	M	68	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/154	M	36	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/155	M	86	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/156	M	91	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/157	M	96	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/158	M	63	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/159	M	68	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/160	M	72	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/161	M	71	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/162	M	37	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/163	M	95	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/164	M	80	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/165	M	85	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/166	M	66	A/A	C/C	C/C	G/G	C/C	T/T	G/G

			Polymorphisms						
Sample ID	**Gender**	**Age**	**rs334**	**rs713040**	**rs10768683**	**HbVar.827**	**HbVar.845**	**rs368604295**	**rs33935983**
CC/167	M	90	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/168	M	37	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/169	M	91	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/170	M	76	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/171	M	-	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/172	M	93	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/173	M	87	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/174	M	78	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/175	M	80	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/176	M	59	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/177	M	85	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/178	M	76	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/179	M	76	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/180	M	95	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/181	M	76	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/184	M	74	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/185	M	68	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/186	M	77	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/187	M	85	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/188	M	64	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/189	M	82	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/190	M	78	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/192	M	75	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/193	M	72	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/194	M	55	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/198	M	86	A/A	T/C	G/C	G/G	C/C	T/T	G/G
CC/198*	M	11	A/A	C/C	C/C	G/G	C/C	T/T	G/G
CC/199	M	66	A/A	T/C	G/C	G/G	C/C	T/T	G/G