Gene 564 (2015) 141–152

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

From Arabia to Iberia: A Y chromosome perspective

María Regueiro [a], Ralph Garcia-Bertrand [a,*], Karima Fadhlaoui-Zid [b], Joseph Álvarez [a], Rene J. Herrera [a]

[a] Biology Department, Colorado College, Colorado Springs, CO 80903, USA
[b] Laboratoire de Genetique, Immunologie et Pathologies Humaines, Faculte des Sciences de Tunis, Campus Universitaire El Manar II, Universite el Manar, Tunis, Tunisia

ARTICLE INFO

Article history:
Received 5 January 2015
Received in revised form 8 February 2015
Accepted 15 February 2015
Available online 17 February 2015

Keywords:
Phylogenetic relationships
Y haplogroups E and J
Y-STR
Muslim migration of Iberia

ABSTRACT

At different times during recent human evolution, northern Africa has served as a conduit for migrations from the Arabian Peninsula. Although previous researchers have investigated the possibility of the Strait of Gibraltar as a pathway of migration from North Africa to Iberia, we now revisit this issue and theorize that although the Strait of Gibraltar, at the west end of this corridor, has acted as a barrier for human dispersal into Southwest Europe, it has not provided an absolute seal to gene flow. To test this hypothesis, here we use the spatial frequency distributions, STR diversity and expansion time estimates of Y chromosome haplogroups J1-P58 and E-M81 to investigate the genetic imprints left by the Arabian and Berber expansions into the Iberian Peninsula, respectively. The data generated indicate that Arabian and Berber genetic markers are detected in Iberia. We present evidence that suggest that Iberia has received gene flow from Northwest Africa during and prior to the Islamic colonization of 711 A.D. It is interesting that the highest frequencies of Arabia and Berber markers are not found in southern Spain, where Islam remained the longest and was culturally most influential, but in Northwest Iberia, specifically Galicia. We propose that Moriscos' relocations to the north during the *Reconquista*, the migration of cryptic Muslims seeking refuge in a more lenient society and/or more geographic extensive pre-Islamic incursions may explain the higher frequencies and older time estimates of mutations in the north of the Peninsula. These scenarios are congruent with the higher diversities of some diagnostic makers observed in Northwest Iberia.

1. Introduction

Western Europe, including the Iberian Peninsula, was first populated by humans relatively recently in the mid upper Paleolithic 35,000–40,000 years ago (ya) (Cavalli-Sforza et al., 1994). Although Europe is characterized by overall genetic homogeneity, in comparison to other biogeographical regions (Lao et al., 2008), Iberia, in particular, has been a recipient as well as a reservoir of human diversity. During the last Ice Age (18,000–80,000 ya), for example, the Peninsula became a human refugium as glaciers advanced, covering most of Europe encapsulating human populations and genetic diversity in what is now Portugal and Spain (Richards et al., 2000; Torroni et al., 2000). When the ice sheet began to retreat, about 15,000 ya, subsequent to the Last Glacial Maximum (LGM), this sanctuary was one of the sources that participated in the repopulation of Europe.

Starting in the late Paleolithic, a number of archeological sites begin to signal the impact of specific pre-historical cultures of diverse origins on the Iberian Peninsula. The most significant groups include the Tartessians (Koch, 2013), of unknown origins that appeared during the 9th century before the common era (BCE) (Marcos Garcia, 1987; Almagro-Gorbea, 2004), Proto-Indo Europeans from the steeps of Eastern Europe (Mallory et al., 1997) or Eastern Mediterranean Islands (Fernández et al., 2014) early in the Neolithic (7500–5500 BCE), Iberians from the Eastern Mediterranean or North Africa in the 6th century BCE (Sanmartí, 2005) and Celts from Central Europe about 450 BCE (Judice Gamito, 1994). In more recent historical times, Phoenicians from the Middle East, Greeks, Romans, Germanic tribes from Central Europe, Vikings from Scandinavia, Arabs from the Persian Gulf region and Berbers from Northwest Africa as well as the Roma from India have contributed to the Iberian genetic make-up (de Hoz, 1982; Gieben, 1991). Several of the above mentioned migrations are clearly reflected in the collage of cultures, linguistic affinities, music and architecture that are on display in the various regions of Iberia today, yet it remains to be ascertained to what extent this extreme cultural diversity is in fact reflected in the gene pools of its extant populations. Gene flow from North Africa to Iberia has been in demonstrated in a number of previous publications (Rando et al., 1998; Pereira et al., 2005; Plaza et al., 2003; Santos et al., 2014) and higher frequency of some Y-

Abbreviations: STR, Short tandem repeat; YA, Years ago; LGM, Last Glacial Maximum; BCE, Before current era; mtDNA, Mitochondrial DNA; A.D., After death; SNP, Single nucleotide polymorphism; MJ, Median joining; MDS, Multi-dimensional Scaling; MP, Maximum parsimony; GD, Gene diversity.

* Corresponding author at: Biology Department, Colorado College, 14 East Cache La Poudre Street, Colorado Springs, CO 80903-3294, USA.
E-mail address: RBertrand@coloradocollege.edu (R. Garcia-Bertrand).

http://dx.doi.org/10.1016/j.gene.2015.02.042
0378-1119/

chromosome North African lineages have been observed in northern Iberia (Adams et al, 2008).

By the late 7th century CE, Arabian and Bedouin forces coming from the Arabian capital of Damascus had reached the far west of North Africa (Maghreb). In 711 CE, a Berber-speaking army under Arabian suzerainty crossed over into the Iberian Peninsula and, within four years, had captured almost the entire Peninsula, with the exception of Asturias, the northern Basque country, Cantabria, Galicia and most of the Pyrenees in the north, which remained largely unoccupied. Arabian and Berber forces then remained in control of most of the Peninsula for more than five centuries, with a subsequent gradual withdrawal toward the southern region of Andalusia driven by the reconquest by Christian forces (*La Reconquista*) (Harvey, 2005). By the end of the 13th century, after almost 800 years of occupation, all Islamic political control ended in Iberia with the fall of the enclave of Granada late December 1492 (Domínguez Ortiz and Vincent, 1979).

Although the Christian reconquest put an end to the Muslim political control in Iberia, people of the Islamic faith were allowed to remain in what is now Spain and Portugal as part of the terms of surrender at Granada in 1492. In fact, it was not until the mid 1520s that the Muslims were forced to convert to Christianity or were expelled from Iberia, and it was not until 114 years later, in 1614, that a final ultimatum was given to the last Muslims residents of *al-Andalus* (the Arabic name given to the region) (Barkai, 1984). Undoubtedly, during the 800 years of Islamic occupation, some degree of bidirectional gene flow between the two communities occurred as interreligious marriages and conversions took place. Further, during the century immediately following the fall of Granada, repatriation of Muslims, to the north of Spain and relocation of Christians to the vacant lands in Andalusia, in the south, provided for additional complexity to this major bidirectional exodus. Specific regions affected by this influx of Muslims that were forced to relocate out of *al Andalus* was the northwest corner of the Peninsula, mountainous and low in population density area, known today as Galicia (Harvey, 2005).

It is thought that approximately half of the Muslim residents of *al-Andalus* selected to convert (Jayyusi, 1994; O'Shea, 2006). Of those that converted, some became practicing Christians adopting new names. Others, the so called crypto-Muslims, were baptized and professed to be Christians but in fact continued practicing the old religion and/or cultural elements of Islam in secret (Stem, 1964; Guettat, 1980; Barkai, 1984; Jayyusi, 1994; Bahrami, 1995; Menocal et al., 2000). Considering the wide geographical extend of the occupation and duration of the Muslim dominion as well as the policy and magnitude of conversions, it is reasonable to expect that the Islamic dominion had a profound impact on the genetic make-up of Iberia. This contention is supported by a recent study of Botigue and colleagues (Botigue et al., 2013) that argues for most of the introgression from North Africa occurring during the past 300 years. Therefore, we hypothesize that the Islamic stay in Iberia affected the genetic constitution of populations in the Peninsula and that due to the policies of relocation of people after the fall of Granada, the distribution of Arab and Berber markers is not uniform within the Iberian Peninsula.

Previous studies have highlighted the genetic characteristics and similarities between the Middle East and northeastern Africa as well as the clear genetic differentiation among Northwest Africa and both Sub-Sahara Africa and Europe, including Iberia. Traditionally, studies accessing the role played by the Strait of Gibraltar tend to indicate that this short but treacherous 14 km stretch of water acted more like a barrier halting bidirectional migration (Cavalli-Sforza et al., 1994), yet more recent work suggests that high rates of gene flow have occurred since pre-Neolithic times (Currat et al., 2010).

Genetic studies indicate that the contemporary populations of Spain are not uniform. The Basque, for example, is characterized by several high frequency genetic markers including Y haplogroup R1b and its derivative R1b1b2 at 87.1% (highest in Western Europe) (Alonso et al., 2005; Balaresque et al., 2010) and the blood groups' Rh negative and O alleles, at 35% and 55% (Cavalli-Sforza et al., 1994; Capelli et al., 2009), respectively, but exhibits one of the lowest frequencies (2%) in Iberia of the Berber marker E1b1b1b1a-M81 on the Y chromosome (Flores et al., 2004). M81 originated in North Africa about 5600 ya and it is thought to signal migrations connected with the Islamic dispersals (Cruciani et al., 2004) and possibly with the Roman and Carthaginian expansions. This marker was found by Flores and collaborators at highest frequencies in Malaga (11.5%), Galicia (10.5%) and Cantabria (8.6%) (Flores et al., 2004). Other investigators have reported commensurate levels of M81 in the provinces of Andalusia (Semino et al., 2004) and Catalonia (Adams et al., 2008), yet reduced levels or no M81 in other regions of Spain (Flores et al., 2004; Semino et al., 2004). E1b1b1c-M123 is another mutation exhibiting geographical partitioning within Iberia. It is thought to have originated in the region of the Near East or Anatolia in the early Neolithic and its dispersal into Europe, the Levant and North Africa mirrors very closely the dissemination of farming during the Neolithic (Semino et al., 2004). Like M81, M123 is seen at its highest frequencies within Iberia in the extreme northwest province of Galicia (5.2%) (Adams et al., 2008), a level as high or higher than in Tunisia (5.2%) and Algeria (3.1%), respectively (Arredi et al., 2004). Much lower levels or no detection have been reported in various regions of central and northern Iberia (Flores et al., 2004; Adams et al., 2008). M123's low frequencies in Western Europe, other than in Iberia, suggest that it may have penetrated the Peninsula from Northern Africa across the Mediterranean Sea and not by land, through continental Europe. Phoenician trading, which flourished from 1200 to 300 BCE could have contributed to the distribution of this marker and its introduction into Iberia. E1b1b1a-M78 is another Y chromosome marker that partitions non-uniformly within Iberia, with the province of Asturias, in the northwest of the Peninsula (just east of Galicia), exhibiting the highest frequency (10.0%). Southern Spain and southern Portugal possess only 3.2% and 4.1%, respectively, of this marker (Cruciani et al., 2007). M78 is thought to have originated in Northeast Africa, specifically in what is now Egypt or Libya (Cruciani et al., 2007) about 17,000–20,000 ya. It is possible that M78 may represent a signature of the Phoenician dominion that spread across the Mediterranean. M78's presence in Iberia dates back to at least the early Neolithic since 7000 year old funeral remains of individuals carrying this mutation were discovered in a Catalonian cave in northeastern Spain (Lacan et al., 2011).

The Y chromosome marker J1a2b-P58 with origin in the Middle East approximately 10,000 ya is associated with the expansion of Semitic herder-hunters into the Arabian Peninsula (Chiaroni et al., 2010). It has been reported that dispersals carrying different J1 subhaplogroups markers entered North Africa in historic times (Semino et al., 2004). Specifically, P58 is tied to the Arabization of Northwest Africa (Ennafaa et al., 2011). J1 subhaplogroups, although highly abundant in Arabia, where it reaches levels of 40%–75%, in Iberia, it has only been detected in Andalusia in southern Spain and at minimal frequency of 1.1% (Semino et al., 2004).

In terms of mtDNA, the North African-specific U6 haplogroup dated to about 50,000 ya has only been observed in the northwest of the Peninsula, with Galicia (2.2%) and northern Portugal (4.3%) exhibiting the highest frequencies (González et al., 2003). Based on this distribution and the higher diversity value (0.014 ± 0.001) in this area compared to North Africa (0.006 ± 0.001), it was suggested that historic events alone, such as the Muslim occupation, cannot be the sole cause of the U6's presence in Iberia and that migrations dating from pre-Neolithic times are also responsible (González et al., 2003). The other maternal lineages that signal North African and Sub-Saharan gene flow belong to macro-haplogroup L. L1 in the Peninsula is attributed to a number of Middle Eastern and Arabic migrations into Northern Africa driven by Phoenician and Arabian occupations, respectively (Cerezo et al., 2012). The highest frequencies of this haplogroup are seen in Cordoba in southern Spain at 8.30% (Casas et al., 2006; Hernández et al., 2014) and Galicia at 3.70% (Achilli et al., 2007) while the lowest are observed