FADS1 and the timing of human adaptation to agriculture

Sara Mathieson[1] & Iain Mathieson[2*]

[1] Department of Computer Science, Swarthmore College, Swarthmore, PA 19081

[2] Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA 19104

[*] Correspondence to: mathi@pennmedicine.upenn.edu

Abstract

Variation at the *FADS1/FADS2* gene cluster is functionally associated with differences in lipid metabolism and is often hypothesized to reflect adaptation to an agricultural diet. Here, we test the evidence for this relationship using both modern and ancient DNA data. We document pre-out-of-Africa selection for both the derived and ancestral *FADS1* alleles and show that almost all the inhabitants of Europe carried the ancestral allele until the derived allele was introduced approximately 8,500 years ago by Early Neolithic farming populations. However, we also show that it was not under strong selection in these populations. Further, we find that this allele, and other proposed agricultural adaptations including variants at *LCT/MCM6*, *SLC22A4* and *NAT2*, were not strongly selected until the Bronze Age, 2,000-4,000 years ago. Similarly, increased copy number variation at the salivary amylase gene *AMY1* is not linked to the development of agriculture although in this case, the putative adaptation precedes the agricultural transition. Our analysis shows that selection at the *FADS* locus was not tightly linked to the development of agriculture. Further, it suggests that the strongest signals of recent human adaptation may not have been driven by the agricultural transition but by more recent changes in environment or by increased efficiency of selection due to increases in effective population size.

Introduction

Human history has seen a number of major transitions in diet (Luca, et al. 2010). The most recent were the transition to a modern "industrialized" diet and, before that, the transition in many parts of the world to a diet heavily based on the products of agriculture. Even outside these periods, differences in diet based on both culture and food source availability would have been a major aspect of environmental differences between human populations–differences that would likely lead to genetic adaptation. Thus, by identifying and studying the evolution of genetic adaptations to diet, we learn not only about historical changes in diet, but also about the genetic basis of diet-related phenotypic differences among present-day humans.

One previously identified adaptation involves the fatty acid desaturase genes *FADS1* and *FADS2*. These genes encode proteins which catalyze key steps in the ω-3 and ω-6 lipid biosynthesis pathways (Nakamura and Nara 2004). These pathways synthesize long-chain (LC) polyunsaturated fatty acids (PUFA) necessary for cell- and, particularly, neuronal-membrane development from short-chain (SC) PUFA (Darios and Davletov 2006). The evolutionary interaction with diet stems from the fact that different diets contain different ratios of SC- and LC-PUFA. Specifically, diets that are high in meat or marine products contain relatively high LC-PUFA levels, and thus may require lower levels of *FADS1* and *FADS2* activity compared to diets that are high in plant-based fats (Ameur, et al. 2012; Mathias, et al.

2012; Fumagalli, et al. 2015; Kothapalli, et al. 2016; Buckley, et al. 2017; Ye, et al. 2017). As a result of this environmental interaction, these genes have been repeatedly targeted by natural selection.

Most dramatically, a derived haplotype ("haplotype D", following Ameur, et al. (2012)) containing the 3' end of *FADS1* is at very high frequency in present-day African populations and intermediate to high frequency in present-day Eurasians (Ameur, et al. 2012; Mathias, et al. 2012), and experienced ancient positive selection in Africa (Ameur, et al. 2012; Mathias, et al. 2012). The derived haplotype increases expression of *FADS1* (Ameur, et al. 2012) and likely represents an adaptation to a high ratio of SC- to LC-PUFA, *i.e.* to a plant-based diet. Direct evidence from ancient DNA has shown that haplotype D was rare in European hunter-gatherers around 10,000 years before present (BP) but has increased in frequency since then, likely due to selection, to its present-day frequency of ~60% in Europe (Mathieson, et al. 2015). This increase was plausibly associated with the adoption, starting around 8,500 BP, of an agricultural lifestyle and diet that would have a higher SC- to LC-PUFA ratio than the previous hunter-gatherer diet (Mathieson, et al. 2015; Buckley, et al. 2017; Ye, et al. 2017). Interestingly, the Altai Neanderthal genome shares at least a partial version of haplotype D (Buckley, et al. 2017; Harris, et al. 2017), suggesting that the functional variation at this locus may predate the split of Neanderthals and modern humans.

Other haplotypes at the locus have been shown to be under selection in different populations in more recent history. In particular, another haplotype that is common in, but largely restricted to, the Greenlandic Inuit population reduces the activity of *FADS1,* is associated with PUFA levels, and is likely an adaptation to a diet that is extremely high in LC-PUFA from marine sources (Fumagalli, et al. 2015). Conversely, a variant that increases expression of *FADS2* has been selected in South Asian populations – and may be a specific adaptation to a vegetarian diet (Kothapalli, et al. 2016). Reflecting its important role in lipid metabolism, variation at the *FADS* locus also contributes significantly to variation in lipid levels in present-day populations. As well as directly contributing to variation in PUFA levels, SNPs in haplotype D are among the strongest genome-wide association study (GWAS) signals for triglyceride and cholesterol levels (Teslovich, et al. 2010). Thus, the complex evolutionary history of the region is not only informative about ancient human diets, but also directly relevant for understanding the distribution of lipid-related disease risk both within and between populations. We therefore aimed to characterize the history of the region, by combining inference from ancient and modern DNA data, in order to understand the evolutionary basis of this important functional variation and its relationship with changes in diet. We also compared the evolutionary history of the *FADS* locus with the histories of other loci involved in dietary adaptation, to see whether we could detect shared patterns of adaptation.

Results

Haplotype structure at *FADS1*

We began by investigating 300 high-coverage whole-genome sequences from the Simons Genome Diversity Project (SGDP) (Mallick, et al. 2016), as well as three archaic human genomes (Meyer, et al. 2012; Prufer, et al. 2014; Prufer, et al. 2017), to fully describe the fine-scale haplotype structure in the region (Figure 1, Supplementary Figure 1). By clustering haplotypes using a graph-adjacency algorithm (Methods), we defined three nested "derived" haplotypes (Figure 1A, Supplementary Table 1, Supplementary Figure 1). Haplotype D extends over 78kb, is specific to Eurasia, and is comparable to the haplotype D defined by Ameur, et al. (2012). Haplotype C is an 18kb portion of haplotype D that is shared between African and Eurasian populations. Finally, haplotype B is a 15kb region that represents the portion of haplotype D that is shared between modern humans and both the Altai (Prufer, et al. 2014) and Vindija (Prufer, et al. 2017) Neanderthals. We denote the modern human haplotype carrying ancestral alleles at key SNPs as haplotype A (the "ancestral" haplotype), and we also analyze the Denisovan (Meyer, et al. 2012) and Chimp (panTro2) (Chimpanzee Sequencing Analysis Consortium 2005) haplotypes. This analysis also allows us to prioritize likely causal SNPs, which are currently unknown (Buckley, et al. 2017; Ye, et al. 2017). If the derived SNP that was selected in modern humans was also selected in Neanderthals, then it must lie in haplotype B, which is defined by just four SNPs (rs174546, rs174547, rs174554 and rs174562).

Using data from the 1000 Genomes project (1000 Genomes Project Consortium 2015), we show that haplotypes in the haplotype B region fall into two clusters (Figure 1B, Supplementary Figure 2), consistent with previously reported analysis of similar regions (Mathias, et al. 2012; Buckley, et al. 2017; Harris, et al. 2017). One cluster contains Eurasian-ancestry individuals that carry haplotype A (and all individuals with Native American ancestry). The other cluster contains Neanderthals, Eurasians that carry haplotype D, and almost all present-day African-ancestry individuals. A small number of intermediate haplotypes are either recombinant or the result of phasing errors. We inferred the phylogeny of the haplotype B region using BEAST2 (Figure 1C) (Bouckaert, et al. 2014). The region is small and has a very low recombination rate, (average 0.03 cM/mb according to the International HapMap Consortium (2007) combined recombination map), so we ignored possible recombination events. Assuming that the common ancestor of human and chimp haplotypes was 7.5-9.5 million years ago (corresponding to a mutation rate of ~$4\text{-}5\times10^{-10}$ per-base per-year), we infer that the most recent common ancestor (MRCA) of the present-day African and European (D) haplotypes was around 80,000 BP, of haplotype C around

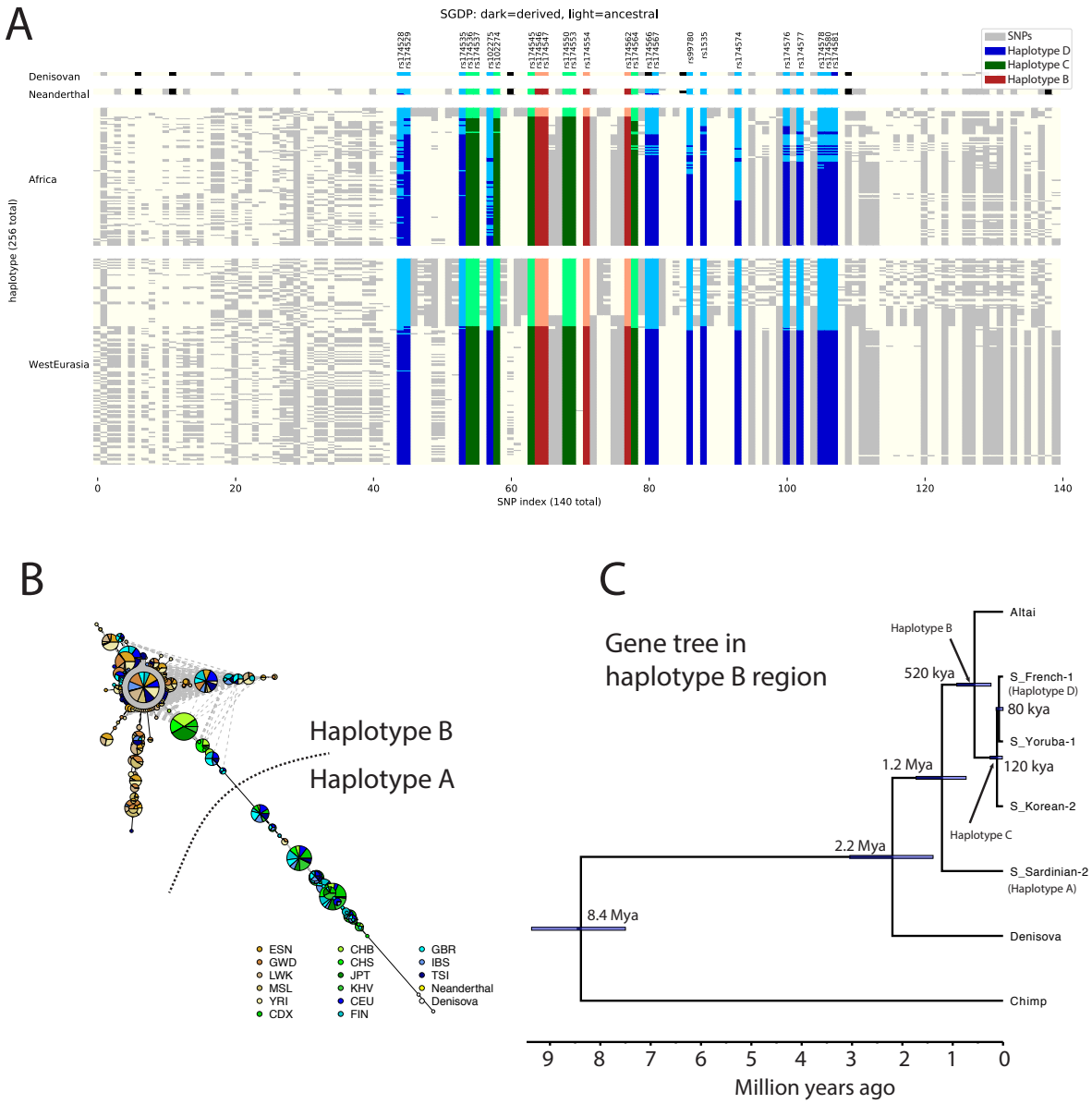


Figure 1: **Haplotype structure at the *FADS1* region**. **A**: Haplotypes from the SGDP African and West Eurasian populations (Mallick, et al. 2016) and the Neanderthal (Prufer, et al. 2014; Prufer, et al. 2017) and Denisovan (Meyer, et al. 2012) genomes. Each column represents a SNP and each row a phased haplotype. Dark and light colors represent derived and ancestral alleles at each SNP and blue, green and red colors indicate SNPs that are part of haplotypes D, C and B, respectively. **B**: Haplotype network for region B constructed from 1000 Genomes (1000 Genomes Project Consortium 2015) and archaic samples. Green, blue and brown indicate East Asian, European and African populations respectively . **C**: Gene tree for the haplotype B region inferred for representative haplotypes.

120,000 BP, of haplotype B around 520,000 BP, and that of haplotypes A and B around 1.2 million BP. These dates are consistent with previous estimates (Harris, et al. 2017) and suggest that the European-specific haplotype D diverged around the time of the out-of-Africa bottleneck and the diversification of non-African lineages (Mallick, et al. 2016). They also suggest that modern human ancestral and derived haplotypes coexisted in the Denisovan-Neanderthal-modern human ancestral population, with the Neanderthal haplotype splitting off around the time of mean human-Neanderthal divergence 550-765,000 BP (Prufer, et al. 2014).

ABC analysis shows ancient selection in Africa and recent selection in Eurasia

Having established the structure and phylogeny of the haplotypes in the region we used approximate Bayesian computation (ABC) (Wegmann, et al. 2010; Peter, et al. 2012) to infer the strength and timing of selection on the derived haplotype in Africa (Table 1, Supplementary Table 2). Using data from the 1000 Genomes project, treating the derived allele of rs174546 as the selected SNP, and fixing the mutation rate to 1.25×10^{-8} per-base per-generation, we infer that selection most likely but not definitively acted on a new mutation (posterior probability 0.55, range 0.25-0.81) and began 136,000-447,000 BP. These estimates are uncertain though, with 95% credible intervals (CrIs) ranging from 44,000-1,251,000 BP. Nonetheless, this analysis suggests that selection began around or before the time of the deepest splits among human populations (Mallick, et al. 2016; Schlebusch, et al. 2017) 200,000-250,000 BP. This is consistent with the observation that the derived allele is shared among all present-day African populations including those, such as Khoe-San and Mbuti, which were substantially diverged from the ancestors of other present-day African populations at least 100,000 BP.

In contrast to this ancient selection in Africa, when we analyze present-day European genomes, we find that strong (selection coefficient s=3.6-5.8%) selection for the derived allele acted on standing variation (posterior probability=1.0) much more recently – starting around 2600 BP (95% CrIs range from 1000-13,000 BP). This recent selection is consistent with published analysis of ancient DNA data (Mathieson, et al. 2015; Buckley, et al. 2017; Ye, et al. 2017) but is hard to reconcile with the suggestion that it is closely linked to the development of agriculture (Mathieson, et al. 2015; Ye, et al. 2017), which dates back at least 10,000 BP (Bellwood 2004). It is also unclear why the derived haplotype was at such low frequency in pre-agricultural Europeans (Mathieson, et al. 2015; Ye, et al. 2017), when ABC results and derived allele sharing between African populations would imply that the derived allele would have been at high frequency before the split of African and non-African ancestral populations. To resolve these questions, we turned to the analysis of ancient DNA data.

	P(New mutation)	Start (years BP)	Start CrI (years BP)
1. Selection for the derived allele in the ancestors of present-day Africans	0.55	246,805	136,389-446,609
2. Selection for the ancestral allele in the ancestors of present-day Europeans	0.04	147,888	58,092-377,659
3. Selection for the derived allele in Europe	0.00	2,664	1,559-4,554

Table 1: **Summary of ABC results for three episodes of selection**. We inferred the probability of selection on a new mutation (as opposed to standing variation), estimated selection-onset time, and 95% credible intervals for this time. See Supplementary Table 1 for more detailed results. Note that for the first episode we infer selection in a population that is ancestral to present-day African populations, but that population is also likely ancestral to present-day European populations, given the timing of the divergence between the two.

Low frequency of the derived *FADS1* allele in Upper Palaeolithic Eurasia

We first investigated why the derived haplotype was at such low frequency in pre-agricultural Europeans, by analyzing data from 46 Upper Palaeolithic and Mesolithic individuals dated between 45,000 and 8,000 BP (Fu, et al. 2016). We infer the presence of haplotype D based on 6 SNPs that were typed on the capture array used in the original study (Fu, et al. 2016). We find that the derived haplotype is rare in all populations (Figure 2A). When we tried to use imputation to increase sample size, the results were inconsistent. Specifically, imputed data suggested a much higher frequency in most of the ancient population groupings, for example around 40% frequency in the Vestonice population. This higher frequency agrees with previous analyses (Ye, et al. 2017). However, imputation is unreliable for these data because many of the Ice Age samples have extremely low coverage. This leads to a high rate of false positive inference of the derived allele because it is at relatively high frequency in the present-day populations used as a reference panel. Older samples tend to have lower coverage and thus higher false positive rates, leading to a spurious inference of a decline in frequency over time. To show this, we estimated the false positive rate as a function of coverage by simulating low coverage data from present-day samples that carry the ancestral allele. Non-zero derived allele frequency estimates in the imputed data are consistent with the estimated false positive rate (Methods, Figure 2B). Therefore, it seems likely that the derived allele was already rare in the "out-of-Africa" population that was ancestral to present-day non-Africans.

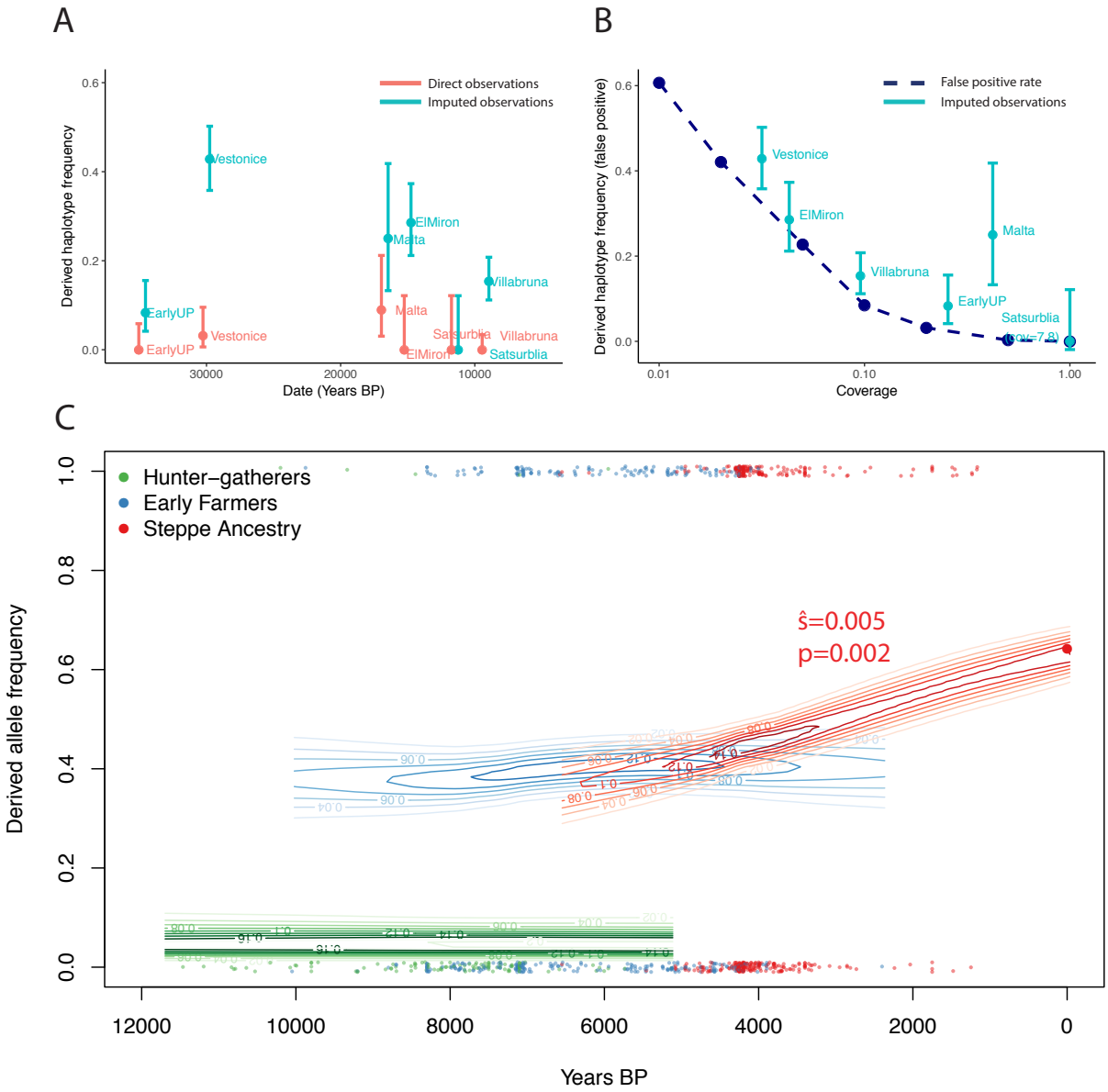


Figure 2: **Direct ancient DNA evidence for the history of *FADS1*. A**: Derived haplotype frequency estimated from direct observation of SNPs on the haplotype (red) and imputed data (blue) in Upper Palaeolithic individuals (45,000-10,000 BP) (Fu, et al. 2016). **B**: Estimated imputation false positive rate as a function of coverage (dashed line). Imputed allele frequencies in Upper Palaeolithic populations plotted for comparison at the median coverage in that population. **C**: Allele frequencies (at rs174546) over the past 12,000 years estimated from 1030 ancient and 99 modern individuals. Each point is an ancient pseudo-haploid individual call, at the bottom of the plot if it is ancestral and the top if it derived. Contours indicate the posterior probability of allele frequencies in the ancient populations and significant selection coefficients are labeled.

This lack of the derived allele in early non-Africans is surprising because our analysis suggests that selection within the ancestral population began at least 246,000 years ago, and the derived allele would therefore be expected to have been at high frequency when African and non-African ancestors diverged. Further, the derived allele is shared between present-day African populations that were genetically isolated before the split of present-day African and non-African ancestors (Mallick, et al. 2016; Schlebusch, et al. 2017). This suggests that there could have been selection for the ancestral allele in the ancestors of Upper Palaeolithic non-Africans. To test this, we used the same ABC approach we used to investigate the derived allele to infer selection on the ancestral allele (Methods, Table 1, Supplementary Table 2). We infer that the ancestral allele was selected from standing variation in the ancestors of present-day Europeans (s=0.2%, 95% CrI=0.1-0.7%), starting around 151,000 BP (95% CrI 58,000-378,000 BP). We also estimated the timing of selection using an independent approach based on haplotype decay (Smith, et al. 2017), confirming ancient selection for the derived allele (average posterior mean 89,000 years with 95% credible intervals ranging from 53,000 to 168,000 years; Supplementary Table 3). This timescale suggests that as the ancestors of Upper Palaeolithic Eurasians diverged from the ancestors of present-day African populations, and certainly before the out-of-Africa bottleneck 40,000-60,000 BP, they experienced differential selection pressures that favored the ancestral rather than the derived allele at *FADS1*. The fixation of the ancestral allele in present-day native Americans has been interpreted as evidence for selection in their Siberian or Beringian ancestors (Amorim, et al. 2017; Harris, et al. 2017; Hlusko, et al. 2018). However, our analysis shows that this may not be the case, because the derived allele was already rare in Eurasia at the time of the Native American-Eurasian split 20-25,000 BP (Raghavan, et al. 2015), and was only reintroduced into Eurasian populations at significant frequency much later.

Selection for the derived *FADS1* allele in Europe was not closely linked to agriculture

The derived haplotype was almost absent in Upper Palaeolithic and Mesolithic Europe (before ~10,000 BP), but is today at a frequency of ~60%. Previous analysis of both ancient and modern DNA has inferred strong selection for the allele over the past 10,000 years (Mathieson, et al. 2015; Field, et al. 2016; Buckley, et al. 2017; Ye, et al. 2017). This has been interpreted to mean that selection for the derived allele was driven by the development of agriculture, around 10,000 BP–a reasonable interpretation since the derived allele is plausibly an adaption to a diet high in plant fats and low in animal fats. However, our ABC analysis suggests that selection in Europe might actually be restricted to the past few thousand years and thus post-date the development of agriculture by many millennia. To test this directly, we analyzed data from 1003 ancient Europeans who lived between 12,000 and 2,000 BP (Keller, et al. 2012; Fu, et al.

2014; Gamba, et al. 2014; Lazaridis, et al. 2014; Olalde, et al. 2014; Raghavan, et al. 2014; Seguin-Orlando, et al. 2014; Allentoft, et al. 2015; Gunther, et al. 2015; Haak, et al. 2015; Jones, et al. 2015; Mathieson, et al. 2015; Cassidy, et al. 2016; Hofmanová, et al. 2016; Martiniano, et al. 2016; Schiffels, et al. 2016; Jones, et al. 2017; Lazaridis, et al. 2017; Lipson, et al. 2017; Mathieson, et al. 2018; Olalde, et al. 2018). We divided these individuals into three populations, based on their genome-wide ancestry (Methods). First, individuals with "hunter-gatherer ancestry" were the Mesolithic inhabitants of Europe or their descendants. Second, "Early farmers" were people from Neolithic Northwest Anatolia or their descendants, possibly admixed with hunter-gatherers, who migrated throughout Europe. Finally, people with "Steppe ancestry" had ancestry that was originally derived from Bronze Age steppe populations like the Yamnaya. (Lazaridis, et al. 2014; Allentoft, et al. 2015; Haak, et al. 2015). Because transitions between these populations involved dramatic genetic discontinuity, with 75-100% of ancestry replaced (Allentoft, et al. 2015; Haak, et al. 2015), we analyzed each of them separately. We estimated the frequency and selection coefficient of the derived allele by fitting a hidden Markov Model to each of the three time series of observations (Methods, Figure 2C) (Mathieson and McVean 2013). We estimate a selection coefficient of 0 in both hunter-gatherer and early farmer populations (i.e. the allele was not under selection), and a selection coefficient of 0.5-0.8% in steppe-ancestry populations, depending on whether we include modern populations or not. This analysis confirms that selection started relatively recently–around 2,000-4,000 BP–and shows that, although the derived allele was present in early farming populations, it was not strongly selected. Plausibly, the derived allele was introduced to the ancestors of the early farmers through admixture with a population that carried "basal Eurasian" ancestry (Lazaridis, et al. 2016) not found in Palaeolithic or Mesolithic Europe.

Patterns of population differentiation at other lipid-associated alleles

SNPs that tag the derived haplotype are among the strongest genome-wide associations with lipid levels (Teslovich, et al. 2010). To test whether the selective events we observe were consequences of more general selection on lipid levels, we investigated patterns of African-European population differentiation between variants associated with three blood lipid traits–triglycerides (TG), high-density cholesterol (HDL) and low-density cholesterol (LDL) (Teslovich, et al. 2010). We find that, for HDL and TG, trait-increasing alleles tend to be more common in African than European populations, while for LDL, the trait increasing allele is more common in European populations (Figure 3A, Methods). These effects are in the opposite direction to those of the *FADS1* haplotype. The derived allele, which is more common in African than European populations, tends to decrease HDL and TG and increase LDL. We conclude that selection on *FADS1* was not driven by its effect on overall blood lipid levels (which, if anything, moved in the

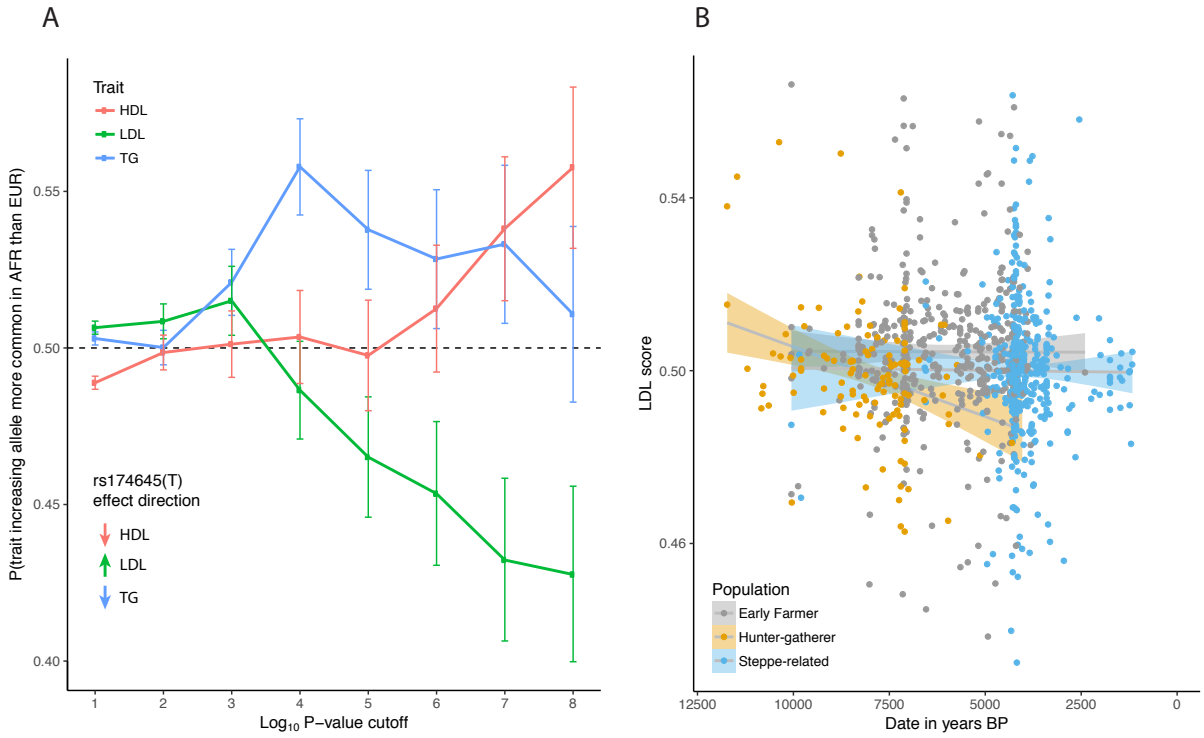


Figure 3: **Trends in lipid-associated allele frequencies A**: Probability that the trait-increasing allele is more common in African than European populations in the 1000 Genomes Project data (1000 Genomes Project Consortium 2015) for lipid-related traits. We show results for each trait with different P-value cutoffs. Vertical bars represent 95% confidence intervals. Inset shows the direction of effect for the derived FADS1 haplotype, using rs174546 as the tag SNP. **B**: LDL score for 1003 ancient individuals, classified according to ancestry, as a function of their age. LDL score is the proportion of significant LDL-associated variants at which the ancient individual carries the trait-increasing allele (Methods).

opposite direction), but by its specific effect on PUFA synthesis. We further find that there is no significant difference in the frequency of LDL-increasing alleles when comparing the three ancient populations (Figure 3B). Since they do differ in the frequency of the *FADS1* allele, this suggests that recent selection on *FADS1* was also not driven by selection more generally on lipid levels.

Patterns of population differentiation at other diet-associated variants

We investigated whether other variants that have been suggested to be associated with the adoption of agriculture showed similar temporal patterns of selection to *FADS1*. The salivary amylase gene *AMY1* is highly copy-number variable among present-day populations, ranging from a diploid copy number of 2 (the ancestral state) to 17 (Groot, et al. 1991; Perry, et al. 2007; Usher, et al. 2015), with a mean of 6.7 copies in present-day Europeans (Usher, et al. 2015). It has been suggested that increased copy number improves the digestion of starchy food and is therefore an adaptation to an agricultural diet that is relatively rich in starch (Perry, et al. 2007). To date, one Early Farmer has been reported to have a relatively high *AMY1* copy number of 16 (Lazaridis, et al. 2014), while six hunter-gatherers had 6-12 copies (Lazaridis, et al. 2014; Olalde, et al. 2014; Gunther, et al. 2018). To investigate this more generally, we called *AMY1* copy number in 76 ancient West Eurasian individuals with published shotgun sequence data (Keller, et al. 2012; Meyer, et al. 2012; Skoglund, et al. 2012; Gamba, et al. 2014; Lazaridis, et al. 2014; Olalde, et al. 2014; Prufer, et al. 2014; Raghavan, et al. 2014; Skoglund, et al. 2014; Fu, et al. 2015; Gunther, et al. 2015; Cassidy, et al. 2016; Kilinc, et al. 2016; Omrak, et al. 2016; Schiffels, et al. 2016; Prufer, et al. 2017; Saag, et al. 2017; Gunther, et al. 2018). We counted the number of reads that mapped to regions around each of the three *AMY1* copies in the human reference genome (Usher, et al. 2015), and compared to bootstrap estimates of mean read depth with a linear correction for GC content (Methods, Figure 4, Supplementary Table 4). We find no significant difference between *AMY1* copy number in present day Europeans and ancient Europeans from the Iron Age/Medieval, Bronze Age, Neolithic, or Mesolithic periods. In particular, Mesolithic pre-agricultural hunter-gatherers have a mean copy number of 7.2 – not statistically different from present-day Europeans. Neolithic Early Farmers have a mean copy number of 8.4 – slightly higher, but not significantly different (after multiple testing correction) from present-day populations ($p=0.02$). We do find that four Upper Palaeolithic individuals dating between ~45,000-20,000 BP have lower *AMY1* copy number than present-day Europeans although, with a small sample size, this is also not statistically significant (mean 3.4, $p=0.02$). These results therefore suggest an expansion in copy number, sometime earlier than ~10,000 BP, but that this expansion predates the development of agriculture.

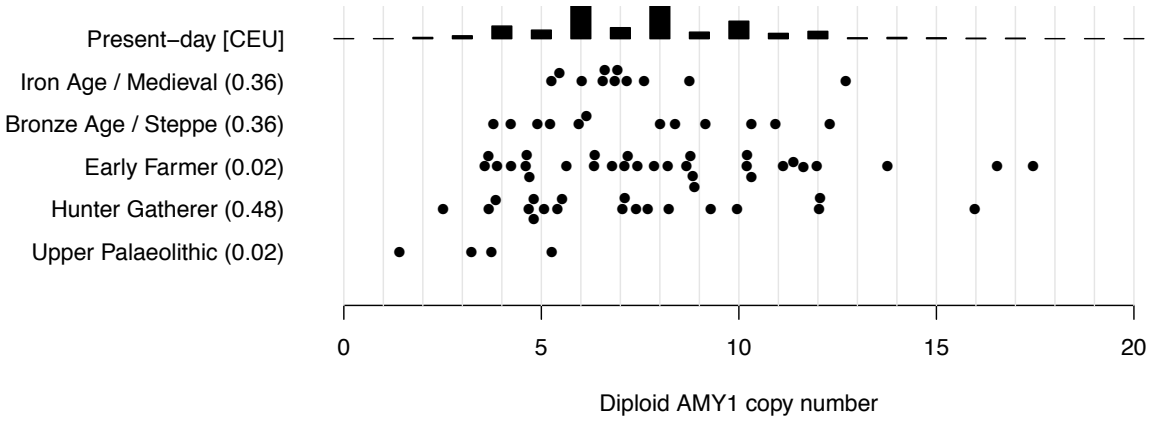


Figure 4: ***AMY1*** **copy number**. Inferred in ancient samples, arranged by time and subsistence strategy, compared with the distribution in a present-day population (CEU) with Northern European ancestry (Usher, et al. 2015). Parentheses: t-test P-values for difference between CEU and ancient populations.

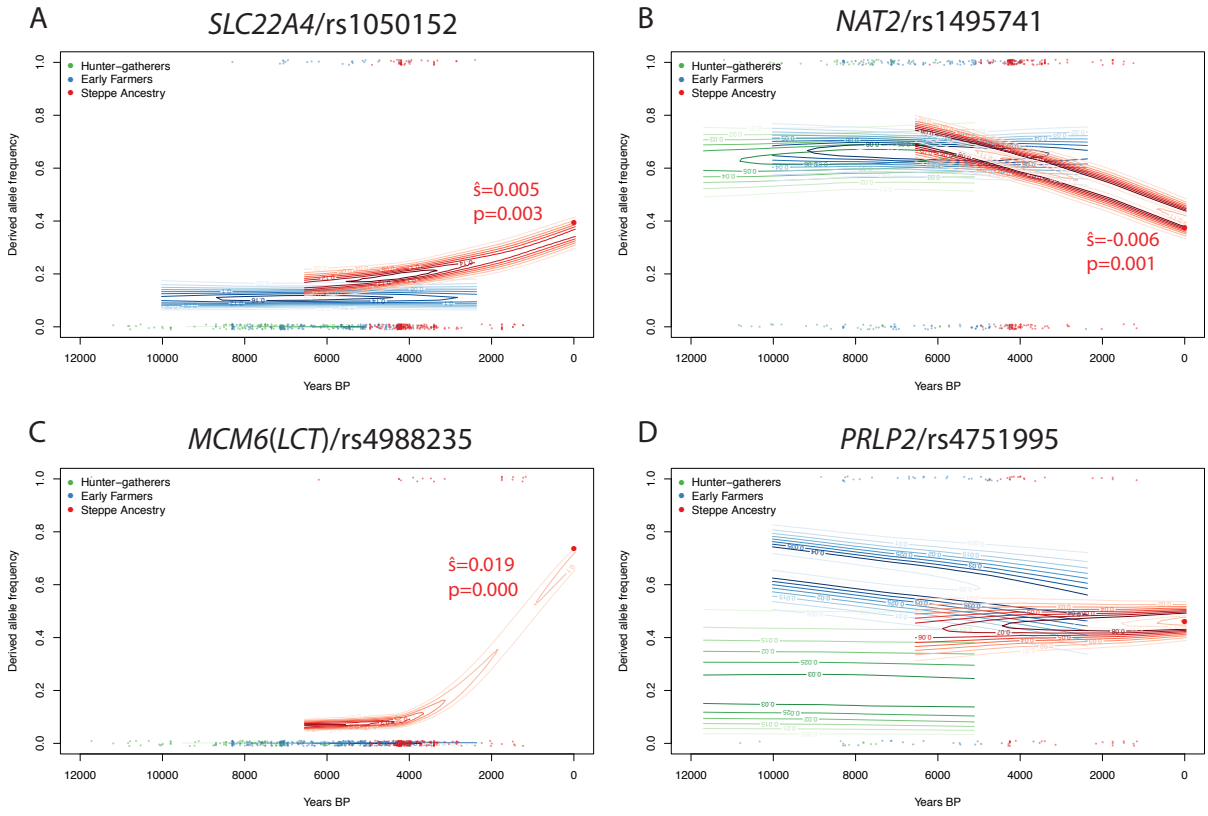


Figure 5: **Allele frequency trajectories for other putative agricultural adaptation variants**. As in Figure 2C, estimated allele frequency trajectories and selection coefficients in different ancient European populations. Significant selection coefficients are labelled.

Finally, we also investigated the history of other mutations that have been suggested to be involved in adaptation to agricultural subsistence. It has been proposed, based on both ancient and modern DNA, that the ergothioneine transporter gene *SLC22A4*–and in particular the nonsynonymous variant 503F (rs1050152)–was targeted by selection in Early Neolithic farming populations (Huff, et al. 2012; Mathieson, et al. 2015). However, analysis of our larger ancient DNA dataset reveals a more complicated story, with an allele frequency trajectory similar to that of the *FADS1* derived allele (Figure 5A). Specifically, the derived allele of rs1050152 is absent in hunter-gatherers, and present at low frequency in Early Farmers and Bronze Age populations. But, similar to the derived *FADS1* allele, it does not increase in frequency–and therefore does not appear to be under selection–in Early Farming populations. Strong selection on this allele can therefore only have operated in the past 2000-4000 years. We find no evidence of selection in early Neolithic farmers, and hypothesize that the allele was introduced, like the derived *FADS1* allele, by basal Eurasian admixture (Lazaridis, et al. 2016) into their ancestors. Similarly, the "slow acetylator" variant of the *NAT2* gene has been suggested to be advantageous in agricultural populations (Luca, et al. 2008; Magalon, et al. 2008; Sabbagh, et al. 2011). We find that a SNP (rs1495741) that tags the "fast acetylator" phenotype (Garcia-Closas, et al. 2011) did not change in frequency between hunter-gatherer and early farmer populations, but did decrease in frequency between the Bronze Age and the present day, i.e. over approximately the past 2000-4000 years (Figure 5B). This timescale is similar to the timescale over which the European lactase persistence variant (Enattah, et al. 2002) became common in Europe (Figure 5C) (Burger, et al. 2007; Allentoft, et al. 2015; Mathieson, et al. 2015). Finally, we find that a variant (rs4751995) in the gene *PLRP2* that is relatively common in present-day populations with a cereal-based diet (Hancock, et al. 2010) is also more common in Early Farmers than hunter-gatherers, although it shows no evidence of selection in any of these populations (Figure 5D).

Discussion

The combination of high-quality genome sequence data from present-day people, and large ancient DNA datasets, provides new opportunities to investigate and dissect the process of human adaptation. In particular, the temporal aspect of ancient DNA data adds another dimension to our analysis, allowing us to make precise inference of the timing of selection. Despite strong ancient selection for the derived *FADS1* allele in Africa, we showed that it was subsequently selected against in the ancestors of Upper Paleolithic West Eurasians, probably before the out-of-Africa bottleneck, to the point that the derived allele was very rare or absent in all European populations, until it was introduced around 8500 BP by migration from early farmers carrying basal Eurasian ancestry. Although the derived allele is plausibly

advantageous in populations that consume a plant-based diet, we find no evidence that it was actually selected in early European Farmers. Conversely, although the low frequency of the derived allele in Upper Palaeolithic Europe is consistent with isotopic evidence that protein intake was dominated by animal protein at this time (Richards 2009), the inferred divergence between the allele frequency in the ancestors of Upper Palaeolithic Europeans and the ancestors of present-day Africans does not appear to correspond to any known dietary transition. Finally, it is not known whether the Neanderthal allele would have had the same function as the derived modern human allele. But if it did, it would not support the claim that the derived allele is associated with a plant-based diet since Neanderthals, like Upper Palaeolithic modern humans, are thought to have consumed a largely meat-based diet (Bocherens 2009). Therefore, the derived/ancestral state at the major *FADS* haplotype does not appear to be a simple marker of plant- vs. meat-based diet or more recent hunter-gatherer vs. farmer subsistence, but likely reflects a more complex pattern of interaction based on unknown dietary and genetic factors.

Due to limited ancient DNA data, we were not able to resolve the history of the *FADS1* allele in East Asia. ABC analysis was very sensitive to the exact demography that we assumed. When we capped recent N_e at 45,000, we found that the ancestral allele was selected in the ancestors of present-day East Asians, although with a large credible interval (7,000-283,000 BP). The 39,000-year-old Tianyuan individual did not carry the derived haplotype, further suggesting that it was absent in the Upper Paleolithic ancestors of East Asians as well as Europeans. We estimated that the derived allele was selected recently, although not as recently as in Europe (i.e. ~16,000 BP in East Asia, with posterior mean estimates ranging from 10,000 to 23,000 BP). The most common East Asian derived haplotype is also an outgroup to the common European and African haplotypes (Figure 1B&C, Supplementary Figure 2), so it may derive from a separate, older, event. More ancient DNA from East Asia will help resolve this question, although we note that agriculture developed later in East Asia than in Western Eurasia, so it is likely that selection on the derived *FADS1* allele was also unassociated with agriculture there as well.

In the case of *FADS1* and all the other examples we investigated, the proposed agricultural adaption was either not temporally linked with agriculture or showed no evidence of selection in agricultural populations. Instead, most of the variants with any evidence of selection were only strongly selected at some point between the Bronze Age and the present day, that is, in a period starting 2000-4000 BP and continuing until the present. This time period is one in which there is relatively limited ancient DNA data, and so we are unable to determine the timing of selection any more accurately. Future research should address the question of why this recent time period saw the most rapid changes in apparently diet-associated genes. One plausible hypothesis is that the change in environment at this time was actually

more dramatic than the earlier change associated with agriculture. Another is that effective population sizes were so small before this time that selection did not operate efficiently on variants with small selection coefficients. For example, analysis of present-day genomes from the United Kingdom suggests that effective population size increased by a factor of 100-1000 in the past 4500 years (Browning and Browning 2015). Ancient effective population sizes less that 10^4 would suggest that those populations would not be able to efficiently select for variants with selection coefficients on the order of 10^{-4} or smaller. Larger ancient DNA datasets from the past 4,000 years will likely resolve this question.

Methods

Identifying and analyzing *FADS1* haplotypes

We defined derived haplotypes using the following procedure. Within the region ±50kb from rs174546 (hg19 chr11:61519830-61619830) there are 140 common (MAF > 0.05) SNPs when considering the SGDP (600 haplotypes) and archaic (6 haplotypes) samples, and restricting to sites where the ancestral allele can be determined based on the chimpanzee genome. For each pair of SNPs within these 140, we compute the number of "mismatches" between ancestral and derived states. For example, if SNP 1 at haplotype *H* is in the ancestral state, but SNP 2 at *H* is in the derived state, then this counts as one mismatch. 1000 Genomes data was used to set the ancestral/derived state for each SNP (we used the chimpanzee allele if the allele was not present in 1000 Genomes). Counting up these mismatches for all pairs of SNPs, we obtain a 140 x 140 symmetric "mismatch" matrix *M*. We transform this matrix into an "adjacency" matrix *A* by setting each entry to 1 if the number of mismatches is below some threshold *t*, and 0 otherwise. In other words, if $M[i,j] <= t$, $A[i,j] = 1$, otherwise $A[i,j] = 0$. This adjacency matrix can then be interpreted as a graph, with SNPs as the nodes and edges between SNPs if they are connected (i.e. have a low number of mismatches).

From this graph, we find the largest clique (connected component where every pair of nodes is connected). This procedure can be interpreted as a way to find a subset of SNPs that are all in high LD with each other. The problem of finding the largest clique in a graph is NP-hard, but we use the Bron-Kerbosch algorithm which is more efficient in practice than brute force (Bron and Kerbosch 1973).

We use the procedure above to define 3 nested haplotypes: one considering all modern non-African populations (haplotype D), one considering all modern human populations (haplotype C) and one considering all modern and Neanderthal samples (haplotype B). For the mismatch thresholds, we use 12, 12, and 3 respectively (this last lower threshold accommodates the small sample size of archaic populations). Haplotype D contains 25 SNPs, haplotype C contains 11 SNPs, and haplotype B 4 SNPs. When there is more than one maximal clique of SNPs to choose from, we select one that is a subset of a larger core. This means that haplotype B is a subset of haplotype C and haplotype C is a subset of haplotype D. Note that our haplotype D differs slightly from the derived haplotype defined by Ameur, et al. (2012) which was 28 SNPs long.

We constructed a haplotype network for the haplotype B region from 1000 Genomes European, East Asian and African haplotypes, using the R package "pegas" (Paradis 2010). We inferred the phylogenetic relationship between the haplotypes by picking a single individual from the SGDP that was homozygous for each of the representative haplotypes and inferring the tree relating the haplotypes using BEAST2 (Bouckaert, et al. 2014). We rooted the tree with chimpanzee and used a uniform [7.5-9.5] million year prior for human-chimp divergence. This corresponds to a mutation rate of ~4-5×10^{-10} per-base per-generation. Human-chimp divergence in the region is around 90% of the genome-wide average, so there does not appear to be a large difference in mutation rate in the region that would bias these estimates.

ABC and startmrca analysis

To quantify the strength and timing of selection in different populations, we ran a simulation study using approximate Bayesian computation (ABC), specifically using *ABCtoolbox* (Wegmann, et al. 2010). We use the pipeline implemented by Peter, et al. (2012) , which uses a model selection method to distinguish selection on a *de novo* mutation (SDN) from selection on standing variation (SSV). In addition to model selection, we estimated two continuous parameters: the selection-onset time and the selection coefficient. To perform the simulations for ABC we used *mbs* (Teshima and Innan 2009), which creates a selected allele frequency trajectory backward in time from a specified present-day frequency. This implicitly creates a range of selection-onset times in the past, which we used as a prior. For the selection strength, we used a uniform prior of [-4,-1] on the log of the selection coefficient.

We chose the derived allele (C) of rs174546 as the putatively selected allele, and analyzed 50kb on either side for a total region length of L=100kb. We fixed the mutation rate at 1.25x10^{-8} per base per generation, the recombination rate at 1.1785x10^{-8} per base per generation [the average rate in this region (International HapMap Consortium 2007)], the current effective population size N_e at 10,000, and the heterozygote advantage at 0.5. For each region (AFR, EAS, and EUR) and each model (SDN and SSV), we simulated 1 million datasets. Each dataset had a sample size of n=170 haplotypes and a length of L=100kb to match the 1000 Genomes data. Within each region we chose four representative populations for further analysis: ESN, GWD, LWK, and YRI for AFR; CDX, CHB, CHS, and JPT for EAS; and CEU, FIN, GBR, and TSI for EUR. For the current frequency of the selected allele, we computed a 95% confidence interval using the four subpopulations within each region and used this as a prior. This resulted in [0.984, 0.992] for AFR, [0.304, 0.670] for EAS, and [0.575, 0.699] for EUR. We also used population-specific demographies from PSMC, based on Yoruba, Han and French individuals from the SGDP (Mallick, et al. 2016).

For each simulated dataset we computed 27 summary statistics (as described in Peter, et al. (2012)). During the ABC estimation phase, we retained the top 500 simulated datasets with statistics closest to each real dataset (12 in total, one for each subpopulation), and computed posterior distributions for the selection-onset time and selection coefficient (Supplementary Table 2). Finally, we combined estimates (Table 1) by assuming a lognormal distribution for the posterior times and computing the distribution of the weighted mean of the eight estimates (four populations each with new/standing variation estimates).

For EAS and EUR, we also wanted to test for ancient selection on the ancestral allele (T) of rs174546. To this end, we merged the subpopulations for both EAS and EUR, and selected 200 haplotypes with the ancestral allele for each. Then we ran our ABC procedure in the same way as before, except with a "current" allele frequency of 0.999999 (*mbs* does not allow 1 for technical reasons). The EAS results were very sensitive to demography–particularly to the value of present-day N_e–so we restricted maximum N_e to the value of 45,000 used by Gravel, et al. (2011).

Startmrca (Smith, et al. 2017) is a method for estimating the selection-onset time of a beneficial allele. It uses an HMM (in the "copying" style of Li and Stephens (2003)) to model present-day haplotypes as imperfect mosaics of the selected haplotype and a reference panel of non-selected haplotypes. We used this method to estimate the selection-onset time for the ancestral allele of rs174546 in the EAS and EUR superpopulations described above. We used YRI individuals that are homozygous for the derived allele as the reference panel. Following Smith, et al. (2017), we used a 1Mb region surrounding rs174546, the HapMap combined recombination map in this region (International HapMap Consortium 2007), 100 haplotypes in the selected population, and 20 haplotypes in the reference population. To obtain selection-onset time estimates for each population, we ran 5 independent MCMC chains, each with 10,000 iterations. We post-processed the results by discarding the first 6000 iterations (burn-in), and retaining the remaining successful iterations over all 5 chains. To obtain credible intervals, we took the 2.5 and 97.5 quantiles of each resulting distribution (Supplementary Table 3.2).

We also analyzed selection for the derived allele of rs174546 in the EAS and EUR superpopulations. In this case, there were not enough AFR individuals homozygous for the ancestral allele to use as a reference population. So instead we used a "local" reference population, i.e. EAS and EUR individuals homozygous for the ancestral allele. We also use a lower prior for the derived allele ([0-4,000] generations vs. [0-20,000] for the ancestral). We did not analyze AFR because the selection is likely too old for the method, and there is no suitable reference population. We note that *startmrca* tends to underestimate the time of

very ancient selective events (the "star" genealogy assumption becomes less appropriate) and does not account for selection on standing variation, which would likely lead to overestimation, so the quantitative results may be unreliable.

Ancient DNA analysis

We first analyzed 46 Upper Palaeolithic and Mesolithic samples (Fu, et al. 2016) for the presence of the derived *FADS1* allele. These samples were typed on a capture array ("1240k capture") that contains 6 of the 25 SNPs that define haplotype D. We divided the samples into broad population groups as defined in the original publication (Fu, et al. 2016) and inferred the allele frequency in each of these populations by maximizing the following likelihood function:

$$\sum_{i=1}^{N} \log\left(p^2 \prod_{j=1}^{6} B(r_{ij}, r_{ij}+a_{ij}, \varepsilon) + 2p(1-p) \prod_{j=1}^{6} B(r_{ij}, r_{ij}+a_{ij}, 0.5+\delta) + (1-p)^2 \prod_{j=1}^{6} B(r_{ij}, r_{ij}+a_{ij}, 1-\gamma) \right)$$

where r_{ij} and a_{ij} are the number of reference and alternative reads from individual *I*at SNP j, N is the number of individuals in the population. $B(x,n,p)$ is the binomial probability of seeing x successes out of n trials with probability p, and ε,δ,γ are small error probabilities, which we set to 0.1, 0, 0.1 for transversions, 0.15, 0.05, 0 for C>T or G>A transitions, and 0, 0.05, 0.15 for T>C or A>G transitions. This implies a conservative 10% rate of contamination or error, and a 5% deamination rate. We computed binomial confidence intervals assuming an effective sample size of $2 - \left(\frac{1}{2}\right)^{\sum_{i,j=1,1}^{N,6}(r_{ij}+a_{ij})-1}$.

To impute the *FADS1* haplotype in the Upper Palaeolithic and Mesolithic samples, we computed genotype likelihoods at each SNP in the 5Mb region around the derived haplotype, for each individual, assuming a binomial distribution of reference and alternative allele counts and a 5% deamination rate. We then imputed using *BEAGLE 4.1* (Browning and Browning 2016), and the 1000 Genomes reference panel downloaded from bochet.gcc.biostat.washington.edu/beagle/1000_Genomes_phase3_v5a. We filtered out imputed SNPs with a genotype probability of less than 0.8 and used the remaining SNPs to determine the presence of the allele. In order to estimate the false positive rate we simulated read data at different coverages for 50 individuals from the 1000 Genomes EAS (East Asian) super-population who are homozygous for the ancestral allele. We computed genotype likelihoods and imputed as for the ancient data, having first removed the 50 test individuals from the reference panel. We performed 10 simulations for each coverage level and used the frequency at which the derived allele was imputed as an estimate of the false positive rate.

To analyze the Holocene history of *FADS1* and other alleles, we assembled a dataset of 1003 published ancient samples, most of which were typed on the "1240k" capture array which targets ~1.24 million SNPs. We used the pseudo-haploid version of these data, where each individual has a single allele at each SNP, from a randomly selected read. We classified these individuals into "hunter-gatherer", "Early Farmer" and "Steppe ancestry" populations as follows. First, we ran supervised ADMIXTURE (Alexander, et al. 2009) with K=4, and the four populations: WHG, EHG, Anatolia_Neolithic and Yamnaya_Samara fixed to have cluster membership 1, as previously reported (Mathieson, et al. 2018). We then classified individuals based on their inferred ancestry. If they had more than 25% ancestry from the Yamnaya_Samara cluster and dated later than 7500 BP, we classified them as "Steppe ancestry". If they had less than 25% ancestry from the Yamnaya_Samara cluster and more then 50% from the Anatolia_Neolithic cluster, we classified them as "Early Farmer". Finally, if they had less than 25% ancestry from the Yamnaya_Samara cluster and less than 50% ancestry from the Anatolia_Neolithic cluster and were dated earlier than 5100 BP, we classified them as "hunter-gatherer". These classifications are informed by previous analysis that combined genetic, chronological and archaeological information, and largely correspond to classifications that would be derived from archaeological context alone. We estimated allele frequency trajectories and selection coefficients separately for each population using a method that fits a hidden Markov model to the observed frequencies (Mathieson and McVean 2013).

To call *AMY1* copy number we assembled a set of 76 ancient genomes with shotgun sequence data that had nonzero mapped coverage at the locus. The majority of published ancient shotgun genomes have zero coverage, presumably because the copy number variable region was masked during alignment. We counted the number of reads that mapped to the any of the three *AMY1* duplicate regions in the human reference genome (Usher, et al. 2015) and compared the total to the average read depth in 1000 random regions of chromosome 1, of the same size as the *AMY1* duplicate regions. We fitted a linear model of coverage as a function of GC content to these 1000 regions, for each individual, and used this to correct our estimates for GC bias.

Analysis of lipid GWAS hits

We tested the directionality of lipid-associated alleles using genome-wide association meta-analysis results for LDL, HDL and TG (Teslovich, et al. 2010). Specifically, we constructed a list of SNPs with P-values below a given cutoff by iteratively selecting the SNP with the lowest P-value and then removing all SNPs within 500kb. For each of these SNPs we extracted allele frequencies in the EUR and AFR

super-populations and then tested whether trait increasing alleles were more common in AFR than EUR (Figure 3A).

To analyze the LDL hits in ancient samples, we first identified all SNPs with an association P-value less than 10^{-6} that were on the capture array used to genotype the majority of the ancient samples. We iteratively removed SNPs within 250kb of the most-associated SNPs to produce an independent set of associated SNPs. For each (pseudo-haploid) individual, we constructed the "LDL score" by counting the proportion of these SNPs at which that individual carried the trait-increasing allele (Figure 3B). We found no significant differences with respect to ancestry when we fitted a binomial generalized linear model with ancestry as a covariate. We also fitted a model including date as a covariate to test for significant differences over time, also with a nonsignificant result.

References

1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. Nature 526:68-74.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655-1664.

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 522:167-172.

Ameur A, Enroth S, Johansson A, Zaboli G, Igl W, Johansson AC, Rivas MA, Daly MJ, Schmitz G, Hicks AA, et al. 2012. Genetic adaptation of fatty-acid metabolism: a human-specific haplotype increasing the biosynthesis of long-chain omega-3 and omega-6 fatty acids. Am J Hum Genet 90:809-820.

Amorim CE, Nunes K, Meyer D, Comas D, Bortolini MC, Salzano FM, Hunemeier T. 2017. Genetic signature of natural selection in first Americans. Proc Natl Acad Sci U S A 114:2195-2199.

Bellwood P. 2004. First Farmers: The Origins of Agricultural Societies: Wiley-Blackwell.

Bocherens H. 2009. Neanderthal Dietary Habits: Review of the Isotopic Evidence. In: Hublin J-J, Richards MP, editors. The Evolution of Hominin Diets: Springer. p. 241-250.

Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol 10:e1003537.

Bron C, Kerbosch J. 1973. Finding All Cliques of an Undirected Graph [H]. Communications of the Acm 16:575-577.

Browning BL, Browning SR. 2016. Genotype Imputation with Millions of Reference Samples. Am J Hum Genet 98:116-126.

Browning SR, Browning BL. 2015. Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent. Am J Hum Genet 97:404-418.

Buckley MT, Racimo F, Allentoft ME, Jensen MK, Jonsson A, Huang H, Hormozdiari F, Sikora M, Marnetto D, Eskin E, et al. 2017. Selection in Europeans on Fatty Acid Desaturases Associated with Dietary Changes. Mol Biol Evol 34:1307-1318.

Burger J, Kirchner M, Bramanti B, Haak W, Thomas MG. 2007. Absence of the lactase-persistence-associated allele in early Neolithic Europeans. Proc Natl Acad Sci U S A 104:3736-3741.

546 Cassidy LM, Martiniano R, Murphy EM, Teasdale MD, Mallory J, Hartwell B, Bradley DG.
547 2016. Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic
548 genome. Proceedings of the National Academy of Sciences 113:368-373.

549 Chimpanzee Sequencing Analysis Consortium. 2005. Initial sequence of the chimpanzee genome
550 and comparison with the human genome. Nature 437:69-87.

551 Darios F, Davletov B. 2006. Omega-3 and omega-6 fatty acids stimulate cell membrane
552 expansion by acting on syntaxin 3. Nature 440:813-817.

553 Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Jarvela I. 2002. Identification of a
554 variant associated with adult-type hypolactasia. Nat Genet 30:233-237.

555 Field Y, Boyle EA, Telis N, Gao Z, Gaulton KJ, Golan D, Yengo L, Rocheleau G, Froguel P,
556 McCarthy MI, et al. 2016. Detection of human adaptation during the past 2000 years. Science
557 354:760-764.

558 Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, Patterson N, Rohland
559 N, Lazaridis I, Nickel B, et al. 2015. An early modern human from Romania with a recent
560 Neanderthal ancestor. Nature 524:216-219.

561 Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, Johnson PLF, Aximu-Petri A,
562 Prufer K, de Filippo C, et al. 2014. Genome sequence of a 45,000-year-old modern human from
563 western Siberia. Nature 514:445-449.

564 Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, Furtwangler A, Haak W, Meyer
565 M, Mittnik A, et al. 2016. The genetic history of Ice Age Europe. Nature 534:200-205.

566 Fumagalli M, Moltke I, Grarup N, Racimo F, Bjerregaard P, Jorgensen ME, Korneliussen TS,
567 Gerbault P, Skotte L, Linneberg A, et al. 2015. Greenlandic Inuit show genetic signatures of diet
568 and climate adaptation. Science 349:1343-1347.

569 Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V,
570 Domboróczki L, Kővári I, Pap I, Anders A, et al. 2014. Genome flux and stasis in a five
571 millennium transect of European prehistory. Nat Commun 5.

572 Garcia-Closas M, Hein DW, Silverman D, Malats N, Yeager M, Jacobs K, Doll MA, Figueroa
573 JD, Baris D, Schwenn M, et al. 2011. A single nucleotide polymorphism tags variation in the
574 arylamine N-acetyltransferase 2 phenotype in populations of European background.
575 Pharmacogenet Genomics 21:231-236.

576 Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, Yu F, Gibbs RA,
577 Genomes P, Bustamante CD. 2011. Demographic history and rare allele sharing among human
578 populations. Proc Natl Acad Sci U S A 108:11983-11988.

579 Groot PC, Mager WH, Frants RR. 1991. Interpretation of polymorphic DNA patterns in the
580 human alpha-amylase multigene family. Genomics 10:779-785.

581 Gunther T, Malmstrom H, Svensson EM, Omrak A, Sanchez-Quinto F, Kilinc GM, Krzewinska
582 M, Eriksson G, Fraser M, Edlund H, et al. 2018. Population genomics of Mesolithic Scandinavia:
583 Investigating early postglacial migration routes and high-latitude adaptation. PLoS Biol
584 16:e2003703.

585 Gunther T, Valdiosera C, Malmstrom H, Urena I, Rodriguez-Varela R, Sverrisdottir OO,
586 Daskalaki EA, Skoglund P, Naidoo T, Svensson EM, et al. 2015. Ancient genomes link early
587 farmers from Atapuerca in Spain to modern-day Basques. Proc Natl Acad Sci U S A 112:11917-
588 11922.

589 Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S,
590 Harney E, Stewardson K, et al. 2015. Massive migration from the steppe was a source for Indo-
591 European languages in Europe. Nature 522:207-211.

592 Hancock AM, Witonsky DB, Ehler E, Alkorta-Aranburu G, Beall C, Gebremedhin A, Sukernik
593 R, Utermann G, Pritchard J, Coop G, et al. 2010. Human adaptations to diet, subsistence, and
594 ecoregion are due to subtle shifts in allele frequency. Proc Natl Acad Sci U S A 107 Suppl
595 2:8924-8930.

596 Harris DN, Ruczinski I, Yanek LR, Becker LC, Becker DM, Guio H, Cui T, Chilton FH, Mathias
597 RA, O'Connor T. 2017. Evolution of Hominin Polyunsaturated Fatty Acid Metabolism: From
598 Africa to the New World. biorXiv:https://doi.org/10.1101/175067.

599 Hlusko LJ, Carlson JP, Chaplin G, Elias SA, Hoffecker JF, Huffman M, Jablonski NG, Monson
600 TA, O'Rourke DH, Pilloud MA, et al. 2018. Environmental selection during the last ice age on
601 the mother-to-infant transmission of vitamin D and fatty acids through breast milk. Proc Natl
602 Acad Sci U S A.

603 Hofmanová Z, Kreutzer S, Hellenthal G, Sell C, Diekmann Y, Díez-del-Molino D, van Dorp L,
604 López S, Kousathanas A, Link V, et al. 2016. Early farmers from across Europe directly
605 descended from Neolithic Aegeans. Proceedings of the National Academy of Sciences 113:6886-
606 6891.

607 Huff CD, Witherspoon DJ, Zhang Y, Gatenbee C, Denson LA, Kugathasan S, Hakonarson H,
608 Whiting A, Davis CT, Wu W, et al. 2012. Crohn's disease and genetic hitchhiking at IBD5. Mol
609 Biol Evol 29:101-111.

610 International HapMap Consortium. 2007. A second generation human haplotype map of over 3.1
611 million SNPs. Nature 449:851-861.

612 Jones ER, Gonzalez-Fortes G, Connell S, Siska V, Eriksson A, Martiniano R, McLaughlin RL,
613 Llorente MG, Cassidy LM, Gamba C. 2015. Upper Palaeolithic genomes reveal deep roots of
614 modern Eurasians. Nature communications 6.

615 Jones ER, Zarina G, Moiseyev V, Lightfoot E, Nigst PR, Manica A, Pinhasi R, Bradley DG.
616 2017. The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early European
617 Farmers. Curr Biol 27:576-582.

618 Keller A, Graefen A, Ball M, Matzas M, Boisguerin V, Maixner F, Leidinger P, Backes C,
619 Khairat R, Forster M, et al. 2012. New insights into the Tyrolean Iceman's origin and phenotype
620 as inferred by whole-genome sequencing. Nat Commun 3:698.

621 Kilinc GM, Omrak A, Ozer F, Gunther T, Buyukkarakaya AM, Bicakci E, Baird D, Donertas
622 HM, Ghalichi A, Yaka R, et al. 2016. The Demographic Development of the First Farmers in
623 Anatolia. Curr Biol 26:2659-2666.

624 Kothapalli KS, Ye K, Gadgil MS, Carlson SE, O'Brien KO, Zhang JY, Park HG, Ojukwu K, Zou
625 J, Hyon SS, et al. 2016. Positive Selection on a Regulatory Insertion-Deletion Polymorphism in
626 FADS2 Influences Apparent Endogenous Synthesis of Arachidonic Acid. Mol Biol Evol
627 33:1726-1739.

628 Lazaridis I, Mittnik A, Patterson N, Mallick S, Rohland N, Pfrengle S, Furtwängler A, Peltzer A,
629 Posth C, Vasilakis A, et al. 2017. Genetic origins of the Minoans and Mycenaeans. Nature
630 548:214-218.

631 Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, Fernandes D, Novak M,
632 Gamarra B, Sirak K, et al. 2016. Genomic insights into the origin of farming in the ancient Near
633 East. Nature 536:419-424.

634 Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber
635 JG, Castellano S, Lipson M, et al. 2014. Ancient human genomes suggest three ancestral
636 populations for present-day Europeans. Nature 513:409-413.

637 Li N, Stephens M. 2003. Modeling linkage disequilibrium and identifying recombination
638 hotspots using single-nucleotide polymorphism data. Genetics 165:2213-2233.

639 Lipson M, Szecsenyi-Nagy A, Mallick S, Posa A, Stegmar B, Keerl V, Rohland N, Stewardson
640 K, Ferry M, Michel M, et al. 2017. Parallel palaeogenomic transects reveal complex genetic
641 history of early European farmers. Nature 551:368-372.

642 Luca F, Bubba G, Basile M, Brdicka R, Michalodimitrakis E, Rickards O, Vershubsky G,
643 Quintana-Murci L, Kozlov AI, Novelletto A. 2008. Multiple advantageous amino acid variants in
644 the NAT2 gene in human populations. PLoS One 3:e3136.

645 Luca F, Perry GH, Di Rienzo A. 2010. Evolutionary adaptations to dietary changes. Annu Rev
646 Nutr 30:291-314.

647 Magalon H, Patin E, Austerlitz F, Hegay T, Aldashev A, Quintana-Murci L, Heyer E. 2008.
648 Population genetic diversity of the NAT2 gene supports a role of acetylation in human adaptation
649 to farming in Central Asia. Eur J Hum Genet 16:243-251.

650 Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N,
651 Nordenfelt S, Tandon A, et al. 2016. The Simons Genome Diversity Project: 300 genomes from
652 142 diverse populations. Nature 538:201-206.

Martiniano R, Caffell A, Holst M, Hunter-Mann K, Montgomery J, Müldner G, McLaughlin RL, Teasdale MD, Van Rheenen W, Veldink JH. 2016. Genomic signals of migration and continuity in Britain before the Anglo-Saxons. Nature communications 7:10326.

Mathias RA, Fu W, Akey JM, Ainsworth HC, Torgerson DG, Ruczinski I, Sergeant S, Barnes KC, Chilton FH. 2012. Adaptive evolution of the FADS gene cluster within Africa. PLoS One 7:e44926.

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.

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 528:499-503.

Mathieson I, McVean G. 2013. Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. Genetics 193:973-984.

Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prufer K, de Filippo C, et al. 2012. A high-coverage genome sequence from an archaic Denisovan individual. Science 338:222-226.

Nakamura MT, Nara TY. 2004. Structure, function, and dietary regulation of delta6, delta5, and delta9 desaturases. Annu Rev Nutr 24:345-376.

Olalde I, Allentoft ME, Sanchez-Quinto F, Santpere G, Chiang CW, DeGiorgio M, Prado-Martinez J, Rodriguez JA, Rasmussen S, Quilez J, et al. 2014. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. Nature 507:225-228.

Olalde I, Brace S, Allentoft ME, Armit I, Kristiansen K, Booth T, Rohland N, Mallick S, Szecsenyi-Nagy A, Mittnik A, et al. 2018. The Beaker phenomenon and the genomic transformation of northwest Europe. Nature.

Omrak A, Gunther T, Valdiosera C, Svensson EM, Malmstrom H, Kiesewetter H, Aylward W, Stora J, Jakobsson M, Gotherstrom A. 2016. Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool. Curr Biol 26:270-275.

Paradis E. 2010. pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics 26:419-420.

Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, et al. 2007. Diet and the evolution of human amylase gene copy number variation. Nat Genet 39:1256-1260.

Peter BM, Huerta-Sanchez E, Nielsen R. 2012. Distinguishing between selective sweeps from standing variation and from a de novo mutation. PLoS Genet 8:e1003011.

688 Prufer K, de Filippo C, Grote S, Mafessoni F, Korlevic P, Hajdinjak M, Vernot B, Skov L, Hsieh
689 P, Peyregne S, et al. 2017. A high-coverage Neandertal genome from Vindija Cave in Croatia.
690 Science 358:655-658.

691 Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G,
692 Sudmant PH, de Filippo C, et al. 2014. The complete genome sequence of a Neanderthal from
693 the Altai Mountains. Nature 505:43-49.

694 Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S,
695 Stafford Jr TW, Orlando L, Metspalu E, et al. 2014. Upper Palaeolithic Siberian genome reveals
696 dual ancestry of Native Americans. Nature 505:87-91.

697 Raghavan M, Steinrucken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, Albrechtsen A,
698 Valdiosera C, Avila-Arcos MC, Malaspinas AS, et al. 2015. Genomic evidence for the
699 Pleistocene and recent population history of Native Americans. Science 349:aab3884.

700 Richards MP. 2009. Stable Isotope Evidence for European Upper Paleolithic Human Diets. In:
701 Hublin J-J, Richards MP, editors. The Evolution of Hominin Diets: Springer. p. 251-257.

702 Saag L, Varul L, Scheib CL, Stenderup J, Allentoft ME, Saag L, Pagani L, Reidla M, Tambets
703 K, Metspalu E, et al. 2017. Extensive Farming in Estonia Started through a Sex-Biased
704 Migration from the Steppe. Curr Biol 27:2185-2193 e2186.

705 Sabbagh A, Darlu P, Crouau-Roy B, Poloni ES. 2011. Arylamine N-acetyltransferase 2 (NAT2)
706 genetic diversity and traditional subsistence: a worldwide population survey. PLoS One
707 6:e18507.

708 Schiffels S, Haak W, Paajanen P, Llamas B, Popescu E, Loe L, Clarke R, Lyons A, Mortimer R,
709 Sayer D. 2016. Iron age and Anglo-Saxon genomes from East England reveal British migration
710 history. Nature communications 7:10408.

711 Schlebusch CM, Malmstrom H, Gunther T, Sjodin P, Coutinho A, Edlund H, Munters AR,
712 Vicente M, Steyn M, Soodyall H, et al. 2017. Southern African ancient genomes estimate
713 modern human divergence to 350,000 to 260,000 years ago. Science 358:652-655.

714 Seguin-Orlando A, Korneliussen TS, Sikora M, Malaspinas A-S, Manica A, Moltke I,
715 Albrechtsen A, Ko A, Margaryan A, Moiseyev V, et al. 2014. Genomic structure in Europeans
716 dating back at least 36,200 years. Science 346:1113-1118.

717 Skoglund P, Malmstrom H, Omrak A, Raghavan M, Valdiosera C, Gunther T, Hall P, Tambets
718 K, Parik J, Sjogren KG, et al. 2014. Genomic diversity and admixture differs for Stone-Age
719 Scandinavian foragers and farmers. Science 344:747-750.

720 Skoglund P, Malmstrom H, Raghavan M, Stora J, Hall P, Willerslev E, Gilbert MT, Gotherstrom
721 A, Jakobsson M. 2012. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in
722 Europe. Science 336:466-469.

723 Smith J, Coop G, Stephens M, Novembre J. 2017. Estimating Time to the Common Ancestor for
724 a Beneficial Allele. Mol Biol Evol 35:1003-1017.

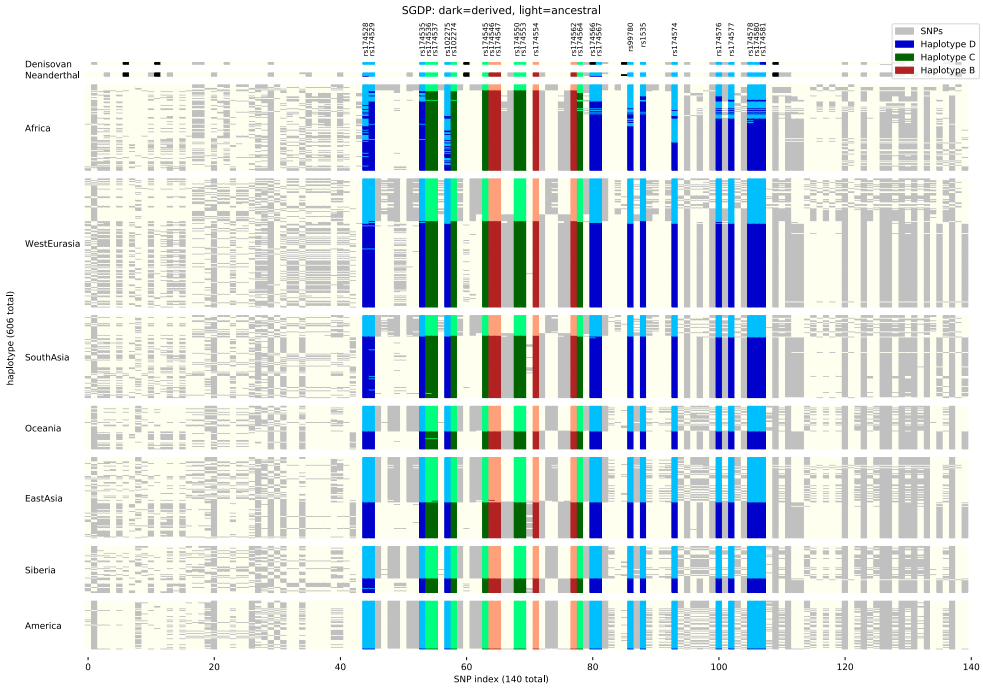
725 Teshima KM, Innan H. 2009. mbs: modifying Hudson's ms software to generate samples of
726 DNA sequences with a biallelic site under selection. BMC Bioinformatics 10:166.

727 Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello
728 JP, Ripatti S, Chasman DI, Willer CJ, et al. 2010. Biological, clinical and population relevance
729 of 95 loci for blood lipids. Nature 466:707-713.

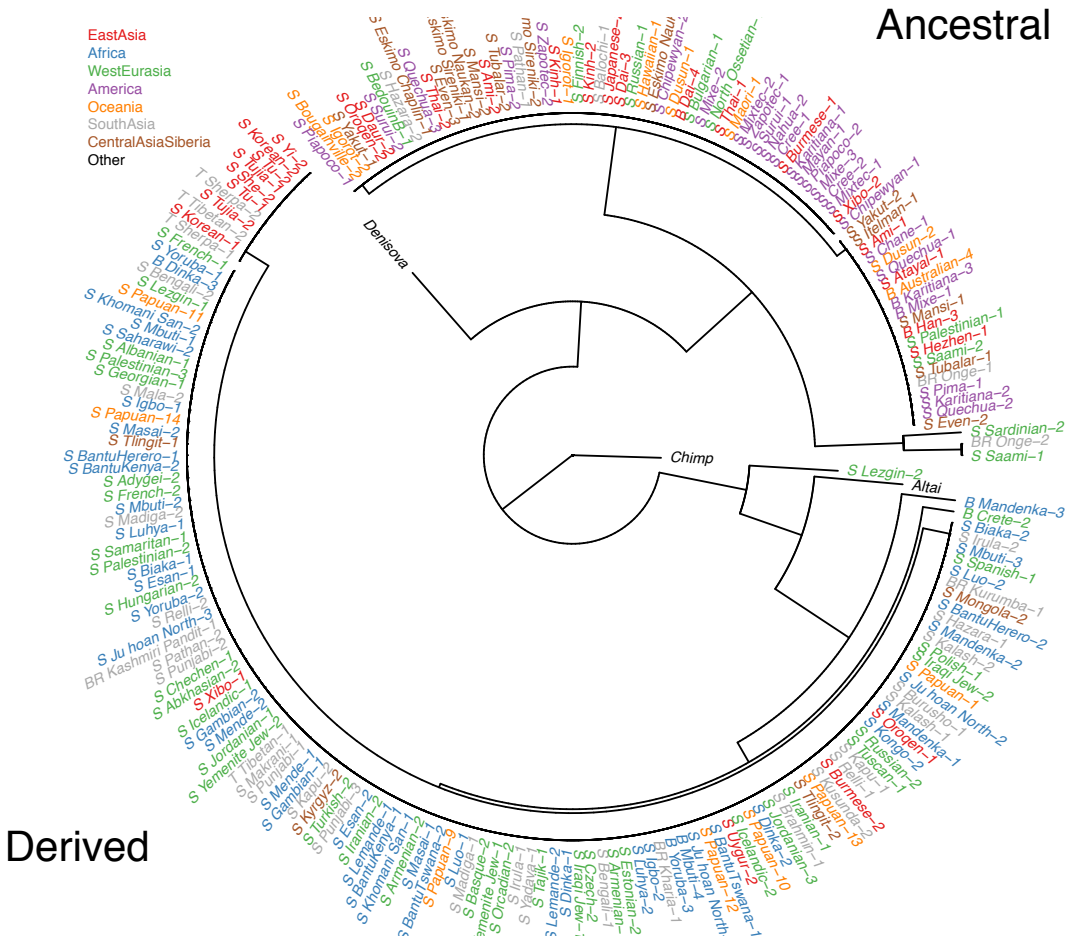
730 Usher CL, Handsaker RE, Esko T, Tuke MA, Weedon MN, Hastie AR, Cao H, Moon JE, Kashin
731 S, Fuchsberger C, et al. 2015. Structural forms of the human amylase locus and their
732 relationships to SNPs, haplotypes and obesity. Nat Genet 47:921-925.

733 Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L. 2010. ABCtoolbox: a versatile
734 toolkit for approximate Bayesian computations. BMC Bioinformatics 11:116.

735 Ye K, Gao F, Wang D, Bar-Yosef O, Keinan A. 2017. Dietary adaptation of FADS genes in
736 Europe varied across time and geography. Nat Ecol Evol 1:167.
737



Supplementary Figure 1: As Figure 1A, including all 300 SGDP individuals.



Supplementary Figure 2: UPGMA tree of haplotype B region including all 300 SGDP individuals