Ana Cláudia Moreira Azevedo Genetic profile of the Iberian Peninsula population: contributions of AIMs

UMinho|2014

Universidade do Minho
Escola de Ciências

Ana Cláudia Moreira Azevedo

Genetic profile of the Iberian Peninsula population: contributions of AIMs

janeiro de 2014

Universidade do Minho
Escola de Ciências

Ana Cláudia Moreira Azevedo

Genetic profile of the Iberian Peninsula population: contributions of AIMs

Dissertação de Mestrado
Mestrado em Genética Molecular

Trabalho realizado sob a orientação da
Doutora Ana Paula Fernandes Monteiro Sampaio Carvalho
e da
Doutora Cristina Maria Pereira dos Santos

janeiro de 2014

Nome: Ana Cláudia Moreira Azevedo

Endereço electrónico: ana.moreirazevdo@gmail.com **Telefone:** +351911985065

Número de Identificação Civil: 13552855

Título da dissertação de mestrado:

Genetic profile of the Iberian Peninsula population: contributions of AIMs

(Perfil genético da população da Península Ibérica: contribuições dos AIMs)

Orientadores: Doutora Ana Paula Fernandes Monteiro Sampaio Carvalho

Doutora Cristina Maria Pereira dos Santos

Ano de conclusão: 2014

Designação do Mestrado:

Mestrado em Genética Molecular

Universidade do Minho, ___/___/______

Assinatura: ____

Acknowledgements/Agradecimentos

No fim desta etapa gostaria de agradecer a todos aqueles que me ajudaram ao longo desta:

À Professora Doutora Cristina Pereira Santos, minha orientadora, por ter aceite orientar uma aluna de outra universidade e que não conhecia, pela partilha de conhecimentos científicos, pela paciência e pelo apoio que sempre me deu,

À Professora Doutora Ana Paula Fernandes Monteiro Sampaio Carvalho, minha co-orientadora, agradeço-lhe pela simpatia e tempo que despendeu comigo,

Aos meus pais, por todo o apoio que me deram e por terem tornado este sonho possível,

À minha irmã, Bia, por todo o apoio e por me ter sempre uma palavra de conforto quando as coisas não corriam como esperado,

A todos os amigos que deixei em Portugal, pelo apoio e carinho que me deram,

Á Tânia por ter sido uma óptima companheira de Erasmus,

À Rut, Mar, Sara e Maria por todo o apoio e à Amanda pela ajuda e paciência,

A todos da Unidade de Antropologia Biológica na Faculdade de Biociências da Universidade Autónoma de Barcelona,

A todos que conheci durante a minha estadia na Catalunha,

A todos o meu Obrigado! Merci! Gracias!

Por fim gostaria de agradecer ao programa Erasmus Placements pelo financiamento para o período de mobilidade.

Abstract

Ancestry informative markers (AIMs) are a type of single nucleotide polymorphism (SNP) used to establish the probability of an individual biogeographical ancestry, as they occur at different frequencies on each population.

The aims of this study are: chapter 1) to select European AIMs that would allow to differentiate Iberian populations and design and validate primers that allow to analyse the selected AIMs by High Resolution Melting (HRM); chapter 2) to preliminarily genetically characterize three Iberian populations using selected SNPs.

For these purposes, twenty-six SNPs were selected in order to capture a north-south genetic gradient observed in Europe and the influence of Sephardic Jews. For the selected SNPs, primers which would allow the amplification of the region that contained the SNP were design. From the initial group, five SNPs were selected and the primers tested in the laboratory, PCR conditions were optimized and then used for analyses by HRM.

In chapter two, three SNPs were genotyped using HRM methodology in three specific Iberian populations, namely, Catalonia, Andalucia and Zamora. The data obtained was then compared to data of other populations already available. This preliminary study allows a glimpse of a possible European north-south axis in the Peninsula, nevertheless an extended work has to be performed using all the selected SNPs and including more Iberian populations.

Resumo

Marcadores informativos de ancestralidade (AIMs) são um tipo de polimorfismo de nucleótido único (SNP) usados para estabelecer a probabilidade de ancestralidade biogeográfica de um individuo, pois ocorrem com diferentes frequências em cada população.

Os objetivos deste estudo são: o capítulo 1) selecionar AIMs europeus que permitam diferenciar as populações ibéricas, desenhando e validando primers que permitam a análise dos AIMs selecionados por High Resolution Melting (HRM), capítulo 2) caracterizar geneticamente, de forma preliminar, três populações ibéricas através dos SNP previamente selecionados.

De forma a atingir os objetivos propostos, selecionaram-se vinte e seis SNPs, para se verificar a possível existência de um gradiente genético, de norte para sul da Europa, e a influência dos judeus sefarditas. Para os SNPs selecionados, foram desenhados primers que permitissem a amplificação da região que contivesse o SNP. Do grupo inicial de vinte e seis SNPs, cinco foram selecionados e os seus primers testados em laboratório, com condições de PCR otimizadas, e em seguida usados na análise pela técnica de HRM.

No capítulo dois, fez-se a genotipagem de três SNPs, pela metodologia de HRM, em três populações ibéricas, nomeadamente, Catalunha, Andaluzia e Zamora. Os dados obtidos foram comparados com os dados de outras populações já disponíveis.

Este estudo permite um vislumbre de um possível eixo europeu norte-sul presente na península ibérica. No entanto, um trabalho mais exaustivo deverá ser realizado utilizando todos os SNPs selecionados e incluindo outras populações ibéricas.

Index

Index of Figures

Index of Tables

I. Introduction

1. Evolution and Diversity

During the ancient Greece there were already reports regarding inherited differences between humans [Vogel et al., 1997]. However, until the 20th century, the ultraconservative ideas enacted by the Christian Religion prevailed. These ideas are derived by the philosophies of Plato and Aristotle, which attributed fixed and immutable characteristics to the species [Futuyma, 1998].

In the 19th century Charles Darwin formulates the theory of evolution [Griffiths et al., 1993]. This theory assumes that the evolution of species is the result of different survival and reproduction rates of individuals' variants that pre-existed in the group. Selection can eliminate some characters in a population, but cannot lead to new forms if they are not present in the original population. This contradiction was a serious impediment to explain the logic of evolution, which could not be accepted until the time that more knowledge about genetics and the processes that mediate the inheritance existed [Boyd et al., 2001]. The initial overcome for these contradictions took place in early 20th century, in the decades of 30 and 40, when the synthesis of the modern theory of evolution was developed. This development was due to a combine work of people from different areas of the science, that atone the theory proposed by Darwin with genetic processes to demonstrate that natural populations are not uniform but there have many variants [Griffiths et al., 1993].

Diversity refers to differences between individuals of the same species. In the case of humans, they differ in thousands of ways. Even in a group of related people there is variation in height, weight, hair colour, eye colour, etc, which refers to changes in the phenotype of organisms. The levels of genetic diversity within populations are affected by the evolutionary forces, such as mutation, natural selection, genetic drift, gene flow and recombination [Boyd et al., 2001].

Although the rate of appearance of new mutations is low, mutation is the ultimate source of genetic variation, producing new alleles. So, without this process evolution wouldn't exist [Raven et al., 2011]. Recombination is the production of new genetic combinations from the initially generated by the mutation, so the variation is increased even more [Boyd et al., 2001].

Natural selection removes variation from population by favouring one allele over other at a gene locus. However, in some circumstances, selection can do exactly the opposite and maintain population variation [Raven et al., 2011; Griffiths et al., 1993]. The modulating effect of natural selection on allelic frequencies, depend on whether a particular trait is favored or disadvantaged in a particular environment. Also, the speed of natural selection depends on the original allelic frequencies and values of fitness for each genotype [Griffiths et al., 1993].

Genetic drift causes random changes in allelic frequencies, for example, if two populations are isolated from each other, both change and, over time, might become genetically different [Griffiths et al., 1993].

While mutation, selection and genetic drift contribute to the differentiation between populations, gene flow acts in the opposite direction, in other words, migration make that over time, a mix of genes occurs, decreasing the differences between populations [Boyd et al., 2001].

All these evolutionary forces acting simultaneously over time, produce complex patterns of variation in allelic frequencies within a population and between different populations. In this sense, considering the recent human evolution, it is interesting to see the genetic changes that occur between different populations and identify the causes which produced such changes [Lewontin, 1974].

1.1- Human populations' genetics

Population genetics describes the genetic variation, and determine how it changes over time and space, on a theoretical and experimental basis [Griffiths et al., 1993].

The traditional physical anthropology began to study the human variability in the 19th century, consisting on the description of the human body, through its measurement [Amorim, 2007]. On 1900, Landsteiner published a short note commenting the agglutination of red cells of some patients by the serum of other patients [Landsteiner,1900], reaction which is now known to be the consequence of antibody and human blood cells interaction. In 1901, the same author published another study describing the human blood groups and revealed the polymorphisms on the phenotype blood system ABO [Landsteiner, 1901], thus becoming the first polymorphic genetic marker applied in humans [Owen, 2000; Cavalli-Sforza, 1996; Jobling et al., 2004] and opening a new era for anthropological studies. On 1917, Hirszfeld and Hirszfeld study the ABO phenotype proportions on different population and came to the conclusion that the frequencies where different on all populations, what nowadays we know that each population show different frequencies not only on the polymorphism ABO but practically in all other genetic polymorphic marker studied [Cavalli-Sforza, 1996; Cavalli-Sforza et al., 2003]. Subsequent analysis of proteins [Pauling et al., 1949; Lewis et al., 1958] enabled the analysis of a large number of genes, where the first results show a higher diversity than expected in the human population, with a large genetic variation within populations (≈85%) and only a small part (≈15%) attributable to differences between populations [Lewontin, 1967, 1972, 1974]. However, protein based genetic was limited due to the low discrimination power of the system [Budowle et al., 2008]. From this moment on, genetic markers become a valuable tool for anthropology, as long as they meet a number of characteristics: to be monophyletic, neutral, permit to establish a direct relationship between antecessor and offspring without the distortion of environmental adaptations and allowing comparisons of interpopulation variability [Susanne et al., 2007]. Finally, the direct analysis of DNA became possible, thanks to advances in techniques such as polymerase chain reaction [Saiki et al., 1985]. Since then, the knowledge on genetic markers have progressed and nowadays the most important markers are based on polymorphisms at the DNA, which have fewer limitations against, for example, the use of proteins [Budowle et al., 2008]. The most common forms of polymorphism used are: restriction length fragment polymorphism (RFLP), minisatellites (Variable Number of Tandem Repeats – VNTR), microsatellites (Short Tandem Repeats - STRs) and single nucleotide polymorphism (SNP) [Budowle et al., 2008; Griffiths et al., 1993].

1.2- Single nucleotide polymorphism (SNP)

A Single Nucleotide Polymorphism (SNP) is a site at which individuals differ by only a single nucleotide and the minority of the variant must have a frequency greater than 1% (fig. 1) [Raven et al., 2011]. This kind of polymorphism is the most common form of variation in the human genome [Sobrino et al., 2005]; and according to the dsSNP database on NCBI the number of SNP already identified on the human genome is 60,560,048 SNPs [NCBI, 2013]. Thus, due to its abundance, they can be useful as genetic markers for numerous genetic applications [Zhao et al., 2000], for examples to search for mutation associated with specific genetic diseases [Scherer, 2008].

SNP reflect past mutations that were probably unique events and two individuals sharing an allelic variant are marked with a common evolutionary heritage. In other words, our genes have ancestors and analyzing shared patterns of SNP variation can give insights on the origin of individuals [Stoneking, 2001]. Thus, the study of this variability allows us to infer the history of human evolution and patterns of migration of populations [Quintáns et al., 2003].

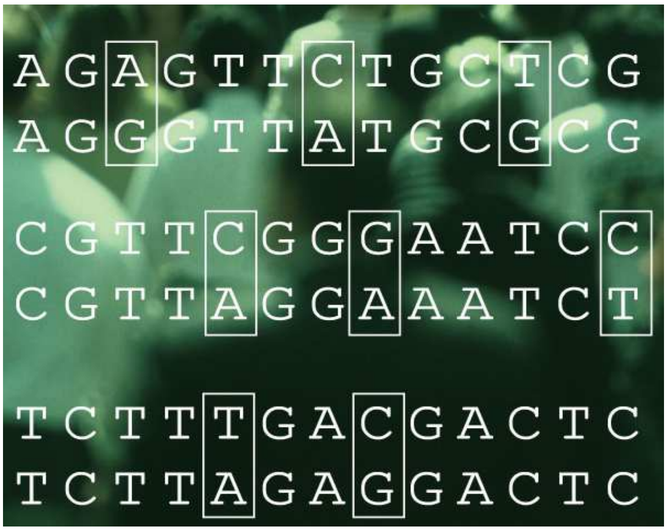

Figure 1 - DNA is a linear combination of four nucleotides, so when comparing two sequences position by position if there are different nucleotides at the same position we are in the present of a SNP [Chakravarti, 2001].

On the forensic level, the use of SNP is still on an experiment phase. However, comparing to short tandem repeats (STR) which are now used routinely, SNP have a lower mutation rate (10^{-8} vs. 10^{-3}), the recovery of information from degraded DNA samples is in theory possible, due to the fact that a smaller target region is needed to SNP typing, and SNP profiling avoids stutter artefacts that complicate STR profile interpretation [Butler et al., 2007; Kayser et al., 2011]. On the other hand, SNP aren't as polymorphic as STR and more SNPs are required to reach equivalent powers of discrimination or random match probabilities [Butler et al., 2007].

On a forensic perspective, SNPs can be divided into different categories as: identity-testing SNPs, lineage informative SNPs, phenotype informative SNPs and ancestry informative SNPs [Budowle et al., 2008]. Weir's unbiased (F_{ST})

Identify-testing SNPs provide genetic information to differentiate people and thus exclude individuals that cannot be the source of an evidentiary sample or cannot be a putative member. The best SNPs for identity testing are those that have the highest heterozygosity and low Weir's unbiased (F_{ST}) levels, because fewer SNPs will be needed to reach high levels of power of discrimination and fewer reference population databases will be required for statistical assessments for forensic casework [Budowle et al., 2008]. A study conducted by Kidd and collaborators shows the efforts and selection criteria necessary to develop the most useful identify-testing SNP panels [Kidd et al., 2006].

Lineage informative SNPs have been identified on the mtDNA genome and Y chromosome, due to the lack of recombination, these SNPs have been quite informative for evolutionary studies and kinship analysis [Budowle et al., 2008]. The most likely use for this class of SNP is for missing cases or mass disaster identification [Budowle et al., 2008; Kayser et al., 2011].

Phenotype informative SNPs are used to established high probability that an individual has a particular phenotypic characteristic, such as skin colour, hair colour and morphology, eye colour or stature [Budowle et al., 2008; Kayser et al., 2011]. Of all the traits studies, eye colour is the most accurately predictable and actually there are a system named IrisPlex that allows predicting eye colour [Kayser et al., 2011].

Ancestry informative SNPs are used to establish high probability of an individual's biogeographical ancestry and to indirectly infer some phenotypic characteristics. It is expected that there will be many forensic cases where this markers will be useful for guiding the police for find unknown persons or victims [Kayser et al., 2011; Phillips et al., 2009; Gu Y., 2011]. Already a small set of autosomal ancestry-informative DNA markers have been developed for DNA base inference of biogeographic ancestry at the level of continental resolution, and in this sense continental biogeographic ancestry may be used carefully in some cases to get a general idea about a person's appearance, but this only be applied to individuals whose ancestors come from a single continental region as can be established with DNA-based ancestry testing [Kayser et al., 2011].

1.3- Ancestry informative markers (AIMs)

Ancestry informative markers (AIMs) are SNPs distributed throughout the human genome, used to establish the probability of an individual biogeographical ancestry, since they occur at different frequencies in different world populations [Budowle et al., 2008; Gu et al., 2011].

A form of studying ancestry is by studying biogeography ancestry, in which's person origin is associated with the geographic location of presumed ancestors inferred by comparison with contemporary populations living in that locations [Royal et al., 2010]. The "Malthusian parameter of ascent" combined with the fact that our specie is relatively young and that our ancestors arrived in various regions of the world, explains the globally complex and mosaic-like landscape of genetic clusters, which strongly influences DNA-based inference of biogeography ancestry [Kayser et al., 2011].

Inference of individual ancestry from genetic markers is helpful in diverse situations, including admixture and association mapping, forensics, prediction of medical risks, wildlife management, studies of dispersal, gene flow and evolutionary history and to confirm genome-wide association studies (GWAS) [Rosenberg et al., 2003; Nassir et al., 2009].

Inference of genetic ancestry is applied to individuals and is based on the analysis of a large collections of ancestry informative markers [Royal et al., 2010], which have attracted quite attention due to their implication in different areas of genetics [Drineas et al., 2010].

AIMs that show large frequency differences among intercontinental groups can be used to detect and correct stratification within the populations [Bauchet et al., 2007]..

1.4- Typing methodologies

The SNP genotyping can be divided into different methods, which can be carried out using different specific techniques. Some of this methods and techniques are presented in figure 2.

Among the different methods, the one chosen and applied in this study was High Resolution Melting (HRM), which relies on the analysis of DNA melting and resulted from the collaboration between academics and industry and was introduced on scientific field in 2002 [QIAGEN®, 2009; Reed et al., 2007]. This technique characterizes the DNA samples according to their dissociation behaviour as they transition from double stranded DNA (dsDNA) to single-stranded DNA (ssDNA) with increasing temperature. Before the HRM, the target sequence must be amplified to a high-copy number in the presence of a dsDNA-binding fluorescent dye (EvaGreen), which actively binds to the dsDNA and fluoresces brightly when bounds. The change in fluorescence can be used to measure the increase in DNA concentration during PCR and then to directly measure thermally-induced DNA melting by HRM. To perform the HRM analysis, the temperature is increased and the fluorescence of the dye will be measured continuously, leading to a plot against the temperature [QIAGEN®, 2009]. This difference of Tm between samples will allow to differentiate samples as homozygous or heterozygous [Gundry et al., 2003; Liew, et al., 2004]. This technique is simple, easier, rapid, inexpensive and above all is a closed-tube system which prevents contamination with PCR products; on the other hand, depends strongly on a good PCR, instruments and dyes [QIAGEN®, 2009; Wittwer, 2009; Reed et al., 2007].

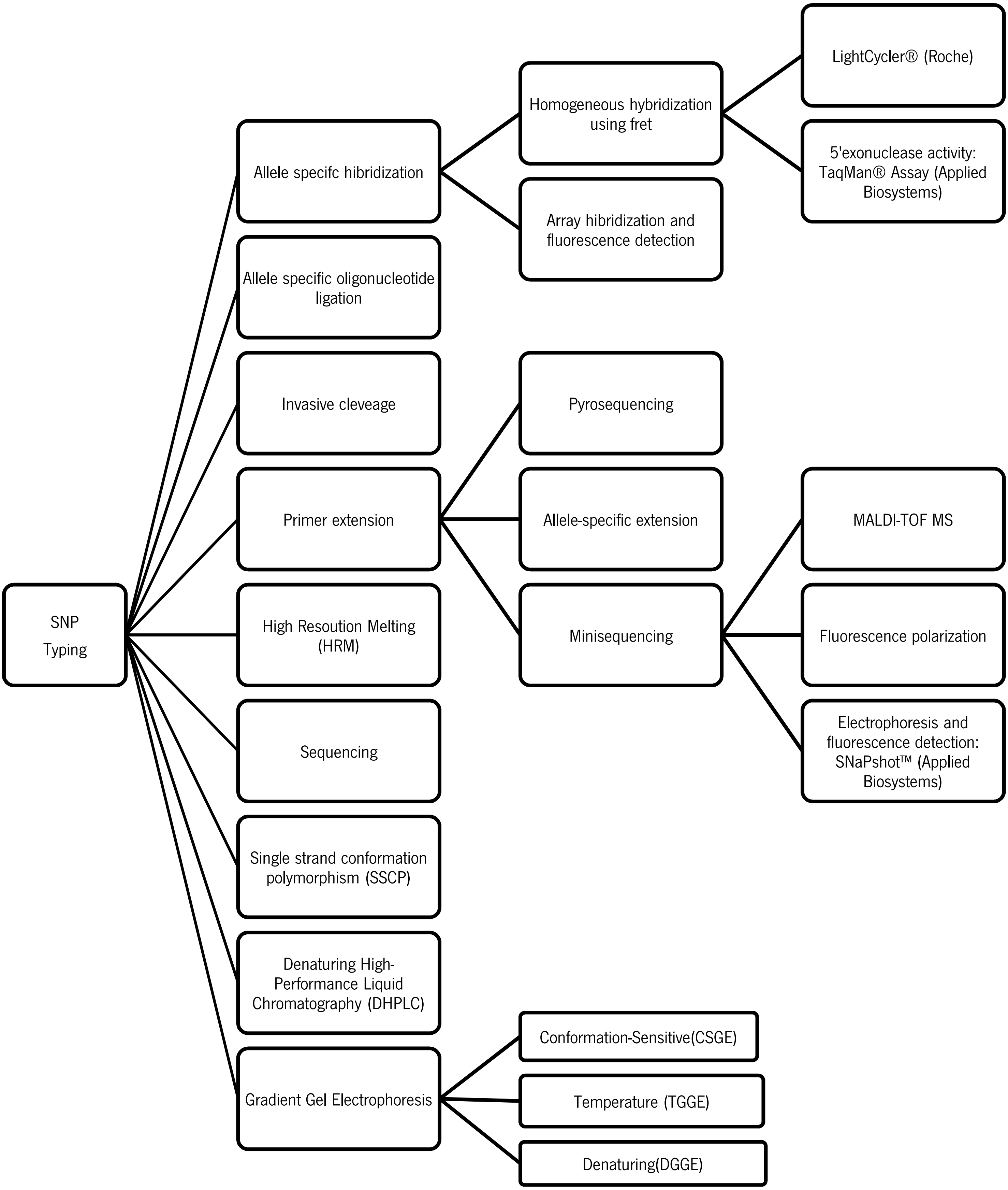
SNP
Typing
Allele specifc hibridization
Allele specific oligonucleotide ligation
Invasive cleveage
Primer extension
High Resoution Melting (HRM)
Sequencing
Single strand conformation polymorphism (SSCP)
Denaturing High-Performance Liquid Chromatography (DHPLC)
Gradient Gel Electrophoresis
Homogeneous hybridization using fret
Array hibridization and fluorescence detection
LightCycler® (Roche)
5'exonuclease activity: TaqMan® Assay (Applied Biosystems)
Pyrosequencing
Allele-specific extension
Minisequencing
MALDI-TOF MS
Fluorescence polarization
Electrophoresis and fluorescence detection: SNaPshot™ (Applied Biosystems)
Conformation-Sensitive(CSGE)
Temperature (TGGE)
Denaturing(DGGE)

Figure 2 - Some of the SNP typing methods used in scientific research. Some typing methods can divided itself on other more specific forms of typing and also differ themselves in the detection methods used [Sobrino et al., 2005; Kwok et al., 2003; Twyman, 2005; Costabile et al., 2006; Syvänen , 2001; Gundry et al., 2003; Liew et al., 2004; Twyman et al., 2003]

2. Iberian Peninsula: Geographical and Historical Considerations

2.1 - Iberian Peninsula

The Iberian Peninsula is situated in the southwest of Europe and it is formed by five different countries: Portugal, Spain, Gibraltar (which is under the United Kingdom domain), Andorra and a very small part France (western and northern slopes of the Pyrenean) (figure 3). However, when the term Iberian Peninsula is used, most of the times it just refers to Portugal and Spain. On this work, when it refer to the Iberian Peninsula is regarding only to Portugal and Spain.

The Iberian Peninsula has a surface of 598,081 km^2 [INE, 2012a] and a population of 56,596,729 inhabitants [INE, 2012b].

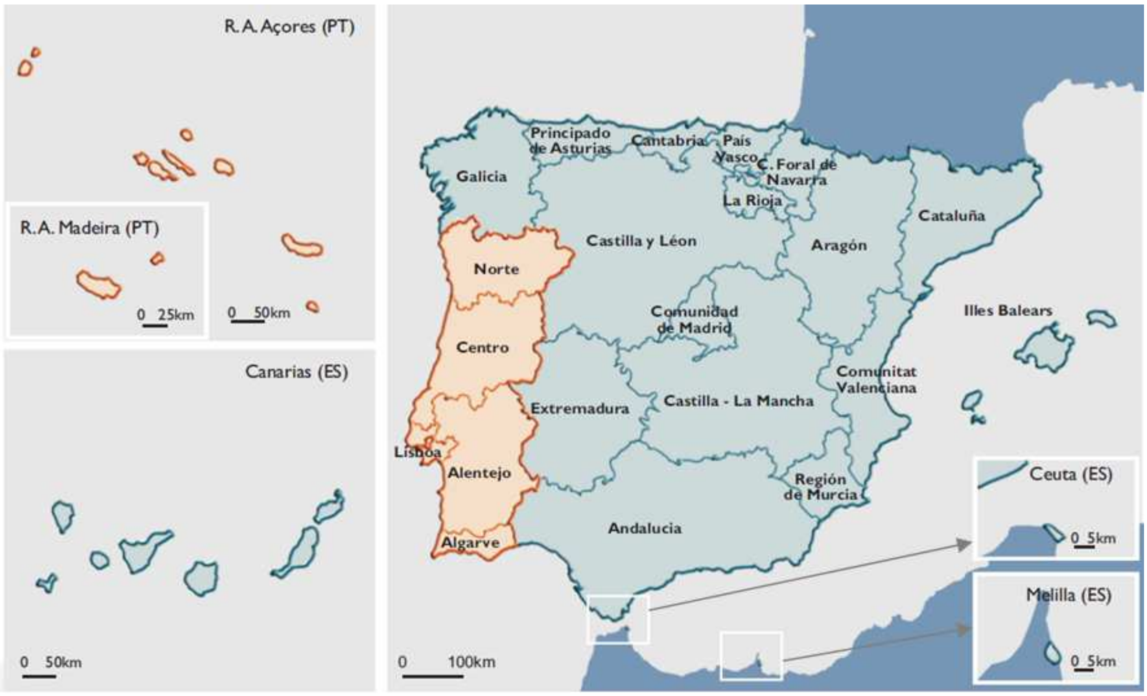

Figure 3 - Map of the Iberian Peninsula, where is differentiated the area that corresponds to Spain (in blue) and the area that corresponds to Portugal (in orange) [INE, 2012b]

The history of Spain and Portugal has always been connected, because for many centuries there wasn't any distinction between the two countries, and the Iberian Peninsula was seen as one geographic entity.

The occupation of the Peninsula by modern humans began 35,000 years ago with settlements in the north and east of the Peninsula. During the period of the last glaciations maxim the human populations were confined to shelters located in Southern Europe, mainly occupying the northern part of the Iberian Peninsula. When this Ice Age ended the re-colonization of Europe began from the southern refuges [Stone, 1990].

The Neolithic period began approximately 7,000 years ago and is characterized by the development of agriculture and livestock. In Europe the first societies appeared in the European south-east between 7,000 b.C and 5,000 b.C., in the central and west Mediterranean from before 6,000 b.C, in central Europe before 5500 b.C and finally in the north-west of Europe before 4,000 b.C. [Ammerman et al., 1984; Whittle, 1996; Chikhi et al., 1998]. To explain this transition, there are two hypotheses: first, called demic diffusion model, that involves the movement of a large human contingent that colonize Europe by displacing or absorbing the European Paleolithic inhabitants, which predicts a major effect on the European gene pool [Semino et al., 1996; Cavalli-Sforza et al., 1997; Quintana-Murci et al., 1999; Simoni et al., 2000; Chikhi et al., 2002; Richards et al., 1996] and the second, called cultural diffusion model, which argues that the transition to agriculture occurred without large movements of populations and without changes in the genetic makeup of the pre-existing populations [Whittle, 1996; Semino et al., 1996; Richards et al., 1996].

In the Bronze Age and Iron Age, there were already several populations settled in the Peninsula: the tartessos, heirs of the megalithic culture and that occupied the south-west of the Peninsula; the iberians, a mix of different populations that extend through the eastern region of the Peninsula; and the celtiberians, people from the celtic culture that arrive in the 13th century a.C. and occupies a broad area of the plateau, reaching the western end when the Iberian tribes occupied the Ebro Valley [Burillo, 2007].

From all this populations, one of the most important in the Peninsula was the Iberians. The called Iberian culture existed from 600 b.C. until 60 a.C., and was characterized by the forms of life of the indigenes societies, enriched by the Greek and Phoenician culture. The called iberos were religious and offered sacrifices to the divinities, develop a consistent economy of agricultural, boosted the exportation of textile crafts, and in 4th a.C. invented their own system of writing [Sobrequés i Callicó 2010; Arribas 1971].

On the 10th century b.C, occurred the first settlement of the first Mediterranean nation: the Phoenicians, whom established in the coast of Andalucia, creating colonies with a commercial nature [Mata, 2002]. The Phoenicians were followed by the Greeks, which were powerful agents of culturalization which lead to the emergence of the Iberian world [Sobrequés i Callicó 2010], the Carthaginians and finally the Romans on the 3th century b.C, that occupied all the Peninsula [Mira, 2000]. The presence of the Roman Empire in the Peninsula was due to the fact that the Greeks, threatened by the Carthaginians, asked Rome for help, which lead to the beginning of their presence in 218 b.C. in Ampurias [Sobrequés i Callicó, 2010].

Between 218 a.C. and 4th century the called Hispanias, name given to the Peninsula by the Romans, were under the Roman domain. The military conquest was followed by a colonizing action that changed the history of the Peninsula. On one hand, the Romanization led to the lost of personality that characterized the indigenous people and have marked the collective identity until today, on the other hand led to the organization of new cities, create a plan of ordering of the Hispania and construct the first network of communication between the different cities [Sobrequés i Callicó, 2010; Tovar et al., 1982].

On the 3th century, the Roman Empire entered in an external and internal crisis and above all, the barbarians started to force the borders of the empire [Tovar et al., 1982; Sobrequés i Callicó, 2010]. The entry of Germanics caused a sudden change in the social and political aspect. This entrance was divided in two phases: the first consisted on Vandals, Alans and Swabians. This last one settled in the northwest of the Peninsula and was the one that achieved greater social and political organization. On a second phase the Visigoths arrived expelling the other Germanic populations on the Peninsula, except the Swabians, nevertheless they were conquered and absorbed by the Visigoth kingdom on the sixth century. The end of the Visigothic period took place at the early eighth century when the peninsula was invaded by Arabs and Berbers from North Africa [Garcia, 1989; Sobrequés i Callicó, 2010].

Islamic domain lasted in the Peninsula for seven centuries, receiving the conquered territory the name of Al-Andalus, during which population from the North Africa transited to the peninsula at different stages [Camps, 1996].

The formation of two sovereign states on the Peninsula occurred due to the need of administrative organization of Leon and Castile, at 1139 the Kingdom of Portugal was established and it borders became definitively established in 1249 [Reilly, 1992].

Later then, the territory known as Al-Andalus began to decrease in extension, due to the successive conquests of the Northern Christian kingdoms. In 1492, the last muslin bastion, Granada, was conquered by the Cristian kingdoms, completing the conquest of the Iberian Peninsula [Quesada, 1989].

The remaining minorities, Jewish and Moorish, that existed in the peninsula suffered a similar end. The Sephardic Jews that did not convert to Catholicism were expelled from the kingdoms of Castile and Aragon in July of 1492 [Quesada, 1989] and although at first they took refuge in Portugal they were later forced to leave this territory in the year 1497 [Stone, 1990]. On the other hand, the Moorish were forced to be baptized in 1502, in an attempt to achieve their integration, however they were banished in 1609 after different conflicts with the Christian population throughout the fifteenth century. Nevertheless the population that choose to converse suffer some problems to integrate in the Christian society [Quesada, 1989].

Ending the Middle Age, all the countries from the occident lived a period of crisis. First, in 1333 a crisis of subsistence, then in 1348 the Black Death took place, with relapses in 1362, 1363 and 1371, which led to a considerable diminution of the population [Sobrequés i Callicó, 2010].

2.1.1 - Catalonia

Catalonia is an autonomous community located on the northeast of the Peninsula, and administratively is divided in four provinces: Barcelona, Gerona, Tarragona and Lérida. The community appears surrounded by France and Andorra on the north, the Mediterranean Sea on the east, at south by the Valencia Community and west by Aragon Community. Cataluña has a surface of 32,091 km^2 and a population of 7,539,618 inhabitants [INE, 2012a].

On 600 b.C., the Greeks create the first colony on the catalan Mediterranean coast, named Emporias (Ampurias), which meant market. In a few years, what began like a small market was then a consolidated urban structure. After the Greeks occupation, existed the Roman Empire, which ruling was retarded in Catalonia by the Iberos, starting only in 195 a.C. after a long battle [Sobrequés i Callicó, 2010].

The Muslim Berbers, coming from the North of Africa, started the occupation of the Catalan territory between 713 and 720, until 1148-1149 when the state of Catalonia was finally recaptured by the Count of Barcelona Ramón Berenguer IV. The beginning of the recaptured of the capital territory was due to the sovereign Carolingian, which ruled from the other side of the

Pyrenees, but in the beginning of the 9th century, the counds started to become independent [Reilly, 1992].

This way, during the 11th century are configured the traits that characterized a national reality, in other words, a common origin, a land, an economical and structured social life defined, a community that expresses themselves in the same language but above all, a knowlodgment by the people of this fact. On the 12th century, the terms "Catalan" and "Catalonia" were already used to nominate the people and land ruled by the Count Ramón Berenguer III [Reilly, 1992].

After this period, Catalonia was during some centuries a region that passed from kingdom to kingdom or even try to have their own independence, which was always attack by other kingdoms. During the 18th and 19th century, the distinct governments of Spain dismantled what was left of the state of Catalonia and tried to erase their characteristics traits [Sobrequés i Callicó, 2010]].

Between 1609 and 1610, around five thousands Catalans Moriscos, namely, ancient Muslims that were forced to convert, left towards Africa, due to the religious intolerance that was lived in the peninsula [Sobrequés i Callicó, 2010].

The autonomy of Catalonia was only approved in 1932, but in 1936 there was still some competence under de power of the government. This ended with the military rebellion on 1936, which led to the civil war of 1936-1939. During the dictatorship, the creation of new industries in Catalonia led people to migrate to Catalonia, leaving regions more undeveloped. The dictatorship led many Catalan to run to France, although their situation deteriorates with the German occupation of France. The general Franco had a special attention on constrain any show of Catalonia (language and symbols) [Sobrequés i Callicó, 2010].

2.1.2 - Andalucía

Andalucía is an autonomous community located on the south of the Peninsula, and it is composed by eight provinces: Huelva, Cadiz, Seville, Córdoba, Malaga, Jaen, Almeria and Granada. The community appears surrounded by west to east by Portugal, Extremadura Community, Castilla-La Mancha Community and Murcia Community, at south by the Mediterranean Sea and the Atlantic Ocean [INE, 2012a].

Around 1000 b.C. the Phoenicians, Greeks and Carthaginians settled on the area of Andalusia. In the sixth century b.C., the Roman Empire nicknamed the area as Andalusia, and during this time the region was rich due to the exportations of wine and olive oil [Tovar et al., 1982; Moreno, 2001].

After the invasions of the Vandals and the Visigoths, the Muslims from the North of Africa invaded Andalusia in 711 a.C. The Muslim presence lasted for eight centuries and left an important and still visible mark in the population and in the culture of Andalusia. The Muslims established an Emirate in Andalusia, with his capital in Cordoba, which then became independent from Damascus. This period represents one of the most flourishing moments of Andalusia, with great social, cultural and economic prosperity [Reilly, 1992; Gill et al., 2008].

On the 11th century, the caliphate became weakened by civil wars, ending in the conquest of the area by the Catholic kings, ending the conquest of Granada in 1492. The current territory of Andalusia just appeared after the War of Alpujarras (1570-1572), when happened the expulsion of all muslims from the region [Reilly, 1992; Gill et al., 2008].

After centuries of splendor lived in Andalusia, the golden period ended which led to the fall of what was an important region of Spain. In 1980, Andalusia acquired it own autonomy and began a slow recovery that latest until today. Andalusia is one of the communities of Spain with the higher unemployment rate, which have create a great amount of emigration, however the region is still looked for by the tourists due to the his important cultural history [Gill et al., 2008; Moreno, 2001].

2.1.3 - Zamora

The province of Zamora is located at the north western edge of the Central Plateau, in the autonomous community of Castile and Leon. The province appears confined to the north by Leon, east of Valladolid, Salamanca on the south, and at west by the province of Orense and Portugal. Zamora has a surface of 10,561 km^2 and a population of 193,383 inhabitants [INE, 2012a].

In the 3th century b.C, when the Roman Empire enter the Peninsula, in the province of Zamora coexisted three types of celtiberians: the vetones [Bosch, 2003]; the vacceos, that are considered the first stable population settled on the area [Wattenberg, 1959]; and the astures, that were the lasts one to the annexed to the Roman Empire [Bosch, 2003].

On the 5th century, with the fall of the Roman Empire the Germanic population enters the Peninsula. The Germanic distributed themselves on different zones of the province: the suevos on the occidental area and the vandalos occupied the rest of the province [Garcia, 1989]. This territorial distribution existed until the entry of the Visigoths, that first occupy the held territory of vandalos, and just in the end of the 5th century the Visigoths although took the area occupied by the suevos [Barbero et al., 1974].

On the beginning of the 8th century, with the entrance of the Muslims, Zamora was only used by them to control the trims and the forts used since the Roman Empire to defend themselves from the north population. Taking advantage of this situation, Alfonso I the Asturias invaded different cities, as Zamora, killing the muslim population and took the Christians to repopulate the Cantabria mountain and the coast zone of Galicia [Barbero et al., 1974]. These factors led to the lack of human settlements in de Douro line, acting this way as a frontier between the Astur kingdom and the muslim Kingdom [Menendez, 1960; García, 1982].

During the next three centuries (9th to 11th) Zamora was definitively annexed to the kingdom of Asturias, in two different stages. During these stages, Zamora suffered different immigration flows, composed by mozárabs from al-Andalus, population coming from Asturias, Leon and Galicia on a first stage; and on a second stage, once again, the mozarabs from al-Andalus, population coming from Asturias, Leon and Galicia, but as well jews from al-Andalus and settlers from France [Lorenzo, 1995].

During the 20th century, Zamora have lost almost a third of its population by different factors: high mortality caused by a flu pandemic, mortality and movements due to the civil war, and the increase of emigration to industrial areas due to the lack of progress, what led to the marginalization of the province until today [Redondo et al., 1995]. All this populations lost didn’t occur in an uniform way through the province, being affected the areas that possessed less resources.

2.2 - Genetic Diversity and Structure of the Iberian Peninsula

As explained previously, over the millenniums, the Iberian Peninsula has been influenced by multiples cultures and people of different origins, which configure it present day genetic architecture.

The first studies within the Iberia Peninsula have shown that, in terms of nuclear gene frequencies, blood groups and enzyme polymorphisms, the Basques are differentiated from neighbouring populations [Bertranpetit et al., 1991]. This has contributed to the ongoing debate on the origin of this isolated population with its linguistic and social distinctions. In accordance, analysis of monoparental genetic markers, the Y-chromosome and mitochondrial DNA, reveal that the Iberian Peninsula reflects a homogenous background, except in some areas like the Basque Country, Pyrenees and South of Portugal [Arroyo-Pardo et al. 2007; Pino-Yanes et al., 2011].

The Y-chromosome is characterized by high levels of haplogroup R1*(xR1a) and lower levels of haplogroups I*-M170, J*-12f2.1 and E*-M35 [Gonçalves et al., 2005; Arroyo-Pardo et al. 2007]. The haplogroups R1* and I* are linked to the Palaeolithic and show the highest frequencies in the Northeast of the Iberian Peninsula, the Basque Country and the Pyrenean populations [Arroyo-Pardo et al. 2007]. The distribution of the African haplogroups is highly variable across the Iberian Peninsula. For example, high frequencies of haplogroup E3b1* have been explained as the result of recent gene flow due to the long-term Islamic occupation of the peninsula, yet the presence of the haplogroup E-M81 in the North of the Peninsula is difficult to explain because the historical background does not support the presence of African individuals on that region. The presence of haplogroup E-M81 has been explained as consequence of prehistoric links between Iberians and North Africans, according to mtDNA analysis [Arroyo-Pardo et al., 2007].

Adams and collaborators [Adams et.al, 2008], using Y-STR haplotypes and admixture analysis showed the presence of high mean levels of admixture on North African and Sephardic Jews patrilineal ancestry in modern Iberian Peninsula population, 10.6% and 19.8%, respectively. This admixture is also showed in a study conducted by Nogueiro and collaborators [Nogueiro et al., 2010], on which analyzed Y-SNP and Y-STR and unveiled that Jew population from Trás-os-Montes (NE of Portugal) show high frequency of haplogroup R1b1b2, which is less frequent in other Jewish populations, indicating a significant level of admixture with non-Jews Iberian populations. This Jew population also showed high frequency of haplogroups T and J, which are uncommon to Portuguese population, yet it reflects their origin in Middle East [Nogueiro et al., 2010].

The mtDNA of the Iberian populations indicates a Palaeolithic origin with a minor Neolithic contribution. Preliminary mtDNA studies showed that the Iberians are different from the rest of the Europeans under genetic diversity, having present unique lineage groups, like the Basque country and Catalonia, and also showed evidence of geographical subdivision. The Peninsula exhibits relatively higher proportion of lineages of African origin, especially haplogroups U6, L3, M1, L1b and L2, which has been interpreted either as a pre-historic African colonization or as historical migrations from Africa [Arroyo-Pardo et al., 2007; Alvarez et al., 2010]. Ruling out the Islamic occupation and the slave trade, no historical migrations are considered to have influenced the mtDNA gene pool of the Iberian Peninsula [Arroyo-Pardo et al., 2007]. Nevertheless the Iberians share the majority of their haplotypes with the Europeans (haplogroups H, U, J, T, W and X) [Arroyo-Pardo et al., 2007; Corte-Real et al., 1996].

Concerning more recent studies [Moorjani et al., 2011] based on genome-wide data, results indicate that the highest proportion of African ancestry in Europe is observed in the Iberian Peninsula, with values of 3,2±0,3% in Portugal and 2,4±0,3% in Spain; which are consistent with the inferences based on Y-chromosome and mtDNA that showed that that within Europe the Southwestern Europeans have the highest haplotype-sharing with Africans [Moorjani et al., 2011; Pino-Yanes et al., 2011].

A study conducted by Gayán and collaborators [Gayán et al., 2010], at which 262,264 SNP were genotyped using Affymetrix NspI 250K chip on the Spanish population, showed that this population is largely homogeneous within itself. However, patterns of micro-structure may be able to predict locations of origin from distant regions. Data obtained allow to differentiate between the two most apart geographically centers [Gayán et al., 2010], which corresponds to a south to north-east axis, which is in line with a south-west to north-east axis previously found in European populations [Seldin et al., 2006; Bauchet et al., 2007; Tian et al., 2008; Novembre et al., 2008; Tian et al., 2009].

II. Aims

The overall aim of this work is to study at what degree European AIMs can be applied in microgeographic differentiation. In this sense, the working line was divided into two steps, which forms in this dissertation two chapters.

The aims of chapter I - Primer design and validation for analysis of European AIMs, are:

1) To select a set of SNP that would allow characterizing and differentiating the populations of the Iberian Peninsula;

2) To design primers to amplify the regions encompassing the selected SNPs and that can be used to HRM methodology;

3) To obtain the three possible genotypes to each SNP, which would be helpful as controls ahead.

The aim of chapter II is to preliminarily genetically characterize different Iberian populations, namely, Catalonia, Andalucia and Zamora, using some of the selected SNPs.

III. Chapter I - Primer design and validation for analysis of European AIMs

1. Material and methods

1.1 - SNP selection

The SNP selection was based on the historical background of the Iberian Peninsula. In this sense, there were considered some articles that mentioned the existence of an northern-southeastern axis that shows substructure in Europe [Seldin et al., 2006; Bauchet et al., 2007; Tian et al., 2008; Novembre et al., 2008; Tian et al., 2009; Gayán et al., 2010]. Moreover, it was considered the genetic contribution of the Sephardi Jews, as their presence was longstanding, with some evidence of their presence predated the Christian era, yet it is believed that many Jews arrived during the Roman Period, as slaves or voluntarily and later due to the Islamic invasion of their homelands in the Near East [Adams et al., 2008].

1.2 - Primer design for HRM

Primer design was performed using the ApE-A Plasmid Editor v1.16 and the Primer Designer version 1.01 software. In order to have a good resolution at the HRM analysis, the criteria used in primer design were: the melting temperature (T_m) must be at least 56°C, the difference of T_m of primers to each SNP can't be higher than 1°C, the content of CG must be between 40% and 60% and the PCR product should have between 70 and 150 base pairs.

To check to the formation of primer-dimers and hairpin structures the Windows32 Primer select 4.05 from DNASTAR package was used. The melting temperature of the primers was adjusted using the Eurofins MWG Operon PCR & Sequencing Primer Tool (http://ecom.mwgdna.com). Finally, to ensure the target region to amplify, the primers were submitted to an In-Silico PCR at the UCSC (University of California, Santa Cruz) (http://genome.ucsc.edu/cgi-bin/hgPcr).

1.3 - Control Samples

A set of 30 random samples of DNA from human populations from the Iberia Peninsula were amplified, purified and automated sequenced in order to obtain samples with a know genotype that would be further used as control samples in HRM typing experiments. Subsequently, a HRM was performed in order to establish the melting curves and temperatures for which genotype and SNP.

1.3.1 – PCR amplification, purification and sequencing

The PCR mix for each sample consisted of 17 µL of water, 1 µL of 10 pmol of primer forward plus 1 µL of 10 pmol of primer reverse, 2.5 µL of 10x NH_4 Reaction Buffer, 1 µL of 50mM $MgCl_2$ solution, 0.25 µL of 100mM DNTPs, 0.25 µL of 5u/µL TaqDNA polymerase and 2 µL of DNA, totalling a final volume of 25 µL.

The PCRs were performed in a Primus 96 plus thermocycler and the amplification program consisted of an initial denaturation step of 5 minutes at 95°C, followed by 35 cycles of PCR [30s at 95°C, 30s at annealing temperature (T_a) and 45s at 72°C], and a final extension step of 5min at 72°C. The PCR amplification results were visualized by electrophoresis in agarose gels (2%).

PCR products were purified using the JETQUICK Spin Column Technique (Genomed), according to the protocol provided by the supplier. The samples were then sequenced using the BigDye® Terminator v3.1 kit. The sequence reaction was performed according to the manufacture specifications and were further purified using an EDTA/Ethanol protocol according to the recommendations available in the BigDye® Terminator v3.1 Cycle Sequencing Kit Protocol. The sequence electrophoresis was run in an ABI 3130XL sequencer at the Servei de Genòmica, Universitat Autònoma de Barcelona.

Sequence quality was analyzed using the Sequence Scanner v1.0 (Applied Biosystems). The alignment of the obtained sequences was performed using the BioEdit sequence alignment editor [Hall, 1999].

1.3.2 - Analysis by High Resolution Melting (HRM)

The PCR reaction mix for each sample consisted of 5 µL of 2x HRM PCR Master Mix, which consists on HotStarTaq Plus DNA polymerase, Type-it HRM PCR Buffer, EvaGreen, Q-Solution, dNTP mix; 0.35 µL of 10 pmol of primer forward plus 0,35 µL of 10 pmol of primer reverse; 3.3 µL of RNase-free water; and 1 µL of DNA, making up a final volume of 10 µL. The samples were then introduced in Rotor-Gene Q, a QIAGEN'S real-time PCR cycler, following the protocol of Quiagen for HRM analysis of SNPs, which consist on a initial PCR activation step for 5min at 95°C, followed by 40 cycles of a two step cycling [first step is to denature for 10s at 95°C; the second step is annealing/extension for 30s at 55°C, during which it is necessary to activate fluorescence data acquisition on the green channel], and finally the step of HRM for 2s with the ramp of temperatures from 65°C to 95°C with 0.1°C increments, during which there is fluorescence data acquisition.

It is recommended that after determining the T_m for each sequence, the subsequent HRM may be run between 5°C below the T_m to above 5°C the T_m, which may reduce the time needed for analysis.

1.4 - Ancient samples

A set of six ancient DNA samples were analyzed by HRM using the primers and conditions optimized with modern samples.

2. Results

2.1 – SNP Selection

The selection of SNPs was done based on the available literature, ending on a total of twenty-six SNP chosen (table 1). Twenty-one SNPs were chosen from the paper of Bauchet and collaborators [Bauchet et al., 2007] and five SNPs from the article of Raskin and collaborators [Raskin et al. 2008].

From the study of Bauchet and collaborators [Bauchet M. et al., 2007], the twenty-one SNPs selected were chosen from a full set of 1200 European AIMs (EuroAIMs). The selected SNPs presented the highest F_{ST} values between the northern and southeastern European populations.

From the article of Raskin and collaborators [Raskin L. et al., 2008], five SNPs which allows to differentiate Sephardi Jewish populations were chosen.

2.2 – Primer design and PCR optimization

A total of twenty-six primer pairs were designed, according to the specifications for HRM PCR primer design (table 2).

During primer design some major problems were found because the pair of primers to each SNP shouldn't differ more that 1°C in Tm and the PCR product should had a size between 70bp and 150bp. Moreover, the formation of primer-dimers and hairpins represent another challenge, since some primers pairs followed all the other specifications but failed by leading to the formation of primer-dimers and hairpins.

The primer pairs obtained were submitted to in-Silico PCR to ensure that the region amplified was the desired region. All the pair primer amplified only the region for which the primers were designed, leading only to the amplification of the target region.

Table 1 - Description of the SNP chosen: chromosome location, sequence and nucleotide mutation

Literature reference	**SNP**	**Chromosome location**	**Sequence and nucleotide mutation**
Bauchet M. et al., 2007	rs988436	5	ACTACAGCTATACTCCTCAAATCTAT[G/T]TGTATAGTCGACACAAGAGTTTTAC
	rs942793	10	ACAGCAAATGGCAAAGGTTCTCAGTC[A/C]CACCTGGCAGACTTGTACCCTAGCT
	rs1368136	8	AAAGGATCAGGAGAAGCTTTCCAGAG[C/T]AGGTAATGTCAGACCTGGGTCTTAA
	rs2060983	8	TTTGTTAGCCTAGGATGCATGAGGTT[A/G]AGAGCAATATTTTTTTTTTCCTCTC
	rs4988235	2	CTGGCAATACAGATAAGATAATGTAG[C/T]CCCTGGCCTCAAAGGAACTCTCCTC
	rs1404402	1	GGTTGAAAATAGACTGAAGGGAAAAA[A/C]GGGTGGAAAGTCAGAGACCCAGATA
	rs1016120	2	CATGGATTTACATTCTTACCAACAGC[C/T]CTCAGTGAACCCTGCCCCTGCCCCC
	rs1414411	1	TAAATTTTCGCTTTGTTGGATTGACT[C/T]AAACTCTTTTAAATACATTTAATGG
	rs2014303	4	CATGCTGAGTGCTGGTTTTGTACTCC[A/C]ACATGCTCTATAGTGAAACTTTCAA
	rs1030626	8	GGATCCAGGGGCTCAACTGATGCTTT[C/G]TGTACGTGCTTCCTTCCCTCCTTGG
	rs1517661	12	GGATGGAAAGCAGCACATGTGTCCTG[A/C]AGAAATGACGACCAGTCTAGAGCAA
	rs764138	16	CCATTAATATTCAGAAAGAATACCCA[A/G]AGAAACCATGAAATCTTTCCTGAAA
	rs2218497	13	CCAGAATCATCCTGATACCAAAACCC[A/G]AAAGAGACATAACAAAAAGGAAACT
	rs725379	2	ATGTCAAAATGCCTCCCGATGACTAA[C/T]TTCAACAGCTAAATGGAAAATGGTG
	rs1377724	15	CAAACGAAAAATATCTAGCACAGATG[C/T]TATATCTCTAGTTTTAACTGCTCTT
	rs1406121	2	CTGCACCCAGGGGGAGGATGAACCCA[A/G]TGGCTTCCATGCATGACAAGTCAGG

	rs869538	4	GGACTGGAGCCCTTGGCTTCTGATAT[A/G]GCATAATACCAAGAAGACAATTTCT
	rs1905471	13	GTTCACTTCAGGCATCCTTCCCTTAC[A/C]AGCTGAATGTCCCTAGGAGATCATA
	rs764681	16	ATCTCGGACATAGGCTTAGCAATCTG[C/T]AGTTTCACCCCTCGCCAGGTAATTC
	rs1280100	4	AATTAAGAAATTGAAATTCATAACAC[A/G]CAAATGTAAGGGAAGAGCTGCTTAC
	rs723211	10	AGCGAAGAGTCACTCTTGGGGTATCA[A/G]TGAGAGAAGAGGCTGGGCATCAGAG
Raskin L. et al., 2008	rs11200014	10	TCCAAAAAAAGATGCACAGAGGGAAG[A/G]TTTTCTCTCTGGTCCCAACACGTGT
	rs2981579	10	CTTGCTTGTTTTGGATACGACCTCTG[C/T]AGACATCCAGGTTATGTATTTCCTC
	rs1219648	10	AAGCACGCCTATTTTACTTGACACAC[A/G]CTCTTCAAGGATGGCCATGGCTTGT
	rs2420946	10	CATCGACCTCCTTCCTGATCCATAAG[C/T]GCATCCACAGATCCCAGGTTGAGCA
	rs2981582	10	TCATCGCCACTTAATGAACCTGTTTG[A/C/G/T]GGAGAGTCCACCTGGTGCCTGCCTG

Table 2 - Primers designed to amplify specific regions, on which existed the SNP previously selected. All the primers obey to the specifications previously described. Melting temperatures (Tm) and fragment length for each pair of primers is also presented

SNP	Fragment length	Primer Forward		Primer Reverse	
		Sequence 5'-3'	Tm (ºC)	Sequence 5'-3'	Tm (ºC)
rs988436	138	CCTATCACTAGAAGAAACAGACAC	59,3	GACAGGTGTCATGTACCTACATA	58,9
rs942793	87	TCTGTCTGGCTTGTGTCTG	56,7	CAGCTAGGGTACAAGTCTG	56,7
rs1368136	114	GGAGAATTCAGAGAGGGATACT	58,4	CTTCCTTAGCAAACTCACAGTC	58,4
rs2060983	128	GTCTTAAAATGGGTGATTGGCCTC	61,0	CCTTCACTGCCTCTAGAAAGAGAA	61,0
rs4988235	95	TTAGACCCTACAATGTACTAGTAGG	59,7	GCAACCTAAGGAGGAGAGTTC	59,8
rs1404402	126	GCAGAGGGATCACAAGATGT	57,3	AGATTATTGCAGGGGCTGTG	57,3
rs1016120	137	GCTGCAGCTAAATTGCCTCTCAAA	61,0	GGAGAAGAATCCCAGTAGAGGTAA	61,0
rs1414411	282*	CTAACCAAATGGCCAAAATTGACTG	59,7	CAGACTGAGCAGATGGCACA	59,4
rs2014303	142	CCACTTTCACTTTTGCTTTTGCC	58,9	AAGCGATGGAGTCATAAGGAGTT	58,9
rs1030626	144	GGAAAGGGAAAGGGAGGATG	59,4	CCATGAATTCCGTACTGCATAAAG	59,3
rs1517661	101	CCAGGAACAATTGTTCTGCTG	57,9	CCTGTTCCTTGGTATTGCTCT	57,9
rs764138	170*	CTCTTGGCCCTAATACCCAT	57,3	CTACAGGGCCATCTTCATGT	57,3
rs2218497	149	GCTGGTACCATTCCTGCCAA	59,4	GTTCATCAAGGATATTGACCTGAAG	59,7
rs725379	135	AAGTAAAATCAGCAGAACCTGG	56,5	CTGATTCTCCAACAGAGGG	56,7

rs1377724	100	GGGTACAGAGAAGGTTGGAA	57,3	GGTCAAGATCTCAGAAGAGC	57,3
rs1406121	115	ATGTCCTGAGATTCCCAGCA	57,3	TGCTCTTCAGGATTTGGGTG	57,3
rs869538	141	GCTACAATCTATGCAGCAATTTCC	59,3	CACAAGGCTAGTATGCCATTAGAA	59,3
rs1905471	94	AGGAGCATGCTCAGTGTGTTT	57,9	GGTATATGATCTCCTAGGGAC	57,9
rs764681	98	CATCACCTGGACACTTATTTGC	58,4	CATGTATCAGAATTACCTGGCG	58,4
rs1280100	102	CACCACAGCGCCAAATGAT	56,7	GTCCAGTTTGTGAAAAGGTAAG	56,5
rs723211	130	GGATGTCACTCAGGACTCA	56,7	ATGATGCTCTGATGCCCAG	56,7
rs11200014	95	GATCTGATGTTTTCGGCTGTTCA	58,9	GACTTTAATACACGTGTTGGGAC	58,9
rs2981579	96	AGAAGAGGCTGGTGGAGGA	58,8	CCGGGGGGAGGAAATACAT	58,8
rs1219648	102	TTCCATGGTACCGGTTTCCCAA	60,3	ATTGGGACAAGCCATGGCCAT	59,8
rs2420946	138	CTCATAGCTCATGGAAACTATAAACCC	61,9	AGGCCTGCTCAACCTGGGAT	61,4
rs2981582	100	GAGAATAAAACGGCAGATCCCA	58,4	GACTGCTGCGGGTTCCTAA	58,8

2.3 – Control Sample analysis

Resorting to the Weir's unbiased (F_{ST}) values for each SNP as a measure of genetic distance between two groups, the five SNPs with the highest F_{ST} values were selected for laboratory validation. The sets of primer for these five SNPs (table 3) were tested for PCR amplification. The conditions of PCR, including the annealing temperature (T_a) (table 3), were optimized and no problems of amplification or specificity were detected for the five regions amplified.

A set of 30 samples from different human populations were randomly selected to establish control samples. These samples were further amplified and sequenced, and analyzed by HRM. This analyze allowed the compilation of the melting temperature (Tm) for each SNP (table 3).

Table 3 - Annealing temperature (T_a) for Polymerase Chain Reaction (PCR) and melting temperature (T_m) for High Resolution Melting (HRM).

SNP	T_a (ºC) for PCR	T_m (ºC) for HRM
rs988436	56,3	78,4
rs942793	53,7	80,3
rs1368136	55,4	78,5
rs2060983	58	76,8
rs4988235	56,7	78,7

The sequence of the region encompassing each SNP in different samples allows classifying the sample according to it genotype (fig. 4). Afterwards, HRMs were performed with the samples which were sequenced and it genotype identified. The HRM allowed establishing for each SNP a melting pattern for the different genotypes. This would be helpful in the future for the classification of other samples. By the analysis of the figure 5 it is possible to infer that in a HRM analysis the homozygous melting curve appear parallel to each other and that the heterozygous curve begins with one of the homozygous and ends with the other, and it cross the space left between the two homozygous. Moreover, the HRM analysis allows determining the melting temperature of HRM experiments for each SNP (table 3), a parameter that would be useful in subsequent experiments reducing the time of analysis.

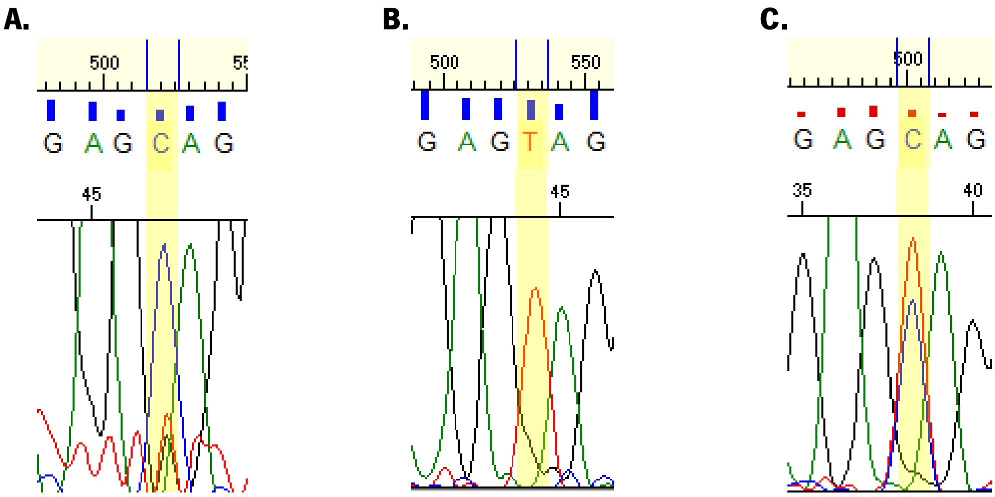

Figure 4 – Sequencing of three control samples for a specific region of one of the SNP study, on which it is possible to identify the three possible genotypes: CC (A); TT (B), CT (C)

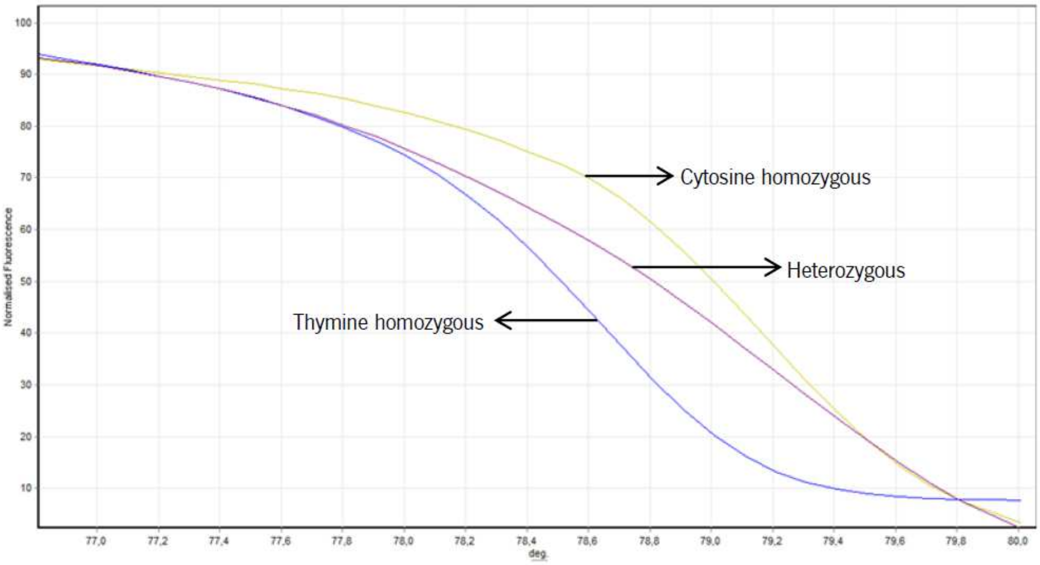

Figure 5 – HRM of three samples with distinct genotypes: TT (blue), TC (violet) and CC (yellow). This graph allows the establishment of a pattern for each genotype for each SNP studied.

2.4 – Ancient samples

The primers and conditions optimized for control samples were tested to amplify ancient samples from Bronze and Iron Age periods and positive amplifications were obtained (fig. 6). These results point that these primers and genotyping method can be used successfully in ancient samples and this point will be investigated in the future.

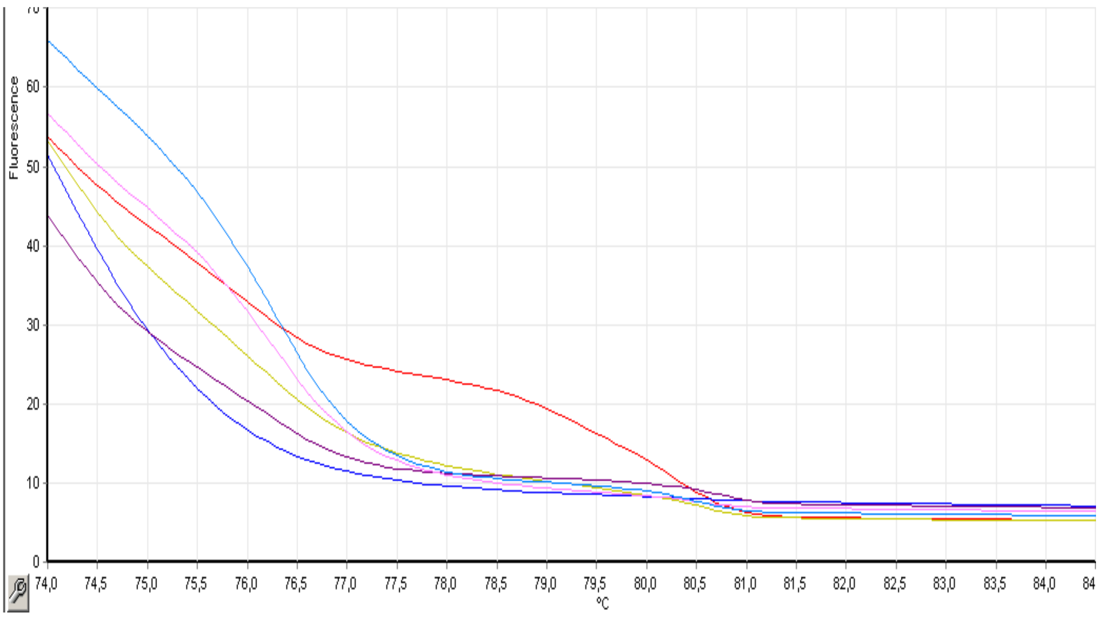

Figure 6 - HRM of samples from Bronze and Iron Age periods, which show positive amplifications.

3. Discussion

The majority of selected SNPs presented the highest F_{ST} values between the northern and southeastern European populations and five allows differentiating the influence of Sephardi Jewish populations [Bauchet et al., 2007; Raskin et al. 2008]. Thus, with this set of twenty-six SNPs we expect to see different patterns of genetic frequency of the SNPs in different Iberian populations.

Relatively to the SNPs of the Sephardi Jewish populations, is expected to see an obvious influence on Catalonia and Zamora, due to the fact that these two regions are described as places of settlement for the Jews during the inquisition time. During this time, they were forced to convert to Catholicism but yet many of them maintained their beliefs in secret [Quesada, 1989; Stone, 1990].

Regarding the non-inclusion of SNPs for the African population, it was due to the fact that the African influence in the peninsula has been studies using mainly monoparental genetic markers.

Among the SNP genotype techniques the one chosen for this study was HRM. This technique presents a number of advantages as: cost advantages derived from the simplicity of the method; doesn't require specific labelled probes; is a closed-tub method where no processing is required between amplification and analysis, on which the risk of contamination is eliminated; hybridization is monitored over a range of temperatures instead of at a single temperature; after PCR amplification, the melting curves are generated by monitoring the fluorescence of a saturating dye that doesn't inhibit PCR; is a non-destructive method which allows a subsequent characterization using gel electrophoresis or sequencing [Tayler, 2010; Vossen, 2009; Reed, 2007].

Comparing the HRM with other techniques, we can identify several disadvantages of other techniques such as SSCP, TGCE, DGGE and DHPLC. All of them need extensive automation, instrumentation and analysis and all require the separation of the gel or matrix, during which there is a high risk of contamination because the PCR product becomes exposed to the environment; other techniques, such as primer extension and invasive cleavage, demand the use of specific probes which have to be design [Reed, 2007; Kwok, 2001].

Thus, and taking into account that the studies that already used this technique [Zhou et al., 2005; Erali et al., 2008; Takano et al., 2008], the selection of HRM as the technique for this study seemed the right one. Taking into consideration all the specification required for the HRM, a set of primers has designed for all the SNP selected. A handicap of this technique is that the SNPs have to be analysed one by one, which depending on the number of SNPs to be analyzed can be a time consuming technique. Although this handicap, when analyzing ancient samples the individual analyze of each SNP is important to ensure obtaining results.

In order to enable the reading of results of the HRM is very important to establish control samples, which are used as reference samples to each genotype. In all subsequent HRM, this reference samples are also embedded facilitating the classification of samples when, for example, genotype assignment is not straightforward. This control samples also allowed to reduce de time of the HRM, as each sequence for a specific SNP has a specific melting temperature, and in further HRM cycles the range of temperature can be decrease, leading to a small time consuming process.

Regarding the ancient samples, the positive amplifications obtained let us to think that HRM can be an alternative to SNP genotyping of degraded. It would be important, however, to further prosecute this technique with ancient samples.

IV. Chapter II - Iberian Human Population Structure: the European AIMs contribution

1. Material and Methods

1.1 – Samples

The number of samples analysed by SNP and population is described in table 4. The samples are from individuals of both sexes and from the autochthon population, checked until the previous two generations. All the participants gave their written informed consent to participate in the study.

The DNA extraction was performed using the Jetquick Blood/Cell Culture DNA Spin Kit (Genomed) according to the manufacture specifications.

For comparative purposes, genotype frequencies for each SNP were obtained from the 1000 genomes projects platform [1000genomes, 2013] for the populations listed in table 5.

Table 4 – Number of samples analyzed taking in consideration the SNP and the population

	rs988436	**rs1368136**	**rs4988135**
Catalonia	31	31	30
Andalucia	49	55	31
Zamora	33	33	32

1.2 - SNP analysed and methodology

The SNP analysed were rs988436, rs1368136 and rs4988235. The genotyping methodology used was High Resolution Melting using the optimized conditions detailed in the previous section *Chapter I - Primer design and validation for analysis of European AIMs*.

1.3 - Data analysis

1.3.1 - Intra-population analysis

Conformity with Hardy-Weinberg equilibrium was tested for each SNP and population. For probability interpretation a Bonferroni correction for multiple tests was applied. Hardy-Weinberg equilibrium states that the original proportion of the genotypes in a population will remain constant from generation to generation, as long as: no mutations take place; no immigration or emigration takes place; random mating is occurring; the population size is very large and no selection occurs. If these parameters are achieved and the genotypes proportions do not change, it is said to be in Hardy-Weinberg equilibrium. If a population is not in Hardy-Weinberg equilibrium one or more evolutionary forces are operating [Raven et al., 2011; Guo et al., 1992].

Nei Gene Diversity [Nei, 1987] was calculated as a molecular diversity indicator. Gene diversity is equivalent to the expected heterozygosity for diploid data and it is defined as the probability that two randomly chosen haplotypes are different in the sample [Excoffier et al., 2005].

All the mentioned analysis were performed using Arlequin ver. 3.0 [Excoffier et al., 2005].

1.3.2 – Inter-population analysis

The population comparisons were made by computing pairwise F_{ST} values [Reynolds et al., 1983; Slatkin, 1995] and by means of an exact test of population differentiation [Raymond et al., 1995]. This analysis were performed using Arlequin ver. 3.0 [Excoffier et al., 2005].

To evaluate the relation between populations using the combined information of all the SNPs analyzed, a Principal Component Analysis (PCA) was used. PCA consists on reducing the dimensionality of a data set consisting of a large number of interrelated variables, while retaining the variation present in the data set [Jolliffe, 2002]

Table 5 - Number of samples for each population used in population comparison taking in consideration the each SNP

Population Description	**Population Code**	**Super Population**	**rs1368136**	**rs988436**	**rs4988235**
Han Chinese in Bejing, China	CHB	East Asian	97	97	97
Japanese in Tokyo, Japan	JPT		89	89	89
Southern Han Chinese	CHS		100	100	100
Utah Residents (CEPH with Northern and Western European ancestry)	CEU	European	85	85	85
Toscani in Italy	TSI		98	98	98
Finnish in Finland	FIN		93	93	93
British in England and Scotland	GBR		89	89	89
Iberian population in Spain	IBS		14	14	14
Yoruba in Ibadan, Nigeria	YRI	African	88	88	88
Luhya in Webuye, Kenya	LWK		97	97	97
Americans of African Ancestry in SW USA	ASW		61	61	61
Mexican Ancestry from Los Angeles, USA	MXL	Ad Mixed American	66	66	66
Puerto Ricans from Puerto Rico	PUR		55	55	55

Colombians from Medellin, Colombia	CLM		60	60	60

2. Results

The values of allelic and genotypic frequencies for the three studied populations are described in table 6, table 7 and table 8 for the SNPs rs988436, rs1368136 and rs4988235 respectively.

For the SNP rs988436 the allele T is the one that shows the highest frequency, ranging from 0.8 to 0.9. Regarding the genotypic frequency, it does not exist homozygous G on any of the three populations and the three populations show higher frequency of homozygous T over the heterozygous.

For the SNP rs1368136 Catalonia and Zamora have higher values for the allele C over the allele T, however the opposite occurs in Andalucia that show a higher level of allele T. Still, Zamora has a bigger difference between the allele frequencies than Catalonia, what also happens it the case of Zamora. Regarding the genotypic frequency, Catalonia shows a higher frequency of homozygous C followed by homozygous T; Andalucia shows the greater frequency on heterozygous followed by homozygous T; and Zamora shows a higher frequency of homozygous C followed by heterozygous.

For the SNP rs4988235 Catalonia showed a higher frequency of the allele T, yet the difference between the two alleles is very low; on Andalucia and Zamora the allele C showed the higher frequency, ranging from 0.60 to 0.67. Regarding the genotypic frequency, the three populations showed the higher frequencies of heterozygous, but on Catalonia the next higher frequency belongs to homozygous T while in Andalucia and Zamora showed of homozygous C and very low frequencies on homozygous T.

Table 6 – Allelic frequency and genotypic frequency of the three studied populations regarding the SNP rs988436

	Allelic frequency		**Genotypic frequency**		
	Allele G	**Allele T**	**GG**	**TT**	**GT**
Catalonia	0.10	0.90	0.00	0.81	0.19
Andalucia	0.16	0.84	0.00	0.67	0.33
Zamora	0.18	0.82	0.00	0.64	0.36

Table 7 - Allelic frequency and genotypic frequency of the three studied populations regarding the SNP rs1368136

	Allelic frequency		Genotypic frequency		
	Allele C	Allele T	CC	TT	CT
Catalonia	0.53	0.47	0.39	0.32	0.29
Andalucia	0.43	0.57	0.22	0.36	0.42
Zamora	0.64	0.36	0.46	0.18	0.36

Table 8 - Allelic frequency and genotypic frequency of the three studied populations regarding the SNP rs4988235

	Allelic frequency		Genotypic frequency		
	Allele C	Allele T	CC	TT	CT
Catalonia	0.48	0.52	0.27	0.30	0.43
Andalucia	0.60	0.40	0.39	0.19	0.42
Zamora	0.67	0.33	0.39	0.06	0.55

The P-values obtain for the Hardy-Weinberg equilibrium test are shown in table 7. After applying a Bonferroni correction for multiples tests all the samples are in Hardy-Weinberg equilibrium.

Table 9 - Hardy-Weinberg equilibrium probability values for the three SNP selected and for the studied populations.
(*not significant after Bonferrroni correction)

	Catalonia	Andalucia	Zamora
rs988436	1.00000	0.32248	0.55785
rs1368136	0.02909*	0.17604	0.26054
rs4988235	0.72024	1.00000	0.26818

Relatively to the values of gene diversity (table 8), for SNP rs 988436 the values range from 0.023 to 0.338, on which the studied populations values are average. Concerning SNP rs1368136 diversity values range from 0.405 to 0.508, on which the studied populations show one of the higher values. For SNP rs4988235 the values range from 0.000 to 0.508, on which the studied populations also show one of the higher values, being Catalonia the population with the highest value of diversity (0.508).

Pairwise F_{ST} values and differentiation test results reveal that the three studied populations didn’t show significantly different frequencies amongst them neither with the Iberian population in Spain (data from previous studies) from the set of comparing populations. With the exception of rs4988235 in which Zamora presents significant differences in relation to Catalonia and to the Iberian population in Spain.

Regarding the pairwise F_{ST} values obtained for the SNP rs988436, Catalonia showed a significant genetic distance with the populations from Toscani in Italy, Finnish in Finland, British in England and Scotland, and Yoruba in Ibadan. Andalucia and Zamora showed a significant genetic distance with the populations from Mexican Ancestry from Los Angeles and with the three African populations. For the rs1368136, the three studied populations showed a significant genetic distance with the Finnish in Finland. Andalucia also showed a significant genetic distance with the three populations from East Asia and Puerto Ricans Puerto Rico. Zamora also showed a significant genetic distance with British in England and Scotland, Luhya in Webuye, Americans of African Ancestry in SW USA, Mexican Ancestry from Los Angeles and Colombians from Medellin. For the rs4988235, the three studied populations showed a significant genetic distance with East Asia population, Utah Residents (CEPH with Northern and Western European ancestry), Toscani in Italy, British in England and Scotland. Catalonia showed a significant genetic distance with Zamora, Ad Mixed American and African populations. Andalucia showed with Ad Mixed American population, Finnish in Finland, Luhya in Webuye and Americans of African Ancestry in SW USA. Zamora showed with Finnish in Finland, Luhya in Webuye, Americans of African Ancestry in SW USA, Mexican Ancestry from Los Angeles.

On what concerns differentiation tests, for rs988436, the three studied populations showed significantly different frequencies with Yoruba in Ibadan; Andalucia and Zamora showed with Luhya in Webuye and Americans of African Ancestry in SW USA, and Andalucia also showed with Mexican Ancestry from Los Angeles. For rs1368136, the three populations showed significantly different frequencies with Finnish in Finland and Mexican Ancestry from Los Angeles. Catalonia showed differences with Luhya in Webuye, Americans of African Ancestry in SW USA, and Puerto Ricans from Puerto Rico. Andalucia showed with the three populations from East Asia, and Puerto Ricans from Puerto Rico. Zamora showed with, Luhya in Webuye, Americans of African Ancestry in SW USA and. For rs4988235, the three populations showed showed a significantly different frequencies with the populations from East Asia, Utah Residents (CEPH with Northern and Western European ancestry), Toscani in Italy, British in England and Scotland, Luhya in Webuye, Americans of African Ancestry in SW USAs, and Ad Mixed American population. Zamora also showed significantly different frequencies with Finnish in Finland.

PCA shows that the three studied populations appear in the same quadrant relatively close from each other and separate from all the comparison populations (fig. 8), a result that is consistent with the result of population differentiation that point to the absence of differentiation within Iberia when these 3 SNPs are analysed.

Table 10 - Gene diversity values obtained for the three studied populations and for the populations used for comparison

Population Description	**rs 988436**	**rs 1368136**	**rs 4988235**
Catalonia	0,178	0,506	0,508
Andalucia	0,281	0,496	0,496
Zamora	0,282	0,47	0,451
Americans of African Ancestry in SW USA	0,094	0,495	0,253
Utah Residents (CEPH with Northern and Western European ancestry)	0,307	0,503	0,413
Han Chinese in Bejing, China	0,27	0,455	0
Southern Han Chinese	0,235	0,448	0
Colombians from Medellin, Colombia	0,245	0,503	0,448
Finnish in Finland	0,342	0,405	0,488
British in England and Scotland	0,338	0,501	0,391
Iberian population in Spain	0,304	0,508	0,508
Japanese in Tokyo, Japan	0,282	0,48	0
Luhya in Webuye, Kenya	0,07	0,483	0
Mexican Ancestry from Los Angeles, USA	0,141	0,481	0,378
Puerto Ricans from Puerto Rico	0,224	0,499	0,323
Toscani in Italy	0,338	0,502	0,184
Yoruba in Ibadan, Nigeria	0,023	0,499	0

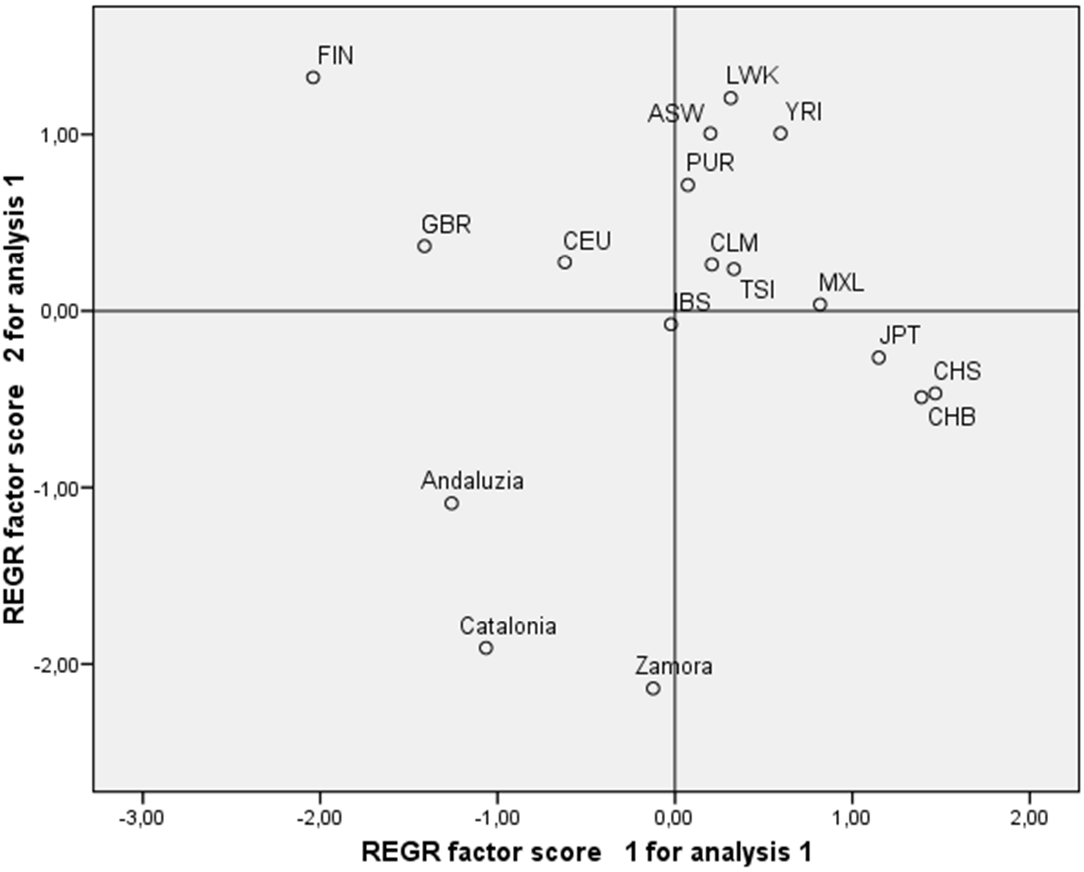

Figure 7 - PCA of the studied populations and the comparison populations. Component 1 explains 49,32% of the variation and Component II explains 32,66%

3. Discussion

The data obtained from the three SNP analysed for Catalonia, Andalucia and Zamora didn't show clear differences between them. This lack of differentiation could be due to the low amount of SNP tested or just because the Iberian Peninsula reflects, to some extent, a homogenous background [Arroyo-Pardo et al., 2007; Pino-Yanes et al., 2011].

One limitation to the interpretation of the results obtained is the analyses of only three SNP.. In the future, it would be important to analyse the remaining selected SNP on the three studied population to allow taking further conclusions.

A constraint to the study of these populations with SNPs is the low amount of studies done with SNPs in Europeans populations, as in this study the only comparing European population we had were Utah Residents with Northern and Western European ancestry, Toscani in Italy, Finnish in Finland, British in England and Scotland and Iberian population in Spain (at which the three studied population belong). In this sense, it would be important to increase the number of European populations studied, leading to the existence of data for each country of Europe, in a first stage, and then study different population within each country taking first into consideration the historical background of the country. Although the lack of information to comparison, the PCA showed an evident separation between the three studied populations and the comparison populations.

For further knowledge it would be important that a large number samples of the three Iberian populations were analyzed, as well it would be important that more Iberian populations and European populations were studied, enabling that some more information was add to the available.

V. Conclusions

Taking into consideration the aims initial proposed, it was possible to select a set of twenty-six SNPs accordingly with the Iberian Peninsula background. For these SNPs specific primers were design, taking into consideration that later on they would be analyzed by HRM.

From this initial set, five SNPs were selected and tested for PCR amplification and optimized the PCR conditions. The analyze by HRM allowed to establish the Tm for each SNP and the sequencing of samples allowed to identify the three possible genotype. These three possible genotype with the HRM pattern for each genotype were used as control samples in subsequent HRM.

Regarding the ancient samples, the positive amplifications obtained were important and further samples should be analyzed, so it would be possible to also obtain the patterns for each genotype. This technique is important for ancient samples, as is a close-tube technique as very little manipulation is required.

In the second chapter, three SNP were analyzed using samples from three Iberian populations (Catalonia, Andalucia and Zamora). Although no specific genetic pattern was found, it would be important to prosecute the study of Iberian populations.

This study was important to primarily understand the genetic background of the Iberian population. Nevertheless, further studies are important to best understand and characterize the genetics of the Iberian Peninsula. It would be also important to study the genetics of the European population, allowing a better genetics landscape of Europe.

Bibliography

1000genomes, <http://www.1000genomes.org> [Accessed: 1st September 2012]

Adams S.M., Bosch E., Balaresque P.L., Ballereau S.J., Lee A.C., Arroyo E., López-Parra A.M., Aler M., Grifo M.S.G, Brion M., Carracedo A., Lavinha J.,Martínez-Jarreta B., Quintana-Murci L., Picornell A., Ramon M., Skorecki K., Behar D.M., Calafell F., Jobling M.A. (2008) *The Genetic Legacy of Religious Diversity and Intolerance: Paternal Lineages of Christians, Jews, and Muslims in the Iberian Peninsula*. The American Journal of Human Genetics 83: 725–736

Alvarez L, Santos C, Ramos A, Pratdesaba R, Francalacci P, Aluja MP. (2010) *Mitochondrial DNA patterns in the Iberian Northern plateau: population dynamics and substructure of the Zamora province*. American Journal of Physical Anthropology 142(4):531-539

Ammerman A.J and Cavalli-Sforza L.L. (1984) *The Neolithic transition and the genetics of populations in Europe*. Princeton: Pricenton Univ. Press.

Amorim A. (2007) *Genetic markers: The interplay between concepts and technology in the Antropological scene*. In: Santos C. and Lima M. (editors) Recent Advances in Molecular Biology and Evolution: Aplications to Biological Anthropology. Kerala, India: Research Singpost

Arribas A. (1971) *Os Iberos*. Editorial Verbo

Arroyo-Pardo E, Baeza C, Fernández E, López-Parra AM. (2007). *Genetic history of the Iberian Peninsula.* In: Santos C, Lima M, editors. Recent Advances in Molecular Biology and Evolution: Aplications to Biological Anthropology. Kerala, India: Research Singpost. p 389-411.

Barbero A., Vigil M. (1974) *Sobre los orígenes sociales de la Reconquista*. Esplugues de Llobregat: Ariel

Bauchet M., McEvoy B., Pearson L. N., Quillen E. E., Sarkisian T., Hovhannesyan K., Deka R., Bradley D. G. and Shriver M. D. (2007) *Measuring European Population Stratification with Microarray Genotype Data*. American Journal of Human Genetics 80: 948 – 956

Bertranpetit J., Cavalli-Sforza L. (1991) *A genetic reconstruction of the history of the population of the Iberian Peninsula*. *Annals of Human Genetics* 55: 51-67

Bosch P. (2003) *Etnologia de la Península Ibèrica*. Cortadella J, editor. Pamplona: Urgoiti

Boyd R., Silk JB. (2001) *Cómo evolucionaron los humanos*. Barcelona: Editorial Ariel, S.A

Budowle B. and Daal A. (2008) *Forensically relevant SNP classes.* BioTechniques 44: 603-610

Burillo F. (2007) *Los celtíberos: Etnias y estados*. Barcelona: Critica

Butler J.M., Coble M.D., Vallone P.M. (2007) *STRs vs. SNPs: thoughts on the future of forensic DNA testing*. Forensic Science, Medicine and Pathology 3: 200-205

Camps G. (1996) *Los bereberes, ¿ mito o realidad?* En: Roque MA (editor). Las culturas del Magreb: Antropologia Historia y sociedad. Barcelona: Icaria Editorial

Cavalli-Sforza L.L. (1996) *Genes, Povos e Línguas*. Epistemologia e sociedade. Instituto Piaget

Cavalli-Sforza L.L., Feldman M. W. (2003) *The application of molecular genetic approaches to the study of human evolution*. Nature Genetics Supplement 33

Cavalli-Sforza L.L., Minch E. (1997) *Paleolithic and Neolithic lineages in the European mitocondrial gene pool*. American Journal of Human Genetics 61(1):247-254

Chakravarti A. (2001) *Single nucleotide polymorphisms: to a future of genetic medicine*; Nature 409

Chikhi L., Destro-Bisol G., Bertorelle G., Pascali V. and Barbujani G. (1998) *Clines of nuclear DNA markers suggest a largely Neolithic ancestry of the European gene pool*. Proceedings of the National Academy of Science 95: 9053-9058

Chikhi L., Nichols RA., Barbujani G., Beaumont MA (2002) *Y genetic data support the Neolithic demic diffusion model*. Proceedings of the National Academy of Science 99(17): 11008-11013

Corte-Real H. B. S. M., Macaulay V.A., Richards M. B., Haritip G., Issad M. S, Cambon-Thomsen A., Papiha S., Bertranpetit J., Sykes B.C. (1996). *Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis.* The American Journal of Human Genetics 60: 331-350

Costabile M., Quach A., Ferrante A. (2006) *Molecular Approaches in the Diagnosis of Primary Immunodeficiency Diseases*. Human Mutation 27(12): 1163-1173

Drineas P., Lewis J., Paschou P. (2010) *Inferring Geographic Coordinates of Origin for Europeans Using Small Panels of Ancestry Informative Markers.* PLoS ONE 5 (8): e11892. doi:10.1371/journal.pone.0011892

dsSNP database <http://www.ncbi.nlm.nih.gov/snp> (Accessed: September 2013)

Erali M., Voelkerding K.V., Wittwer C.T. (2008) *High resolution melting applications for clinical laboratory medicine*. Experimental and Molecular Pathology 85: 50–58

Excoffier, L. G. Laval, and S. Schneider (2005) Arlequin ver. 3.0: *An integrated software package for population genetics data analysis*. Evolutionary Bioinformatics Online 1:47-50

Futuyma D. (1998) *Evolutionary Biology*. Sunderland, Massachuset: Sinauer associates, Inc.

García A.B. (1982) *Toponomástica e historia: notas sobre la despoblación en la zona meridional del Duero*. In España Medieval 2:115-134

García L.A. (1989) *Historia de España visigoda*. Madrid: Cátedra

Gayán J., Galan J.J., González-Pérez A., Sáez M.E., Martínez-Larrad M.T., Zabena C., Rivero M.C.,Salinas A., Ramírez-Lorca R., Morón F.J., Royo J.L., Moreno-Rey C., Velasco J., Carrasco J.M., Molero E., Ochoa C., Ochoa M.D., Gutiérrez M., Reina M., Pascual R., Romo-Astorga A., Susillo-González J.L., Vázquez E., Real L.M., Ruiz A. and Serrano-Ríos M. (2010) *Genetic Structure of the Spanish Population.* BMC Genomics, 11:326

Gill J., Gill J.J., (2008) *Andalucía: A cultural history*. Oxford University Press

Gonçalves R., Freitas A., Branco M., Rosa A., Fernandes A.T., Zhivotovsky L.A., Underhill P.A., Kivisild T., Brehm A. (2005). *Y-chromosome Lineages from Portugal, Madeira and Açores Record Elements of Sephardim and Berber Ancestry*. Annals of Human Genetics 69: 443–454

Griffiths A., Miller H., Suzuki D., Lewontin R. and Gelbart R. (1993) *Population Genetics: An Introduction to Genetic Analysis*. New York: W.H. Freeman and Company. 737-769

Gu Y., Yun L., Zhang L., Yang F., Hou Y (2011) *The portencial forensic utility of two single nucleotide polymorphisms in predicting biogeographical ancestry*. Forensic Science International:Genetics Supplement Series 3(1): e105-e106

Gundry C.N., Vandersteen J.G., Reed G.H., Pryor R.J., Chen J., Wittwer C.T. (2003) *Amplicon Melting Analysis with Labeled Primers: A Closed-Tube Method for Differentiating Homozygotes and Heterozygotes*. Clinical Chemistry 49(3): 396–406

Guo, S. and Thompson, E. (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361-372

Hall, T.A. 1999. BioEdit*: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT*. Nucleic Acids Symposium Series 41:95-98.

INE (2012a) *España en cifras 2012*. Instituto Nacional de Estadistica, España

INE (2012b) *Península Ibérica en cifras 2011*. Instituto Nacional de Estadistica, España

Jobling M.A., Gill P. (2004) *Encoded evidence: DNA in forensic analysis*. Nature Reviews Genetics 5: 739-751

Jolliffe I.T. (2002) Principal component analysis, 2nd Ed. Springer

Kayser M., Knijff P. (2011) *Improving human forensics through advances in genetics, genomics and molecular biology*; Nature Review 12: 179-192

Kidd K.K., Pakstis A.J., Speed W.C., Grigorenko E.L., Kajuna S.L.B, Karoma N.J., Kungulilo S., Kim J.J., Lu R.B., Odunsi A., Okonofua F., Parnas J., Schulz L.O., Zhukova O.V., Kidd J.R. (2006) *Developing a panel for forensic identification of individuals*. Forensic Science International 164: 20-32

Kwok P.Y and Chen X (2003) *Detection of Single Nucleotide Polymorphisms*. Current Issues of Molecular Biology 5: 43-60

Kwok P.Y. (2001) *Methods for genotyping single nucleotide polymorphisms.* Annu. Rev. Genomics Hum. Genet. 2:235–58

Landsteiner K. (1900). *Zur Kenntnis der antifermentativen,lytisichen und agglutinierenden Wirkungen desBlutserums und der Lymphe*. Zentbl. Bakt. Orig. 27: 357–362

Landsteiner K. (1901). *Ueber Agglutinationserscheinungen normalen menschlichen Blutes*. Wien. Klin. Wochenschr. 14**:** 1132–1134 [Translation: On agglutination phenomena of normal human blood, in S. H. Boyer (Editor), 1963, *Papers on Human Genetics*, pp. 27–31. Prentice-Hall, Englewood Cliffs, NJ.]

Lewis JH., Walters D., Didisheim P., Merchant WR. (1958) *Aplication of continous flow electrophoresis to the study of the blood coagulation proteins and the fibrinolytic enzyme system. Normal human materials*. The Journal of Clinical Investigation 37(9): 1323-1331

Lewontin RC: (1967) *An estímate of average heterozyosity in man*. American Journal of Human Genetics 19(5): 681-685

Lewontin R.C. (1972) *The apportionment of human diversity*. Journal of Evolutionary Biology 6: 381-398

Lewontin R.C. (1974) *The problem of genetic diversity*. Harvey lectures 70 series: Wiley

Liew M., Pryor R., Palais R., Meadows C., Erali M., Lyon E., Wittwer c. (2004) *Genotyping of Single-Nucleotide Polymorphisms by High-Resolution Melting of Small Amplicons*. Clinical Chemistry 50(7): 1156–1164

Lorenzo A.V. (1995) *Población y poblamiento de Zamora en la Edad Media*. In: Alba JC (editor). Historia de Zamora: Diputacion de Zamora, Instituto de Estudios Zamoranos Florian de Ocampo

Mata D.R. (2002). *The Beginnings of the Phoenician Presence in South Western Adndaluzia*. In: Bierling M.R., Giton S. editors. The Phoenicians in Spain. Eisenbrauns USA

Menendez R. (1960) *Repoblación y tradición en la Cuenca del Duero*. Madrid: Enciclopedia Lingüística Hispánica

Mira M. (2000) *Cartago contra Roma. Las Guerras Púnicas*. Madrid: Alderabán Ediciones, S.L.

Moorjani P., Patterson N., Hirschhorn J.N., Keinan A., Hao L., Atzmon G., Burns E., Ostrer H., Price A.L., Reich D., (2011) *The History of African Gene Flow into Southern Europeans, Levantines, and Jews*. PLoS Genetics 7(4): e1001373. doi:10.1371/journal.pgen.1001373

Moreno R.S. (2001) *Historia breve de Andalucia;* Madrid, Silex

Nassir R., Kosoy R., Tian C., White P., Butler L. M., silva G., Kittles R., Alarcon-Riquelme M., Gregersen P. K., Belmont J. W., De La Vega F. M. and Seldin M. F. (2009) *An ancestry informative marker set for determining continental origin: validation and extension using human genome diversity panels*. BMC Genetics 10:39

National Center for Biotechnology Information (NCBI) <http://www.ncbi.nlm.nih.gov/snp> [Accessed: September 2013]

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

Nogueiro I., Manco L., Gomes V., Amorim A., Gusmão L. (2010) P*hylogeographic analysis of paternal lineages in NE Portuguese Jewish communities*. American Journal of Physical Antropology 141(3): 373-381

Novembre J., Johnson T., Bryc K., Kutalic Z., Boyko A.R., Auton A., Indap A., King K.S., Bergmann S., Nelson M.R., Stephens M and Bustamante C.D. (2008) *Genes mirror geography within Europe*. Nature 456(7218):98-101

Owen R. (2000) *Karl Landsteiner and the first human marker locus*. Genetics 155(3): 995-998

Pauling L., Itano HA., Singer SJ., Wells C. (1949) *Sickle cell anemina, a molecular disease*. Science 109(2835): 543-548

Phillips C., Prieto L., Fondevila M., Salas A., Gómez-Tato A., Álvarez-Dios J., Alonso A., Blanco-Verea A., Brión M., Montesino M., Carracedo A. , Lareu M.V. (2009) *Ancestry Analysis in the 11-M Madrid Bomb Attack Investigation*. PLoS ONE 4(8): e6583. doi:10.1371/journal.pone.0006583

Pino-Yanes M, Corrales A, Basaldáa S, Hernández A, Guerra L, et al. (2011) *North African Influences and Potential Bias in Case-Control Association Studies in the Spanish Population*. PLoS ONE 6(3): e18389. doi:10.1371/journal.pone.0018389

QIAGEN®(2009) Type*-it® HRM™ PCR Handbook - For detection of gene mutations and SNPs by high-resolution melting (HRM) analysis*. QIAGEN

Quesada M.A.L. (1989) *Los Reys Católicos: La Corona y la Unidad de España*. Asociacion Francisco Lopez de Gomara, Valencia

Quintana-Murci L., Semino O., Bandelt HJ., Passarino G., McElreavey K., and Santachiara-Benerecetti AS. (1999) *Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa*. Nature Genetics 23(4): 437-441

Quintáns B., Álvarez-Iglesias V., Salas A., Phillips C., Lareu M. V., Carracedo A., (2003) *Typing of mitocond*rial *DNA coding region SNPs of forensic and anthropologial interest using SNaPshot minisequencing*. Forensic Science International 140: 251-257

Raskin L., Pinchev M., Arad C., Lejbkowicz F., Tamir A., Rennert H.S., Rennert G. and Gruber S.B. (2008) *FGFR2 is a Breast Cancer Susceptibility Gene in Jewish and Arab Israeli Populations.* Cancer Epidemiology Biomarkers & Prevention 17: 1060-1065

Raven P. H., Johnson G. B., Mason K. A., Loss J.B., and Singer S.R. (2011) *Biology 9th Ed*. McGraw-Hill

Raymond M., Rousset F. (1995) *An exact tes for population differentiation*. Evolution 49:1280-1283

Redondo J.C., Fernández de Diego E. (1995) *Formación y evolución histórica de la provincia*. En: Alba JC (editor). Historia de Zamora: Diputacion de Zamora, Instituto de Estudios Zamoranos Florian de Ocampo

Reed G.H., Kent J.O., Wittwer C.T. (2007) *High-Resolution DNA melting analysis for simple and efficient molecular diagnostics*. Pharmacogenomics 8(6): 597-608

Reilly B. (1992) *Cristãos e Muçulmanos: A Luta pela Península Ibérica*. Editorial Teorema, Lisboa

Reynolds, J., Weir, B.S., and Cockerham, C.C. 1983 Estimation for the coancestry coefficient: basis for a short-term genetic distance. Genetics 105:767-779.

Richards M., Côrte-Real H., Forster P., Macaulay V., Wilkinson-Herbots H., Demaine A., Papiha S., Hedges R., Bandelt HJ. and Sykes B. (1996) *Paleolithic and Neolithic Lineages in the European Mitochondrial Gene Pool*. The American Journal of Human Genetics 59:185-203

Rosenberg N. A., Li L. M., Ward R., Pritchard J. K. (2003) *Informativeness of Genetic Markers for Inference of Ancestry*. The American Journal of Human Genetics 73: 1402-1422

Royal C. D., Novembre J., Fullerton S. M., Goldstein D. B., Long J. C., Bamshad M. J. and Clark A. G. (2010) *Inferring Genetic Ancestry: Opportunities, Challenges and Implications*; The American Journal of Human Genetics 86: 661-673

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

Scherer S. (2008) *What is the Frequency of Single Nucleotide Polymorphisms in the Genome? In: A Short Guide to the Human Genome*, ch.9

Seldin M.F., Shigeta R., Villoslada P., Selmi C., Tuomilehto J., Silva G., Belmont J.W., Klareskog L., Gregersen P.K. (2006) *European Population Substructure: Clustering of Northern and Southern Populations*. PLoS Genet 2(9): e143. DOI: 10.1371/journal.pgen.0020143

Semino O., Passarino G., Brega A., Fellous M., and Santachiara-Benerecetti AS. (1996) *A view of the neolithic demic diffusion in Europe through two Y-chormossome-specific markers*. The American Journal of Human Genetics 59(4):964-968

Simoni L., Calafell F., Pettener D., Bertranpetit J., and Barbujani G. (2000) *Geographic patterns of mtDNA diversity in Europe*. American Journal of Human Genetics 66(1):262-278

Slatkin, M. 1995 A measure of population subdivision based on microsatellite allele frequencies. Genetics 139: 457-462

Sobrequés i Callicó J. (2010) *Historia de Cataluña,* 2nd Ed.,Editorial Base, Barcelona

Sobrino B., Brión M., Carracedo A. (2005) *SNP in forensic genetics: a review on SNP typing methodologies*. Forensic Science International 154: 181-194

Stone N. (1990) *Atlas of World History.* Barraclough G, editor. London: Times Books Limited

Stoneking M. (2001) *Single nucleotide polymorphisms: From the evolutionary past...* ; Nature 409: 821-822

Susanne C., Rebato E. (2007) *Molecular anthropology: Advances and problems*. In: Santos C and Lima M (editors) Recent Advances in Molecular Biology and Evolution: Aplications to

Syvänen A.C. (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nature Reviews Genetics 2: 930-942

Takano E.A., Mitchell G., Fox S.B., Dobrovic A. (2008). *Rapid detection of carriers with BRCA1 and BRCA2 mutations using high resolution melting analysis*. BMC Cancer 8:59

Tayler S., Scott R., Krutz R., Fisher C., Patel V., Bizouarn F. (2010) *A pratical guide to High Resolution Melt analysis genotyping.* Bulletin 6004, Bio-Rad Laboratories Inc.

Tian C., Plenge R.M., Ransom M., Lee A., Villoslada P., Selmi C., Klareskog L., Pulver A.E., Qi L., Gregersen P.K., Seldin M.F. (2008). *Analysis and Application of European Genetic Substructure Using 300 K SNP Information*. PLoS Genet 4(1): e4. doi:10.1371/journal.pgen.0040004

Tian C., Kosoy R., Nassir R., Lee A., Villoslada P., Klareskog L., Hammarstrom L., Garchon H.J., Pulver A. E., Ransom M., Gregersen P.K. and Seldin M.F. (2009) *European Population Genetic Structure: Further Definition of Ancestry Informative Markers for Distinguishing among Diverse European Ethnic Groups*. Journal of Molecular Medicine 15(11-12):371-378

Tovar A. and Blazquez J.M. (1982) *Historia de la Hispania Romana: La Península Ibérica desde 218 a.C hasta el siglo V*. Madrid: Alianza Editorial

Twyman R.M. (2005) Single Nucleotide Polymorphism (SNP) Genotyping Techniques—An Overview. Encyclopedia of Diagnostic Genomics and Proteomics

Twyman R.M. and Primrose S.B. (2003) Techniques patents for SNP genotyping. Pharmacogenomics 4:1, 67-79

Vogel F., Motulsky A.G. (1997) *Human genetics: problems and approaches*. 3rd Ed. Springe

Vossen R.H.A.M., Aten E., Roos A., Den Dunnem J.T. (2009) *High-Resolution Melting Analysis (HRMA) – More Than Just Sequence Variant Screening*. Human Mutation 30(6): 860–866

Wattenberg F. (1959) La *Región Vaccea: celtiberismo y romanización en la cuenca media del Duero*. Madrid: CSIC

Whittle A. (1996) *Europe in the Neolithic*. Cambridge: Cambridge Univ Press

Wittwer C.T. (2009) High-*Resolution DNA Melting Analysis: Advancements and Limitations*. Human Mutation 30(6): 857-859

Zhao L. P., Zarbl H. (2000) *Single Nucleotide Polymorphisms and Future Applications. In: SNP and Microsatellite Genotyping: Markers for Genetic Analysis*, edited by A. Hajeer, J. Worthington and S. John

Zhou L., Wang L., Palais R., Pryor R., Wittwer C.T. (2005) *High-Resolution DNA Melting Analysis for Simultaneous Mutation Scanning and Genotyping in Solution*. Clinical Chemistry 51(10): 1770 –1777