Demographic insights of Human North African populations using genetic data

Laura Rodríguez Botigué

TESI DOCTORAL UPF / 2012

DIRECTOR DE LA TESI

Dr. David Comas

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

UNIVERSITAT POMPEU FABRA

Al Vito

Agraïments

Sou molts els que m'heu ajudat a arribar fins on sóc ara, que heu confiat en mi i m'heu donat empenta quan ho he necessitat. Per començar tu, Vito. Han estat uns anys plens d'imprevistos i de grans canvis, però crec que en tot moment això que tenim ha estat la constant que ens ha donat la força i la decisió per seguir endavant, i això és cosa de dos. Així que gràcies. També als meus pares, que teniu una confiança cega en mi que a vegades fa por. Sembla que per vosaltres no hi hagi res impossible de fer, i aquest suport incondicional ha estat molt important per mi.

Ixa, compartir amb tu el despatx ha estat una de les millors coses d'aquesta tesis. Has estat col·lega, amiga, consellera, confident... i una llarga llista de coses més, totes bones. M'has ajudat a resoldre molts (molts) dubtes, i la teva tenacitat i perseverància m'han fet millor persona. Bego, Sando, vosotras habéis dado la chispa al ambiente del despacho. Y tu Karlita, también me has acogido y cuidado los meses que he estado en Stanford. Decirte que has sido buena anfitriona, junto con Andrés, sería decir poco. Sobre todo este último año, lleno de emociones, será difícil de olvidar, así como *Inspiración* y el jilguero Moreno-Estrada por la mañana, tarde y noche. Gracias. Urkito, Txemita, m'encanteu. Són moltes les coses bones que he après de vosaltres, però crec que la millor és que sovint m'obligueu a veure el món des d'una altra perspectiva, m'obligueu a plantejar-me preguntes sobre coses que donava per descomptat, i això es molt i molt saludable. Merci. Oscarin, el noble de la colla i també apassionat de les bèsties. Crec que compartim el voler regir-nos per un codi d'honor personal que és una mica similar i que ens agrada fer volar la imaginació. M'acabaré llegint *El nombre del viento*, només l'he de demanar als altres amics de l'Hospi. Monix, la sensibilitat personificada. Sempre patint per si... ara per tu comença una nova etapa, igual que per mi, però espero que no m'oblidis, perquè tant el Vito com jo et portarem a tu al cor. Chungui Power, saps de que m'alegro més? De sentir la teva rialla i veure els teus ulls plorosos després de tant riure a l'hora del cafè. Són un bàlsam per a qualsevol que tingui un mal dia. Gràcies. Ludovi, tu i la Ixa sou les dones més fortes que he conegut a Bioevo. Per mi sou les dues un exemple a seguir, i m'alegro de tot el temps que hem passat juntes.

I en general a tots els Bioevos amb qui he pogut compartir un moment o altre. Al futbol, al volei, quan he fet de nòmada pels despatxos, als Journal Club i als Book Club. Sou únics en la vostra espècie així que gràcies a tots, perquè per molt que t'apassioni la feina que fas, si no hi ha un bon ambient els dilluns es fan costa amunt, i això és una cosa que en quatre anys no he experimentat.

Jordi, Clara, Dani, Bernat, Edu, Joan, Salo i Carol, vosaltres heu estat claus en aquesta tesi. M'heu fet riure fins que no m'han saltat les llàgrimes, m'heu acompanyat a passejar les bèsties quan necessitava amics i em faltava temps, i m'heu escoltat quan tenia problemes i les coses no anaven com volia. Em sento molt orgullosa d'haver-vos tingut a prop durant aquests quatre anys. Fins i tot quan era a 1,200 km de distància he sentit la vostra proximitat. Per tots els sopars, sortides, passejades, cafès, mudances i partides, aquesta tesi és també una mica vostra.

Els amics més antics que tinc, Núria, Melina, Xavi, Laura i ara Sergi, Bruna i Louis. Ens veiem poc, però quan ho fem deixem petja! Ens hem fet grans junts i hem experimentat i superat problemes similars a la vegada. Tenim tants records compartits acumulats que sembla que no puguem quedar-nos mai sense conversa. Aquest lligam no es trenca fàcilment, gairebé m'atreveixo a dir que és per sempre. Vosaltres sou part de l'essència del que sóc ara, perquè ens hem construit junts. Així que aquesta tesis també és mèrit vostre. Gràcies!

Biòlegs! Escampats pel món i vivint aventures. Vosaltres sou la meva part salvatge. Joana, Alba, gràcies per no rendir-vos i fer-me un truc quan feia temps que no donava senyals de vida. Sandra, Cons, Clara, Enric, Titus, Eli, Carles, Rastes, *tots*... Us tinc al cor i al cap, i vosaltres també m'heu ajudat a construir una part molt important de mi mateixa. Sabeu que sou especials, i que teniu una manera particular i genuina de veure el món. Quan us vaig conèixer em vau trasbalsar la vida i fer la carrera amb vosaltres ha estat molt més enriquidor del que mai m'hagués pogut imaginar. Gràcies per ser com sou.

Jeff, Brenna, Simon and Martin. With you I have experienced the population genetics field at a further level. Your passion and knowledge has been a constant challenge and an example to follow

for me. Thanks for sharing your time and moments, where I have learned so many things and I have realized I still need to learn so many other things. I look forward to seing you again, and having the chance to share new experiences together.

Victor, Deli, gràcies per ser-hi. Tonino, Pina, Anna, Pasquale, Michela, Leo, Gian-Marco, Pier-Nicolo, Elizabeth, Paolo, grazie pure a voi per farmi la vita un po piu facile in Italia.

Finalment, David gràcies per ser darrere d'aquest projecte. Per donar-me la llibertat quan la necessitava i el toc d'atenció quan me n'anava per les branques. I gràcies també a tu, Tomàs. Tu m'has donat molts consells pràctics i els dos m'heu guiat en la meva etapa successiva.

No vull acabar sense agrair a tots els que han fet possible que el refectori del convent classense de Ravenna fos una sala d'estudi. Aquí he tingut l'honor d'escriure les últimes línies de la tesis.

Abstract

The history of North Africa is extremely complex, and it has been difficult to assess from genetic and archeological data whether early populations were replaced by later migrations or if there has been continuous settlement of the region. To resolve the history of human origin and migrations in North Africa, I have used two main forms of genetic data, the maternally inherited mtDNA and 730,000 genome-wide SNPs from a genotype array in a sample set representative of the region. I have discovered that North Africa is a mosaic of an autochthonous component dating back to the Paleolithic and at least four other ancestries, two recent ancestries from sub-Saharan Africa and the others from Europe and the Near East. We have also discovered extensive North African gene flow to the Iberian Peninsula, and minor proportions in the rest of the Europe.

Resum

La història del Nord d'Àfrica és extremadament complexa, i fins ara ha estat molt difícil determinar a partir de la genètica o l'arqueologia si els primers pobladors van ser reempleçats per migracions posteriors, o si el poblament de la regió ha estat continuat al llarg del temps. Per tal d'investigar els orígens i les migracions de l'home al Nord d'Àfrica he fet servir dos marcadors genètics en un grup de poblacions representatives de la regio, el marcador heretat per via materna, el DNA mitocondrial (mtDNA), i 730,000 SNPs de tot el genoma genotipats amb un xip. He descobert que el Nord d'Àfrica és un mosaic format per un component autòcton amb origens en el Paleolític i un mínim de quatre components més, dos d'ells recents d'origen sub-Saharià i els altres Europeu i d'Orient Proper. També hem descobert un flux genic recent d'origen Nord Africà molt elevat a la Península Ibèrica, i en menor quantitat a Europa.

Preface

North African populations are distinct from sub-Saharan Africans based on cultural, linguistic and phenotypic attributes. The history of North Africa is extremely complex, with evidence of anatomically modern humans already settling in the region 160,000 ya. It is difficult to assess from the archaeological record whether early populations were replaced by later migrations or if there has been continuous settlement of the region. The time and the extent of genetic divergence between populations North and South of the Sahara remain poorly understood, as well as African connections with Near Eastern populations.

To resolve the history of human origin and migrations in North Africa, I have used two main forms of genetic data. First, in collaboration with colleagues, I have analyzed the genetic landscape of North Africa using maternally inherited mitochondrial DNA data (mtDNA) from previously published populations and from general Libyan population. Second, I have analyzed new genome-wide SNP genotyping array data (730,000 autosomal markers) from seven North African populations spanning from Morocco to Egypt and four Spanish populations.

Analysis of high-throughput genotyping in North African populations has answered a long-standing question about the region: ancestors of contemporary North Africans that inhabited in the region date from Paleolithic times, and the Neolithic wave of expansion did not entail a complete nor relevant genetic replacement in the entire region. Also, it has been possible to assess proportions of ancestry from neighboring regions, including up to four different ancestries. Of special interest is the Near Eastern influence detected mainly in Egypt and Libya, which is in agreement with the results shown using mitochondrial DNA. The study of this region using dense genotype data has also allowed investigating the substantial amount of North African admixture we find in the Iberian Peninsula and in minor proportions in the rest of Europe. Also, it has been possible to determine the time of migration and the nature of Sub-Saharan gene flow using haplotype information extracted from a local ancestry assignment analysis.

Index

1. INTRODUCTION

1.1. North African demography: Paleontological, archeological, historical and linguistic evidences.

North Africa, the northernmost region of the continent, is separated from Europe by the Mediterranean Sea and from the remainder of the African continent by the Sahara desert. These natural barriers have acted as demographic barriers as well, isolating North African people from populations from surrounding regions. However, it has also been reported that in some occasions climate conditions changed into a much wetter environment, allowing some contact between North Africa and sub-Saharan Africa.

During the last decade North Africa has drawn a special interest in the anthropological field thanks to the publication of some important archaeological findings (Balter, 2011; Smith et al, 2007). In summary, peopling in North Africa by modern humans is more ancient than it was thought, which has led to novel hypotheses regarding the origin of modern humans and the Out of Africa theory. In addition, new dating techniques have established new time ranges of the different techno-complexes that characterize the region throughout history (Balter, 2006). As a result, North African ancient history is now being revisited.

In this section I aim to briefly describe the actual state of the art on these fields.

a) North Africa during the Pleistocene (126,000 - 12,000 ya[1])

Traditionally, researchers distinguish *Homo sapiens* from his predecessors basically by the cranial and dental features observed (Smith et al, 2007). *Homo sapiens* is unique among hominines in a number of cranial and skeletal characteristics, being the most outstanding ones the proportions of the skull in which the brain is housed and the slenderness of its skeleton when compared to other more robust hominid species. Nonetheless, it is noteworthy to point out that an undoubtedly gradation in morphology exists between *H.*

---

[1] years ago

sapiens and his predecessors, being sometimes difficult to set a boundary between taxa. Furthermore, it is also a taxon extremely diverse in terms of morphological traits: variation in the later *Homo sapiens* fossil record is too great to be accommodated in a single taxon. Indeed, for years, paleoanthropologists have designated as "archaic" *Homo sapiens* a very heterogeneous assortment of relatively large brain hominids that are fairly recent but are clearly not of modern human morphology (Tattersall, 2009). Some researchers have assigned some of these remains to separate species like *Homo helmei* or a *Homo sapiens* subspecies, *Homo idaltu (White et al, 2003)*. Here I will refer to as "early *H. sapiens"* or "early humans" all hominid fossils that possess all or nearly all the diagnostic skull features of *H.sapiens* except for the brow and/or chin characteristics (Tattersall, 2009). In the same way, I will refer to as "modern *H. sapiens"* or "modern humans" those hominine fossils that are indistinguishable from the morphology found in at least one regional population of modern humans (Wood, 2010).

Earliest evidences of the existence of early humans are found in Ethiopia, East Africa. The hominine remains, Omo I, date from 195,000 years ago (ya), based on the crystals from the stratigraphic level in which it was found (McDougall et al, 2005). A number of other early human remains are found in East and South Africa with an inferred time range between 160,000 and 90,000 ya (**Figure 1**).

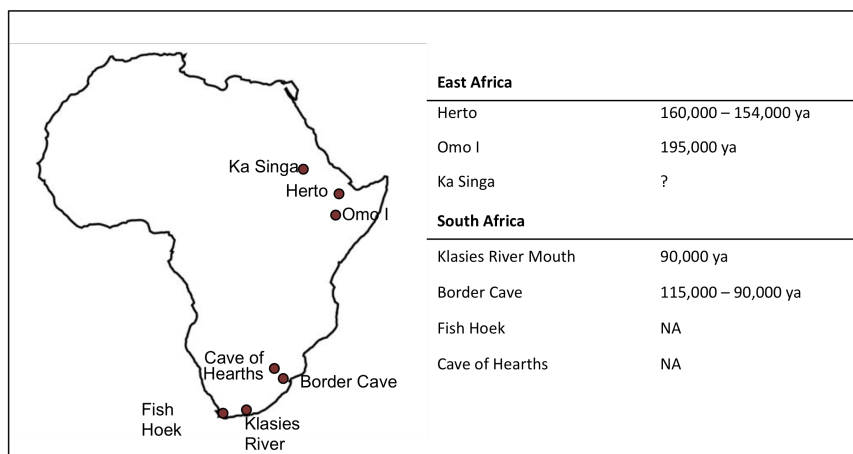| East Africa	
Herto	160,000 – 154,000 ya
Omo I	195,000 ya
Ka Singa	?
South Africa	
Klasies River Mouth	90,000 ya
Border Cave	115,000 – 90,000 ya
Fish Hoek	NA
Cave of Hearths	NA

Figure 1 The oldest hominin sites attributed to the first early modern *Homo sapiens.* The locations are indicated in the map and in the right their estimated dates. Adapted from (O'Neil, 2012).

Outside the African continent there are no fossil evidences that can be convincingly assigned to *H. sapiens* before 92,000 ya (Tattersall, 2009). These proofs are in agreement with the hypothesis established around 25 years ago by Bräuer stating that the abundance and the old age of early human remains in Africa suggested that modern humans originated around 200,000 ya somewhere in Sub-Saharan Africa and around 100,000 ya would have left the continent and colonized the rest of the world in a process now known as the Out of Africa (Bräuer, 1984; Stringer & Andrews, 1988).

Until recent, North Africa had been left apart in the expeditions focused on investigating early humans during the Upper Pleistocene for several reasons. First of all, even if more than 100 archaeological sites have been excavated in the region, most of them lack of hominid remains from the Upper Pleistocene. In such cases, researchers extract information from the material found on those sites, usually manufactured tools and other artifacts related to the manufacturing process. The Upper Pleistocene in North Africa is associated with either Mousterian or Aterian assemblages (**Figure 2**). Hand axes, recloirs and points are the characteristic tools of a Mousterian industry, which is flint flake-based and displays some use of Levallois technology. The Aterian industry includes novel tools not seen in Mousterian assemblages. Examples of these are stemmed artifacts, bifacial and unifacial points, and worked bones (Jacobs et al, 2011).
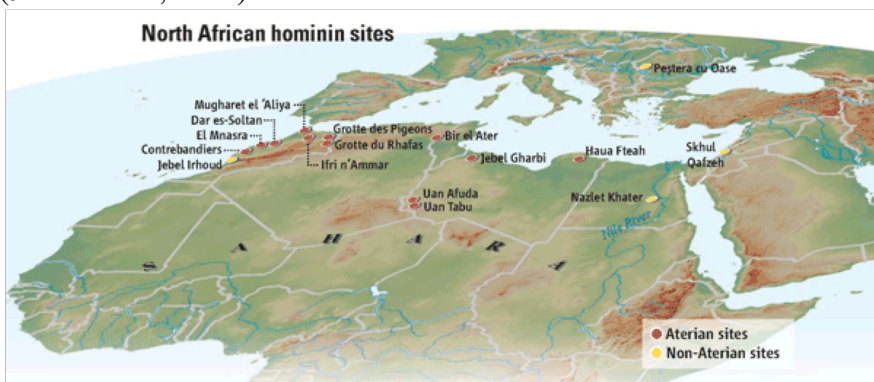

Figure 2 Hominin sites in North Africa in the Upper Pleistocene (Balter, 2011).

Another fundamental innovation associated with Aterian is the presence of elements more related to a symbolic behavior like shell

beads (Bouzouggar et al, 2007; d'Errico et al, 2009), pigment use (Nespoulet et al, 2008) and structured fireplaces (Nespoulet et al, 2008). Symbolic behavior is especially relevant because it is used as one of the markers of modernity associated to *Homo sapiens*, and would thus establish the difference between early and modern humans. Traditionally, the archeological material was dated using ^{14}C. However, it has been shown that this technique does not allow a trustable dating beyond 50,000 ya (Balter, 2006). For years it was considered that Aterian first appeared around 40,000 ya, because these were the most trustable dates. Older estimates were discarded because they were at the limit of the reliable radiocarbon technique. In summary, chronological uncertainty, the scarcity of stratified Mousterian and Aterian sites, and the poverty of the hominid fossil record, has prevented researchers to build a historical framework of North Africa during that period.

However, changes began five years ago, with novel discoveries on the fossil record and the appearance of new dating techniques. In 2001, hominines from Jebel Irhoud (Morocco), which in a first moment seemed to be Neandertals, turned out to share a number of sinapomorphies with modern humans, and were thus classified as "archaic" *H. sapiens* (Hublin, 2001). Some years later a study based on the dental development of a juvenile remain, Irhoud 3, showed that the pattern of dental growth was similar to that exclusive of modern humans (Smith et al, 2007) establishing the presence of early humans in North Africa around 130,000 - 190,000 ya. A debate regarding the classification of Irhoud 3 emerged, with some authors stating that Irhoud 3 is an early human specimen based on its morphology, and others claiming that the teeth development proofs that it can be considered a modern human. On top of this, three years later, in 2010, the remains of another child hominine were found in Grotte des Contrebandiers, Morocco. The first analysis dated the skull and partial skeleton in 108,000 ya (Balter, 2011), though a scientific study on this specimen has not yet been published.

The antiquity of the human remains recently found in North Africa has triggered a long-lasting debate regarding the causes by which modern humans existed in North Africa at that time and their possible evolution. On the light of these results some researchers

claim that modern humans might have first originated in North Africa and later on might have spread to the remainder of the African continent. However, most researchers are cautious and state that it is possible that those hominids belonged to a group of modern humans that were not involved in the Out of Africa migrations and did neither leave any genetic heritage in extant human populations (Balter, 2011). In summary, findings in North Africa have opened a debate based on two hypotheses, the first one claiming unique East African or North African origin of modern humans, and the second one claiming a multiregional origin with an *in situ* evolution in North Africa. Nonetheless, it cannot be denied that these findings have shown that the origins of modern humans are much more complex than it was once believed.

In a parallel way, during the last five years, a number of new techniques to date stratigraphic layers have also been developed. Two examples are the Optically Stimulated Termoluminiscence (OSL) dating of sediments (Jacobs & Roberts, 2007) and Termoluminiscence (TL) dating of burnt objects (Pagonis et al, 2011). These methods are characterized by going far beyond the dating based on ^{14}C range and therefore have higher accuracy in dating older sites. Surprisingly, the application of these techniques into the North African Aterian assemblages, gave results much older than 50,000 ya. As a result, Aterian is now considered to have appeared at least 80,000 ya, probably going as far as 120,000 ya. In a similar way, Aterian sites with evidences of symbolic behavior have also been re-dated and turned out to be much older than previously thought (Bouzouggar et al, 2007; d'Errico et al, 2009). Surprisingly, they turned out to be older than the oldest evidences of symbolic behavior found at the moment (Henshilwood et al, 2004; Holden, 2004).

New chronologies now available stress not only the relevance of the Aterian industry in North Africa, but also the antiquity of this technology and the symbolic behavior associated to it. The longevity of this culture over time and its wide distribution in the region suggest that people living during the Aterian belonged to a solid and strong population network. It has been suggested that from 80,000 to 60,000 ya groups associated with Aterian techno-complexes underwent a major expansion in combination with

behavioral innovations (Van Peer, 1998; Van Peer & Vermeersch, 2000). Some authors associate this hypothesis to the evolution of modern human behavior and dispersal in the region (Jacobs et al, 2012), and have hypothesized a pan-African theory in which fully modern human behavior would have taken place first in North Africa and afterwards would have spread into the rest of the continent.

In summary, investigations carried out during the last decade in North Africa regarding the Upper Pleistocene have shown that it is a region very abundant in archeological content and perhaps it played a key role in the origin and evolution of modern humans. First, human remains dating between 130,000 and 190,000 ya have been found in the western part of the region. Also, it has been shown that Aterian culture was more ancient than previously thought spanning probably from 120,000 to 20,000 ya and is spread across the region. Finally, objects associated to a symbolic behavior are found, and they are not only the oldest objects of this type found at the date, but also quite frequent all over the region. All these characteristics are indices of a well-structured society and not a cul-de-sac, which would in turn support the hypothesis that North Africans living in the Aterian would have had more chances to last for successive periods or expand beyond the North African region.

Finally, new insights on climate conditions in Africa would show that human migration between North Africa and Sub-Saharan Africa was possible. Around 120,000 ya the Nile river was not the only river to go through the Sahara desert. A number of river corridors and lakes located in Libya also used to cross the desert at that time (Drake et al, 2011; Osborne et al, 2008). This has been proposed as a possible route Out of Africa, but another important consequence is that contact between North African and Sub-Saharan African societies would be possible and therefore cultural influences might have taken place as well. Also, three episodes of human occupation in Morocco can be correlated with contemporaneous phases of wetter climate and expanded grassland habitat, whereas the gaps in occupation may represent drier periods and a contraction of grassland in the Sahara (Jacobs et al, 2012). As a result, it has been suggested that population mobility within North Africa was linked to climatic conditions (Garcea & Giraudi, 2006).

Particularly, authors claim that Aterian-related groups spread throughout North Africa during the Last Interglacial, between 125,000 and 74,000 ya, developing some local or regional diversification. Furthermore, variability in the Aterian assemblages would be explained by changes in the environmental conditions of the Sahara during the interglacial, in which a drier climate would favor isolation. On the contrary, during wetter periods contacts between groups would have been established, which would also be compatible with the pan-African hypothesis of modern human evolution. A question that remains unanswered, however, is the reason why the Aterian tool production outlived this climate changes without any clear sign of cultural evolution, a pattern shown in successive techno-complexes.

The disappearance of the Aterian culture, around 20,000 ya, remains unexplained. Several hypotheses exist, such as the model of an *in situ* transition towards the cultures of the Holocene (Stoetzel et al, 2011), but large techno-typological differences between industries do not support this model. In North Africa, the transition between Aterian and Iberomaurusian (the following industry) was relatively sudden, and the disappearance of the Aterian culture and associated human populations remains unexplained (Debénath, 2000). During the Upper Pleistocene the industry found in Egypt is different from that found in the rest of North Africa, and it is called the Nubian Complex. It spans from 128,000 to 74,000 ya in desert areas and from 240,000 to 50,000 nearby the Nile Valley (Olszewski et al, 2010). This differences in time ranges coincide with a period characterized by wetter climate, and it is suggested that with the improvement of weather conditions populations expanded into previously arid areas and remained there well after that the typically arid climate came back (Mercier et al, 1999). Nubian technology is thought to have originated in Sub-Saharan Africa and to be used by modern humans, even if no human remains have been found associated to any Nubian Complex assemblages. However, this tool industry is highly relevant due to its distribution (**Figure 3**). It is found in Egypt, northern Sudan, some eastern Sahara oases and recently in Saudi Arabia as well, dating 106,000 ya (Rose et al, 2011). This has triggered once again the debate of when did the Out of Africa migration take place and which route did it follow, given that the Nile River is an obvious corridor to

many scientists. It is also worth to notice that local and chronological differences have been detected in Nubian assemblages, what raises the question of which of these forms of industry was linked to OOA modern humans (Olszewski et al, 2010; Rots et al, 2011). Unfortunately, little is known on the successive time, from 40,000 to 15,000 ya. Different cultures are found, characterized by the production of stone and bone tools, a reduction of the tools dimensions and lately by the use of burines, microblades and microlithic industry, but there are still a lot of gaps and informative sites are scarce. As a consequence researcher cannot establish a historical framework of that time period in Egypt.
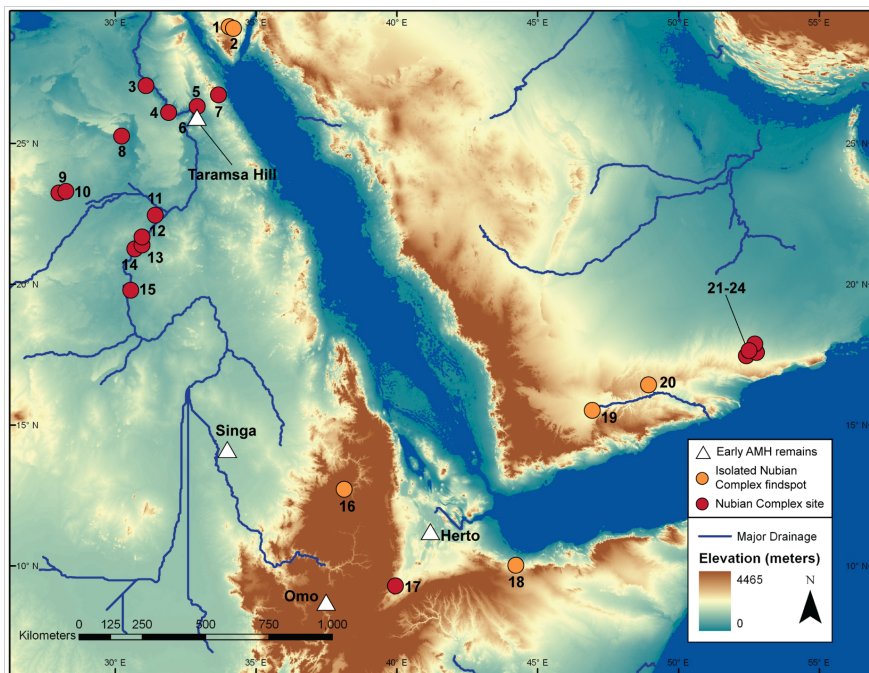
Figure 3 Map of Nubian Complex occurrences in Northeast Africa and Arabia (Rose et al, 2011)**.**

Until recently, it was thought that North Africa had little to contribute to major questions regarding the origins of modern humans, their dispersal and the origins of symbolic behavior. Such a view is now being rapidly transformed due to novel techniques that allow a more accurate study of the very rich archaeological collections available throughout this region. As a result, new ideas and advances on the origins of humanity have been proposed including North Africa in their framework. Also, new data available

from this region has given new insights in the understanding of other major cultural transitions in the late Pleistocene and early Holocene.

b) North Africa during the prehistoric Holocene (20,000 - 3,000 ya)

The Holocene is the geological period that follows the Pleistocene starting after the Last Glacial Maximum, around 20,000 ya and going on until present. In the beginning of the Holocene the great majority of the technological assemblages found in North Africa are known as the Iberomaurusian. The tool industry is characterized by microlithic backed, partially backed obtuse-ended, and other bladelets. It has been suggested that the use of microlithic bladelet tradition supposed a major break with preceding technologies (Close & Wendorf, 1990). Many archaeological sites where Iberomaurusian assemblages have been found contain human burials as well. The examination of the human remains showed that Iberomaurusians had a distinctively robust skeleton, though some authors specify that a high variability in their morphology exists. Another characteristic is the deliberate extraction of the incisors during the lifetime of the individual, a procedure known as tooth ablation.

The origins of Iberomaurusian people and technology are still a subject of debate. First evidences of the use of Iberomaurusian industry appear in North Africa after a long period of aridity during the Last Glacial Maximum, around 20,000 - 18,000 ya (Bouzouggar et al, 2008; Garcea & Giraudi, 2006; Hunt et al, 2010). The change in the climate conditions into a much arid environment is associated with the emergence of this new technique but it is uncertain whether it was the major cause, given that in the Aterian big climate changes happened with no detectable changes in the industry or the hunting strategies (Stoetzel et al, 2011). The diversity and extent of the Aterian, together with reported changes in settlement patterns (Garcea & Giraudi, 2006) before the Last Glacial Maximum make difficult the confirmation of population continuity or, alternatively, population replacement. Some authors have suggested an *in situ* evolution of the Iberomaurusian, based on similar settlement patterns detected in Libya from the Aterian through the rest of the

Pleistocene, into the Holocene and until the present day (Garcea & Giraudi, 2006). Resemblances between Iberomaurusians and European Cro-Magnon are compatible with a European migration hypothesis, and finally, a West Asian and a sub-Saharan African migration have also been hypothesized.

Around 13,000 ya evidence of accumulation of massive shell midden deposits are found in Iberomaurusian assemblages in Morocco and Libya (Bouzouggar et al, 2008; Taylor et al, 2011). The study of the shells from Taforalt in Morocco revealed that great amounts of terrestrial snails were collected, cooked, and consumed at that period (Taylor et al, 2011), compatible with a more sedentary way of living, which would represent a remarkable change in subsistence strategies. The examination of burials has showed that Iberomaurusian people were, like in the Aterian, a very heterogeneous society, more complex than it was thought. The method of interment appears to have varied during the Iberomaurusian lifespan and in different geographical places. Funerary activity is highly variable in these assemblies, such that it is not possible to define a characteristic tradition (Humphrey et al, 2012). Secondary depositions in the burials sometimes seem to be intentional and others just the consequences of ongoing activity in the site. Also, many different body positions are found across North Africa, and in some cases there are also evidences of the deliberated inclusion of funerary artifacts. It is noticeable the presence of ochre and cut marks on some bones. This can be explain by the performance of some rituals, which denote a certain profundity of thinking about life and death (Mariotti et al, 2009).

Like its origins, the end of the Iberomaurusian culture and the Iberomaurusian people is also controversial. In Western Maghreb and Libya it is replaced by the Neolithic period, a worldwide phenomenon that supposed a dramatic change on the economy, the culture and the lifestyle of human populations. In Tunisia and eastern Algeria a new industry is found, named the Capsian, which was ultimately substituted as well by the Neolithic culture. Capsian archaeological assemblages succeed the Iberomaurusian ones in eastern Maghreb and are dated between 10,400 and 6,000 ya. Capsian people are characterized by a slender skeleton, which would be the main reason to hypothesize an origin out of North

Africa and independent from Iberomaurusians (Camps, 1974; Hiernaux, 1975). However, recent studies based on a general evolutionary approach, the dental features and the funerary traditions have found similarities between both cultures, making plausible the hypothesis that there was continuity between them (Irish, 2000; Mariotti et al, 2009; Sheppard & Lubell, 1990).

Capsian sites are made up of large accumulations of ashes, burned stones, knapped flint and, especially, of land snail shells, which have given them the name of *escargotières*. They are often open air, sometimes under rock-shelters, but seldom in caves. Material culture from Capsian assemblages is very diversified, and a part from lithic and bone industries, all sorts of decorative objects are found, like worked ostrich eggshells, engraved stones, ochre staining of lithic artifacts, and human bones used for both ritual and utilitarian purposes. All these objects reveal an elaborate Capsian art, and reflect the possible presence of a ritual or symbolic system associated with complex social interactions. Typological studies have divided Capsian assemblages into two categories named Typical Capsian, characterized by large tools fabrication, a poor bone industry and a microlithic component, and Upper Capsian, with much smaller tools, abundance of bone industry and the development of geometric microliths. Studies (Jackes & Lubell, 2008; Rahmani, 2004) based on a dynamic modern analytical method that takes into account not only the tools, but also the complete *chaîne opératoire,* and also based on the use of improved dating techniques, have concluded that the Upper Capsian is the evolution of the Typical Capsian. A novel technique appears in North Africa during the Upper Capsian, which is the pressure-flaking technique, which is considered for some authors as a chronostratigraphical marker (Barton et al, 2008; Rahmani, 2004). The appearance of pressure-flaking coincides with a more arid climate change beginning 8,400 ya and lasting well beyond 8,200 ya. Indeed, it has been suggested that this technique would be an *in situ* invention as the result of an already existing stimulus (in opposition to a technique developed by chance) in the Typical Capsian (Rahmani, 2004). Moreover, the pressure-flaking technique would have allowed people to be more independent of the raw material, and thus be able to expand to places where it was less

abundant. Also, they found evidences of specialization and storage of the material, both signals of an evolving society.

Finally, it is noticeable that Capsian lasted in this region long after the Neolithic had spread in the rest of the North African region. Capsian latest sites are dated from around 6,000 ya, and Neolithic sites have been found in Morocco and Sudan 6,000 and 9,000 ya. The reasons of why a hunter-gatherer society lasted in the middle of a region where agriculture and complex societies had settled are still unknown.

In Egypt it seems that during the LGM and soon after that animals gathered nearby the Nile River, and vegetables endured as well in that area. This allowed populations to establish in the Nile Valley and adopt a more settled lifestyle given the proximity of the game and the plant species. Around 13,000 ya and until 9,000 ya the Qatan culture is found in southern Upper Egypt. Some Qatan tools show traces of sickle gloss, an effect of manipulating vegetables commonly associated to agriculture. Pollen from *Graminea* is found in the region and in some areas also wild barley. This has raised some debate about a possible Nilotic origin of agriculture, previous and independent from that of the Near East, though the sudden disappearance of this culture and successive evidences of hunter-gathering populations in the same region do not support that theory. It is also worthy to mention that the Qatan culture is also characterized by some funerary rituals similar to those found in Neolithic cultures.

The Neolithic period can easily be determined by the production of ceramics, though its most important aspect is the social and economic changes that populations from all over the world experienced. It meant the beginning of food and animal domestication, which eventually ended up with the introduction of agriculture and the end of the hunter-gatherer lifestyle in exchange of sedentism for most of the human populations worldwide. This ultimately caused cultural, and economic changes in populations and the beginning of socio-political complexity. Models detailing the Neolithic expansion can be divided in two categories according to whether they focus on its origins or whether they focus on its dispersal. The first ones are called *provenance models*, and aim to

explain the origins of the Neolithic culture in a given region. The second ones are grouped as *dispersal models*, with the scope of explaining how the Neolithic was spread in a given region (**Figure 4**). For a deeper review of the Neolithic models proposed, one can refer to Linstädter and collaborators (Linstädter et al). The classical theory on the origins of the Neolithic in North Africa states that it was spread from the Near East, where crop domestication was first developed, into Egypt and from there to the remainder of the North African region under a wave of advance model, and in a parallel way to Europe (Ammerman & Cavalli-Sforza, 1984). It is hypothesized that it involved some genetic replacement in autochthonous populations.
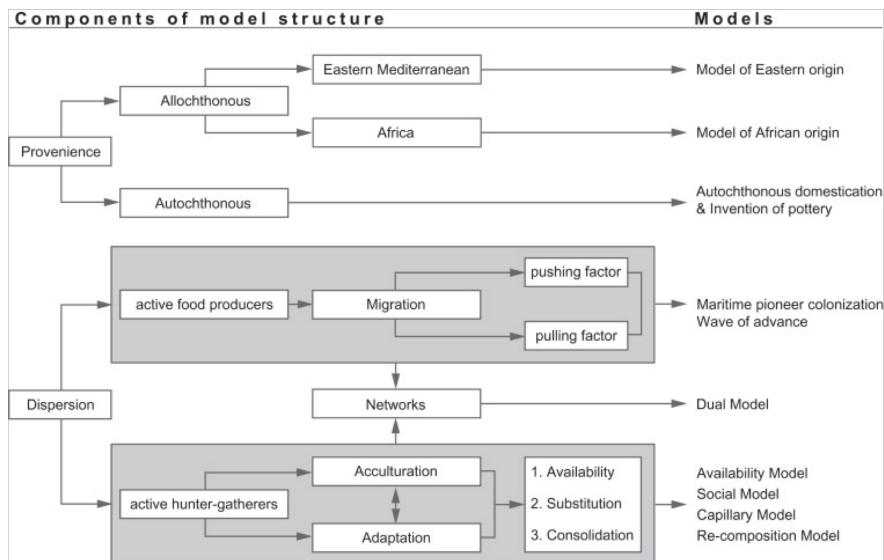
Figure 4 Diagram showing the different Neolithic models proposed (Linstädter et al)**.**

An unsolved and hotly debated question regarding this hypothesis is whether this genetic replacement actually existed and, if it did, to which extent it shaped the genetic structure of extant populations. Many aspects of this theory seem to be in agreement with data found in Egypt. Nonetheless, there are many aspects of this theory that do not fit the data found in North Africa.

Around 7,500 ya the definitive transition to agriculture takes place in the Nile Valley, being the most accepted hypothesis that it was a consequence of the Near Eastern influence. The climate optimum

that characterized this period is thought to have favored animal and plant domestication in the area, and in turn this accelerated the production of tissues like linen and animal skins, as well as ceramics and basketry. Nonetheless, improvements in hunting-related tool industry were also carried out, mainly in arrowheads and bone harpoons production, suggesting that gaming was still an important source of meat obtaining. Overall Neolithic in Egypt had a deep impact in the population, increasing the social complexity and the work specialization. Funerary practices also increased its complexity and among other characteristics implied the provisioning of food and ceramic to the deceased, his orientation to the west and his location beyond the cultivable lands, symbolizing the exit of the world of the living and his preparation for the afterlife. Of note is that during that period some differences are detected between northern and southern Egypt, being the first one more advanced in terms of tool fabrication and the second one more sophisticated in pottery production. The unification of the two regions would start 4,500 BCE and last for more than one thousand years.

In the rest of North Africa, however, Neolithic evidences are more challenging to interpret. As noted above, in North West Africa and Libya, Neolithic replaces Iberomaurusian. There are no signs of both cultures, occupying contemporaneously the same site in Morocco, but some funerary traditions, tool fabrication techniques, and the continuation of the tooth ablation practice are not incompatible with population and cultural continuity (Barton et al, 2008; Nespoulet et al, 2008). Latest evidences of Iberomaurusian assemblages are found in the Moroccan Mediterranean coast and are reliably dated to $8,900 \pm 1,100$ ya based on TL (Barton et al, 2008), whereas first evidences of early Neolithic in North Africa are dated around 7,200 ya, based on the presence of Cardial pottery and other ceramics. In Morocco, the Neolithic is first evidenced by the presence of ceramics (Nespoulet et al, 2008) and later by the presence of cereals. A gap of around 3,000 years has been reported between the origins of ceramic production, 9,000 ya, and the origins of food domestication, 6,000 ya, in North Africa (Garcea, 2006). Furthermore, archaeozoological data based on vertebrate remains show that this period supposed a climatic optimum characterized by a warm and wet climate, and during which dietary resources of

Neolithic populations were more diversified than in previous periods, with a changes towards a greater consumption of marine resources (Stoetzel et al, 2011). On the light of these facts, some authors have suggested that in North Africa sedentism would actually predate agriculture and would be associated to pottery, exploitation of aquatic resources and consumption of boiled food. This would be in agreement of an evolution of the changes previously discussed associated to Capsian populations and their *escargotières*. Here, dietary habits would have shifted to a regular consumption of stew and porridge. On top of this, it is suggested that in North Africa food domestication started with pastoralism, not agriculture, what would have demanded a further change to a nomadic settlement system related to the need of grasslands. It would not be until 4,000 ya that agriculture would have been developed in the region, cultivating millet and sorghum, vegetables suitable for making stew and porridge (Garcea, 2006; Haaland, 2005).

Other authors have noted similarities and coexistence (around 7,500 ya) in Western Iberian Neolithic ceramics and African Mediterranean ones, being different from pottery found in other places in Europe and in the eastern Iberian Peninsula (Cortés Sánchez et al; Linstädter et al), and have suggested a common African origin of this culture that would have spread to western Iberia through the strait of Gibraltar. Moreover, the existence of ceramics predating agriculture is interpreted as a "step-by-step" acculturation model. Also, the Cardial ceramics present in the early Neolithic in North Africa are also frequent along the Mediterranean and Atlantic coasts. It has been suggested that ceramics ubiquity in the Mediterranean and the progress of navigation skills would evidence some communication between both shores of the Mediterranean sea during that period (El Idrissi, 2011; Straus, 2001).

On top of this, it must be noted that in the case in which Egyptian populations transmitted the agriculture to the rest of North African populations, they did not transmit other traits of their culture, as the funerary traditions and the tool production remained different from that of Egypt.

The Neolithic is also characterized by extensive creation of art paintings that sometimes bring some information about the life-style of Neolithic populations (**Figure 5**). Most of the rock art has not been reliably dated, but first paintings date probably from around 14,000 - 12,000 ya or alternatively around 9,000 - 7,000 ya, as the landscape usually depicted shows a Sahara with lakes and rivers and a savannah ecosystem. The UNESCO World Heritage has divided cave paintings into five periods following a chronological order and respecting differences between styles. The first one corresponds to the *Naturalistic period*, characterized by the depiction of fauna from the savannah. Among the wild animals represented there are giraffes, elephants, hippos and the extinct aurochs and long-horned buffalos. Some depictions are in real scale and the anatomic details of the animals evidence a naturalistic representation and an excellent knowledge of the species represented (Gauthier et al, 1996). However, some species are far more represented than what it would be expected according to their abundance.
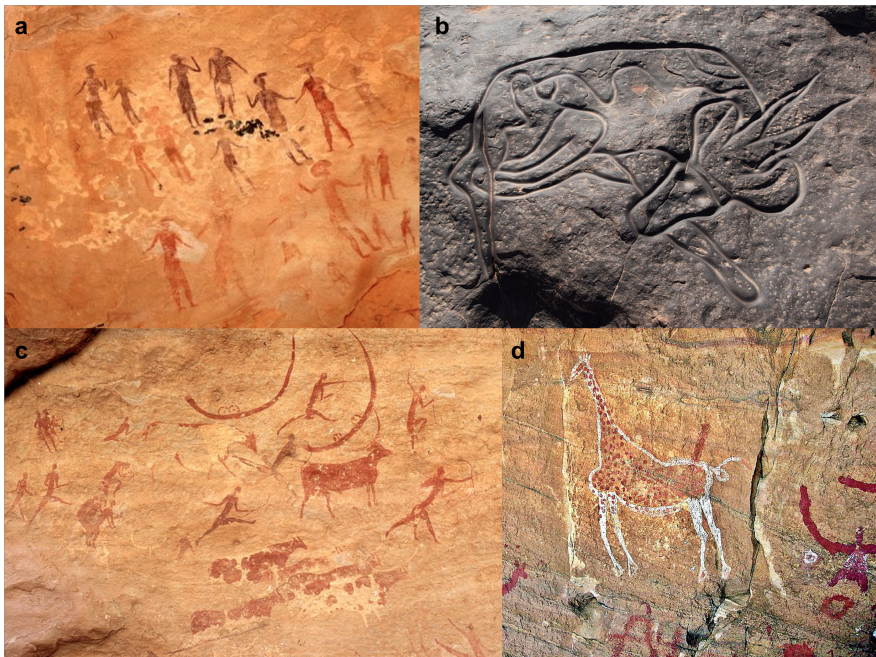

Figure 5 Cave paintings in North Africa. a) Community scene; b) Caprid or Antilope representation; c) Hunting scene; d) Painting of a giraffe

Big mammals and ostriches are frequently depicted, in opposition to other birds and smaller mammals. There is a general agreement in

stating that these depictions are not a simple representation of the environment. Some authors defend that this would be a way of communicating between populations, on the basis that often caves do not show signs of inhabitation and they are of difficult access. Other authors think that they might be the results of ritual practices led to favor the hunting. Following this period there is the *Archaic period*, which would be associated to a more symbolic oriented rock art. Fantastic animals can be found, being chimers from different animals, and colossal human figures wearing masks or with jackals or lycaons heads are depicted showing magic powers. Between 6,000 and 3,000 ya there is the *Bovine period*. It is the more rich in terms of painting representations and they belong to the best-known form of prehistoric mural art. During this period bovine herds and day-life is depicted. Interestingly, scenes of animal farming are often represented together with scenes of animal hunting, with would be in agreement with hypothesis stating that animal domestication preceded plant domestication in North Africa. Domestic animals designed often wear collars or are shown forming herds around a basement (Gauthier et al, 1996). The fourth period covers the end of the Neolithic and the beginning of historicity. It is the *Equine period* and corresponds to the first appearance of the horse in North Africa, around 5,000 - 3,000 ya and the disappearance of savannah-related species due to the aridity of the weather and the associated dryness. During this period the painting style is much more schematic, without evidences of symbolism. Finally there is the *Cameline period*, during the first centuries after Christianity.

In summary, the connection between the different cultures existing in North Africa during the Holocene is still not well established. At every intermediate period one can find researchers supporting continuity or population replacement. Also, some researchers state that climate conditions influenced in the cultural changes, whether others think it was independent. Finally, it is noteworthy mentioning that cultural continuity does not necessarily have to be linked with genetic continuity, so that conclusions on population continuity or replacement based only archaeological data should be taken with caution. About who were the first Berbers, it is commonly accepted that Berber origins can be traced back in the Capsian culture around 9,000 ya (Camps, 1997), based on

morphological similarities between Capsian human remains and extant North Africans. After that, possible interactions with Iberomaurusians and Neolithic people might have taken place though due to the absence of evidences, they are at the moment only conjectures. Finally some proto-Berber tribes would have started to expand and settle all over the region. The Neolithic period would last in North Africa until the Phoenician colonization around 3,100 ya, which would introduce the iron and bronze production techniques. Evidences of Bronze Age European culture have been found in Neolithic sites in Morocco (2010), suggesting some potential trade.

The Neolithic in Egypt is associated to the predynastic period. It starts with the incorporation of copper, (even if scarcely used) and ranges between 6,000 and 3050 BCE. Several sub-stages typically divide the predynastic period, namely the primitive period and the three Naqada periods. However, they establishment is quite arbitrary and there is not a noticeable dramatic change between one and another. During the primitive predynastic period differences between north and south Egypt are still found, and different cultures define the regions. The north conducts more hunting and fishing activities, while the south is nutritionally sustained largely on agriculture. Nonetheless, common advances in furniture, agricultural equipment, and funerary practices took place in both regions.

Figure 6 Ceramics during the predynastic period in Egypt a) Woman dancer b) Vase incorporating geometric motifs

The main novelty of this period is that tombs begin to acquire a more solid architectural appearance, and around 4,500 BCE the deceased start to carry into the afterlife models of houses and mud bricks, establishing the existence of urban architecture between 4,500 and 4,000 BCE. Jewelry, bone and ivory objects, and amulets representing animals and human figures are also found (**Figure 6a**). Also, ceramics evolve and start to be decorated with geometric motifs inspired in plants and acquire new forms (**Figure 6b**). The end of the predynastic period starts with the dynastic period and the Pharaohs rule.

c) North Africa during historicity[23]

Historicity might be defined as the period in which events start to be recorded, being thus directly associated with the development of scripture. In Egypt historicity starts well before than in the rest of North Africa, thanks to the development of hieroglyph writing

---

[2] In this section dates will be designated under the form of *common era* (CE), which is respect to the year 0 from the Gregorian (occidental) calendar

[3] References for this section are Claramunt Rofríguez S (1991) La formació i l'expansió de l'Islam. In *Història Universal*, Rubio H (ed), Vol. 2. Barcelona: Edicions 92 S.A.
, Padró i Parcerisa J (1993) L'Egipte faraònic.Ibid. Vol. 1, pp 183-205. Editorial 92 S.A.

system. Regarding the rest of the region, it was Greeks and Romans who recorded most of the historical events that took place, with the intrinsic bias that this implies.

Around 4,000 BCE ceramics were often decorated with animal representations that would be subsequently found in the early pictograms. It is not clear if those representations represent indeed historical documents or they have just a purely emblematic function, but in any case during the last years of this period the hieroglyphically scripture was consolidated. Simultaneously, the scarcity of natural resources obliged populations in Egypt to establish commercial relationships with its neighboring regions. This would tighten the contact between north and south, and in the end a cultural admixed population would appear around 3,500 ya. The early dynastic period, referred to as Thinite period, starts around 3,100 BCE after the unification of the two lands. Little is known about this period, mainly because of the absence of documented information. From this moment on, however, the classical characteristic traits of the Egyptian civilization are developed, like the worship to totemic half-animal gods, the building of mastabas that will evolve into pyramids as a funerary practice, and the tight relationship between politics and religions with the rule of the pharaohs as sons of god.

There are mainly three Kingdom periods in which the rule of the Pharaohs was in their magnificence, separated by intermediate periods where the Pharaonic rule was loosed in opposition to local rule or foreign rule. Around 550 BCE Greeks establish a small colony in the Nile Delta, followed by a Persian conquest of the whole region and their cessation to Alexander the Great. Macedonian culture did not supplant the Egyptian culture, and a new dynasty of Pharaohs descendants of Alexander the Great started and lasted over three hundred years.

First reports of North African history refer to the Phoenician establishment around 900 BCE. They settled in Tunisia and about 100 years later build the city of Carthage. As time passed, they founded small settlements along the western coast, where they built market towns and use them to anchor trading ships.
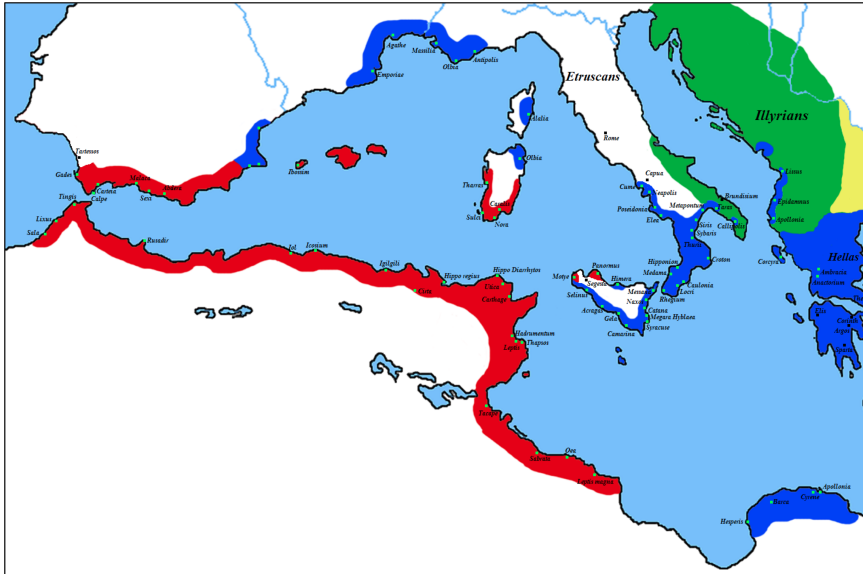
Figure 7 North African colonies in 550 BCE. In red are represented colonies from Phoenicia and in blue colonies from Greece.

Eventually, Carthage would become the wealthiest city in the Mediterranean thanks to its trade and commerce. Phoenicians ruled the trading across the Mediterranean Sea for some centuries and once in North Africa, trading between Carthaginians and Berbers is also reported (**Figure 7**).

Until 550 BCE Carthage had to pay a rent to Libyan Berber tribes in order to use the land in the city surroundings, and a century later it stopped definitively when Carthaginians started to expand inland. Some conflicts with Libyans, Numidians and Mauri Berbers are recorded, but in the end Phoenicians did not conquer the territory, even if they enslaved Berbers or recruited them to belong to the Carthaginian army. In 264 BCE the Punic wars started between Carthage and rising Rome, with the involvement of those recruited Berbers. In the end of the third Punic war (146 BCE) Carthage was defeated, destroyed, and under Roman rule. Macedonians would also lose Egyptian and Libyan settlements around 30 BCE. This would give Romans the hegemony of the Mediterranean Sea, and the beginning of the Roman rule in North Africa.

With the decline of Carthage some Berber Kingdoms emerged inlands. Since that moment, there might have been several fights, fusions and divisions of these Kingdoms, though little information

is available. Mauretania rose as a Roman client kingdom occupying the present Morocco, though best-known Berber kingdom was Numidia (202 to 46 BCE), in today Algeria and western Tunisia. Since its creation the Kingdom would alternate between being sovereign and under the Roman rule. The first King, Massinissa, would have helped the Romans to win the Punic wars, after changing side. He aimed to unify all semi-nomad Berber tribes from the region to build a strong state, and to that end he introduced Carthaginian agriculture techniques and forced many Numidians to settle as peasant farmers. After the death of the King in 148 BCE the territory was divided among his descendants and constant quarreling among them and with the Roman Empire avoided a new reunification of the kingdom. From that moment on Numidia would alternate periods of independence with periods under the Roman control or even some territory transfers to the Mauretanian kingdom.

During the first centuries of Roman occupation influence on North African territory was minor, but in 27 BCE, with the transformation of Rome to an Empire, changes were made and North Africa was divided in two provinces, one north and the other south. This would cause the Romanization of the region and of part of the Berber population. Most North African towns became notably prosperous thanks to agriculture, and grain, olive oil and other crops were produced and exported towards other regions of the Empire, in addition of pottery and the trade of exotic animals as well as the trade of Sub-Saharan slaves. By the earliest centuries ACE North Africa was one of the wealthiest Roman provinces, and thus appealed the migration of numerous Romans, most of them veterans from the army. Also, Christians and Jews started to settle in the region. This had an especially deep impact in Egypt, where Christianity started to substitute the Egyptian culture until it completely disappeared. Some local populations were also Romanized during this period, with Berbers actively participating in the imperial security forces.

In 439 Vandals settled in the Iberian Peninsula invaded North Africa and founded their own kingdom which included the African provinces and the Mediterranean western Islands. Nonetheless, they could not held interior territories and finally they abandoned them

in hands of non Romanized Berber tribes. In 533 Western Romans (or Byzantines) conquered back the North African territory, though they had to repel constant attacks from Berber tribes. They were able to hold the territory for more than a century, but in the beginning of the 6th century the Muslim Umayyad Caliphate conquered Egypt and in 698 an Egyptian Muslim army attacked and conquered Carthage. By 709 all North Africa was under the Arab caliphate control and their population converted to the Islam. Soon after the conquest of North Africa, a mixed group of Berbers, Arabs and sub-Saharan Africans, known as the Moors migrated to the Iberian Peninsula where they would rule until 1492.

This would suppose a dramatic change in the culture of North African populations, with the Islamization of the Berbers and the adoption of the Arab language by most of the population in opposition to the Berber language. There is some debate about to which extent there is a correlation between Arab-speaking North Africans being actually from Arabic descent and Berber-speaking North African being autochthonous populations already existing in North Africa previous to the Arabic expansion. Until now genetic studies have not been able to differentiate between Arab-speaking and Berber-speaking populations.

In year 739 there was a Great Berber Revolt, which supposed the secession of the Arab caliphate. Umayyad succeeded in maintaining their rule in Al-Andalus (Spain) and Ifriqiya (Tunisia), but lost the remaining Maghrebi territory, which from that moment was divided into small states and ruled by local tribal chieftains and Kharijite imams, an Islamic school of thought. From that moment on, a number of dynasties ruled different regions of North Africa during different periods of time. Of special mention are the Almoravids, a Moroccan Berber dynasty with its origins in nomadic tribes from the Sahara that ruled in Western Maghreb and Al-Andalus during the 11th century and prevented the fall of Al-Andalus to the Christian kingdoms. Almoravids were defeated by the Almohads, which ruled in Iberia until 1212, where they lost against an alliance of Christian princes.

After the middle Ages most of North Africa with the exception of present-day Morocco was under Ottoman rule, and in the 18th and

19th centuries Spanish, French and English countries colonized North Africa. In 1914 the whole African continent (including North Africa) was divided into the main European powers defining the present-day actual boundaries. North African countries would not be independent until the 20th century.

In summary, recorded history describes North Africa as a key settlement to rule the Mediterranean, which is evidenced by the numerous conquerors that have dominated the region during the last two millennia. However, the proportion of descendants that these conquerors left in today North Africans is not clear. During the conflicts and rule changes it is probable that most of the former inhabitants fled from the region or died once the new invader had won. On the other hand, nomad Berber tribes living in the inlands have always lived independently from what was going on in the coast, not likely being affected by the different invaders.

d) Languages in North Africa[4]

The family of languages and dialects spoken by people in North Africa and the Sahara Desert prior to the Arab expansion is the Berber, which is indigenous to the region, together with ancient Egyptian, which was spoken only in Egypt. They both belong to the Afro-Asiatic branch together with Semitic languages such as Arabic or biblical Hebrew. During the Arab Caliphate rule the Arabic language started to spread replacing the use of the indigenous Berber languages. As a result, ancient Egyptian became extinct and Berber languages in the rest of the region underwent a dramatic reduction in usage. In the present day Berber is still dominant in specific areas from North West Africa, being spoken by a great number of communities in Morocco, Western Sahara and Algeria. On the contrary, the rest of North African populations mainly speak Arabic (**Figure 8**).

---

[4] References for this section: Camps G (1997) *Les Berbères: Mémoire et identité*, Paris: Actes Sud.
, Heine B, Nurse D (2000) *African Languages: An Introduction*: Cambridge University Press.

26
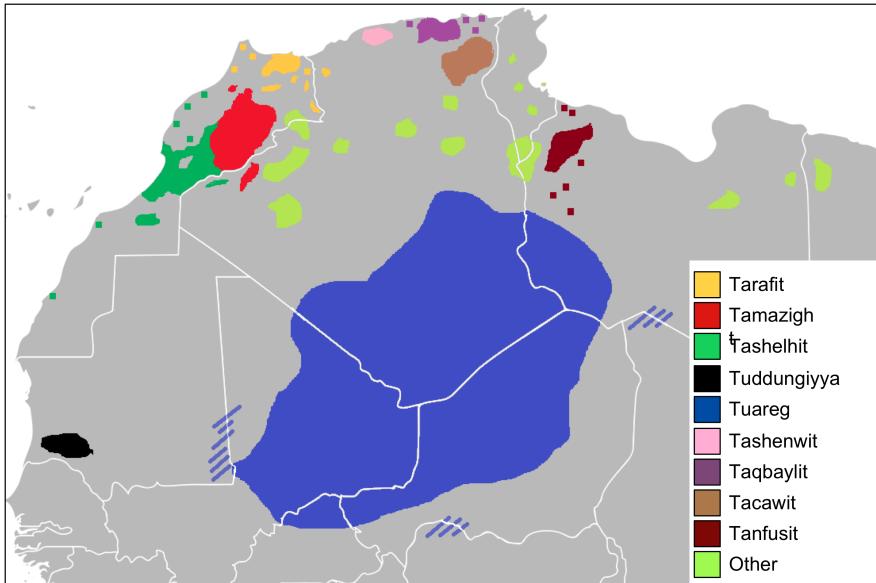
Figure 8 Present day distribution of Berber languages in Africa.

It is unknown and a long-standing debate to which extent the language replacement involved a demographic replacement as well. On one hand historic literature often refers to an expulsion of Berber tribes to the Atlas and to the Desert by the invading armies during the Arab expansion, coinciding with the areas where Berber is spoken today. On the other hand, many North Africans report Berber ancestry, independently of their native language, and some Berber-speaking North African identify themselves as Arabs. Some genetic studies have been undertaken to address this question, and they will be discussed in the next section.

27

1.2. North African demography: genetic evidences

1.2.1 Evidences revealed by autosomal markers

a) Evidences from classical polymorphisms

Classical polymorphisms were the first markers used in detecting genetic variation among individuals, and constitute all those elements detected by means of antigen - antibody reactions or by protein electrophoresis. In the year 1900 the ABO blood group system was described by Karl Landsteiner (Landsteiner, 1900; Owen, 2000), and it is now considered to be the first human genetic polymorphism to be discovered. Landsteiner found out mixing red blood cells and blood serum from different individuals that red blood cells could present antigens A, B or none (O). When the individuals had different (and incompatible) blood group, there would be an antigen-antibody reaction, and the mix of red cells and serum would form a lattice network. In the following years, a number of other blood group systems and HLA systems were defined using the same technology. From the second half of the century on, the introduction of protein electrophoresis allowed the separation of proteins according to their weight, and analysis of many blood proteins revealed variation across individuals on those proteins too, increasing the number of classic markers available to describe variation in humans and other species.

Bosch and collaborators (1997) compiled a database of allele frequencies based on classic markers in North African populations. Data was extracted from previous studies (Cavalli-Sforza et al, 1994; Mourant et al, 1976; Roychoudhury & Nei, 1988; Tills et al, 1983) and included blood groups, red cell enzymes, serum proteins, and HLA antigens, entailing a total of 62 different loci. Principal Component Analysis, Neighbor-Joining trees based on genetic distances and a Delaunay network were performed. The main finding is that genetic structure within North Africa exists, with a sharp genetic differentiation between a group formed by Egypt and Libya, and the rest of North African populations (including Mauritania) (Figure 9). Authors claim that an isolation-by-distance model could generate only part of this differentiation, and that other

factors such as directional migrations from east to west must have taken place as well. Specifically, they associate this pattern of genetic variation with the Neolithic expansion that originated in the Fertile Crescent. They hypothesize that the agriculture would have easily reached the Nile River Valley and its surroundings, but on the contrary it would have taken a couple of centuries to reach the west due to the harshness of the terrain, extending the Hunter-Gatherer lifestyle. This is in agreement with the findings regarding the Capsian techno-complex discussed in the former section. Authors also point out that the Arabic expansion may have contributed to the differentiation between eastern and western North Africa in addition to the Neolithic expansion.
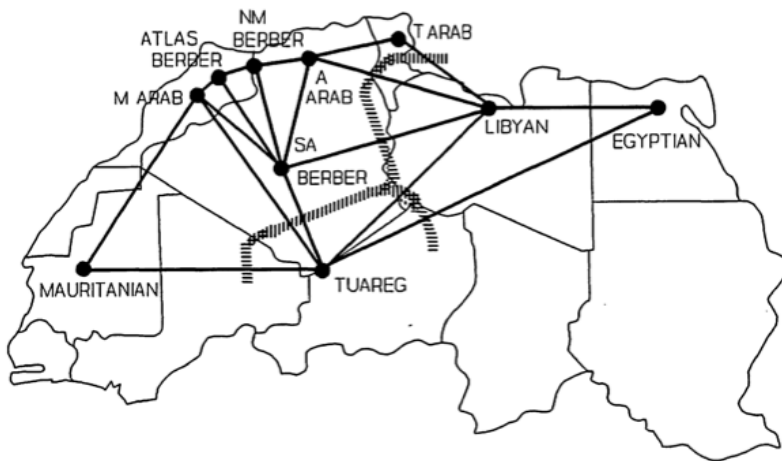
Figure 9 Delaunay triangulation between populations in North Africa (solid lines) and the two most significant genetic boundaries (shaded lines). Abbreviated names are read as follows: NM BERBER North Moroccan Berber; M ARAB Moroccan Arab; A ARAB Algerian Arab; SA BERBER South Algerian Berber; TARAB Tunisian Arab. (Bosch et al. 1997)

Interestingly, the Neighbor-Joining tree shows that Arab speakers are genetically closer to Libya and Egypt than Berber speakers, which suggests some degree of admixture between the local Berbers and Arab migrants.

Other findings are that genetic relationship between North Africa and both, sub-Saharan Africa and Europe, is weak. The Arabic expansion into Iberia had a limited genetic impact, in agreement with historical records, as shown in the Neighbor-Joining tree,

where Libya is genetically closer to Andalusia than Western Arabic speakers. Similar results are found in a dataset of Mediterranean populations, where a North-South differentiation can be appreciated based on the analysis of 11 classical polymorphisms (Harich et al, 2002). Finally, the contribution of sub-Saharan Africa into North African populations seems to be relatively small. The differentiation of Mauritanians, Tuareg and south Algerian are explained by the effect of drift on small isolated populations (Bosch et al, 1997), and no significant differences are detected between Berber-speaking and Arab-speaking groups (Coudray et al, 2006; Harich et al, 2002).

Nevertheless, it must be noticed that in Bosch et al (1997) analyses that compare North Africa with its neighboring regions, North African samples are grouped according to geographic and cultural criteria, which may cause spurious results if there is genetic substructure within the grouped populations. This could explain that Berbers appear closer to sub-Saharan populations, because many Berber populations are located in the south and most likely they have had some gene flow from sub-Saharan Africa.

b) Results from autosomal repetitive elements: Alu sequences and STRs

Alu elements (**Figure 10**) are the commonest members of the SINEs (Short Interspersed Nuclear Elements) family of dispersed repeats. It is estimated that SINEs comprise 13% of the genome, with around 1.5 million repeats per haploid genome, each repeat having a length of up to 100-300 base pairs (bp) (Houck et al, 1979; Jobling et al, 2004; Watkins et al, 2001; Witherspoon et al, 2006).

Figure 10 Structure of an *Alu* dimer (adapted from Jobling et al. 2004).
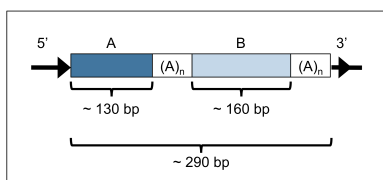
The possibility of sequencing the human genome and the improved understanding of the evolutionary relationships of different families of Alu elements, as defined by sequence variants, allowed more systematic attempts to isolate Alu polymorphisms. As a matter of fact, a large number of Alu insertion polymorphisms have been discovered in the human genome. In

addition they have a number of advantages, like that they are easy to type, most of them are thought to be neutral and have well established ancestral states (El Moncer et al, 2010). Also, every Alu insertion in the genome is most likely a unique event in human evolution, and thus shared Alu polymorphisms among individuals are considered identical by descent and free of homoplasy. In summary, Alu polymorphisms are markers of a widely contrasted informative nature, and as a consequence in the last 25 years they have become a very common genetic marker.

STRs (Short Tandem Repeats) are also referred to as microsatellites, and are tandem arrays of repeat units ranging between 1 and 6 bp length (e.g. AC, ATT, CAG, AGAT). Those used as genetic markers are present typically between 10 and 30 times in the genome. Microsatellites with some specific repeat units show clustering, but most of them are distributed throughout the genome (Jobling et al, 2004). Their mutation rates have been estimated to be 10^{-3} - 10^{-4} (per locus per generation) based on direct pedigree analysis (Brinkmann et al, 1998) and mutant detection in small populations of sperm DNA molecules (Di Rienzo et al, 1998). Important properties of the mutation process in microsatellites have determined their usefulness as population genetics markers. First, most of the mutations in microsatellites involve the loss or gain of a single repeat unit (Brinkmann et al, 1998; Xu et al, 2000), so differences in the number of repeat units between populations is directly related to how closely related these populations are. Second, mutation rate increases proportionally to array length, though below a given threshold of repeat number the mutation rate is undetectable (Xu et al, 2000). This deep knowledge of mutation mechanisms can be easily taken into account in the study of microsatellites. The same authors found out that expansion occurs equally throughout the array size range, but contraction increases as the array becomes larger, so that allele lengths have a stable distribution. Third, dinucleotide microsatellites mutate faster than trinucleotide or tetranucleotide (Webster et al, 2002); and fourth, uninterrupted repeat arrays mutate also faster than interrupted ones containing variant repeats (Jobling et al, 2004). Most studies based on STRs are focused on the Y chromosome, but some of them are also autosomal and analyzed together have the advantage of representing several loci in the genome, and thus overcome the problem of stochastic processes acting at a single locus.

Initially, studies on population genetics based on autosomal markers focused on either Alu insertions or STR polymorphisms. However, some STR haplotypes were found to be in linkage disequilibrium with specific Alu insertions in human populations. The analysis of these Alu/STR compound systems provides a remarkably higher degree of qualitative information and allows focusing in different span times taking advantage of the different mutation rates of the two markers. Most of the studies using Alu insertions and STR polymorphisms on North African populations are interested in investigating the general genetic structure of the Mediterranean basin. Reassuringly, they have remarkably similar results, though some differences exist regarding the hypotheses and the conclusions stated.

Overall, Mediterranean populations are genetically distinct from Central European and sub-Saharan African populations, as seen in multidimensional scaling representation based on genetic distances (El Moncer et al, 2010; Gonzalez-Perez et al, 2010) (**Figure 11**). Also, different studies detect sharp differences between North and
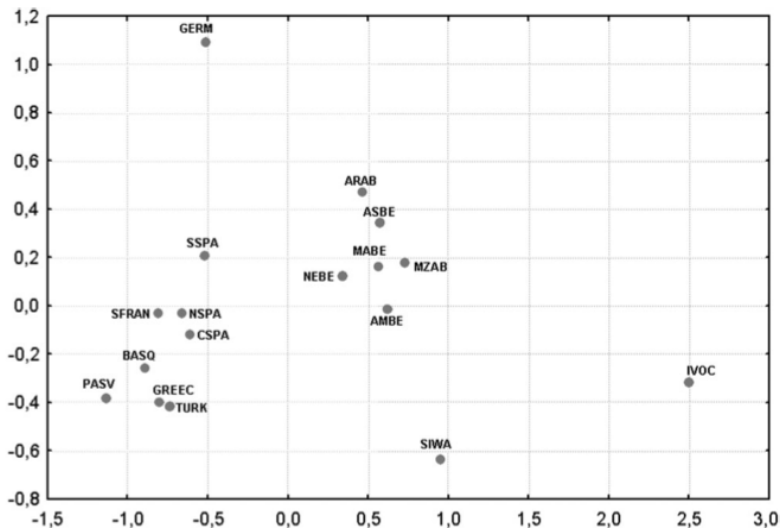
Figure 11 MDS plot (stress 0.049) applied to the Reynold's genetic distance matrix based on three *Alu*/STR compund systems. Abbreviations are read as follows: GERM Germany; SSPA South Spain; SFRAN South France; NSPA North Spain; CSPA Central Spain; BASQ Basques; PASV Pasiego Valley; GREEC Greece; TURK Turkey; ARAB Arab speakers; ASBE Asni Berbers; MABE Middle Atlas Berbers; AMBE Amizmimz Berbers; MZAB Mozabites; SIWA Siwa; IVOC Ivory Coast (Gonzalez-Perez et al. 2010)

South Mediterranean populations (Bosch et al, 2000; Comas et al, 2000; El Moncer et al, 2010; Gonzalez-Perez et al, 2010), though statistical significance in a hierarchical AMOVA is reached only in the last study, in which the compound systems Alu/STR are considered. In a similar way, genetic differences between eastern and western North Africa have also been described (Bosch et al, 2000; Flores et al, 2001). Discrepancies are found regarding the overall homogeneity of the Mediterranean basin.

Measures of genetic differentiation and variance in the Alu loci frequencies in Comas et al (2000) show that the region is relatively homogeneous, whereas in Gonzalez-Perez et al (2010) significant heterogeneity is detected with a high correlation between genetic and geographic distances, following an isolation by distance model. This inconsistency might be caused either by the different number of markers used or by the different geographic range considered. Finally, low genetic differentiation is found between Arab and Berber-speaking groups in North Africa, and distance from these two groups to European populations is very similar (Bosch et al, 2000; Comas et al, 2000; El Moncer et al, 2010).

In what gene flow is concerned, evidences of North African admixture with sub-Saharan populations are found in all mentioned studies except for Bosch et al (2000), where the only sample with known sub-Saharan ancestry is constituted by African-Americans. The proportion of sub-Saharan ancestry present in North Africa ranges from 13% to 46% depending on whether Alu, STR/Alu or STR markers are used to measure variability. It is expected that given the high mutation rate of STRs, the upper bound of admixture proportions will be an overestimation produced by repeated homoplasic mutations. Generally, gene flow from North Africa to Europe is considered to be weak or absent in all studies. For instance, proportions of admixture are estimated to be around 6% using STRs but negligible using Alu or STR/Alu compounds (Gonzalez et al, 2007), and in Bosch et al (2000) and Comas et al (2000) evidences of gene flow in the Iberian Peninsula are found only in Andalusians and Portuguese. A remarkable exception to this is the work presented by Flores et al (2000), where they estimate around 25% of North African admixture in Iberian populations, using a variety of methods. Nevertheless, it must be noticed that their work is centered on a single locus, and therefore results may

be just the result of stochastic processes and not representative of the whole population demography.

Finally, some population outliers have also been found in North Africa (Bosch et al, 2000; Gonzalez-Perez et al, 2010). They are populations that bear low levels of genetic diversity when compared to the rest of the region and it is thought that the low variability is the consequence of the strong effect of genetic drift. Also, these populations are known to be either geographically or culturally isolated, which explains why the main force of genetic variation is drift.

On the light of these results, several hypotheses have been presented. Some studies justify differences between both shores of the Mediterranean by a differential effect in population replacement during the Neolithic wave (Bosch et al, 2000; Comas et al, 2000). In this way, extant North African populations would trace back to a Paleolithic origin, whereas in Europe the Neolithic revolution would have entailed a demographic replacement. This is compatible with the hypothesis presented in Gonzalez-Perez et al (2010), in which the geographic structure seen in the Mediterranean would be the consequence of longer periods of isolation in North Africa compared to Europe. However, in this study no time frames are specified. Also, all studies find evidences of ancient rather than recent sub-Saharan gene flow into North Africa in the Alu/STR haplotype frequency distribution. Dates range between 9,000 and 5,000 years ago (El Moncer et al, 2010; Gonzalez-Perez et al, 2010), coinciding with the Neolithic expansion and a period of milder climatic conditions in the desert. Finally, genetic differentiation between northern and southern Mediterranean reinforce the hypothesis that the Islamic expansion had a minor effect on actual Iberian Peninsula populations. However, Alu insertions and STR polymorphisms in the CD4 locus show that Iberia bears higher heterozygosity levels, has more haplotypes and less linkage than other European populations (Flores et al, 2000), and haplotypes of most likely sub-Saharan origin are present across all northern Mediterranean (Gonzalez-Perez et al, 2010) coast. These facts evidence that North African gene flow to Europe exists, and that it is most likely more ancient than the Islamic expansions.

1.2.2 Genetic structure in North Africa revealed by uniparental markers

a) A brief introduction to uniparental markers: Y chromosome and mitochondrial DNA

In humans, as well as in other diploid organisms, the great majority of the DNA in parental and maternal gametes undergoes a process of independent assortment and recombination that ensures that each gamete is genetically unique, being a mosaic of their respective paternal and maternal lineages. As a consequence, investigating the genetic contribution of each individual's ancestor or building phylogenetic trees of a given locus is extremely difficult. Exceptions to this are mitochondrial DNA (mtDNA) and Y-chromosome, which are respectively matrilinearly and patrilinearly inherited without recombination (with the exception of a part of the chromosome Y that does recombine with the X chromosome) (Giles et al, 1980; Jobling & Tyler-Smith, 2003) (**Figure 12)**.

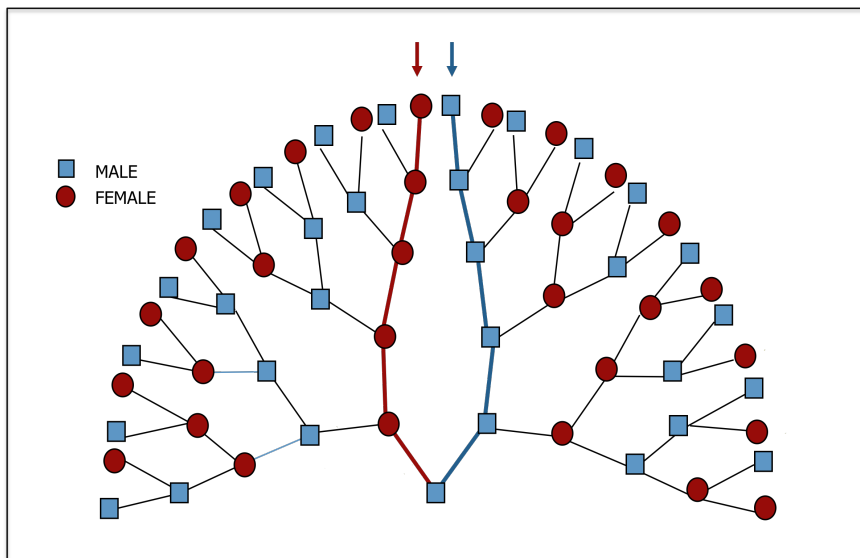
Figure 12 Diagram representing the inheritance of uniparental markers. The genealogy of a man (bottom square) is drawn, including the last five generations. It can be noticed that even if all the 32 ancestors from the fifth generation contributed to the genome of this individual, only two of them, a male and a female, are responsible of its Y chromosome and mitochondrial DNA, respectively.

The absence of recombination entails first, that haplotypes shared between individuals are identical by descent, i.e. have the same common ancestor, and second that variation in these lineages comes solely by the accumulation of mutations. As a consequence, phylogenetic trees (**Figure13, 14**) are built and are used to infer the level of structure among populations and the order and time of their descent (Underhill & Kivisild, 2007). Also, the study of these phylogenetic trees together with geography allows making inferences about migration history (Jobling, 2012). The polymorphisms used to classify sequences into haplogroups are Single Nucleotide Polymorphisms (SNPs), mutations that occur at a single locus and are almost always biallelic. The mutation rate of these markers is very low, of the order of 10^{-8}, and therefore sequences sharing a SNP have most likely a common ancestor. Also, knowing the mutational rate of this and other polymorphisms related to uniparental markers have established a *molecular clock* that helps dating demographic processes. However, some discrepancies exist regarding the reliability of this clock, especially for the Y chromosome (Zhivotovsky et al, 2004), so dates are often taken with caution. In summary, haploid characteristics of these genetic markers allow the successful application of both phylogenetic and phylogeographic (Ashlock, 1974) approaches to the population genetics field, and therefore constitute a powerful tool to compare maternal and paternal lineages among human populations. Finally, the advent of PCR amplification and sequencing of the DNA has made easier and faster the study of these markers, and in the last 20 years a great number of studies have been performed.

The first mtDNA complete sequence was available in 1981 (Anderson et al, 1981) and is now known as CRS (Cambridge Reference Sequence). This sequence was revised by Andrews et al. (1999) rCRS, and recently a new nomenclature of the mtDNA complete sequence has been proposed (Behar et al, 2012). After PCR sequencing was available, the great majority of studies used only the genetic information contained in a specific region of the mitochondrial genome, the hypervariable sequences I and II (HVS I and II), and sometimes some informative polymorphisms found in the coding region were also genotyped and included in the analyses. As more human sequences were described the mtDNA phylogenetic

tree was updated, the accuracy within haplogroup branches increased, and the global picture of this haploid marker improved.
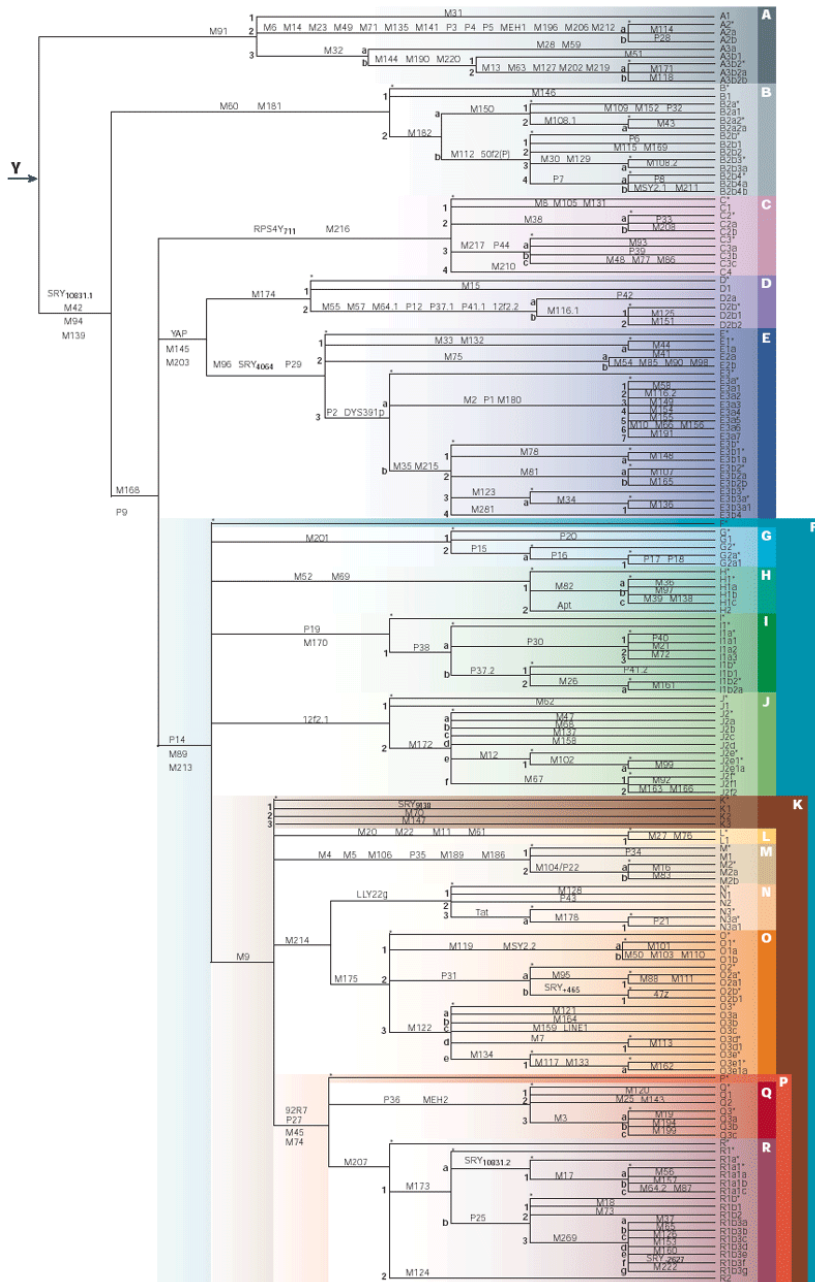

Figure 13 Y chromosome phylogenetic tree. Described mutations are shown in the branches and haplogroup names at the tip of the branches (Jobling 2003)
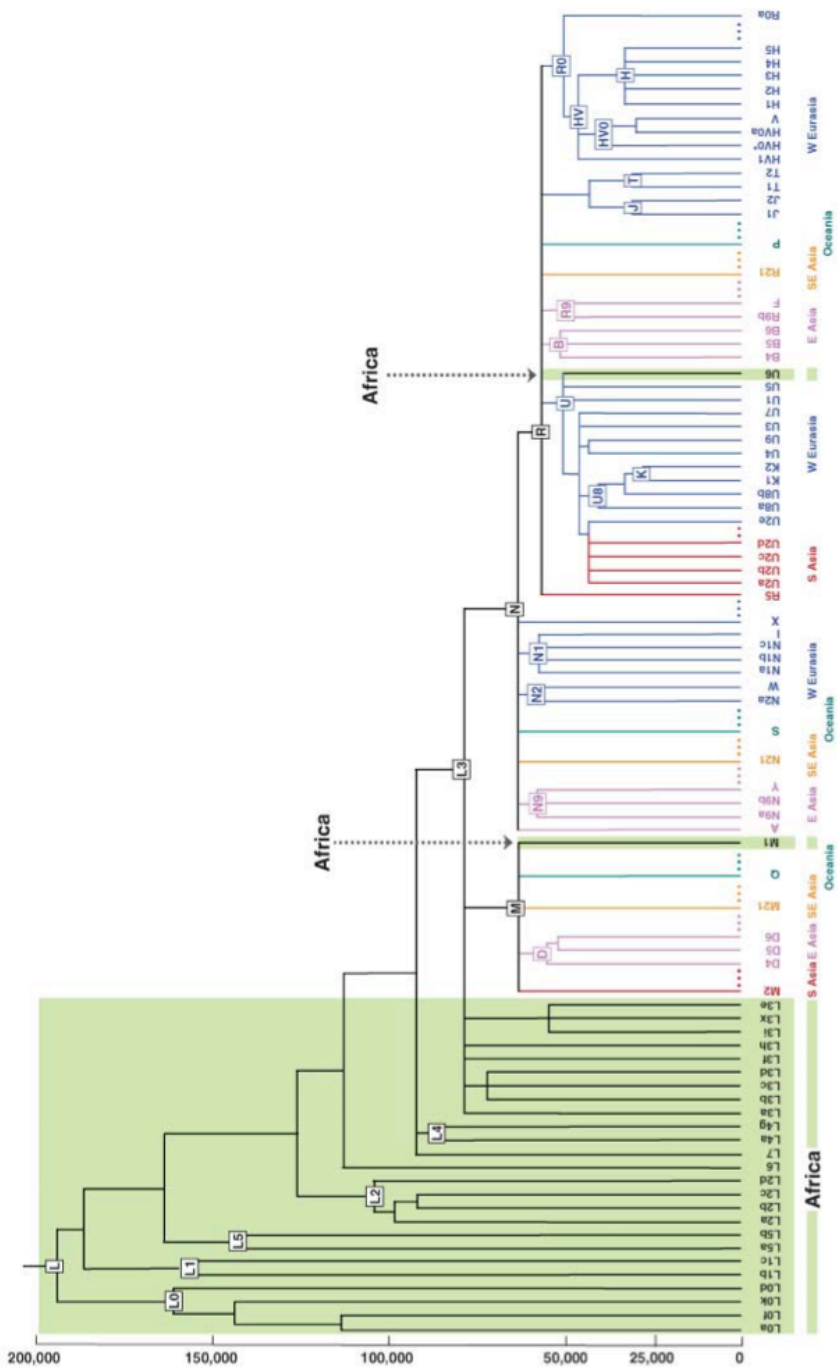
Figure 14 Mitochondrial DNA phylogenetic tree. As in Figure 13, each branch represents a haplogroup. In the left part of the plot, a time scale (in years) is included (Olivieri et al. 2006).

Finally, in the last years some analyses based on complete mitochondrial sequences have been performed, which have helped to overcome problems of intraspecific homoplasy (as a result of the high mutation rate) which in turn introduces complications to the tree building, especially when only the hypervariable region is taken into account (Kivisild et al, 2006).

The Y chromosome is the chromosome that determines the male sex function in humans and other mammals, and thus it is uniparentally inherited from father to sons. 95% of the Y chromosome does not recombine and the regions that do recombine are in the termini of the chromosome forming the pseudoautosomal regions (Jobling et al, 2004). In opposition to mtDNA, Y chromosome is characterized by having several types of polymorphisms, and the ones used as genetic markers are SNPs, STRs and indels. In a similar way to mtDNA, SNPs in the Y chromosome are used to define haplogroups. Underhill *et al.* (2000) developed a set of markers and typed a large set of samples from worldwide populations, providing a well-established Y-chromosome phylogeny, with very detailed information, against which any particular new population could be evaluated. This worldwide phylogenetic tree has been recently updated and now contains around 600 SNPs defining 311 lineages or haplotypes (Karafet et al, 2008). Also, SNPs and STRs are known to have different mutation rates, and turn out to be very useful in investigating demographic events at different geographic and time scales. In this way, haplogroups, which are defined by SNPs, will typically embrace a wide geographic range, whereas variation within haplogroups, characterized by STRs, will be more locally restricted (Underhill & Kivisild, 2007).

In summary, the scientific community has made extensive efforts in sequencing worldwide samples, and it has been shown that some genetic variants have arisen locally and are able to distinguish between the continental gene pools (Underhill & Kivisild, 2007). Also, uniparental markers have a relatively low effective population size, are more sensitive to effects related to rapid population divergence compared to autosomal markers (Jorde et al, 1998), and consequently are especially suitable as a genetic marker in the investigation of recent migrations. Nonetheless, it must be noticed that the smaller effective population size of uniparental markers has also a main disadvantage. Small and isolated populations will have

an increased genetic differentiation detected through those markers because the effect of drift will be much stronger. Also, though a lot of variation can be found in uniparental markers, they represent two single loci, and therefore they are also more prone to stochastic effects from drift. Finally, there are evidences that both mitochondrial DNA (Baudouin et al, 2005; Topf et al, 2007) and Y chromosome (Sezgin et al, 2009) are indeed subject to natural selection, which may obscure inferences of demography.

b) Results from mtDNA analysis

A lot of studies have been performed using mtDNA markers in the population genetics field, but few of them have investigated North African populations. Those that have focused on this region are based on relatively small communities in Morocco, Tunisia and Egypt, whereas studies that include Algeria and Libya or that have an integrative point of view of the region are very scarce. In opposition to the findings revealed by autosomal Alu and STR markers, mtDNA results show that, broadly, North Arica is a highly admixed region. Surprisingly, the bulk of mitochondrial haplotypes have a European or a Near Eastern origin. Sub-Saharan haplogroups are also found in North Africa, conforming between 15 and 30% of the population genetic pool. The remaining sequences, representing around 10% of total mtDNA lineages, belong to the back-to-Africa haplogroups U6 and M1, and are often considered as autochthonous lineages because they reach highest frequencies in North Africa (M1 also in Ethiopia), even though they originated elsewhere. Interestingly, the contribution of these back-to-Africa haplogroups varies a lot across populations. This complex maternal landscape can be visualized in analyses like Principal Component or Multidimensional Scaling based on either haplogroup frequencies or genetic distances, where North Africans are located between European and sub-Saharan populations but more closely related to Europeans (**Figure 15**). Also, AMOVA (Analysis of MOlecular VAriance) on sub-Saharan, European, Near Eastern and North African populations have shown that populations within groups are relatively homogeneous and that groups are significantly differentiated between them. Finally, Fst values show strong diversifications among groups (Coudray et al, 2009).

Mitochondrial DNA studies in North Africa have been centered in three main questions. (i) Ascertaining gene flow between North Africa and its neighboring regions (i.e. sub-Saharan Africa, Europe and the Near East), (ii) comparing the genetic background of the Arab speaking and Berber speaking populations, and (iii) finding out the phylogeographic structure and time estimates of the U6 and M1 back-to-Africa haplogroups.
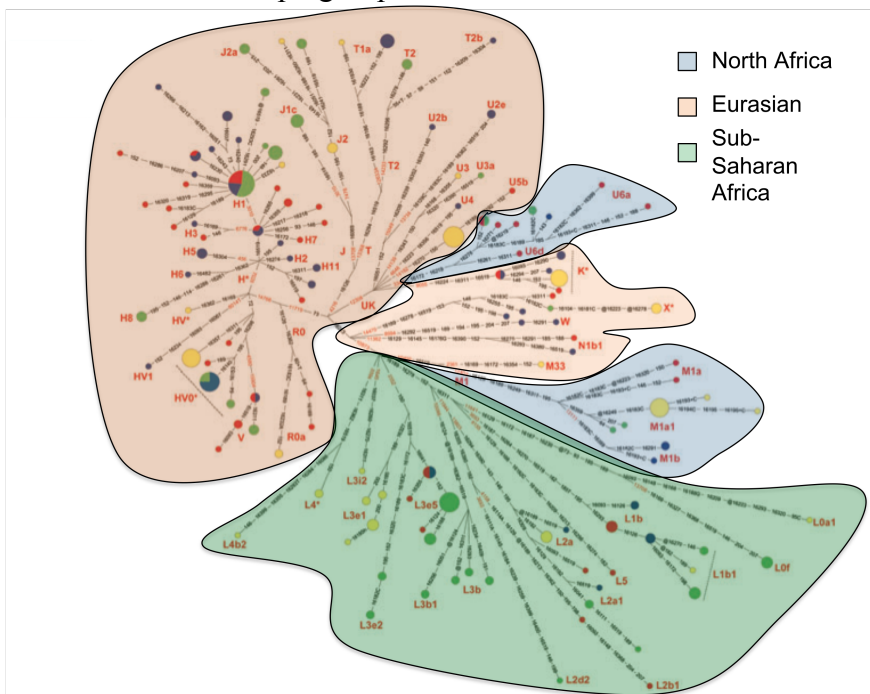
Figure 15 Haplogroup phylogenetic network of four North African populations from Morocco and Egypt. Circles are proportions to the number of individuals bearing that given haplogroup. Shaded areas represent the geographic origin of these haplogroups (adapted from Coudray et al. 2009)

• *Gene flow between North Africa and neighboring regions*
Studies based on autosomal markers had shown a slight degree of gene flow from sub-Saharan Africa towards North Africa, and this subject was also investigated using mitochondrial DNA. A pronounced decreasing gradient of sub-Saharan mitochondrial L lineages has been detected from north to south, in western (Brakez et al, 2001; Rando et al, 1998) and eastern (Krings et al, 1999) North Africa. Particularly, in eastern North Africa evidences of a bidirectional corridor have been found, based on the frequency and diversity of typical northern and southern sequences along the Nile

River Valley. Nonetheless it is interesting to notice that this north | south gradient is interrupted in western North Africa by Arab-speaking populations, that seem to be more permeable to sub-Saharan gene flow than Berber-speaking populations (Coudray et al, 2009; Loueslati et al, 2006; Plaza et al, 2003; Rando et al, 1998). Also, AMOVA analysis comparing European, North Western African and sub-Saharan populations has shown that sub-Saharan lineages in North Africa contribute in part to the genetic differentiation found between Europe and North West Africa (Plaza et al, 2003). Results are further supported in a Principal Component Analysis the authors performed, in which North Africa and Iberian populations cluster together when sub-Saharan lineages are not included in the analysis. This is in clear disagreement with most studies based on autosomal markers, where a clear genetic barrier was depicted between both shores of the Mediterranean.

Finally, it is not clear the time since admixture between sub-Saharan and North African populations. Fadhlaoui-Zid *et al.* (2004) point to an *in situ* evolution of some sub-haplogroups that have a coalescence age of 10,500 years which would be in agreement with an Upper Paleolithic migration. However, a more recent study suggests that trans-Saharan slave trade is the main source of sub-Saharan lineages present in North Africa, based on its continental distribution (Harich et al, 2010).

Concerning gene flow with western Eurasia, the accurate dissection of haplogroup H (one of the most common haplogroups) into several monophyletic sub-haplogroups has been of capital relevance. It has changed the vision of the continent from a rather uniform landscape into one with several regional peaks and clinal variations: H1 and H3 are the lineages more common in Western Europe (Achilli et al, 2004a; Torroni et al, 2001b), whereas H6, H5 and H2 are more abundant in the east (Loogvali et al, 2004; Roostalu et al, 2007). In a similar way to Europe, this eastern cline of haplogroup H and of sub-haplogroup H1 frequencies is also observed in North Africa (Coudray et al, 2009; Ennafaa et al, 2009). In fact, coalescence ages estimated in North Africa for these sub-haplogroups coincide with a Paleolithic migration from the Franco-Cantabrian refuge after the Last Glacial Maximum. Interestingly, Tunisians, Tunisian Berbers and Moroccan Berbers seem to have received more influence from the Near East, whereas western Moroccan and Saharan North African populations have more influence from the Iberian Peninsula, based on the ratio of sub-

haplogroup frequencies (Ennafaa et al, 2009). As expected, Egyptian populations have a larger proportion of eastern European and Near Eastern H sub-haplogroups (Kujanova et al, 2009; Saunier et al, 2009), whereas Libyan Tuareg, an isolated population with extremely low levels of genetic diversity, has high frequencies of western European haplogroup H1 (Ottoni et al, 2009). A deeper phylogeographic analysis of H sequences on North Africa, Europe and the Near East could help to confirm whether the introgression of H1 and H3 into North Africa corresponds to a Last Glacial Maximum expansion, and whether a similar process took place from eastern Europe or the Near East at the same time or more recently.

- *Genetic relationship between Arab and Berber-speaking populations*

The question of whether genetic differences exist between Arab-speaking and Berber-speaking communities in North Africa has been a long-standing object of interest and an approach to estimate the demographic impact of the Islamic expansion in the region. Results based on classic polymorphisms did detect a differentiation between the two groups, whereas Alu and STRs sequences did not show a significant difference between them. In agreement with the repetitive elements, results based on mtDNA do not distinguish the two groups. AMOVA analyses do not show significant genetic differences between them (Coudray et al, 2009; Maca-Meyer et al, 2003), and methods like PCA or MDS based on haplogroup frequencies or genetic distances clearly do not form different clusters. However, it is interesting to notice that values of pairwise sequence differences are generally lower in Berber populations, which suggest that they are more subject to the effects of genetic drift, probably due to small sample sizes and isolation. This also explains the high genetic variance among geographically close Berber groups, and reflects their heterogeneity (Fadhlaoui-Zid et al, 2004), as well as the detection of some populations with clearly different mitochondrial gene pools that behave as population outliers.

Nonetheless, some discrepancies exist on whether genetic differentiation caused by cultural barriers between Arab and Berber groups is high enough to overcome genetic differentiation caused by an *isolation by distance* model. Evidences of cultural differences

being the main force that differentiates populations in North Africa are found in Cherni et al (2005), where geographically close Berber and Arab populations share only old mtDNA lineages, reflecting more likely a common ancestry rather than posterior gene flow. Also, Loueslati et al (2006) found that in the Tunisian Jerba island genetic differentiation between Arab and Berber speaking groups is similar to that found between European countries. On the contrary, other studies that compare eastern and western populations show evidences that there are more affinities between ethnic groups from the same region than across North Africa (Coudray et al, 2009; Ennafaa et al, 2009), in agreement with the *isolation by distance* model. Specifically, these studies find evidences that eastern populations have more influence from the Near East and Eastern Africa, whereas western North African populations are genetically closer to the Iberian Peninsula and western sub-Saharan Africa, and that this influence from neighboring regions is stronger that differences detected between geographically close Arab and Berber-speaking groups.

- *Haplogroups U6 and M1*

Another subject of investigation related to the origins of Berber people was the study of U6 and later on M1 clades. Rando and collaborators (1998) described U6 haplogroup on Moroccan Berber populations and called it the *Berber motif*. They hypothesized it was a sister haplogroup to U5, a haplogroup common in Europe and with a Near Eastern origin. Posterior studies based on complete mitochondrial DNA sequences confirmed the hypothesis and determined that U6 is the signature of a back-to-Africa migration wave that took place between 39,000 and 52,000 ya, during Paleolithic times (Maca-Meyer et al, 2001; Plaza et al, 2003), and it most likely has a Near Eastern origin rather than a European one (Maca-Meyer et al, 2003; Olivieri et al, 2006).

Table 1 Estimated ages for different subgroups of U6 haplogroup based on coding and HVSI regions (Maca-Meyer et al. 2003)

Subgroup	N	Age[24]	Age[45]	N	Age
U6b1	-			45	5,830 ± 4,551
U6b	4	8,568 ± 2,856	11,560 ± 3,853	62	24,411 ± 15,200
U6a	8	24,277 ± 4,831	32,754 ± 6,518	128	27,590 ± 13,576
U6a1	3	15,232 ± 5,677	20,552 ± 7,659	82	13,289 ± 5,470
U6c	2	5,712 ± 3,298	7,707 ± 4,450	16	17,658 ± 12,862
U6	14	25,297 ± 4,543	34,130 ± 6,130	206	66,222 ± 25,269

This study also states that this haplogroup has since then been in North Africa, and coalescence age estimates show it is the most ancient haplogroup in the region. Studies in North West Africa showed a patched distribution of this haplogroup (Plaza et al, 2003), though this could be explained by the high genetic drift of Berber populations. Haplogroup U6 is also found in Iberian populations, though at very low frequencies (less than 5%). Plaza et al (2003) compared U6 frequency differences on both shores of the Mediterranean and estimated the contribution of North Africa into the Iberian genetic background to be 18%. Maca-Meyer et al (2003) also hypothesize a northward expansion based on U6 phylogeography, though they suggest that not one but multiple migration waves must have taken place to explain the frequencies of the different U6 sub-haplogorups in Iberia.

Finally, a deeper investigation on a broader geographic area showed that interestingly, even if U6 has a western Eurasian origin, its frequency in North Africa decays eastward (Fadhlaoui-Zid et al, 2004). A more specific study on the main sub-haplogroups distribution, U6a and U6a1 lineages, showed that while U6a has the focus of expansion on North West Africa, U6a1 has its focus on North East Africa, though its overall frequencies in the region are lower when compared to U6a (Maca-Meyer et al, 2003). Thus U6a and U6a1 could be signals of different expansive demographic events that have taken place within North Africa.
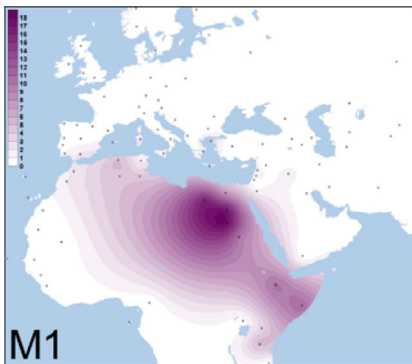
Figure 16 Map of M1 haplogroup frequencies (Olivieri et al. 2006)

The other haplogroup mostly found in North Africa, M1, displays its highest frequencies in Ethiopia, and therefore it was first hypothesized to have originated in East Africa (Quintana-Murci et al, 1999). However, posterior studies showed evidences that this haplogroup is, together with U6, the signal of a back-to-Africa event during the Paleolithic (Gonzalez et al, 2007; Olivieri et al, 2006). However, in an opposite way to U6, M1 shows a decay in frequency westwards (Coudray et al, 2009). Actually, frequencies in southern Egypt are similar to those found in Ethiopia (Stevanovitch et al, 2004). Unfortunately,

no extensive phylogeographic studies have been performed on M1 in North Africa, and thus the information available on this haplogroup is scarce.

c) Results from Y chromosome analysis

Studies based on Y chromosome data in North Africa are fewer than those in mtDNA, and most of them are centered in North West Africa and the Iberian Peninsula. Around 70% of modern societies are patrilocal (Burton et al, 1996; Murdock, 1967) (i.e. women tend to move to their husbands' location). As a consequence, it will be more likely that men live closer to their birthplace than women, and therefore it is expected that Y chromosome analysis will show more structure than mitochondrial DNA. In addition to this, Y chromosome mutational processes offer a high phylogeographic resolution (Jobling & Tyler-Smith, 2003), which together with patrilocality ensure that this marker is suitable for studying human migrations. In agreement to this, studies focusing in North Africa have centered in dating human origins in the region and disentangling past migration events with its neighboring regions, with special attention to the Iberian Peninsula. As expected genetic structure is found in the North African male lineage, results showing that around 75% of Y chromosome sequences belong to haplogroup E3b*-M35 and its derived clades E3b1-M78 and E3b2-M81, the last one being the more frequent. This genetic background seems to have originated independently to that from Europe, in opposition to mitochondrial DNA, where 60% of the North African lineages have a Eurasian origin.

- *Origins of Y chromosome in North Africa*

The worldwide frequency distribution of haplogroups in the E3b branch shows that the ancestral haplogroup E3b-M35 reach highest frequencies in sub-Saharan Africa, whereas E3b1-M78 is more frequent in Eastern Africa and E3b2-M81 in North Africa. The overall picture supports that the ancestral lineage, E3b-M35, originated somewhere in sub-Saharan Africa and was subsequently introduced from eastern Africa into North Africa. A recent phylogeographic study of the E3b branch, accounting for biallelic and STR variation across populations, suggests that the derivate lineage E3b1-M78 originated in eastern North Africa between

18,000 and 6,000 ya (Cruciani et al, 2007). These dates are in agreement with another study from Semino et al (2004). On the other hand, the same authors consider that the other derivate lineage E3b2-M81, the commonest in North Africa, has a much more recent origin than its sister clade. This is in agreement with TMRCA estimates of these haplogroups in North Africa, which coincide with the Neolithic expansion (Arredi et al, 2004). Also, patterns of haplotype and STR diversity within E3b2-M81 haplogroup support an east to west migration, which is in agreement with the Neolithic expansion. In contrast, other studies support an Upper Paleolithic expansion, based on TMRCA using a different dating method (Bosch et al, 2001).
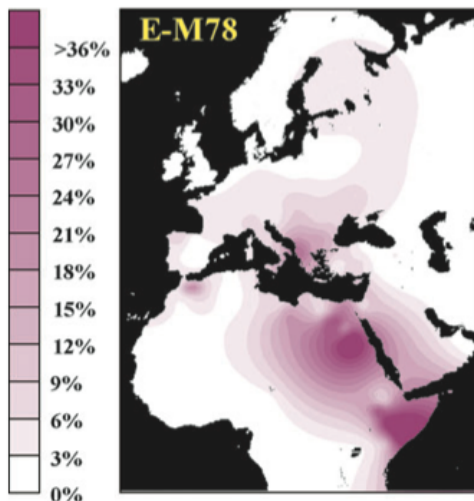
Figure 17 Map of E3b1-M78 haplogroup frequencies (Cruciani et al. 2007)

On the other hand, all studies coincide in that haplogroup J*-12f2, present at moderate frequencies in North Africa, has most likely a Middle Eastern origin and entered North Africa during a Neolithic expansion. The presence of J2-M127 in Europe and its also likely Middle Eastern origin, have been explained by a parallel Neolithic expansion on both shores of the Mediterranean (Bosch et al, 2001), though a more deep analysis of J2 sub-branches seems to tell a more complex story of European colonization, involving several expansions (Semino et al, 2004). Overall, a strong geographical structure has been detected in North Africa, and correlation analysis between genetic and geographic distance have shown to be strong and significant (Arredi et al, 2004).

- *Gene flow between North Africa and neighboring regions*

All studies carried out in North Africa agree in that the extant genetic background in the region originated independently from that in the Iberian Peninsula and sub-Saharan Africa, given that Y

chromosome sequences come from totally different super-haplogroup branches (Arredi et al, 2004; Bosch et al, 2001). Nonetheless, shared haplogroups at low frequencies exist, suggesting limited gene flow. An example of it is the presence of haplogroup E3b2-M81 at low frequencies in the Iberian Peninsula, whose overall contribution is estimated to be around 7%, showing a south-to-north clinal pattern (Bosch et al, 2001; Cruciani et al, 2004). In a similar way, sub-haplogroups of E3b1-M78 are also found at low frequencies in Europe in coastal Mediterranean populations, but the lack of substructure does not allow determining a time frame of gene flow from North Africa towards Europe.
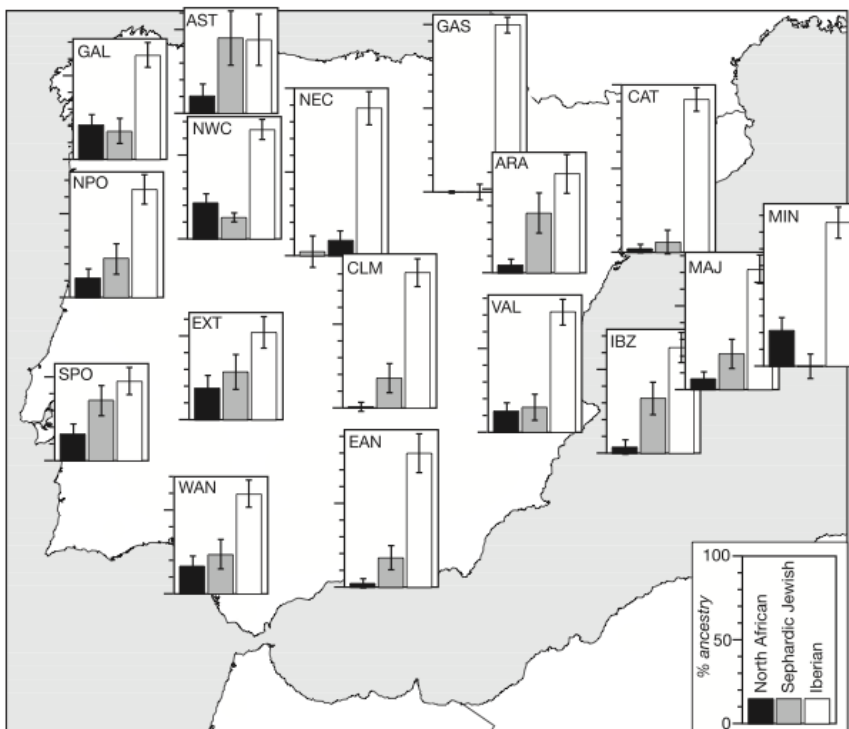
Figure 18 Admixture proportions in the Iberian Peninsula. In black are shown the North African admixture proportions based on mY estimator based on Moroccan parental populations. Error bars indicate standard deviations, and three letter codes indicate populations. GAL Galicia; AST Asturias; GAS Gascony; NPO North Portugal; NWC North West Castile; NEC North East Castile; ARA Aragon; CAT Catalonia; SPO South Portugal; EXT Extremadura; CLM Castilla La Mancha; VAL Valencia, WAN West Andalusia; EAN East Andalusia; IBZ Ibiza; MAJ Majorca; MIN Minorca. (Adams et al. 2008)

However, it is unlikely that these haplogroups entered Europe through the Levant, given the sub-haplogroup distribution (Cruciani et al, 2004). In a wider frame, it is estimated that the overall

contribution of North African lineages in Iberia is around 10% (Adams et al, 2008a). Also, the presence of haplogroups from the branch R1* in western North Afirca, support the idea that gene flow from Iberia towards North Africa has also taken place, though the time in which this migration took place remains uncertain due to the lack of substructure (Arredi et al, 2004; Bosch et al, 2001).

Some discrepancies exist in ascertaining the period during which this gene flow took place. Patterns of STR diversity within haplogroups support a very recent migration event (Arredi et al, 2004; Semino et al, 2004), being the most likely period the Islamisation of the Iberian Peninsula (Adams et al, 2008a). Related to this, it must be noticed that the homogeneity observed in North African Y chromosomes also include Arab and Berber-speaking communities, which in this case support the idea that, in opposition to what happened in Iberia, the Islamisation in North Africa was a cultural rather than a demographic process (Arredi et al, 2004); Bosch et al (2001).

Regarding sub-Saharan Africa, levels of gene flow are evident between North Eastern Africa and Eastern Africa, where the Nile River Valley would have acted as a genetic corridor (Cruciani et al, 2007). Authors claim that if the lineage E3b1-M78 originated in North East Africa and is also found in Eastern Africa, it must be consequence of a back-migration process. This is in agreement with findings concerning mitochondrial DNA from Krings et al (1999) and Olivieri et al (2006). Unfortunately, the period in which North African lineages entered Eastern Africa cannot be established with precision due to the STRs nature, and it ranges from the Upper Paleolithic and the Neolithic. Interestingly, evidences of gene flow also exist on the opposite direction, testified by the low frequencies of haplogroup E3a*, present in Africa and the Near East (Luis et al, 2004).

1.2.3 Results from genome wide autosomal data

In 2005 the Phase I of the International HapMap Project was completed, and results were published (The International HapMap Consortium, (2005)). The main goal of the project was to offer a guide for the design and prioritization of SNP genotyping assays for

disease association studies. In this way, they discovered over a million of *tag* SNPs, i.e. common SNPs that are in high *linkage disequilibrium* (LD) with nearby SNPs and are therefore informative of the variation in the region (Sachidanandam et al, 2001). Autosomal biparentally inherited markers like SNPs offer information of both ancestries. Also, every tag SNP is an independent marker, so that the survey of hundreds of SNPs will offer the perspective of genetic variation under a selectively neutral model (it is not likely that natural selection acts on all the SNPs) and the effect of genetic drift will be much less than in uniparental markers (Jobling, 2012). Contemporary to HapMap, new genotyping techniques were developed, and soon several companies built cost- efficient chips or arrays that carried out the simultaneous genotyping of SNPs, present in the HapMap Project and in other SNP databases (Ragoussis, 2009).

The availability of these genotyping assays has facilitated the whole-genome association (WGA) studies, but has also contributed in the human population genetics field. It is now possible to investigate geographic structure at very high resolution, and it has been shown, for instance, that genetic mirrors geography at a very fine scale in Europe (Lao et al, 2008; Novembre et al, 2008). However, with the analysis of markers alone it is not possible to estimate the time in which this structure was generated. As a consequence, in the following years methods based on the analysis of LD decay across populations (Moorjani et al, 2011; Pool & Nielsen, 2009) or based on other characteristics (Gravel, 2012b; Pugach et al, 2011) have been developed to estimate the time of admixture events. It is clear, in summary, that the availability of hundreds of thousands of markers has changes the approach in studying population genetics. With autosomal data population structure and gene flow can be more accurately defined and admixture events can be more reliably estimated.

Nonetheless, dense genotype data entails a major problem. The great majority of SNP discovery projects have been carried mainly in European populations. This creates an important ascertainment bias, i.e. diversity in populations of European ancestry is greatly overrepresented when compared to populations with other ancestries (Jobling, 2012). Also, it must be noticed that the SNP discovery in the HapMap project (whose SNPs are present in many

array platforms) required that the markers have to be common, having a minor allele frequency greater than 5% (Clark et al, 2005; Nielsen et al, 2004). This will also affect analysis related with the site frequency spectrum, like some selection scans or simulations.

One of the first studies (Li et al, 2008) performed with dense genotype data used the Human Genome Diversity Panel (HGDP-CEPH) (Cann et al, 2002), and includes the North African Mozabites. The main goal of the study was to describe worldwide human relationships, but we will focus on results regarding the Mozabites. The unsupervised clustering algorithm analysis performed with *frappe* (Tang et al, 2005) shows that Mozabites have three different ancestral components.
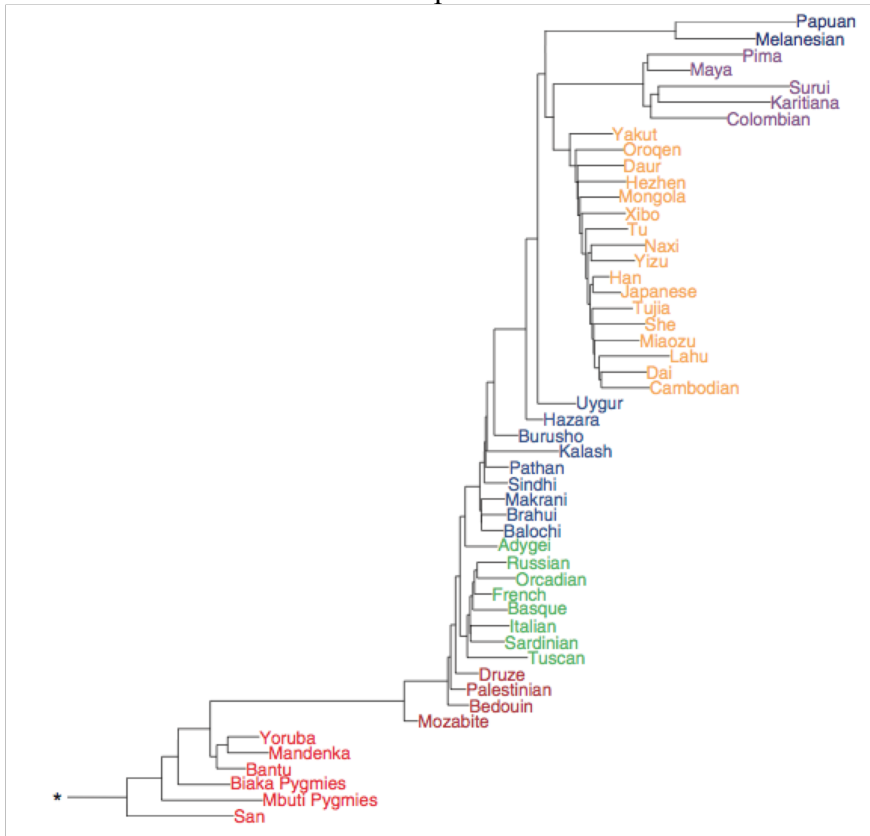
Figure 19 Maximum likelihood tree of 51 worldwide human populations. It can be noticed that Mozabites are an outgroup of Middle Eastern populations, more close to sub-Saharan popoulations. * indicates the root of the tree also where the chimpanzee branch is located (Li et al. 2008)

They share around 20% of their genome with sub-Saharan Africa populations, around 30% with European populations and the rest (50%) of their genomes has a common ancestry with Middle Eastern populations. This is probably why authors classify Mozabites as a Middle Eastern population. However, it must be noticed that in the population dendogram constructed by a maximum likelihood approach (**Figure 19**), they are located close to Middle Eastern populations but as an outlier, and in a PCA of the region, the first component separates the Mozabites from the rest of the Middle Eastern populations. In conclusion, the region most similar to North Africa is the Middle East, though it is not clear that both regions can be considered as a single cluster. Also it must be noticed that this study is based on a single North African population that in previous studies showed some degree of inbreeding compared to other populations from the region (Bosch et al, 2000).

2. OBJECTIVES

Objectives

The main goal of my Thesis is to obtain a global picture of the genetic structure of extant human North African populations and understand it in the framework of its neighboring regions: Europe, sub-Saharan Africa and the Near East. In order to do so, I had the following objectives:

1) Obtain a region-wide picture of the maternal genetic landscape in North Africa.

2) Determine when did the ancestors of modern North Africans arrive for the first time in the region.

3) Assess at which proportions this ancestry is present in present-day populations.

4) Establish the relationship between present-day North Africans and sub-Saharan African populations, and, if possible, estimate the time of gene flow to the region.

5) Determine the influence of Near Eastern populations in North Africa.

6) Investigate recent patterns of gene flow between North Africa and Europe with the ultimate goal of establishing the degree of interaction of the two regions.

3. RESULTS

3.1 Mitochondrial DNA structure in North Africa reveals a genetic discontinuity in the Nile Valley

Fadhlaoui-Zid K, Rodríguez-Botigué L, Naoui N, Benammar-Elgaaied A, Calafell F, Comas D. Mitochondrial DNA structure in North Africa reveals a genetic discontinuity in the Nile Valley. American Journal of Physical Anthropology 2011;145(1):107-17.

3.2 Genomic ancestry of North Africans supports a back-to-Africa migration

Henn BM, Botigué LR, Gravel S, Wang W, Brisbin A, Byrnes JK, Fadhlaoui-Zid K, Zalloua PA, Moreno-Estrada A, Bertranpetit J, Bustamante CD, Comas D. Genomic ancestry of North Africans supports back-to-Africa migrations. PLoS Genetics 2012;8(1):e1002397.

3.3 Gene flow from North Africa contributes to differential genetic diversity in Southern Europe

Title: Gene flow from North Africa contributes to differential genetic diversity in Southern Europe

Authors: Laura R. Botigué[1*], Brenna M. Henn[2*], Simon Gravel[2], Erik Corona[3,4], Christopher R. Gignoux[5], Gil Atzmon[6,7], Edward Burns[6], Harry Ostrer[7,8], Carlos Flores[9,10], Jaume Bertranpetit[1], David Comas[1¤], Carlos D. Bustamante[2¤]

Affiliation:

[1] Institut de Biologia Evolutiva (CSIC-UPF), Departament de Ciències Experimental i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain.

[2] Department of Genetics, Stanford University, Stanford, CA, USA.

[3] Lucile Packard Children's Hospital, Stanford CA, USA.

[4] Department of Pediatrics, Stanford University, Stanford, CA, USA.

[5] University of California San Francisco, San Francisco, CA, USA.

[6] Department of Medicine, Albert Einstein College of Medicine, Bronx, NY 10461, USA.

[7] Department of Genetics, Albert Einstein College of Medicine, Bronx, NY 10461, USA.

[8] Department of Pathology, Albert Einstein College of Medicine, Bronx, NY 10461, USA.

[9] Research Unit, Hospital Universitario N.S. de Candelaria, Santa Cruz de Tenerife, Spain

[10] CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain

* Both authors contributed equally

¤ Co-corresponding authors and joint supervisors

Word Count: 3,500

Running Title: Gene flow from North Africa to Europe

Abstract:

Human genetic diversity in southern Europe is higher than in other regions of the continent. This difference has been attributed to post-glacial expansions, the demic diffusion of agriculture from the Near East and levels of gene flow from Africa. Using single nucleotide polymorphism (SNP) data from 2,099 individuals in 43 populations, we show that estimates of recent shared ancestry between Europe and Africa are substantially increased (up to 20%, with fine-scale geographic differences) when gene flow from North Africans, rather than Sub-Saharan Africans, is considered. The gradient of North African ancestry accounts for previous observations of low levels of sharing with Sub-Saharan Africa and is independent of recent gene flow from the Near East. The source of genetic diversity in southern Europe has important biomedical implications; we find that most disease risk alleles from genome-wide association studies follow expected patterns of divergence between Europe and North Africa, with the principal exception of multiple sclerosis.

Introduction:

Clinal gradients of human genetic diversity and genetic markers in Europe have been attributed to directional migration patterns, climate, natural selection and isolation by distance models (Lao et al, 2008; Novembre et al, 2008; Novembre & Stephens, 2008; Pickrell et al, 2009). In particular, southern European populations exhibit the highest levels of genetic diversity, which declines in northern latitudes. Three main hypotheses have been proposed to explain this phenomenon. Under the first hypothesis, populations retreated to glacial refugia in southern Europe about 20,000 years ago (ya), but when these populations later re-colonized the continent, only a subset of the genetic diversity was carried into northern regions (Forster, 2004). The second hypothesis is that gene flow from the Near East, associated with the demic diffusion of agriculture, differentially affected geographic regions and in particular introduced additional genetic diversity to southeastern Europe (Cavalli-Sforza et al, 1994; Currat & Excoffier, 2005). The third hypothesis suggests that increased genetic diversity is the result of migrations from the African continent into southern Europe (Auton et al, 2009; Moorjani et al, 2011). These hypotheses are not mutually exclusive; however, we focus on testing a

hypothesis of gene flow from Africa to Europe, which has received the least amount of attention and may be the easiest to detect due to the recent timeframe of the proposed demographic event.

About 20,000 ya during the Last Glacial Maximum, populations in Europe retreated into the glacial refugia located in the Mediterranean peninsulas, where climate conditions were milder. Differences in genetic diversity in European extant populations have been explained by a continental re-colonization from these glacial refugia at the end of the glacial period, a process during which only a subset of the genetic diversity from the refugia would expand to the rest of the region. For instance, radiocarbon dates suggest that re-colonization in Britain took place around 14,700 ya (Jacobi & Higham, 2009). The geographic distribution and ages of mtDNA haplogroup HV0, V, H1 and H3 in European populations reflect that pattern of postglacial human recolonization from the Franco-Cantabrian refugia (Achilli et al, 2004b; Pereira et al, 2005; Torroni et al, 2001a), and a similar pattern has also been detected in Y chromosome as the case of haplogroup I (Rootsi et al, 2004). Differential gradients of genetic diversity in many other species within Europe (e.g. grasshoppers, brown bears and oak trees) have also been attributed to post-glacial expansions during this time (Hewitt, 1999).

Changes in genetic diversity in European populations have also been associated to the Neolithic expansion from the Near East (Cavalli-Sforza et al, 1994). The relative effect of demic diffusion of early agriculture on the genetic composition of European populations remains a hotly contested topic (Bramanti et al, 2009; Gignoux et al, 2011; Haak et al, 2005). It has been suggested that Near Eastern Neolithic mtDNA lineages comprise almost one quarter of the extant European haplogroups (Richards et al, 2000b) and Y chromosome genetic diversity also retains a strong signal from the Near East (Rosser et al, 2000). Beginning about 8,000 years ago, extensive archaeological data document the spread of the Neolithic across southern Europe; for example, at this time similar Neolithic pottery is found in both Europe and the Near East. A notable exception is southern Iberia: around 7,500 ya, strong similarities are found in pottery production in this region and North West Africa. Additionally, the existence of "maritime pioneers" in the Mediterranean Sea during this period has been hypothesized (Zilhão, 2001). Some authors have interpreted this as a proof of the

existence of Neolithic networks joining the European and African shores of western Mediterranean Sea (Linstädter et al).

Lastly, three recent studies highlight the possibility of genetic exchange between Europe and North Africa. Moorjani *et al.* estimated that about 1-3% of recent Sub-Saharan African ancestry is present in multiple southern European populations, Cerezo *et al.* find evidence of older (11,000 ya) Sub-Saharan gene flow towards Europe based on mtDNA genomes, and Auton *et al.* found that short haplotypes were shared between the Yoruba Nigerians and southwestern Europeans. However, given the geographic barrier imposed by the Sahara Desert between North Africa and Sub-Saharan Africa, and the proximity of North Africa to Europe, it is plausible that gene flow from Africa to Europe actually originated in North Africa. North Africans are significantly genetically diverged from Sub-Saharan populations (Fadhlaoui-Zid et al, 2011; Henn et al, 2012), and hence previous studies may not have accurately estimated the proportion or range of admixture in Europe by using a Sub-Saharan sample as a source population. During historical times, the settlement of Phoenicians, Greeks and Romans along the North African coast is well established. Most recently, the Moorish Berber conquest in Iberia began in the 8th century C.E. and lasted for more than five hundred years; this conquest has been suggested as a potential source of gene flow from North Africa towards the Iberian Peninsula. The Y chromosome haplogroup E3b2-M81 distribution is in agreement with a recent North African gene flow at that period (Adams et al, 2008b).

Here we analyze recently published SNP data from 7 North African populations (Henn et al, 2012), together with data from 30 European populations (Henn et al, 2012; Nelson et al, 2008) (including new Affymetrix 6.0 data for 3 Spanish populations: Galician, Andalucian, Canary Islands), 2 European Jewish populations (Atzmon et al, 2010), 1 Near Eastern population (Hunter-Zinck et al, 2010) and HapMap3 Sub-Saharan African populations (**Supplementary Table 1**). Our objective was to quantify the extent and pattern of recent gene flow between European and African populations. We use both allele frequency and haplotype identity by descent (IBD) approaches to estimate North African ancestry proportions in European populations. In order to quantify the variance in ancestry in European populations and obtain bounds on the time since admixture, we use a

quantitative model for the decrease in ancestry variance with the time since admixture (Gravel, 2012a). Recent gene flow between populations can also be detected by long haplotypes shared IBD with high density SNP genotyping data (Browning & Browning, 2010; Gusev et al, 2012). We investigate regional patterns of haplotype sharing between North Africa, Sub-Saharan Africa, Near East and Europe in detail, and observe a significant latitudinal gradient of North African ancestry within Europe characterized by a dramatic difference between the Iberian Peninsula and the neighboring regions.

Results:

Estimating Gene Flow between Africa and Europe
Ancestry proportions:

To estimate allele-based sharing between Africans and Europeans, we applied an unsupervised clustering algorithm, *ADMIXTURE* (Alexander et al, 2009) to data from all populations. We assumed *k*=2-10 ancestral populations, performed 10 iterations for each *k* and examined the cross-validation performance of each *k* (**Supplementary Fig. 1,2**). After filtering for missing loci and minor allele frequency, a total of 243,000 SNPs remained. Importantly, we make no assumption in this analysis that any given source population is an un-admixed population; that is, the analysis is run unsupervised where the current population configurations are not known (so that any sub-Saharan African ancestry could be detected in both North Africans and Europeans). Estimates of admixture show little bias using an unsupervised approach when the ancestral populations are substantially diverged (e.g. at an F_{st} of 0.05) and using several hundred thousand markers previous work suggests that Europeans and North Africans are segregated by an F_{st}=0.06. Evidence At *k*=4, ancestry assignment differentiated between non-Jewish European populations (from now on referred to as "European*"*), European Jews, Sub-Saharan Africans, and a group formed by Near Eastern and North African populations. At *k*=5,6 components mainly assigned to North African populations and Tunisian Berbers, respectively, clearly appear. European populations sharing this North African ancestral component appear almost exclusively in southern Europe (**Fig. 1, Supplementary Fig 3**).

Southern European populations have a high proportion (5-35%) of joint Near Eastern | North African ancestry assigned at k=4. However, identification of distinct Near Eastern and North African ancestries in k=5,6 differentiates southeastern from southwestern Europe. Southwestern European populations average between 4-20% of their genomes assigned to North African ancestral cluster (**Supplementary Fig. 3**), while this value does not exceed 2% in southeastern European populations. Contrary to past observations, Sub-Saharan ancestry is detected at <1% in Europe, with the exception of the Canary Islands. In summary, when North African populations are included, allele frequency-based clustering indicates a better assignment to North African rather than Sub-Saharan ancestry, and estimates of African ancestry in European populations increase relative to previous studies (**Supplementary Fig. 3**).

An average amount of 10% of European ancestry is also detected in North African populations. However, it is virtually absent in many individuals from Tunisia and Western Sahara and ranges between 4 – 16% in the rest of North Africa, with notable intra-population variation. Levels of European ancestry detected in the Near East are low (3%), which suggests that direct gene flow from Europe to North Africa may be more likely than from the Near East.

Long identical by descent haplotypes:

Recent gene flow between populations results in long shared haplotypes; we investigated differences in African ancestry among European populations by analyzing these genomic segments inferred to be identical by descent between Sub-Saharan Africa, North Africa, Europe and the Near East (**Supplementary Methods**). Migration from one endogamous population to another generates genetic segments that share a recent common ancestor (and in short time spans are IBD) between the two populations; the distribution and length of IBD segments are informative of recent migration. We restrict our analysis to IBD segments greater than 1.5 cM identified using fastIBD (Browning & Browning, 2011). We first examined the extent to which detection of long IBD segments is conditioned on marker density. We compared IBD performance between a high-density dataset (HDD) of 641,884 SNPs with a low-density dataset (LDD) of 280,462 SNPs from our primary analysis. Results can be found in the **Supplementary Information**. Long IBD segments are far less sensitive to ascertainment bias than are

allele frequency estimates, and by analyzing segments >1.5 cM, we help minimize background linkage disequilibrium (problematic for the identification of short haplotypes) (Conrad et al, 2006). We calculated a summary statistic informative of the level of gene flow between two populations, though not the directionality: "W_{EA}" is the sum of lengths (in cM) of all DNA segments inferred to be shared identical by descent between a given European population "*E*" and North African | Sub-Saharan African populations "*A*", normalized by the possible number of pairwise comparisons between both populations (Atzmon et al, 2010). Also, it has been shown that extensive IBD sharing may be a signal of positive selection shared among populations (Albrechtsen et al, 2010). To confirm that the IBD geographic pattern was not due to natural selection, we removed all IBD segments extensively shared and recalculated W_{EA} statistic (**Supplementary Material**).

A gradient of shared IBD segments is observed from southern to northern Europe (based on W_{EA}, **Figure 3, Supplementary Table 3**). This gradient is highest in the Iberian Peninsula for both North Africa and Sub-Saharan African IBD segments (with the exception of the Basques who show similar levels of sharing to other European populations). Additionally, we emphasize that the amount of IBD sharing detected between Sub-Saharan Africa and Europe is almost one order of magnitude lower than that between North Africa and Europe. We regressed the North African-European IBD metric (W_{EA}) on the sine of latitude to evaluate the strength of this gradient, and find a significant relationship across southern-to-northern Europe, $p=5.5 \times 10^{-8}$ (**Figure 4**).

To pinpoint which specific North African regions exchanged migrants with Europe, we calculated W_{EA} between a given European population and each of the 7 North African populations (**Figure 5, Supplementary Table 3**). Southwestern European populations, and in particular the Canary Islands, show the highest levels of IBD sharing with northwestern African populations (i.e., the Maghreb: Morocco, Western Sahara, Algeria and Tunisia). While inferred IBD sharing does not indicate directionality, the North African samples that have highest IBD sharing with Iberian populations also tend to have the lowest proportion of the European cluster in *ADMIXTURE* (**Figure 1**), e.g. Saharawi, Tunisian Berbers and South Moroccans. For example, the Canary Islanders share

more IBD segments with the Tunisians and Western Saharans (**Figure 5**), who present extremely minimal levels of European ancestry in contrast to populations like North Moroccans or Algerians who have greater estimated European ancestry. This suggests that gene flow occurred from Africa to Europe rather than the other way around.

We next aimed to rule out a model where observed sharing between Europe and North Africa was the result of recent gene flow from the Near East into both regions. We compared IBD between Qatari individuals (the best Near Eastern representatives genotyped with the Affymetrix platform currently available, **Supplementary Fig. 5**) and both Europe and North Africa. Southwestern Europe has more IBD segments shared with the Maghreb than Qatar. Interestingly, eastern Mediterranean populations share more segments IBD with the Near East than with North Africa, and the northern European populations show only limited IBD sharing with both North Africa and the Near East (**Figure 3c, 5**). The southwest to northeast gradient of North African IBD sharing (**Fig. 3a**) and the distinct peak in sharing between Iberia and the Maghreb (**Figure 5**) indicate that sharing in southwestern Europe is independent of gene flow from the Near East.

ADMIXTURE vs. IBD estimated proportions:

We next compared the frequency-based vs. haplotype sharing methods of estimating North African ancestry proportions. If gene flow has occurred in the distant past (i.e. more than 30 generations ago), ancestry detected by allele frequencies may exceed estimates of ancestry from recent long haplotypes IBD. **Supplementary Figures 4,6** show the correlations and proportions of sharing detected with ADMIXTURE for $k=$ 4 to 6 compared with estimates from fastIBD. The correlation between IBD and allele-based estimates of North African ancestry is highest at $k=6$ ($R^2=0.85$). Allele estimates of joint Near Eastern and North African from $k=4$ substantially over-estimate IBD proportions. Concordance between IBD and $k=6$ allele-based estimates is particularly clear for southwestern European populations (**Supplementary Figure 6**).

However, for Central and North European populations, the $k=6$ cluster estimates of North African ancestry tend to be 0%, while haplotype sharing estimates average about 2% of the European genome. This difference is possibly due to either false positive matching with short IBD tracts (although all were greater

than 1.5 cM) or remnants of Near Eastern ancestry that are also shared with North Africa. The latter possibility is supported by the better concordance between the joint *k*=4 Near Eastern | North African ancestry with Central and North European IBD estimates (**Supplementary Figure 6**).

Implications of Gene Flow from North Africa to Europe
Time since admixture estimates:

The variance in ancestry assignments for individuals *within* a population depends on the total ancestry proportions, the timing and duration of gene flow, population structure and/or assortative mating within the population, and errors in assignment. We used variances in ancestry proportions across populations estimated with ADMIXTURE to infer effective admixture times, i.e. the times required to achieve the observed variance in the population given a single gene flow event in a randomly mating population (see model from (Gravel, 2012a)). Focusing on the North African component at *k*=6, we found that a migration event from North Africa to Europe would have occurred at least 8-10 generations ago (approximately 240-300ya) in Spain, and at least 6-7 generations ago in France and Italy (**Figure 2**). Since population structure, continuous gene flow, assortative mating, and errors in assignments may considerably increase the variance (and thus reduce the effective migration time), we consider these time estimates to be lower bounds: under all the proposed variance-increasing scenarios, there must be a substantial proportion of migration that has occurred before the effective migration time, possibly much earlier.

Disease Risks:

We asked whether the migrations between North Africa and Europe affected the pattern of alleles associated with disease risk in these regions. By drawing on a database of GWAS risk alleles, we determined the cumulative risk for 134 diseases in each European and African population for which we had high-density SNP data (**Methods**).

We studied the deviations from random drift for all diseases with a False Discovery Rate (FDR) < 0.05. Pairwise q-values controlling for the FDR of all possible population comparisons within each disease (not across all diseases) were also calculated. The vast majority of disease alleles reflect expected patterns of neutral divergence (assessed with F_{st}) among populations. Interestingly, we

found that the multiple sclerosis (MS) risk calculated from 53 independent loci under a multiplicative risk model displayed a significant deviation from random drift for several North African populations. Maghrebi populations (e.g. Moroccans, **Supplementary Figure 11**) had a significantly elevated risk for MS, while the Canary Islanders, the population with highest inferred North African ancestry, had a significantly decreased risk for MS. While MS prevalence is thought to increase along south-to-north latitudinal gradients in the northern hemisphere, prevalence data for North Africans are extremely limited (Rosati, 2001). Our results suggest that North African Maghrebi have a greater *genetic risk* than expected under a neutral model, although presentation of MS could be attenuated by environmental variables such as UV exposure (Handel et al, 2010). Based on our model, we would predict individuals with high North African ancestry living in Europe to have a higher genetic risk for MS (see supporting evidence for North African immigrants in France in (Kurtzke et al)).

Discussion:

Using genome-wide SNP data from over 2,000 individuals, we characterize broad clinal patterns of recent gene flow between Europe and Africa that have a perceptible effect on genetic diversity of European populations. We have shown that recent North African ancestry is highest in southwestern Europe and decreases in northern latitudes, with a sharp difference between the Iberian Peninsula and France, where Basques show low North African influence (as suggested in (Martinez-Cruz et al, 2012)). Our estimates of shared ancestry are much higher than previously reported (up to 20% of the European individuals' genomes). This increase in inferred African ancestry in Europe is due to our inclusion of 7 North African, rather than Sub-Saharan African populations. Specifically, elevated shared African ancestry in Iberia and the Canary Islands can be traced to populations in the North African Maghreb like Morocco, Western Sahara and the Tunisian Berbers. Our results, based on both allele-frequencies and long shared haplotypes, support the hypothesis that recent migrations from North Africa contributed substantially to the higher genetic diversity in southwestern Europe. Previous Y-chromosome data have highlighted examples of male-biased gene flow from Africa to Europe, such as the eastern African slave ancestry in Yorkshire,

England (King et al, 2007) and the legacy of Moors in Iberia (Adams et al, 2008a). Here we show that gene flow from Africa to Europe is not merely reflected on the Y-chromosome but corresponds to a much broader effect.

Alternative models of gene flow: Migration(s) from the Near East likely have had an effect on genetic diversity between southern and northern Europe (discussed below), but do not appear to explain the gradients of African ancestry in Europe. A model of gene flow from the Near East into both Europe and North Africa, such as a strong demic wave during the Neolithic, could result in shared haplotypes between Europe and North Africa. However, we observe haplotype sharing between Europe and the Near East follows a southeast to southwest gradient, while sharing between Europe and the Maghreb follows the opposite pattern (**Figure 3**); this suggests that gene flow from the Near East cannot account for the sharing with North Africa.

We do detect low levels of IBD and allele sharing between the Near East and the majority of the European continent. Both IBD and allele sharing with the Near East appear elevated in southeastern Europe (e.g. Italy, Yugoslavia, Cyprus), although estimates of Near Eastern ancestry in northern Europe tend to decrease at higher level ancestral k's (**Supplementary Figure S1**). It is possible that these patterns reflect more ancient migrations, perhaps dating back to the Neolithic, which resulted in a low level of short Near Eastern haplotypes across much of Europe. This hypothesis is further supported by results of the time since admixture estimate based variances in ancestry proportions (**Figure 2**), which suggest that Near Eastern ancestry is older than the North African one and therefore they did not enter Europe on the same migration wave.

Another possible hypothesis to explain the increased diversity in southern Europe could have been that an influx of Jewish ancestry had a heterogeneous effect on genetic diversity in Europe. However, in most European populations here, virtually no Jewish ancestry was detected. On average, 1% of Jewish ancestry is found in Tuscan HapMap population and Italian Swiss, as well as Greeks and Cypriots. This may reflect the higher sharing with Near Eastern populations in the Italian peninsula and southeastern Europe (**Figure 3c**) or low levels of gene flow with the early Italian Jewish communities (Atzmon et al, 2010).

Disease Risk Implications:

The observation that the majority of disease risk alleles in this study follow an expected pattern of neutral drift among populations is consistent with the interpretation that these common alleles are not strongly affected by natural selection. We note that alleles identified in GWASs of individuals of largely northern European descent have limited portability to neighboring populations because the tagged GWAS SNPs may no longer be in linkage disequilibrium with the causative variant. Thus, estimates of genetic risk for these diseases in North Africans are likely inaccurate because North African specific risk SNPs are missing. With these caveats, we note that one disease, multiple sclerosis does not conform to a pattern of neutral genetic drift and we cannot rule out a role for natural selection affecting the frequency of these risk variants. Our results show an increased genetic risk for multiple sclerosis in North African populations. West Saharans and North Moroccans carry higher frequencies of MS alleles that deviate from neutral expectations of divergence among European and African populations. However, the Canary Islands, while displaying the highest amount of North African ancestry, have the lowest predicted genetic risk for MS. The complexity of these results serves to emphasize the importance of conducting disease associations in many diverse populations (Bustamante et al, 2011). The significant gene flow from North Africa into southern Europe will result in a miscalculation of genetic disease risk in certain European populations, if North African specific risk variants are not taken into account.

Materials and Methods:

Data: Recently published and new single nucleotide polymorphism (SNP) data were used to build a database of 43 populations and 2,099 individuals. The database includes 7 North African populations (Henn et al, 2012), together with data from 27 European populations (Altshuler et al, 2010; Henn et al, 2012; Nelson et al, 2008), 2 European Jewish populations (Atzmon et al, 2010), 1 Near Eastern population (Hunter-Zinck et al, 2010) and HapMap3 Sub-Saharan African populations (**Supplementary Table 1**). Additionally, new data for 3 Spanish populations: Galician (NW Spain), Andalusian (S Spain), and the Canary Islands, was included in the database. Informed consent was obtained from all newly collected Spanish populations. Samples were genotyped on the Affymetrix 6.0 chip, and quality control filtered for missing loci and close relatives. Data from these new populations can be found at: *bhusers.upf.edu/dcomas/*

ADMIXTURE analysis: An unsupervised clustering algorithm ADMIXTURE 1.21 (Alexander et al, 2009) was used to determine allele-based sharing in a dataset of 243,000 markers formed by a total of 41 populations (29 Europeans, 2 Jewish Europeans, 1 Near Eastern, the Qatari, 7 North Africans and 2 Sub-Saharan Africans). For the sake of equal representation, a random subset of 15 individuals was chosen for any population having a much larger sample size. Ten ancestral clusters (*k*= 2 through 10) in total were tested successively, running 10 iterations for each ancestral cluster (**Supplementary Figure 1**) and calculating cross validation errors for every run (**Supplementary Figure 2**). Moreover, for *k*=4 through 6,200 bootstraps were performed by resampling subsets of each chromosome, so that standard errors for each ancestral cluster estimate could be obtained (Alexander et al, 2009).

Modeling migration timing: According to Gravel *et al*. (Gravel, 2012a) the relationship between variance in ancestry Var(X^p), time since admixture *T* and migration proportions *m* is simply:

$$Var\left(X^p\right) = \frac{m(1-m)}{2^{T-1}} + \frac{2m(1-m)\left(1-1/2N\right)^{T-1}}{2n+2(T-2)L}$$

where *L* is the total length of the

genome (in Morgans), *n* is the number of chromosomes (22), and *N* is the population size. Here we used *n*=22 and *L*=3,500 cM.

IBD detection: The analysis of IBD sharing was conducted using all the populations in the dataset (**Supplementary Table 1**) with the exclusion of the European Jewish populations. We note that in the *ADMIXTURE* analysis at K=3 there is shared ancestry between Europeans and Jewish populations, however, this could represent either shared ancestral variation or gene flow. Higher levels of K>3 showed very little recent Jewish ancestry in European populations. The removal of Jewish populations from the dataset increased the number of common markers from 243,000 to 274,000 and to a total of 41 populations. A preliminary test of IBD sharing was calculated with both GERMLINE (Gusev et al, 2009) and fastIBD (Browning & Browning, 2011). Results showed fastIBD as the more accurate method to detect IBD in our dataset (**Supplementary Methods**), so that further analyses were done using this algorithm.

Correction for sample size: In order to compare between the different statistics calculated from the IBD results, we correct for sample size, given that in European populations there are differences in sample size of two orders of magnitude. We follow Atzmon *et al*. (Atzmon et al, 2010) calculation of IBD sharing metrics. Suppose we want to calculate a parameter related to the IBD between *PopA* and *PopB*. Our environment will be a list of all the individuals belonging to *PopA* and *PopB*. The correction factor is the total number of possible pairs where one individual is from *PopA* and the other is from *PopB,* which will depend on the population sizes, *n* and *m*, respectively. This is:

$$W_{AB} = \frac{\sum_{a \in A} \sum_{b \in B} W^{ab}}{nm}$$

Standard deviation from W_{AB} statistic was obtained on the basis of the standard deviation in IBD sharing between *PopA* and *PopB*.

Risk allele frequencies: We asked whether the migrations between North Africa and Europe affected the pattern of alleles associated with disease risk in these regions. Using a database developed in Corona *et al*. (Corona et al, 2010) after manual curation from published literature, SNPs associated with a disease in a genome wide association study (GWAS) and having a p-value below 1×10^{-6}

with a reported risk allele were included in this analysis. Candidate gene studies were not included due to the large p-values that are reported and the resulting skepticism they cause. Only single SNP GWAS hits were included (as opposed to haplotype blocks associated with a disease). In cases where different disease risk alleles were reported for the same disease in different studies, the risk allele in the study with the largest sample size (disease + non-disease individuals) was used. Since many GWAS SNPs are in linkage disequilibrium (LD) with the actual causal SNP, we filtered SNPs with LD r^2 value greater than 0.2 to insure only one SNP per associated region was used. SNPs with the largest odds ratio were retained during local filtering as they are more likely to reflect the risk associated with the actual causal SNP. When the odds ratio was not reported, retention of SNPs with the largest sample size in the study were prioritized. Cumulative risk allele frequency results for each population and 134 diseases are plotted online at geneworld.stanford.edu/africa_hapmap. The cumulative risk allele frequency is number of risk alleles present in each population across all SNPs associated with the disease divided by the total number of alleles.

Figure Legends:

Figure 1 Legend: *Allele-based estimates of ancestry in Europe, North Africa, Sub-Saharan Africa, Jews and the Near East.* Unsupervised *ADMIXTURE* results for k=4:6. Cross-validation indicated k=4 as the best fit, but higher density datasets (Henn et al, 2012) and higher values of k continue to identify population-specific ancestries (**Supplementary Fig. 2**); we therefore conservatively focused on k=3:6 ancestral populations.

Figure 2 Legend: *Variance in ancestry proportions within populations depends on the overall ancestry proportions in the population and the timing of gene flow.* (**a**) Using the proportion of Near Eastern | North African ancestry inferred at k=4 with *ADMIXTURE*, we estimated the variance in ancestry within each of 11 European populations, indicated here by abbreviations. The grey lines show the expected relationship between ancestry proportions (X-axis) and variances (left Y-axis), under a single pulse model occurring at generation g (right Y-Axis). Departures from single-pulse models tend to increase the variance in ancestry, and so the corresponding effective times should be thought of as lower bounds: significant migration must have occurred before the effective times (see text). (**b**) Estimating the effective time of migration based on variance in North African ancestry proportions inferred under the k=6 model (**Figure 1**).

Figure 3 Legend: *Amount of genetic sharing (in cM) between*

Europe and the Africa shows that the highest sharing is with the Iberian Peninsula. **(a)** Genetic sharing represented as a density map with Sub-Saharan ancestry, **(b)** North African ancestry and **(c)** Near Eastern ancestry. The Canary Islands are shown in the top left.

Figure 4 Legend: *Significance of latitudinal gradient of IBD sharing within Europe.* In order to determine if there was a significant relationship between latitude and mean IBD count (W_{EA}) within Europe, we regressed W_{EA} on *log(sin(latitude))*. The sine of the latitude was used to obtain distance-appropriate vertical values; then we log-transformed these value to obtain the expected decay of allele sharing in 2-dimensional habitats (Rousset, 1997). The p-value of the regression for IBD shared between North Africa and Europe is 5.5×10^{-8} (the p-value for sharing between Europe and Sub-Saharan Africa is 8.7×10^{-9}, data not shown).

Figure 5 Legend: *Population-specific estimates of haplotype sharing (cM) between North Africa and Europe.* Estimates of W_{EA} between each European population (labeled on the X-axis) and each of the 7 North African populations and the Qatari are represented by colors and symbols. A substantial increase in haplotype sharing is detected between southwestern European populations and Maghrebi populations (i.e. Morocco, Western Sahara and Tunisia) in comparison to the remainder of the European continent. The excess of sharing between the Near East and southern central and Eastern Europe is also noteworthy.
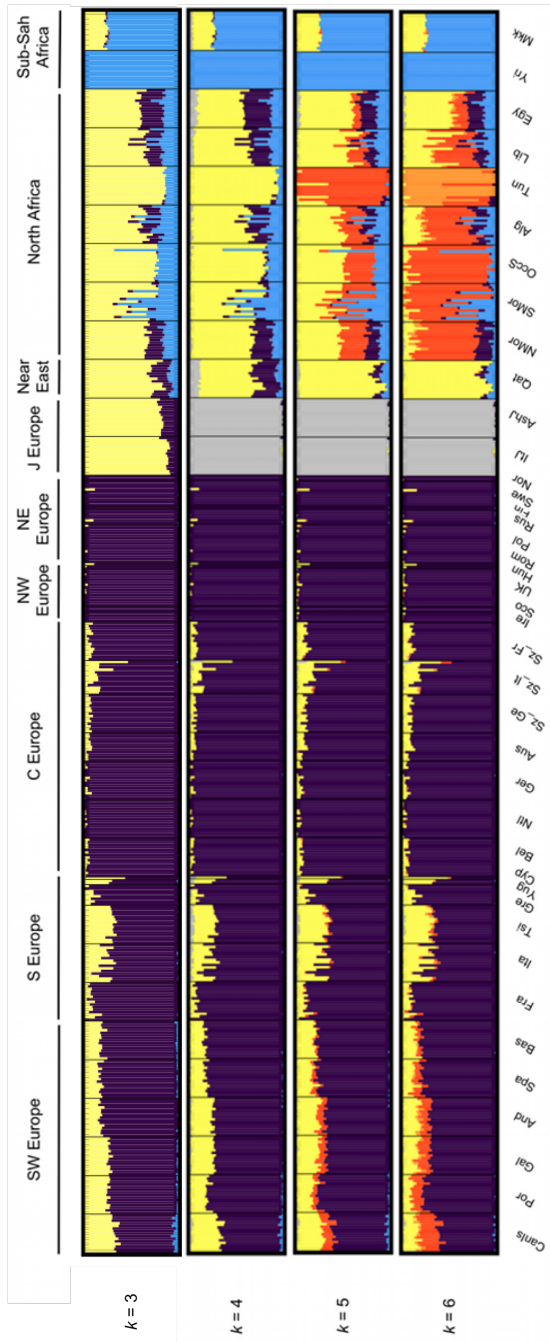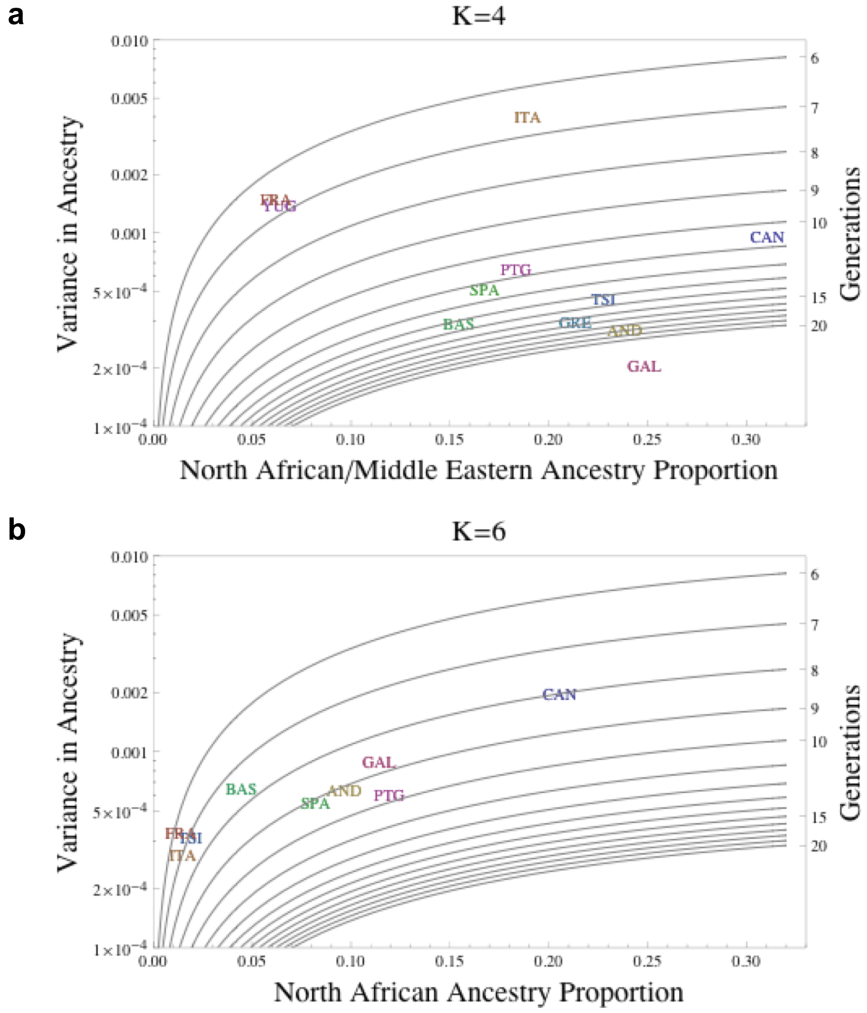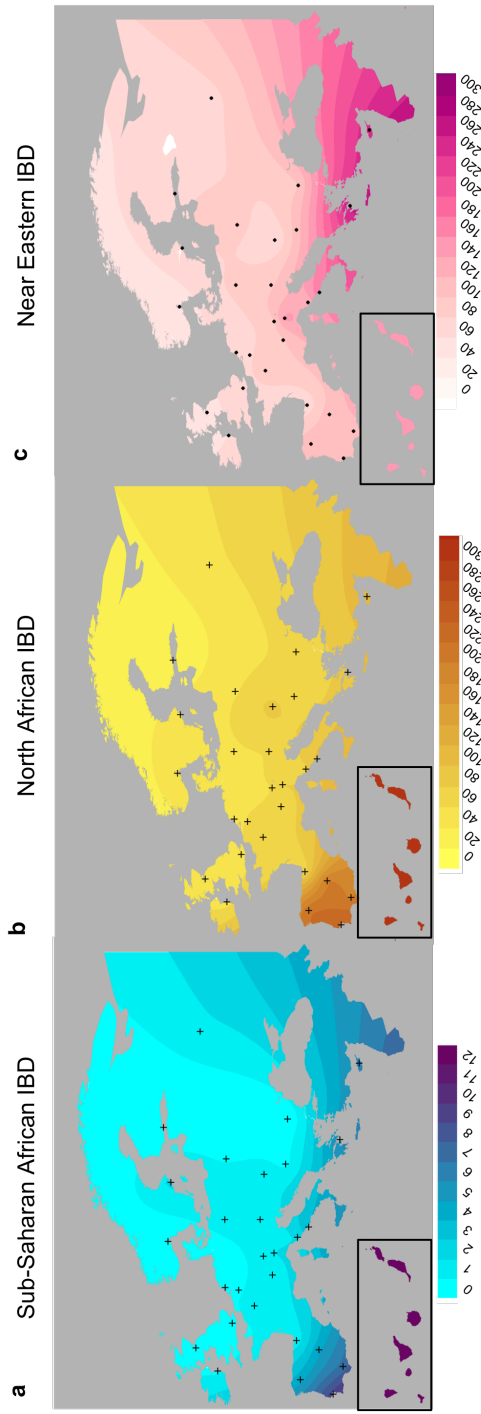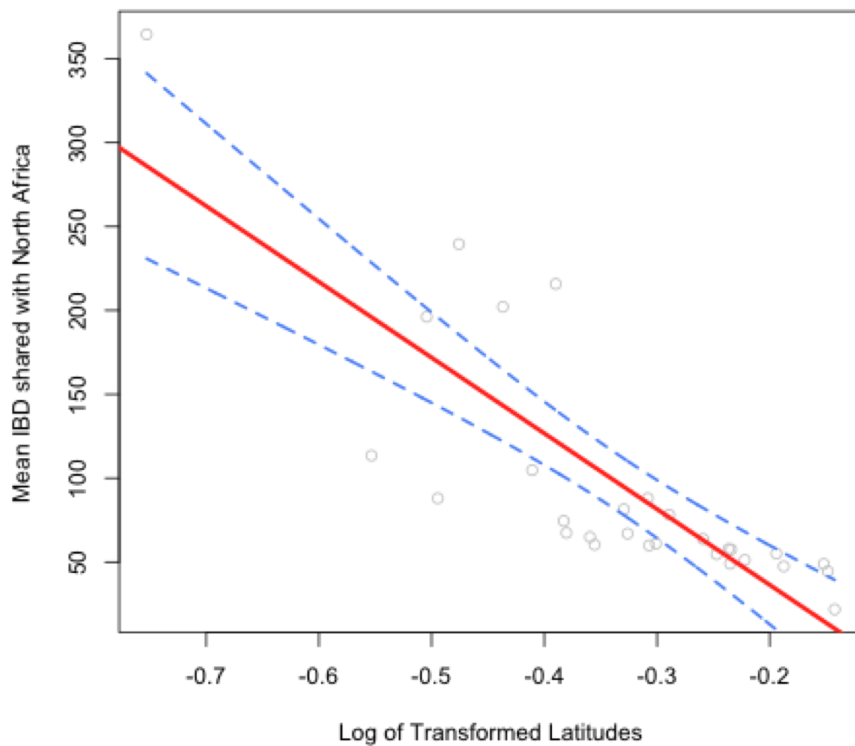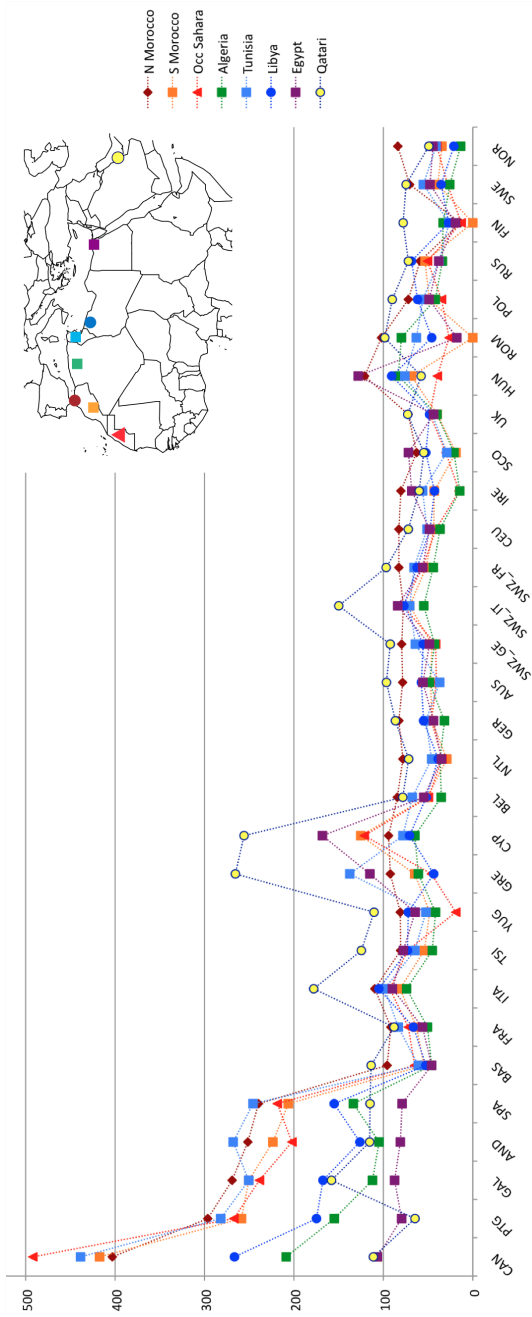
Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

83

References:

1.   Novembre J & Stephens M (2008) Interpreting principal component analyses of spatial population genetic variation. *Nat Genet* 40(5):646-649.
2.   Pickrell JK*, et al.* (2009) Signals of recent positive selection in a worldwide sample of human populations. *Genome Res* 19(5):826-837.
3.   Lao O*, et al.* (2008) Correlation between genetic and geographic structure in Europe. *Curr Biol* 18(16):1241-1248.
4.   Novembre J*, et al.* (2008) Genes mirror geography within Europe. *Nature* 456(7218):98-101.
5.   Forster P (2004) Ice Ages and the mitochondrial DNA chronology of human dispersals: a review. *Philos Trans R Soc London [Biol]* 359(1442):255-264.
6.   Currat M & Excoffier L (2005) The effect of the Neolithic expansion on European molecular diversity. *Proc Biol Sci* 272(1564):679-688.
7.   Cavalli-Sforza LL, Menozzi P, & Piazza A (1994) *The history and geography of human genes* (Princeton University Press).
8.   Auton A*, et al.* (2009) Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Res* 19(5):795-803.
9.   Moorjani P*, et al.* (2011) The History of African Gene Flow into Southern Europeans, Levantines, and Jews. *PLoS Genet* 7(4):e1001373.
10.   Jacobi RM & Higham TFG (2009) The early Lateglacial re-colonization of Britain: new radiocarbon evidence from Gough's Cave, southwest England. *Quaternary Science Reviews* 28(19–20):1895-1913.
11.   Torroni A*, et al.* (2001) A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* 69(4):844-852.
12.   Achilli A*, et al.* (2004) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet* 75(5):910-918.

13. Pereira L, *et al.* (2005) High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res* 15(1):19-24.

14. Rootsi S, *et al.* (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in europe. *Am J Hum Genet* 75(1):128-137.

15. Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* 68(1-2):87-112.

16. Haak W, *et al.* (2005) Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science (New York, N.Y.)* 310(5750):1016-1018.

17. Bramanti B, *et al.* (2009) Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science (New York, N.Y.)* 326(5949):137-140.

18. Gignoux CR, Henn BM, & Mountain JL (2011) Rapid, global demographic expansions after the origins of agriculture. *Proc Natl Acad Sci U S A* 108(15):6044-6049.

19. Richards M, *et al.* (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 67(5):1251-1276.

20. Rosser ZH, *et al.* (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* 67(6):1526-1543.

21. Zilhão J (2001) Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. *Proceedings of the National Academy of Sciences* 98(24):14180-14185.

22. Linstädter J, Medved I, Solich M, & Weniger G-C (Neolithisation process within the Alboran territory: Models and possible African impact. *Quaternary International* (0).

23. Cerezo M, *et al.* (2012) Reconstructing ancient mitochondrial DNA links between Africa and Europe. *Genome Res* 27:27.

24. Fadhlaoui-Zid K, *et al.* (2011) Mitochondrial DNA structure in North Africa reveals a genetic discontinuity in the Nile Valley. *Am J Phys Anthropol* 145(1):107-117.

25. Henn BM, *et al.* (2012) Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. *PLoS Genet* 8(1):12.
26. Adams SM, *et al.* (2008) The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *Am J Hum Genet* 83(6):725-736.
27. Nelson MR, *et al.* (2008) The Population Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics research. *Am J Hum Genet* 83(3):347-358.
28. Atzmon G, *et al.* (2010) Abraham's children in the genome era: major Jewish diaspora populations comprise distinct genetic clusters with shared Middle Eastern Ancestry. *Am J Hum Genet* 86(6):850-859.
29. Hunter-Zinck H, *et al.* (2010) Population genetic structure of the people of Qatar. *Am J Hum Genet* 87(1):17-25.
30. Gravel S (2012) Population Genetics Models of Local Ancestry. *Genetics* 4:4.
31. Browning SR & Browning BL (2010) High-resolution detection of identity by descent in unrelated individuals. *Am J Hum Genet* 86(4):526-539.
32. Gusev A, *et al.* (2012) The Architecture of Long-Range Haplotypes Shared within and across Populations. *Mol Biol Evol* 29(2):473-486.
33. Alexander DH, Novembre J, & Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655-1664.
34. Browning BL & Browning SR (2011) A fast, powerful method for detecting identity by descent. *Am J Hum Genet* 88(2):173-182.
35. Conrad DF, *et al.* (2006) A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet* 38(11):1251-1260.
36. Albrechtsen A, Moltke I, & Nielsen R (2010) Natural selection and the distribution of identity-by-descent in the human genome. *Genetics* 186(1):295-308.
37. Rosati G (2001) The prevalence of multiple sclerosis in the world: an update. *Neurol Sci* 22(2):117-139.

38. Handel AE, Giovannoni G, Ebers GC, & Ramagopalan SV (2010) Environmental factors and their timing in adult-onset multiple sclerosis. *Nat Rev Neurol* 6(3):156-166.
39. Kurtzke JF, Delasnerie-Laupretre N, & Wallin MT (1998) Multiple sclerosis in North African migrants to France. *Acta Neurol Scand* 98(5):302-309.
40. Martinez-Cruz B, *et al.* (2012) Evidence of Pre-Roman Tribal Genetic Structure in Basques from Uniparentally Inherited Markers. *Mol Biol Evol* 6:6.
41. King TE, *et al.* (2007) Africans in Yorkshire? The deepest-rooting clade of the Y phylogeny within an English genealogy. *Eur J Hum Genet* 15(3):288-293.
42. Adams SM, *et al.* (2008) The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *Am J Hum Genet* 83(6):725-736.
43. Bustamante CD, Burchard EG, & De la Vega FM (2011) Genomics for the world. *Nature* 475(7355):163-165.
44. Altshuler DM, *et al.* (2010) Integrating common and rare genetic variation in diverse human populations. *Nature* 467(7311):52-58.
45. Gusev A, *et al.* (2009) Whole population, genome-wide mapping of hidden relatedness. *Genome Res* 19(2):318-326.
46. Corona E, Dudley JT, & Butte AJ (2010) Extreme evolutionary disparities seen in positive selection across seven complex diseases. *PLoS One* 5(8).
47. Gravel S (2011).
48. Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145(4):1219-1228.

4. DISCUSSION

4.1 Exploring North Africa with dense genotype data

The opportunity of investigating the region of North Africa with genotype array data was promising because long-standing questions like the origins of extant inhabitants or gene flow to and from other regions could be now approached by the analysis of hundreds of thousands of markers. One of the reasons is that, the overall picture that this markers show is less biased than when using other markers, because stochastic effects are buffered in the *ensemble* of data.

On the other hand, the analysis of genomic data allows the genetic structure of human populations to be analyzed beyond the continental scale, revealing patterns of structure even in small regional areas (Gayan et al, 2010; Hunter-Zinck et al, 2010; Nelis et al, 2009; O'Dushlaine et al, 2010) (**Figure 20**).
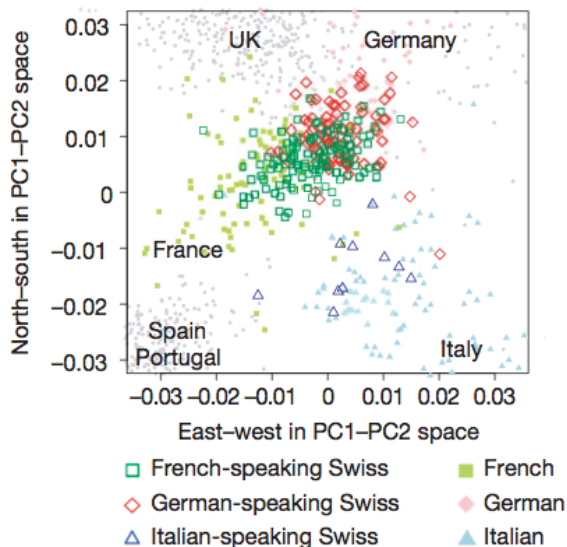
Figure 20 Magnification of a Principal Component Analysis of European populations that highlights the region around Switzerland. Swiss individuals are colored as a function of their language. (Novembre et al. 2008)

As a consequence, populations with a complex demographic history are now studied with unprecedented accuracy, and the advent of past demographic events can be investigated at a deeper detail. In other words, proportions of gene flow and the original location of a

given migrant population can now be addressed more specifically. Now it is possible to establish relationships between specific populations rather than between regions (Bryc et al, 2010), which provides a better understanding of human genetic diversity and human evolution, whereas previously, with the analysis of few markers it was only possible to establish gene flow between broad regions, and the directionality, the time and the place of origin of this gene flow was very difficult to estimate if not impossible.

As a consequence, in the last years new algorithms that estimate proportions of ancestry at an individual level have been developed (Alexander et al, 2009; Pritchard et al, 2000; Yang et al, 2012), which has helped obtaining a finer picture of populations' genetic background. More recently, other algorithms have been developed that allow inferring ancestry at the chromosome level (Brisbin et al, 2010; Price et al, 2009; Sankararaman et al, 2008), which has supposed a great step forward in establishing genetic relationships between populations.

In turn, we have used local ancestry assignment to infer recent time of gene flow or admixture events between populations, based on the length of the segments of a given ancestry. This has shed light for the first time to the genetic impact of historical migrations in North Africa, as estimates of gene flow coincide with those events.

Nonetheless, North Africa has shown to be a region with a particular demographic structure and displaying significant levels of gene flow from neighboring regions. This framework has sometimes entailed a challenge when applying novel algorithms. Here, we discuss some of the challenges we have faced in the analysis of North African populations and how we interpret the main findings.

4.2 The North African ancestors of extant North Africans

In Fadhlaoui-Zid et al (2011) we found in agreement with previous studies that the oldest haplogroups in North Africa are the back-to-Africa U6 and M1 haplogroups. Nonetheless, they are present at low frequencies (around 7%) and show a clear geographic structure:

M1 is more frequent in the east whereas U6 is more frequent in the west. Other Paleolithic haplogroups are H (Achilli et al, 2004b) and its sub-haplogroups H1 and H3 have been associated with the end of the Paleolithic, the LGM expansion (Achilli et al, 2004b; Torroni et al, 2001a). Haplogroup H is thought to have spread from the Near East in Europe, and it is possible that during this process it also entered North Africa. On the other hand, H1 and H3 are thought to have a southwestern European origin, specifically in the Franco-Cantabrian refuge.

The analysis of autosomal data (Henn et al, 2012) has shown that a Maghrebi ancestry, different from that of the Near East and Europe can be identified, and is present at very high frequencies in some populations, specially in the west. Dating the time split between North Africans and both Near Eastern and European populations, we find that this Maghrebi component has at least between 18,000 and 38,000 years ago, which is in the same time range of U6 and M1 coalescence times (44,000 ± 21,600 and 23,000 ± 9,200 ya, respectively) and coincides as well with the time during which haplogroup H is supposed to have expanded into Europe (Richards et al, 2000b).

It must be noticed that these mtDNA haplogroups do not show a star-like shape, so that the dating could be uncertain (Richards et al, 2000b) given to the effect of genetic drift. In the case of autosomal analysis, the caveat is that date estimates from the Near East are based only on a single Arabic population, which may not be the most appropriate one to make this test considering the historic episode of the Arabic expansion. However, genetic distances from southern European populations, which have more thoroughly been studied, are also present in the analysis and time intervals coincide with a hypothetical Near Eastern source and posterior expansion into Europe and North African in a parallel way.

It is noticeable in any case that both dense genotype data and mtDNA analysis give similar results, and show that during Paleolithic times there was a migration from the Near East towards North Africa which ultimately represents the autochthonous genetic architecture of present day North Africans.

Of special mention is the case of Tunisian Berbers, which in the *ADMIXTURE* analysis most of their genome is assigned to a single ancestry that is barely present in other individuals. This could be interpreted as Tunisians having another ancestry independent from that of other North Africans, though previous studies based on uniparental markers have shown that this population displays very low levels of genetic diversity (Fadhlaoui-Zid et al, 2004), which we have further investigated in Henn et al (2012). It is noticeable also that *ADMIXTURE* shows that its most closely related ancestral component is the autochthonous North African component.

On the basis of studies using uniparental markers I have suggested in section 1.2.2 that Berber-speaking populations show more evidences of genetic drift than Arab-speaking populations, and now for the first time autosomal data has found further evidences of this. However, now it has been possible as well to demonstrate that this Tunisian Berbers do not show evidences of admixture with North African neighbors. We interpret that all their genetic variation derives actually from the autochthonous Maghrebi component, representing a subset of it. It would be of great interest to investigate whether other Berber-speaking populations from other locations in the region show similar genetic differentiation that Tunisian Berbers or whether the effect of genetic drift led to a different derivation of the Maghrebi component.

4.3 Sub-Saharan gene flow into North Africa.

a) Detecting sub-Saharan ancestry in North Africa and determining its origin

In the case of mtDNA assessing sub-Saharan ancestry in North Africa (or elsewhere) is straightforward, given that haplogroups with a sub-Saharan origin are very well defined (Salas et al, 2002). L-lineages proportions in North Africa are very variable across populations, though on average they represent around 23% of the mitochondrial lineages. In (Fadhlaoui-Zid et al, 2011) we find significant differences in the distribution of eastern and western sub-Saharan lineages in North Africa, following a diffusion model. This is in agreement with previous studies (Harich et al, 2010; Salas et al, 2002) that show a differential east-west pattern as well.

In the case of autosomal data, detecting sub-Saharan ancestry is also fairly simple, because the genetic difference between Africa and Out-of-Africa populations is enough to be easily detected using these algorithms. However, most algorithms that assign ancestry proportions at an individual level are based on the dataset one is working with. Therefore, one must take into consideration information available and determine its completeness to select the populations that need to be included. In this way, if a representative of a given parental population is absent, its genetic contribution to the admixed population will be reflected in the closest population included in the dataset, leading to an overestimation of this particular population.

Ideally, in the study of North Africa all sub-Saharan African populations neighboring North Africa should be included, specially because it is known that in Africa a clear genetic structure exists even between close populations (Henn et al, 2011), so it is possible that the populations we are using are not the closest ones to the real populations that migrated into the region. Unfortunately, sampling in Africa is still far from complete, and array data is not available for all populations. A further caveat is that even if one can obtain genotype data for a given set of populations, the process of merging several datasets that have been genotyped in different platforms or even at different centers leads to the loss of a lot of markers due mainly to quality control processes.

Individual ancestry assignment performed with *ADMIXTURE* in Henn et al. 2012 with a variety of sub-Saharan populations had some surprising results. All North African populations were more closely related with eastern sub-Saharan populations, Luhya and Maasai, than with other western sub-Saharan African populations. Specifically, western North African individuals had more genetic proximity with Luhya whereas eastern North Africans were more closely related with Maasai.

Considering the mtDNA results mentioned before and taking into account that Luhya are the only Bantu-speakers included in the analysis, we hypothesize that sub-Saharan component from western North Africans most likely comes from a western sub-Saharan population closely related with Bantu. The inclusion of further

Bantu populations could help to assess more accurately the origin of sub-Saharan migrants admixing with North Africans.

b) Dating the time of gene flow to North Africa

Sub-Saharan African mtDNA haplogroups that we found in North Africa are thought to have entered the region recently, in agreement with (Harich et al, 2010). Specially, a very consistent pattern with trans-Saharan routes has been found (**Figure 21**).
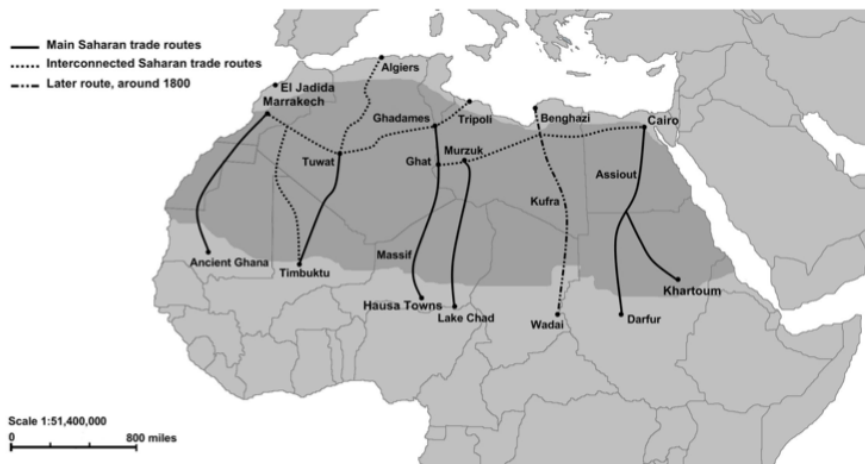
Figure 21 Routes from trans-Saharan slave trade (Harich et al. 2010).

Dense genotype data allows performing local ancestry assignment to make inferences about time of gene flow. The underlying idea is that after an admixture event, migrant haplotypes are broken by recombination at each generation with the consequent decrease in its length (Pool & Nielsen, 2009). Analyzing the length distribution of migrant haplotypes obtained by local ancestry allows inferring the time since admixture between two populations (Gravel, 2012b). However, this relies on the accuracy of algorithms in assigning ancestries across the chromosome, which in turn depends on how distinct are the populations considered and the number of markers used.

Reliable results are obtained when determining ancestry in African Americans (Shriner et al, 2011) **Figure 22**. In this case the scenario is relatively simple. Only two ancestries are involved, they are highly differentiated between them (Fst ~ 0.15), and the admixture

event is recent, so that tracts of ancestry are long enough to be easily detected. The case of Mexican Americans, though involving three different ancestries (Native America, European and African), still comprises highly differentiated ancestries. However, in this case it has already been shown that results are strongly dependent on how good is the approximation of the parental population (Tang et al, 2005).
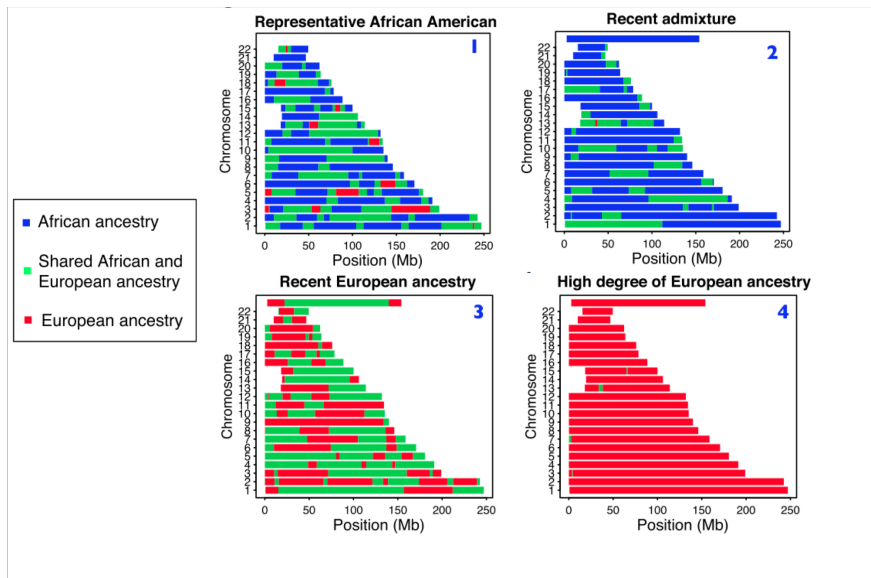
Figure 22 Individual ancestry estimates of four representative African-Americans. The colors represent two chromosomes of West African ancestry (blue), two chromosomes of European ancestry (red), or one chromosome of West African and one chromosome of European ancestry (green) (Bryc et al. 2010)

In the case of North Africa the scenario is much more complex, given that five different ancestries are involved. However, we decided to choose those North African populations that display levels of mainly a single sub-Saharan ancestry, and considering that the genetic distance between this and the other Out-of-Africa ancestries is greater than 0.09, we could reliably use the inferred sub-Saharan tracts so that they could be used to infer time of gene flow.

Results have shown in agreement with mtDNA studies, that sub-Saharan gene flow into North Africa is recent. In western North Africa it dates 1,200 years ago, coinciding with the rising of the

Ghana Empire, which also coincides with the trans-Saharan slave trade period. Regarding sub-Saharan gene flow in Egypt, a significant increase in gene flow is detected around likely occurred after the Arabic expansion into North Africa 1,400 ya. Nonetheless, caution must be taken in interpreting these dates, given that it is likely that populations geographically closer to North Africa could tell us a slightly different picture, in a similar way to which the inclusion of North Africa in detecting gene flow between Africa and Europe has also shown different results to those previously found (Botigué et al; Moorjani et al, 2011).

In conclusion, the development of these methods has confirmed that recent history has had indeed a perceptible impact in the genetic structure of contemporary populations (Botigué et al; Campbell et al, 2012; Moorjani et al, 2011). However, it is not likely that all the genetic architecture of actual populations is the consequence of migrations during historicity. The next step of great interest would be to infer former demographic events, like the ones associated to Neolithic processes, one of the periods of more relevance in human evolution but still not fully understood. With the availability of complete sequences

4.4 Studying North Africa and other Out-of-Africa populations.

Our Mitochondrial DNA studies suggest, as with previous studies including those using other markers, a close genetic relationship between Eurasian and North African populations as demonstrated by most of the haplogroups having a Eurasian origin. Related to this, Richards et al (2000a) suggest a major Neolithic origin for many Eurasian haplogroups. One interesting results found in Fadhlaoui-Zid is that, Eurasian haplogroups show a pattern similar to that shown by sub-Saharan haplogroups: eastern North Africa has a differential contribution of Near Eastern haplogroups, whereas westsern North Africa has a greater proportion of south western European lineages. Specifically, in agreement with an isolation by distance model, the Near Eastern influence we find in North Africa decreases from east to west (excepting the Tunisians).

Two major events have been associated with a possible expansion from the Near East to North Africa, the agripastoralism expansion during the Neolithic and the Islamic expansion during the sixth century. On the other hand, we find that European lineages in western North Africa might be associated with the LGM, in agreement with other studies (Achilli et al, 2004a; Torroni et al, 2001b).

Nonetheless, as pointed in the former section, genetic differences between populations is a key factor to be able to carry out precise estimations of admixture event using genomic data. In this case, we are dealing with populations with an average Fst of 0.05. As a consequence, even if at the genomic scale they can be differentiated, when trying to distinguish segments of these ancestries the density of markers available is not sufficient to distinguish haplotypes. Therefore, it may be that some segments are miss-assigned and undetectable. In the case of Near Eastern ancestry, a further issue is that only a single population is available to represent Near Eastern ancestry.

In general I think that assigned ancestry of Out of Africa populations should be taken with caution and considered as a broad picture of the pattern of genetic diversity, but not used as a tool to carry out further analysis.

Even if no formal analysis can be performed with the available data to confirm which event is the one that explains the pattern observed in eastern North Africa, I think that evidences are more supportive of a Neolithic expansion rather than the Islamic one. The *ADMIXTURE* analysis in Henn et al (2012) does not show high levels of variation in the Near Eastern ancestry across individuals within populations, which has been interpreted as an older rather than recent admixture event (Gravel, 2012b). On the other hand, in the introduction we have shown that most of the studies find evidences of a weak genetic impact of the Arabic expansion in western North Africans, and we do not find evidences of extensive haplotype sharing between eastern North African and Arabic lineages (Fadhlaoui-Zid et al, 2011).

In summary, I hypothesize that the Near Eastern component found mainly in Egypt and Libya is most likely to be of Neolithic origin.

A greater marker density and a better representation of Near Eastern populations would undoubtedly help answering this question.

Regarding gene flow between North Africa and Europe, we took advantage of the fact that extensive efforts have been made in terms of sampling in Europe, and that dense genotype data is available for most of the regions (Nelson et al, 2008) to investigate this subject. Also, a study appeared recently estimating proportions of African ancestry in Europe (Moorjani et al, 2011), and greater genetic diversity in southern Europe compared to northern Europe has also been detected (Auton et al, 2009). Finally, the genetic relationship between North Africa and the Iberian Peninsula has also been a long-standing question, with studies supporting extensive gene flow and others showing evidences of isolation between the regions, as has been reviewed in the Introduction.

We decided to complement results of individual ancestry performed by *ADMIXTURE* with Identity by Descent (IBD) analysis, with the scope of investigating patterns of recent gene flow. We have detected substantial gene flow from western North Africa and the Iberian Peninsula, with similar proportions estimated by both algorithms. Given that *ADMIXTURE* has the further advantage of distinguishing the directionality of the gene flow, we have determined that all the IBD segments shared between regions are the result of recent gene flow from North Africa to Europe. On the other hand, I think it is highly unlikely that genomic evidences of a LGM migration from the Iberian Peninsula to western North Africa are detected in shared IBD segments from array data.

5. CONCLUDING REMARKS

Concluding remarks

North African populations are distinct from sub-Saharan Africans based on cultural, linguistic and phenotypic attributes. The history of North Africa is extremely complex, with evidence of anatomically modern humans already settling in the region 160,000 ya. It is difficult to assess from the archaeological record whether early populations were replaced by later migrations or if there has been continuous settlement of the region. The time and the extent of genetic divergence between populations North and South of the Sahara remain poorly understood, as well as African connections with Near Eastern populations.

To resolve the history of human origin and migrations in North Africa, I have used two main forms of genetic data. First, in collaboration with colleagues, I have analyzed the genetic landscape of North Africa using maternally inherited mitochondrial DNA data (mtDNA) from previously published populations and from general Libyan population. Second, I have analyzed new genome-wide SNP genotyping array data (730,000 autosomal markers) from seven North African populations spanning from Morocco to Egypt and four Spanish populations.

Analysis of high-throughput genotyping in North African populations has answered a long-standing question about the region, that is, ancestors of contemporary North Africans that inhabited in the region date from Paleolithic times, and the Neolithic wave of expansion did not entail a complete nor relevant genetic replacement in the entire region. Also, it has been possible to assess proportions of ancestry from neighboring regions, including up to four different ancestries. Of special interest is the Near Eastern influence detected mainly in Egypt and Libya, which is in agreement with the results shown using mitochondrial DNA. The study of this region using dense genotype data has also allowed investigating the amount of North African admixture we find above all in the Iberian Peninsula and in minor proportions in the rest of Europe. Also, it has been possible to determine the time of migration and the nature of Sub-Saharan gene flow using haplotype information extracted from a local ancestry assignment analysis.

References

(2005) A haplotype map of the human genome. *Nature* **437:** 1299-1320

(2010) Archaeologists find 5,000 year-old skeletons in Morocco. Middle East online, http://www.middle-east-online.com/english/?id=38879.

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, Moral P, Dugoujon JM, Roostalu U, Loogvali EL, Kivisild T, Bandelt HJ, Richards M, Villems R, Santachiara-Benerecetti AS, Semino O, Torroni A (2004a) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet* **75:** 910-918

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, Moral P, Dugoujon JM, Roostalu U, Loogvali EL, Kivisild T, Bandelt HJ, Richards M, Villems R, Santachiara-Benerecetti AS, Semino O, Torroni A (2004b) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet* **75:** 910-918

Adams SM, Bosch E, Balaresque PL, Ballereau SJ, Lee AC, Arroyo E, Lopez-Parra AM, Aler M, Grifo MS, Brion M, Carracedo A, Lavinha J, Martinez-Jarreta B, Quintana-Murci L, Picornell A, Ramon M, Skorecki K, Behar DM, Calafell F, Jobling MA (2008a) The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *Am J Hum Genet* **83:** 725-736

Adams SM, Bosch E, Balaresque PL, Ballereau SJ, Lee AC, Arroyo E, Lopez-Parra AM, Aler M, Grifo MS, Brion M, Carracedo A, Lavinha J, Martinez-Jarreta B, Quintana-Murci L, Picornell A, Ramon M, Skorecki K, Behar DM, Calafell F, Jobling MA (2008b) The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *Am J Hum Genet* **83:** 725-736

Albrechtsen A, Moltke I, Nielsen R (2010) Natural selection and the distribution of identity-by-descent in the human genome. *Genetics* **186:** 295-308

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19:** 1655-1664

Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Bonnen PE (2010) Integrating common and rare genetic variation in diverse human populations. *Nature* **467:** 52-58

Ammerman A, Cavalli-Sforza LL (1984) *The Neolithic transition and the genetics of populations in Europe*, Princeton: Princeton University Press.

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290:** 457-465

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N *Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA*: Nat Genet. 1999 Oct;23(2):147.

Arredi B, Poloni ES, Paracchini S, Zerjal T, Fathallah DM, Makrelouf M, Pascali VL, Novelletto A, Tyler-Smith C (2004) A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa. *Am J Hum Genet* **75:** 338-345

Ashlock PD (1974) The Uses of Cladistics. *Annual Review of Ecology and Systematics* **5:** 81-99

Atzmon G, Hao L, Pe'er I, Velez C, Pearlman A, Palamara PF, Morrow B, Friedman E, Oddoux C, Burns E, Ostrer H (2010) Abraham's children in the genome era: major Jewish diaspora populations comprise distinct genetic clusters with shared Middle Eastern Ancestry. *Am J Hum Genet* **86:** 850-859

Auton A, Bryc K, Boyko AR, Lohmueller KE, Novembre J, Reynolds A, Indap A, Wright MH, Degenhardt JD, Gutenkunst RN, King KS, Nelson MR, Bustamante CD (2009) Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Res* **19:** 795-803

Balter M (2006) Radiocarbon Dating's Final Frontier. *Science* **313:** 1560-1563

Balter M (2011) Was North Africa the Launch Pad for Modern Human Migrations? *Science* **331:** 20-23

Barton N, Bouzouggar A, Humphrey L, Berridge P, Collcutt S, Gale R, Parfitt S, Parker A, Rhodes E, Schwenninger J-L (2008) Human Burial Evidence from Hattab II Cave and the Question of Continuity in Late Pleistocene–Holocene Mortuary Practices in Northwest Africa. *Cambridge Archaeological Journal* **18:** 195-214

Baudouin SV, Saunders D, Tiangyou W, Elson JL, Poynter J, Pyle A, Keers S, Turnbull DM, Howell N, Chinnery PF (2005) Mitochondrial DNA and survival after sepsis: a prospective study. *Lancet* **366:** 2118-2121

Behar DM, Harmant C, Manry J, van Oven M, Haak W, Martinez-Cruz B, Salaberria J, Oyharcabal B, Bauduer F, Comas D, Quintana-Murci L (2012) The Basque paradigm: genetic evidence of a maternal continuity in the Franco-Cantabrian region since pre-Neolithic times. *Am J Hum Genet* **90:** 486-493

Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, Bertranpetit J (2001) High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. *Am J Hum Genet* **68:** 1019-1029

Bosch E, Calafell F, Perez-Lezaun A, Clarimon J, Comas D, Mateu E, Martinez-Arias R, Morera B, Brakez Z, Akhayat O, Sefiani A, Hariti G, Cambon-Thomsen A, Bertranpetit J (2000) Genetic structure of north-west Africa revealed by STR analysis. *Eur J Hum Genet* **8:** 360-366

Bosch E, Calafell F, Perez-Lezaun A, Comas D, Mateu E, Bertranpetit J (1997) Population history of north Africa: evidence from classical genetic markers. *Hum Biol* **69:** 295-311

Botigué L, Henn B, Gravel S, Corona E, Gignoux C, Atzmon G, Burns E, Ostrer H, Flores C, Bertranpetit J, Comas D, Bustamante C Gene flow from North Africa contributes to differential genetic diversity in Southern Europe. *Proceedings of the National Academy of Sciences* **under review**

Bouzouggar A, Barton N, Vanhaeren M, d'Errico F, Collcutt S, Higham T, Hodge E, Parfitt S, Rhodes E, Schwenninger J-L, Stringer C, Turner E, Ward S, Moutmir A, Stambouli A (2007) 82,000-year-old shell beads from North Africa and implications for the origins of modern human behavior. *Proceedings of the National Academy of Sciences* **104:** 9964-9969

Bouzouggar A, Barton R, Blockley S, Bronk-Ramsey C, Collcutt S, Gale R, Higham T, Humphrey L, Parfitt S, Turner E, Ward S (2008) Reevaluating the Age of the Iberomaurusian in Morocco. *African Archaeological Review* **25:** 3-19

Brakez Z, Bosch E, Izaabel H, Akhayat O, Comas D, Bertranpetit J, Calafell F (2001) Human mitochondrial DNA sequence variation in the Moroccan population of the Souss area. *Ann Hum Biol* **28:** 295-307

Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, Tambets K, Antanaitis-Jacobs I, Haidle MN, Jankauskas R, Kind CJ, Lueth F, Terberger T, Hiller J, Matsumura S, Forster P, Burger J (2009) Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science (New York, NY)* **326:** 137-140

Bräuer G (1984) In *The origins of modern humans: a world survey of the fossil evidence*, Smith F, Spencer F (eds). New York: Liss, Alan R.

Brinkmann B, Klintschar M, Neuhuber F, Huhne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet* **62:** 1408-1415

Brisbin A, Weissman MM, Fyer AJ, Hamilton SP, Knowles JA, Bustamante CD, Mezey JG (2010) Bayesian linkage analysis of categorical traits for arbitrary pedigree designs. *PLoS One* **5**

Browning BL, Browning SR (2011) A fast, powerful method for detecting identity by descent. *Am J Hum Genet* **88:** 173-182

Browning SR, Browning BL (2010) High-resolution detection of identity by descent in unrelated individuals. *Am J Hum Genet* **86:** 526-539

Bryc K, Auton A, Nelson MR, Oksenberg JR, Hauser SL, Williams S, Froment A, Bodo JM, Wambebe C, Tishkoff SA, Bustamante CD (2010) Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc Natl Acad Sci U S A* **107:** 786-791

Burton ML, Moore CC, Whiting JWM, Romney AK, Aberle DF, Barcelo JA, Dow MM, Guyer JI, Kronenfeld DB, Levy JE, Linnekin J (1996) Regions Based on Social Structure. *Current Anthropology* **37:** 87-123

Bustamante CD, Burchard EG, De la Vega FM (2011) Genomics for the world. *Nature* **475:** 163-165

Campbell CL, Palamara PF, Dubrovsky M, Botigue LR, Fellous M, Atzmon G, Oddoux C, Pearlman A, Hao L, Henn BM, Burns E, Bustamante CD, Comas D, Friedman E, Pe'er I, Ostrer H (2012) North African Jewish and non-Jewish populations form distinctive, orthogonal clusters. *Proc Natl Acad Sci U S A* **109:** 13865-13870

Camps G (1974) *Les civilisations préhistoriques de l'Afrique du Nord et du Sahara*: Doin.

Camps G (1997) *Les Berbères: Mémoire et identité*, Paris: Actes Sud.

Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A, Chen Z, Chu J, Carcassi C, Contu L, Du R, Excoffier L, Ferrara

GB, Friedlaender JS, Groot H, Gurwitz D, Jenkins T, Herrera RJ, Huang X, Kidd J, Kidd KK, Langaney A, Lin AA, Mehdi SQ, Parham P, Piazza A, Pistillo MP, Qian Y, Shu Q, Xu J, Zhu S, Weber JL, Greely HT, Feldman MW, Thomas G, Dausset J, Cavalli-Sforza LL (2002) A human genome diversity cell line panel. *Science* **296:** 261-262

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) *The History and Geography of Human Genes*: Princeton University Press.

Cerezo M, Achilli A, Olivieri A, Perego UA, Gomez-Carballa A, Brisighelli F, Lancioni H, Woodward SR, Lopez-Soto M, Carracedo A, Capelli C, Torroni A, Salas A (2012) Reconstructing ancient mitochondrial DNA links between Africa and Europe. *Genome Res* **27:** 27

Cherni L, Loueslati BY, Pereira L, Ennafaa H, Amorim A, El Gaaied AB (2005) Female gene pools of Berber and Arab neighboring communities in central Tunisia: microstructure of mtDNA variation in North Africa. *Hum Biol* **77:** 61-70

Claramunt Rofríguez S (1991) La formació i l'expansió de l'Islam. In *Història Universal*, Rubio H (ed), Vol. 2. Barcelona: Edicions 92 S.A.

Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R (2005) Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res* **15:** 1496-1502

Close AE, Wendorf F (1990) North Africa at 18,000 BP. In *The world at 18,000 BP: Low latitudes*, Gamble C, Soffer O (eds), Vol. 2, pp 41-57. London: Unwin Hyman

Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, Batzer MA, Bertranpetit J, Sajantila A (2000) Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. *Hum Genet* **107:** 312-319

Conrad DF, Jakobsson M, Coop G, Wen X, Wall JD, Rosenberg NA, Pritchard JK (2006) A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet* **38:** 1251-1260

Corona E, Dudley JT, Butte AJ (2010) Extreme evolutionary disparities seen in positive selection across seven complex diseases. *PLoS One* **5**

Cortés Sánchez M, Jiménez Espejo FJ, Simón Vallejo MD, Gibaja Bao JF, Carvalho AF, Martinez-Ruiz F, Gamiz MR, Flores J-A, Paytan A, López Sáez JA, Peña-Chocarro L, Carrión JS, Morales Muñiz A, Roselló Izquierdo E, Riquelme Cantal JA, Dean RM, Salgueiro E, Martínez Sánchez RM, De la Rubia de Gracia JJ, Lozano Francisco MC, Vera Peláez JL, Rodríguez LL, Bicho NF The Mesolithic–Neolithic transition in southern Iberia. *Quaternary Research*

Coudray C, Guitard E, Kandil M, Harich N, Melhaoui M, Baali A, Sevin A, Moral P, Dugoujon JM (2006) Study of GM immunoglobulin allotypic system in Berbers and Arabs from Morocco. *Am J Hum Biol* **18:** 23-34

Coudray C, Olivieri A, Achilli A, Pala M, Melhaoui M, Cherkaoui M, El-Chennawi F, Kossmann M, Torroni A, Dugoujon JM (2009) The complex and diversified mitochondrial gene pool of Berber populations. *Ann Hum Genet* **73:** 196-214

Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V, Colomb EB, Zaharova B, Lavinha J, Vona G, Aman R, Cali F, Akar N, Richards M, Torroni A, Novelletto A, Scozzari R (2004) Phylogeographic analysis of haplogroup E3b (E-M215) y chromosomes reveals multiple migratory events within and out of Africa. *Am J Hum Genet* **74:** 1014-1022

Cruciani F, La Fratta R, Trombetta B, Santolamazza P, Sellitto D, Colomb EB, Dugoujon JM, Crivellaro F, Benincasa T, Pascone R, Moral P, Watson E, Melegh B, Barbujani G, Fuselli S, Vona G, Zagradisnik B, Assum G, Brdicka R, Kozlov AI, Efremov GD,

Coppa A, Novelletto A, Scozzari R (2007) Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosomal haplogroups E-M78 and J-M12. *Mol Biol Evol* **24:** 1300-1311

Currat M, Excoffier L (2005) The effect of the Neolithic expansion on European molecular diversity. *Proc Biol Sci* **272:** 679-688

d'Errico F, Vanhaeren M, Barton N, Bouzouggar A, Mienis H, Richter D, Hublin JJ, McPherron SP, Lozouet P (2009) Out of Africa: modern human origins special feature: additional evidence on the use of personal ornaments in the Middle Paleolithic of North Africa. *Proc Natl Acad Sci U S A* **106:** 16051-16056

Debénath A (2000) Le peuplement préhistorique du Maroc : Données récentes et problèmes. *Anthropologie* **104:** 131-145

Di Rienzo A, Donnelly P, Toomajian C, Sisk B, Hill A, Petzl-Erler ML, Haines GK, Barch DH (1998) Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. *Genetics* **148:** 1269-1284

Drake NA, Blench RM, Armitage SJ, Bristow CS, White KH (2011) Ancient watercourses and biogeography of the Sahara explain the peopling of the desert. *Proc Natl Acad Sci U S A* **108:** 458-462

El Idrissi A (2011). Neolithization and Early Neolithic of Morocco. *Transition in Mediterranean, how hunters became farmers (Epipaleolithic, Mesolithic, Early Neolithic)*; Toulouse. Museum de Toulousse.

El Moncer W, Esteban E, Bahri R, Gaya-Vidal M, Carreras-Torres R, Athanasiadis G, Moral P, Chaabani H (2010) Mixed origin of the current Tunisian population from the analysis of Alu and Alu/STR compound systems. *J Hum Genet* **55:** 827-833

Ennafaa H, Cabrera VM, Abu-Amero KK, Gonzalez AM, Amor MB, Bouhaha R, Dzimiri N, Elgaaied AB, Larruga JM (2009)

Mitochondrial DNA haplogroup H structure in North Africa. *BMC Genet* **10:** 8

Fadhlaoui-Zid K, Plaza S, Calafell F, Ben Amor M, Comas D, Bennamar El gaaied A (2004) Mitochondrial DNA heterogeneity in Tunisian Berbers. *Ann Hum Genet* **68:** 222-233

Fadhlaoui-Zid K, Rodriguez-Botigue L, Naoui N, Benammar-Elgaaied A, Calafell F, Comas D (2011) Mitochondrial DNA structure in North Africa reveals a genetic discontinuity in the Nile Valley. *Am J Phys Anthropol* **145:** 107-117

Flores C, Maca-Meyer N, Gonzalez AM, Cabrera VM (2000) Northwest African distribution of the CD4/Alu microsatellite haplotypes. *Ann Hum Genet* **64:** 321-327

Flores C, Maca-Meyer N, Perez JA, Hernandez M, Cabrera VM (2001) Y-chromosome differentiation in Northwest Africa. *Hum Biol* **73:** 513-524

Forster P (2004) Ice Ages and the mitochondrial DNA chronology of human dispersals: a review. *Philos Trans R Soc London [Biol]* **359:** 255-264

Garcea EAA, Giraudi C (2006) Late Quaternary human settlement patterning in the Jebel Gharbi. *Journal of Human Evolution* **51:** 411-421

Garcea EEA (2006) Semi-permanent foragers in semi-arid environments of North Africa. *World Archaeology* **38:** 197 - 219

Gauthier Y, Gauthier C, Morel A, Tillet T (1996) *L'art du Sahara: Archives des sables*, France: Seuil.

Gayan J, Galan JJ, Gonzalez-Perez A, Saez ME, Martinez-Larrad MT, Zabena C, Rivero MC, Salinas A, Ramirez-Lorca R, Moron FJ, Royo JL, Moreno-Rey C, Velasco J, Carrasco JM, Molero E, Ochoa C, Ochoa MD, Gutierrez M, Reina M, Pascual R, Romo-Astorga A, Susillo-Gonzalez JL, Vazquez E, Real LM, Ruiz A, Serrano-Rios M

(2010) Genetic structure of the Spanish population. *BMC Genomics* **11:** 326

Gignoux CR, Henn BM, Mountain JL (2011) Rapid, global demographic expansions after the origins of agriculture. *Proc Natl Acad Sci U S A* **108:** 6044-6049

Giles RE, Blanc H, Cann HM, Wallace DC (1980) Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A* **77:** 6715-6719

Gonzalez AM, Larruga JM, Abu-Amero KK, Shi Y, Pestano J, Cabrera VM (2007) Mitochondrial lineage M1 traces an early human backflow to Africa. *BMC Genomics* **8:** 223

Gonzalez-Perez E, Esteban E, Via M, Gaya-Vidal M, Athanasiadis G, Dugoujon JM, Luna F, Mesa MS, Fuster V, Kandil M, Harich N, Bissar-Tadmouri N, Saetta A, Moral P (2010) Population relationships in the Mediterranean revealed by autosomal genetic data (Alu and Alu/STR compound systems). *Am J Phys Anthropol* **141:** 430-439

Gravel S (2012a) Population Genetics Models of Local Ancestry. *Genetics* **4:** 4

Gravel S (2012b) Population genetics models of local ancestry. *Genetics* **191:** 607-619

Gusev A, Lowe JK, Stoffel M, Daly MJ, Altshuler D, Breslow JL, Friedman JM, Pe'er I (2009) Whole population, genome-wide mapping of hidden relatedness. *Genome Res* **19:** 318-326

Gusev A, Palamara PF, Aponte G, Zhuang Z, Darvasi A, Gregersen P, Pe'er I (2012) The Architecture of Long-Range Haplotypes Shared within and across Populations. *Mol Biol Evol* **29:** 473-486

Haak W, Forster P, Bramanti B, Matsumura S, Brandt G, Tanzer M, Villems R, Renfrew C, Gronenborn D, Alt KW, Burger J (2005) Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science (New York, NY)* **310:** 1016-1018

Haaland R (2005). Africa and the Near East: pot and porridge bread and oven: two food systems maintained over 10,000 years. *12th Congress of the Panafrican Archaeological Association for Prehistory and related studies*; Gaborone. University of Botswana.

Handel AE, Giovannoni G, Ebers GC, Ramagopalan SV (2010) Environmental factors and their timing in adult-onset multiple sclerosis. *Nat Rev Neurol* **6:** 156-166

Harich N, Costa MD, Fernandes V, Kandil M, Pereira JB, Silva NM, Pereira L (2010) The trans-Saharan slave trade - clues from interpolation analyses and high-resolution characterization of mitochondrial DNA lineages. *BMC Evol Biol* **10:** 138

Harich N, Esteban E, Chafik A, Lopez-Alomar A, Vona G, Moral P (2002) Classical polymorphisms in Berbers from Moyen Atlas (Morocco): genetics, geography, and historical evidence in the Mediterranean peoples. *Ann Hum Biol* **29:** 473-487

Heine B, Nurse D (2000) *African Languages: An Introduction*: Cambridge University Press.

Henn BM, Botigue LR, Gravel S, Wang W, Brisbin A, Byrnes JK, Fadhlaoui-Zid K, Zalloua PA, Moreno-Estrada A, Bertranpetit J, Bustamante CD, Comas D (2012) Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. *PLoS Genet* **8:** 12

Henn BM, Gignoux CR, Jobin M, Granka JM, Macpherson JM, Kidd JM, Rodriguez-Botigue L, Ramachandran S, Hon L, Brisbin A, Lin AA, Underhill PA, Comas D, Kidd KK, Norman PJ, Parham P, Bustamante CD, Mountain JL, Feldman MW (2011) Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc Natl Acad Sci U S A* **108:** 5154-5162

Henshilwood C, d'Errico F, Vanhaeren M, van Niekerk K, Jacobs Z (2004) Middle Stone Age Shell Beads from South Africa. *Science* **304:** 404

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68:** 87-112

Hiernaux J (1975) *The people of Africa*: Scribner.

Holden C (2004) Oldest Beads Suggest Early Symbolic Behavior. *Science* **304:** 369

Houck CM, Rinehart FP, Schmid CW (1979) A ubiquitous family of repeated DNA sequences in the human genome. *J Mol Biol* **132:** 289-306

Hublin JJ (2001) In *Human Roots*, L B, K R-B (eds), pp 99-121. Bristol, United Kingdom: Western Academic and Specialist

Humphrey L, Bello SM, Turner E, Bouzouggar A, Barton N (2012) Iberomaurusian funerary behaviour: Evidence from Grotte des Pigeons, Taforalt, Morocco. *Journal of Human Evolution* **62:** 261-273

Hunt C, Davison J, Inglis R, Farr L, Reynolds T, Simpson D, el-Rishi H, Barker G (2010) Site formation processes in caves: The Holocene sediments of the Haua Fteah, Cyrenaica, Libya. *Journal of Archaeological Science* **37:** 1600-1611

Hunter-Zinck H, Musharoff S, Salit J, Al-Ali KA, Chouchane L, Gohar A, Matthews R, Butler MW, Fuller J, Hackett NR, Crystal RG, Clark AG (2010) Population genetic structure of the people of Qatar. *Am J Hum Genet* **87:** 17-25

Irish JD (2000) The Iberomaurusian enigma: north African progenitor or dead end? *J Hum Evol* **39:** 393-410

Jackes M, Lubell D (2008) Early and Middle Holocene Environments and Capsian Cultural Change: Evidence from the Télidjène Basin, Eastern Algeria. *African Archaeological Review* **25:** 41-55

Jacobi RM, Higham TFG (2009) The early Lateglacial re-colonization of Britain: new radiocarbon evidence from Gough's

Cave, southwest England. *Quaternary Science Reviews* **28:** 1895-1913

Jacobs Z, Meyer MC, Roberts RG, Aldeias V, Dibble H, El Hajraoui MA (2011) Single-grain OSL dating at La Grotte des Contrebandiers ('Smugglers' Cave'), Morocco: Improved age constraints for the Middle Paleolithic levels. *Journal of Archaeological Science* **38:** 3631-3643

Jacobs Z, Roberts RG (2007) Advances in optically stimulated luminescence dating of individual grains of quartz from archeological deposits. *Evolutionary Anthropology: Issues, News, and Reviews* **16:** 210-223

Jacobs Z, Roberts RG, Nespoulet R, El Hajraoui MA, Debenath A (2012) Single-grain OSL chronologies for Middle Palaeolithic deposits at El Mnasra and El Harhoura 2, Morocco: Implications for Late Pleistocene human-environment interactions along the Atlantic coast of northwest Africa. *J Hum Evol* **12:** 12

Jobling MA (2012) The impact of recent events on human genetic diversity. *Philos Trans R Soc Lond B Biol Sci* **367:** 793-799

Jobling MA, Hurles M, Tyler-Smith C (2004) *Human Evolutionary Genetics: Origins, Peoples & Disease*: Garland Science.

Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* **4:** 598-612

Jorde LB, Bamshad M, Rogers AR (1998) Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *Bioessays* **20:** 126-136

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* **18:** 830-838

King TE, Parkin EJ, Swinfield G, Cruciani F, Scozzari R, Rosa A, Lim SK, Xue Y, Tyler-Smith C, Jobling MA (2007) Africans in

Yorkshire? The deepest-rooting clade of the Y phylogeny within an English genealogy. *Eur J Hum Genet* **15:** 288-293

Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis K, Passarino G, Underhill PA, Scharfe C, Torroni A, Scozzari R, Modiano D, Coppa A, de Knijff P, Feldman M, Cavalli-Sforza LL, Oefner PJ (2006) The role of selection in the evolution of human mitochondrial genomes. *Genetics* **172:** 373-387

Krings M, Salem AE, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, Sajantila A, Paabo S, Stoneking M (1999) mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration? *Am J Hum Genet* **64:** 1166-1176

Kujanova M, Pereira L, Fernandes V, Pereira JB, Cerny V (2009) Near eastern neolithic genetic input in a small oasis of the Egyptian Western Desert. *Am J Phys Anthropol* **140:** 336-346

Kurtzke JF, Delasnerie-Laupretre N, Wallin MT (1998) Multiple sclerosis in North African migrants to France. *Acta Neurol Scand* **98:** 302-309

Landsteiner K (1900) Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe. *Zentralblatt Bakteriologie* **27:** 357-362

Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, Caliebe A, Balascakova M, Bertranpetit J, Bindoff LA, Comas D, Holmlund G, Kouvatsi A, Macek M, Mollet I, Parson W, Palo J, Ploski R, Sajantila A, Tagliabracci A, Gether U, Werge T, Rivadeneira F, Hofman A, Uitterlinden AG, Gieger C, Wichmann HE, Ruther A, Schreiber S, Becker C, Nurnberg P, Nelson MR, Krawczak M, Kayser M (2008) Correlation between genetic and geographic structure in Europe. *Curr Biol* **18:** 1241-1248

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319:** 1100-1104

Linstädter J, Medved I, Solich M, Weniger G-C Neolithisation process within the Alboran territory: Models and possible African impact. *Quaternary International*

Loogvali EL, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, Metspalu E, Tambets K, Reidla M, Tolk HV, Parik J, Pennarun E, Laos S, Lunkina A, Golubenko M, Barac L, Pericic M, Balanovsky OP, Gusar V, Khusnutdinova EK, Stepanov V, Puzyrev V, Rudan P, Balanovska EV, Grechanina E, Richard C, Moisan JP, Chaventre A, Anagnou NP, Pappa KI, Michalodimitrakis EN, Claustres M, Golge M, Mikerezi I, Usanga E, Villems R (2004) Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol Biol Evol* **21:** 2012-2021

Loueslati BY, Cherni L, Khodjet-Elkhil H, Ennafaa H, Pereira L, Amorim A, Ben Ayed F, Ben Ammar Elgaaied A (2006) Islands inside an island: reproductive isolates on Jerba island. *Am J Hum Biol* **18:** 149-153

Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioglu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ (2004) The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *Am J Hum Genet* **74:** 532-544

Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* **2:** 13

Maca-Meyer N, Gonzalez AM, Pestano J, Flores C, Larruga JM, Cabrera VM (2003) Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. *BMC Genet* **4:** 15

Mariotti V, Bonfiglioli B, Facchini F, Condemi S, Belcastro MG (2009) Funerary practices of the Iberomaurusian population of Taforalt (Tafoughalt; Morocco, 11–12,000 BP): new hypotheses based on a grave by grave skeletal inventory and evidence of deliberate human modification of the remains. *Journal of Human Evolution* **56:** 340-354

Martinez-Cruz B, Harmant C, Platt DE, Haak W, Manry J, Ramos-Luis E, Soria-Hernanz DF, Bauduer F, Salaberria J, Oyharcabal B, Quintana-Murci L, Comas D (2012) Evidence of Pre-Roman Tribal Genetic Structure in Basques from Uniparentally Inherited Markers. *Mol Biol Evol* **6:** 6

McDougall I, Brown FH, Fleagle JG (2005) Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature* **433:** 733-736

Mercier N, Valladas H, Froget L, Joron JL, Vermeersch PM, Van Peer P, Moeyersons J (1999) Thermoluminescence Dating of a Middle Palaeolithic Occupation at Sodmein Cave, Red Sea Mountains (Egypt). *Journal of Archaeological Science* **26:** 1339-1345

Moorjani P, Patterson N, Hirschhorn JN, Keinan A, Hao L, Atzmon G, Burns E, Ostrer H, Price AL, Reich D (2011) The History of African Gene Flow into Southern Europeans, Levantines, and Jews. *PLoS Genet* **7:** e1001373

Mourant AE, Kopec AC, Domaniewska-Sobczak K (1976) *The distribution of the human blood groups and other polymorphisms*, London u.a.: Oxford Univ. Press.

Murdock GP (1967) *Ethnographic atlas*: University of Pittsburgh Press.

Nelis M, Esko T, Magi R, Zimprich F, Zimprich A, Toncheva D, Karachanak S, Piskackova T, Balascak I, Peltonen L, Jakkula E, Rehnstrom K, Lathrop M, Heath S, Galan P, Schreiber S, Meitinger T, Pfeufer A, Wichmann HE, Melegh B, Polgar N, Toniolo D, Gasparini P, D'Adamo P, Klovins J, Nikitina-Zake L, Kucinskas V, Kasnauskiene J, Lubinski J, Debniak T, Limborska S, Khrunin A, Estivill X, Rabionet R, Marsal S, Julia A, Antonarakis SE, Deutsch S, Borel C, Attar H, Gagnebin M, Macek M, Krawczak M, Remm M, Metspalu A (2009) Genetic structure of Europeans: a view from the North-East. *PLoS One* **4:** 8

Nelson MR, Bryc K, King KS, Indap A, Boyko AR, Novembre J, Briley LP, Maruyama Y, Waterworth DM, Waeber G, Vollenweider P, Oksenberg JR, Hauser SL, Stirnadel HA, Kooner JS, Chambers JC, Jones B, Mooser V, Bustamante CD, Roses AD, Burns DK, Ehm MG, Lai EH (2008) The Population Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics research. *Am J Hum Genet* **83:** 347-358

Nespoulet R, El Hajraoui M, Amani F, Ben Ncer A, Debénath A, El Idrissi A, Lacombe J-P, Michel P, Oujaa A, Stoetzel E (2008) Palaeolithic and Neolithic Occupations in the Témara Region (Rabat, Morocco): Recent Data on Hominin Contexts and Behavior. *African Archaeological Review* **25:** 21-39

Nielsen R, Hubisz MJ, Clark AG (2004) Reconstituting the frequency spectrum of ascertained single-nucleotide polymorphism data. *Genetics* **168:** 2373-2382

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, Stephens M, Bustamante CD (2008) Genes mirror geography within Europe. *Nature* **456:** 98-101

Novembre J, Stephens M (2008) Interpreting principal component analyses of spatial population genetic variation. *Nat Genet* **40:** 646-649

O'Dushlaine CT, Morris D, Moskvina V, Kirov G, Consortium IS, Gill M, Corvin A, Wilson JF, Cavalleri GL (2010) Population structure and genome-wide patterns of variation in Ireland and Britain. *Eur J Hum Genet* **18:** 1248-1254

O'Neil D. (2012) Evolution od modern humans. Vol. 1990, p. A Survey of the Biological and Cultural Evolution of Archaic and Modern Homo sapiens.

Olivieri A, Achilli A, Pala M, Battaglia V, Fornarino S, Al-Zahery N, Scozzari R, Cruciani F, Behar DM, Dugoujon JM, Coudray C, Santachiara-Benerecetti AS, Semino O, Bandelt HJ, Torroni A

(2006) The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. *Science* **314:** 1767-1770

Olszewski DI, Dibble HL, McPherron SP, Schurmans UA, Chiotti L, Smith JR (2010) Nubian Complex strategies in the Egyptian high desert. *Journal of Human Evolution* **59:** 188-201

Osborne AH, Vance D, Rohling EJ, Barton N, Rogerson M, Fello N (2008) A humid corridor across the Sahara for the migration of early modern humans out of Africa 120,000 years ago. *Proc Natl Acad Sci U S A* **105:** 16444-16447

Ottoni C, Martinez-Labarga C, Loogvali EL, Pennarun E, Achilli A, De Angelis F, Trucchi E, Contini I, Biondi G, Rickards O (2009) First genetic insight into Libyan Tuaregs: a maternal perspective. *Ann Hum Genet* **73:** 438-448

Owen R (2000) Karl Landsteiner and the First Human Marker Locus. *Genetics* **155:** 995-998

Padró i Parcerisa J (1993) L'Egipte faraònic. In *Història Universal*, Rubio H (ed), Vol. 1, pp 183-205. Barcelona: Editorial 92 S.A.

Pagonis V, Chen R, Kitis G (2011) On the intrinsic accuracy and precision of luminescence dating techniques for fired ceramics. *Journal of Archaeological Science* **38:** 1591-1602

Pereira L, Richards M, Goios A, Alonso A, Albarran C, Garcia O, Behar DM, Golge M, Hatina J, Al-Gazali L, Bradley DG, Macaulay V, Amorim A (2005) High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res* **15:** 19-24

Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, Srinivasan BS, Barsh GS, Myers RM, Feldman MW, Pritchard JK (2009) Signals of recent positive selection in a worldwide sample of human populations. *Genome Res* **19:** 826-837

Plaza S, Calafell F, Helal A, Bouzerna N, Lefranc G, Bertranpetit J, Comas D (2003) Joining the pillars of Hercules: mtDNA sequences

show multidirectional gene flow in the western Mediterranean. *Ann Hum Genet* **67:** 312-328

Pool JE, Nielsen R (2009) Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics* **181:** 711-719

Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, Ruczinski I, Beaty TH, Mathias R, Reich D, Myers S (2009) Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet* **5:** 19

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155:** 945-959

Pugach I, Matveyev R, Wollstein A, Kayser M, Stoneking M (2011) Dating the age of admixture via wavelet transform analysis of genome-wide data. *Genome Biol* **12:** 25

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

Ragoussis J (2009) Genotyping technologies for genetic research. *Annu Rev Genomics Hum Genet* **10:** 117-133

Rahmani N (2004) Technological and Cultural Change Among the Last Hunter-Gatherers of the Maghreb: The Capsian (10,000–6000 B.P.). *Journal of World Prehistory* **18:** 57-105

Rando JC, Pinto F, Gonzalez AM, Hernandez M, Larruga JM, Cabrera VM, Bandelt HJ (1998) Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Ann Hum Genet* **62:** 531-550

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M,

Rychkov S, Rychkov O, Rychkov Y, Golge M, Dimitrov D, Hill E, Bradley D, Romano V, Cali F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Di Rienzo A, Novelletto A, Oppenheim A, Norby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozari R, Torroni A, Bandelt HJ (2000a) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* **67:** 1251-1276

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Golge M, Dimitrov D, Hill E, Bradley D, Romano V, Cali F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Di Rienzo A, Novelletto A, Oppenheim A, Norby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozari R, Torroni A, Bandelt HJ (2000b) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* **67:** 1251-1276

Roostalu U, Kutuev I, Loogvali EL, Metspalu E, Tambets K, Reidla M, Khusnutdinova EK, Usanga E, Kivisild T, Villems R (2007) Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. *Mol Biol Evol* **24:** 436-448

Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, Barac L, Pericic M, Balanovsky O, Pshenichnov A, Dion D, Grobei M, Zhivotovsky LA, Battaglia V, Achilli A, Al-Zahery N, Parik J, King R, Cinnioglu C, Khusnutdinova E, Rudan P, Balanovska E, Scheffrahn W, Simonescu M, Brehm A, Goncalves R, Rosa A, Moisan JP, Chaventre A, Ferak V, Furedi S, Oefner PJ, Shen P, Beckman L, Mikerezi I, Terzic R, Primorac D, Cambon-Thomsen A, Krumina A, Torroni A, Underhill PA, Santachiara-Benerecetti AS, Villems R, Semino O (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in europe. *Am J Hum Genet* **75:** 128-137

Rosati G (2001) The prevalence of multiple sclerosis in the world: an update. *Neurol Sci* **22:** 117-139

Rose JI, Usik VI, Marks AE, Hilbert YH, Galletti CS, Parton A, Geiling JM, Černý V, Morley MW, Roberts RG (2011) The Nubian Complex of Dhofar, Oman: An African Middle Stone Age Industry in Southern Arabia. *PLoS ONE* **6:** e28239

Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman G, Beckman L, Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper G, Corte-Real HB, de Knijff P, Decorte R, Dubrova YE, Evgrafov O, Gilissen A, Glisic S, Golge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T, Kravchenko SA, Krumina A, Kucinskas V, Lavinha J, Livshits LA, Malaspina P, Maria S, McElreavey K, Meitinger TA, Mikelsaar AV, Mitchell RJ, Nafa K, Nicholson J, Norby S, Pandya A, Parik J, Patsalis PC, Pereira L, Peterlin B, Pielberg G, Prata MJ, Previdere C, Roewer L, Rootsi S, Rubinsztein DC, Saillard J, Santos FR, Stefanescu G, Sykes BC, Tolun A, Villems R, Tyler-Smith C, Jobling MA (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* **67:** 1526-1543

Rots V, Van Peer P, Vermeersch PM (2011) Aspects of tool production, use, and hafting in Palaeolithic assemblages from Northeast Africa. *Journal of Human Evolution* **60:** 637-664

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145:** 1219-1228

Roychoudhury AK, Nei M (1988) *Human polymorphic genes: world distribution*: Oxford University Press.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D (2001) A map of human genome

sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409:** 928-933

Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sanchez-Diz P, Macaulay V, Carracedo A (2002) The making of the African mtDNA landscape. *Am J Hum Genet* **71:** 1082-1111

Sankararaman S, Sridhar S, Kimmel G, Halperin E (2008) Estimating local ancestry in admixed populations. *Am J Hum Genet* **82:** 290-303

Saunier JL, Irwin JA, Strouss KM, Ragab H, Sturk KA, Parsons TJ (2009) Mitochondrial control region sequences from an Egyptian population sample. *Forensic Sci Int Genet* **3:** 5

Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, Santachiara-Benerecetti AS (2004) Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am J Hum Genet* **74:** 1023-1034

Sezgin E, Lind JM, Shrestha S, Hendrickson S, Goedert JJ, Donfield S, Kirk GD, Phair JP, Troyer JL, O'Brien SJ, Smith MW (2009) Association of Y chromosome haplogroup I with HIV progression, and HAART outcome. *Hum Genet* **125:** 281-294

Sheppard PJ, Lubell D (1990) Early Holocene Maghreb prehistory: An evolutionary approach. *Sahara* **3:** 63 - 69

Shriner D, Adeyemo A, Ramos E, Chen G, Rotimi CN (2011) Mapping of disease-associated variants in admixed populations. *Genome Biol* **12:** 30

Smith TM, Tafforeau P, Reid DJ, Grun R, Eggins S, Boutakiout M, Hublin JJ (2007) Earliest evidence of modern human life history in North African early Homo sapiens. *Proc Natl Acad Sci U S A* **104:** 6128-6133

Stevanovitch A, Gilles A, Bouzaid E, Kefi R, Paris F, Gayraud RP, Spadoni JL, El-Chenawi F, Beraud-Colomb E (2004) Mitochondrial DNA sequence diversity in a sedentary population from Egypt. *Ann Hum Genet* **68:** 23-39

Stoetzel E, Marion L, Nespoulet R, El Hajraoui MA, Denys C (2011) Taphonomy and palaeoecology of the late Pleistocene to middle Holocene small mammal succession of El Harhoura 2 cave (Rabat-Témara, Morocco). *Journal of Human Evolution* **60:** 1-33

Straus LG (2001) Africa and Iberia in the Pleistocene. *Quaternary International* **75:** 91-102

Stringer CB, Andrews P (1988) Genetic and fossil evidence for the origin of modern humans. *Science* **239:** 1263-1268

Tang H, Peng J, Wang P, Risch NJ (2005) Estimation of individual admixture: analytical and study design considerations. *Genet Epidemiol* **28:** 289-301

Tattersall I (2009) Human origins: Out of Africa. *Proceedings of the National Academy of Sciences* **106:** 16018-16021

Taylor VK, Barton RNE, Bell M, Bouzouggar A, Collcutt S, Black S, Hogue JT (2011) The Epipalaeolithic (Iberomaurusian) at Grotte des Pigeons (Taforalt), Morocco: A preliminary study of the land Mollusca. *Quaternary International* **244:** 5-14

Tills D, Kopeć AC, Tills RE, Mourant AE (1983) *The distribution of the human blood groups, and other polymorphisms*: Oxford University Press.

Topf AL, Gilbert MT, Fleischer RC, Hoelzel AR (2007) Ancient human mtDNA genotypes from England reveal lost variation over the last millennium. *Biol Lett* **3:** 550-553

Torroni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, Rengo C, Martinez-Cabrera V, Villems R, Kivisild T, Metspalu E, Parik J, Tolk HV, Tambets K, Forster P, Karger B, Francalacci P, Rudan P, Janicijevic B, Rickards O, Savontaus ML, Huoponen K,

Laitinen V, Koivumaki S, Sykes B, Hickey E, Novelletto A, Moral P, Sellitto D, Coppa A, Al-Zaheri N, Santachiara-Benerecetti AS, Semino O, Scozzari R (2001a) A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* **69:** 844-852

Torroni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, Rengo C, Martinez-Cabrera V, Villems R, Kivisild T, Metspalu E, Parik J, Tolk HV, Tambets K, Forster P, Karger B, Francalacci P, Rudan P, Janicijevic B, Rickards O, Savontaus ML, Huoponen K, Laitinen V, Koivumaki S, Sykes B, Hickey E, Novelletto A, Moral P, Sellitto D, Coppa A, Al-Zaheri N, Santachiara-Benerecetti AS, Semino O, Scozzari R (2001b) A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* **69:** 844-852

Underhill PA, Kivisild T (2007) Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu Rev Genet* **41:** 539-564

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* **26:** 358-361

Van Peer P (1998) The Nile Corridor and the Out-of-Africa Model An Examination of the Archaeological Record. *Current Anthropology* **39:** S115-S140

Van Peer P, Vermeersch PM (2000) The Nubian complex and the dispersal of modern humans in North Africa. In *Recent research into the Stone Age of Notheastern Africa*, Krzyzaniak L, Kroeper K, Kobusiewicz M (eds), pp 47-60. Poznan: Poznan Archaeological Museum

Watkins WS, Ricker CE, Bamshad MJ, Carroll ML, Nguyen SV, Batzer MA, Harpending HC, Rogers AR, Jorde LB (2001) Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms. *Am J Hum Genet* **68:** 738-752

Webster MT, Smith NG, Ellegren H (2002) Microsatellite evolution inferred from human-chimpanzee genomic sequence alignments. *Proc Natl Acad Sci U S A* **99:** 8748-8753

White TD, Asfaw B, DeGusta D, Gilbert H, Richards GD, Suwa G, Clark Howell F (2003) Pleistocene Homo sapiens from Middle Awash, Ethiopia. *Nature* **423:** 742-747

Witherspoon DJ, Marchani EE, Watkins WS, Ostler CT, Wooding SP, Anders BA, Fowlkes JD, Boissinot S, Furano AV, Ray DA, Rogers AR, Batzer MA, Jorde LB (2006) Human population genetic structure and diversity inferred from polymorphic L1(LINE-1) and Alu insertions. *Hum Hered* **62:** 30-46

Wood B (2010) Colloquium paper: reconstructing human evolution: achievements, challenges, and opportunities. *Proc Natl Acad Sci U S A* **2:** 8902-8909

Xu X, Peng M, Fang Z (2000) The direction of microsatellite mutations is dependent upon allele length. *Nat Genet* **24:** 396-399

Yang MA, Malaspinas AS, Durand EY, Slatkin M (2012) Ancient Structure in Africa Unlikely to Explain Neanderthal and Non-African Genetic Similarity. *Mol Biol Evol* **10:** 10

Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L (2004) The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet* **74:** 50-61

Zilhão J (2001) Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. *Proceedings of the National Academy of Sciences* **98:** 14180-14185

APENDIX

Sanchez-Quinto F, Botigue LR, Civit S, Arenas C, Avila-Arcos MC, Bustamante CD, et al. North African populations carry the signature of admixture with Neandertals. PLoS One. 2012;7(10): e47765.