Article: Discoveries

Population Genetics and Signatures of Selection in Early Neolithic European Farmers

Ainash Childebayeva[1,2,*], Adam Benjamin Rohrlach[1,2,3], Rodrigo Barquera[1,2], Maïté Rivollat[1,4], Franziska Aron[1], András Szolek[5,9], Oliver Kohlbacher[5,6,7,8], Nicole Nicklisch[10,11], Kurt W. Alt[10,11], Detlef Gronenborn[12], Harald Meller[11], Susanne Friederich[11], Kay Prüfer[1,2], Marie-France Deguilloux[4], Johannes Krause[1,2], Wolfgang Haak[1,2,*]

[1]*Archaeogenetics Department, Max Planck Institute for the Science of Human History, Kahlaische Straße 10, D-07745 Jena, Germany*

[2]*Archaeogenetics Department, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany*

[3]*ARC Centre of Excellence for Mathematical and Statistical Frontiers, School of Mathematical Sciences, The University of Adelaide, Adelaide, Australia*

[4]*Université de Bordeaux, CNRS, PACEA-UMR 5199, 33615 Pessac, France*

[5]*Applied Bioinformatics, Dept. of Computer Science, University of Tübingen, Tübingen, Germany.*

[6]*Institute for Bioinformatics and Medical Informatics, University of Tübingen, Tübingen, Germany.*

[7]*Translational Bioinformatics, University Hospital Tübingen, Tübingen, Germany.*

[8]*Biomolecular Interactions, Max Planck Institute for Developmental Biology, Tübingen, Germany.*

[9]*Department of Immunology, Interfaculty Institute for Cell Biology, University of Tübingen, Tübingen, Germany*

[10]*Center of Natural and Cultural Human History, Danube Private University, Krems-Stein, Austria*

[11]*State Office for Heritage Management and Archaeology Saxony-Anhalt - State Museum of Prehistory, Halle (Saale), Germany*

[12]*Römisch-Germanisches Zentralmuseum, Leibniz Research Institute for Archaeology, Ernst-Ludwig-Platz 2, 55116 Mainz, Germany* https://orcid.org/0000-0002-6467-5643

**Correspondence to: ainash_childebayeva@eva.mpg.de; wolfgang_haak@eva.mpg.de*

Keywords: Selection, Neolithization, Ancient DNA, Linear Pottery Culture

ABSTRACT

Human expansion in the course of the Neolithic transition in western Eurasia has been one of the major topics in ancient DNA (aDNA) research in the last ten years. Multiple studies have shown that the spread of agriculture and animal husbandry from the Near East across Europe was accompanied by large-scale human expansions. Moreover, changes in subsistence and migration associated with the Neolithic transition have been hypothesized to involve genetic adaptation. Here we present high quality genome-wide data from the Linear Pottery Culture (LBK) site Derenburg Meerenstieg II (DER) (N=32 individuals) in Central Germany. Population genetic analyses show that the DER individuals carried predominantly Anatolian Neolithic-like ancestry and a very limited degree of local hunter-gatherer admixture, similar to other early European farmers. Increasing the LBK cohort size to ~100 individuals allowed us to perform various frequency- and haplotype-based analyses to investigate signatures of selection associated with changes following the adoption of the Neolithic lifestyle. In addition, we developed a new method called AIMLESS (*Admixture-informed Maximum-likelihood Estimation for Selection Scans*) that allowed us test for selection signatures in an admixture-aware fashion. Focusing on the intersection of results from these selection scans, we identified various loci associated with immune function (*JAK1*, HLA-DQB1) and metabolism (*LMF1, LEPR, SORBS1*), as well as skin color (*SLC24A5, CD82*) and folate synthesis (*MTHFR*, *NBPF3)*. Our findings shed light on the evolutionary pressures, such as infectious disease and changing diet, that were faced by the early farmers of Western Eurasia.

INTRODUCTION

The Neolithic transition, a shift from a foraging to a farming-based subsistence, marks one of the most substantial social, ecological and economic transformations in the prehistory of West Eurasia. The process of neolithization has interested and fascinated archaeologists, anthropologists, demographers, epidemiologists, specialists in evolutionary medicine, and the general public for more than a hundred years (Bocquet-Appel and Bar-Yosef 2008; Bogucki 1999; Chamberlain 2006; Dawkins 1894; Johnson and Earle 2000; Lubbock 1866). In retrospect, the Neolithic was a transformative period in which a mostly sedentary lifestyle, domestication of plants and animals, and the associated relative independence from nature with reliance on sustained food supplies came into being. A manufacturing economy with good yields, surplus, and stockpiling caused an increase in fertility, with earlier age at weaning and subsequent shorter birth intervals which brought about an exponential population increase (Bocquet-Appel and Bar-Yosef 2008; Deevey 1960). The negative effects of neolithization manifested in an increased prevalence and transmission of infectious, metabolic, and nutritional deficiency diseases, favored by sedentary lifestyles, growing populations, and close contact with domesticated animals (Barrett et al. 1998; Cordain 1999; Gering et al. 2019; Harper and Armelagos 2010).

Archaeological research has shown that agriculture and animal husbandry originated in the Near East around 10,000 BCE (Zeder 2011), from where these practices spread to North Africa, Europe and South Asia. With the help of ancient DNA (aDNA) studies, it was shown that the spread of agriculture across Europe was not a transfer of ideas and technologies, but was instead mediated by the expansion of the early farmers, i.e. through demic diffusion (Gamba et al. 2014; Haak et al. 2010; Haak et al. 2015; Hofmanova et al. 2016; Lazaridis et al. 2014a; Mathieson et al. 2015; Skoglund et al. 2014; Skoglund et al. 2012). By ~6,000 BCE, sedentary farming was established in the Aegean (Horejs et al. 2015), southeastern Europe and the Carpathian Basin (Bánffy 2019), and direct connections have been established between the early European farmers (EEF) from Neolithic Hungary and Greece and the early farmers from western Anatolia (Hofmanova et al. 2016; Lazaridis et al. 2016; Lipson et al. 2017; Mathieson et al. 2015). Confirming previous archaeological findings (Gronenborn 2014; Zvelebil 2001), genetic studies have reported evidence for two routes of early farmer expansion from the Near East: the inland Central European route; and the coastal (Mediterranean) route to western Europe, Iberia (Brunel et al. 2020; Olalde et al. 2015; Rivollat et al. 2020), and eventually the Atlantic Archipelago (Brace et al. 2019; Cassidy et al. 2016; Rohrlach et al. 2021).

The Linear Pottery Culture (or LBK after German *Linearbandkeramik*) is associated with early accounts of farming in Neolithic Europe (Bánffy 2019; Petrasch 2020; Zvelebil 2004). The LBK culture originated in the western Carpathian Basin in today's Hungary and Slovakia between 5500 and 5400 BCE, and spread across the European loess plains to the Paris Basin in Western Europe (Bickle 2009), and Ukraine in eastern Europe. Archaeologically, LBK emerged from interactions between Starčevo-Körös-Criș complex farmers and local Mesolithic hunter-gatherers (HGs) (Gronenborn 2007). Numerous LBK sites have been found in Central Germany, particularly in the Mittelelbe-Saale (MES) region, a biogeographical region that was attractive to the early Neolithic farmers due to its fertile soils, waterways and adequate levels of precipitation.

The emergence of agriculture in Western Eurasia has been linked to new environmental and cultural pressures. During this time, HG groups in the area of the Fertile Crescent transitioned to a more sedentary lifestyle associated with increased population sizes, a change in diet and, importantly, increased exposure to infectious diseases from animals, e.g. *Salmonella enterica* (Armelagos et al. 1991; Barrett et al. 1998; Harper and Armelagos 2010; Key et al. 2020). Successful settlement in new geographic regions, such as central Europe, would have required rapid adaptation to new environmental, economic, and social conditions. Lower meat consumption and increased caries lesions in teeth indicative of cereal consumption have been reported for the early Neolithic compared to later time periods (Münster et al. 2018; Nicklisch et al. 2016). Moreover, increased prevalence of *cribra orbitalia* and porotic hyperostosis, which indicate either a significant burden of infectious diseases, low quality diet, or a combination of both (Walker et al. 2009), was found at LBK sites (Ash et al. 2016; Nicklisch 2017). A previous study of selection in Europe, using aDNA, in general reported evidence of selection for light skin color, infectious disease resistance, and fatty acid metabolism, among other signals (Mathieson et al. 2015). Additionally, Neolithic Aegeans have been shown to carry alleles associated with reduced skin pigmentation and type 2 diabetes susceptibility, and a number of inflammatory disease-associated loci (Hofmanova et al. 2016). Taken together, early Neolithic farmers likely faced selective pressures from the increased exposure to pathogens and the change in diet.

Here we report new genome-wide data from human remains associated with an LBK burial site Derenburg Meerenstieg II (Wernigerode, Saxony-Anhalt, Germany). The Derenburg (DER) burial ground is located in Saxony-Anhalt, close to the Holtemme tributary of the Bode River, in the northern Harz foreland (Fig. 1). Previous studies using stable isotopes have shown that mean adult human values for carbon and nitrogen isotopes were typical for the region and

suggested a mixed farming diet including domesticated plant and animal products (Münster et al. 2018; Oelze et al. 2011). A previous study of mitochondrial DNA (mtDNA) lineages from the EEF, including the individuals in this study, showed that the mtDNA haplogroups of LBK individuals and their frequency distribution is more similar to the present-day population of Anatolia and the Near East (Haak et al. 2010). This indirectly argued for demic diffusion from Anatolia to Europe, and a degree of genetic continuity between the two regions, which was later confirmed using genome-wide studies (Hofmanova et al. 2016; Lazaridis et al. 2016; Lipson et al. 2017; Mathieson et al. 2015; Rivollat et al. 2020).

One of our aims was to provide an updated genomic portrait of DER in the context of genome-wide variation of the EEF. We explored the genomic ancestry of LBK individuals comparing them to other EEF and Anatolian Early farmers (AEF), as well as local European Western Hunter-Gatherers (WHG) on a broader scale. We also investigated the intra-site biological relatedness and explored the demography of LBK groups by comparing DER individuals to previously published individuals from nearby sites, such as Halberstadt (only ~10-15 km apart), and other LBK sites from Germany, Poland, Austria, and Hungary (Lipson et al. 2017; Mathieson et al. 2018; Rivollat et al. 2020).

NEW APPROACHES

Leveraging the substantially increased cohort size of the early European farmers, we performed various scans of selection to determine specific genes and pathways that were under selection in LBK individuals, as well as the early farmers in general. To our knowledge, genetic adaptation associated with the Neolithic lifestyle has not been formally tested using aDNA. We performed allele frequency- (LSBL and our own newly developed method called *Admixture-informed Maximum-likelihood Estimation for Selection Scans* (AIMLESS)) and haplotype-based selection scan (XP-EHH), and analyses of the HLA class I and II alleles, to investigate whether the LBK population was under an evolutionary pressure as a result of the Neolithization and population migration/expansion. Moreover, we performed selection tests comparing early farmer individuals to modern-day African populations and the local HGs to test general adaptation to the Neolithization in Western Eurasia.

The ability to detect true loci that are under selection is based on how well the population structure is accounted for, especially in a test that depends on comparisons between populations. Thus, we developed AIMLESS to test for selection in an ancestry-aware fashion. This method is a modification of the admixture-aware linear model-based test developed by Mathieson et al. (2015), however, our test is specifically designed for two population admixture scenarios. AIMLESS is similar to other LRT-based tests such as *AdaptMix* (Mendoza-Revilla et

al. 2021) and Ohana (Cheng et al. 2022), but does not rely on *a priori* ancestry estimation, and determines the mixture components for the ancestries considered for each SNP independently instead of using a genome-wide estimate.

RESULTS and DISCUSSION

***Data generation and authentication*.** We generated genome-wide single-nucleotide polymorphism (SNP) capture data for ~1.24 million variant sites (1240k SNP array) across the genome (Fu et al. 2015; Mathieson et al. 2015), mitochondrial genome capture (Maricic et al. 2010), Y-chromosome capture (YMCA) (Rohrlach et al. 2021) and 3Mb immune-capture data (Immel et al. 2021) from N=32 individuals. In addition, we sequenced the complete genomes of two DER individuals (DER002 2.26X and DER009 3.07X), and a local HG individual from Bad Dürrenberg, Germany (BDB001 13x), for which genome-wide SNP capture data was previously reported (Rivollat et al. 2020). A subset of DER individuals (N=10) was radiocarbon dated to ca. 5300-4800 cal. BCE (Sup. Table 1) or ~7000 years before present (BP), which is consistent with the younger phase of the LBK period in Germany (Gronenborn et al. 2014).

The authenticity of human aDNA from the DER individuals was confirmed using MapDamage (Ginolhac et al. 2011), wherein all aDNA fragments were shown to have deamination patterns consistent with degradation over time, and unambiguous sex determination. Male samples were checked for X chromosome contamination from modern sources, using ANGSD (Korneliussen et al. 2014). Mitochondrial DNA contamination was estimated using Schmutzi (Renaud et al. 2015) and contamMix (Fu et al. 2013). No significant contamination (>5%) from modern sources was determined in the DER cohort (Sup. Table 2). Based on the genetic sex determination, we identified 18 females and 14 males (54.5% female).

***Genome-wide data analysis*.** We genotyped the 1240k SNP positions using pileupCaller (https://github.com/stschiff/sequenceTools/tree/master/src-pileupCaller) obtaining pseudohaploid calls for each variant position on the capture. We merged the 1240k SNPs with the modern-day Human Origins dataset, as well as relevant ancient samples (https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data, version 44.3), to perform a Principal Component Analysis (PCA). Due to the incomplete nature of the ancient data, ancient individuals were projected onto the genetic variation of 1363 modern West Eurasians (Fig. 1A). Based on the position in PC space, we found that the DER individuals fall together with published LBK individuals, as well as contemporaneous early Neolithic individuals from southeastern Europe and western Anatolia (Fig. 1A). We then performed an unsupervised ADMIXTURE analysis and found that at K=12

DER represent a mixture of an ancestry component maximized in Anatolian Neolithic individuals and a smaller proportion of ancestry component maximized in WHG (Fig. 1B).

Following the PCA and ADMIXTURE analyses, we calculated various F-statistics to formally test the ancestry components identified in the admixture analysis and inferred from the position on the PCA plot. We first performed an f4-test of the form f4(Mbuti, HG; DER, Anatolia_N) in order to test whether DER individuals carried additional HG ancestry compared to Anatolia_N (Fig. 2A). Significantly negative f4-values (|Z|>3) for models in which England_Mesolithic, Villabruna, Loschbour HG, France_Rochedane, and France_Chaudardes were used as HG, indicated excess WHG ancestry in DER. We also obtained a significantly negative f4-value for the Iron Gates HGs (Mathieson et al. 2018).

In a follow-up f4-analysis of the form f4(Mbuti, HG; EEF, Anatolia_N) we iterated through a number of EEF groups and tested the following two HG sources specifically: Loschbour HG (WHG) and Iron Gates HGs (Sup. Fig. 1). Additionally, we tested BDB001 (Rivollat et al. 2020) as a local HG ancestry proxy from the MES region. The DER individuals showed significantly negative f4-results suggesting more HG ancestry compared to Anatolia_N in the first two cases (Loschbour HG and Iron Gates), but not for the local HG proxy BDB001. This suggests that DER individuals had received additional HG admixture upon their expansion into Europe, but not HG ancestry immediately related to the local BDB001 source. When comparing all LBK groups, many were cladal with Anatolia_N, indicating a rather swift expansion with only occasional contribution of additional (and local) HG ancestry.

In turn, when comparing the HG sources to each other, we could not identify which source contributed the most to DER (Sup. Fig. 2). However, we did see an increased affinity of DER to the Loschbour HG representing the WHG ancestry, compared to Iron Gates HGs who have additional eastern HG ancestry (Sup. Fig. 2) (Mathieson et al. 2018).

We then used qpAdm to quantify the relative contribution of HG ancestry represented by the Loschbour HG and the Anatolia_N ancestry represented by Neolithic individuals from Barcin, western Anatolia (Mathieson et al. 2015). When modelled individually, we found that DER individuals had between 0%-20% (SE 6.6%) European HG ancestry (Fig. 2B, Sup. Table 3). When modeled as a group, we found that on average DER individuals had ~7% HG and ~93% western Anatolian Neolithic ancestries, indicating limited HG contribution (Fig. 2C).

Using the same set of sources (Anatolia_N and Loschbour HG), we estimated the time of admixture between the Anatolian farmers and WHG using DATES v. 753 (Moorjani et al. 2016) to have occurred 24.286 +/- 2.784 generations ago (Sup. Fig. 3), or ~680 calendar years before the time of the DER individuals assuming a generation time of 28 years (Fenner 2005),

which falls at the beginning of the Starčevo-Körös-Criș complex (~6,200-5,200 BCE) of the southeastern European Neolithic (Porčić et al. 2020).

According to both the model-free ADMIXTURE and the qpAdm results with a defined set of sources, we can show very limited HG ancestry influx in the DER LBK group in addition to the HG ancestry already carried by Anatolian farmers. This suggests minimal interaction between European HG and expanding early LBK farmers from southeastern Europe during the Neolithic, corroborating findings by other studies (Lipson et al. 2017; Mathieson et al. 2018; Rivollat et al. 2020).

Uniparentally inherited markers. Mitochondrial DNA and Y-chromosome haplogroups of the DER individuals are generally consistent with the genetic variation reported from early Neolithic farmers (Sup. Table 1) (Brandt et al. 2013; Szecsenyi-Nagy et al. 2015), and the mitochondrial genomes match the HVR-I haplotypes reported earlier (Haak et al. 2010). We primarily found G2a2 and H2 Y haplogroups associated with Neolithic farmers (Haak et al. 2015; Lacan et al. 2011; Rivollat et al. 2020; Rohrlach et al. 2021). We also observed one male with the Y-chromosome I haplotype that is generally common among European HGs and therefore considered as a signal of male HG ancestry contribution, but also reported in lower frequency from Iberian and French Neolithic individuals (Lipson et al. 2017; Rivollat et al. 2020).

Biological Relatedness*.* Based on the analysis of biological relatedness via READ (Monroy Kuhn et al. 2018) and lcMLkin (Lipatov et al. 2015), we identified three first degree relationships (two-parent offspring, one sibling pair) and second-degree related pairs (Sup. Table 1). The mother-daughter pair (DER019 mature female - DER018 infant) was buried in the same grave, while the mother-son pair (DER022 mature female - DER011 mature male) was not. Despite the integration of contextual anthropological and archaeological evidence such as age-at-death and uniparentally-inherited haplogroups, we were not able to construct the true single pedigree of the 2nd degree pairs, as several alternatives were possible.

We performed Wilcoxon Rank Sum tests on the mean measure of relatedness for each individual to the rest of the group as calculated from READ to see whether there is a general skew towards males or females from the site being more closely related to each other, which would indicate either potential patri- or matrilocality. We compared the pair-wise relatedness of all males to all females (p-value=0.34), all adult males to females (p-value=0.12), and the subadults (p-value=0.35), none of which significantly differed from each other (Sup. Fig. 4).

We determined the degree of inbreeding in the DER cohort, and compared it to other Neolithic EEF populations using hapROH 0.1a6 (Ringbauer et al. 2021). Based on the analysis of runs of homozygosity (ROH), we found evidence of low levels of background relatedness as

indicated by low values for the sum of short ROH segments. Moreover, compared to other EEF populations, we did not see different levels of inbreeding among individuals at DER (Fig. 2E). In general, we found consistently low levels of inbreeding across all early Neolithic groups in Central Europe, suggesting that early European farmers lived in larger groups or groups with an extended mating network, and ultimately stemmed from a source deme with a large effective population size (Ringbauer et al. 2020).

The biological relatedness at DER is similar to other LBK sites from Germany, e.g. Halberstadt-Sonntagsfeld, where one first- and two first- or second-degree related pairs were observed out of 24 total individuals (Lipson et al. 2017). Considering that the DER burial site is archaeologically a closed find and that the temporal spread of the calibrated years before present (calBP) is around 300 years, 6848.5-7141calBP (Sup. Table 1), the few first- and second-degree relationships among a total of 32 genotyped DER individuals suggest that the site was used over a longer period of time and not restricted to (a) particular biological kin group(s).

***LBK inter-site comparisons*.** We further tested the observations on LBK population structure inferred from biological relatedness results using inbreeding coefficients and effective population size comparisons of LBK sites in Germany with other already published EEFs. We used READ (Monroy Kuhn et al. 2018) to test for relatedness between individuals from DER and other LBK sites in Germany, mainly already published data from Halberstadt-Sonntagsfeld (Germany_LBK_HBS, N=29) and Stuttgart-Mühlhausen (Germany_LBK_SMH, N=29) (Lipson et al. 2017; Rivollat et al. 2020). In order to do this, we first estimated the median relatedness in each German LBK group separately, and found them to be comparable to each other (LBK_HBS = 0.24; LBK_SMH=0.25; LBK_DER=0.24). We then performed a joint biological relatedness analysis of all German LBK sites using READ, and found no evidence of first- or second-degree relatedness between individuals from different sites.

To investigate this further, we estimated the effective population size (N_e) in LBK as implemented in (Fernandes et al. 2021). Using the maximum likelihood to fit N_e, we estimated the effective population size to have been ~5000 individuals (3688–6778 95% CI) for all LBK sites from Germany, which is consistent with effective population size estimates for a relatively large population (Mele et al. 2012; Tenesa et al. 2007). Together with the results from our analysis of the lack of inbreeding and inter-site biological relatedness, this further suggests that the early Neolithic LBK farmers were part of a larger population, did not practice close biologically related mating, and prevented inbreeding. In comparison, based on limited biological

relatedness among individuals in intramural burials, it has been suggested that Neolithic Anatolia was not a strictly kin-based society (Pilloud and Larsen 2011; Yaka et al. 2021).

Phenotypic analysis. We determined genotype likelihoods for a select number of SNPs associated with genetic adaptation, including lactase persistence, metabolic adaptation, among others, as well as SNPs associated with phenotypic traits as determined by the H-IrisplexS platform (Chaitanya et al. 2018; Walsh et al. 2014) (Sup. Fig. 5). Based on the analysis of the SNPs associated with phenotypes, we see the highest frequencies of brown (33%) eye color, intermediate skin color (21%) and brown (21%) hair color. We report these data being fully aware of the fact that the analysis of phenotypic variation can be a contentious issue, due to the fact that it was developed based on modern individuals of predominantly European ancestry, and does not perform as well in admixed individuals and populations outside of Europe (Dembinski and Picard 2014; 2016; Yun et al. 2014). Thus, we note that these data, and specifically eye, skin, and hair color, are frequent requests of museums and other organizations to illustrate reconstruction life histories of past societies.

***HLA class-I and II allele analysis*.** HLA Class-I and II haplotypes in the DER individuals were analyzed using Optitype (Szolek et al. 2014). Based on the analysis, we report the highest allele frequencies for *HLA-A* to be A*24:02 (0.44) and A*02:01 (0.38), B*51:01 (0.31) and B*27:05 (0.25) for *HLA-B*, and C*02:02 (0.25) and C*15:02 (0.23) for *HLA-C* (Sup. Tables 4 and 5). In comparison, Late Neolithic individuals from the Wartberg culture show similar distributions of the HLA class I alleles, with HLA-B*27:05 at a frequency of 23% and HLA-C*02:02 at 17% (Immel et al. 2019), which is higher than the modern frequencies of the two alleles in Europeans being 3.73% (Wu et al. 2021) and 5.6% (UK pop 3 from (Gonzalez-Galarza et al. 2021)), respectively. Interestingly, the two highest frequency *HLA-B* alleles in our sample, HLA-B*51:01 and HLA-B*27:05, have been associated with Behçet's disease susceptibility (Kirino et al. 2013; Verity et al. 1999) and ankylosing spondylitis (AS) (Cauli et al. 2013; Chen et al. 2017; International Genetics of Ankylosing Spondylitis et al. 2013). Based on the skeletal analysis of the individuals, the carriers of the HLA allele B*27:05 among the DER individuals did not show skeletal signs of AS, even though approximately 90% of the AS patients are HLA-B*27^{+} (Brewerton et al. 1973; Schlosstein et al. 1973; Woodrow and Eastmond 1978). While only 1-5% of the HLA-B*27 positive individuals develop ankylosing spondylitis in the general population (Reveille et al. 2010; van der Linden et al. 1983), it is recognized as one of the most important monogenic associations for this disease known to date (Karnes et al. 2017; Reveille et al. 2010; van der Linden et al. 1983). It is possible that some of the carriers in our sample could have still developed the condition later in life (Feldtkeller et al. 2003). The most frequent HLA class II

alleles are HLA-DRB1*11:01 (0.42), -DRB3*02:02 (0.48), -DQA1*05:05 (0.44), and -DQB1*03:01 (0.52) (Sup. Tables 4 and 5). All these alleles show similar frequencies as they are all part of the same HLA class II haplotype (43.75%), which was reported in relatively high frequencies in present-day central European populations such as Czech Republic (15.30%) (Zajacova et al. 2016), but also in Nganasan from Siberia (18.80%) (Uinuk-Ool et al. 2004).

Selection scans. The transition from hunting and gathering to agriculture and animal husbandry in the Neolithic has been linked to new selective pressures due to an increased pathogen burden (Armelagos et al. 1991; Key et al. 2020) and a shift to a cereal-based diet together with decreased nutritional diversity (Luca et al. 2010). We performed two types of selection scans (frequency- and haplotype-based) to detect signatures of selection associated with Neolithization in the newly generated DER cohort together with already published data. We first describe the settings and cut-offs used in each analysis and then address the intersection of the results in a joint discussion.

We first used a locus-specific branch length (LSBL) test for selection (Shriver et al. 2004). The LSBL test is based on the pair-wise genetic distances between three populations, wherein each SNP is considered independently. SNPs with the longest Fst branches in one population compared to the other two are considered as potential selection candidates. We performed three different LSBL tests to determine whether the transition from hunting and gathering to an agricultural lifestyle and the population expansion from the Near East was associated with adaptation by natural selection to new environments in higher latitudes of Europe and to the Neolithic way of life in general.

First, we compared three cohorts consisting of LBK (N=97), Neolithic individuals from western Anatolia (N=37), and autochthonous HGs as local control group (N=86), (see Sup. Fig. 6 for the PCA with the individuals used in our cohorts, and Sup. Table 6 for the list of the individuals). We chose to compare the LBK to the Anatolian farmers because these populations are known to be closely related based on our PCA and Admixture analyses, as well as previous studies (Lazaridis et al. 2016; Mathieson et al. 2015). SNPs with the largest pair-wise Fst values between Anatolian farmers and the LBK, as well as between LBK and HGs were prioritized, since these SNPs could indicate selection for the new environmental or lifestyle conditions associated with the expansion into and across Europe, and ongoing Neolithization, respectively. Based on the LSBL analysis, we prioritized N=9549 genetic loci under the empirical quantile < 0.01 (top 1%) that was calculated based on the distribution of the LSBL statistic following previous studies (Bigham et al. 2010).

Second, to test for selection since the deeper split of LBK and Anatolia_N from WHG, we merged the ancient farmer populations (LBK and Anatolia) into a single Early Farmer (EF) group, and compared it to WHG, and modern Esan (ESN) individuals from the 1KG dataset (Genomes Project et al. 2015) as a non-European farming group comparison. In a third analogous LSBL analysis, we also compared EF to WHG and an African HG cohort represented by Biaka and Mbuti from the SGDP and HGDP datasets, respectively (Bergström et al. 2020; Mallick et al. 2016). During this deeper time frame we expect the Neolithization in Western Eurasia to have had a substantial effect due to the change in subsistence and increased population densities and disease susceptibility.

We also performed an ancestry-aware selection scan using our newly developed method AIMLESS. For each position on the genome for which we had allele frequencies for all three groups (LBK, Anatolia_N and WHG) we performed a maximum likelihood estimation (MLE) of allele frequencies in the LBK based on the observed data (unconstrained model), which we compared to the MLE based on the most likely admixture of Anatolia_N and WHG ancestries (constrained model). We performed a likelihood ratio test comparing two models, and determined p-values based on a chi-square distribution with 2 degrees of freedom, and considered loci with p-value<1e-8 as significant based on simulation studies (Sup. Figs. 7 and 8).

To look for more recent signals of adaptation (Szpiech and Hernandez 2014) to the Neolithic lifestyle through the comparison of LBK to WHG, we also explored a haplotype-based selection test called *Cross Population Extended Haplotype Homozogysity* (XP-EHH), which can detect selective sweeps in one population compared to another (Sabeti et al. 2007). To do so, we first used GLIMPSE (Rubinacci et al. 2021) to impute and phase the genome-wide data from DER and already published LBK individuals (Lazaridis et al. 2014b; Lipson et al. 2017; Mathieson et al. 2015; Rivollat et al. 2020), as well as Iron Gates HGs (Mathieson et al. 2018). To understand the coverage cutoff for imputation, we performed a down-sampling experiment, wherein we down-sampled a high-coverage (92x) 1240k SNP captured LBK individual from Stuttgart Mühlhausen, which was obtained as a result of routinely using this library as positive control in capture experiments. Based on our test of the genotype agreement between the down-sampled and diploid data, we decided on a cutoff of 0.5x (LBK: N=31, Iron Gates: N=27) for the inclusion of the imputed samples in the downstream selection scans with selscan (Szpiech and Hernandez 2014) (Sup. Fig. 9). We identified 52,648 sites with the absolute XPEHH value greater than 2.

Due to the limitations of missing data inherent in aDNA and to further overcome the potential issue of false positive results based on imputation or factors like drift and population structure, we prioritized results that were overlapping between XP-EHH, LSBL, and AIMLESS, since it has been shown that a combination of multiple tests is more powerful to detect true selection signals (Grossman et al. 2010). An important caveat is that we used the 1KG data as modern reference panel (Genomes Project et al. 2015) to impute the ancient data. This is not ideal since imputation with a modern reference can introduce spurious genotypes and make the individuals imputed more similar to the reference panel used. We are also aware that the European gene pool underwent further transformation during the Bronze Age (Allentoft et al. 2015; Haak et al. 2015). This is why the selection results based on imputed data have to be analyzed and interpreted with caution.

As predicted, we identified many SNPs associated with immune function and metabolism among the LSBL top 1% hits from the comparison of LBK vs. Anatolia_N and WHG (Sup. Table 7). To determine if any biological pathways are over-represented among the top SNPs identified by the LSBL test of LBK vs Anatolia_N and WHG, we used Gowinda (Kofler and Schlotterer 2012) for a gene-set enrichment analysis. We identified 110 significant GO (Gene Ontology) terms after a correction for multiple comparisons using the False Discovery Rate (FDR) cutoff 0.05. Among the significant results, we found 8 out of 110 GO terms associated with lipid metabolism (OR 1.82 compared to all GO terms) (GO0061365: positive regulation of triglyceride lipase activity, GO0051006: positive regulation of lipoprotein lipase activity, GO0060193: positive regulation of lipase activity, GO0015909: long-chain fatty acid transport, GO0015908: fatty acid transport, among others) and 6 out of 110 with the immune system (OR 0.65 compared to all GO terms) (GO0006911: phagocytosis, GO0002376: immune system process, GO0050900: leukocyte migration, GO0001776: leukocyte homeostasis, and others) (Sup. Table 8). We summarized and visualized the GO terms with REVIGO (Supek et al. 2011) using the default parameters (Sup. Fig. 10). Among the GO term 'superclusters' we identified: 'long-chain fatty acid transport', 'leukocyte homeostasis', 'positive regulation of lipoprotein lipase activity', 'immune system process'.

Overall, we found 227 overlapping positions between the three LSBL tests (Sup. Tables 9-11). Notably, those included SNPs associated with *APOM* rs805297 (linked to rheumatoid arthritis and lipid signaling), *BDNF* rs11030104 (BMI), *IGSF21* rs12031938 and *ILDR2* rs10489574 (immune response), *LMF1* rs12933840 and rs8062983 (lipid metabolism), MHC SNPs rs2844482 and rs9277027 (immune function), and *PDE2A* rs11235559 (susceptibility to viral infections) (Fumagalli et al. 2010). Our findings from the LSBL comparisons highlight the

immune and lipid metabolism pathways as strong selection candidates. This is not surprising, since immune function adaptation has been shown before in previous analyses of natural selection in Western Europe from the Neolithic to modern time (Chekalin et al. 2019; Mathieson et al. 2015). To our knowledge, this is the first time that the continuous selection on the immunity related pathways is shown with regards to the Neolithic expansion from the Near East to Europe.

Among the SNPs that were identified by the LSBL tests for deeper selection in the early Neolithic farmers, we find loci associated with skin color (*SLC24A5, CD82*) adaptation and folate synthesis (*MTHFR*, *NBPF3*) (Sup. Tables 9 and 10). Notably, *SLC24A5,* has been also identified by previous studies of selection in Europeans using aDNA (Ju and Mathieson 2021; Mathieson et al. 2015). Skin color variation is one of the most striking features of human phenotypic diversity. Skin color is highly correlated with latitude and the resulting ultraviolet radiation (UVR) intensity, and populations closer to the equator have darker skin color, while the populations located away from the equator have lighter skin color (Jablonski and Chaplin 2000). It has been long hypothesized that skin color variation has evolved as a balance between vitamin D synthesis and folate synthesis, as well as skin cancer protection. On one hand, vitamin D is synthesized upon UVR exposure (Jablonski and Chaplin 2010), and, on the other hand, folate has to be protected from photolysis from UVR (He et al. 2009). Our findings support the idea of coevolution of folate synthesis and lighter pigmentation in the EF, compared to the HGs and the African outgroup populations. Moreover, our study narrows down the timing of selection on *SLC24A5* to after the split between the European Hunter-Gatherers and the ancestors of the early Neolithic farmers. Of note, European HGs are characterized by a darker skin color, even though they occupied higher latitudes with low UVR exposure. The current hypothesis is that the diet based on meat/fish served as a sufficient source of vitamin D in higher latitudes for the HGs, similar to what has been shown in the Inuit and other Arctic indigenous populations (Kolahdooz et al. 2013; Schaebel et al. 2015).

We analyzed SNPs that are associated with pigmentation in the ancient populations separately (LBK, Anatolia_N, WHG) (Table 1), since we hypothesized that migration to a higher latitude from Anatolia to central Europe could be associated with positive selection for light skin color phenotype. Based on the comparison between LBK and Anatolia_N, we did not see a major shift in skin color SNPs that are involved in determining lighter skin pigmentation, suggesting that this adaptation already happened in Anatolia/the Near East during the transition to a sedentary farming lifestyle.

The results from the XP-EHH test confirmed well-known selection signals linked to skin color adaptation in the gene *SLC24A5* (Lamason et al. 2005) (Sup. Fig. 11). Selection signature around the *SLC24A5* core SNP rs1426654 (xpehh=5.13) suggests that the haplotype that is associated with lighter skin color is highly prevalent in the LBK individuals, and is also under selection in the Iron Gates HGs. We also identified *FADS1* as a potential selection candidate by the LSBL test comparing LBK individuals to Anatolia_N and WHG, and by XP-EHH for SNP rs174546 (xpehh=-2.05) (Sup. Fig. 12). The *FADS1/FADS2* locus has been identified before by previous studies of selection using ancient individuals. It has been shown that FADS was under selection both before and after the Neolithic transition (Ye et al. 2017), with the ancestral allele under selection before (Mathieson 2020) and the derived after (Martiniano et al. 2017).

A novel gene of interest that was also identified by both LSBL and XP-EHH is *LMF1* (Fig. 4A). LMF1 or lipase maturation factor 1 is involved in the lipase system, and is required for the maturation of the lipoprotein lipase, hepatic lipase, and endothelial lipase (for review see (Peterfy 2012)). LMF1 activity is associated with levels of triglycerides, and its mutations are linked to hypertriglyceridemia, or the presence of high levels of triglycerides, and lipase deficiency (Peterfy et al. 2007). LMF1 is present at high levels in mice embryos, suggesting a potential, although currently unknown, role in embryonic development (Ehrhardt et al. 2014). Plant-based diets are characterized by a lower lipid content (Yokoyama et al. 2017), and efficient lipid metabolism could have been under selection especially in LBK individuals, who have been shown to have higher rates of caries than individuals in the Late Neolithic, indicative of the reliance on cereals (Nicklisch et al. 2016). Another novel selection candidate is *JAK1* (Fig. 4B). The JAK-STAT pathway is involved in the tuning of immune response, and is associated with immune disorders (Witalisz-Siepracka et al. 2018). Janus kinases are involved in immune cell signaling, and complete deletions of Jak1 and Jak2 are not compatible with life in mice (Ghoreschi et al. 2009). Polymorphisms in *JAK1* have been linked to blood cell type composition (Astle et al. 2016), white blood cell count (Kichaev et al. 2019).

Lastly, we explored all 971 loci that are shared between the three selection tests AIMLESS, XP-EHH and LSBL (LBK vs. Anatolia_N and HGs). These include SNPs associated with *LEPR*, *IL13*, HLA-A, HLA-B, HLA-DQB2, *CHRAC1*, *BMP1*, *SORBS1*, *ELMO1*, and others (Sup. Table 12, Fig. 3), among which SNPs associated with immune function and metabolism such as *HLA-DQB1*, *SORBS1,* and *LEPR* stood out (Sup. Table 12). The HLA-DQ molecule is a major histocompatibility complex (MHC) class II protein, expressed in B lymphocytes, dendritic cells, and macrophages. Diseases that are associated with the genes HLA-DQA1 and -DQB1 coding for HLA-DQ include Celiac Disease. (Megiorni and Pizzuti 2012; Zubillaga et al. 2002),

autoimmune diseases (Badenhoop et al. 1995; Spurkland et al. 1991), and viral resistance (Huang et al. 2019). Interestingly, *HLA-DQB1* showed a strong signature of balancing selection in regions where Celiac Disease is common (Sams and Hawks 2014; Solberg et al. 2008), indicating its involvement in the response to pathogens, potentially offsetting the negative selective pressures from Celiac Disease (Hale 2017). *SORBS1* is associated with obesity and the development of type 1 and 2 diabetes (Germain et al. 2015; Lin et al. 2001), as well as reduced Tuberculosis bacterial growth (Agrawal et al. 2016). *LEPR* encoding the leptin receptor is another gene of interest, due to its association with obesity-related outcomes in modern individuals (Clement et al. 1998; Farooqi et al. 2007), as well as its involvement in the immune system (De Rosa et al. 2015). Importantly, *LEPR* has been reported as a selection candidate using both ancient and modern data (Colbran et al. 2021; Voight et al. 2006).

CONCLUSIONS

On the basis of comparative analyses of genome-wide genetic variation at ~1.24M SNPs in EEF, we find similar ancestry proportions at Derenburg Meerenstieg II, as we see in other LBK sites across Germany and the broader LBK distribution zone (Hofmanova et al. 2016; Lipson et al. 2017; Rivollat et al. 2020). Overall, we find that LBK individuals closely resemble western Anatolian early Neolithic farmers, suggesting a rapid population expansion from western Anatolia to Central Europe. Moreover, the results from our LBK individuals suggest that there was no, or very limited, admixture between the local European HG and the incoming farmers during the pioneer phase of the early Neolithic in Central Europe. We show that all early farmers belonged to a relatively large source population compared to the local HGs, and found no signal of recent inbreeding, which means that larger population sizes were initially maintained. The relative genetic similarities between early farmers allowed us to group them together, leveraging the increased cohort size for selection scans. Despite there being a limited difference in time between the LBK and Anatolian Neolithic individuals, and potential ongoing gene flow from Anatolia and southeastern Europe, the overlapping results of the three selection scans we explored and of the HLA class-I and II analysis suggest that Early Neolithic individuals in central Europe were experiencing an ongoing metabolic and immune function adaptation to the Neolithic lifeways as a result of the increased disease burden in group sizes that exceeded those of hunter-gatherers. Moreover, we see an adaptation to higher latitudes in Neolithic individuals compared to modern Africans and HGs, as evidenced by allele frequency changes of variants associated with pigmentation and folate synthesis. The Neolithization has been considered one of the strongest natural selective forces acting on the human genome. Our results show that there are multiple signals of selection that can be observed in the LBK and the

early farmers in general in association with the transition to an agricultural lifestyle and sedentism. However, our findings also highlight the challenges of studying selection using ancient DNA, especially in light of complex admixture or demographic histories and limited time that elapsed since the split of populations under study. We also stress the need for increased cohort sizes and refinement of statistical tools that can take these confounding factors into account.

MATERIALS AND METHODS

Archaeological background/site description. Located in the shadow of the Harz (Hercynian) mountain range, the amount of precipitation in the area around DER in the Mittelelbe-Saale region is just under 530 nm/a and the loess-shaped slope area is covered by thick layers of topsoil, in some cases up to 1 meter, which has favored the preservation of skeletal material. The archaeological site was excavated between 1997-1999, and represents a continuous settlement from the Early (Linear Pottery Culture, LBK) and Middle Neolithic (Rössen, Ammensleben) to Bronze and Iron Age (Fritsch et al. 2010). The LBK-associated graveyard represents a closed find and encompassed 43 graves, as well as two separate graves outside the graveyard. A total of 32 single grave, one double, one triple burial, two burials in settlement pits, two burials with secondary inhumations, and one empty grave were found. The majority of DER individuals were buried in East-West orientation in a varying flexed position, which was a common practice in the Neolithic. Slightly more than half of the skeletons exhibited a very good to moderate degrees of preservation (grade 1-3 = 100-50% preserved from skeleton), while other burials were be poorly (grade 4 = 25-50%) or very poorly preserved (grade 5 = < 25%) (Nicklisch 2017).

Sample preparation. Samples were processed using the following aDNA extraction protocol (Velsko et al. 2020), with the UDG-half treatment. Samples (N=37) were first shotgun sequenced followed by the 1240k capture on the samples (N=32) with more than 0.1% endogenous DNA for select SNPs (Sup. Table 1). The raw fastq files were processed and mapped to the human reference hs37d5 with default parameters from EAGER (Peltzer et al. 2016). Minimum mapping and base qualities of 30 were used.

Contamination and ancient DNA authentication. aDNA status was authenticated using MapDamage2.0 (Jonsson et al. 2013). Two base pairs were trimmed from the UDG-half bam files following the aDNA damage pattern using trimBam as implemented in BamUtil 1.0.13 (Jun et al. 2015). X-chromosome contamination estimation in males was performed using ANGSD (Korneliussen et al. 2014) following the software manual. First step: ‘angsd -r X:5000000-

154900000 -doCounts 1 -iCounts 1 -minMapQ 30 -minQ 30', second step: 'contamination -a angsdCounts -h HapMapChrX.gz 2> out'. Contamination estimates from samples with more than ~200 SNPs on the X chromosome covered were considered as reliable (Sup. Table 1). Mitochondrial DNA contamination was estimated using Schmutzi (Renaud et al. 2015) and contamMix (Fu et al. 2013).

Genotyping. Pseudohaploid genotyping was performed using SAMtools mpileup (Li and Durbin 2009) and PileupCaller (https://github.com/stschiff/sequenceTools/blob/master/src/SequenceTools/PileupCaller.hs) for the 1240k sites. Pseudohaploid genotypes were merged with the Human Origin Affymetrix dataset of modern individuals around the globe (Patterson et al. 2012), as well as published ancient West Eurasian individuals (https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data, version 44.3).

SmartPCA from EIGENSOFT v7.2.1 (http://www.hsph.harvard.edu/alkes-price/software/) was used to perform a PCA projection of ancient samples on top of the variation of the modern data. Unsupervised admixture was performed on k components from 1 to 17 with random seeding after LD pruning in plink with the following settings (-indep-pairwise 200 25 0.4). Cross-validation (CV) values were compared, and the k component 11 with the lowest CV error was selected.

F-statistics. R package *admixr* was used to perform f-statistics and *qpAdm* (Petr et al. 2019). We performed f4 tests to determine if there were differences in HG ancestry in the individuals from DER compared to other EEF groups, as well as to test whether LBK individuals had additional HG ancestry compared to individuals from Anatolia Neolithic (Bar8 and Bar31 (Hofmanova et al. 2016)). Based on the admixture analysis, we formally tested the admixture components using *qpAdm* as implemented in the *admixr* package (Petr et al. 2019) using a 12 sample outgroup (Mbuti, Papuan, Onge, Han, Karitiana, Mota, Ust-Ishim, Russia_Mal'ta1, Vestonice, CHG, Israel_Natufian, Villabruna) + additional European HG sources (Villabruna, Russia_EHG, GoyetQ2). The admixture event between WHG and Anatolian farmers was determined using DATES v. 753 (Moorjani et al. 2016).

Biological relatedness and Runs of Homozygosity analyses. READ (Monroy Kuhn et al. 2018) and lcMLkin (Lipatov et al. 2015), were used to perform tests of biological relatedness. ROH analysis was performed using HapROH on samples with more than 600k SNPs based on the 1240k capture (Ringbauer et al. 2020).

Mitochondrial and Y haplotype analysis. To determine mitochondrial DNA haplogroups, bam files were mapped to the mtDNA rCRS sequence (Andrews et al. 1999). Haplogrep 2.0 was

used to assign mitochondrial haplogroups. Y-haplotype analysis was performed using the method described in (Rohrlach et al. 2021).

Shotgun data. Three individuals (DER002, DER009, BDB001) were shotgun sequenced (PE75) to high coverage.

HLA typing. A development version of OptiType 2 (Szolek et al. 2014) was used to determine HLA Class I and II alleles from ≥30bp single-end sequencing data retrieved from the immunocapture libraries (https://github.com/FRED-2/OptiType, tag DER). Fifteen out of 432 allele calls were overruled in favor of runner-up alleles based on anomalous coverage patterns induced by short reads cross-mapping with alleles of different loci.

Selection scans. LSBL (locus-specific branch length) was used on the 1240k genotype data to determine sites in the genome that have undergone recent selective sweeps. This method relies on pair-wise Fst values for three populations to identify loci with extreme Fst difference between population of interest and two other groups used for comparison. Resulting LSBL values were ranked and empirical p-values were determined. Resulting SNPs were annotated using the R package BioMart (Durinck et al. 2009).

We also performed a pathway enrichment analysis using gowinda (Kofler and Schlotterer 2012) comparing the top 1% LSBL hits to all loci analyzed (1240k SNPs after QC). This method performs gene-set enrichment base on Gene Ontology (GO) classifications, and designed specifically for GWA studies. Gowinda was designed to take into account the assumptions that are often violated in genome-wide associations studies, such as longer genes having more SNPs, and overlapping genes being sampled in clusters (Kofler and Schlotterer 2012). Gowinda performs internal permutations by randomly sampling SNPs from the total set of SNPs. By recording the associated genes and performing the permutation multiple times, gowinda establishes an empirical null distribution, and the significant overrepresentation of candidate SNPs is determined compared to the empirical null distribution (Kofler and Schlotterer 2012).

AIMLESS Admixture-Informed Maximum-Likelihood Estimation for Selection Scans.

For site i on the genome, we calculate the allele frequency for two source populations and a target population, denoted $f_i^{S_1}$, $f_i^{S_2}$ and f_i^T, respectively. We then calculate the optimal value of the mixing parameter $\hat{\gamma}_i \epsilon [0,1]$ such that we minimize the value of the function

$$f\left(\gamma, f_i^T, f_i^{S_1}, f_i^{S_2}\right) = \left(f_i^T - \left[\gamma f_i^{S_1} + (1-\gamma) f_i^{S_2}\right]\right)^2,$$

and define the "constrained allele frequency estimator to be $\hat{f}_i^T = \hat{\gamma}_i f_i^{S_1} + (1-\hat{\gamma}_i) f_i^{S_2}$. Note that if $f_i^{S_1} = f_i^{S_2}$, then we simply let $\hat{\gamma}_i = 1$.

We then calculate log-likelihood values for the unconstrained model and the constrained model assuming that $X_i \sim B(N_i, \hat{f}_i^T)$ for the constrained model, and that $X_i \sim B(N_i, f_i)$ for the unconstrained model, denoted l_1 and l_0. Naturally, as f_i is the maximum-likelihood estimator for a binomial observation, $l_0 \geq l_1$.

Finally, we perform a likelihood ratio test (LRT), with test statistic $\Lambda_i = -2(l_1 - l_0)$, where $\Lambda_i \sim X_2^2$. Via a simulation study based on the observed allele frequencies in Anatolia_N and WHG populations, we identified a p-value cut-off of 1×10^{-8} that returned a false discovery rate of zero, with the highest positive discovery rate.

Note that we filter sites for which we have observed allele counts from at least 10 individuals from each population, and for which the allele frequency in the source populations is greater than zero.

Haplotype-based scan. Genome-wide 1240k capture data from Derenburg, LBK and Iron Gates HG individuals were imputed using GLIMPSE following the standard procedures (Rubinacci et al. 2021). Genotype likelihoods were computed using bcftools mpileup (Li 2011). The GLIMPSE_phase function was used to run imputation in chunks with window size of 2,000,000 bps and buffer size of 200000 bps. The chunks were then ligated using GLIMPSE_ligate and haplotypes were sampled using GLIMPSE_sample. The 1000G reference panel was used for imputation (Genomes Project et al. 2015). Samples with endogenous DNA coverage on the 1240k capture of above 0.5x were included in the analysis. REHH package was used to run haplotype-based selection scans (Gautier and Vitalis 2012).

Data Availability. Genomic data (BAM and fastq format) are available at the European Nucleotide Archive under accession number PRJEB52488, genotypes in eigenstrat format can be found at https://edmond.mpdl.mpg.de/dataset.xhtml?persistentId=doi:10.17617/3.HOKI5I The code for AIMLESS is available at https://edmond.mpdl.mpg.de/dataset.xhtml?persistentId=doi:10.17617/3.HOKI5I.

ACKNOWLEDGEMENTS

We thank the members of the Archaeogenetics Department of the Max Planck Institute for the Science of Human History, especially Ellen Späth for sampling, the Population Genetics group for helpful feedback, and Guido Brandt, Antje Wissgott and Stephen Clayton for sequencing and handling of the raw data. We are indebted to State Office for Heritage Management and

Archaeology Saxony-Anhalt – State Museum of Prehistory, Germany, and all archaeologists and students involved in the excavations at Derenburg-Meerenstieg II. We would also like to thank Iain Mathieson, Gaspard Kerner and Vanessa Villalba-Mouco for helpful discussions during the preparation of the manuscript.

FUNDING

The study was funded by the Max Planck Society, and the French (ANR) and German (DFG) Research Foundations, INTERACT project, ANR-17-FRAL-0010, DFG-HA-5407/4-1, 2018-2021 (to M.-F.D., W.H.), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement no. 771234 – PALEoRIDER (to W.H.).

AUTHOR CONTRIBUTIONS

Conceptualization: W.H., A.C.; Data generation: F.A., R.B.; Data curation: W.H., A.C.; Formal analysis: A.C., A.B.R.; Funding acquisition: W.H., M.-F.D.; Investigation: A.C., W.H.; Writing – Original Draft Preparation: A.C., W.H.; Writing – Review & Editing: A.C., W.H., A.B.R.

TABLE 1. SNPs associated with pigmentation and other phenotypes of interest.

Gene	Phenotype	SNP	REF	ALT	EFF	LBK	Anatolia	WHG	AFR	EAS	EUR
FADS1/FADS2	Fatty Acid Metabolism	rs174546	C	T	T	0.70	0.71	0.97	0.02	0.57	0.35
TYR	Light Skin (WE)	rs10831496	A	G	A	0.28	0.42	0.13	0.17	0.25	0.68
TYR	Light Skin (WE)	rs1042602	A	C	A	0.17	0	0	0.01	0.00	0.37
OCA2	Light Skin (EE)	rs1800414	C	T	C	0	0	0	0.01	0.60	0
OCA2	Light Skin (WE)	rs1800404	C	T	T	0.73	0.85	0.83	0.13	0.39	0.79
OCA2	Blue Eyes (WE)	rs12913832	G	A	G	0.42	0.30	0.55	0.03	0.00	0.64
APBA2	Light Skin (WE)	rs4424881	C	T	C	0.90	0.95	0.71	0.10	0.45	0.87
SLC24A5	Light Skin (WE)	rs1426654	A	G	A	0.89	1	0.47	0.07	0.01	1.00
SLC24A5	Blue Eyes (WE)	rs2470102	A	G	A	0.90	1	0.44	0.07	0.25	0.99
MC1R	Light Skin (EE)	rs2228479	A	G	A	0.03	0	0.08	0.00	0.29	0.07
MHC	Immunity	rs2269424	A	G	A	0.73	0.89	0.68	0.01	0.14	0.26

FIGURES

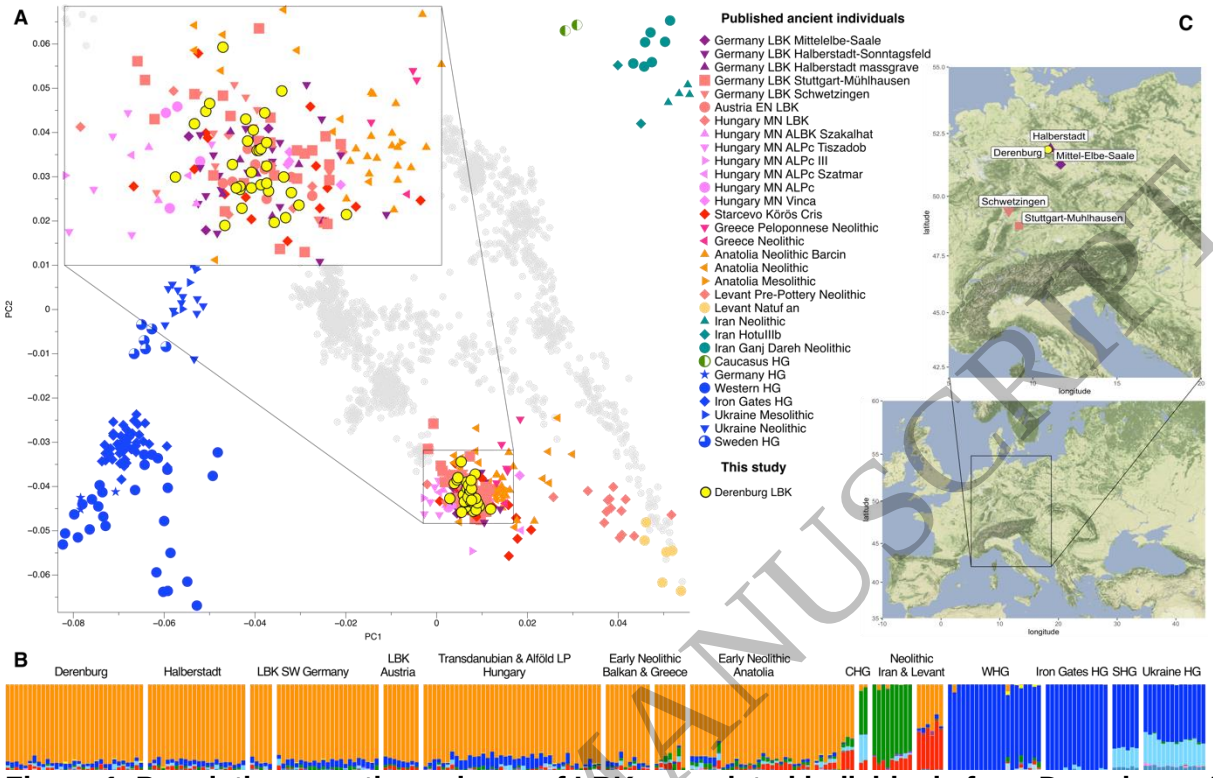

Figure 1. Population genetic analyses of LBK-associated individuals from Derenburg. A) Principal component analysis of DER individuals and relevant published ancient groups, projected on West Eurasian genetic variation. B) Unsupervised ADMIXTURE results (k=12) of a subset of individuals shown in panel A. C) Map of Europe and zoom in on LBK sites in today's Germany.

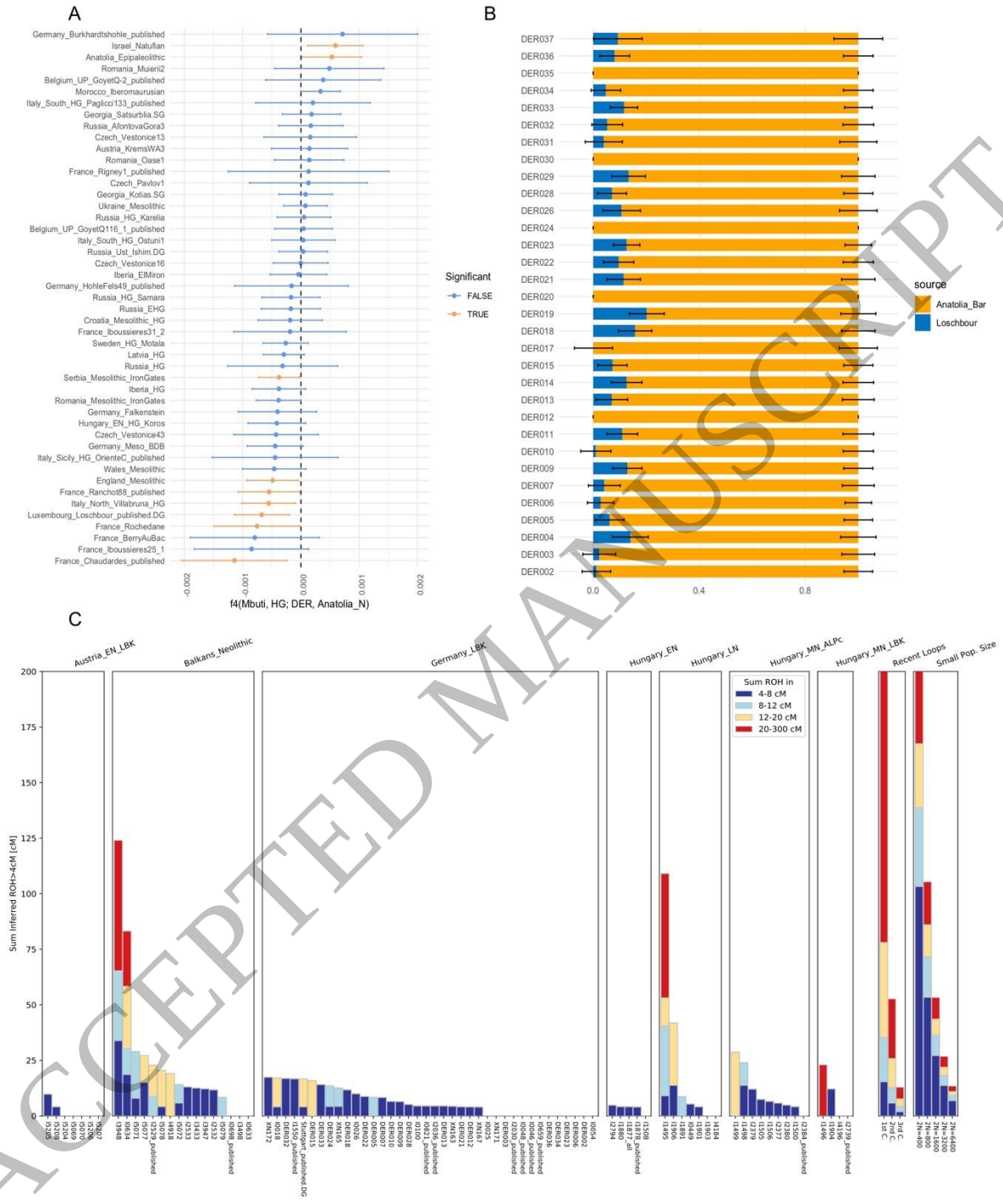

Figure 2. Hunter-Gatherer admixture modeling. (A) F4 analysis of the form *f4*(Mbuti, European_HG; DER, Anatolia_N). Each hunter-gatherer individual tested is represented on the y-axis. Tests with z-scores > 3 are shown in orange compared to rest in blue; (B) *qpAdm* test results for DER individuals using the Loschbour HG and Anatolia_N_Barcin individuals as admixture sources. All models have p-values greater than 0.05, which indicates models that cannot be rejected formally in case of *qpAdm*; (C) HapROH output for German LBK individuals compared to EEF from Hungary, Austria, and the Balkans. Only samples with more than 400k called 1240k SNPs were included.

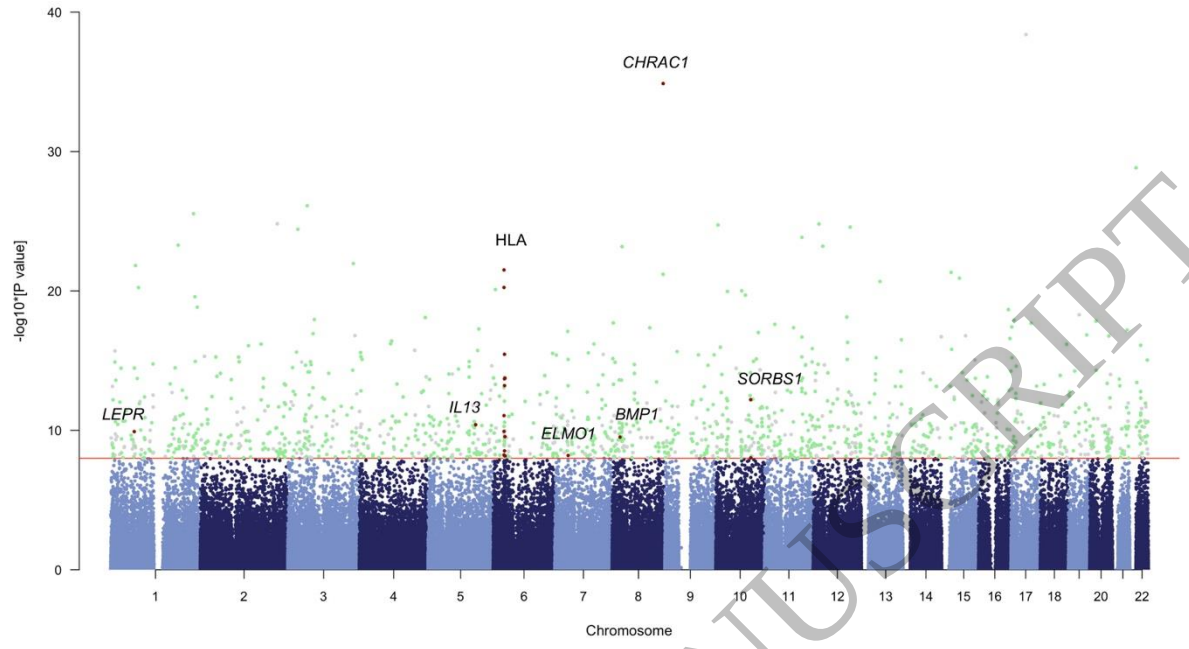

Figure 3. Manhattan plot of the results of the selection scans. Chromosomes are shown on the x axis, while the -$\log_{10}$(p-values) from AIMLESS are represented on the y axis. (A) Significant loci that are shared between AIMLESS, LSBL (LBK vs Anatolia_N and HG), and XP-EHH are highlighted in green. Loci that do not overlap between the three tests are shown in grey. The loci related to immunity and metabolic pathways are highlighted in red.

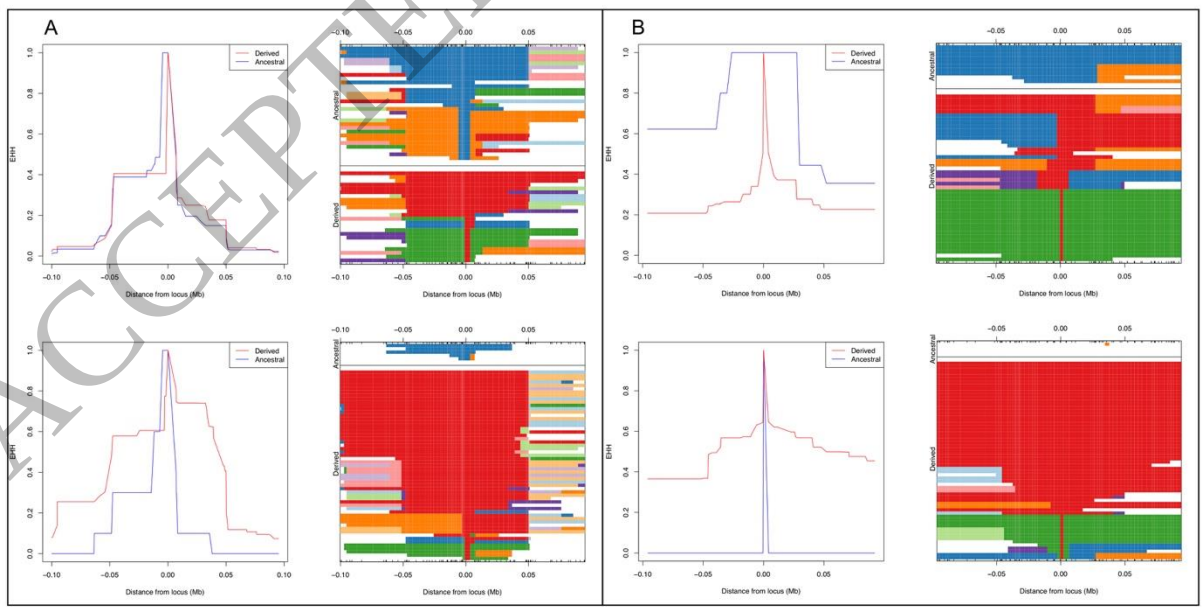

Figure 4. Extended haplotype homozygosity. (A) LMF SNP rs12933840, (B) JAK1 SNP rs3790541. Top panels show the EHH for the ancestral and the derived alleles for HG, and the bottom panels for LBK.

REFERENCES

Agrawal N, Bhattacharyya C, Mukherjee A, Ullah U, Pandit B, Rao KVS, Majumder PP. 2016. Dissecting host factors that regulate the early stages of tuberculosis infection. Tuberculosis. 100:102-113.

Allentoft ME, Sikora M, Sjogren KG, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlstrom T, Vinner L et al. 2015. Population genomics of bronze age eurasia. Nature. 522(7555):167-172.

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the cambridge reference sequence for human mitochondrial DNA. Nat Genet. 23(2):147.

Armelagos GJ, Goodman AH, Jacobs KH. 1991. The origins of agriculture: Population growth during a period of declining health. Population and Environment. 13(1):9-22.

Ash A, Francken M, Pap I, Tvrdy Z, Wahl J, Pinhasi R. 2016. Regional differences in health, diet and weaning patterns amongst the first neolithic farmers of central europe. Sci Rep. 6:29458.

Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, Mead D, Bouman H, Riveros-Mckay F, Kostadima MA et al. 2016. The allelic landscape of human blood cell trait variation and links to common complex disease. Cell. 167(5):1415-1429 e1419.

Badenhoop K, Walfish PG, Rau H, Fischer S, Nicolay A, Bogner U, Schleusener H, Usadel KH. 1995. Susceptibility and resistance alleles of human-leukocyte antigen (hla) dqa1 and hla dqb1 are shared in endocrine autoimmune-disease. J Clin Endocr Metab. 80(7):2112-2117.

Bánffy E. 2019. First farmers of the carpathian basin: Changing patterns in subsistence, ritual and monumental figurines. Oxbow Books.

Barrett R, Kuzawa CW, McDade T, Armelagos GJ. 1998. Emerging and re-emerging infectious diseases: The third epidemiologic transition. Annual review of anthropology. 27(1):247-271.

Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, Chen Y, Felkel S, Hallast P, Kamm J et al. 2020. Insights into human genetic variation and population history from 929 diverse genomes. Science. 367(6484):eaay5012.

Bickle P. 2009. Creating communities: New advances in central european neolithic research. Oxbow Books.

Bigham A, Bauchet M, Pinto D, Mao X, Akey JM, Mei R, Scherer SW, Julian CG, Wilson MJ, Lopez Herraez D et al. 2010. Identifying signatures of natural selection in tibetan and andean populations using dense genome scan data. PLoS Genet. 6(9):e1001116.

Bocquet-Appel J-P, Bar-Yosef O. 2008. The neolithic demographic transition and its consequences. Springer Science & Business Media.

Bogucki P. 1999. The origins of human society. Blackwell Publishers Oxford.

Brace S, Diekmann Y, Booth TJ, van Dorp L, Faltyskova Z, Rohland N, Mallick S, Olalde I, Ferry M, Michel M et al. 2019. Ancient genomes indicate population replacement in early neolithic britain. Nature Ecology & Evolution. 3(5):765-771.

Brandt G, Haak W, Adler CJ, Roth C, Szecsenyi-Nagy A, Karimnia S, Moller-Rieker S, Meller H, Ganslmeier R, Friederich S et al. 2013. Ancient DNA reveals key stages in the formation of central european mitochondrial genetic diversity. Science. 342(6155):257-261.
Brewerton DA, Hart FD, Nicholls A, Caffrey M, James DC, Sturrock RD. 1973. Ankylosing spondylitis and hl-a 27. Lancet. 1(7809):904-907.
Brunel S, Bennett EA, Cardin L, Garraud D, Barrand Emam H, Beylier A, Boulestin B, Chenal F, Ciesielski E, Convertini F et al. 2020. Ancient genomes from present-day france unveil 7,000 years of its demographic history. Proc Natl Acad Sci U S A. 117(23):12791-12798.
Cassidy LM, Martiniano R, Murphy EM, Teasdale MD, Mallory J, Hartwell B, Bradley DG. 2016. Neolithic and bronze age migration to ireland and establishment of the insular atlantic genome. Proceedings of the National Academy of Sciences. 113(2):368.
Cauli A, Shaw J, Giles J, Hatano H, Rysnik O, Payeli S, McHugh K, Dessole G, Porru G, Desogus E et al. 2013. The arthritis-associated hla-b*27:05 allele forms more cell surface b27 dimer and free heavy chain ligands for kir3dl2 than hla-b*27:09. Rheumatology (Oxford). 52(11):1952-1962.
Chamberlain AT. 2006. Demography in archaeology. Cambridge University Press.
Chekalin E, Rubanovich A, Tatarinova TV, Kasianov A, Bender N, Chekalina M, Staub K, Koepke N, Ruhli F, Bruskin S et al. 2019. Changes in biological pathways during 6,000 years of civilization in europe. Molecular biology and evolution. 36(1):127-140.
Chen B, Li J, He C, Li D, Tong W, Zou Y, Xu W. 2017. Role of hla-b27 in the pathogenesis of ankylosing spondylitis (review). Mol Med Rep. 15(4):1943-1951.
Cheng JY, Stern AJ, Racimo F, Nielsen R. 2022. Detecting selection in multiple populations by modeling ancestral admixture components. Molecular biology and evolution. 39(1).
Clement K, Vaisse C, Lahlou N, Cabrol S, Pelloux V, Cassuto D, Gourmelen M, Dina C, Chambaz J, Lacorte JM et al. 1998. A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. Nature. 392(6674):398-401.
Colbran LL, Johnson MR, Mathieson I, Capra JA. 2021. Tracing the evolution of human gene regulation and its association with shifts in environment. Genome Biol Evol. 13(11).
Cordain L. 1999. Cereal grains: Humanity's double-edged sword. World Rev Nutr Diet. 84:19-73.
Dawkins B. 1894. On the relation of the palaeolithic to the neolithic period. The Journal of the Anthropological Institute of Great Britain and Ireland. 23:242-251.
De Rosa V, Galgani M, Santopaolo M, Colamatteo A, Laccetti R, Matarese G. 2015. Nutritional control of immunity: Balancing the metabolic requirements with an appropriate immune function. Semin Immunol. 27(5):300-309.
Deevey ES. 1960. The human population. Scientific American. 203(3):194-205.
Durinck S, Spellman PT, Birney E, Huber W. 2009. Mapping identifiers for the integration of genomic datasets with the r/bioconductor package biomart. Nat Protoc. 4(8):1184-1191.
Ehrhardt N, Bedoya C, Peterfy M. 2014. Embryonic viability, lipase deficiency, hypertriglyceridemia and neonatal lethality in a novel lmf1-deficient mouse model. Nutr Metab (Lond). 11:37.
Farooqi IS, Wangensteen T, Collins S, Kimber W, Matarese G, Keogh JM, Lank E, Bottomley B, Lopez-Fernandez J, Ferraz-Amaro I et al. 2007. Clinical and molecular genetic spectrum of congenital deficiency of the leptin receptor. N Engl J Med. 356(3):237-247.

Feldtkeller E, Khan MA, van der Heijde D, van der Linden S, Braun J. 2003. Age at disease onset and diagnosis delay in hla-b27 negative vs. Positive patients with ankylosing spondylitis. Rheumatol Int. 23(2):61-66.
Fenner JN. 2005. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. American journal of physical anthropology. 128(2):415-423.
Fernandes DM, Sirak KA, Ringbauer H, Sedig J, Rohland N, Cheronet O, Mah M, Mallick S, Olalde I, Culleton BJ et al. 2021. A genetic history of the pre-contact caribbean. Nature. 590(7844):103-110.
Fritsch B, Claßen E, Müller U, Dresely V. 2010. Die linienbandkeramischen gräberfelder von derenburg ‘‘meerenstieg ii’’und halberstadt ‘‘sonntagsfeld’’. Lkr Harz Jahresschr Mitteldt Vorgesch.
Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, Patterson N, Rohland N, Lazaridis I, Nickel B et al. 2015. An early modern human from romania with a recent neanderthal ancestor. Nature. 524(7564):216-219.
Fu Q, Mittnik A, Johnson PLF, Bos K, Lari M, Bollongino R, Sun C, Giemsch L, Schmitz R, Burger J et al. 2013. A revised timescale for human evolution based on ancient mitochondrial genomes. Curr Biol. 23(7):553-559.
Fumagalli M, Pozzoli U, Cagliani R, Comi GP, Bresolin N, Clerici M, Sironi M. 2010. Genome-wide identification of susceptibility alleles for viral infections through a population genetics approach. PLoS Genet. 6(2):e1000849.
Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V, Domboroczki L, Kovari I, Pap I, Anders A et al. 2014. Genome flux and stasis in a five millennium transect of european prehistory. Nat Commun. 5:5257.
Gautier M, Vitalis R. 2012. Rehh: An r package to detect footprints of selection in genome-wide snp data from haplotype structure. Bioinformatics. 28(8):1176-1177.
Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA et al. 2015. A global reference for human genetic variation. Nature. 526(7571):68-74.
Gering E, Incorvaia D, Henriksen R, Conner J, Getty T, Wright D. 2019. Getting back to nature: Feralization in animals and plants. Trends Ecol Evol. 34(12):1137-1151.
Germain M, Pezzolesi MG, Sandholm N, McKnight AJ, Susztak K, Lajer M, Forsblom C, Marre M, Parving HH, Rossing P et al. 2015. Sorbs1 gene, a new candidate for diabetic nephropathy: Results from a multi-stage genome-wide association study in patients with type 1 diabetes. Diabetologia. 58(3):543-548.
Ghoreschi K, Laurence A, O'Shea JJ. 2009. Janus kinases in immune cell signaling. Immunol Rev. 228(1):273-287.
Ginolhac A, Rasmussen M, Gilbert MT, Willerslev E, Orlando L. 2011. Mapdamage: Testing for damage patterns in ancient DNA sequences. Bioinformatics. 27(15):2153-2155.
Gonzalez-Galarza FF, McCabe A, Melo Dos Santos EJ, Jones AR, Middleton D. 2021. A snapshot of human leukocyte antigen (hla) diversity using data from the allele frequency net database. Hum Immunol. 82(7):496-504.
. Beyond the models:Neolithisation'in central europe. PROCEEDINGS-BRITISH ACADEMY; 2007: OXFORD UNIVERSITY PRESS INC.

Gronenborn D. 2014. The persistence of hunting and gathering: Neolithic western temperate and central europe. Oxford University Press. p. 787-804.
Gronenborn D, Strien HC, Dietrich S, Sirocko F. 2014. 'Adaptive cycles' and climate fluctuations: A case study from linear pottery culture in western central europe. Journal of Archaeological Science. 51:73-83.
Grossman SR, Shlyakhter I, Karlsson EK, Byrne EH, Morales S, Frieden G, Hostetter E, Angelino E, Garber M, Zuk O et al. 2010. A composite of multiple signals distinguishes causal variants in regions of positive selection. Science. 327(5967):883-886.
Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, Adler CJ, Der Sarkissian CS, Brandt G, Schwarz C, Nicklisch N et al. 2010. Ancient DNA from european early neolithic farmers reveals their near eastern affinities. PLoS Biol. 8(11):e1000536.
Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardson K et al. 2015. Massive migration from the steppe was a source for indo-european languages in europe. Nature. 522(7555):207-211.
Hale NE. 2017. Are we genetically adapted to the neolithic diet? Journal of Evolution and Health: A joint publication of the Ancestral Health Society and the Society for Evolutionary Medicine and Health. 2(3).
Harper K, Armelagos G. 2010. The changing disease-scape in the third epidemiological transition. International journal of environmental research and public health. 7(2):675-697.
He YY, Wang XC, Jin PK, Zhao B, Fan X. 2009. Complexation of anthracene with folic acid studied by ftir and uv spectroscopies. Spectrochim Acta A Mol Biomol Spectrosc. 72(4):876-879.
Hofmanova Z, Kreutzer S, Hellenthal G, Sell C, Diekmann Y, Diez-Del-Molino D, van Dorp L, Lopez S, Kousathanas A, Link V et al. 2016. Early farmers from across europe directly descended from neolithic aegeans. Proc Natl Acad Sci U S A. 113(25):6886-6891.
Horejs B, Milić B, Ostmann F, Thanheiser U, Weninger B, Galik A. 2015. The aegean in the early 7th millennium bc: Maritime networks and colonization. Journal of World Prehistory. 28(4):289-330.
Huang JT, Xu R, Wang M, Liao Q, Huang K, Shan ZG, You QZ, Li CY, Rong X, Fu YS. 2019. Association of hla-dqb1*03:01 and drb1*11:01 with spontaneous clearance of hepatitis c virus in chinese li ethnicity, an ethnic group genetically distinct from chinese han ethnicity and infected with unique hcv subtype. Journal of Medical Virology. 91(10):1830-1836.
Immel A, Key FM, Szolek A, Barquera R, Robinson MK, Harrison GF, Palmer WH, Spyrou MA, Susat J, Krause-Kyora B et al. 2021. Analysis of genomic DNA from medieval plague victims suggests long-term effect of yersinia pestis on human immunity genes. Molecular biology and evolution.
Immel A, Rinne C, Meadows J, Barquera R, Szolek A, Pierini F, Susat J, Böhme L, Dose J, Bonczarowska J et al. 2019. Neolithic genomes reveal a distinct ancient hla allele pool and population transformation in europe. bioRxiv.
International Genetics of Ankylosing Spondylitis C, Cortes A, Hadler J, Pointon JP, Robinson PC, Karaderi T, Leo P, Cremin K, Pryce K, Harris J et al. 2013. Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. Nat Genet. 45(7):730-738.

Jablonski NG, Chaplin G. 2000. The evolution of human skin coloration. J Hum Evol. 39(1):57-106.
Jablonski NG, Chaplin G. 2010. Colloquium paper: Human skin pigmentation as an adaptation to uv radiation. Proc Natl Acad Sci U S A. 107 Suppl 2:8962-8968.
Johnson AW, Earle TK. 2000. The evolution of human societies: From foraging group to agrarian state. Stanford University Press.
Jonsson H, Ginolhac A, Schubert M, Johnson PL, Orlando L. 2013. Mapdamage2.0: Fast approximate bayesian estimates of ancient DNA damage parameters. Bioinformatics. 29(13):1682-1684.
Ju D, Mathieson I. 2021. The evolution of skin pigmentation-associated variation in west eurasia. Proc Natl Acad Sci U S A. 118(1).
Jun G, Wing MK, Abecasis GR, Kang HM. 2015. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. Genome Res. 25(6):918-925.
Karnes JH, Bastarache L, Shaffer CM, Gaudieri S, Xu Y, Glazer AM, Mosley JD, Zhao S, Raychaudhuri S, Mallal S et al. 2017. Phenome-wide scanning identifies multiple diseases and disease severity phenotypes associated with hla variants. Sci Transl Med. 9(389).
Key FM, Posth C, Esquivel-Gomez LR, Hubler R, Spyrou MA, Neumann GU, Furtwangler A, Sabin S, Burri M, Wissgott A et al. 2020. Emergence of human-adapted salmonella enterica is linked to the neolithization process. Nat Ecol Evol. 4(3):324-333.
Kichaev G, Bhatia G, Loh PR, Gazal S, Burch K, Freund MK, Schoech A, Pasaniuc B, Price AL. 2019. Leveraging polygenic functional enrichment to improve gwas power. Am J Hum Genet. 104(1):65-75.
Kirino Y, Bertsias G, Ishigatsubo Y, Mizuki N, Tugal-Tutkun I, Seyahi E, Ozyazgan Y, Sacli FS, Erer B, Inoko H et al. 2013. Genome-wide association analysis identifies new susceptibility loci for behcet's disease and epistasis between hla-b*51 and erap1. Nat Genet. 45(2):202-207.
Kofler R, Schlotterer C. 2012. Gowinda: Unbiased analysis of gene set enrichment for genome-wide association studies. Bioinformatics. 28(15):2084-2085.
Kolahdooz F, Barr A, Roache C, Sheehy T, Corriveau A, Sharma S. 2013. Dietary adequacy of vitamin d and calcium among inuit and inuvialuit women of child-bearing age in arctic canada: A growing concern. PLoS One. 8(11):e78987.
Korneliussen TS, Albrechtsen A, Nielsen R. 2014. Angsd: Analysis of next generation sequencing data. BMC Bioinformatics. 15:356.
Lacan M, Keyser C, Ricaut FX, Brucato N, Duranthon F, Guilaine J, Crubezy E, Ludes B. 2011. Ancient DNA reveals male diffusion through the neolithic mediterranean route. Proc Natl Acad Sci U S A. 108(24):9788-9791.
Lamason RL, Mohideen MA, Mest JR, Wong AC, Norton HL, Aros MC, Jurynec MJ, Mao X, Humphreville VR, Humbert JE et al. 2005. Slc24a5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science. 310(5755):1782-1786.
Lazaridis I, Nadel D, Rollefson G, Merrett DC, 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(7617):419-424.

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M et al. 2014a. Ancient human genomes suggest three ancestral populations for present-day europeans. Nature. 513(7518):409-413.
Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M et al. 2014b. Ancient human genomes suggest three ancestral populations for present-day europeans. Nature. 513(7518):409-413.
Li H. 2011. A statistical framework for snp calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 27(21):2987-2993.
Li H, Durbin R. 2009. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics. 25(14):1754-1760.
Lin WH, Chiu KC, Chang HM, Lee KC, Tai TY, Chuang LM. 2001. Molecular scanning of the human sorbin and sh3-domain-containing-1 (sorbs1) gene: Positive association of the t228a polymorphism with obesity and type 2 diabetes. Human Molecular Genetics. 10(17):1753-1760.
Lipatov M, Sanjeev K, Patro R, Veeramah KR. 2015. Maximum likelihood estimation of biological relatedness from low coverage sequencing data. bioRxiv.023374.
Lipson M, Szecsenyi-Nagy A, Mallick S, Posa A, Stegmar 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(7680):368-372.
Lubbock J. 1866. Address delivered to the section of “primæval antiquities” at the london meeting of the archaeological institute, july, 1866. Archaeological Journal. 23(1):190-211.
Luca F, Perry GH, Di Rienzo A. 2010. Evolutionary adaptations to dietary changes. Annu Rev Nutr. 30:291-314.
Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A et al. 2016. The simons genome diversity project: 300 genomes from 142 diverse populations. Nature. 538(7624):201-206.
Maricic T, Whitten M, Paabo S. 2010. Multiplexed DNA sequence capture of mitochondrial genomes using pcr products. PLoS One. 5(11):e14004.
Martiniano R, Cassidy LM, O'Maolduin R, McLaughlin R, Silva NM, Manco L, Fidalgo D, Pereira T, Coelho MJ, Serra M et al. 2017. The population genomics of archaeological transition in west iberia: Investigation of ancient substructure using imputation and haplotype-based methods. PLoS Genet. 13(7):e1006852.
Mathieson I. 2020. Limited evidence for selection at the fads locus in native american populations. Molecular biology and evolution. 37(7):2029-2033.
Mathieson I, Alpaslan-Roodenberg S, Posth C, Szecsenyi-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(7695):197-203.
Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M et al. 2015. Genome-wide patterns of selection in 230 ancient eurasians. Nature. 528(7583):499-503.
Megiorni F, Pizzuti A. 2012. Hla-dqa1 and hla-dqb1 in celiac disease predisposition: Practical implications of the hla molecular typing. J Biomed Sci. 19:88.

Mele M, Javed A, Pybus M, Zalloua P, Haber M, Comas D, Netea MG, Balanovsky O, Balanovska E, Jin L et al. 2012. Recombination gives a new insight in the effective population size and the history of the old world human populations. Molecular biology and evolution. 29(1):25-30.
Mendoza-Revilla J, Chacón-Duque JC, Fuentes-Guajardo M, Ormond L, Wang K, Hurtado M, Villegas V, Granja V, Acuña-Alonzo V, Jaramillo C et al. 2021. Disentangling signatures of selection before and after european colonization in latin americans. bioRxiv.2021.2011.2015.467418.
Monroy Kuhn JM, Jakobsson M, Gunther T. 2018. Estimating genetic kin relationships in prehistoric populations. PLoS One. 13(4):e0195491.
Moorjani P, Sankararaman S, Fu Q, Przeworski M, Patterson N, Reich D. 2016. A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. Proc Natl Acad Sci U S A. 113(20):5652-5657.
Münster A, Knipper C, Oelze VM, Nicklisch N, Stecher M, Schlenker B, Ganslmeier R, Fragata M, Friederich S, Dresely V et al. 2018. 4000 years of human dietary evolution in central germany, from the first farmers to the first elites. PLOS ONE. 13(3):e0194862.
Nicklisch N. 2017. Spurensuche am skelett: Paläödemografische und epidemiologische untersuchungen an neolithischen und frühbronzezeitlichen bestattungen aus dem mittelelbe-saale-gebiet im kontext populationsdynamischer prozesse; kulturwandel. Landesamt für Denkmalpflege und Archäologie Sachsen-Anhalt, Landesmuseum für ….
Nicklisch N, Ganslmeier R, Siebert A, Friederich S, Meller H, Alt KW. 2016. Holes in teeth – dental caries in neolithic and early bronze age populations in central germany. Annals of Anatomy - Anatomischer Anzeiger. 203:90-99.
Oelze VM, Siebert A, Nicklisch N, Meller H, Dresely V, Alt KW. 2011. Early neolithic diet and animal husbandry: Stable isotope evidence from three linearbandkeramik (lbk) sites in central germany. Journal of Archaeological Science. 38(2):270-279.
Olalde I, Schroeder H, Sandoval-Velasco M, Vinner L, Lobon I, Ramirez O, Civit S, Garcia Borja P, Salazar-Garcia DC, Talamo S et al. 2015. A common genetic origin for early farmers from mediterranean cardial and central european lbk cultures. Molecular biology and evolution. 32(12):3132-3142.
Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. Genetics. 192(3):1065-1093.
Peltzer A, Jager G, Herbig A, Seitz A, Kniep C, Krause J, Nieselt K. 2016. Eager: Efficient ancient genome reconstruction. Genome Biol. 17:60.
Peterfy M. 2012. Lipase maturation factor 1: A lipase chaperone involved in lipid metabolism. Biochim Biophys Acta. 1821(5):790-794.
Peterfy M, Ben-Zeev O, Mao HZ, Weissglas-Volkov D, Aouizerat BE, Pullinger CR, Frost PH, Kane JP, Malloy MJ, Reue K et al. 2007. Mutations in lmf1 cause combined lipase deficiency and severe hypertriglyceridemia. Nat Genet. 39(12):1483-1487.
Petr M, Vernot B, Kelso J. 2019. Admixr-r package for reproducible analyses using admixtools. Bioinformatics. 35(17):3194-3195.
Petrasch J. 2020. Settlements, migration and the break of tradition: The settlement patterns of the earliest bandkeramik and the lbk and the formation of a neolithic lifestyle in western central europe. Quaternary International.

Pilloud MA, Larsen CS. 2011. “Official” and “practical” kin: Inferring social and community structure from dental phenotype at neolithic çatalhöyük, turkey. American journal of physical anthropology. 145(4):519-530.
Renaud G, Slon V, Duggan AT, Kelso J. 2015. Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biol. 16:224.
Reveille JD, Sims AM, Danoy P, Evans DM, Leo P, Pointon JJ, Jin R, Zhou X, Bradbury LA, Appleton LH et al. 2010. Genome-wide association study of ankylosing spondylitis identifies non-mhc susceptibility loci. Nat Genet. 42(2):123-127.
Ringbauer H, Novembre J, Steinrucken M. 2021. Parental relatedness through time revealed by runs of homozygosity in ancient DNA. Nat Commun. 12(1):5425.
Ringbauer H, Novembre J, Steinrücken M. 2020. Human parental relatedness through time - detecting runs of homozygosity in ancient DNA. bioRxiv.2020.2005.2031.126912.
Rivollat M, Jeong C, Schiffels S, Kucukkalipci I, Pemonge MH, Rohrlach AB, Alt KW, Binder D, Friederich S, Ghesquiere E et al. 2020. Ancient genome-wide DNA from france highlights the complexity of interactions between mesolithic hunter-gatherers and neolithic farmers. Sci Adv. 6(22):eaaz5344.
Rohrlach AB, Papac L, Childebayeva A, Rivollat M, Villalba-Mouco V, Neumann GU, Penske S, Skourtanioti E, van de Loosdrecht M, Akar M et al. 2021. Using y-chromosome capture enrichment to resolve haplogroup h2 shows new evidence for a two-path neolithic expansion to western europe. Sci Rep. 11(1):15005.
Rubinacci S, Ribeiro DM, Hofmeister RJ, Delaneau O. 2021. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. Nature Genetics. 53(1):120-126.
Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R et al. 2007. Genome-wide detection and characterization of positive selection in human populations. Nature. 449(7164):913-918.
Sams A, Hawks J. 2014. Celiac disease as a model for the evolution of multifactorial disease in humans. Human Biology. 86(1):19-35.
Schaebel LK, Bonefeld-Jorgensen EC, Laurberg P, Vestergaard H, Andersen S. 2015. Vitamin d-rich marine inuit diet and markers of inflammation - a population-based survey in greenland. J Nutr Sci. 4:e40.
Schlosstein L, Terasaki PI, Bluestone R, Pearson CM. 1973. High association of an hl-a antigen, w27, with ankylosing spondylitis. N Engl J Med. 288(14):704-706.
Shriver MD, Kennedy GC, Parra EJ, Lawson HA, Sonpar V, Huang J, Akey JM, Jones KW. 2004. The genomic distribution of population substructure in four populations using 8,525 autosomal snps. Hum Genomics. 1(4):274-286.
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(6185):747.
Skoglund P, Malmström H, Raghavan M, Storå J, Hall P, Willerslev E, Gilbert MTP, Götherström A, Jakobsson M. 2012. Origins and genetic legacy of neolithic farmers and hunter-gatherers in europe. Science. 336(6080):466.

Solberg OD, Mack SJ, Lancastera AK, Single RM, Tsai Y, Sanchez-Mazas A, Thomson G. 2008. Balancing selection and heterogeneity across the classical human leukocyte antigen loci: A meta-analytic review of 497 population studies. Human Immunology. 69(7):443-464.
Spurkland A, Ronningen KS, Vandvik B, Thorsby E, Vartdal F. 1991. Hla-dqa1 and hla-dqb1 genes may jointly determine susceptibility to develop multiple-sclerosis. Human Immunology. 30(1):69-75.
Supek F, Bosnjak M, Skunca N, Smuc T. 2011. Revigo summarizes and visualizes long lists of gene ontology terms. PLoS One. 6(7):e21800.
Szecsenyi-Nagy A, Brandt G, Haak W, Keerl V, Jakucs J, Moller-Rieker S, Kohler K, Mende BG, Oross K, Marton T et al. 2015. Tracing the genetic origin of europe's first farmers reveals insights into their social organization. P Roy Soc B-Biol Sci. 282(1805).
Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. 2014. Optitype: Precision hla typing from next-generation sequencing data. Bioinformatics. 30(23):3310-3316.
Szpiech ZA, Hernandez RD. 2014. Selscan: An efficient multithreaded program to perform ehh-based scans for positive selection. Molecular biology and evolution. 31(10):2824-2827.
Tenesa A, Navarro P, Hayes BJ, Duffy DL, Clarke GM, Goddard ME, Visscher PM. 2007. Recent human effective population size estimated from linkage disequilibrium. Genome Res. 17(4):520-526.
Uinuk-Ool TS, Takezaki N, Derbeneva OA, Volodko NV, Sukernik RI. 2004. Variation of hla class ii genes in the nganasan and ket, two aboriginal siberian populations. Eur J Immunogenet. 31(1):43-51.
van der Linden S, Valkenburg H, Cats A. 1983. The risk of developing ankylosing spondylitis in hla-b27 positive individuals: A family and population study. Br J Rheumatol. 22(4 Suppl 2):18-19.
Velsko I, Skourtanioti E, Brandt G. 2020. Ancient DNA extraction from skeletal material. protocols.io.
Verity DH, Marr JE, Ohno S, Wallace GR, Stanford MR. 1999. Behcet's disease, the silk road and hla-b51: Historical and geographical perspectives. Tissue Antigens. 54(3):213-220.
Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. PLoS Biol. 4(3):e72.
Walker PL, Bathurst RR, Richman R, Gjerdrum T, Andrushko VA. 2009. The causes of porotic hyperostosis and cribra orbitalia: A reappraisal of the iron-deficiency-anemia hypothesis. American journal of physical anthropology. 139(2):109-125.
Witalisz-Siepracka A, Klein K, Prinz D, Leidenfrost N, Schabbauer G, Dohnal A, Sexl V. 2018. Loss of jak1 drives innate immune deficiency. Front Immunol. 9:3108.
Woodrow JC, Eastmond CJ. 1978. Hla b27 and the genetics of ankylosing spondylitis. Ann Rheum Dis. 37(6):504-509.
Wu X, Wang G, Zhang L, Xu H. 2021. Genetics of ankylosing spondylitis-focusing on the ethnic difference between east asia and europe. Front Genet. 12:671682.
Yaka R, Mapelli I, Kaptan D, Dogu A, Chylenski M, Erdal OD, Koptekin D, Vural KB, Bayliss A, Mazzucato C et al. 2021. Variable kinship patterns in neolithic anatolia revealed by ancient genomes. Curr Biol. 31(11):2455-2468 e2418.
Ye K, Gao F, Wang D, Bar-Yosef O, Keinan A. 2017. Dietary adaptation of fads genes in europe varied across time and geography. Nat Ecol Evol. 1:167.

Yokoyama Y, Levin SM, Barnard ND. 2017. Association between plant-based diets and plasma lipids: A systematic review and meta-analysis. Nutrition Reviews. 75(9):683-698.
Zajacova M, Kotrbova-Kozak A, Cerna M. 2016. Hla-drb1, -dqa1 and -dqb1 genotyping of 180 czech individuals from the czech republic pop 3. Hum Immunol. 77(4):365-366.
Zeder MA. 2011. The origins of agriculture in the near east. Current Anthropology. 52(S4):S221-S235.
Zubillaga P, Vidales MC, Zubillaga I, Ormaechea V, Garcia-Urkia N, Vitoria JC. 2002. Hla-dqa1 and hla-dqb1 genetic markers and clinical presentation in celiac disease. J Pediatr Gastroenterol Nutr. 34(5):548-554.
Zvelebil M. 2001. The agricultural transition and the origins of neolithic society in europe. Documenta Praehistorica. 28:1–26.
Zvelebil M. 2004. The many origins of the lbk. LBK Dialogues Studies in the formation of the Linear Pottery Culture Oxford: Archaeopress.183-205.