Report

Current Biology

Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years

Highlights

- Genomic data from Tollense, the oldest large-scale conflict site north of the Alps
- Novel method indicates Bronze Age warriors represent an unstructured population
- Lactase persistence frequency in Tollense (7.1%) is significantly lower than today
- Selection coefficient estimate of 6% over the last 3,000 years

Authors

Joachim Burger, Vivian Link, Jens Blöcher, ..., Mark G. Thomas, Krishna R. Veeramah, Daniel Wegmann

Correspondence

jburger@uni-mainz.de (J.B.), daniel.wegmann@unifr.ch (D.W.)

In Brief

Burger et al. report the first genomic data from the oldest known battlefield north of the Alps. With additional data from 55 individuals from sites in Southern and Eastern Europe dating to the Bronze Age, they find evidence for a strong and ongoing selection on lactase persistence in various parts of Europe over the last 3,000 years.

Burger et al., 2020, Current Biology *30*, 1–9
November 2, 2020
https://doi.org/10.1016/j.cub.2020.08.033

CellPress

Please cite this article in press as: Burger et al., Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.08.033

Current Biology

CellPress

Report

Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years

Joachim Burger,[1,17,18,*] Vivian Link,[2,3,17] Jens Blöcher,[1,17] Anna Schulz,[4] Christian Sell,[1] Zoé Pochon,[2,3] Yoan Diekmann,[1] Aleksandra Žegarac,[5] Zuzana Hofmanova,[2,3] Laura Winkelbach,[1] Carlos S. Reyna-Blanco,[2,3] Vanessa Bieker,[6] Jörg Orschiedt,[7,8] Ute Brinker,[8,9] Amelie Scheu,[1] Christoph Leuenberger,[10] Thomas S. Bertino,[11] Ruth Bollongino,[12] Gundula Lidke,[13] Sofija Stefanović,[5,14] Detlef Jantzen,[9] Elke Kaiser,[7] Thomas Terberger,[15] Mark G. Thomas,[16] Krishna R. Veeramah,[11] and Daniel Wegmann[2,3,*]

[1]Palaeogenetics Group Institute of Organismic and Molecular Evolution (iomE), Johannes Gutenberg University Mainz, 55099 Mainz, Germany
[2]Department of Biology, University of Fribourg, 1700 Fribourg, Switzerland
[3]Swiss Institute of Bioinformatics, 1700 Fribourg, Switzerland
[4]Centre for the Study of Manuscript Cultures, Cluster of Excellence Understanding Written Artefacts, Hamburg University, 20354 Hamburg, Germany
[5]Laboratory of Bioarchaeology, Faculty of Philosophy, University of Belgrade, 11000 Belgrade, Serbia
[6]Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology (NTNU), 7012 Trondheim, Norway
[7]Institut für Prähistorische Archäologie, Freie Universität Berlin, Fabeckstr. 23-25, 14195 Berlin, Germany
[8]Landesamt für Denkmalpflege und Archäologie Sachsen-Anhalt, Richard-Wagner-Straße 9, 06114 Halle (Saale), Germany
[9]State Agency for Heritage Service of Mecklenburg-Vorpommern, Domhof 4, 19055 Schwerin, Germany
[10]Department of Mathematics, University of Fribourg, 1700 Fribourg, Switzerland
[11]Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245, USA
[12]108a Central Road, Upper Moutere 7175, New Zealand
[13]Schloßstraße 50, 14059 Berlin, Germany
[14]BioSense Institute, University of Novi Sad, Bulevar Zorana Dindica 1, Novi Sad 21000, Serbia
[15]Seminar for Pre- and Protohistory, University of Göttingen, Nikolausberger Weg 15, 37073 Göttingen, Germany
[16]UCL Genetics Institute, Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK
[17]These authors contributed equally
[18]Lead Contact
*Correspondence: jburger@uni-mainz.de (J.B.), daniel.wegmann@unifr.ch (D.W.)
https://doi.org/10.1016/j.cub.2020.08.033

SUMMARY

Lactase persistence (LP), the continued expression of lactase into adulthood, is the most strongly selected single gene trait over the last 10,000 years in multiple human populations. It has been posited that the primary allele causing LP among Eurasians, rs4988235-A [1], only rose to appreciable frequencies during the Bronze and Iron Ages [2, 3], long after humans started consuming milk from domesticated animals. This rapid rise has been attributed to an influx of people from the Pontic-Caspian steppe that began around 5,000 years ago [4, 5]. We investigate the spatiotemporal spread of LP through an analysis of 14 warriors from the Tollense Bronze Age battlefield in northern Germany (~3,200 BP), the oldest large-scale conflict site north of the Alps. Genetic data indicate that these individuals represent a single unstructured Central/Northern European population. We complemented these data with genotypes of 18 individuals from the Bronze Age site Mokrin in Serbia (~4,100 to ~3,700 BP) and 37 individuals from Eastern Europe and the Pontic-Caspian Steppe region, predating both Bronze Age sites (~5,980 to ~3,980 BP). We infer low LP in all three regions, i.e., in northern Germany and South-eastern and Eastern Europe, suggesting that the surge of rs4988235 in Central and Northern Europe was unlikely caused by Steppe expansions. We estimate a selection coefficient of 0.06 and conclude that the selection was ongoing in various parts of Europe over the last 3,000 years.

RESULTS AND DISCUSSION

The Majority of the Tollense Samples Are Male and Unrelated

We enriched DNA from 21 samples from the Tollense battlefield for 5 MB of putatively neutral regions and 487 phenotypically informative loci associated with metabolic syndrome, adult lactase persistence (LP), non-infectious and inflammatory diseases, and eye, skin, and hair pigmentation [6]. Of the 21 samples, two showed evidence of high contamination (>9% based on mtDNA) and one was not involved in the battle but rather dated to the Neolithic period. We analyzed the data of 14 of

Please cite this article in press as: Burger et al., Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.08.033

Table 1. Samples Analyzed and Frequency of the LP Allele (rs4988235-A) in Different Cultures

	Sample	Reference	Sample Age (calBP)	N	13.910*T Allele Frequency	90% CI	Genotyping
East European (Steppe)	Eneolithic	this study	5,200–5,978	2	0	0.000–0.602	PCR
	Usatovo	this study	4,950–5,450	1	0	0.000–0.842	PCR
	Yamnaya	this study	4,480–5,378	13	0	0.000–0.132	PCR
	Early catacomb culture	this study	4,341–4,674	8	0	0.000–0.206	PCR
	Developed catacomb culture	this study	4,250–4,564	2	0	0.000–0.602	PCR
	Yamnaya-Poltavkinskaja	this study	4,450–4,884	6	0	0.000–0.265	PCR
	Late catacomb culture	this study	3,975–4,706	5	0	0.000–0.3085	PCR
Q14	East European Steppe: 3,600-2,300 BCE	[2–4, 7–12]	4,277–5,471	37	0	0.000–0.040	NGS (PH)
	Area of Corded Ware Culture: 2,900-2,300 BCE	[2–4, 9–11, 13–18]	4,250–4,833	55	0,018	0.000–0.027	NGS (PH)
Rest of the continent	Mokrin (Bronze Age Serbia)	[19]	3,700–4,100	18[a]	0.046	0.001–0.145[a]	NGS
	Tollense	this study	3,100–3,200	14	0.071	0.009–0.235	NGS
	Prague (Jinonice, Zahradnictví and Kobylisy, Ke Stírce Street)	[3]	3,700–4,200	14	0.101	0.031–0.227	NGS
	Bedfordshire, Biddenham Loop	[3]	3,130–3,206	6	0.140	0.028–0.358	NGS
Q15	Lichtenstein Cave, Late Bronze Age in Germany	[20, 21]	2,700–3,000	34[a]	0.294	0.2–0.42[a]	PCR
	Kivutkalns, Baltic Bronze Age in Latvia	[22]	2,560–2,730	8	0.575	0.364–0.77	NGS
	Early Medieval Bavarians	[6]	~1,500	21	0.524	0.4–0.647	NGS
	Szolad (northern ancestry), Early Medieval in Hungary	[23]	1,500	13	0.727	0.567–0.856	NGS

Frequencies including 90% credible intervals (CIs) for NGS data were estimated based on genotype likelihoods using ATLAS. Frequencies based on pseudo-haploid (PH) calls and PCR data were estimated from allele counts taken from genotypes, and a 90% CI was calculated based on the beta distribution. Sample location as well as replication status of the LP genotype are given in Table S3. N, number of individuals.
[a]Family graves with high degrees of relatedness (see also Table S1)

the remaining individuals for which we successfully enriched DNA at targeted regions to a mean depth >4× (Tables 1 and S1). We detected no close relatives among these 14 individuals (Figure S1). Surprisingly, two of these individuals were women, consistent with a male-dominated, but not exclusively male, battle.

Tollense Sample Shows No Structure

Projected onto a principle-component analysis (PCA) plot trained on modern samples, all Tollense individuals fall within the range of Central and Northern European variation, suggestive of little or no genetic substructure (Figure 1). Multiple lines of evidence corroborate that impression: first, the spread of the Tollense samples on the PCA matches that of other ancient population samples of similar age, namely the Lech Valley in Bavaria (~4,700 to ~3,250 BP) [13] and Mokrin in Serbia (~4,100 to ~3,700 BP) [19]. Second, only seven (0.4%) out of the 1,638 D-statistics [24] of the form D (Tollense1 and Tollense2; Test_Population and YRI) were statistically significant with an absolute *Z* score cutoff of 3, and none were significant after correcting for multiple testing (Table S2). Among the ten highest D-statistic values, six indicated that WEZ35 was more distant from various European population samples than the other battlefield samples. We additionally tested for differences between visually suggestive upper and lower clusters in the PCA, but no D-statistic of the form D (YRI, Test_Population, Tollense_lower, and Tollense_upper) was significant. Third, F_{st} between random partitions of the Tollense sample were generally lower than those between random partitions of a modern 1000 Genomes CEU sample of the same size (P[F_{st} Tollense < F_{st} CEU] = 65.57%, 65.42%, 65.65%, and 65.35% for partitions of 4/10, 5/9, 6/8, or 7/7 individuals). Finally, we failed to reject Hardy-Weinberg equilibrium (HWE) using a novel method that accounts for genotyping uncertainty in ancient samples (see STAR Methods). Specifically, we tested for a deficit of heterozygous genotypes as quantified by the inbreeding coefficient F, which is expected to be positive among samples from a structured population due to the Wahlund effect but for which we estimated F = 0.0 (maximum *a posteriori* [MAP], 90% credible interval [CI] 0.0–$1.6 \cdot 10^{-4}$; Figure 2A). Contrasting this model with strict HWE, the latter received 99.94% posterior support. The Tollense individuals thus conform to a sample from a single unstructured population.

CURBIO 16792

Please cite this article in press as: Burger et al., Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.08.033

CellPress

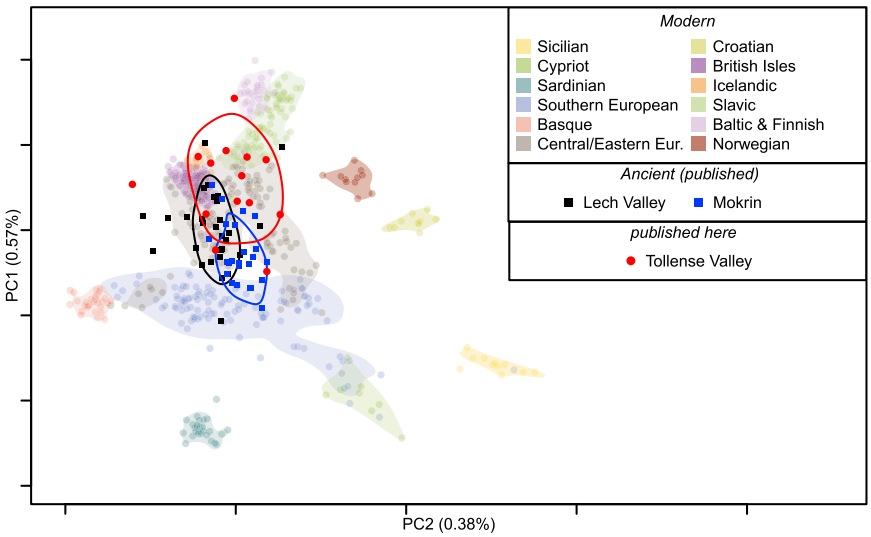

Figure 1. PCA of Modern and Ancient Europeans

Black squares, ancient samples from the Lech Valley in Bavaria; blue squares, ancient samples from Mokrin in Serbia; red circles, Tollense samples from this study. Ancient samples are projected onto modern reference sample space. Colored circles correspond to the 70% density contour lines of the ancient groups.

Tollense Population Is from Central/Northern Europe

The ancestry of the 14 Tollense individuals is Central to Northern European, as indicated by several analyses. First, the 70% density contour in the PCA shows a slightly more northern center for the Tollense sample than for those of the Bavarian Lech Valley and Serbian Mokrin individuals (Figure 1). Second, the Tollense sample appears closest to 5th century Bavarians [6] and modern Central and Northern Europeans when quantified with F_{st}, although all European populations have overlapping confidence intervals (Figure 2B). Finally, the same clustering was obtained with a TreeMix analysis (Figures 2C and S2) at the maximum-likelihood estimate of five migration events, of which an event between the Finnish and an East Asian population was the only one involving Europeans.

Based on archeological finds at the battlefield, it has been hypothesized that some of the warriors came from southern Central Europe, i.e., South-eastern Germany and Bohemia [25–27]. Carbon and strontium isotope ratios further suggest a mixture of locals and non-locals [28]. Although it is possible that our data are insufficient to resolve genetic affinities at such a small geographic scale, it appears likely that the Tollense individuals were sampled from a relatively homogeneous population with a high degree of continuity to those living in the same broad region today. But we caution that a strong correspondence between the material culture and genetic ancestry is only expected under conditions of strict and enduring population separation.

Major Allele Frequency Increase before the Medieval Period

To compare LP frequencies temporally and spatially, we determined rs4988235-A allele frequencies (also known as −13,910*T) using a Bayesian estimator in various ancient, post-Neolithic population samples with more than five individuals per site. We estimated low frequencies for both Tollense (7.1%; N = 14) and Mokrin (4.6%; N = 18) samples ($P[f_{Mokrin} < f_{Tollense}] = 0.889$). However, we caution that various individuals at the Mokrin site are close relatives. These estimates are consistent with those of other Bronze Age populations, albeit with a lack of high-quality data for comparison (Table 1). For instance, data from somewhat older sites in the UK (N = 6) [3] and Czechia (N = 14) [3] indicate allele frequencies of 17% and 10%, respectively, which are not significantly different from those of Tollense ($P[f_{UK} < f_{Tollense}] = 0.225$; $P[f_{Czechia} < f_{Tollense}] = 0.33$). However, these samples do not originate from a single time point or location and are based on extremely low allelic depth, few individuals, or both. The highest Bronze Age frequency (29%; N = 34) [20, 21] was observed for the Lichtenstein Cave in Germany, a family grave site a few centuries younger than Tollense.

Markedly higher frequencies are observed from more recent samples, suggesting ongoing strong selection on the rs4988235-A allele during the intermediate period. A sample from Latvia, dating to about 2,730–2,560 BP ($P[f_{Latvia} < f_{Tollense}] < 10^{-3}$; N = 8) [22], and a Medieval sample from Southern Germany (~1,500 BP; $P[f_{Germany} < f_{Tollense}] < 10^{-3}$; N = 21) [6] both have a frequency of roughly 57%, and a sample with Northern European ancestry from an Early Medieval cemetery in Hungary has a frequency of 73% ($P[f_{Latvia} < f_{Tollense}] < 10^{-3}$; N = 13) [23].

These frequency estimates indicate that, although the rs4988235-A allele had reached frequencies that make it detectable in small-sample-size ancient DNA studies by the beginning of the Bronze Age in various parts of Europe, it had not reached the frequencies observed in population samples dating from the Iron Age or later. Such a pattern is consistent with selection starting as early as the Neolithic but also indicates continuing strong selection during and particularly after the Bronze Age.

Selection Is Ongoing and Strong after the Neolithic Period

To quantify the selection strength on the rs4988235-A allele since the Bronze Age, we treated the Tollense individuals as sampled from an unstructured population with broad continuity to modern Central/Northern Europeans and examined allele frequency changes in the intervening 130 generations for all phenotypically relevant loci genotyped. We used the unrelated 1000 Genomes CEU samples as a modern reference population sample and employed two complementary methods: (1) a Bayesian approach (ApproxWF) [29] under an additive model (dominance coefficient of h = 0.5) and a constant population, the size of which we estimated based on allele frequencies at neutral loci, and (2) forward simulations under a model of exponential population growth [30].

Both methods identify two alleles influenced by strong positive selection: the LP causing allele rs4988235-A and rs7570971-C,

CURBIO 16792

Please cite this article in press as: Burger et al., Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.08.033

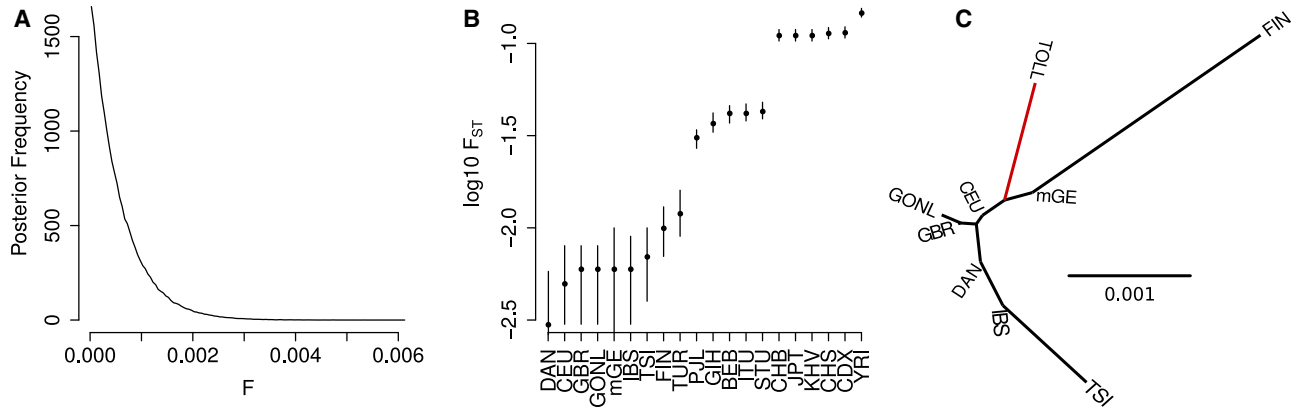

Figure 2. Tollense Samples Show No Population Structure

(A) Posterior density of inbreeding coefficient F (see also Figure S3).

(B) F_{ST} between Tollense samples at 5K neutralome against 5th century Bavarians (mGE), Dutch (GONL), Danish (DAN), Turkish (TUR), and Eurasian+YRI 1000 Genomes populations, excluding transition SNPs.

(C) TreeMix analysis with zoom on European populations allowing five migration events with an unrooted tree. CEU, Utah residents with northern and western European ancestry (see also Figure S2).

which occurs almost exclusively on the same haplotype as rs4988235-A and additionally plays a role in total serum cholesterol reduction [31]. Persuasive evidence for a functional role in LP strongly indicates that selection is acting on the rs4988235-A allele rather than on the linked rs7570971-C allele [32]. Within 3,000 years, sample counts for these alleles increased from 2/26 to 144/192 and 2/28 to 134/190, respectively. ApproxWF estimates a selection coefficient of around 5% for rs7570971-C and 6% for rs4988235-A, also when assuming dominance (h = 1; Figures 3B and 3C). Consistent with the low frequency we report here for the Bronze Age, these estimates are a bit higher than recent estimates from modern (e.g., 1.6%) [33] and ancient data (e.g., 1.8%) [34].

The forward simulation approach additionally rejected genetic drift as the only explanation for the inferred frequency change of the derived allele at rs5743810 ($p < 0.001$). This locus is located in the toll-like receptor gene complex (TLR6) and is associated with pathogen pattern recognition and innate immune response. Notably, none of the other 440 tested loci showed a significant signal of selection between the period of the Tollense site (3,200 BP) and today.

Eastern Europeans Steppes Are Not the Source for Lactase Persistence

The spread of the rs4988235-A allele has previously been attributed to a Steppe-associated expansion during the early Bronze Age [35]. Based on imputed data, Allentoft et al. [4] reported a low allele frequency (5%) of the rs4988235-A allele in Bronze Age Europeans, similar to that reported here, but a much higher frequency (~25%) among Bronze Age samples from the Pontic-Caspian Steppe, indicating a possible steppe origin of lactase persistence. Because imputing allele presence in ancient samples using modern reference individuals may be problematic in regions of strong recent selection, we investigated this hypothesis by genotyping the rs4988235 locus using PCR in Eneolithic and Early Bronze Age samples from Eastern Europe and the Steppe region. The majority of the sampled individuals were buried in barrow graves dating from the end of the 4th mill. BCE to the end of the 2nd mill. BCE and are representative for the Early Bronze Age in Eastern Europe (Figure 3A; Table 1). We could not detect the rs4988235-A allele among any of these samples (N = 37), suggesting that the frequency of this allele was very low, possibly close to zero, and almost certainly lower than the 5.4% previously reported for a geographically, culturally, and temporally diverse sample with "Steppe ancestry" [2]. Additionally, we re-analyzed published data from the Eastern European steppe area (5,600–4,300 BP) and that of the Corded Ware Culture in Central and North-eastern Europe (4,900–4,300 BP)—based on pseudo-haploid random allele picking—obtaining frequencies of 0% and 1.8%, respectively; this corresponds to a single LP-associated allele in 92 individuals (Table 1). Although these estimates are not directly informative about the origin of the rs4988235-A allele, they appear inconsistent with a major contribution of the Steppe-associated expansion to the high frequencies observed after the Bronze Age in Europe.

The Time Course of Selection

Although the LP-causing rs4988235-A allele has been under very strong natural selection at the broad geographic scale of Western Eurasia, considerable uncertainty and debate remains concerning the underlying drivers and spatiotemporal distribution of that selection and the role demographic processes played in shaping allele frequencies [36]. When considering the underlying drivers of selection, it is important to recognize that the date or origin of the rs4988235-A allele, the timing of selection on LP (whether constant from some point in the past or episodic), and the first observation of that allele in ancient DNA data are distinct and probably all separated by thousands of years. Linkage disequilibrium studies [37–39] indicate an allele origin in the last 20,000 years, perhaps in the Holocene. Evidence of milk consumption from analysis of fatty acids deposited on pot sherds [40–42] and from archaeological herd kill-off profiles [43] dates back to the Early Neolithic in Anatolia, the Levant, and South-eastern Europe. However, even if strong natural

Please cite this article in press as: Burger et al., Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.08.033

CellPress

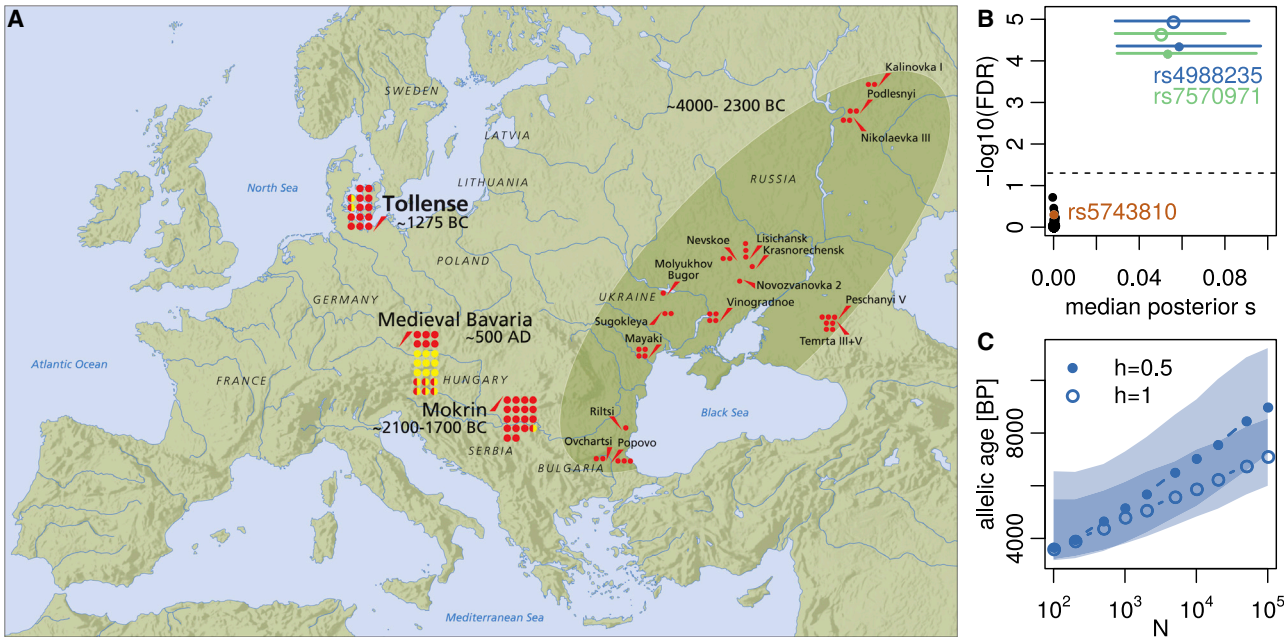

Figure 3. Lactase Persistence Distribution and Selection in Europe

(A) Map of lactase persistence in the Bronze Ages sites of Tollense and Mokrin, in Eneolithic and Early Bronze Age samples from Eastern Europe and the Pontic-Caspian Steppe area, and an Early Medieval sample from Bavaria. Red corresponds to the ancestral and yellow to the derived allele associated with LP (see also Table S3). The shaded shape encompasses the samples from the east European Steppe.

(B) For each phenotypic locus, we show the posterior selection coefficient and the false discovery rate (FDR) as inferred with ApproxWF. Loci identified as under selection by forward simulations approach are colored. Open circles show estimates obtained under a dominant model (h = 1). The dashed line indicates the 5% FDR cutoff used.

(C) Posterior densities estimated by ApproxWF under the additive (h = 0.5, solid) and dominant (h = 1, dashed) model for the loci identified as under selection by both methods.

selection favoring LP has been operating since the Early Neolithic, we would expect sigmoidal allele frequency trajectories and so low and barely detectable frequencies for up to several thousand years. Thus, low-frequency estimates until the later Bronze Age [2–4, 13, 33, 44, 45] (this study) are not necessarily inconsistent with selection on LP starting when or shortly after milk became a significant dietary component.

To illustrate this, we inferred the age of the rs4988235-A allele from the Tollense data using a novel Bayesian method and assuming a hard sweep [35] with selection coefficient of 6% since the mutation occurred (Figure 3C). The inferred allele age is highly dependent on the unknown effective population size between the time of mutation and the Bronze age, as well as the dominance coefficient h. Assuming dominance (h = 1), the MAP estimate varied between 3,550 (90% CI 3,200–5,100) and 7,140 (90% CI 5,525–7,650) years BP for population sizes $2N = 10^2$ and $2N = 10^5$, respectively, and was about 20% older when assuming additivity (h = 0.5). In line with the low frequency for the Tollense sample, these estimates are at the lower end of the age range [35] (3,280–23,100 BP) or considerably younger [33] (17,500 BP) than recent estimates from modern data. However, we note that all such estimates make strong assumptions about the demographic history, the strength of selection, as well as the origin of the selected allele.

The Bronze Age therefore represents an important waypoint at which the rs4988235-A allele became sufficiently common to be detectable in reasonably sized ancient DNA samples. Data from the Bronze Age can then serve as a starting point to quantify selection reliably in subsequent times. The Tollense sample, consisting of individuals that most likely died in a battle within a short period of time, possibly on the same day, is well suited for this. The inferred selection coefficient of 6% between Tollense (~3,200 BP) and today is high, especially considering that the advancement of agricultural skills and the increase in dietary breadth since the Neolithic should have produced alternatives to milk consumption.

Drivers of Selection

Beyond milk being a nutrient-dense and relatively nutrient-balanced food, various explanations have been offered for the strong selection inferred for LP. These include improved calcium absorption by supplementing vitamin-D-poor diets at high latitudes [46], the supply of a relatively pathogen-free fluid [47–49], the suppression of malaria symptoms through a reduction of p-aminobenzoic acid consumption [50, 51], improvements in gut health through galactose and galacto-oligosaccharides reshaping the colonic microbiome [36, 45, 52–55], avoidance of diarrhea under famine conditions [56], and increased economic efficiency of calorie production for dairy farming [57]. It is unlikely that any single one of these factors has acted alone over the whole period from the Neolithic to modern times. However, the inference of an ongoing and possibly increased selection coefficient from the Bronze Age to at least Medieval times could be interpreted as favoring factors related to increases in population

Please cite this article in press as: Burger et al., Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.08.033

and settlement density, such as those concerning pathogen loads [36, 56]. In this context, it is interesting to note that, among the other >400 functional loci we examined, the only other allele reaching significance in terms of a signature of selection is the derived allele at rs5743810 in the toll-like receptor gene complex (TLR6), which is associated with pathogen pattern recognition and innate immune response. Although difficult to test explicitly, there remains a strong possibility that selection on the rs4988235-A allele is modulated by other genetic factors through a pleiotropic network, possibly in relation to epidemic disease resistance in the context of milk drinking.

Demography and Selection

Aside from strong natural selection, demographic processes, such as migration and population range expansion—leading to allele surfing [58, 59]—will have shaped the distribution of the rs4988235-A allele [60]. It is not clear whether the northwestern distribution of this allele in Europe is primarily the result of these processes or spatially structured selection strengths shaped by, for example, different traditions of milk use, different climates and ecologies, or different levels of incident ultraviolet radiation. However, it is noteworthy that there has been little genetic turnover between the Tollense population sampled in this study and populations in the same region today, yet we still infer a high selection coefficient. The same applies to the Mokrin site in Serbia, where a major population shift is equally unlikely—leaving natural selection as the main explanation for the observed allele frequency change over the last few millennia.

Final Remark

The somewhat similar inferred rs4988235-A allele frequencies between the early Middle Ages and today at different locations does not exclude the possibility of ongoing selection during that period [61], because this likely represents the transitional phase of the sigmoid frequency trajectory expected for a dominant trait. For LP, it now seems likely that the phase of most rapid frequency rise was between 4,000 BP and 1,500 BP. We contend that research should be focused on this phase to better understand the evolutionary history of the most strongly selected single gene trait in Holocene Western Eurasia and many other parts of the world.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Archaeological and morphological information
- METHOD DETAILS
  - Sample preparation
  - Library construction and amplification
  - Illumina sequencing
  - 5 MB capture assay
  - Read mapping, alignment cleaning and merging
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Sex Determination
  - Ancient DNA Authenticity
  - Haplogroups of uniparental Markers
  - Genotype likelihoods in Tollense samples
  - Pseudo-haploid calls
  - Selection inference
  - Selection inference with ApproxWF
  - Selection inference with forward simulations
  - Population StructurePrincipal component analysis
  - F_{ST} and TreeMix
  - D-statistics with with ancient individuals
  - Population structure within the Tollense > 4x sample
  - Inferring global inbreeding coefficient F
  - Model
  - Priors
  - Inference
  - Allele frequency estimation for NGS data
  - Allele frequency estimation for PCR data
  - Estimating Allele Age
  - Inference

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2020.08.033.

ACKNOWLEDGMENTS

This work was funded by grants BO4119/1-1 of the German Science Foundation to R.B. and J. Burger and 31003A_173062 of the Swiss National Science Foundation to D.W. We thank the Genome of the Netherlands Consortium for access to BAM files. Parts of this research were conducted using the supercomputer Mogon and/or advisory services offered by Johannes Gutenberg University Mainz (https://hpc.uni-mainz.de), which is a member of the AHRP (Alliance for High Performance Computing in Rhineland Palatinate; https://www.ahrp.info) and the Gauss Alliance e.V. Genomic data are available at the European Nucleotide Archive under the accession no. PRJEB38406 in BAM and fastq format.

AUTHOR CONTRIBUTIONS

J. Burger, D.W., T.T., and R.B. initiated this research. A. Schulzd, A.Ž., L.W., and R.B. produced data. C.S., J. Blöcher, K.R.V., V.L., J. Burger, Z.H., J.O., A. Scheua, V.B., Z.P., Y.D., and C.L. analyzed data. J. Burger, V.L., D.W., K.R.V., and M.G.T. wrote the article with the help of all co-authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 4, 2020
Revised: July 7, 2020
Accepted: August 7, 2020
Published: September 3, 2020

REFERENCES

1. Enattah, N.S., Jensen, T.G.K., Nielsen, M., Lewinski, R., Kuokkanen, M., Rasinpera, H., El-Shanti, H., Seo, J.K., Alifrangis, M., Khalil, I.F., 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.

2. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499–503.

3. Olalde, I., Brace, S., Allentoft, M.E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., et al. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. Nature *555*, 190–196.

4. Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of bronze age eurasia. Nature *522*, 167–172.

5. Furholt, M., Grossmann, R., and Szmyt, M. (2016). Corded Ware and Bell Beakers: A Practice-Based Perspective on Local Communities, Transregional Interaction and Social Heterogeneity in Late Neolithic Europe (Transitional Landscape), pp. 117–131.

6. Veeramah, K.R., Rott, A., Groß, M., van Dorp, L., López, S., Kirsanow, K., Sell, C., Blöcher, J., Wegmann, D., Link, V., et al. (2018). Population genomic analysis of elongated skulls reveals extensive female-biased immigration in Early Medieval Bavaria. Proc. Natl. Acad. Sci. USA *115*, 3494–3499.

7. Järve, M., Saag, L., Scheib, C.L., Pathak, A.K., Montinaro, F., Pagani, L., Flores, R., Guellil, M., Saag, L., Tambets, K., et al. (2019). Shifts in the genetic landscape of the Western Eurasian Steppe associated with the beginning and end of the Scythian dominance. Curr. Biol. *29*, 2430–2441.e10.

8. Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., et al. (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. Nature *551*, 368–372.

9. Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkhoshbacht, N., Candilio, F., Cheronet, O., et al. (2018). The genomic history of southeastern Europe. Nature *555*, 197–203.

10. Narasimhan, V.M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., Lazaridis, I., Nakatsuka, N., Olalde, I., Lipson, M., et al. (2019). The formation of human populations in South and Central Asia. Science *365*, eaat7487.

11. Schroeder, H., Margaryan, A., Szmyt, M., Theulot, B., Włodarczak, P., Rasmussen, S., Gopalakrishnan, S., Szczepanek, A., Konopka, T., Jensen, T.Z.T., et al. (2019). Unraveling ancestry, kinship, and violence in a Late Neolithic mass grave. Proc. Natl. Acad. Sci. USA *116*, 10705–10710.

12. Wang, C.C., Reinhold, S., Kalmykov, A., Wissgott, A., Brandt, G., Jeong, C., Cheronet, O., Ferry, M., Harney, E., Keating, D., et al. (2019). Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. Nat. Commun. *10*, 590.

13. Mittnik, A., Massy, K., Knipper, C., Wittenborn, F., Friedrich, R., Pfrengle, S., Burri, M., Carlichi-Witjes, N., Deeg, H., Furtwängler, A., et al. (2019). Kinship-based social inequality in bronze age europe. Science *366*, 731–734.

14. Fernandes, D.M., Strapagiel, D., Borówka, P., Marciniak, B., Żądzińska, E., Sirak, K., Siska, V., Grygiel, R., Carlsson, J., Manica, A., et al. (2018). A genomic Neolithic time transect of hunter-farmer admixture in central Poland. Sci. Rep. *8*, 14879.

15. Gamba, C., Jones, E.R., Teasdale, M.D., McLaughlin, R.L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kővári, I., Pap, I., Anders, A., et al. (2014). Genome flux and stasis in a five millennium transect of European prehistory. Nat. Commun. *5*, 5257.

16. Jones, E.R., Zarina, G., Moiseyev, V., Lightfoot, E., Nigst, P.R., Manica, A., Pinhasi, R., and Bradley, D.G. (2017). The Neolithic transition in the Baltic was not driven by admixture with early European farmers. Curr. Biol. *27*, 576–582.

17. Sánchez-Quinto, F., Malmström, H., Fraser, M., Girdland-Flink, L., Svensson, E.M., Simões, L.G., George, R., Hollfelder, N., Burenhult, G., Noble, G., et al. (2019). Megalithic tombs in western and northern Neolithic Europe were linked to a kindred society. Proc. Natl. Acad. Sci. USA *116*, 9469–9474.

18. Skoglund, P., Malmström, H., Omrak, A., Raghavan, M., Valdiosera, C., Günther, T., Hall, P., Tambets, K., Parik, J., Sjögren, K.G., et al. (2014). Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. Science *344*, 747–750.

19. Žegarac, A., Winkelbach, L., Blöcher, J., Diekmann, Y., Gavrilović, M.K., Porčić, M., Stojković, B., Milašinović, L., Schreiber, M., Wegmann, D., et al. (2020). Kinship, acquired and inherited status, and population structure at the Early Bronze Age Mokrin necropolis in northern Serbia. bioRxiv. https://doi.org/10.1101/2020.05.18.101337.

20. Schilz, F. (2006). Molekulargenetische verwandtschaftsanalysen am prähistorischen skelettkollektiv der Lichtensteinhöhle. PhD thesis (University of Göttingen). http://webdoc.sub.gwdg.de/diss/2006/schilz/schilz.pdf.

21. Seidenberg, V. (2016). Ein bronzezeitlicher familienclan als genetisches archiv – morphologisch-paläogenetische bearbeitung des skelettkollektivs aus der Lichtensteinhöhle. PhD thesis (University of Göttingen). https://ediss.uni-goettingen.de/bitstream/handle/11858/00-1735-0000-002B-7D55-C/Dissertation_Seidenberg.pdf?sequence=1.

22. Mittnik, A., Wang, C.-C., Pfrengle, S., Daubaras, M., Zariņa, G., Hallgren, F., Allmäe, R., Khartanovich, V., Moiseyev, V., Tõrv, M., et al. (2018). The genetic prehistory of the Baltic Sea region. Nat. Commun. *9*, 442.

23. Amorim, C.E.G., Vai, S., Posth, C., Modi, A., Koncz, I., Hakenbeck, S., La Rocca, M.C., Mende, B., Bobo, D., Pohl, W., et al. (2018). Understanding 6th-century barbarian social organization and migration through paleogenomics. Nat. Commun. *9*, 3547.

24. Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y., et al. (2010). A draft sequence of the Neandertal genome. Science *328*, 710–722.

25. Uhlig, T., Krüger, J., Lidke, G., Jantzen, D., Lorenz, S., Ialongo, N., and Terberger, T. (2019). Lost in combat? A scrap metal find from the Bronze Age battlefield site at Tollense. Antiquity *93*, 1211–1230.

26. Terberger, T., and Heinemeier, J. (2014). Die Ernährungsweise der bronzezeitlichen Menschen aus dem Tollensetal im Spiegel ihrer 13C und 15N Isotopie - erste Ergebnisse. Tod im Tollensetal—-Forschungen zu den Hinterlassenschaften eines bronzezeitlichen Gewaltkonfliktes in Mecklenburg-Vorpommern *1*, 209–214.

27. Jantzen, D., Brinker, U., Orschiedt, J., Heinemeier, J., Piek, J., Hauenstein, K., Krüger, J., Lidke, G., Lübke, H., Lampe, R., et al. (2011). A Bronze Age battlefield? weapons and trauma in the Tollense Valley, north-eastern Germany. Antiquity *85*, 417–433.

28. Price, T.D., Frei, R., Brinker, U., Lidke, G., Terberger, T., Frei, K.M., and Jantzen, D. (2019). Multi-isotope proveniencing of human remains from a Bronze Age battlefield in the Tollense Valley in northeast Germany. Archaeol. Anthropol. Sci. *11*, 33–49.

29. Ferrer-Admetlla, A., Leuenberger, C., Jensen, J.D., and Wegmann, D. (2016). An approximate markov model for the wright–fisher diffusion and its application to time series data. Genetics *203*, 831–846.

30. Wilde, S., Timpson, A., Kirsanow, K., Kaiser, E., Kayser, M., Unterländer, M., Hollfelder, N., Potekhina, I.D., Schier, W., Thomas, M.G., et al. (2014). Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. Proc. Natl. Acad. Sci. USA *111*, 4832–4837.

31. Zhang, G., Muglia, L.J., Chakraborty, R., Akey, J.M., and Williams, S.M. (2013). Signatures of natural selection on genetic variants affecting complex human traits. Appl. Transl. Genomics *2*, 78–94.

32. Lewinsky, R.H., Jensen, T.G.K., Møller, J., Stensballe, A., Olsen, J., and Troelsen, J.T. (2005). T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity in vitro. Hum. Mol. Genet. *14*, 3945–3953.

33. Stern, A.J., Wilton, P.R., and Nielsen, R. (2019). An approximate full-likelihood method for inferring selection and allele frequency trajectories from DNA sequence data. PLoS Genet. *15*, e1008384.

34. Mathieson, S., and Mathieson, I. (2018). FADS1 and the timing of human adaptation to agriculture. Mol. Biol. Evol. *35*, 2957–2970.

35. Satta, Y., and Takahata, N. (2020). Population genomics on the origin of lactase persistence in Europe and South Asia. bioRxiv. https://doi.org/10.1101/2020.06.30.179432.

36. Walker, C., and Thomas, M.G. (2019). The evolution of lactose digestion. In Lactose: Evolutionary Role, Health Effects, and Applications, M. Paques, and C. Lindner, eds. (Academic), pp. 1–48.

37. Bersaglieri, T., Sabeti, P.C., Patterson, N., Vanderploeg, T., Schaffner, S.F., Drake, J.A., Rhodes, M., Reich, D.E., and Hirschhorn, J.N. (2004). Genetic signatures of strong recent positive selection at the lactase gene. Am. J. Hum. Genet. *74*, 1111–1120.

38. Gallego Romero, I., Basu Mallick, C., Liebert, A., Crivellaro, F., Chaubey, G., Itan, Y., Metspalu, M., Eaaswarkhanth, M., Pitchappan, R., Villems, R., et al. (2012). Herders of Indian and European cattle share their predominant allele for lactase persistence. Mol. Biol. Evol. *29*, 249–260.

39. Liebert, A., López, S., Jones, B.L., Montalva, N., Gerbault, P., Lau, W., Thomas, M.G., Bradman, N., Maniatis, N., and Swallow, D.M. (2017). World-wide distributions of lactase persistence alleles and the complex effects of recombination and selection. Hum. Genet. *136*, 1445–1453.

40. Craig, O.E., Chapman, J., Heron, C., Willis, L.H., Bartosiewicz, L., Taylor, G., Whittle, A., and Collins, M. (2005). Did the first farmers of central and eastern Europe produce dairy foods? Antiquity *79*, 882–894.

41. Evershed, R.P., Payne, S., Sherratt, A.G., Copley, M.S., Coolidge, J., Urem-Kotsu, D., Kotsakis, K., Ozdoğan, M., Ozdoğan, A.E., Nieuwenhuyse, O., et al. (2008). Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. Nature *455*, 528–531.

42. Salque, M., Bogucki, P.I., Pyzel, J., Sobkowiak-Tabaka, I., Grygiel, R., Szmyt, M., and Evershed, R.P. (2013). Earliest evidence for cheese making in the sixth millennium BC in northern Europe. Nature *493*, 522–525.

43. Vigne, J.-D., and Helmer, D. (2007). Was milk a "secondary product" in the Old World Neolithisation process? Its role in the domestication of cattle, sheep and goats. Anthropozoologica *42*, 9–40.

44. Burger, J., Kirchner, M., Bramanti, B., Haak, W., and Thomas, M.G. (2007). Absence of the lactase-persistence-associated allele in early Neolithic Europeans. Proc. Natl. Acad. Sci. USA *104*, 3736–3741.

45. Segurel, L., Guarino-Vignon, P., Marchi, N., Lafosse, S., Laurent, R., Bon, C., Fabre, A., Hegay, T., and Heyer, E. (2020). Why and when was lactase persistence selected for? Insights from Central Asian herders and ancient DNA. PLoS Biol. *18*, e3000742.

46. Flatz, G., and Rotthauwe, H.W. (1973). Lactose nutrition and natural selection. Lancet *2*, 76–77.

47. Cook, G.C., and al-Torki, M.T. (1975). High intestinal lactase concentrations in adult Arbs in Saudi Arabia. BMJ *3*, 135–136.

48. Cook, G.C. (1978). Breath hydrogen concentrations after oral lactose and lactulose in tropical malabsorption and adult hypolactasia. Trans. R. Soc. Trop. Med. Hyg. *72*, 277–281.

49. Gerbault, P., Liebert, A., Itan, Y., Powell, A., Currat, M., Burger, J., Swallow, D.M., and Thomas, M.G. (2011). Evolution of lactase persistence: an example of human niche construction. Philos. Trans. R. Soc. Lond. B Biol. Sci. *366*, 863–877.

50. Cordain, L., Hickey, M.S., and Kim, K. (2012). Malaria and rickets represent selective forces for the convergent evolution of adult lactase persistence. In Biodiversity in Agriculture: Domestication, Evolution, and Sustainability, P. Gepts, T.R. Famula, R.L. Bettinger, S.B. Brush, A.B. Damania, P.E. McGuire, and C.O. Qualset, eds. (Cambridge University), pp. 299–308.

51. Lokki, A.I., Järvelä, I., Israelsson, E., Maiga, B., Troye-Blomberg, M., Dolo, A., Doumbo, O.K., Meri, S., and Holmberg, V. (2011). Lactase persistence genotypes and malaria susceptibility in Fulani of Mali. Malar. J. *10*, 9.

52. Cederlund, A., Kai-Larsen, Y., Printz, G., Yoshio, H., Alvelius, G., Lagercrantz, H., Strömberg, R., Jörnvall, H., Gudmundsson, G.H., and Agerberth, B. (2013). Lactose in human breast milk an inducer of innate immunity with implications for a role in intestinal homeostasis. PLoS ONE *8*, e53876.

53. Gibson, G.R., Hutkins, R., Sanders, M.E., Prescott, S.L., Reimer, R.A., Salminen, S.J., Scott, K., Stanton, C., Swanson, K.S., Cani, P.D., et al. (2017). Expert consensus document: The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics. Nat. Rev. Gastroenterol. Hepatol. *14*, 491–502.

54. Wahlqvist, M.L. (2015). Lactose nutrition in lactase nonpersisters. Asia Pac. J. Clin. Nutr. *24* (*Suppl 1*), S21–S25.

55. Knol, J., Scholtens, P., Kafka, C., Steenbakkers, J., Gro, S., Helm, K., Klarczyk, M., Schöpfer, H., Böckler, H.M., and Wells, J. (2005). Colon microflora in infants fed formula with galacto- and fructo-oligosaccharides: more like breast-fed infants. J. Pediatr. Gastroenterol. Nutr. *40*, 36–42.

56. Sverrisdóttir, O.O., Timpson, A., Toombs, J., Lecoeur, C., Froguel, P., Carretero, J.M., Arsuaga Ferreras, J.L., Götherström, A., and Thomas, M.G. (2014). Direct estimates of natural selection in Iberia indicate calcium absorption was not the only driver of lactase persistence in Europe. Mol. Biol. Evol. *31*, 975–983.

57. Ingold, T. (1980). Hunters, Pastoralists and Ranchers: Reindeer Economies and Their Transformations (Cambridge University).

58. Klopfstein, S., Currat, M., and Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. Mol. Biol. Evol. *23*, 482–490.

59. Edmonds, C.A., Lillie, A.S., and Cavalli-Sforza, L.L. (2004). Mutations arising in the wave front of an expanding population. Proc. Natl. Acad. Sci. USA *101*, 975–979.

60. Itan, Y., Powell, A., Beaumont, M.A., Burger, J., and Thomas, M.G. (2009). The origins of lactase persistence in Europe. PLoS Comput. Biol. *5*, e1000491.

61. Montalva, N., Adhikari, K., Liebert, A., Mendoza-Revilla, J., Flores, S.V., Mace, R., and Swallow, D.M. (2019). Adaptation to milking agropastoralism in Chilean goat herders and nutritional benefit of lactase persistence. Ann. Hum. Genet. *83*, 11–22.

62. Church, D.M., Schneider, V.A., Graves, T., Auger, K., Cunningham, F., Bouk, N., Chen, H.-C., Agarwala, R., McLaren, W.M., Ritchie, G.R.S., et al. (2011). Modernizing reference genome assemblies. PLoS Biol. *9*, e1001091.

63. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R.; 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.

64. Alkan, C., Kavak, P., Somel, M., Gokcumen, O., Ugurlu, S., Saygi, C., Dal, E., Bugra, K., Güngör, T., Sahinalp, S.C., et al. (2014). Whole genome sequencing of Turkish genomes reveals functional private alleles and impact of genetic interactions with Europe, Asia and Africa. BMC Genomics *15*, 963.

65. Besenbacher, S., Liu, S., Izarzugaza, J.M., Grove, J., Belling, K., Bork-Jensen, J., Huang, S., Als, T.D., Li, S., Yadav, R., et al. (2015). Novel variation and de novo mutation rates in population-wide de novo assembled Danish trios. Nat. Commun. *6*, 5969.

66. Genome of the Netherlands Consortium (2014). Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat. Genet. *46*, 818–825.

67. Lazaridis, I., Mittnik, A., Patterson, N., Mallick, S., Rohland, N., Pfrengle, S., Furtwängler, A., Peltzer, A., Posth, C., Vasilakis, A., et al. (2017). Genetic origins of the Minoans and Mycenaeans. Nature *548*, 214–218.

68. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*, pdb.prot5448.

69. Link, V., Kousathanas, A., Veeramah, K., Sell, C., Scheu, A., and Wegmann, D. (2017). ATLAS: analysis tools for low-depth and ancient samples. bioRxiv. https://doi.org/10.1101/105346.

70. Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. *8*, e1002967.

71. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.

72. Wang, C., Zhan, X., Bragg-Gresham, J., Kang, H.M., Stambolian, D., Chew, E.Y., Branham, K.E., Heckenlively, J., Fulton, R., Wilson, R.K., et al.; FUSION Study (2014). Ancestry estimation and control of population stratification for sequence-based association studies. Nat. Genet. *46*, 409–415.

73. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.

74. Poznik, G.D., Xue, Y., Mendez, F.L., Willems, T.F., Massaia, A., Wilson Sayres, M.A., Ayub, Q., McCarthy, S.A., Narechania, A., Kashin, S., et al.; 1000 Genomes Project Consortium (2016). Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. Nat. Genet. *48*, 593–599.

75. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

76. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

77. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

78. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.

79. Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L.F., Aximu-Petri, A., Prüfer, K., de Filippo, C., et al. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. Nature *514*, 445–449.

80. Lidke, G., Jantzen, D., Lorenz, S., and Terberger, T. (2018). The Bronze Age battlefield in the Tollense Valley, Northeast Germany. Conflict scenario research. In Conflict Archaeology. Materialities of Collective Violence from Prehistory to Late Antiquity. EAA Themes in Contemporary Archaeology 5, M. Fernández-Götz, and N. Roymans, eds. (Routledge), pp. 61–68.

81. Brinker, U., Flohr, S., Piek, J., and Orschiedt, J. (2015). Human remains from a Bronze Age site in the Tollense Valley: victims of a battle? In The Routledge Handbook of the Bioarchaeology of Human Conflict, C. Knüsel, and M.J. Smith, eds. (Routledge), pp. 146–160.

82. Brinker, U., Schramm, A., Flohr, S., Jantzen, D., Piek, J., Hauenstein, K., and Orschiedt, J. (2016). The Bronze Age battlefield in the Tollense Valley, Mecklenburg-Western Pomerania, northeast Germany – Combat marks on human bones as evidence of early warrior societies in northern Middle Europe? In Later Prehistory to the Bronze Age. 1. The Emergence of Warrior Societies and Its Economic, Social and Environmental Consequences. 2. Imports and Aegeo-Mediterranean Influences on the Continental European Tombs in the Bronze and Iron Ages. Proceedings of the XVII UISPP World Congress (1–7 September 2014, Burgos, Spain) 9, F. Coimbra, D. Delfino, V. Sirbu, and C. Schuster, eds. (Archaeopress), pp. 39–56.

83. Brinker, U., Harten-Buga, H., Staude, A., Jantzen, D., and Orschiedt, J. (2018). Perimortem lesions on human bones from the Bronze Age battlefield in the Tollense Valley: An interdisciplinary approach. In Prehistoric Warfare and Violence. Quantitative and Qualitative Approaches, A. Dolfini, R. Crellin, C. Horn, and M. Uckelmann, eds. (Springer), pp. 39–60.

84. Terberger, T., Jantzen, D., Krüger, J., and Lidke, G. (2018). Das bronzezeitliche Kampfgeschehen im Tollensetal – ein Großereignis oder wiederholte Konflikte? In Bronzezeitliche Burgen zwischen Taunus und Karpaten / Bronze Age Hillforts between Taunus and Carpathian Mountains. UPA 319, S. Hansen, and R. Krause, eds. (Habelt), pp. 103–122.

85. Kachalova, N.K. (2002). O sootnoshenii Yamnoi i Poltavkinskoi kultur. In Problemy arkheologii Evrazii. K 80-letiyu N, P.M. Munchaev, ed. (Ya. Merperta), pp. 117–126.

86. Bollongino, R., Nehlich, O., Richards, M.P., Orschiedt, J., Thomas, M.G., Sell, C., Fajkosová, Z., Powell, A., and Burger, J. (2013). 2000 years of parallel societies in Stone Age Central Europe. Science *342*, 479–481.

87. Bramanti, B., Thomas, M.G., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M.N., Jankauskas, R., Kind, C.J., et al. (2009). Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. Science *326*, 137–140.

88. Aronesty, E. (2011). ea-utils: "Command-line tools for processing biological sequencing data.". https://github.com/ExpressionAnalysis/ea-utils.

89. Skoglund, P., Storå, J., Götherström, A., and Jakobsson, M. (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. J. Archaeol. Sci. *40*, 4477–4482.

90. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

91. Kousathanas, A., Leuenberger, C., Link, V., Sell, C., Burger, J., and Wegmann, D. (2017). Inferring heterozygosity from ancient and low coverage genomes. Genetics *205*, 317–332.

92. Dimitrieva, S., and Bucher, P. (2013). UCNEbase–a database of ultraconserved non-coding elements and genomic regulatory blocks. Nucleic Acids Res. *41* (Database issue, D1), D101–D109.

93. Waples, R.K., Albrechtsen, A., and Moltke, I. (2019). Allele frequency-free inference of close familial relationships from genotypes or low-depth sequencing data. Mol. Ecol. *28*, 35–48.

94. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, arXiv:1303.3997v2. https://arxiv.org/abs/1303.3997.

95. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. Nature *536*, 419–424.

96. Bhatia, G., Patterson, N., Sankararaman, S., and Price, A.L. (2013). Estimating and interpreting F_{ST}: the impact of rare variants. Genome Res. *23*, 1514–1521.

97. Busing, F.M.T.A., Meijer, E., and Van Der Leeden, R. (1999). Delete-m jackknife for unequal m. Stat. Comput. *9*, 3–8.

98. Leseva, M.N., Grand, R.J., Klett, H., Boerries, M., Busch, H., Binder, A.M., and Michels, K.B. (2018). Differences in DNA methylation and functional expression in lactase persistent and non-persistent individuals. Sci. Rep. *8*, 5649. Q12

99. Wilde, S. (2015). Populationsgenetik kupfer-und bronzezeitlicher Bevölkerungen der osteuropäischen Steppe. PhD thesis (Universitätsbibliothek Mainz). Q13

Please cite this article in press as: Burger et al., Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.08.033

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Tollense capture raw data	This study	PRJEB38406
PCR genotype estimates of 37 individuals from the following cultures: Eneolithic, Usatovo, Yamnaya, Early Catacomb Culture, Developed Catacomb Culture, Yamnaya-Poltavkinskaja, Late Catacomb Culture	This study	Table S1
Human reference genome NCBI build 37	[62]	GRCh37
Unrelated CEU from 1000 Genomes	[63]	NA06984, NA06985, NA06986, NA06989, NA06994, NA07000, NA07037, NA07048, NA07051, NA07056, NA07346, NA07347, NA07357, NA10847, NA10851, NA11829, NA11830, NA11831, NA11832, NA11840, NA11843, NA11881, NA11892, NA11893, NA11894, NA11918, NA11919, NA11920, NA11930, NA11931, NA11932, NA11933, NA11992, NA11993, NA11994, NA11995, NA12003, NA12004, NA12005, NA12006, NA12043, NA12044, NA12045, NA12046, NA12058, NA12144, NA12154, NA12155, NA12156, NA12234, NA12249, NA12273, NA12275, NA12282, NA12283, NA12286, NA12287, NA12340, NA12341, NA12342, NA12347, NA12348, NA12383, NA12399, NA12400, NA12413, NA12414, NA12489, NA12546, NA12716, NA12717, NA12718, NA12748, NA12749, NA12750, NA12751, NA12760, NA12763, NA12775, NA12778, NA12813, NA12814, NA12815, NA12827, NA12828, NA12829, NA12830, NA12842, NA12843, NA12872, NA12873, NA12874, NA12889, NA12890, NA12891, NA12892
1000 Genomes Eurasian Yoruba	[63]	NA18486, NA18488, NA18489, NA18498, NA18499, NA18501, NA18502, NA18504, NA18505, NA18507, NA18508, NA18510, NA18511, NA18516, NA18517, NA18519, NA18520, NA18522, NA18523, NA18853, NA18856, NA18858, NA18861, NA18864, NA18865, NA18867, NA18868, NA18870, NA18871, NA18873, NA18874, NA18876, NA18877, NA18878, NA18879, NA18881, NA18907, NA18908, NA18909, NA18910, NA18912, NA18915, NA18916, NA18917, NA18923, NA18924, NA18933, NA18934, NA19092, NA19093, NA19095, NA19096, NA19098, NA19099, NA19102, NA19107, NA19108, NA19113, NA19114, NA19116, NA19117, NA19118, NA19119, NA19121, NA19129, NA19130, NA19131, NA19137, NA19138, NA19141, NA19143, NA19144, NA19146, NA19147, NA19149, NA19152, NA19153, NA19159, NA19160, NA19171, NA19172, NA19175, NA19184, NA19185, NA19189, NA19190, NA19197, NA19198, NA19200, NA19201, NA19204, NA19206, NA19207, NA19209, NA19210, NA19213, NA19214, NA19222, NA19223, NA19225, NA19235, NA19236, NA19238, NA19239, NA19247, NA19248, NA19256, NA19257
Turkish genomes	[64]	

(*Continued on next page*)

Please cite this article in press as: Burger et al., Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.08.033

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
		06A010111, 08P210611, 24D220611, 25A220611, 31P140611, 32A140611, 33M140611, 34S291210, 35C240511, 38I220611, 42S291210, 48S210611, 50G301210, 52C130611, 57M220611, 65A220611
5th century Bavarian individuals	[6]	AED_106, AED_1119, AED_1135, AED_204, AED_249, AED_432, AED_92, ALH_10, ALH_1, ALH_2, ALH_3, NW_255, STR_241, STR_248, STR_266, STR_300, STR_316, STR_393, STR_480, STR_486, STR_502
Prague (Jinonice, Zahradnictví and Kobylisy, Ke Stírce Street)	[3]	I7195, I7196, I7197, I7198, I7199, I7200, I7201, I7202, I7203, I4886, I4888, I4889, I4890, I4891, I4945
Bedfordshire, Biddenham Loop	[3]	I7575, I7576, I7577, I7578, I7580, I7626, I7628
Kivutkalns, Baltic Bronze Age in Latvia	[22]	Kivutkalns153, Kivutkalns164, Kivutkalns194, Kivutkalns19, Kivutkalns207, Kivutkalns215, Kivutkalns222, Kivutkalns25, Kivutkalns42
Szolad (northern ancestry), Early Medieval in Hungary	[23]	SZ2, SZ3, SZ4, SZ5, SZ9, SZ11, SZ13, SZ15, SZ16, SZ18, SZ30, SZ38, SZ45
Genome DK (Danish)	[65]	only allele frequencies available
GONL (Dutch)	[66]	gonl-100c, gonl-101c, gonl-102c, gonl-103c, gonl-104c, gonl-105c, gonl-106c, gonl-107c, gonl-108c, gonl-109c, gonl-10c, gonl-110c, gonl-111c, gonl-112c, gonl-113c, gonl-114c, gonl-115c, gonl-116c, gonl-117c, gonl-118c, gonl-119c, gonl-11c, gonl-120c, gonl-121c, gonl-122c, gonl-123c, gonl-124c, gonl-125c, gonl-126c, gonl-127c, gonl-128c, gonl-129c, gonl-12c, gonl-130c, gonl-131c, gonl-132c, gonl-133c, gonl-134c, gonl-135c, gonl-136c, gonl-137c, gonl-138c, gonl-139c, gonl-13c, gonl-140c, gonl-141c, gonl-142c, gonl-143c, gonl-144c, gonl-145c, gonl-146c, gonl-147c, gonl-148c, gonl-149c, gonl-14c, gonl-150c, gonl-151c, gonl-152c, gonl-153c, gonl-154c, gonl-155c, gonl-156c, gonl-157c, gonl-158c, gonl-159c, gonl-15c, gonl-160c, gonl-161c, gonl-162c, gonl-163c, gonl-164c, gonl-165c, gonl-166c, gonl-167c, gonl-168c, gonl-169c, gonl-16c, gonl-170c, gonl-171c, gonl-172c, gonl-174c, gonl-175c, gonl-176c, gonl-177c, gonl-178c, gonl-179c, gonl-17c, gonl-180c, gonl-181c, gonl-182c, gonl-183c, gonl-184c, gonl-185c, gonl-186c, gonl-187c, gonl-188c, gonl-189c, gonl-18c, gonl-190c, gonl-191c, gonl-192c, gonl-193c, gonl-194c, gonl-195c, gonl-196c, gonl-197c, gonl-198c, gonl-199c, gonl-19c, gonl-1c, gonl-200c, gonl-201c, gonl-202c, gonl-203c, gonl-204c, gonl-205c, gonl-206c, gonl-207c, gonl-208c, gonl-209c, gonl-20c, gonl-210c, gonl-212c, gonl-213c, gonl-214c, gonl-215c, gonl-216c, gonl-217c, gonl-218c, gonl-219c, gonl-21c, gonl-220c, gonl-221c, gonl-222c, gonl-223c, gonl-224c, gonl-225c, gonl-226c, gonl-227c, gonl-228c, gonl-229c, gonl-22c, gonl-230c, gonl-231c, gonl-232c, gonl-233c, gonl-234c, gonl-235c, gonl-236c, gonl-237c, gonl-238c, gonl-239c, gonl-23c, gonl-240c, gonl-241c, gonl-242c, gonl-243c, gonl-244c, gonl-245c, gonl-246c, gonl-247c, gonl-248c, gonl-249c, gonl-24c,

(*Continued on next page*)

Please cite this article in press as: Burger et al., Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.08.033

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
		gonl-250c, gonl-25c, gonl-26c, gonl-27c, gonl-28c, gonl-29c, gonl-2c, gonl-30c, gonl-31c, gonl-32c, gonl-33c, gonl-34c, gonl-35c, gonl-36c, gonl-37c, gonl-38c, gonl-39c, gonl-3c, gonl-40c, gonl-41c, gonl-42c, gonl-43c, gonl-44c, gonl-45c, gonl-46c, gonl-47c, gonl-48c, gonl-49c, gonl-4c, gonl-50c, gonl-51c, gonl-52c, gonl-53c, gonl-54c, gonl-55c, gonl-56c, gonl-57c, gonl-58c, gonl-59c, gonl-5c, gonl-60c, gonl-61c, gonl-62c, gonl-63c, gonl-64c, gonl-65c, gonl-66c, gonl-67c, gonl-68c, gonl-69c, gonl-6c, gonl-70c, gonl-71c, gonl-72c, gonl-73c, gonl-74c, gonl-75c, gonl-76c, gonl-77c, gonl-78c, gonl-79c, gonl-7c, gonl-80c, gonl-81c, gonl-82c, gonl-83c, gonl-84c, gonl-85c, gonl-86c, gonl-87c, gonl-88c, gonl-89c, gonl-8c, gonl-90c, gonl-91c, gonl-92c, gonl-93c, gonl-94c, gonl-95c, gonl-96c, gonl-97c, gonl-98c, gonl-99c, gonl-9c
Lech Valley Bavarians	[13]	AITI_119, AITI_2, AITI_40, AITI_43, AITI_50, AITI_72, AITI_78, AITI_98, AITI_65adult, AITI_66, AITI_92, AITI_95, OBKR_80, OBKR_117, OBKR_2, OBKR_50, OBKR_67, OBKR_9A, POST_28, POST_44, POST_50, POST_6, POST_35, POST_99, UNTA58_153, UNTA85_1412, WEHR_1192SkB, WEHR_1415adult, WEHR_1586
Mokrin	[19]	MOK10B, MOK12, MOK13, MOK14, MOK15, MOK16A, MOK17A, MOK18A, MOK19A, MOK20, MOK21A, MOK22, MOK23, MOK24A, MOK25A, MOK26A, MOK27, MOK28A, MOK29A, MOK30, MOK31, MOK32, MOK33, MOK9B
Ancient Greek individuals used in the D-statistics analysis	[67]	I0070 I0071 I0073 I0074 I9005 I9006 I9010 I9033 I9041 I9123 I9127 I9128 I9129 I9130 I9131
Ancient European individuals used in the D-statistics analysis	[2]	I0047 I0059 I0060 I0099 I0112 I0113 I0115 I0116 I0117 I0164 I0171 I0803 I0804 I0805 I0806 I1271 I1276 I1280 I1282 I1284 I1300 I1303 I1314 I1502 I1504 I1532 I1534 I1536 I1538 I1540 I1542 I1544 I1546 I1549 RISE109 RISE150 RISE154 RISE175 RISE179 RISE210 RISE247 RISE254 RISE276 RISE349 RISE371 RISE373 RISE374 RISE431 RISE47 RISE486 RISE71 RISE97 RISE98
Ancient Baltic Sea Region individuals used in the D-statistics analysis	[22]	Kivutkalns153 Kivutkalns164 Kivutkalns19 Kivutkalns194 Kivutkalns207 Kivutkalns209 Kivutkalns215 Kivutkalns222 Kivutkalns25 Kivutkalns42 Turlojiske1 Turlojiske1932 Turlojiske3 Turlojiske5
Ancient European Europe individuals used in the D-statistics analysis	[3]	E09537_d E09538 E09568_d E09569 E09613_d E09614_d I0258 I0260 I0261 I0262 I0263 I0453 I0459 I0460 I0461 I0823 I0825 I0839 I0840 I1381 I1388 I1390 I1391 I1553 I1765_d I1767 I1770 I1775 I1970 I1979 I2364 I2365 I2416 I2418 I2421 I2445 I2446 I2447 I2450 I2452 I2453 I2454 I2455 I2457 I2458 I2459 I2461 I2462 I2463 I2464 I2477 I2478 I2565 I2567 I2568 I2569 I2573 I2574 I2575 I2597 I2598 I2601 I2602 I2604 I2609 I2610 I2612 I2618 I2639 I2653 I2654 I2655 I2656_d I2741 I2786 I2787 I2859 I2860

(*Continued on next page*)

Please cite this article in press as: Burger et al., Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.08.033

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
		I2861 I2981 I3082 I3130 I3132 I3255 I3529 I3587 I3588 I3589 I3590 I3592 I3594 I3596 I3600 I3601 I3602 I3604 I3607 I3874 I3875 I4067 I4069 I4070 I4071 I4073 I4075 I4076 I4178 I4229 I4245 I4247 I4249 I4250 I4251 I4252 I4253 I4884 I4886 I4888 I4889 I4890 I4891 I4892 I4895 I4896 I4930 I4933 I4936 I4945 I4946 I4947 I4950 I5014 I5015 I5017 I5019 I5020 I5022 I5023 I5024 I5025 I5035 I5037 I5042 I5043 I5044 I5364 I5367 I5373 I5376 I5377 I5379 I5382 I5383 I5385 I5441 I5469 I5470 I5471 I5473 I5512 I5513 I5514 I5515 I5516 I5519 I5521 I5523 I5524 I5525 I5527 I5529 I5530 I5531 I5655 I5656 I5658 I5659 I5661 I5663 I5665 I5666 I5748 I5750 I5755 I5757 I5759 I5833 I5834 I5836 I6466 I6467 I6468 I6471 I6472 I6475 I6476 I6480 I6481 I6482 I6531 I6534 I6537 I6538 I6539 I6542 I6543 I6579 I6580 I6582 I6583 I6584 I6585 I6587 I6589 I6590 I6591 I6596 I6604 I6605 I6608 I6609 I6610 I6612 I6613 I6617 I6622 I6623 I6624 I6626 I6628 I6629 I6679 I6680 I6774 I6775 I6778 I7040 I7042 I7043 I7044 I7045 I7195 I7196 I7197 I7198 I7199 I7200 I7201 I7202 I7203 I7205 I7210 I7211 I7213 I7249 I7250 I7251 I7269 I7270 I7272 I7275 I7276 I7278 I7279 I7280 I7281 I7282 I7286 I7287 I7288 I7289 I7290 I7568 I7569 I7570 I7571 I7572 I7573 I7574 I7575 I7576 I7577 I7578 I7580 I7626 I7627 I7628 I7630 I7635 I7638 I7639 I7640
East European Steppe: 3,600-2,300 BCE	[2–4, 7–12]	RISE547.SG, RISE552.SG, I0374, I1917, I2105, I4110, I5118, I5119, ILK001, ILK002, ILK003, I5882, I8745, I11501, I11531, I11732, I11735, I11736, I11734, RISE1166.SG, AY2001, AY2003, RK1001, RK1003, RK4001, RK4002, SA6001, SA6002, SA6003, SA6004, SA6010, SA6013, SIJ002, SIJ003, ZO2002, I2791_published, I5884_publishedRISE547.SG, RISE552.SG, I0374, I1917, I2105, I4110, I5118, I5119, ILK001, ILK002, ILK003, I5882, I8745, I11501, I11531, I11732, I11735, I11736, I11734, RISE1166.SG, AY2001, AY2003, RK1001, RK1003, RK4001, RK4002, SA6001, SA6002, SA6003, SA6004, SA6010, SA6013, SIJ002, SIJ003, ZO2002, I2791_published, I5884_published
Area of Corded Ware Culture: 2,900-2,300 BCE	[2–4, 9–11, 13–18]	I7207, I7208, I7209, I7278, I7289, I7212, I6695, I6696, I7040, I7044, I7045, I7195, I7196, I7200, I7201, I7202, I7203, I7213, I7286, I0805, Ajvide52.SG, Ajvide58.SG, RISE1.SG, RISE94.SG, I3529, I2365, I4178, I2741, I4629, I5015, I5043, I5520, I5521, I5523, I5525, I5529, I6534, I6581, I2786, N44.SG, N45.SG, N47.SG, N49.SG, RISE1159.SG, RISE1162.SG, RISE1163.SG, RISE1164.SG, RISE1167.SG, RISE1170.SG, RISE1171.SG, RISE1172.SG, RISE1173.SG, RISE1248.SG, RISE1249.SG, RISE1250.SG, ans016.SG
Ancient individuals used in D-Statistics	[2, 3, 9, 22, 67]	Kivutkalns209, I5769, Turlojiske5, Kivutkalns164, Kivutkalns19, Kivutkalns207, Kivutkalns215, I5364, Kivutkalns153, Kivutkalns194, Kivutkalns25, I3313, Kivutkalns222, RISE276, Kivutkalns42, I2859, Turlojiske1, I2860, I2861, I3130, I2656_d, Turlojiske3, I7580, I7575, I5383, I7577, I0099, I7626, RISE175, I1504, I9010, I9041, I7576, I7628, I7627, I7578, I2574, I7574, Turlojiske1932, I7571, I9006, I2655, I9033, I7640, I2458, I2653, I2654, I7568, I2573, I7572, RISE210, RISE47, I9123, I3082, I2639, I5470, I7569, I2981,

(*Continued on next page*)

Please cite this article in press as: Burger et al., Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.08.033

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
		I5469, I7573, I4332, I4331, I7570, I1775, Bul6, Bul8, I1770, I2464, I5516, I2163, I6680, RISE247, RISE374, I4070, I4071, I7639, I4067, I7630, I5377, I2602, I7195, I7196, I7197, I7198, I7199, I7200, I7201, I7202, I7203, I2463, I0070, I0071, I0073, I0074, I9005, I4892, RISE150, RISE373, I6623, I5441, I4884, I2610, I6610, I2601, I6542, I6622, I6626, I4076, I2421, I6538, I6608, I2609, I5515, I7638, RISE154, RISE109, RISE486, RISE179, I0115, RISE349, I4950, I5512, I5513, I2567, I3132, I4069, I6679, I2462, I5025, I5666, I6476, I5037, I5042, I5043, I5044, I5035, I6468, I6480, I1381, I6778, I9127, I9128, I9129, I9130, I9131, I4930, I4933, I4936, RISE97, I5750, I6604, I4073, I2447, I2455, I2569, RISE254, I2445, I1979, I2477, I0164, I4075, I2478, I6543, I6612, I6613, I3875, I3255, I2618, I0047, I2598, RISE371, I0116, I7635, I1767, I5373, I2452, I0803, I5382, I2454, I5665, I5755, I5757, I5759, I0804, I4947, I5514, I4890, I5519, I5521, I4249, I5655, I5656, I5658, I5659, I5661, I5663, I5833, I5834, I6590, I6591, I6624, I5019, I5523, I5524, I5525, I5014, I5527, I5022, I5023, I5529, I5530, I5531, I6481, I6482, I5836, I6775, I6539, I6584, I6587, I6589, I6471, I6472, I6475, I5471, I4888, I2457, I2604, I5024, I2597, I3874, RISE71, I2568, I6531, I2453, I1765_d, I6537, I4886, I6774, E09614_d, I6579, I4895, RISE98, I6583, I6582, I1391, I4889, I4891, I2364, RISE431, I4229, I1546, I1549, I1532, I1534, I1536, I1538, I1540, I1542, I1544, I0117, I1502, I5473, I1388, I6585, I4252, I2446, I3590, E09613_d, I2461, I5376, I4896, I6609, I4945, I0113, I4946, I2450, I5385, I2565, I0453, E09569, I4250, I6534, I3587, I3588, I3589, I3594, I3596, I3600, I3601, I3602, I2459, I6580, I4251, I0171, I2416, I2741, I0059, I7249, I7250, I7251, I7269, I7270, I7272, I7275, I7276, I7278, I7279, I7280, I7281, I7205, I7210, I7211, I7213, I7282, I7286, I7287, I7288, I7289, I7290, I2418, I3529, I7040, I7042, I7043, I7044, I7045, I4178, I0840, I0459, I2787, I0461, I3592, I5020, I0805, I0112, I2786, I0839, I5017, I2365, I4253, I2612, I0460, I2575, I1390, E09568_d, E09538, I5015, I4247, I1271, I1276, I1280, I1282, I1284, I1300, I1303, I1314, I0806, I0060, I5367, I5748, I3604, I3607, I0258, I0260, I0261, I0262, I0263, I0823, I1553, I6605, I4245, I5379, E09537_d, I0825, I1970, I6466, I6467, I6596, I6617, I6628, I6629
Oligonucleotides		
MYBait kit	Arbor biosciences; https://arborbiosci.com/genomics/targeted-sequencing/mybaits/mybaits-custom/	N/A
P5 and P7	[68] IDT, Leuven, Belgium	N/A
IS4, IS5, IS6 and IS7	[68] IDT, Leuven, Belgium	N/A
Software and Algorithms		
ATLAS	[69]	N/A
TreeMix	[70]	N/A

(*Continued on next page*)

Please cite this article in press as: Burger et al., Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.08.033

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
plink2	[71]	N/A
approxWF	[29]	N/A
laser2	[72]	N/A
ADMIXTOOLS	[73]	N/A
yHaplo*	[74]	N/A
scipy.stats python library	https://docs.scipy.org/doc	N/A
numpy python library	https://numpy.org/	N/A
bwa aln	[75]	N/A
GATK	[76]	N/A
Samtools	[77]	N/A
VCFtools	[78]	N/A
picard-tools	https://broadinstitute.github.io/picard/	N/A
contamMix	[79]	N/A
Other		
Agilent 2100 Expert Bioanalyzer System and High Sensitivity DNA Analysis Kit	Agilent Technologies	N/A
Amicon Ultra-15 Centrifugal Filter Units	Merck Millipore, Darmstadt, Germany	N/A
AmpliTaq Gold ® Buffer II (10x)	Life Technologies ™	N/A
AmpliTaq Gold ® DNA Polymerase	Life Technologies ™	N/A
ATP Solution (100 mM)	Life Technologies ™	N/A
Bovine Serum Albumin (BSA) (20 mg/ml)	Roche Diagnostics	N/A
Bst Polymerase, Large Fragment (8 U/μl)	New England Biolabs GmbH	N/A
dNTPs (each 10 mM)	QIAGEN, Hilden, Germany	N/A
dNTPs (each 25 mM)	Agilent Technologies	N/A
EDTA (0.5 M), pH 8.0	Ambion/Applied Biosystems, Life Technologies ™, Darmstadt, Germany	N/A
Herculase II Fusion ® DNA Polymerase	Agilent Technologies	N/A
Herculase II Reaction Buffer	Agilent Technologies	N/A
$MgCl_2$ (25 mM)	Life Technologies ™	N/A
MinElute ® PCR Purification Kit	QIAGEN, Hilden, Germany	N/A
MSB ® Spin PCRapace	Invitek, Stratec Molecular, Berlin, Germany	N/A
Sodium N-lauryl sarcosinate	Merck Millipore, Darmstadt, Germany	N/A
Nuclease-free H_2O	Life Technologies ™	N/A
PEG-4000	Thermo Scientific ™	N/A
Proteinase K	Roche Diagnostics, Mannheim, Germany	N/A
Phenol/chloroform/isoamylalcohol (25:24:1)	Roth, Karlsruhe, Germany	N/A
Qubit ® Fluorometric quantitation and dsDNA HS Assay Kit	Invitrogen ™	N/A
T4 DNA Ligase (5 U/μl)	Thermo Scientific ™	N/A
T4 DNA Ligase Buffer (10X)	Thermo Scientific ™	N/A
T4 DNA Polymerase (5 U/μl)	Thermo Scientific ™	N/A
T4 Polynucleotide Kinase	Invitrogen ™	N/A
Tango Buffer (10x)	Life Technologies ™	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ThermoPol Buffer (10X)	New England Biolabs GmbH	N/A
Trichlormethan/Chloroform	Roth, Karlsruhe, Germany	N/A

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contacts, Joachim Burger (jburger@uni-mainz.de) and Daniel Wegmann (daniel.wegmann@unifr.ch).

Materials Availability

Genomic data are available at the European Nucleotide Archive under the accession number PRJEB38406 in BAM and FASTQ format.

Data and Code Availability

All genomic data used in this study is publicly available at the European Nucleotide Archive under the accession number PRJEB38406 or at the sources listed in the Key Resources Table. The code used to plot Figure S1 is available on Bitbucket at https://bitbucket.org/wegmannlab/atlas/downloads/Relatedness.R, the code used to estimate allelic age is available at https://bitbucket.org/wegmannlab/atlas/downloads/Allele_Age_MC_ABC_Sampling_V2.R.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Archaeological and morphological information

Sometime after 1,300 BC, the Tollense Valley saw an armed conflict of a scale hitherto unimagined for the European Bronze Age. Likely more than 1,000, perhaps even more than 2,000 combatants fought each other in the river valley. Long-distance as well as close-range weapons (swords, spears/lances, bow and arrow, wooden clubs and hammers) were used, and the number of casualties, attested to by a large number of perimortem lesions, was high [80]. The bones show perimortem as well as healed lesions, documenting the actual fighting in the valley, but in some cases also violent encounters in previous life. Cases include arrowheads still embedded in bone, small cuts and funnel-shaped lesions indicating arrow shots and stab wounds, as well as large impression fractures, caused by blunt objects [81–83]. These lesions prove hand to hand combat as well as long distance arrow shots. According to morphological criteria the skeletons are very heterogeneous and consist mainly of young adult males. According to the number of individuals and supported by Sr-isotope analyses the skeletal samples include local and non-local warriors [28].

The Tollense Valley (Mecklenburg-Western Pomerania) is characterized by a low moorland environment. Neighboring ground moraine areas have been favored settlement areas since the Neolithic. In 2007, systematic archaeological research in the valley was initiated, including excavations, underwater surveys, and metal detector surveys, as well as scientific analyses on bones and other finds. Human skeletal remains are discovered from a stretch of river of 2.5 km length. Based on a series of more than 100 radiocarbon dates on skeletal and wooden remains from the Tollense Valley, the finds horizon of the Bronze Age battle can now be dated to the first half of the 13th century BC. With one exception (WEZ16 was directly C^{14}-dated to the Neolithic period) all samples are associated with the Bronze Age (c. 1,300- 1,200 calBC). The event is interpreted as one major conflict, not a series of chronologically divergent skirmishes, due to dating results and the general appearance of the finds layer with comparable disarticulation as well as lesion patterns of the human skeletal remains at all analyzed sites [84]. From 2009 to 2015, extensive excavations funded by the German Research Foundation (DFG) were conducted at the main site Weltzin 20 (exposing an area of c. 462 m^2), less extensive excavations were conducted at Weltzin 32; in both cases supported by comprehensive underwater surveys. Archaeological research on the Bronze Age site in the Tollense Valley has uncovered remains of a minimum number of more than 140 individuals thus far, predominantly of young adult men. These remains represent a special Bronze Age skeletal sample due to the large number of individuals and their mortality profile.

The sampled burials from the Eneolithic and Bronze Age period in the Steppe region date from the second half of the 4th millennium and the 3rd millennium B.C. They are associated with the cultures of the Eneolithic, Yamnaya, and Catacomb grave. Yamnaya Culture graves from the 1st half of 3rd mill. BCE were sampled in a region between the Urals and the Balkan (Figure 3; Table S3). Skeletons associated with the Catacomb Culture were sampled from the area between the rivers Wolga and Prut. While the early phase of the Catacomb Grave Culture is still synchronous to late Yamnaya, the developed and late phases of the Catacomb Culture date to 2,500-2,000 BCE,. There is still an ongoing debate on the cultural interpretation of burials in the Middle Wolga region which are often described as Poltavkinskaya Culture to emphasize characteristic differences regarding funerary customs [85].

All sampled burials of the East European Steppe date to the 3rd mill. BCE except three earlier graves attributed to the Late Eneolithic, including the Usatovo Culture. Only one burial (sample No MOB 1) was a flat grave, all others were excavated in burial mounds (Table S3).

Please cite this article in press as: Burger et al., Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.08.033

METHOD DETAILS

Sample preparation

Sample preparation and paleogenetic analyses were conducted at the Institute of Organismic and Molecular Evolution (iomE) at Johannes Gutenberg University, Mainz. Strategies to avoid and detect contaminations were applied as described elsewhere [86, 87] and included sample analysis in a dedicated ancient DNA facility separated from post-PCR laboratories, decontamination of samples, workspace and labware, independent DNA extraction as well as co-analysis of blank controls.

Tooth material or pieces of right femora were used for ancient DNA analyses to avoid double typing. For the isolation of DNA molecules 0.2-0.5 g of bone powder was incubated for 48h in 7 mL lysis buffer containing 0.48 M EDTA (pH 8.0), 0.2% sodium N-lauryl sarcosinate, and 0.1 mg/ml Proteinase K at 37°C followed by phenol-chloroform extraction. Samples were washed with nuclease-free water and concentrated on 50-kDA Amicon Ultra-15 Centrifugal Filter Units (Millipore) to a final volume of 150-350 μl.

Library construction and amplification

Libraries were constructed as detailed in Meyer and Kircher [68] with minor changes. After blunt end repair, samples were purified with the MinElute PCR purification kit (QIAGEN) following the manufacturer's instructions. *Bst* polymerase was heat inactivated for 20 minutes at 80°C [15]. Indexing PCR was performed with 1x Ampli Taq Gold ® Buffer II, 2.5 mM $MgCl_2$, 0.4 mg/ml BSA, 200 μM dNTPs, 0.05 U/μl AmpliTaq Gold® Polymerase, 0.2 μM index primer and either 0.2 μM IS4 or IS7, and 8μl of DNA library in a final volume of 50 μl. Libraries were amplified in five parallels and 10 to 15 PCR cycles followed by a second amplification in a final volume of 100 μl with 1x Herculase II Reaction Buffer (Agilent Technologies), 250 μM dNTPs (Agilent Technologies), 0.3 μM Primer, 1 μl of Herculase II Fusion® DNA Polymerase (Agilent Technologies) and 10 μl of DNA library. Thermal conditions were 2 minutes at 95°C followed by eight cycles at 95°C, 60°C, and 72°C for 30 s each, and a final elongation at 72°C for 5 minutes. The DNA libraries were quantified using Qubit ® Fluorometric quantitation with the dsDNA HS Assay kit and their quality was assessed on the Agilent 2100 Expert Bioanalyzer using the High Sensitivity DNA Analysis kit according to the manufacturer's instructions.

Illumina sequencing

Prior to capture enrichment, DNA libraries were screened for endogenous DNA content on an Illumina MiSeq™ platform at StarSEQ GmbH (Mainz, Germany). Libraries were pooled equimolar and sequenced in a 50 bp single-end run. Endogenous DNA content varied between 0.4 and 48%. Libraries enriched for nuclear loci were pooled to 200 ng and sequenced on an Illumina HiSeq™ platform at the sequencing facilities of the Johannes Gutenberg University (Institute of Molecular Genetics, Mainz, Germany).

5 MB capture assay

The capture followed the protocol described in Veeramah et al. [6]. It was designed to cover 4,905,256 nucleotides (0.16% of the whole genome), including 4,687 independent neutral 1-kb regions in addition to 429 neutral regions of 500 bp length, as well as 486 phenotypic informative markers correlating with pigmentation, metabolic syndrome, diabetes risk, lactase persistence, and markers associated with non-infectious and inflammatory diseases.

Read mapping, alignment cleaning and merging

Residual adaptor sequences were trimmed from the raw reads with a custom made python script. Only reads with a minimum of 95% of all bases with quality score > 15 were kept. Reads originating from paired-end sequencing were joined using fastq-join from ea-utils [88] with default parameters. All reads were aligned with bwa aln [75] to hg19. During SAM/BAM conversion, alignments were filtered and only reads with a minimum mapping quality of 25 were kept. PCR duplicates were removed from the files using MarkDuplicates from picard-tools (https://broadinstitute.github.io/picard/), prior to local re-alignment with GATK [76].

QUANTIFICATION AND STATISTICAL ANALYSIS

Sex Determination

We determined the molecular sex of each sample following Skoglund et al. [89] by calculating the number of reads aligned to the Y chromosome as a fraction of the total number of reads aligned to both sex chromosomes.

Ancient DNA Authenticity

To estimate the authenticity of the ancient sample DNA, we used contamMix, as described in Fu et al. [79]. Therefore, all reads mapping to the mitochondrial genome were extracted from the alignments and realigned against the rCRS (NC_012920.1).

Haplogroups of uniparental Markers

mt-DNA

In order to determine haplogroups for uniparental markers, majority allele calls were created with ATLAS ([69], commit 401af70, task = majorityBase). The resulting VCF files for the mitochondrial genomes were merged and uploaded to the Haplogrep 2.0 web page (https://haplogrep.uibk.ac.at/). See Table S1 for details.

Please cite this article in press as: Burger et al., Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.08.033

Y chromosome haplogroups

Y chromosome haplogroups were determined based on majority allele calls, created by ATLAS ([69], commit 401af70, task = majorityBase). The resulting VCF files were converted to the text based plink format [90], which is the base for the input files necessary for the yhaplo script [74]. Prior to the conversion, the VCF files were filtered in two steps, with a minimum depth of two as well as five alleles per position. Haplogroup determination was therefore done three times, with all available SNPs and for each of the two filter steps for each sample. See Table S1 for details.

Genotype likelihoods in Tollense samples

Post mortem damage patterns

We used ATLAS ([69], commit d5b0aee, task = PMD) to infer empirical PMD patterns as described in Kousathanas et al. [91] from the tabulated mismatches between the raw reads and the reference genome. These patterns were inferred independently for all individuals and read groups.

Base quality score recalibration

We used ATLAS ([69], commit d5b0aee, task = recal) to recalibrate base quality scores with the method described in Kousathanas et al. [91]. This approach exploits known homozygous sites. As known homozygous sites, we used the 10 million sites with highest GERP conservation scores as calculated across the multiple sequence alignments of 88 mammals and provided by Ensembl (http://ftp.ensembl.org/pub/release-96/bed/ensembl-compara/88_mammals.gerp_constrained_element/). For recalibration, we pooled all samples to increase power and estimated common recalibration parameters for all samples generated with the same sequencing run (with option poolReadGroups), but used individual damage patterns.

Genotype likelihoods

We used ATLAS ([69], commit 52a2276, task = GLF) to obtain genotype likelihoods for all individuals at all sites. We included the damage patterns and the recalibration parameters. We then created a set of GLF files for the phenotypic SNP's and a set of GLF files for the captured regions, restricting the analysis to all Bronze Age Tollense samples with low contamination levels and depth > 4x (denoted as Tollense > 4x in the following). For the male samples we further specified chromosomes X,Y and MT to be haploid and for the female samples chromosomes Y and MT.

Pseudo-haploid calls

Post mortem damage patterns

We used ATLAS ([69], commit 401af70, task = PMD) to infer empirical PMD patterns as described in Kousathanas et al. [91] from the tabulated mismatches between the raw reads and the reference genome. These patterns were inferred independently for all individuals and all read groups.

Base quality score recalibration

We used ATLAS ([69], commit 401af70, task = recal) to recalibrate base quality scores with the method described in Kousathanas et al. [91]. This approach exploits known homozygous sites and we used the UCNEbase regions [92]. For recalibration, we pooled all samples to increase power and estimated common recalibration parameters for all samples generated with the same sequencing run (with option poolReadGroups), but used individual damage patterns.

Genotype calling

We generated two sets of calls for all Tollense > 4x to be used in downstream analyses, both based on genotype likelihoods calculated with ATLAS ([69], commit 401af70, task = call, method = MLE): 1) Diploid Maximum Likelihood calls and 2) pseudo-haploid calls by identifying the allele with the highest support as given by the genotype likelihoods.

Kinship analysis

We used ATLAS ([69], commit d5b0aee, task = geneticDist) to estimate the squared genetic distances between all pairs of Tollense > 4x samples, using the GLF files created in step "Estimate genotype likelihoods." We then used a custom R script to apply the method of Waples et al. [93] to transform these distances into estimates of genetic relatedness. This script can be found on the ATLAS wiki (https://bitbucket.org/wegmannlab/atlas/downloads/Relatedness.R). For more details see Figure S1.

Selection inference

Allele count estimation

Under the assumption of population continuity, we identified loci under selection by comparing allele frequencies in the Tollense Valley population to those of a modern reference population using two complementary methods (see below). In absence of direct data from Northern or Central Germany, we used the CEU population from the 1000Genomes project [63] as a modern proxy.

We downloaded the BAM files for 96 unrelated CEU genomes from EBI (http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000_genomes_project/). These BAM files were based on the GRCh38 human reference while the Tollense samples were based on GRCh37. Thus, we used the UCSC Genome Browser (https://genome.ucsc.edu/) liftover tool to translate the coordinates of our capture loci from GRCh37 to GRCh38. We used GRCh38 coordinates to extract the reads that mapped to our regions of interest from the CEU BAM files, and mapped these reads to the GRCh37 reference using BWA mem [94]. Since the downloaded BAM files already had recalibrated quality scores, we did not recalibrate the BAM files with the ATLAS pipeline but directly created GLF files, as specified above for the Tollense samples.

Please cite this article in press as: Burger et al., Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.08.033

In a first step, we used ATLAS ([69], commit 52a2276, task=majorMinor) to identify the major and minor allele and to create a VCF containing genotype likelihoods of all Tollense > 4x and CEU individuals. Then, we used ATLAS ([69], commit 52a2276, task=alleleCounts) to estimate allele counts in the Tollense > 4x and CEU populations separately.

To ensure positive selection coefficients, we estimated selection coefficients for the allele with a higher frequency in CEU than Tollense > 4x.

Selection inference with ApproxWF

Population size estimation

We used the estimated allele counts at the putatively neutral regions to estimate the population size and sequencing error rate with ApproxWF ([29], commit a19e439) under a constant size model and assuming no selection. We assumed that 130 generations have passed between the Tollense battle and the modern CEU reference population, implying a generation time of roughly 25 years. We ran the MCMC for 100,000 iterations and specified that the error rate should be estimated. This resulted in posterior medians for the population size of 2,503 and the error rate of 0.02.

Estimating selection coefficients

We used ApproxWF to estimate selection coefficients for all loci while fixing the population size and sequencing error rate to the values inferred from neutral loci (see above). We ran the estimation assuming a dominance coefficient h = 0.5 (additive) for all loci. We then reran the estimation also with h = 1 (dominant) for the loci associated with lactase persistence.

Selection inference with forward simulations

To test if allele frequency changes over time are the result of drift, or if selection has to be invoked, a forward simulation approach assuming exponential population growth was used, as described in Wilde et al. [30]. For each simulation run, the initial allele frequencies were drawn at random from a beta (n_p + 1, n_q + 1) distribution, where n_p and n_q denote the estimated allele counts in the Tollense > 4x population. Allele frequencies in each following generation were simulated by binomial sampling from the previous generation, while the estimations for the final generation were saved for each simulation run.

Since the size of the population from which the Tollense > 4x sample originates is unknown, we ran the analysis for 50 initial population sizes 10^2 and 10^5, evenly spaced on the log scale. Population sizes were then assumed to grow exponentially over 130 discrete generations. The final population size was set to $8 \cdot 10^6$, which corresponds to approximately one tenth of the Census population of Germany (~83 million, 31.12.2018, https://de.wikipedia.org/wiki/Deutschland).

Under this demographic model, we generated 10^4 allele frequency trajectories for each combination of an initial population size and 51 evenly spaced selection coefficients between −10 and 10. Finally, we calculated a Z-score to determine if the frequency in the modern reference population was outside the distribution of allele frequencies obtained with these simulations. The Z-score was then translated into a two-tailed p value.

Population StructurePrincipal component analysis

We generated PCA's using LASER (version 2.04 [72];). In a first step, a reference space was generated, with a principal components analysis on the genotype data of modern individuals, with imputation of missing entries by averaging over all individual genotypes at that position (data published as part of reference [95]; Southern European [Italian North/South, Spanish /North, Canary Islander, Maltese, Greek], Sicilian, Basque, Sardinian, Cypriot, Central- and Eastern European [Albanian, Bulgarian, Romanian, Hungarian, Czech, German, French], Croatian, Slavic [Russian, Ukrainian, Belarusian, Polish, Sorb, Mordovian], Baltic and Finnish [Estonian, Lithuanian, Finnish], British Isles [English, Orcadian, Scottish, Irish /Ulster, Shetlander], Icelandic, Norwegian). In the following step, samples with depth > 4x were projected onto the reference space based on ten replicates, using pileup files generated with Samtools (version 1.9 [77],) with the suggested filter criteria of minimum mapping quality 30 and minimum base quality 20.

F_{ST} and TreeMix

Reference samples

We extracted all SNP's from whole genome sequences generated from European (CEU, FIN, GBR, IBS and TSI, together named EUR), SouthEast Asian (GIH, PJL, BEB, STU, and ITU, together named SAS), East Asian (CHB, JPT, CHS, CDX and KHV, together named EAS) and Yoruban (YRI) populations that are part of the 1000 Genomes Project. 1000 Genomes SNPs were obtained using VCFtools [78]. We complemented this dataset with the Genome DK (Danish), GONL (Dutch) and Turkish modern samples [64], as well as 5th century Bavarians [6]. Datasets were merged using PLINK [71] and custom Python scripts.

We restricted our analysis to transversion SNPs within the 5K neutralome and to those whose frequencies are likely to vary due to drift (rather than novel mutations) by ascertaining variable sites in the Yoruba population (n = 31,494 SNPs). We used pseudo-haploid calls (see above) for the Tollense > 4x sample, but diploid calls for the reference populations.

F_{ST} analysis

We estimated genome-wide SNP based pairwise F_{ST} between the Tollense > 4x and reference samples as described in Bhatia et al. [96]. Standard errors and 95% confidence intervals were estimated using the delete-m jackknife method [97], with all m-SNPs within each of the 1kb locus successively removed (see also Figure S2).

Please cite this article in press as: Burger et al., Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.08.033

Treemix analysis

We inferred a phylogenetic tree relating all populations using TreeMix 1.12 [70]. Though we recognize that there is unlikely to be a strict bifurcating structure to the population history of this data, this analysis can help extract useful relationships among these populations. To ensure there were no biases due to sample size differences among populations, all allele frequencies were resampled down to 14 haploid alleles to match the Tollense > 4x dataset. We assumed no migration and grouped SNPs into blocks of 100 to control for linkage disequilibrium (option -k 100) (see also Figure S2).

D-statistics with with ancient individuals

In order to examine differentiation among Tollense > 4x individuals in their relationship to other populations, we performed D-statistic analyses with ADMIXTOOLS 5.1 [73] in the form of D(Tollense1, Tollense2; Test_Population, YRI) for each Tollense pair and a reference population. The reference populations were individuals from our ancient reference set used for F_{ST} and TreeMix with additional ancient reference samples [2, 3, 9, 22, 67], with ages within 1,000 years of the Tollense samples grouped based on modern country borders (see also Table S2). To be consistent with how the ancient reference samples were called, pseudo-haploid genotypes were called for each Tollense individual by randomly sampling a read at each of the 1.2 million target SNPs. We note that this analysis relies on off-target reads in each Tollense sample, and as a consequence the SNP numbers involved in each test are low (mean 11,078.2 SNPs). We used a custom python script that allowed multithreading to perform all tests. In addition, the same reference populations were used to test the partition of our samples suggested by the PCA into upper (WEZ24, WEZ40, WEZ51, WEZ56, WEZ58, WEZ71, WEZ83) and lower (WEZ35, WEZ39, WEZ54, WEZ57, WEZ59, WEZ61) clusters.

Population structure within the Tollense > 4x sample

Inbreeding coefficient F

We used ATLAS ([69], commit 52a2276, task = majorMinor) to identify the major and minor allele and to create a VCF containing genotype likelihoods of all Tollense > 4x from the GLF files (see above). We then used ATLAS ([69], commit 52a2276, task = inbreeding) to quantify population structure among the Tollense > 4x sample by estimating the inbreeding coefficient F, running the MCMC for 10^6 iterations with the parameters minVariantQuality = 40, numBurnins = 10 burninLength = 500. We reran this analysis also with argument probMovingToModelNoF = 0 to restrict the analysis to a model with F > 0.

F_{st} permutation test

We created a multisample v*cf.* file from the MLE-calls for the Samples > 4x to calculate F_{st}'s for every possible combination of samples in group sizes from four to seven, toward the remaining samples using Hudson's estimator [96]. Therefore a set of SNPs was selected for which a minimal depth of two reads, a minimal genotype quality of 15, a minor allele frequency of ten or higher, as well as a call rate of 100% was found (1947 SNPs). For comparison the same was done for 14 samples drawn at random from CEU from the 1000 Genomes project (NA06986, NA07357, NA10847, NA11840, NA11881, NA11994, NA11995, NA12286, NA12287, NA12347, NA12414, NA12760, NA12777, NA12873). Results for the ancient samples were compared to the results for CEU by randomly sampling a F_{st} from each distribution of results and subtracting the F_{st} found in CEU from the F_{st} found in the ancient samples. After 100,000 repetitions the fraction of values < 0 was calculated. Thereby we were able to approximate the probability of obtaining a smaller F_{st} in internal comparisons of the ancient samples, compared to a modern population.

Inferring global inbreeding coefficient F

We implemented a method to quantify the Wahlund Effect with the inbreeding coefficient *F*.

Model

Let *F* be defined as the global deficit in heterozygotes in a population of individuals $i = 1,\ldots,I$ at all loci $l = 1,\ldots,L$. The genotype of an individual (i) at a locus (l) is denoted by g_{il} and the alternative allele's frequency in the population by p_l.

We define two sub-models for *F*: M_F, where $0 \leq F \leq 1$ and the genotype are distributed according to Hardy-Weinberg Equilibrium with inbreeding (Equation 1), and M_{HWE}, where the genotype frequencies are determined by the allele frequencies according to Hardy-Weinberg Equilibrium (Equation 2).

$$P(g_{il} \vee M_F, F, p_l) = \begin{cases} (1-F)(1-p_l)^2 + F(1-p_l) & \text{if } g_{il} = 0 \\ (1-F)2p_l(1-p_l) & \text{if } g_{il} = 1 \\ (1-F)p_l^2 + Fp_l & \text{if } g_{il} = 2 \end{cases} \quad \text{Equation 1}$$

$$P(g_{il} \vee M_{HWE}, F, p_l) = \begin{cases} (1-p_l)^2 & \text{if } g_{il} = 0 \\ 2p_l(1-p_l) & \text{if } g_{il} = 1 \\ p_l^2 & \text{if } g_{il} = 2 \end{cases} \quad \text{Equation 2}$$

In order to infer *F* and *p* from the sequencing data (d) we assume independence of loci and individuals.

Please cite this article in press as: Burger et al., Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.08.033

We define two submodules for the allele frequency: $M_p$for polymorphic loci, where $p \sim B(\gamma)$, and M_0 for monomorphic loci. The likelihood function of our model is shown in Equations 3 and 4, where the products runs over all loci in model M_p.

$$P(g_{il} \vee M_F, F, p_l, \gamma) = \prod_{M_p^{(l)}} P(p_l|\gamma) \left[\prod_i \sum_g P(d_{il}|g) P(g|F, p_l) \right] \quad \text{Equation 3}$$

$$P(g_{il} \vee M_{HWE}, p_l, \gamma) = \prod_{M_p^{(l)}} P(p_l|\gamma) \left[\prod_i \sum_g P(d_{il}|g) P(g|p_l) \right] \quad \text{Equation 4}$$

We estimate posterior distributions of the parameters $\theta = \{F, p, \gamma, \pi\}$ with a reversible-jump Metropolis-Hastings algorithm.

Priors

We assume $P(M_F) = P(M_{HWE})$ and $F \sim U(0,1)$. We define $P(M_p) = \pi$ and $P(M_0) = 1 - \pi$ and $\pi \sim U(0,1)$. The polymorphic loci are modeled according to a symmetric beta distribution: $p_l \sim B(\gamma)$. We assume extreme allele frequencies to be the most probable, i.e., γ to be small. For ease of implementation we update $c = log(\gamma)$, where $c \sim U(-\infty, \infty)$. This variable transformation has the benefit of allowing the use of the symmetric uniform prior for γ instead of the asymmetric exponential distribution.

Inference

We used ATLAS ([69], commit 52a2276, task = majorMinor) to identify the major and minor allele and to create a VCF containing genotype likelihoods of all Tollense > 4x from the GLF files (see above). We then used ATLAS ([69], commit 52a2276, task = inbreeding) to quantify population structure among the Tollense > 4x sample by estimating the inbreeding coefficient F, running the MCMC for 10^6 iterations with the parameters minVariantQuality = 40, numBurnins = 10 burninLength = 500. We reran this analysis also with argument probMovingToModelNoF = 0 to restrict the analysis to a model with F > 0.

Estimating LP allele frequencies

■■■

Allele frequency estimation for NGS data

We estimated the allele frequencies in Table 1 using ATLAS ([69], commit 52a2276, task alleleFreq, doBayesian). Whole genome sequencing was available for Tollense, CEU and Mokrin and we used the genotype likelihoods in the VCFs created as described in section Allele count estimation as input. For the other NGS populations in Table 1, we downloaded the published BAM files and analyzed them with the standard GATK pipeline [76] to obtain genotype likelihoods for rs4988235-A. Frequencies for rs4988235-A for East European Steppe: 3600-2300 BCE and Area of Corded Ware Culture: 2900-2300 BCE are based on pseudo-haploid calls, taken from the V42.4 version of the downloadable genotypes of present-day and ancient DNA data (compiled from published papers), provided by the Reich Lab (https://reich.hms.harvard.edu/downloadable-genotypes-present-day-and-ancient-dna-data-compiled-published-papers).

We used ATLAS ([69], task = compareAlleleFreq, commit 8af19b9) to obtain posterior probabilities of the allele frequency of Tollense to be smaller than the allele frequency of other populations. For this, we assumed the allele frequencies to follow a Beta distribution with parameters alpha and beta set to 0.7. The results only changed minimally when using an uninformative prior with alpha and beta set to 1 and did not change the fact that Medieval populations had higher frequencies for rs4988235-A than the Bronze Age populations.

Allele frequency estimation for PCR data

In Europe, LP is caused by a single C to T transition 13,910 bp upstream of the LCT gene (−13,910*T, rs4988235). The LCT data was obtained through single or multiplex PCR amplification and Sanger and/or 454 sequencing (see Table S3 for details). The laboratory work was performed in dedicated ancient DNA facilities of the Palaeogenetics Group at Johannes-Gutenberg University Mainz. Methods for sample preparation, extraction, amplification and sequencing, including decontamination methods and criteria for replication and validation, followed Wilde et al. [30]. The primer sequences used in the single and multiplex PCR setups described therein are as follows: LCTa_U 5′CTGCGCTGGCAATACAGATAAGA3′ and LCTa_L 5′CAAATGCAACCTAAGGAGGAGAGTT3′.

In brief, at least four amplifications of two independent extractions were required to consider a genotype as fully replicated. For homozygous individuals, all four amplifications need to show the same allele, for heterozygous individuals, the alternate allele or a heterozygous signal has to be present at least twice and once in each extraction. Only nine of the 37 samples presented here do not entirely fulfill our own strict criteria but miss them by one amplification (see Table S3). We include them because the associated blank controls and additional mitochondrial and nuclear marker amplifications do not show any sign of contamination (see [30]).

We obtained the allele frequencies by dividing the derived allele count by the total amount of alleles. The Credible Intervals were determined as the 5th/95th percentile of a beta distribution, defined by $\beta(n_p + 1, n_q + 1)$, where n_p and n_q were the number of ancestral and derived alleles.

Please cite this article in press as: Burger et al., Low Prevalence of Lactase Persistence in Bronze Age Europe Indicates Ongoing Strong Selection over the Last 4,000 Years, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.08.033

Estimating Allele Age

Consider two alleles A and a segregating in a population of size $2N$. The frequency of Allele A in generation t is a random variable $X(t)$ taking values between 0 and $2N$. Under the Wright-Fisher model with selection, these frequencies form a time-homogeneous Markov chain with transition probabilities

$$P_{ij} := P(X(t) = j | X(t-1) = i) = \binom{2N}{k} p_i^j (1-p_i)^{2N-j} \quad \text{Equation 5}$$

where

$$p_i = \frac{w_{AA} i^2 + w_{Aa} i(2N-i)}{w_{AA} i^2 + 2 w_{Aa} i(2N-i) + w_{aa}(2N-i)^2},$$

and $w_{AA} = 1 + s$, $w_{Aa} = 1 + hs$, and $w_{aa} = 1$ are defined by the selection coefficient s and the dominance coefficient h.

Let us denote by t_0 the time at which genomic data $D = (D_1, \ldots, D_n)$ is available for n samples. Further, let T denote the age of the selected allele A, i.e., $t_0 - T$ denotes the time at which a mutation changed its frequency from 0 to 1. We aim at inferring the posterior probability

$$P(T | D, 2N, s, h) \propto P(D | T, 2N, s, h) P(T).$$

Given known parameters $2N$, s and h. We assume here an improper prior $P(T) \propto 1$. Since the true allele frequency at t_0 is unknown, we have

$$P(D | T, 2N, s, h) = \sum_{x=1}^{2N} P(D | X(t_0) = x) P(X(t_0) = x | T, 2N, s, h).$$

We use ATLAS to determine

$$P(D | X(t_0) = x) = \prod_{i=1}^{n} P(D_i | X(t_0) = x)$$

with

$$P(D_i | X(t_0) = x) = \sum_{g=0}^{2} P(D_i | g) P(g | X(t_0) = x),$$

where $P(D_i|g)$ are the genotype likelihoods calculated by ATLAS and $P(g|X(t_0) = x)$ are the Hardy-Weinberg proportions.

To determine $P(X(t_0) = x | T, 2N, s, h)$ we will resort to a Monte-Carlo scheme. Specifically, we will sample $k = 1, \ldots, K$ allele frequency trajectories x_k by repeating the following steps:

- Set $x_k(0) = 1$ and time $t = 0$.
- Simulate the next frequency $x_k(t+1)$ according to the transition probabilities in Equation 1. Increment t.
- Repeat Step 2 for t_{max} generations or until the chain is absorbed at 0.

Let T be a matrix with elements T_{xt} representing the number of simulated trajectories that had frequency x at time t. From this matrix, we get the estimate

$$P(X(t_0) = x | T, 2N, s, h) \approx \frac{T_{xT}}{\sum_{y=1}^{2N} T_{yT}}.$$

Inference

We set the parameter to: $t_{max} = 100000$, $s = 0.06$, $n_0 = 1$, $N \in \{100, 200, 500, 1000, 2000, 5000, 10000, 20000, 50000, 100000\}$, $h \in \{0.5, 1\}$ and ran the simulations for 10000 iterations. The allele age was transformed from generations to the Before Present (BP) time-scale in the following way:

$$age = 25 * generations + 3150,$$

where 25 is the number of years per generation and 3150 is the age of Tollense in BP.