

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

⁴IS, 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 processes^{1,2}.

Lactase persistence (LP) trait or lactose tolerance exhibits non-random spatial variation patterns in present-day human populations. In north-western Europeans, the LP prevalence varies between 75–95% decreasing in gradient towards the south and east of the continent (among others see^{3,4}). High LP frequencies are also recorded in other continents but in specific local regions (e.g. Africa) where some population groups either have developed a

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^a 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^{5, 6}.

In Europe, the most common genetic 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*)⁷. 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^{2, 8}. 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^{3, 9}.

In the last decade, studies dealing with paleogenomic and archaeogenetic of the lactase gene (*LCT*) have provided evidence on prehistoric movements as well as signals of adaptation of lactase persistence¹⁰. 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^{11–13}. 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 BP¹⁴. 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 Spain¹⁰ and central/northern Europe dating ~4000 years BP^{12, 15, 16} 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^{17–20}. 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 phenotype². 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^{21, 22}. (Figure 1)

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 counting²³. Summary statistics, such as the observed and expected heterozygosity, respectively, H_o and H_e and F fixation indices were calculated using the *GenAlEx* package²⁴.

^a Genetics Home Reference, NIH, National Library of Medicine, <https://ghr.nlm.nih.gov/gene/LCT>

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: *-13907 C/G*, *-13910 C/T*, *-13915 T/G*, and *-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 analysed the *-13910 C/T* variant ($N=114$ populations; $N=19,440$ 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. *-13907*, *-13915*, *-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 analyse distribution patterns of LP and some genetic SNP variants known to underlie the LP phenotype in different continents. Following Itan et al.²⁵, 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* 2²⁶ was used to perform spatial autocorrelation analysis aimed to statistically contrasting the existence of relationships between *-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

SNP	rs	Intron	Ancestral state	Derived state	Genetic geography of the derived state allele
-14010	rs145946881	13	G	C	Sub-Saharan Africa
-13915	rs41380347	13	T	G	East sub-Saharan Africa, Arabia and Near/Middle East
-13910	rs4988235	13	C	T	Europe, North of Africa and East sub-Saharan Africa
-13907	rs41525747	13	C	G	Sub-Saharan Africa

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 | θ), read “the likelihood of the data given θ ”, which is the parameter that describes the probability distribution of the observable data. In the present study, the modelled parameter has been θ = 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^{27,28}. 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:

$$\text{Data} \mid \theta \sim \text{Binomial}(N, \theta) \quad (1)$$

where “N” is the number of Bernoulli trials

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

$$\theta \sim \text{Beta}(\alpha, \beta) \quad (2)$$

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

$$\text{Posterior}(\theta \mid \text{Data}) = \frac{\text{lik}(\text{Data} \mid \theta) \times \text{prior}(\theta)}{\text{prob}(\text{Data})} \quad (3)$$

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

$$(\theta \mid \text{Data}) \sim \text{Beta}(\alpha', \beta') \quad (4)$$

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 *-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; Figui 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 popula-

tions. 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 Portuguese (16/50: 32%)]. In the case of Berber groups studied here, the relative frequency of *CT-13910* genotypes was 27% (Asni), 18% (Bouhria), 35% (Figui) 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 Moroccan global sample that proportion declined to 28%.

TABLE 2

GENOTYPE AND ALLELE FREQUENCIES FOR *-13910 C>T*, *-13915 T>G*, *-13907 C>G* AND *-14010 G>C* LP-ASSOCIATED ALLELES AMONG SOUTHERN IBERIAN (ANDALUSIA AND SOUTH PORTUGAL) AND MOROCCAN BERBER POPULATIONS

Population	N	LP	<i>-13910 C>T</i>				<i>-13915 T>G</i>				<i>-13907 C>G</i>				<i>-14010 G>C</i>			
			CC	CT	TT	*T	TT	TG	GG	*G	CC	CG	GG	*G	GG	GC	CC	*C
Southern Iberia	372	0.63	136	161	75	0.418	372	0	0	0	372	0	0	0	372	0	0	0
E. Andalusia	170	0.61	68	68	34	0.400	170	0	0	0	170	0	0	0	170	0	0	0
W. Andalusia	152	0.67	50	77	25	0.418	152	0	0	0	152	0	0	0	152	0	0	0
S. Portugal	50	0.64	18	16	16	0.480	50	0	0	0	50	0	0	0	50	0	0	0
Northwestern Africa (Morocco)	98	0.35	71	25	2	0.148	91	7	0	0.036	98	0	0	0	98	0	0	0
Asni	40	0.37	27	11	2	0.188	38	2	0	0.025	40	0	0	0	40	0	0	0
Bouhria	38	0.31	31	7	0	0.092	33	5	0	0.066	38	0	0	0	38	0	0	0
Figui	20	0.35	13	7	0	0.175	20	0	0	0	20	0	0	0	20	0	0	0

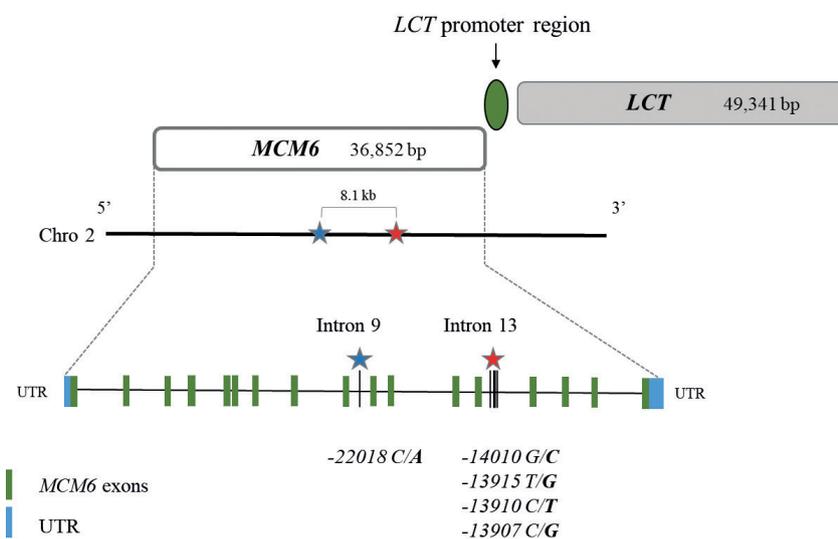


Fig. 1. Molecular architecture of *MCM6* and *LCT* genomic regions (2q21). Main single nucleotide polymorphisms (on the right) associated with the LP trait are shown in their gene location.

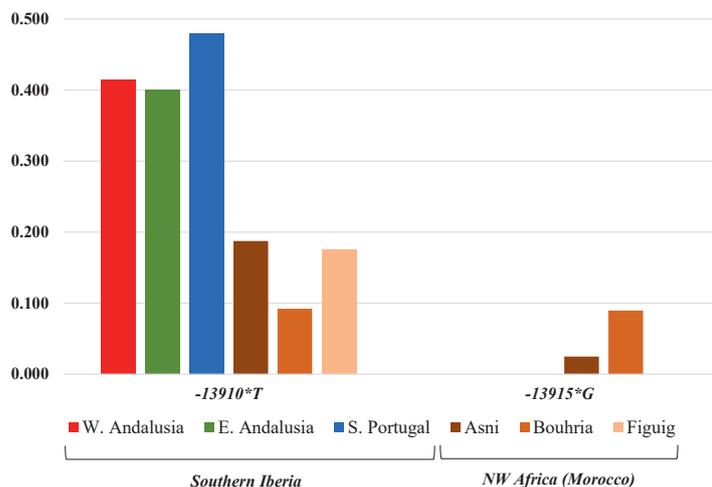


Fig. 2. Allele frequencies for the derived states of -13910^*T and -13915^*G alleles in the MCM6 genomic region associated to LP trait in the western Mediterranean.

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% (2/40) and 13.2 % (5/38), 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 trait³². The $*T$ allele, exhibits a negative gradient in the direction northern-southern/eastern of the continent³. 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 radiation³³. (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 Moran³⁴ 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)³⁵. (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 $\sim 26\%$ ³⁹.

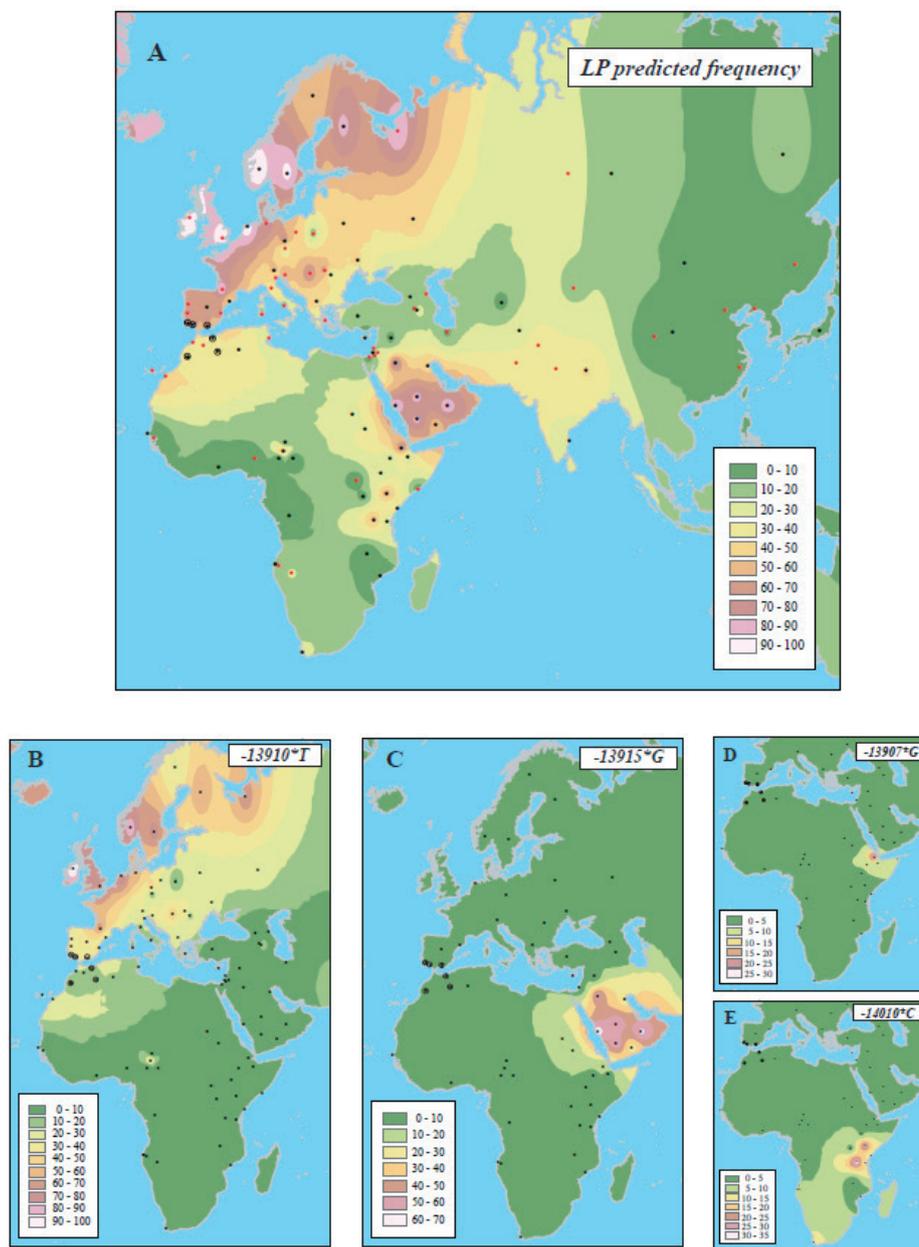


Fig. 3. Contour maps of Lactase Persistence (LP) predicted frequency (A) and LP associated variants: -13910^*T (B), -13915^*G (C), -13907^*G (D) and -14010^*C (E) in Europe, Africa and southwestern Asia. Frequencies (%) are shown in a color scale. Red dots indicate populations in which only is available 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.

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 re-

current migratory movements that happened between Iberia and northern Africa from Neolithic and Bronze Ages⁴⁰. This population dynamic continued in later pre-historic 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

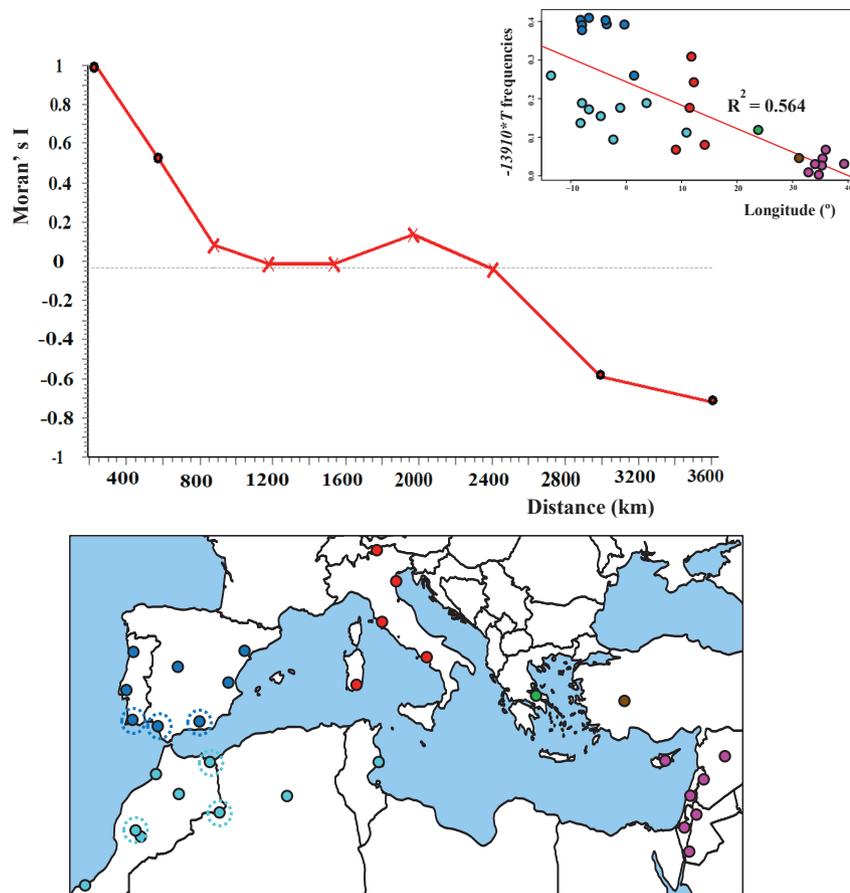


Fig. 4. Correlogram plot based on Lactase Persistence trait (LP) associated to -13910^*T allele frequencies across Mediterranean space. Similar results are provided using regression model (see upper right corner). The Iberian (in dark blue) and Moroccan (in light blue) population samples analyzed in the present study are highlighted with a double circle.

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 period²². 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

TABLE 3

OBSERVED (H_o) AND EXPECTED HETEROZYGOSITY (H_e) AND FIXATION INDEX (F) FOR THE LP VARIANTS $-13910 C/T$ AND $-13915 G/T$ IN SOUTHERN IBERIAN AND MOROCCAN BERBER POPULATIONS

Populations (n)	$-13910 C/T$			$-13915 T/G$		
	H_o	H_e	F	H_o	H_e	F
E Andalusia (170)	0.40	0.48	0.167	0	0	–
W Andalusia (152)	0.50	0.48	–0.041	0	0	–
South Portugal (50)	0.32	0.49	0.359	0	0	–
Asni (40)	0.27	0.30	0.097	0.05	0.04	–0.026
Bouhria (38)	0.18	0.16	–0.101	0.13	0.12	–0.070
Figuig (20)	0.35	0.28	–0.212	0	0	–

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.

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 Arabia⁴⁵] (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 diffusion⁴⁶. 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 Sudan⁵¹. Positive selection could have had a relevant role in maintaining these LP-associated variants in some pastoralist groups, characterized by milk consumption⁹. 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 Valley⁷.

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 β -globin gene (HBB gene) and originating β -Thalassemia, we observed similar patterns than that exhibited by the -13910^*T allele in the western Mediter-

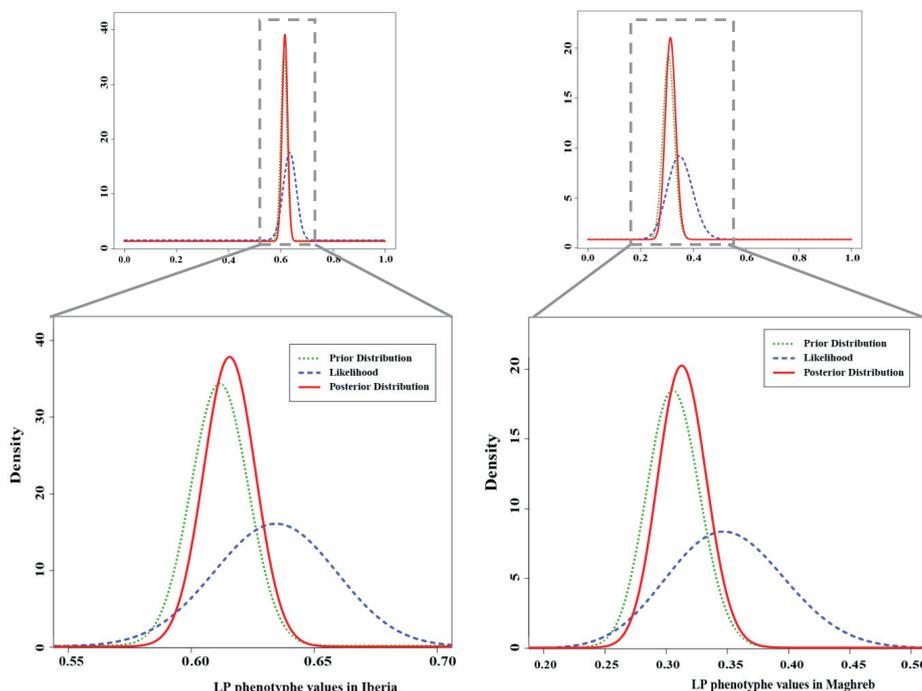


Fig. 5. Distribution curves of the Bayesian Beta-Binomial model of LP phenotype values in Iberia ($n_1= 10$ populations; $n_2= 2122$ individuals) and Maghreb ($n_1= 9$ populations; $n_2= 552$ individuals). The posterior distributions are presented in red color.

reanean (Calderón et al. unpublished data). Interestingly, the incidence of β -Thalassemia *IVS I-1* and *-13910*T* LP-associated allele are found in appreciable, similar frequencies in southern Iberia: 0.410 and 0.418, respectively. This condition might be extrapolated to the neighbouring Morocco, where the *IVS I-1* allele yielded a value of 0.069, comparable in order of magnitude to that observed by the *-13910*T* allele (0.148). This scenario would give further evidence to the close and long-standing relationships between Iberian Peninsula and north-western Africa.

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.

REFERENCES

1. LUCA F, PERRY GH, DI RIENZO A, *Ann Rev Nutr*, 21 (2010) 291. DOI: 10.1146/annurev-nutr-080508-141048. — 2. RANCIARO A, CAMPBELL MC, HIRBO JB, KO WY, FROMENT A, ANAGNOSTOU P, KOTZE MJ, IBRAHIM M, NYAMBO T, OMAR SA, TISHKOFF SA, *Hum Genet*, 94 (2014) 496. DOI: 10.1016/j.ajhg.2014.02.009. — 3. INGRAM CJ, MULCARE CA, ITAN Y, THOMAS MG, SWALLOW DM, *Hum Genet*, 124 (2009) 579. DOI: 10.1007/s00439-008-0593-6. — 4. LIEBERT A, LÓPEZ S, JONES BL, MONTALVA N, GERBAULT P, LAU W, THOMAS MG, BRADMAN N, MANIATIS N, SWALLOW DM, *Hum Genet*, 136 (2017) 1445. DOI: 10.1007/s00439-017-1847-y. — 5. HEYER E, BRAZIER L, SÉGUREL L, HEGAY T, AUSTERLITZ F, QUINTANA-MURCI L, GEORGES M, PASQUET P, VEUILLE M, *Hum Biol*, 83 (2011) 379. DOI: 10.3378/027.083.0304. — 6. GERBAULT P, ROFFET-SALQUE M, EVERSHED RP, THOMAS MG, *IUBM Life*, 65 (2013) 983. DOI: 10.1002/iub.1227. — 7. SCHLEBUSCH CM, JAKOBSSON M, *Ann Rev Genomics Hum Genet*, 31 (2018) 405. DOI: 10.1146/annurev-genom-083117-021759. — 8. SÉGUREL L, BON C, *Ann Rev Genomics Hum Genet*, 18 (2017) 297. DOI: 10.1146/annurev-genom-091416-035340. — 9. FRIEDRICH DC, SANTOS SE, RIBEIRO DOS-SANTOS ÁK, HUTZ MH, *PLoS One*, 7 (2012) 9. DOI: 10.1371/journal.pone.0046520. — 10. PLANTINGA TS, ALONSO S, IZAGIRRE N, HERVELLA M, FREGEL R, VAN DER MEER JW, NETEA MG, DE LA RÚA C, *Eur J Hum Genet*, 20 (2012) 778. DOI: 10.1038/ejhg.2011.254. — 11. GAMBA C, JONES ER, TEASDALE MD, MCCLAUGHLIN RL, GONZALEZ-FORTES G, MATTIANGELI V, DOMBORÓCZKIL, KÓVÁRI I, PAP I, ANDERSA, WHITTLE A, DANI J, RACZYK P, HIGHAM TF, HOFREITER M, BRADLEY DG, PINHASI R, *Nat Commun*, 21 (2014) 5257. DOI: 10.1038/ncomms6257. — 12. ALLENTOFT ME, SIKORAM, SJÖGREN KG, RASMUSSEN S, RASMUSSEN M, STENDERUP J, DAMGAARD PB, SCHROEDER H, AHLSTRÖM T, VINNER L, MALASPINAS AS, MARGARYANA, HIGHAM T, CHIVALL D, LYNNERUP N, HARVIG L, BARON J, DELLA CASA P, DĄBROWSKI P, DUFFY PR, EBEL AV, EPIMAKHOV A, FREI K, FURMANEK M, GRALAK T, GROMOV A, GRONKIEWICZ S, GRUPE G, HAJDU T, JARYSZ R, KHARTANOVICH V, KHOKHLOV A, KISS V, KOLÁŘ J, KRIŠKA A, LASAKI, LONGHI C, MCGLYNN G, MERKEVICIUS A, MERKYTEI, METSPALU M, MKRTCHYAN R, MOISEYEV V, PAJA L, PÁLFI G, POKUTTA D, POSPIESZNY E, PRICE TD, SAAG L, SABLIN M, SHISHLINA N, SMRČKA V, SOENOV VI, SZEVEŘÉNYI V, TÓTH G, TRIFANOVA SV, VARUL L, VICZE M, YEPISKOPOSYAN L, ZHITENEV V, ORLANDO L, SICHERITZ-PONTÉN T, BRUNAK S, NIELSEN R, KRISTIANSEN K, WILLERSLEV E, *Nature*, 522 (2015) 167. DOI: 10.1038/nature14507. — 13. JONES BL, OLJIR, T, LIEBERT A, ZMARZ P, MONTALVA N, TAREKEYN A, EKONG R, THOMAS MG, BEKELE E, BRADMAN N, SWALLOW DM, *Hum Genet*, 134 (2015) 917. DOI: 10.1007/s00439-015-1573-2. — 14. OLALDE I, ALLENTOFT ME, SÁNCHEZ-QUINTO F, SANTPERE G, CHIANG CW, DEGIORGIO M,

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 – Fondo Europeo de Desarrollo 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-FED-ER-007274).

PRADO-MARTINEZ J, RODRÍGUEZ JA, RASMUSSEN S, QUILEZ J, RAMÍREZ O, MARIGORTA UM, FERNÁNDEZ-CALLEJO M, PRADAME, ENCINAS JM, NIELSEN R, NETEA MG, NOVEMBRE J, STURM RA, SABETI P, MARQUÈS-BONET T, NAVARRO A, WILLERSLEV E, LALUEZA-FOX C, *Nature*, 507 (2014) 225. DOI: 10.1038/nature12960. — 15. HAAK W, LAZARIDIS I, PATTERSON N, ROHLAND N, MALLICK S, LLAMAS B, BRANDT G, NORDENFELT S, HARNEY E, STEWARDSON K, FU Q, MITTNIKA, BÁNFFY E, ECONOMOU C, FRANCKEN M, FRIEDERICH S, PENA RG, HALLGREN F, KHARTANOVICH V, KHOKHLOV A, KUNST M, KUZNETSOV P, MELLER H, MOCHALOV O, MOISEYEV V, NICKLISCH N, PICHLER SL, RISCH R, ROJO GUERRA MA, ROTH C, SZÉCSÉNYI-NAGY A, WAHL J, MEYER M, KRAUSE J, BROWN D, ANTHONY D, COOPER A, ALT KW, REICH D, *Nature*, 522 (2015) 207. DOI: 10.1038/nature14317. — 16. MATHIESON I, LAZARIDIS I, ROHLAND N, MALLICK S, PATTERSON N, ROODENBERG SA, HARNEY E, STEWARDSON K, FERNANDES D, NOVAK M, SIRAK K, GAMBA C, JONES ER, LLAMAS B, DRYOMOV S, PICKRELL J, ARSUAGA JL, DE CASTRO JM, CARBONELL E, GERITSEN F, KHOKHLOV A, KUZNETSOV P, LOZANOM, MELLER H, MOCHALOV O, MOISEYEV V, GUERRA MA, ROODENBERG J, VERGÈS JM, KRAUSE J, COOPER A, ALT KW, BROWN D, ANTHONY D, LALUEZA-FOX C, HAAK W, PINHASI R, REICH D, *Nature*, 528 (2015) 499. DOI: 10.1038/nature16152. — 17. PEREIRA L, CUNHA C, ALVES C, AMORIM A, *Human Biol*, 77 (2005) 213. — 18. COUDRAY C, OLIVIERI A, ACHILLI A, PALA M, MELHAOUI M, CHERKAOUI M, EL-CHENNAWI F, KOSSMANN M, TORRONI A, DUGOUJON JM, *Ann Hum Genet*, 73 (2009) 196. DOI: 10.1111/j.1469-1809.2008.00493.x. — 19. AMBROSIO B, HERNÁNDEZ C, NOVELLETTO A, DUGOUJON JM, RODRÍGUEZ JN, CUESTA P, FORTES-LIMA C, CALDERÓN R, *Coll Antropol*, 34 (2010) 1215. — 20. HERNÁNDEZ CL, SOARES P, DUGOUJON JM, NOVELLETTO A, RODRÍGUEZ JN, RITO T, OLIVEIRA M, MELHAOUI M, BAALI A, PEREIRA L, CALDERÓN R, *PLoS One*, 10 (2015) 10. DOI: 10.1371/journal.pone.0139784. — 21. HOLLOX EJ, POULTER M, ZVARIK M, FERAQ V, KRAUSE A, JENKINS T, SAHA N, KOZLOV AI, SWALLOW DM, *Am J Hum Genet*, 68 (2001) 160. DOI: 10.1086/316924. — 22. TISHKOFF SA, REED FA, RANCIARO A, VOIGHT BF, BABBITT CC, SILVERMAN JS, POWELL K, MORTENSEN HM, HIRBO JB, OSMAN M, IBRAHIM M, OMAR SA, LEMA G, NYAMBO TB, GHORI J, BUMPSTEAD S, PRITCHARD JK, WRAY GA, DELOUKAS P, *Nat Genet*, 39 (2007) 31. DOI: 10.1038/ng1946. — 23. HARTL D, CLARK A, *Principles of Population Genetics* (Sinauer Associates, Sunderland, 2006). — 24. PEAKALL R, SMOUSE PE, *Bioinformatics*, 28 (2012) 2537. DOI: 10.1093/bioinformatics/bts460. — 25. ITAN Y, JONES BL, INGRAM CJ, SWALLOW DM, THOMAS MG, *Evol Biol*, 10 (2010) 16. DOI: 10.1186/1471-2148-10-36. — 26. ROSENBERG M, ANDERSON C, *Methods Ecol Evol*, 2 (2011) 229. DOI: 10.1111/j.2041-210X.2010.00081.x.

- 27. LEE PM, *Bayesian Statistics: An introduction* (Wiley, New York, 2012). — 28. GELMAN A, CARLIN JB, STERN HS, RUBIN DB, *Bayesian Data Analysis* (Chapman and Hall, London, 2014). — 29. CORELLA D, ARREGUI M, COLTELL O, PORTOLÉS O, GUILLEM-SÁIZ P, CARRASCO P, SORLÍ JV, ORTEGA-AZORÍN C, GONZÁLEZ JI, ORDOVÁS JM, *Obesity*, 19 (2011) 1707. DOI: doi: 10.1038/oby.2010.320. — 30. MANCO L, PIRES S, LOPES AI, FIGUEIREDO I, ALBUQUERQUE D, ALVAREZ M, ROCHA J, ABADE A, *Ann Hum Biol*, 40 (2013) 205. DOI: 10.3109/03014460.2012.754943. — 31. MULCARE CA, *The evolution of the lactase persistence phenotype*. MS Thesis. (University College London, London, 2006). — 32. STONEKING M, *An Introduction to Molecular Anthropology* (Wiley Blackwell, New Jersey, 2017). — 33. GERBAULT P, MORET C, CURRAT M, SANCHEZ-MAZAS A, 2009. Impact of selection and demography on the diffusion of lactase persistence. *PLoS One*. 4(7):e6369. — 34. MORAN P, *Biometrika*, 37 (1950) 17. — 35. DE FANTI S, SAZZINI M, GIULIANI C, FRAZZONI F, SARNO S, BOATTINI A, MARASCO E, MANTOVANI V, FRANCESCHI C, MORAL P, GARAGNANI P, LUISELLI D, *Am J Phys Anthropol*, 158 (2015) 708. DOI: 10.1002/ajpa.22814. — 36. BERSAGLIERI T, SABETI PC, PATTERSON N, VANDERPLOEG T, SCHAFFNER SF, DRAKE JA, RHODES M, REICH DE, HIRSCHHORN JN, 2004. Genetic Signatures of Strong Recent Positive Selection at the Lactase Gene. *Am J Hum Genet*. 74(6):1111-20. — 37. MYLES S, BOUZEKRI N, HAVERFIELD E, CHERKAOUI M, DUGOUJON JM, WARD R, *Hum Genet*, 117 (2005) 34. DOI: 10.1007/s00439-005-1266-3. — 38. BEN HALIMA Y, KEFI R, SAZZINI M, GIULIANI C, DE FANTI S, NOUALI C, NAGARA M, MENGOZZI G, ELOUEJ S, ABID A, JAMOUSHI H, CHOUCANE L, ROMEO G, ABDELHAK S, LUISELLI D, *Genes Nutr*, 12 (2017) 20. DOI: 10.1186/s12263-017-0573-3. — 39. ENATTAH NS, TRUDEAU A, PIMENOFF V, MAIURI L, AURICCHIO S, GRECO L, ROSSI M, LENTZE M, SEO JK, RAHGOZAR S, KHALIL I, ALIFRANGIS M, NATAH S, GROOP L, SHAAT N, KOZLOVA A, VERSCHUBSKAYA G, COMAS D, BULAYEVA K, MEHDI SQ, TERWILLIGER JD, SAHI T, SAVILAHTI E, PEROLA M, SAJANTILA A, JARVELA I, PELTONEN L, *Am J Hum Genet*, 81 (2007) 615. DOI: 10.1086/520705. — 40. VEERAMAH KR, *Curr Opin Genet Dev*, 53 (2018) 83. — 41. MANEN C, MARCHAND G, CARVALHO AF, *Le Néolithique ancien de la péninsule Ibérique: vers une nouvelle évaluation du mirage africain?* In: *Actes du XXVI congrès préhistorique de France*, Avignon 3, 133–151 (Société Préhistorique Française, 2007). — 42. LINSTÄDTER J, MEDVED I, SOLICH M, WENIGER GC, *Quaternary International*, 274 (2012) 219. DOI: 10.1016/j.quaint.2012.01.013. — 43. OTTONI C, PRIMATIVO G, HOOSHIAR-KASHANI B, ACHILLI A, MARTÍNEZ-LABARGA C, BIONDI G, TORRONI A, RICKARDS O, *PLoS One*, 5 (2010) 10. DOI: 10.1371/journal.pone.0013378. — 44. HERNÁNDEZ CL, DUGOUJON JM, NOVELLETO A, RODRÍGUEZ JN, CUESTA P, CALDERÓN R, *BMC Genetics*, 19 (2017) 46. DOI: 10.1186/s12863-017-0514-6. — 45. IMTIAZ F, SAVILAHTI E, SARNES-TO A, TRABZUNID, AL-KAHTANI K, KAGEVI I, RASHED MS, MEYER BF, JÄRVELÄ I, *J Med Genet*, 44 (2007) 10. DOI: 10.1136/jmg.2007.051631. — 46. CHARATI H, PENG MS, CHEN W, YANG XY, JABBARI ORI R, AGHAJANPOUR-MIR M, ESMAILIZADEH A, ZHANG YP, *Hum Genomics*, 13 (2019) 7. DOI: 10.1186/s40246-019-0195-5. — 47. ENATTAH NS, JENSEN TG, NIELSEN M, LEWINSKI R, KUOKKANEN M, RASINPERA H, EL-SHANTI H, SEO JK, ALIFRANGIS M, KHALIL IF, NATAH A, ALI A, NATAH S, COMAS D, MEHDI SQ, GROOP L, VESTERGAARD EM, IMTIAZ F, RASHED MS, MEYER B, TROELSEN J, PELTONEN L, *Am J Human Genet*, 82 (2008) 57. DOI: 10.1016/j.ajhg.2007.09.012. — 48. AL-ABRI AR, AL-RAWAS O, AL-YAHYAE S, AL-HABORI M, AL-ZUBAIRI AS, BAYOUMI R, *Hum Biol*, 84 (2012) 271. DOI: 10.3378/027.084.0310. — 49. ČERNÝ V, ČÍŽKOVÁ M, POLONI ES, AL-MEERI A, MULLIGAN CJ, *Am J Phys Anthropol*, 159 (2018) 607. DOI: 10.1002/ajpa.22920. — 50. FERNANDES V, BRUCATO N, FERREIRA JC, PEDRO N, CAVADAS B, RICAUT FX, ALSHAMALI F, PEREIRA L, *Mol Biol. Evol*, 36 (2019) 575. — 51. TISHKOFF, SA, REED, FA, FRIEDLAENDER, FR, EHRET, C, RANCIARO, A, FROMENT, A, HIRBO, JB, AWOMOYI, AA, BODO, JM, DOUMBO, O, IBRAHIM, M, JUMA TA, KOTZE MJ, LEMA G, MOORE JH, MORTENSEN H, NYAMBO TB, OMAR SA, POWELL K, PRETORIUS GS, SMITH MS, THERA MA, WAMBEBE C, WEBER JL, WILLIAMS SM, *Science*, 324 (2009) 1035. DOI: 10.1126/science.1172257.

R. Calderón

Faculty of Biology, Complutense University, Avda. de Séneca, 2 Ciudad Universitaria 28040 Madrid, Spain
e-mail: rcalfer@bio.ucm.es

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 poluotoka 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 gradijentno 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.

