Lactase Persistence in Southern Iberia and Northwestern Africa: New Insights into the Population Structure and History of the Western Mediterranean

Rosario Calderón1, Candela L. Hernández1, Luis J. Sánchez-Martínez1, Javier Cabau1, Andrea Novelletto2, Jean-Michel Dugoujon3, Luisa Pereira4

1Department of Biodiversity, Ecology and Evolution. Faculty of Biology, Complutense University, Madrid, Spain
2Department of Biology, University “Tor Vergata”, Rome, Italy
3CNRS UMR 5288 Laboratory of Molecular Anthropology and Synthesis Imaging University “Paul Sabatier Toulouse III”, Toulouse, France
4i3S, Institute of Research and Innovation in Health, Porto University, Porto, Portugal

ABSTRACT

The development of lactose tolerance in humans represents a fair example of interaction between biology and culture. Lactase persistence (LP) during adulthood is strongly associated with several genetic variants (SNPs) in the MCM6 region (2q21.3). The −13910 C/T is the most widely analysed variant in present-day human populations and its origin has been postulated in central Europe during the Neolithic. To deepen the knowledge on the incidence of LP trait in the western Mediterranean, a number of 470 adult individuals autochthonous from southern Iberia and north-western Africa (Morocco) were genotyped for the SNPs −13907 C/G, −13910 C/T, −13915 T/G, and −14010 G/C related to LP phenotype. Allele and genotype frequencies were estimated as well as proportions of predicted LP phenotype. Besides, contour maps based on allele frequencies were constructed, and a Bayesian Model was implemented in order to analyse the geographic distribution patterns of LP trait and associated mutations across Europe, Africa, Mediterranean Basin and neighbouring areas. Frequencies of the European −13910*T derived allele among southern Iberian populations reached a mean value of 0.418 whereas in Moroccan Berbers figures were comparatively lower: 0.092 (Bouhria), 0.175 (Figuig) and 0.188 (Asni). The −13915*G variant, with a probable origin in the Arabian Peninsula, was observed in the study of Moroccan Berber samples (not in the Iberia Peninsula) with frequencies varying between 0.025 (Asni) and 0.066 (Bouhria). The Bayesian model provided a frequency value for LP trait in the Iberian Peninsula of 0.615 (95% PI: 0.594 – 0.636) whereas in the Maghreb it was 0.313 (95% PI: 0.275 – 0.352). Our results give further support to an European origin for −13910*T allele, and unveils a negative gradient of its frequencies from Iberia towards Maghreb, giving evidence again to the recurrent prehistoric and historical human movements and admixture processes through the Gibraltar Strait and its adjacent coasts.

Key words: Human Evolution, Human Genetic Diversity, LCT, −13910*T, −13915*G, Iberian Peninsula, Andalusia, North-western Africa.

Introduction

The development of lactose tolerance in humans is understood as an evolutionary process with a key cultural dimension. In terms of human evolution, biocultural traits have a potential of strong impact on populations because they allow an adaptive evolution that could be faster than that driven by strictly biological processes1,2.
pastoralist socio-economy or have been in contact with milk production and other dairy products. Therefore, LP in humans represents a fair example of interaction between biology and culture.

The high prevalence of LP phenotype registered in northern European latitudes have been related to low doses of solar radiation as well as to diminished biological capacities to synthesize vitamin D in people living in those areas. Hence, the ability to adequately digest the disaccharide lactose is assumed as selectively advantageous (e.g. protection against rickets) against environmental conditions of low UV radiation. In this context, the contribution of LPH protein synthesized by the LCT gene is well accepted because it allows the use of beneficial products for health, as is the case of flavonoids, conferring adaptive advantages. Nevertheless, controversies exist regarding the ways in which natural selection would have worked on adaptation of lactase activity in human adulthood after weaning.

In Europe, the most common variant responsible for the LP phenotype is the −13910 C/T. However, other independent variants of lactase persistence close to the abovementioned SNP have been identified at high frequencies in Arabian Peninsula (i.e. −13915 T/G) and in certain African pastoralist and non-pastoralist groups (i.e. −13907 C/G, −14010 G/C). In consequence, it can be assumed that LP and its associated genetic variants would be linked to different geographic origins revealing a strong cultural component and convergent evolution. The high correlation observed between geography and LP-associated alleles, turns these markers into important tools for tracking migratory movements and admixture among population groups, being that population dynamic presumably favoured by natural selection.

In the last decade, studies dealing with paleogenomic and archaegenetic of the lactase gene (LCT) have provided evidence on prehistoric movements as well as signals of adaptation of lactase persistence. Molecular analyses on ancient DNA (aDNA) European samples have revealed the presence of the mutant LP −13910*T allele in the continent. In pre-Neolithic Europe, the allele *T seems not to be present yet in appreciable frequencies. The earliest occurrence of the −13910*T has been reported in a Mesolithic individual’s DNA from La Braña site (northern Meseta, Spain) dated to about 7000 years BP. The authors, however, warned about the stated finding as it might be related to aDNA damage processed (cytosine deamination). In Late Neolithic sites (~5000–4300 BP) from northern Spain and central/northern Europe dating ~4000 years BP the −13910*T allele has also been detected.

In the present survey, we report the first data on genetic variants associated to LP trait in southern Iberia and Morocco. Western Mediterranean conceived as the closest physical connection between Europe and North of Africa has played a crucial role for human migrations across continents for the last millennia. Our results offer further hints on migratory events, mainly from Neolithic period, in this crucial geographical enclave of the world, and its impact on the genetic composition and structure of this metapopulation.

Materials and methods

Population samples

LP-associated SNPs were investigated in 470 unrelated, healthy, autochthonous individuals (until the third generation) from western Mediterranean. The overall sample studied has familiar origins in southern Iberia (eastern Andalusia: Granada province, N=170; western Andalusia: Huelva province, N=152; southern Portugal, N=50) and Morocco (Berbers) (Asni, N=40, Bouhria, N=38 and Figuig, N=20). A detailed description of these population groups can be found elsewhere. The written informed consent was collected from each individual who participated in the study. The Bioethics Committee of the Universidad Complutense of Madrid (UCM) approved the research protocols used for this research.

Genomic architecture of LCT and MCM6 genes and SNPs genotyping

Nowadays, the architecture of LCT gene (~50 kilobases (kb) length) located on chromosome 2 (2q21.3) is well known (Figure 1). It encodes the protein of enzymatic nature LPH (Lactase-Phlorizin Hydrolase) responsible, in turn, for catalysing the dissociation reaction of the disaccharide lactose in glucose and galactose. Upstream the LCT gene is positioned the adjacent MCM6 gene (~36 kb length) which represents an enhancer region that affects lactase promoter activity, and thus involved in LCT transcription. MCM6 gene contains at least five SNPs (Single Nucleotide Polymorphisms) or variants that are responsible for the LP phenotype. The MCM6 variants in combination with other additional markers in the region, shape different haplotypes for the LCT gene, whose frequencies in human populations provide useful information to understand interpopulation genetic relationships within and between continents.

The SNPs associated to LP phenotype that have been molecularly characterized in the present study (see Table 1) were performed by means of TaqMan real-time assays (Thermo Fisher Scientific) in a StepOne instrument. All analyses were Custom TaqMan SNP Genotyping Assays, with the exception of SNP −13910 C/T (rs4988235) which was predesigned (reference C_2104745_10).

Genetic data analysis

Allele and genotype frequencies for the LP-associated variants were estimated by direct gene counting. Summary statistics, such as the observed and expected heterozygosity, respectively, Ho and He and F fixation indices were calculated using the GenAlEx package.
A population database for genetic MCM6 variants was constructed by using in PubMed the following Medical Subject Headings (MeSH) terms or keywords: LCT, lactase, lactase persistence, anthropological genetics, Eurasia, Iberian Peninsula, Mediterranean, Africa, Near East. For each selected survey, some punctual information was taken: population and geographic area, sample size, and allele frequencies of target SNPs: $\sim$13907 C/G, $\sim$13910 C/T, $\sim$13915 T/G, and $\sim$14010 G/C as key variants. Besides, geographical coordinates defining specific localities or regions were considered as well. The gathered LP information comprised 114 populations containing 19,440 samples. The dataset was explored at different levels. Firstly, those population genetic studies which only comprised 114 populations containing 19,440 individuals were selected. Besides, we considered those surveys (N= 66 populations; N= 9615 individuals) which provided a more complete view of LP trait, that is, where other LP associated genetic variants were investigated (i.e. $\sim$13907, $\sim$13915, $\sim$14010) (underlined in Supplementary Table 1). The map showing the geographic distribution of populations used here for the sake of comparisons can be found in Supplementary Figure 1.

Contour (surface) maps were built to analyze distribution patterns of LP and some genetic SNP variants known to underlie the LP phenotype in different continents. Following Itan et al., the expected frequencies of LP phenotype were estimated from the total sum of each MCM6 SNP loci genotyped here, containing the derived allele (homozygous and heterozygous genotypes). Contour maps were generated by using the Spatial Analyst extension and the IDW (Inverse Distance Weighted) method in ArcGIS 10.6 software (ESRI, Redlands, CA: Environmental Systems Research Institute). In parallel, the statistical program PASSaGE was used to perform spatial autocorrelation analysis aimed to statistically contrasting the existence of relationships between $\sim$13910*T derived allele frequencies and population topology.

### TABLE 1
MAIN SNPs INTERVENING IN THE REGULATION OF LCT GENE AND ITS INCIDENCE IN SOME WORLDWIDE HUMAN POPULATIONS

<table>
<thead>
<tr>
<th>SNP</th>
<th>rs</th>
<th>Intron</th>
<th>Ancestral state</th>
<th>Derived state</th>
<th>Genetic geography of the derived state allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sim$14010</td>
<td>rs145946881</td>
<td>13</td>
<td>G</td>
<td>C</td>
<td>Sub-Saharan Africa</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>East sub-Saharan Africa, Arabia and Near/Middle East Europe, North of Africa and East sub-Saharan Africa</td>
</tr>
<tr>
<td>$\sim$13915</td>
<td>rs41380347</td>
<td>13</td>
<td>T</td>
<td>G</td>
<td>Sub-Saharan Africa</td>
</tr>
<tr>
<td>$\sim$13910</td>
<td>rs4988235</td>
<td>13</td>
<td>C</td>
<td>T</td>
<td>East sub-Saharan Africa, Arabia and Near/Middle East Europe, North of Africa and East sub-Saharan Africa</td>
</tr>
<tr>
<td>$\sim$13907</td>
<td>rs41525747</td>
<td>13</td>
<td>C</td>
<td>G</td>
<td>Sub-Saharan Africa</td>
</tr>
</tbody>
</table>

### Bayesian Model
In order to shape LP trait frequencies registered in Iberia and Maghreb, a Bayesian Beta-Binomial model has been implemented using the statistical software R 3.5.1. Bayesian methodology uses a likelihood function (data | $\theta$), read “the likelihood of the data given $\theta$”, which is the parameter that describes the probability distribution of the observable data. In the present study, the modelled parameter has been $\theta$ = Frequency of LP trait in Iberian and Maghreb populations. Some common likelihood distributions fit well the normal- or bell-shaped curve, the binomial distribution for binary variables, and the Weibull distribution for event times. Therefore, binomial distribution has been used to model the likelihood (1) (see formula below) since the parameter of interest is a proportion, that is, frequency of individuals with LP trait in a given population:

$$
\text{Data} \mid \theta \sim \text{Binomial} (N, \theta)
$$

where “N” is the number of Bernoulli trials

A Bayesian model also includes a prior probability distribution, prior ($\theta$), to describe previous knowledge on the parameter $\theta$ of probability before observing the data. In the present study the prior ($\theta$) distribution has been modelled with a beta distribution, with parameters $\alpha$ (number of individuals with LP trait) and $\beta$ (number of individuals without LP trait) (2) based on available data from previous studies.

$$
\theta \sim \text{Beta}(\alpha, \beta)
$$

Bayesian statistical methods use the observed data in the ongoing study to learn about the distribution of the $\theta$ parameter by applying Bayes’ Theorem (3), which combines the prior and the likelihood distribution by computing the posterior distribution:

$$
\text{Posterior}(\theta \mid \text{Data}) = \frac{\text{lik}(\text{Data} \mid \theta) \times \text{prior}(\theta)}{\text{prob}(\text{Data})}
$$

The posterior distribution of the parameter of interest ($\theta$), conditioned to the data is a new beta distribution (4), which combines the prior and likelihood distribution.

$$
(\theta \mid \text{Data}) \sim \text{Beta}(\alpha', \beta')
$$

### Results and Discussion
Genotype and allele frequencies for LP- genetic variants and predicted proportions of LP phenotype in southern Iberia and Morocco are shown in Table 2. The frequency distributions of those LP-derived alleles observed in the analysed western Mediterranean metapopulation are represented in Figure 2. The incidence of projected LP phenotype values in southern Iberia is 63% and this condition seems to be strictly associated to the major $\sim$13910
C/T variant, with a mean frequency of 0.418 for the 13910*T-derived allele [range: 0.400 (E. Andalusia: Granada) to 0.480 (southern Portugal)]. Our findings closely agree with those registered in southwestern Europe (~40%, 29,30) which are comparatively lower than that recorded in the northern latitudes of the continent31. The other characterized LP-SNPs −13907 C>G, −13915 T>G and −14010 G>C were absent in the whole studied Iberian sample. By contrast, LP phenotype in north-western Africa (e.g. Morocco 35%, present study) would be interpreted as a combination of two LP-associated variants −13910 C/T and −13915 T/G. The former occurs at much lower frequency (Bourhia: 0.092; Figuig and Asni: 0.180) than in Europeans; the latter (see Table 2) was only present in seven of 91 analysed samples (0.036) being all them heterozygous TG-13915. No other LP-variants were detected. The −13915 T/G is considered a genetic marker of Arabian Peninsula populations. The correlation or concordance between predicted LP trait and −13910*T allele is closer in southern Iberia (0.63 vs 0.42) than in the analysed Moroccan Berber samples (0.35 vs 0.15). (Table 2, Figure 2)

The distribution of −13910 C/T genotypes showed a rather high proportion of lactase-persistence heterozygous for the CT-13910 genotype among southern Iberian samples [western Andalusians (77/152: 50.7%); eastern Andalusians (68/170: 40%) and southern Portugese (16/50: 32%)]. In the case of Berber groups studied here, the relative frequency of CT-13910 genotypes was 27% (Asni), 18% (Bouhria), 35% (Figuig) and only two donors (5%) from Asni were distinguished as homozygous (TT). A high representation of heterozygous genotypes leads to an increase of the expected proportion of LP phenotype (e.g. western Andalusia, 51% vs 68%); in the analysed Morocccan global sample that proportion declined to 28%.

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GENOTYPE AND ALLELE FREQUENCIES FOR −13910 C&gt;T, −13915 T&gt;G, −13907 C&gt;G AND −14010 G&gt;C LP-ASSOCIATED ALLELES AMONG SOUTHERN IBERIAN (ANDALUSIA AND SOUTH PORTUGAL) AND MOROCCAN BERBER POPULATIONS</strong></td>
</tr>
<tr>
<td>Population</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Southern Iberia</td>
</tr>
<tr>
<td>E. Andalusia</td>
</tr>
<tr>
<td>W. Andalusia</td>
</tr>
<tr>
<td>S. Portugal</td>
</tr>
<tr>
<td>Northwestern Africa (Morocco)</td>
</tr>
<tr>
<td>Asni</td>
</tr>
<tr>
<td>Bouhria</td>
</tr>
<tr>
<td>Figuig</td>
</tr>
</tbody>
</table>

**Fig. 1.** Molecular architecture of MCM6 and LCT genomic regions (2q21). Main single nucleotide polymorphisms (on the right) associated with the LP trait are shown in their gene location.
The plot of Figure 2 makes clearly visible how the $-13910^*T$-derived allele is noticeably represented among southern Iberians, as an example of European population. Our results are in agreement with other Spanish samples analysed from Catalonia, Valencia and general population where only were distinguished by the presence of the $-13910 C/T$ variant\(^4,29\). As it can be observed, the $-13915^*G$ allele is present in Berbers and at polymorphic frequencies (>0.01) (i.e. Asni: 0.025; Bouhria: 0.066). Heterozygous genotypes $TG-13915$ were 5% (240) and 13.2% (538), respectively. Bouhria Berbers seem to reveal a more exclusive LP-allelic profile with the lowest and the highest frequencies of $-13910^*T$ (0.092) and $-13915^*G$ (0.066).

Table 3 shows the observed ($H_o$) and expected ($H_e$) heterozygosity as well as fixation indices ($F$) for the LP-associated polymorphisms $-13910 C/T$ and $-13915 T/G$. The expected heterozygosity among eastern Andalusians and southern Portuguese is greater than that observed. This scenario is reversed for western Andalusians; Bouhria and Figuig samples stand out for negative $F$ fixation index as well.

Lactase persistence trait represents a clear example of convergent evolution as different allele variants lead to the same scenario, that is, they give rise to the same LP phenotype in different regions of the world, as it is the case of Europe, Arabian Peninsula and East Africa\(^22,31,32\). A mosaic of surface maps are displayed in Figure 3, where continents or specific continental regions with frequency records of LP trait are explained by the presence of LP-associated SNPs $-13910$ (Figure 3B), $-13915$ (Figure 3C), $-13907$ (Figure 3D) and $-14010$ (Figure 3E) among other variants. All of these representations were sourced from the same population dataset. Lactase persistence phenotype seems to be especially represented in European populations, with the highest frequencies mainly concentrated along northern Atlantic European coasts just where $-13910^*T$-derived allele is highest in frequency as well. Therefore, the $-13910^*T$ would be responsible for most (if not all) of the LP trait\(^32\). The $^*T$ allele, exhibits a negative gradient in the direction northern-southern/eastern of the continent\(^3\). Natural selection would have acted favouring the high prevalence of the $13910^*T$ mutation in northern European latitudes, not only for the caloric intake advantage that represent milk consumption but also by considering milk as a source of vitamin D, and thus facilitating calcium assimilation as alternative to low levels of UV radiation\(^35\). (Figure 3)

Figure 4 presents spatial patterns of variation of the $-13910^*T$ allele across the Mediterranean Basin outlined through a profile shaped by a correlogram. The plot suggests a marked population genetic structure. Moran’s $I$ index of Moran\(^3\) was devised to measure the degree of genetic similarity between populations, and checks whether genetic markers exhibit a well-defined spatial structure for the set of populations used in the correlation analysis. In our study case, the correlogram is distinguished by significant and positive autocorrelation values for short distances and significant and negative autocorrelation values for long distances. Autocorrelation values are expressed in relation to the geographical distances between populations. Our outcomes confer high support to the observed negative gradient for the $-13910^*T$ allele with frequencies decreasing across the Mediterranean space in a western/eastern axis. Within the Italian peninsula, the $-13910^*T$ allele has also revealed an apparent negative gradient from the north (0.310) to the south (0.080)\(^10\). (Figure 4)

The spreading process of $-13910^*T$ allele outside Europe, facilitated by migrations and gene flow could reach easily the Maghreb (mainly Morocco) through the Iberian Peninsula. Frequencies of this major European LP marker are far lower among Maghrebi territories and populations [e.g. Moroccan Berbers, 0.17 (N=5 population samples); Algerian Mozabites, 0.19 (N=1) and North-Central-South Tunisia, 0.11] (N=1)) (see \(^36–38\)). In Saharauis, the $-13910^*T$ occur at frequencies ~26%\(^9\).
The presence of $-13910^T$ allele in the Maghreb (mainly in its western side) and its absence in other surrounding more southern African regions (e.g., in sub-Saharan Africa has not still been reported) would support a scenario in which Maghrebi populations would have shared a dairying origin with southern Europeans. Moreover, the dominantly negative values of $F$ fixation index in some Berber human groups, such as it is shown in Table 3, would indicate a higher observed than expected heterozygosity. These results should be interpreted in the light of the recurrent migratory movements that happened between Iberia and northern Africa from Neolithic and Bronze Ages\(^{40}\). This population dynamic continued in later prehistoric and protohistoric times followed by a steady continuity during the ancient history onwards. Clear similarities in the Neolithization process across the Strait of Gibraltar and surrounding areas have been previously stated\(^{41,42}\). The arrival of the $-13910^T$ allele in northern Africa could be also caused by the establishment of the Roman Empire on both sides of western Mediterranean

---

**Fig. 3.** Contour maps of Lactase Persistence (LP) predicted frequency (A) and LP associated variants: $-13910^T$ (B), $-13915^G$ (C), $-13907^G$ (D) and $-14010^C$ (E) in Europe, Africa and southwestern Asia. Frequencies (%) are shown in a color scale. Red dots indicate populations in which only is available information for the $-13910^T$ variant. See information on the populations set used for comparative analysis in Supplementary Table 1. Iberian and Moroccan samples analyzed in the present study are highlighted with a double circle.
where contacts between the Iberian Peninsula and north-western Africa were particularly intense in that epoch. In consequence, trans-Mediterranean migrations established between southern Europe and the Maghreb from ancient times through would have favoured the dispersion of the \(-13910^T\) allele, and thus leaving signals in the genome of many present-day North African populations\(^{43,44}\). (Table 3)

It has been postulated that the origin of \(-13910^C/T\) variant was likely in Europe and it seems to be related with the Neolithic period\(^{22}\). In Africa and other surrounding non-African areas, the presence of this LP allele (\(-13910^T\)) has been attributed to different evolutionary histories. Estimates for the most common recent ancestor (TMRCA) associated to \(-13910^T\) have provided results in accordance with aDNA analyses, recognizing the role of selective effects in shaping the present-day population frequencies for this polymorphism (selection coefficient, \(s=0.04-0.05\)). In Basque and Finnish populations age estimates were similar, respectively 5150 years BP; 95% CI 4950–5425 and 5000 years BP; 95% CI 4800-5225. In north-western Africans (Morocco) evolutionary ages showed a later temporal frame (3110 years BP; 95% CI 2950-3225); even younger results were obtained when analysing aDNA samples from the Near East and south-western Asia (i.e. Iran 1875 years BP; 95% CI 0-2325)\(^{39}\).

**TABLE 3**

<table>
<thead>
<tr>
<th>Populations (n)</th>
<th>(-13910^C/T)</th>
<th>(-13915^T/G)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(H_o)</td>
<td>(H_e)</td>
</tr>
<tr>
<td>E Andalusia (170)</td>
<td>0.40</td>
<td>0.48</td>
</tr>
<tr>
<td>W Andalusia (152)</td>
<td>0.50</td>
<td>0.48</td>
</tr>
<tr>
<td>South Portugal (50)</td>
<td>0.32</td>
<td>0.49</td>
</tr>
<tr>
<td>Asni (40)</td>
<td>0.27</td>
<td>0.30</td>
</tr>
<tr>
<td>Bouhria (38)</td>
<td>0.18</td>
<td>0.16</td>
</tr>
<tr>
<td>Figuig (20)</td>
<td>0.35</td>
<td>0.28</td>
</tr>
</tbody>
</table>

Fig. 4. Correlogram plot based on Lactase Persistence trait (LP) associated to \(-13910^T\) allele frequencies across Mediterranean space. Similar results are provided using regression model (see upper right corner). The Iberian (in dark blue) and Moroccan (in light blue) population samples analyzed in the present study are highlighted with a double circle.
These findings would be signifying a complex history of \(-13910^* T\) allele in Eurasia related with the Neolithization process and the domestication of cattle.

Another LP-variant \(-13915 \, T/G\), its mutant allele \(*G\) is significantly represented in Arabian Peninsula [e.g. 0.59 in Saudi Arabia\(^{45}\)] (see Figure 3C) and in lesser extent in Africa. Paradoxically in the neighbour Iranian population the \(-13915^* G\) occurs in extremely low frequencies, challenging then the view that LP distribution in Iran resulted from the demic diffusion\(^{46}\). Traces of this LP-SNP variant have also been detected in the Near East, East Africa and North African populations, as is the case of Berber groups studied here.

Presumably, the \(-13915^* T\) was originated in Arabia and widespread towards East Africa within the last 1500 years, as a result of Islamic expansion\(^2,47\) but with a considerably lower influence in north-western Africa. In our studied samples, this variant appears moderately in Moroccan Berbers and is absent in southern Iberia, showing the scarce admixture of the Arabian Peninsula population into the Iberian Peninsula during the Islamic expansion there. Recent studies on mtDNA and genome-wide analysis have pointed out a complex evolutionary history linked to the Arabian Peninsula populations, showing evidence for multiple admixture events from sub-Saharan Africa. Still, the Levantine/European influence and the Islamization process were important in shaping the genetic composition of the region\(^48,49,50\).

The adaptation of lactase persistence is found also in Africa in some particular ethnic groups that have usually practiced herding for generations. Specifically, \(-13907^* C/G\) and \(-14010^* G/C\) variants (see Figure 3D, 3E), the least investigated in terms of population genetics, have a clear focus in Eastern Africa, specifically in Ethiopia, Kenya, Tanzania and Sudan\(^51\). Positive selection could have had a relevant role in maintaining these LP-associated variants in some pastoralist groups, characterized by milk consumption\(^9\). In this line, it should be noted that there is evidence of an independent development of agriculture – outside of the Levantine focus- in several geographic areas of the Sahara/Sahel, Ethiopia, western Africa and the Nile Valley\(^7\).

Results concerning Beta Binomial model for the LP trait frequency in the Iberian Peninsula and Maghreb are presented in Figures 5A and 5B, respectively. Inferences obtained for these two geographic areas are notably different when recorded values are analysed on the x-axis. The point frequency of LP trait in Iberian populations (N= 9), estimated from the posterior distribution, was 0.615 (95% PI: 0.594 – 0.636) whereas in the Maghreb declined to 0.313 (95% PI: 0.275 – 0.352). Curiously, we still observed how distributions drawn from our own data (blue dashed lines) are displaced to the right, revealing slightly higher point estimates than the ones calculated for all Iberia and Maghreb. (Figure 5)

Using the same DNA sample set than in the present work, results emerging from the molecular characterization of the G/A point mutation \(IVS\, I-1\), located in the intron 1 of the \(\beta\)-globin gene (HBB gene) and originating \(\beta\)-Thalassemia, we observed similar patterns than that exhibited by the \(-13910^* T\) allele in the western Mediter-
The logical influence of the Arabian Peninsula populations was, which contrasts with the reduced biological influence of the Arabian Peninsula populations into the Iberian genome.

In summary, the lactase persistence in the region around the Strait of Gibraltar shows how notable and prolonged gene flow between the Iberian and Maghrebi populations was, which contrasts with the reduced biological influence of the Arabian Peninsula populations into the Iberian genome.

Acknowledgements

We thank to the participants who provided their blood samples in this study. The Spanish Ministry of Economy and Competitiveness has financed this work through a research project granted to RC (CGL2014-53985-R). CLH was supported by a research contract during three years linked to this Project. I3S is financed by FEDER – Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020 – Competitiveness and Internationalization Operational Programme (POCI), Portugal 2020, and by Portuguese funds through FCT/ Ministério da Ciência, Tecnologia e Inovação in the framework of the project “Institute for Research and Innovation in Health Sciences” (POCI-01-0145-FEDER-007274).

References

PERZISTENCIJA LAKTAZE NA JUGU PIRINEJSKOG POLUOTOKA I U SJEVEROZAPADNOJ AFRICI: NOVI UVIDI U POPULACIJSKU STRUKTURU I POVIJEST ZAPADNOG SREDOZEMLJA

SAŽETAK

Razvoj tolerancije na laktozu kod ljudi predstavlja dobar primjer interakcije između biologije i kulture. Postojanost laktaze (PL) u odrasloj dobi izrazito je povezana s nekoliko genetskih varijanti (SNP) u MCM6 području (2q21.3). U suvremenim ljudskim populacijama najviše je istraživana varijanta –13910 C/T, a njezino podrijetlo je vjerojatno u srednjoj Europi tijekom neolitika. Kako bi produbili znanje o učestalosti PL-a na zapadnom Sredozemlju, 470 odraslih osoba porijeklom sa južnog dijela Pirinejskog popuotoka i iz sjeverozapadne Afrike (Maroko) genotipizirano je za varijante SNP –13907 C / G, –13910 C / T, –13915 T / G i –14010 G / C koje se odnose na PL fenotip. Procijenjene su frekvencije alela i genotipa, kao i stupanj prediktivne valjanosti PL fenotipa. Osim toga, izrađene su okvirne karte temeljene na frekvencijama alela i pomoću Bayesovog modela analizirana je geografska distribucija LP obilježja i povezanih mutacija u Europi, Africi, Sredozemnom bazenu i susjednim područjima. Učestalost europskog alela –13910 * T je među populacijama južnog Pirinejskog poluotoka dosegla srednju vrijednost od 0,418 dok su vrijednosti među marokanskim Berberima bile niže: 0,092 (Bouhria), 0,175 (Figuig) i 0,188 (Asni). Varijanta –13915 * G, koja vjerojatno potječe s Arapskog poluotoka, uočena je u istraživanju marokanskih Berbera (ne na Pirinejskom poluotoku) s frekvencijama koje variraju između 0,025 (Asni) i 0,066 (Bouhria). Bayesovim modelom dobivena je vrijednost frekvencije za PL na Pirinejskom poluotoku od 0,615 (95% PI: 0,594–0,636) a na Magrebu 0,313 (95% PI: 0,275–0,352). Naši rezultati dodatno podupiru tezu o europskom podrijetlu alela –13910 * T i ukazuju na gradjeno opadanje njegovih frekvencija od Pirinejskog poluotoka prema Magrebu, pružajući nove dokaze o stalnim prapovijesnim i povijesnim kretanjima i miješanjima ljudskih populacija u Gibraltarskom tjesnacu i obalnim područjima oko njega.