

Direct Estimates of Natural Selection in Iberia Indicate Calcium Absorption Was Not the Only Driver of Lactase Persistence in Europe

Oddný Ósk Sverrisdóttir,^{*1} Adrian Timpson,^{2,3} Jamie Toombs,⁴ Cecile Lecoeur,^{5,6} Philippe Froguel,^{5,6,7} Jose Miguel Carretero,⁸ Juan Luis Arsuaga Ferreras,⁹ Anders Götherström,^{1,10} and Mark G. Thomas²

¹Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

²Research Department of Genetics, Evolution and Environment, University College London, London, United Kingdom

³Institute of Archaeology, University College London, London, United Kingdom

⁴Department of Molecular Neuroscience, Institute of Neurology, University College London, London, United Kingdom

⁵European Genomic Institute for Diabetes (EGID), FR3508, Lille, France

⁶CNRS UMR8199, Biology Institute of Lille 1, Lille, France

⁷Department of Genomics of Common Disease, School Of Public Health, Imperial College London, London, United Kingdom

⁸Laboratorio de Evolución Humana, Edificio I+D+i, Universidad de Burgos, Burgos, Spain

⁹Paleontology Department, Universidad Complutense de Madrid, Madrid, Spain

¹⁰Archaeological Research Laboratory, Stockholm University, Stockholm, Sweden

*Corresponding author: E-mail: oddny.o.sverrisdottir@ebc.uu.se.

Associate editor: Beth Shapiro

Abstract

Lactase persistence (LP) is a genetically determined trait whereby the enzyme lactase is expressed throughout adult life. Lactase is necessary for the digestion of lactose—the main carbohydrate in milk—and its production is downregulated after the weaning period in most humans and all other mammals studied. Several sources of evidence indicate that LP has evolved independently, in different parts of the world over the last 10,000 years, and has been subject to strong natural selection in dairying populations. In Europeans, LP is strongly associated with, and probably caused by, a single C to T mutation 13,910 bp upstream of the lactase (*LCT*) gene (-13,910*T). Despite a considerable body of research, the reasons why LP should provide such a strong selective advantage remain poorly understood. In this study, we examine one of the most widely cited hypotheses for selection on LP—that fresh milk consumption supplemented the poor vitamin D and calcium status of northern Europe's early farmers (the *calcium assimilation hypothesis*). We do this by testing for natural selection on -13,910*T using ancient DNA data from the skeletal remains of eight late Neolithic Iberian individuals, whom we would not expect to have poor vitamin D and calcium status because of relatively high incident UVB light levels. None of the eight samples successfully typed in the study had the derived T-allele. In addition, we reanalyze published data from French Neolithic remains to both test for population continuity and further examine the evolution of LP in the region. Using simulations that accommodate genetic drift, natural selection, uncertainty in calibrated radiocarbon dates, and sampling error, we find that natural selection is still required to explain the observed increase in allele frequency. We conclude that the *calcium assimilation hypothesis* is insufficient to explain the spread of LP in Europe.

Key words: lactase persistence, Iberia, Neolithic, forward simulation, ancient DNA, natural selection.

Introduction

The intestinal enzyme lactase (lactase-phlorizin hydrolase) is required for digestion of lactose, the main carbohydrate in milk, into the readily absorbable monosaccharides glucose and galactose (Swallow and Harvey 1993). While lactase is present in infants, its production is usually downregulated following weaning and it is near absent in about 68% of adults worldwide (Itan et al. 2010). Continued production of lactase into adulthood—lactase persistence (LP)—is a genetically determined, Mendelian dominant trait that shows a highly structured global geographic distribution. It is common in Europe, particularly northwestern Europe, parts

of the Indian subcontinent, as well as in many African and Middle Eastern pastoralist groups (Itan et al. 2010; Gallego Romero et al. 2012). It is associated with different allelic variants in different regions (Enattah et al. 2002; Ingram et al. 2007; Tishkoff et al. 2007; Enattah et al. 2008; Ingram et al. 2009) indicating convergent evolution. In Europe (Itan et al. 2009; Itan et al. 2010), the Indian subcontinent (Gallego Romero et al. 2012), and some African peoples (Mulcare et al. 2004), a C to T transition 13,910 bp upstream of the lactase gene (*LCT*), called -13,910*T (Enattah et al. 2002), predominates, whereas in the Middle East and most parts of Africa three other variants are commonly found in LP individuals (13,915*G, 13,907*G, and 14,010*C) (Ingram et al.

2007; Tishkoff et al. 2007). In addition, the presence of several additional alleles in Ethiopian LP pastoralists suggests the possibility of a recent soft selective sweep (Ingram et al. 2009).

In addition to evidence for the convergent evolution of LP, a low frequency or absence of -13,910*T in Early Neolithic central European farmers (Burger et al. 2007), Middle Neolithic Scandinavian hunter-gatherers (Malmstrom et al. 2010), Early Neolithic farmers from northeast Iberia (Lacan, Keyser, Ricaut, Brucato, Tarrus et al. 2011), and Late Neolithic farmers from southern France (Lacan, Keyser, Ricaut, Brucato, Duranthon et al. 2011), as well as long haplotypes surrounding *LCT* (Bersaglieri et al. 2004; Voight et al. 2006; Sabeti et al. 2007; Tishkoff et al. 2007; Gallego Romero et al. 2012; Schlebusch et al. 2013), low variation in closely linked microsatellites (Coelho et al. 2005), and demic (Gerbault et al. 2009) and spatially-explicit (Itan et al. 2009) simulations, all indicate a recent origin for, and strong natural selection acting on, LP-associated alleles.

There is a clear association between the presence of LP and a history of pastoralism in living populations (Holden and Mace 1997), and estimates of the ages of major LP-associated alleles—specifically -13,910*T in Europe and -14,010*C in Africa—bracket dates for the domestication of milkable animals (Peters et al. 1999; Vigne et al. 2001; Helmer et al. 2005; Peters et al. 2005; Bollongino et al. 2012) and evidence of dairying (Craig et al. 2005; Vigne 2006; Evershed et al. 2008) based in archaeological data.

The reasons why LP should provide such a strong selective advantage have been the subject of considerable speculation (Gerbault et al. 2011). The *calcium assimilation hypothesis* (Flatz and Rotthauwe 1973) focuses on key nutritional components in milk, specifically vitamin D and calcium. Vitamin D is essential for calcium absorption in the gut and can be obtained from a few dietary sources (e.g., fish and eggs) or by the action of UVB light on the skin, which photoconverts 7-dehydrocholesterol into cholecalciferol (vitamin D₃). At high latitudes, incident UVB levels are often too low for adequate vitamin D production (Jablonski and Chaplin 2010), so, in the absence of sufficient dietary sources of vitamin D, early northern European farmers may have been calcium deficient and in danger of developing rickets. Milk is a modest source of vitamin D but an excellent source of calcium, and the *Calcium Assimilation Hypothesis* (Flatz and Rotthauwe 1973) proposes that LP was selectively favorable in northern Europe as a means of avoiding poor calcium and vitamin D status.

While the *Calcium Assimilation Hypothesis* (Flatz and Rotthauwe 1973) may explain the high frequencies of LP observed in northern Europe, UVB levels in southern Europe, the Middle East, the Indian subcontinent, and Africa should be sufficient to avoid vitamin D deficiency (Jablonski and Chaplin 2010). Despite this, available estimates of selection coefficients for high-UVB non-European regions are also typically high (Bersaglieri et al. 2004; Tishkoff et al. 2007; Gallego Romero et al. 2012; Schlebusch et al. 2013). In modern Iberians, the -13,910*T allele is found at a frequency of around 37% (Rasinerpa et al. 2005), consistent with an estimated LP frequency of 66% (Leis et al. 1997; Itan et al. 2010). If

these frequencies are also the result of natural selection after the establishment of the Neolithic, then the *Calcium Assimilation Hypothesis* should not be invoked for this region of Europe.

Although a number of statistical methods that use contemporary data have been developed for detecting natural selection acting on novel alleles (Sabeti et al. 2006), all of these indirect approaches have poor sensitivity and temporal resolution, most are confounded by past demographic processes, and many are insensitive to selection acting on standing variation (Peter et al. 2012). In principle, none should be as powerful as direct comparison allele frequencies at different times in a continuous population. Given such data and conditions, any test for natural selection needs only discount genetic drift as the null explanation for allele frequency change. Furthermore, such an approach should permit temporal resolution of episodic selection given sufficient ancient DNA data. However, in central Europe and in Scandinavia, population continuity between Early Neolithic farming and modern (Bramanti et al. 2009) and between Late Neolithic hunter-gatherer and modern (Malmström et al. 2009) populations, respectively, has been rejected using ancient DNA data. While rejection of population continuity does not constitute evidence of complete population replacement, it clearly indicates that a degree of inward gene flow has taken place (Skoglund et al. 2012).

In this study, we report -13,910 C/T allele frequencies from Late Neolithic and Early Bronze Age northeastern Iberia. Using previously published mitochondrial data from Neolithic farmers from the same region (Sampietro et al. 2007; Lacan, Keyser, Ricaut, Brucato, Tarrus et al. 2011), as well as modern northeast Iberian sequences (Crespillo et al. 2000), we show by coalescent simulation that population continuity in the intervening period cannot be rejected. We then apply a novel forward simulation approach that accommodates genetic drift, natural selection, uncertainty in calibrated radiocarbon dates, and ancient and modern sampling errors to test whether natural selection is still required to explain that rise in Iberian -13,910*T allele frequencies over the last 7,000 years. Additionally, we perform the same analyses on previously published mtDNA and -13,910*T allele frequency data from southern France (Lacan, Keyser, Ricaut, Brucato, Duranthon et al. 2011).

Results

PCR and Sequencing

Of the 18 individual skeletal remains examined from the Iberian site of El Portalón, eight yielded reproducible sequences (see table 1). In total, we pyrosequenced these eight samples 60 times; each sample was sequenced between five and ten times (sequences are available from authors upon request). Out of these 60 genotypings, 53 indicated homozygosity for the *LCT* -13,910*C allele, and 7 indicated heterozygosity (in five of the eight samples). However, for all eight samples, the majority of genotypings consistently indicated homozygosity for the -13,910*C allele (the worst case was ATP6, where 25% of the genotypings indicated heterozygosity

and 75% homozygosity for the -13,910*C allele). Previous studies on ancient cattle single-nucleotide polymorphisms using the pyrosequencing approach have demonstrated the presence of damaged bases (Svensson et al. 2007), and after numerous repeat sequencing of our samples we concluded that most of the few apparent heterozygote genotypes are due to postmortem cytosine deamination or type-two damage (Gilbert et al. 2003). Further support for this can be seen from the molecular behavior of considerably older cave bear (*Ursus deningeri*) (Dabney, Knapp et al. 2013) and hominin (Meyer et al. 2014) sequences recovered from the Sima de los Huesos cave, which is part of the Atapuerca cave system to which El Portalón forms the entrance. In both sets of bone material, nucleotide misincorporation rates in sequencing reads generated on Illumina's HiSeq 2500 platform were elevated at the 3' and 5' ends, indicating authentic ancient DNA (Malmstrom et al. 2007; Dabney, Meyer et al. 2013). Negative controls used in batches showed no or negligible evidence of contamination. However, it should be noted that despite

stringent anti-contamination measures, modern contamination can never be completely discounted (Gilbert et al. 2005; Malmstrom et al. 2005; Malmstrom et al. 2007; Malmström et al. 2009).

Population Continuity

Coalescence simulations were unable to reject population continuity in northeastern Iberia from the Neolithic to the present (see fig. 1a) using 18 ancient HVR1 mtDNA sequences (Sampietro et al. 2007; Lacan, Keyser, Ricaut, Brucato, Tarrus et al. 2011) and 118 modern sequences (Crespillo et al. 2000). Likewise, we were unable to reject population continuity in southeastern France between the Late Neolithic and today (see fig. 1b) using 29 ancient HVR1 mtDNA sequences (Lacan, Keyser, Ricaut, Brucato, Duranthon et al. 2011) and 109 modern sequences (Dubut et al. 2004).

Because our test for population continuity in northeastern Iberia from the Neolithic to the present involved using ancient mtDNA data from two independent studies (Sampietro et al. 2007; Lacan, Keyser, Ricaut, Brucato, Tarrus et al. 2011), raising the possibility of local structuring, we also performed coalescent simulations to test for continuity using those two samples independently (see supplementary fig. S2a and b, Supplementary Material online). As can be seen, for both samples we are still unable to reject population continuity over the vast majority of combinations of the N_{UP} and N_N parameters.

Selection on LCT -13,910*T

In all three comparisons of LCT -13,910*T allele frequency between ancient and modern samples, neutrality was rejected

Table 1. The Number of Times Each Genotype at Position -13910 in the Lactase Gene Was Successfully Obtained in the Iberian Sample.

Sample	C/C	T/T	C/T
ATP 3	5		
ATP 4	6		1
ATP 6	6		2
ATP 7	7		
ATP 8a	8		2
ATP 8b	8		1
ATP 9	8		
ATP 11	5		1

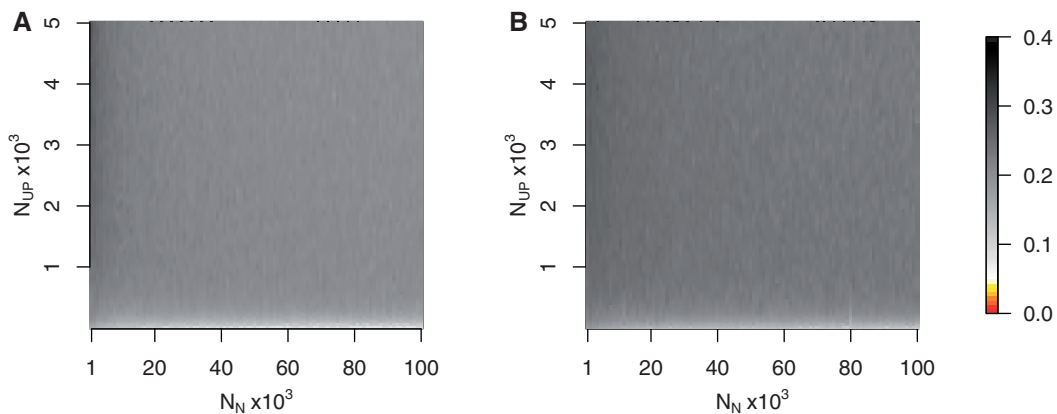


Fig. 1. Probabilities of obtaining simulated molecular F_{ST} s greater than those observed between Neolithic and modern samples. (A) Comparison of 118 modern mtDNA hypervariable region I (HVR1) sequences from northeastern Iberia (Crespillo et al. 2000) and a combined sample of 11 homologous sequences from Neolithic remains from Granollers (Catalonia, northeast Spain) dated to 5,500 years BP (Sampietro, et al. 2007) and 7 sequences from Neolithic remains excavated from a Spanish funeral cave dating to around 7,000 years BP (Lacan, Keyser, Ricaut, Brucato, Tarrus, et al. 2011) ($F_{ST} = 0.0101$; $P = 0.154$). (B) Comparison of a combined modern sample of 37 mtDNA hypervariable region I sequences from the department of Var in the region Provence-Alpes-Côte d'Azur in Provence and 69 sequences from the region of Périgord Limousin, straddling three western-central departments (northern Dordogne, western Corrèze, and southern Haute-Vienne) (Dubut, et al. 2004), and 29 sequences from a Late Neolithic necropolis in Treilles (Lacan, Keyser, Ricaut, Brucato, Duranthon, et al. 2011) ($F_{ST} = 0.00553$; $P = 0.201$). Two phases of exponential growth were modeled, the first after the initial colonization of Europe 45,000 years ago, of assumed effective female population size N_{UP} (y axis), and ending with the Neolithic transition in Iberia 7,500 years ago (López de Pablo and Gómez Puche 2009), of assumed effective female population size N_N (x axis), and the second leading up to the present, with an assumed a female effective population size of 369,715. P values greater than 0.05 are shaded on a gray scale.

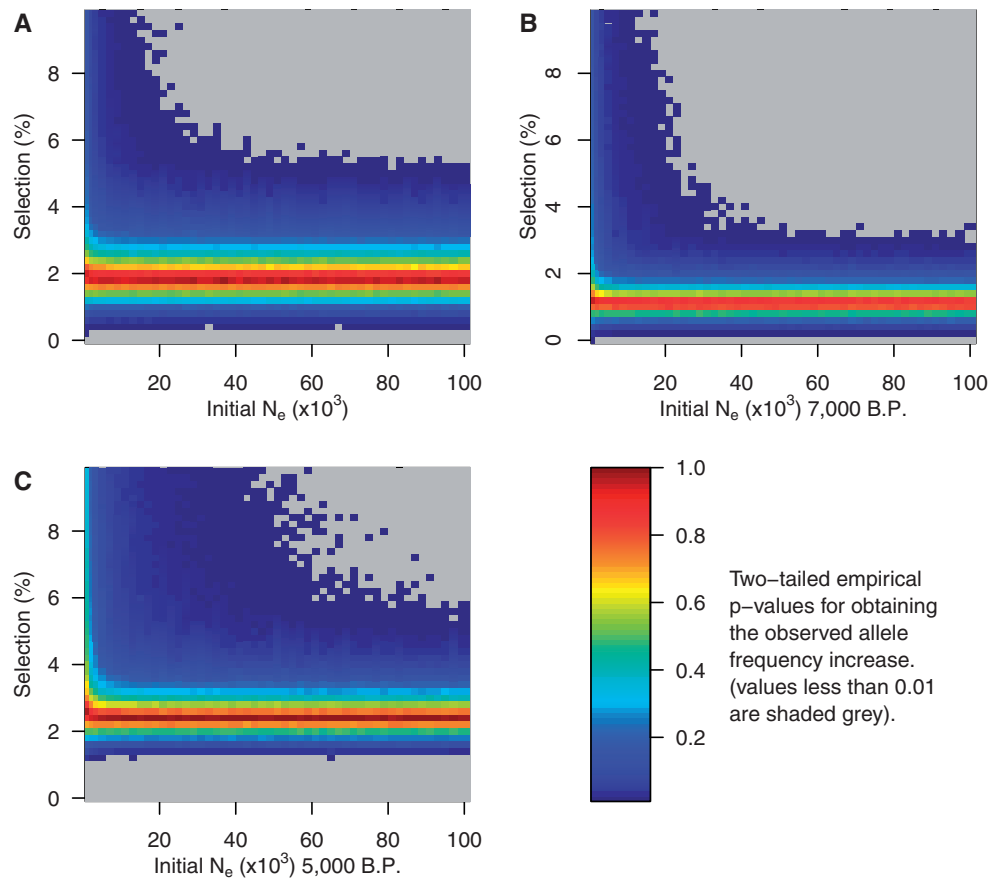


Fig. 2. Two-tailed empirical P values for the similarity between the observed modern LCT -13,910*T frequency and the distribution of simulated frequencies, given the demographic and natural selection model parameters modeled between Neolithic and modern populations. Allele frequencies were compared between (A) the ancient northeastern Iberian population sampled here and modern northeastern Iberians (Rasinpera, et al. 2005), (B) the ancient northeastern Iberian population (sampled by Lacan, Keyser, Ricaut, Brucato, Tarrus, et al. 2011) and the same modern northeastern Iberians as above (Rasinpera et al. 2005), and (C) a population from the Late Neolithic necropolis in Treilles, southern France (Lacan, Keyser, Ricaut, Brucato, Duranthon, et al. 2011) and modern French from the same region (Balkau 1996). P values less than 0.01 are shaded gray.

for all simulated values of ancestral population size (see fig. 2). We performed two separate analyses of LCT -13,910*T allele frequency change in northeastern Iberia, the first between the ancient sample presented here and a modern data set (Rasinpera et al. 2005) and the second between the ancient LCT data presented by Lacan et al (2011) and the same modern Iberian data set. It is notable that the selection values giving the highest P values for the observed allele frequency increase are greater for the ancient LCT data presented here (1.8%) than the data presented by Lacan et al. (2011) (1.2%). This is almost certainly because the LCT data presented here comes from individuals living more recently (mean cal. BP = 3,735 years) than those analyzed by Lacan et al. (2011) (7,000 years BP), and because the LCT -13,910*T allele was not observed in either ancient Iberian sample, the rise in frequency over a shorter period requires greater selection pressures. The selection value (2.4%) giving the highest P values for the observed allele frequency increase in southern France (Balkau 1996; Lacan, Keyser, Ricaut, Brucato, Duranthon et al. 2011) is somewhat higher than that for the Iberian data presented here (1.8%).

Discussion

We have formally tested for population discontinuity between a combined Early (Lacan, Keyser, Ricaut, Brucato, Tarrus et al. 2011) and Middle Neolithic (Sampietro et al. 2007) sample and a modern sample (Crespillo et al. 2000) from northeast Iberia. For virtually all combinations of N_{UP} and N_N , a significant proportion of simulated F_{ST} values were equal to or greater than that observed (fig. 1), and so we were unable to reject population continuity. These results are reminiscent of those presented by Gamba et al. (2012) except that we used a combined ancient sample. If the two ancient northeast Iberian samples considered were from a relatively unstructured population, then combining them should increase scope for rejecting continuity, thus making our test more conservative. However, to allow for the possibility that those samples were from differentiated populations, we also performed coalescent simulations with each ancient sample independently (see supplementary fig. S2a and b, Supplementary Material online) and were still unable to reject population continuity over the majority of combinations of the N_{UP} and N_N parameters. We were likewise unable

to reject population continuity between Late Neolithic and modern populations in southeastern France, using a larger ancient DNA sample (Lacan, Keyser, Ricaut, Brucato, Duranthon et al. 2011). To examine the sensitivity of our continuity tests to the modern population size parameter, we performed identical coalescent simulations to those described earlier except that we set the modern effective population sizes to 0.1 times and 10 times the values assumed from census data (see [supplementary fig. S1a–d, Supplementary Material](#) online). Varying this parameter by an order of magnitude each way made little difference to our results and were still unable to reject population continuity across all combinations of the N_{UP} and N_N parameters.

Since continuity is the null hypothesis, failure to reject it does not constitute a demonstration of it. However, we note that population continuity has been rejected between Neolithic farmers and modern populations from central Europe (Bramanti et al. 2009) and between Late Neolithic hunter-gatherers and modern populations from Scandinavia (Malmström et al. 2009) using smaller ancient sample sizes. It thus seems likely that population turnover since or shortly after the Neolithic transition has been less severe in southwestern Europe than in central or northern Europe (Pinhasi et al. 2012).

Our test for neutrality of the LCT -13,910*T allele is contingent on population continuity and accounts for genetic drift, ancient and modern sampling errors, and calibrated radiocarbon date uncertainty. We believe that our accommodation of the latter is a first in evolutionary inference using serial DNA samples. An important caveat to the selection coefficients estimates presented here is the assumption that selection is constant and starts at the time of the ancient sample used. This is illustrated by the differences between the point estimate of $S = 1.2\%$ using the Early Neolithic data of Lacan et al. (2011) and the point estimate of $S = 1.8\%$ using the Late Neolithic/Early Bronze Age LCT data presented here; higher selection coefficients are required to increase allele frequency by the same amount but over a shorter period of time, because they start with zero observed -13,910*T alleles. Thus, our estimates of S should be considered conservative—the true values may be higher.

The LCT data presented here comes from an archaeological site where faunal remains from the same layers have a mean age of 3,735 years cal. BP. Although the sample is relatively small, using the classical formula $p = 1 - \alpha^{(1/n)}$, where α is the probability of no occurrences of an allele in n observations (Burger et al. 2007), we estimate that the maximum frequency, p , of the LCT -13,910*T allele in this population should be no more than 0.17 with 95% confidence and no more than 0.25 with 99% confidence and could have been zero. If frequencies were very low then this is intriguing given the relatively recent date range and suggests a late introduction but rapid rise in frequency of LCT -13,910*T in Iberia. This, in turn, suggests that LCT -13,910*T was not introduced—or was only introduced at very low frequencies—with arrival of the Neolithic in Iberia. As the mtDNA analyses presented here and elsewhere (Izaguirre and de la Rúa 1999; Lacan, Keyser, Ricaut, Brucato, Tarrus et al. 2011; Gamba et al. 2012;

Hervella et al. 2012; Pinhasi et al. 2012) indicate little population turnover in Iberia since at least the Middle Neolithic, it is difficult to account for the observed change in LCT -13,910*T allele frequency by gene flow from neighboring regions, unless mediated primarily by males, further supporting the strong selection coefficients estimated here.

If the selection coefficients estimated for southeastern France and northwestern Iberia are correct, or underestimates, it is unlikely that they are explained by the *Calcium Assimilation Hypothesis* (Flatz and Rotthauwe 1973) as incident UVB levels in both regions permit adequate vitamin D production (Jablonski and Chaplin 2010). It is not currently possible to compare the selection coefficients estimated here with those for northern Europe using the same selection detection approach because, although LCT -13,910*T allele frequency data are available for ancient samples from northern Europe and Scandinavia (Burger et al. 2007; Malmstrom et al. 2010), in both regions population continuity between Neolithic and modern populations has been rejected (Bramanti et al. 2009; Malmström et al. 2009; Skoglund et al. 2012). However, the selection coefficients estimated here are remarkably similar to those presented in a recent study (Peter et al. 2012) for the LCT -13,910*T allele (2.5%; 95% CI 0.4–20%) using a Finnish sample and approximate Bayesian computation (Beaumont et al. 2002), which also concluded that LCT selection acted on *de novo*, rather than standing variation. The selection coefficient estimate mode reported by Itan et al. (2009) was 9.53%, but this was assumed to apply to less than half of the population of Europe (dairying farmers) and to be spatiotemporally constant. To a first order of approximation, this could be interpreted as indicating a Europe-wide selection coefficient of 4% or 5%, which is about twice the levels we estimate here for southwestern Europe. Thus, it is conceivable that additional factors, possibly including *calcium assimilation*, led to different selection coefficients in Northern Europe. Alternatively, as suggested by Itan et al. (2009), higher frequencies of LCT -13,910*T in northwestern Europe may be the result of a combination of strong selection and allele surfing (Edmonds et al. 2004; Klopstein et al. 2006) on a wave of advance of farmers spreading across the continent.

A temporal dimension to estimates of selection coefficients on LCT -13,910*T is still lacking, but it is certainly possible that selection on this allele was episodic (Itan et al. 2009), possibly in response to oscillations in food supply (Gerbault et al. 2011). While the data that we present is insufficient to detect episodic selection, the method we present here should be adequate for this purpose given sufficient serial sampled data.

Despite overwhelming evidence for strong selection on LP alleles, including that presented here, the reasons for that selection remain elusive (Gerbault et al. 2011). In addition to the *calcium assimilation hypothesis* (Flatz and Rotthauwe 1973) various other hypotheses have been proposed. Apart from the obvious nutrient value of milk in terms of protein, fat, calories, and various vitamin and mineral micronutrients, milk production may have offered significant economic advantages (Benecke 1994) and a good source of relatively

uncontaminated fluid, particularly in arid environments (Cook and al-Torki 1975), although this may be less relevant in Europe (Gerbault et al. 2011). Others have suggested that a milk-rich diet may offer some protection against malaria (Anderson and Vullo 1994; Cordain et al. 2012) or that LP spread as prestige class behavior (Durham 1991). Alternatively, dairy-based economies are likely to have been more buffered against food supply fluctuations than cereal-based economies, both because of the boom-and-bust of growing seasons and the possibility of crop failure (Gerbault et al. 2011). In mixed economies when crops have failed, fermented milk products are an alternative source of nutrients to LP and non-LP individuals alike. However, as lower lactose content milk products are consumed, so high lactose content products (milk and yoghurt) would be left. In famine conditions, the consequences of high lactose food consumption in non-LP individuals (particularly diarrhea) would be more severe than in well-nourished non-LP individuals, perhaps leading to high but episodic selection differentials.

Material and Methods

Samples

We attempted to extract DNA from 18 human mixed skeletal remains, out of which 8 yielded usable DNA. The remains were excavated from the Holocene site of El Portalón, the present day entrance to the Cueva Mayor Karst complex, in the Sierra de Atapuerca, 15 km east of Burgos, Spain (Carretero et al. 2008). There are no direct radiocarbon dates for these samples, but 13 radiocarbon dates are available from horse and cattle remains from the same layers (Lira et al. 2010).

DNA Extraction

Samples were handled in a clean laboratory under positive air pressure, UV irradiated daily, and isolated from post-polymerase chain reaction (PCR) work. Personnel wore disposable facemasks, zip suits, and sterile latex gloves at all times and do not enter the facility if they have worked in a post-PCR environment the same day. Equipment, nonorganic buffers, and reagents were UV irradiated.

Bone samples were individually irradiated in a crosslinker (Ultra-Violet Products Ltd., Cambridge, UK) with 6.0J/cm². From each bone, a millimeter of approximately 1 cm surface area was removed by drilling. Two 0.10 g samples of bone powder were taken from each bone using a drill hood which was UV radiated between sample batches. DNA was then extracted from the bone powder using a modification (Malmstrom et al. 2005) of the silica spin method (Yang et al. 1998). One in every four samples included an extraction control (milli-Q water).

Genotyping

A 53 bp fragment containing the -13,910 C/T polymorphism was amplified using primers “GCTGGCAATACAGATAAGATAATG” and “GAGGAGAGTTCCTTTGAGGC” (Tag Copenhagen). The forward primer was biotinylated (Malmstrom et al. 2005). Amplification was performed using 5 µl of DNA extract, 300 nM primers, and two units

of Smart Taq Hot Start (Naxo). One in every three PCRs was a negative control containing only milli-Q water. PCR amplification conditions were 15 min at 95 °C, followed by 50 cycles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s. The -13,910* alleles were identified through pyrosequencing using primer “CCTTTGAGGCCAGGG”, as previously described (Anderung et al. 2005). Dispensation order for the Pyrosequencer was CATCAGT/CGATG.

Data Analysis

To test for population continuity in northeastern Iberia from the Neolithic until present, we first calculated the molecular F_{ST} (Excoffier et al. 1992) between 118 modern mtDNA hypervariable region I (HVR1) sequences from northeastern Iberia (Crespillo et al. 2000) and a combined sample of 18 homologous sequences from Neolithic remains from comprising 11 sequences from Granollers (Catalonia, northeast Spain) dated to 5,500 years BP (Sampietro et al. 2007) and 7 sequences from remains excavated from a Spanish funeral cave dating to around 7,000 years BP (Lacan, Keyser, Ricaut, Brucato, Tarrus et al. 2011). The region of overlap for these three HVR1 samples was 310 bp (from 16,053 to 16,362 on the Cambridge reference sequence [Anderson et al. 1981]). The F_{ST} between the ancient and modern sequence samples (0.0101; $P = 0.154$) was calculated using the *amova()* function in the “R” (R Core Team 2012) library *ade4* (Dray and Dufour 2007). P values for these F_{ST} estimates were calculated by permuting the data 100,000 times. We then examined whether this observed F_{ST} was greater than expected under a model of population continuity for a range of combinations of female effective population sizes at the Upper Palaeolithic transition in Europe (N_{UP}) 45,000 years ago (Powell et al. 2009), and the Neolithic transition in Iberia (N_N) 7,500 years ago (López de Pablo and Gómez Puche 2009), as described previously (Bramanti et al. 2009; Malmström et al. 2009), assuming exponential growth between 45,000 years ago and 7,500 years ago, and between 7,500 years ago and the present, and a generation time of 25 years. We assumed a modern female effective population size of 369,715 (one-tenth of the modern female population size of Catalonia and Aragon combined [census estimates for 1999, source: <http://www.google.co.uk/publicdata/>, last accessed February 4, 2014] for the region from which the modern sequences were sampled [Crespillo et al. 2000]). Ten thousand coalescent simulations were performed for each of all 10,000 combinations of 100 equally spaced values of N_{UP} , ranging from 10 to 5,000, and 100 equally spaced values of N_N , ranging from 1,000 to 100,000, using *Fastsimcoal* (Excoffier and Foll 2011). Simulated F_{ST} values were calculated using the same *amova()* function as used to calculate the observed F_{ST} values, and the proportion of simulated F_{ST} values equal to or greater than the observed F_{ST} were plotted for each combination of N_{UP} and N_N .

The above analysis was repeated for southeastern France using a combined modern sample of 106 mtDNA hypervariable region I sequences comprising 37 sequences from the department of Var in the region Provence-Alpes-Côte d’Azur

in Provence and 69 sequences from the region of Périgord Limousin, straddling three western-central departments (northern Dordogne, western Corrèze, and southern Haute-Vienne) (Dubut et al. 2004), and 29 ancient sequences from a Late Neolithic necropolis in Treilles (Lacan, Keyser, Ricaut, Brucato, Duranthon et al. 2011) ($F_{ST} = 0.00553$; $P = 0.201$). The region of overlap was 308 bp, from 16,055 to 16,362 on the Cambridge reference sequence (Anderson et al. 1981). Simulations were performed as described above except that a modern effective female population size of 480,598 was assumed (one-tenth of the modern female population of the combined departments of Var, Hérault, Dordogne, Corrèze, Haute-Vienne, Bouches-du-Rhône, Gard, Lozère, Aveyron, Cantal, Lot, Alpes-de-Haute-Provence, Vaucluse, Ardèche, Drôme, Haute-Loire, Puy-de-Dôme, Creuse, Charente, Lot-et-Garonne, Tarn-et-Garonne, Tarn [source: 2006 Population estimates from <http://www.citypopulation.de/>, last accessed February 4, 2014]).

To test whether changes in the -13,910*T allele frequency can be explained by genetic drift, or whether natural selection needs to be invoked, and to estimate the strength of natural selection where appropriate, we used a forward simulation approach. For the Iberian ancient sample reported here, we first obtained 10,000 Markov chain Monte Carlo samples (chain length = 1,100,000, burn in = 100,000, every tenth sample retained) from the calibration process of 13 faunal samples from the site of El Portalón de Cueva Mayor (Sierra de Atapuerca, Burgos) that were found at the same layer as the human samples for which we report LCT -13,910*T allele frequency estimates here (Lira et al. 2010), using the R library Bchron (Haslett and Parnell 2008) and the IntCal09 calibration curve (Heaton et al. 2009). These MCMC samples form a random sample of the summed calibrated date distribution for all 13 faunal samples, permitting integration of date uncertainty into our forward simulations.

To reflect uncertainty in ancestral allele frequencies, assuming a uniform allele frequency prior, estimates were drawn from a random Beta($n_p + 1, n_q + 1$) distribution, where n_p and n_q were equal to the number of -13,910*T alleles and the number of -13,910*C alleles in the ancient sample, respectively. Then, drift and natural selection were forward simulated by binomial sampling across generations, with N equal to the population size in each generation, plus a standard selection equation assuming dominance (Maynard Smith 1998), respectively. Exponential population growth was modeled from a range of values of N_e at the time the ancient sample (50 equally spaced values between 1,000 and 100,000) to a modern N_e of 681,332 (one-tenth census population size of Catalonia in 2004 = 6,813,319; source: <http://www.idescat.cat/>, last accessed February 4, 2014). The number of generations forward simulated was drawn at random from the pool of 10,000 MCMC samples from each of the 13 calibrated radiocarbon dates (Lira et al. 2010) assuming a generation time of 25 years. In the final generation, the simulated modern sample allele frequencies were picked from a random binomial with N equal to the modern sample size ($n = 442$ alleles) (Rasinpera et al. 2005) and the simulated distribution of frequencies compared with the observed frequencies using

the formula $1 - 2 \times |0.5 - P|$, where P is the proportion of simulated modern allele frequencies that are greater than that observed. This yields a two-tailed empirical P value for the similarity between the observed modern allele frequency and the distribution of simulated allele frequencies, given the demographic and natural selection model parameters modeled (Voight et al. 2005).

The above analysis was repeated for published ancient LCT -13,910*T allele frequency data from seven Neolithic remains excavated from a Spanish funeral cave dating to around 7,000 years BP (Lacan, Keyser, Ricaut, Brucato, Tarrus et al. 2011), using the same northeast Iberian modern -13,910*T allele frequency data (Rasinpera et al. 2005), and from 26 human samples from a Late Neolithic necropolis in Treilles, southern France (Lacan, Keyser, Ricaut, Brucato, Duranthon et al. 2011) using a combined modern French -13,910*T allele frequency data set of 270 chromosomes from individuals unrelated at the second cousin level (Balkau 1996).

The Sequences obtained from the eight individuals from this study are available from the authors upon request.

Supplementary Material

Supplementary figures S1 and S2 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

O.O.S. is funded by an EU Marie Curie FP7 Framework Programme grant (LeCHE, grant ref. 215362-2) awarded to A.G., M.G.T., and others not involved in this study. The authors acknowledge the use of the UCL Legion High Performance Computing Facility (Legion@UCL), and associated support services, in the completion of this work. Atapuerca project is funded by Spanish Government project CGL2012-38434-C03-01 and field work at the Atapuerca sites by the Junta de Castilla y León. The authors thank David Balding and Douglas Speed for their helpful advice and all the participants of the D.E.S.I.R study—genotyping for the modern French LCT data was supported by the Conseil Régional Nord-Pas-de-Calais Fonds Européen de Développement Economique et Regional CPER axe Cartdiodiabète 2010–2011 grant to Nabila Bouatia-Naji.

References

- Anderson B, Vullo C. 1994. Did malaria select for primary adult lactase deficiency? *Gut* 35:1487–1489.
- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465.
- Anderung C, Bouwman A, Persson P, Carretero JM, Ortega AI, Elburg R, Smith C, Arsuaga JL, Ellegren H, Gotherstrom A. 2005. Prehistoric contacts over the Straits of Gibraltar indicated by genetic analysis of Iberian Bronze Age cattle. *Proc Natl Acad Sci U S A*. 102:8431–8435.
- Balkau B. 1996. [An epidemiologic survey from a network of French Health Examination Centres, (D.E.S.I.R.): epidemiologic data on the insulin resistance syndrome]. *Rev Epidemiol Sante Publique*. 44:373–375.
- Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162:2025–2035.

- Benecke N. 1994. Archäozoologische Studien zur Entwicklung der Haustierhaltung in Mitteleuropa und Südkandinavien von den Anfängen bis zum ausgehenden Mittelalter. Schriften zur Ur- und Frühgeschichte 46. Berlin, Germany: Akademie Verlag.
- 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.* 74:1111–1120.
- Bollongino R, Burger J, Powell A, Mashkour M, Vigne JD, Thomas MG. 2012. Modern taurine cattle descended from small number of Near-Eastern founders. *Mol Biol Evol.* 29: 2101–2104.
- Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, Tambets K, Antanaitis-Jacobs I, Haidle MN, Jankauskas R, Kind CJ, et al. 2009. Genetic Discontinuity Between Local Hunter-Gatherers and Central Europe's First Farmers. *Science* 326:137–140.
- Burger J, Kirchner M, Bramanti B, Haak W, Thomas MG. 2007. Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proc Natl Acad Sci U S A.* 104:3736–3741.
- Carretero JM, Ortega AI, Juez L, Pérez-González A, Arsuaga JL, Pérez-Martínez R, Ortega MC. 2008. A late Pleistocene-Early Holocene archaeological sequence of Portalón de Cueva Mayor (Sierra de Atapuerca, Burgos, Spain). *MUNIBE* 59:67–80.
- Coelho M, Luiselli D, Bertorelle G, Lopes AI, Seixas S, Destro-Bisol G, Rocha J. 2005. Microsatellite variation and evolution of human lactase persistence. *Hum Genet.* 117:329–339.
- Cook GC, al-Torki MT. 1975. High intestinal lactase concentrations in adult Arabs in Saudi Arabia. *Br Med J.* 3:135–136.
- Cordain L, Hickey MS, Kim K. 2012. Malaria and rickets represent selective forces for the convergent evolution of adult lactase persistence. In: Gepts P, Famula TR, Bettinger RL, Brush SB, Damania AB, McGuire PE, Qualset CO, editors. Biodiversity in agriculture: domestication, evolution, and sustainability. Cambridge, New York: Cambridge University Press. p. 299–308.
- Craig OE, Chapman J, Heron CP, Willis LH, Taylor G, Whittle A, Collins MJ. 2005. Did the first farmers of central and eastern Europe produce dairy foods? *Antiquity* 79:882–894.
- Crespillo M, Luque JA, Paredes M, Fernandez R, Ramirez E, Valverde JL. 2000. Mitochondrial DNA sequences for 118 individuals from north-eastern Spain. *Int J Legal Med.* 114:130–132.
- Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, Nickel B, Valdiosera C, Garcia N, Paabo S, Arsuaga JL, et al. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci U S A.* 110:15758–15763.
- Dabney J, Meyer M, Paabo S. 2013. Ancient DNA damage. *Cold Spring Harb Perspect Biol.* 5:a012567.
- Dray S, Dufour AB. 2007. The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw.* 22:1–20.
- Dubut V, Chollet L, Murail P, Cartault F, Beraud-Colomb E, Serre M, Mogentale-Profizi N. 2004. mtDNA polymorphisms in five French groups: importance of regional sampling. *Eur J Hum Genet.* 12: 293–300.
- Durham WH. 1991. Cultural mediation: the evolution of adult lactose absorption. In: Coevolution: genes, culture and human diversity. Stanford, CA: Stanford University Press. p. 226–285.
- Edmonds CA, Lillie AS, Cavalli-Sforza LL. 2004. Mutations arising in the wave front of an expanding population. *Proc Natl Acad Sci U S A.* 101:975–979.
- Enattah NS, Jensen TG, Nielsen M, Lewinski R, Kuokkanen M, Rasinpera H, El-Shanti H, Seo JK, Alifrangis M, Khalil IF, et al. 2008. Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *Am J Hum Genet.* 82:57–72.
- Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Jarvela I. 2002. Identification of a variant associated with adult-type hypolactasia. *Nat Genet.* 30:233–237.
- Evershed RP, Payne S, Sherratt AG, Copley MS, Coolidge J, Urem-Kotsu D, Kotsakis K, Ozdogan M, Ozdogan AE, Nieuwenhuys O, et al. 2008. Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. *Nature* 455:528–531.
- Excoffier L, Foll M. 2011. fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 27:1332–1334.
- Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.
- Flatz G, Rottshauwe HW. 1973. Lactose nutrition and natural selection. *Lancet* 2:76–77.
- Gallego Romero I, Basu Mallick C, Liebert A, Crivellaro F, Chaubey G, Itan Y, Metspalu M, Easwarkhanth M, Pitchappan R, Vilems R, et al. 2012. Herders of Indian and European Cattle Share Their Predominant Allele for Lactase Persistence. *Mol Biol Evol.* 29: 249–260.
- Gamba C, Fernandez E, Tirado M, Deguilloux MF, Pemonge MH, Utrilla P, Edo M, Molist M, Rasteiro R, Chikhi L, et al. 2012. Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonization by first farmers. *Mol Ecol.* 21:45–56.
- Gerbault P, Liebert A, Itan Y, Powell A, Currat M, Burger J, Swallow DM, Thomas MG. 2011. Evolution of lactase persistence: an example of human niche construction. *Philos Trans R Soc Lond B Biol Sci.* 366: 863–877.
- Gerbault P, Moret C, Currat M, Sanchez-Mazas A. 2009. Impact of selection and demography on the diffusion of lactase persistence. *PLoS One* 4:e6369.
- Gilbert MT, Bandelt HJ, Hofreiter M, Barnes I. 2005. Assessing ancient DNA studies. *Trends Ecol Evol.* 20:541–544.
- Gilbert MT, Willerslev E, Hansen AJ, Barnes I, Rudbeck L, Lynnerup N, Cooper A. 2003. Distribution patterns of postmortem damage in human mitochondrial DNA. *Am J Hum Genet.* 72:32–47.
- Haslett J, Parnell A. 2008. A simple monotone process with application to radiocarbon-dated depth chronologies. *J R Stat Soc Ser C Appl Stat.* 57:399–418.
- Heaton TJ, Blackwell PG, Buck CE. 2009. A Bayesian approach to the estimation of radiocarbon calibration curves: the IntCal09 methodology. *Radiocarbon* 51:1151–1164.
- Helmer D, Gourichon L, Monchot H, Peters J, Sana Segui M. 2005. Identifying early domestic cattle from pre-pottery Neolithic sites on the Middle Euphrates using sexual dimorphism. In: Vigne JD, Helmer D, Peters J, editors. First steps of animal domestication. New archaeological approach. Oxford: Oxbow Books. p. 86–95.
- Hervella M, Izagirre N, Alonso S, Fregel R, Alonso A, Cabrera VM, de la Rúa C. 2012. Ancient DNA from hunter-gatherer and farmer groups from Northern Spain supports a random dispersion model for the Neolithic expansion into Europe. *PLoS One* 7:e34417.
- Holden C, Mace R. 1997. Phylogenetic analysis of the evolution of lactose digestion in adults. *Hum Biol.* 69:605–628.
- Ingram CJ, Elamin MF, Mulcare CA, Weale ME, Tarekegn A, Raga TO, Bekele E, Elamin FM, Thomas MG, Bradman N, et al. 2007. A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? *Hum Genet.* 120:779–788.
- Ingram CJ, Raga TO, Tarekegn A, Browning SL, Elamin MF, Bekele E, Thomas MG, Weale ME, Bradman N, Swallow DM. 2009. Multiple rare variants as a cause of a common phenotype: several different lactase persistence associated alleles in a single ethnic group. *J Mol Evol.* 69:579–588.
- Itan Y, Jones B, Ingram C, Swallow D, Thomas M. 2010. A worldwide correlation of lactase persistence phenotype and genotypes. *BMC Evol Biol.* 10:36.
- Itan Y, Powell A, Beaumont MA, Burger J, Thomas MG. 2009. The origins of lactase persistence in Europe. *PLoS Comput Biol.* 5:e1000491.
- Izagirre N, de la Rúa C. 1999. An mtDNA analysis in ancient Basque populations: implications for haplogroup V as a marker for a major paleolithic expansion from southwestern Europe. *Am J Hum Genet.* 65:199–207.
- Jablonski NG, Chaplin G. 2010. Human skin pigmentation as an adaptation to UV radiation. *Proc Natl Acad Sci U S A.* 107:8962–8968.

- Klopfstein S, Currat M, Excoffier L. 2006. The fate of mutations surfing on the wave of a range expansion. *Mol Biol Evol.* 23:482–490.
- Lacan M, Keyser C, Ricaut FX, Brucato N, Duranthon F, Guilaine J, Crubezy E, Ludes B. 2011. Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. *Proc Natl Acad Sci U S A.* 108:9788–9791.
- Lacan M, Keyser C, Ricaut FX, Brucato N, Tarrus J, Bosch A, Guilaine J, Crubezy E, Ludes B. 2011. Ancient DNA suggests the leading role played by men in the Neolithic dissemination. *Proc Natl Acad Sci U S A.* 108:18255–18259.
- Leis R, Tojo R, Pavon P, Douwes A. 1997. Prevalence of lactose malabsorption in Galicia. *J Pediatr Gastroenterol Nutr.* 25:296–300.
- Lira J, Linderholm A, Olaria C, Brandstrom Durling M, Gilbert MT, Ellegren H, Willerslev E, Liden K, Arsuaga JL, Gotherstrom A. 2010. Ancient DNA reveals traces of Iberian Neolithic and Bronze Age lineages in modern Iberian horses. *Mol Ecol.* 19:64–78.
- López de Pablo JF, Gómez Puche M. 2009. Climate change and population dynamics during the Late Mesolithic and the Neolithic transition in Iberia. *Neolithic Studies 16. Documenta Praehistorica* 36: 67–96.
- Malmström H, Gilbert MT, Thomas MG, Brandstrom M, Stora J, Molnar P, Andersen PK, Bendixen C, Holmlund G, Gotherstrom A, et al. 2009. Ancient DNA reveals lack of continuity between neolithic hunter-gatherers and contemporary Scandinavians. *Curr Biol.* 19: 1758–1762.
- Malmstrom H, Linderholm A, Liden K, Stora J, Molnar P, Holmlund G, Jakobsson M, Gotherstrom A. 2010. High frequency of lactose intolerance in a prehistoric hunter-gatherer population in northern Europe. *BMC Evol Biol.* 10:89.
- Malmstrom H, Stora J, Dalen L, Holmlund G, Gotherstrom A. 2005. Extensive human DNA contamination in extracts from ancient dog bones and teeth. *Mol Biol Evol.* 22:2040–2047.
- Malmstrom H, Svensson EM, Gilbert MT, Willerslev E, Gotherstrom A, Holmlund G. 2007. More on contamination: the use of asymmetric molecular behavior to identify authentic ancient human DNA. *Mol Biol Evol.* 24:998–1004.
- Maynard Smith J. 1998. *Evolutionary genetics.* Oxford, New York: Oxford University Press. xiv, p. 330.
- Meyer M, Fu Q, Aximu-Petri A, Glocke I, Nickel B, Arsuaga JL, Martinez I, Gracia A, de Castro JM, Carbonell E, et al. 2014. A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* 505:403–406.
- Mulcare CA, Weale ME, Jones AL, Connell B, Zeitlyn D, Tarekegn A, Swallow DM, Bradman N, Thomas MG. 2004. The T allele of a single-nucleotide polymorphism 13.9 kb upstream of the lactase gene (LCT) (C-13.9kbT) does not predict or cause the lactase-persistence phenotype in Africans. *Am J Hum Genet.* 74:1102–1110.
- Peter BM, Huerta-Sanchez E, Nielsen R. 2012. Distinguishing between selective sweeps from standing variation and from a de novo mutation. *PLoS Genet.* 8:e1003011.
- Peters J, Helmer D, Von Den Driech A, Sana Segui M. 1999. Early Animal Husbandry in the Northern Levant. *Paleorient* 25:27–47.
- Peters J, Von Den Driech A, Helmer D. 2005. The Upper Euphrates-Tigris Basin: cradle of agro-pastoralism? In: Vigne JD, Helmer D, Peters J, editors. *First steps of animal domestication. New archaeozoological approach.* Oxford: Oxbow Books. p. 96–124.
- Pinhasi R, Thomas MG, Hofreiter M, Currat M, Burger J. 2012. The genetic history of Europeans. *Trends Genet.* 28:496–505.
- Powell A, Shennan S, Thomas MG. 2009. Late Pleistocene demography and the appearance of modern human behavior. *Science* 324: 1298–1301.
- R Core Team. 2012. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.
- Rasinpera H, Forsblom C, Enattah NS, Halonen P, Salo K, Victorzon M, Mecklin JP, Jarvinen H, Enholm S, Sellick G, et al. 2005. The C/C-13910 genotype of adult-type hypolactasia is associated with an increased risk of colorectal cancer in the Finnish population. *Gut* 54:643–647.
- Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. 2006. Positive natural selection in the human lineage. *Science* 312:1614–1620.
- Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, et al. 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* 449:913–918.
- Sampietro ML, Lao O, Caramelli D, Lari M, Pou R, Marti M, Bertranpetit J, Lalueza-Fox C. 2007. Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proc Biol Sci.* 274: 2161–2167.
- Schlebusch CM, Sjödin P, Skoglund P, Jakobsson M. 2013. Stronger signal of recent selection for lactase persistence in Maasai than in Europeans. *Eur J Hum Genet.* 21:550–553.
- Skoglund P, Malmstrom H, Raghavan M, Stora J, Hall P, Willerslev E, Gilbert MT, Gotherstrom A, Jakobsson M. 2012. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* 336:466–469.
- Svensson EM, Anderung C, Baubliene J, Persson P, Malmstrom H, Smith C, Vretemark M, Daugnora L, Gotherstrom A. 2007. Tracing genetic change over time using nuclear SNPs in ancient and modern cattle. *Anim Genet.* 38:378–383.
- Swallow DM, Harvey C. 1993. *Genetics of adult-type hypolactasia.* In: Auricchio S, Semenza G, editors. *Common food intolerances 2: milk in human nutrition and adult-type hypolactasia.* Basel, London: Karger. p. 85–92.
- Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, et al. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet.* 39:31–40.
- Vigne J-D. 2006. *Maitrise et usages de l'élevage et des animaux domestiques au Néolithique: quelques illustrations au Proche-Orient et en Europe.* In: Guilaine J, editor. *Populations néolithiques et environnements, Vol. 1.* Paris: Errance Éd. p. 87–114.
- Vigne JD, Carrere I, Guilaine J, editors. 2001. *Actes du Colloque International Organisé par le Département des Antiquités de Chypre et l'Ecole Française d'Athènes; 2001 17–19 May; Nicosia (Cyprus).*
- Voight BF, Adams AM, Frishe LA, Qian Y, Hudson RR, Di Rienzo A. 2005. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc Natl Acad Sci U S A.* 102:18508–18513.
- Voight BF, Kudavalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72.
- Yang DY, Eng B, Wayne JS, Dudar JC, Saunders SR. 1998. Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *Am J Phys Anthropol.* 105:539–543.