Genetic Heterogeneity between Berbers and Arabs

Lara R Arauna, Institut de Biologia Evolutiva (CSIC-UPF), Departament de Ciències Experiments i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain

David Comas, Institut de Biologia Evolutiva (CSIC-UPF), Departament de Ciències Experiments i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain

The human population history of North Africa has been different from the rest of the continent, and it has been characterised by population replacements, extensive continuous gene flow, and differential admixture from neighbouring regions. This complex demographic landscape has yielded a large degree of genetic heterogeneity among North African populations. Recent historical admixture processes have been inferred from genome-wide data; no correlation between genetics and ethnic groups has been described, pointing to a lack of genetic differentiation between Berber and Arab groups in North Africa. This complex genetic population structure should be taken into account when designing biomedical approaches in North African groups.

Introduction

Little is known about the initial human settlement in North Africa, the movements during the Upper Paleolithic (known as the ‘Late Stone Age’ in the study of African prehistory) and the possible introgression from archaic humans, such as Neanderthals (Green et al., 2010) and Denisovans (Reich et al., 2010) or other unknown African hominins (Hammer et al., 2011), to the populations inhabiting this area. Recent description of forms of Homo sapiens with a mosaic of modern and primitive characteristics in North Africa dated back to ~300,000 ya (years ago) has even challenged the geographical origins of our species (Hublin et al., 2017). However, the first presence of anatomically modern humans (AMHs) in North Africa has been estimated 190,000–130,000 ya by archaeological studies (much older than previous estimates that dated the first occupation of the region ~45,000 ya) (Smith et al., 2007). The Aterian is the first prehistoric industry characterised in North Africa (Barton et al., 2009), which has been suggested to have started around 120,000 ya. Although paradoxical, the Sahara desert has been reported to be a corridor for the movement of people during the ‘green Sahara’ periods thanks to its watercourses. The old dates of AMH remains and the Sahara corridor leave many open questions about the role of North Africa in the origin of modern humans and their dispersal out of the continent. Different hypotheses have been proposed: Was North Africa just a stop on the road out of Africa, or did it play a more important role in the evolution of modern humans?

No clear connections have been established between this first human industry and subsequent cultures in the region, such as the Iberomaurusian industry (22,000–9500 ya) (Newman, 1995). The prefix ‘Ibero-’ refers to the ancient presumption that this culture extended into Iberia, although an origin in the Nile River valley is widely accepted (Camps, 1995). The Iberomaurusian culture was followed by the Capsian industry (10,000–4700 ya) (Newman, 1995) that persisted well into the Neolithic, which began around 5500 ya in the region. The prehistoric cultural changes in North Africa were independent of the change dynamics on the European shores of the Mediterranean.

Based on archaeology and linguistics, it has been argued that the peopling of North Africa has not been continuous and that population replacements may have occurred in different moments. Genetic data (Henn et al., 2012) has supported this population replacement hypothesis in North Africa, suggesting a back-to-Africa migration from the Middle East in pre-Holocene times, older than 12,000 ya. The ancestors of today’s North African populations derive at least partially from this migration wave; however, it is not clear if this replacement was complete or there might be some traces of ancient continuity in the region. Therefore, many open questions remain about the demography of the region in relation to the demographic continuity of humans since the first occupation of North Africa by modern humans, such as how many times a population replacement took place and which are the demographic parameters such as population size of the migrants and where did they come from.

The challenges of the understanding of North African populations are not limited to prehistoric times: in historic times, North Africa has also experienced a very complex history that is reflected in its human demography. Historical records document trade routes across the Sahara and contacts between both Mediterranean shores and the Middle East. Phoenicians (814 B.C.) and
Romans (146 B.C.) occupied part of coastal North Africa with limited population contributions, which were even less significant for Vandals (A.D. 429) and Byzantines (A.D. 533) (Newman, 1995). One of the major historical events in the region that changed the cultural and demographic landscape of North Africa is the Arabisation. The first Arab invasion, initially confined to Egypt, started in A.D. 643 and may have involved only a few thousand individuals (McEvedy, 1995). The Arabs began to impose their religion and language over the Berber autochthonous population, a process that culminated with a second and larger Arab wave in which the Bedouin reached the Maghreb (northwest Africa) in the eleventh century. The Islamic expansion even went on to engulf the Iberian Peninsula (A.D. 711). The later arrivals to North Africa in colonial times include Europeans and Ottoman Turks, mainly in Egypt.

Nowadays in North Africa, two main groups of languages are spoken: Berber and Arab dialects. Both belong to the Afroasiatic family, which is a group of languages that might have originated in the Mesolithic, between 15,000 and 10,000 ya. Although it is not clear, one of the most plausible places of origin of Afroasiatic languages is the ‘grassland east of the Nile Valley’ (Newman, 1995). However, it is unknown if the arrival of the Afroasiatic languages was mainly a cultural diffusion process or it also implied a relevant demographic movement.

The Berbers (Amazighen, as they call themselves) are considered the autochthonous people of North Africa. As can be elucidated from the complex and unsolved prehistory and history of North Africa, it is difficult to establish an origin for the Berbers, and many hypotheses have been proposed. Historical records reveal their presence before the Phoenician arrival to the region (Camps, 1998), and archaeologically, it has been suggested a link between the Capsian culture and the Berber people (Camps, 1995). However, the Berbers might not have been a uniform and homogeneous group, since it has been known that the populations that inhabited North Africa by the time of the Phoenician arrival were a composite group known as Libyco-Berbers, where we find the Garamantes, Bavares, Mauri, Gaetuli, among many others (Newman, 1995). In addition to the uncertainties about the origins of the Berbers, the arrival of other people in the region, especially the influence of Arabs, makes the understanding of the population history of Berbers more challenging.

Considering these previous points, the population genetics definition of today’s Berber and Arab groups is a complex task. Different points of view can be considered to determine whether a population is considered as Berber or not. Berbers can be understood as non-Semitic autochthonous populations in North Africa, although it is challenging to determine which current populations derive from an ancestral autochthonous group in North Africa. Another way to identify Berbers and Arabs is regarding their cultural characteristics, but again, the cultural exchange in recent history of North Africa has been common, and it is difficult to differentiate, for example, an Arabised Berber society from an Arab population in North Africa, among many other combinations. Given this combination of circumstances, the best option for a classification may be based on linguistics, Berber populations being those who speak Berber languages and Arabs being those speaking Arab (Dugoujon et al., 2009). This classification could be the way of having less ‘false positives’; however, this classification is not free of caveats, and populations identified as Arabs or non-Berbers could in fact have a large Berber genetic ancestry. It is important to emphasise that the purpose of this classification is to understand the genetics of North African populations and in any case to establish socio-cultural groups, which should take into consideration many other aspects (traditions, identity, etc.).

Despite the scanty genetic data on North African groups, population genetics might shed some light on the genetic structure of Berbers and Arabs, trying to establish the demographic scenarios that have modelled the current genetic landscape in North Africa.

Genetic Diversity in Berber and Arab Groups

The human genetic data of North Africa is scarce compared to other regions of the African continent, and most of the African genetic diversity studies have been focused on the origin of our species and the first dispersions out of Africa (see, for instance, Tishkoff et al., 2009), the genetic diversity of North African populations and surrounding groups being poorly characterised.

The analyses based on frequencies of classical genetic polymorphisms (blood groups, red cell enzymes and serum proteins) have shown that the genetic landscape in North Africa presents an east-west pattern of variation without differences between Arabs and Berbers, pointing to a sizeable Upper Paleolithic component in current North African populations, whereas the Neolithic diffusion in the region was more a cultural than a demic process (Barbujani et al., 1994; Bosch et al., 1997). These classical studies showed little contribution from sub-Saharan populations in North Africans and genetic differentiation from southern Europeans, including Iberians (Bosch et al., 1997). However, the analysis of mtDNA lineages has shown an important sub-Saharan contribution, although most haplogroups in North Africa are of west-Eurasian origin (Fadhlaoui-Zid et al., 2004, 2011b; Krings et al., 1999; Plaza et al., 2003; Randi et al., 1998). Some of them can be traced to ancient Paleolithic times (such as haplogroups U6, M1, which are almost specific of North African populations); however, some maternal lineages have been the result of a more recent acquisition from Europe or the Middle East (such as haplogroups U5, V, R0a, J1b, U3) (González et al., 2007; Maca-Meyer et al., 2003; Olivieri et al., 2006) (Figure 1). In addition, a large degree of genetic heterogeneity has been shown in North African maternal and paternal lineages compared to other geographical regions such as Europe (Fadhlaoui-Zid et al., 2004, 2011a; Plaza et al., 2003). The analysis of Y-chromosome lineages has shown a high frequency of two specific North African haplogroups (E-M81 and E-M78) (Figure 1), although their origins have been controversial since some analyses have suggested a Paleolithic component (Bosch et al., 2001), whereas others have pointed to a Neolithic origin (Arredi et al., 2004; Cruciani et al., 2004, 2007; Semino et al., 2004). The data on autosomal markers in North African populations based on some short tandem repeats (STRs) (Bosch et al., 2000) and Alu polymorphisms (Comas et al., 2000; Flores et al., 2000; González-Pérez et al., 2010) has been analysed.
Figure 1 Haplogroup frequencies for different Berber and Arab populations for (a) mitochondrial DNA (data from Coudray et al., 2009 and Plaza et al., 2003) and (b) Y chromosome (data from Arredi et al., 2004).
in few North African samples pointing to a closer genetic affinity to West Eurasian populations. Y-chromosome lineages and autosomal markers show a clear differentiation of North African populations with some sub-Saharan gene flow and almost negligible European contact (Bosch et al., 2000, 2001; Comas et al., 2000; González-Pérez et al., 2010). Nonetheless, the contribution of North African female and male lineages in the Iberian Peninsula (Adams et al., 2008; Plaza et al., 2003) should be stressed.

Genome-wide analyses have recently become the most powerful tool to unravel the genetic history of populations. The first large genome-wide study in North Africa analysing several North African and surrounding populations (Henn et al., 2012) suggested the existence of an autochthonous North African component whose origin is located out of Africa before the Holocene (>12,000 years ago) and a complex admixture from the Middle East, Europe and sub-Saharan Africa. This autochthonous North African component was found in high frequencies in the only Berber sample (Berber Tunisian from Chenini) analysed in the study, suggesting that Berbers might be responsible for this autochthonous component in North African samples. However, recent genome-wide data with additional Berber samples have shown a different scenario, with some Berber groups presenting less frequencies of the autochthonous component compared to some Arab groups (Arauna et al., 2017). In addition to the autochthonous component in North Africa, it has been possible to quantify and estimate the recent dates of the sub-Saharan admixture in North Africa (Henn et al., 2012) as well as the genetic contacts from North Africa mainly restricted to south-western Europe (Botigué et al., 2013). Concerning the introgression of archaic hominins into North Africans, some Neanderthal admixture is detected in these groups independent of their European component (Sánchez-Quinto et al., 2012).

Most of the genetic studies in North African populations agree with a limited or non-existent correlation between genetics and geography, and therefore, they show a high population heterogeneity in the region (Bosch et al., 2000; Fadhlaouï-Zid et al., 2004; Flores et al., 2011; González-Pérez et al., 2010). However, an east-west genetic pattern of variation has been observed (Arauna et al., 2017; Bosch et al., 1997; Harich et al., 2001; Henn et al., 2012), which has been explained by Middle Eastern gene flow towards North Africa according to genome-wide analyses (Arauna et al., 2017; Henn et al., 2012). The genetic heterogeneity is not only due to the wide extension of North Africa, since it is also found at microgeographical level when considering more restricted areas, such as Algeria (Bekada et al., 2015) or Tunisia (Fadhlaouï-Zid et al., 2011a; Kefi et al., 2015), for instance. Regardless of the genetic markers analysed, North African populations have been described as a mosaic of North African, Middle Eastern, European and sub-Saharan ancestries. The differential admixture with these four ancestry sources explains the current genetic structure in North Africa, characterised by diverse and heterogeneous populations. This differential admixture explains why nearby populations or even individuals inhabiting the same location might be genetically more distant than groups of people in geographically distant populations. For instance, North African individuals with similar amounts of sub-Saharan admixture present high genetic similarities regardless of their geographical origin. Neither geography, linguistics nor ethnic affiliation is correlated with a specific pattern of admixture proportions. A completely different scenario was shown when analysing North African Jews, who are genetically closer to other Jewish populations than to their geographic neighbours (Campbell et al., 2012). There are many possible scenarios that might explain the differential admixture in North Africa, socio-cultural structure and historical events being the main reasons. However, more complex socio-cultural scenarios, including socio-economic factors, a more precise ethnic or linguistic definition, particularities of the history of each local area and rural versus urban lifestyles, should be considered in order to find a correlation with the complex genetic pattern observed (Arauna et al., 2017).

Concerning the genetic differences between Berber and Arab groups, classical and uniparental markers have highlighted the lack of a genetic structure that could differentiate Berber and Arab populations (Bekada et al., 2015; Fadhlaouï-Zid et al., 2004; Harich et al., 2001; Kefi et al., 2015). Most of these authors have agreed on a cultural rather than a demographic impact of the Arabisation after the Arab expansion (Bosch et al., 2000, 2001; Ennafaa et al., 2011; Flores et al., 2011). However, as stated earlier, an East-to-West genetic gradient of a Middle Eastern component has been described in genome-wide analyses, and admixture events have been dated to the Arabisation period (Arauna et al., 2017; Henn et al., 2012), suggesting that the Arabisation might have had more demographic impact than previously thought.

The first widespread genome-wide study in North Africa (Henn et al., 2012) suggested that the single Berber sample analysed (Tunisian Berber from Chenini) presented high frequencies of an autochthonous North African component (named in the study the ‘Maghrébi’ component) and negligible presence of the Middle Eastern component. This lack of Middle Eastern component in Berbers and its higher frequencies in Arabs pointed to a genome differentiation between both North African groups. However, the posterior genome-wide analysis of additional Berber samples has challenged this genetic differentiation idea (Arauna et al., 2017). This study has corroborated the lack of genetic differentiation between Berbers and Arabs; however, it has shown that the Arabisation had also a demographic impact in North Africa (Figure 2).

It is difficult to exactly quantify the magnitude of the Arab genetic contribution during the Arab expansion in the seventh century, among other reasons because the influence from Middle East has been continuous into North Africa (Camps, 1995). Therefore, this data supports the lack of significant differences between Berber and Arabs, but points to a new explanation: a high and differential level of admixture since the Arab expansion that has affected most North African groups regardless of their Berber or Arab identity.

The other genetic components that are present in North African populations also contribute to the heterogeneity between Berber and Arabs. Both peoples have admixed with sub-Saharan populations in the last two millennia regardless of their ethnicity, which has contributed to the lack of differentiation between both groups (Flores et al., 2000; Harich et al., 2010; Henn et al., 2012). However, it has been shown that the sub-Saharan admixture is most frequent in the southern groups, and in fact, a South–North gradient of sub-Saharan admixture has been described in North Africa (Comas et al., 2000; Plaza et al., 2003), which sometimes can mimic differences between Arab and Berber populations that are
mainly explained by a differential sub-Saharan contribution (Flores et al., 2011).

The high heterogeneity found in North African populations is also present within ethnic groups. Berbers are a composite of groups with very different demographic characteristics, and some Berber groups are genetic outliers compared to other Berber or North African populations. This is the case of the Tunisian Berber population from Chenini, which has been analysed for uniparental and genome-wide markers, and in both cases, it has been described as an outlier group compared to their North African populations (Ennafaa et al., 2011; Fadhlaoui-Zid et al., 2004, 2011a; Henn et al., 2012). This genetic singularity is not exclusive of this group since other Tunisian Berber groups have also shown similar outlier characteristics (Ennafaa et al., 2011; Fadhlaoui-Zid et al., 2004, 2011a). Another example are the Mozabite, an Algerian Berber group included in the Human Genome Diversity Panel (HGDP) (Cann et al., 2002), which have been extensively analysed in many genetic studies and have also been described as an outlier group within North Africa and also within Berbers (Bekada et al., 2015; Plaza et al., 2003). It is important to note that Mozabite have been considered as the representative North African group in a large number of studies due to the fact that they are the only North African group in the HGDP, which should be taken with caution given their genetic outlier characteristics. Furthermore, the Siwa Berber population from Egypt is genetically distant from other Berber groups from West North Africa (Coudray et al., 2009), suggesting a differentiation between Egyptian Berbers and Western Berbers. However, there are many other Berber groups that are closely related between them and to other North African groups. All these evidences support a high diversity within Berbers, which could be explained by isolation processes and inbreeding. For example, the Chenini population is an isolated population that shows high levels of inbreeding (Arauna et al., 2017; Henn et al., 2012). It has been described that some of the diversity observed within Berbers is due to the belonging to an urban or rural area. Rural areas favour isolation leading to genetic differentiated populations (Brakez et al., 2001; Ennafaa et al., 2011; Frigi et al., 2006). On the other hand, urban areas that are less isolated are more favourable to high admixture levels and lead to the lack of differentiation between Berber and Arab groups. Finally, it is important to notice the wide extension and fragmentation of the territory where the Berbers live, which could lead to population differentiation because of many reasons apart from ethnicity (simply by isolation by distance or differential gene flow). Then, it is important to compare the Berber groups to geographically close populations in order to make inferences of genetic distances.

The genetic heterogeneity of North African samples, the complex population admixture patterns and the lack of correlation between genetics and ethnic groups should be taken into account when addressing biomedical approaches in North Africa. Ignoring these population aspects might lead to artefacts and false-positive results when searching for associations between genetic markers and biomedical phenotypes, including disease association studies. Furthermore, population genetic studies in North Africa should also take this diversity into account. It is important to trace many different variables when studying North African populations, such as the social and demographic characteristics of the studied groups (if it is urban or rural, if it is geographically isolated, etc.) and their history (the Arab influence in the region, slave trade routes, etc.). The genetic structure of North African population is determined by many different factors apart from a simple ethnic dichotomy Arab–Berber, and including those variables sheds lights in the understanding of the populations.

**Related Articles**

- **Reconstructing Human History Using Autosomal, Y-Chromosomal and Mitochondrial Markers**
- **Gene Flow, Haplotype Patterns and Modern Human Origins**
- **Genetic Diversity in Africa**
- **Humans: Demographic History**

**Glossary**

**Admixture** The process of mixing of different populations within a species.

**Cultural diffusion** The spread of cultural practices from another group of people without genetic exchange.
Demic diffusion  The spread of a population in an advancing wave.

Gene flow  The movement of genetic variation from one population to another.

Haplogroup  A group of similar allele combinations that share a common ancestor.

Inbreeding  Reproduction that involves genetically related individuals.

Introgression  The process of mixing of different species.

References


**Further Reading**
