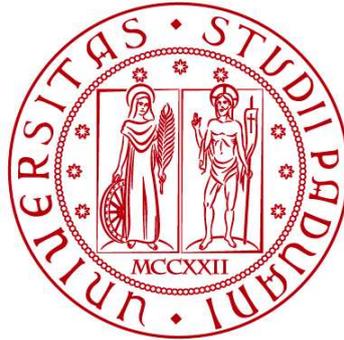


UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOLOGIA

Corso di Laurea in Biologia



ELABORATO DI LAUREA

**THE NEOLITHIC TRANSITION: SIGNS OF SELECTION AND
LEGACY OF ANCIENT GENOMES IN PRESENT-DAY
EUROPEANS**

Tutor: Prof. Massimo Mezzavilla
Dipartimento di Biologia

Laureanda: Margherita Vincenti

ANNO ACCADEMICO 2024/2025

ABSTRACT

The onset of the Neolithic era, approximately 12,000 years ago, marked a pivotal period in human history, characterized by profound transformations in lifestyle, subsistence strategies, and population dynamics. The shift from a nomadic hunter-gatherer lifestyle to a sedentary one, based on agriculture and animal domestication, resulted in substantial demographic and economic reconfigurations, and introduced unprecedented selective pressures on human populations.

Advancements in ancient and modern DNA analyses have enabled the identification of genomic regions subjected to strong positive selection in response to these environmental and cultural changes. The strongest signs of selection have been found on the MCM6/LCT locus, on the FADS gene cluster and on HLA region, which seem to be due respectively to processes of adaptation to famine and dairy consumption, changes in diet, and pathogen exposure.

Furthermore, the expansion of early Neolithic farming groups from Southwest Asia significantly reshaped the genetic landscape of European populations. This process involved both the replacement of local hunter-gatherer groups and admixture with them, resulting in a complex mosaic of ancestries. Contemporary European populations thus exhibit heterogeneous patterns of genetic diversity, shaped by region-specific admixture proportions and ancestral contributions. These genetic patterns not only inform our understanding of human population history, but also serve as the basis for the development of ancestry-specific polygenic risk scores for complex traits.

INDEX

1. INTRODUCTION	1
2. THE EARLY FARMERS	3
2.1 Diffusion of Neolithic ways of life	3
3. THE NEOLITHIC DEMOGRAPHIC TRANSITION	7
3.1 The disadvantages of sedentism	9
3.2 Reproductive trade-offs	11
4. THE SELECTION LANDSCAPE	15
4.1 Signatures of selection	16
4.2 Evolution of lactase persistence in Europe	17
4.3 Signs of selection on the FADS gene cluster	22
4.4 Selection on immune-associated loci	24
5. DISTRIBUTION OF ANCIESTRIES IN PRESENT-DAY EUROPEANS	26
5.1 Phenotypic legacy of ancient Eurasians	27
6. METHODS	31
7. CONCLUSIONS	33
BIBLIOGRAPHY	35

1. INTRODUCTION

From the beginning, the Holocene witnessed some of the most extraordinary changes and revolutions in human societies. Along their quite short history, humans have been for the majority of time hunter-gatherers and just around 12.000 years ago they became producers. Around that time, the domestication of plants and animals took place: agricultural revolution appeared independently in multiple locations, representing a watershed between hunter-gatherer and farming-pastoralists human populations.

This chapter of human history is known as Neolithic, defined by agriculture and cattle husbandry as dominant subsistence system, permanent settlements and deep social changes. As regards for European experience of Neolithization, it all started in southeast Asia and then spread into Europe by cultural and mainly demic diffusion, leading to admixture events between farming and local hunter-gatherers communities, and to major changes in the gene pool of human populations.

Agricultural and pastoral practices gradually became the primary means of subsistence, inherently linked to the shift toward a sedentary lifestyle. Furthermore, sedentism and food surplus were responsible for profound changes in the architecture of social practices, leading to a steep demographic growth. However, increased population density, proximity to humans' and animals' waste deposits and famine due to crop failure led to a decreased well-being and health.

These challenging conditions impacted the evolutionary forces acting on the gene pool of early farmers, who were subjected to unprecedented selective pressures. As a consequence, many regions of the human genome containing diet- and immune-associated loci were targeted by natural selection, leaving molecular fingerprints that nowadays can be used to study the adaptive responses to the exposure to changing environmental conditions.

This work wants to explore the fascinating phenomenon of the agricultural revolution. Of particular interest was to understand which are the strongest signs of selection in the human gene pool that might be correlated to the drastic social and economic transition of Neolithic. Furthermore, it was worth highlighting how and to what extent ancient Eurasian populations contribute to present-day European genetic background, and if different ancestries and admixture proportions can explain extant population phenotypic distribution and risk for diseases.

2. THE EARLY FARMERS

The start of the Neolithic period is defined by a drastic change in the ways of life of many human communities, who adopted a lifestyle based on the exploitation and domestication of plants and animals, combined with higher degrees of sedentism. The Neolithization arrived in Europe through several processes of migration and cultural exchanges, but it originated in southwest Asia about 12,000 years ago. It was established by at least two genetically distinct hunter-gatherer populations, one in southern Levant and the other in a region extending from Iran to the Caucasus [1], and was subsequently brought into Europe along different routes and in different geographical and chronological contexts [1].

2.1 Diffusion of Neolithic ways of life

For many years, there has been debate over which model of diffusion might be the most plausible. Two candidate models (independently or in combination) have been proposed to explain the spread of Neolithic ways of life: demic and cultural diffusion models [2]. The demic diffusion model assumes large-scale migration events from areas where farming first appeared toward areas where it had not yet been established [3]. On the other hand, the cultural diffusion model proposes that the spread of agricultural technologies was mainly driven by the exchange of ideas and knowledge, with minimal involvement of population movement [3].

Analysis of modern DNA during the last decade of the previous century, and more recently the analysis of ancient DNA seems to suggest the primacy of demic over cultural diffusion model [2]. This suggestion comes from a variety of analysis approaches, such as calculating genetic distances, estimating divergence times, applying admixture models [2]. The main contribution of demic diffusion has been furthermore attested by genome-wide studies, showing that

more than 90% of the genetic ancestry of early European farmers comes from Anatolia [2].

However, some authors believe that the establishment of Neolithic societies across Europe cannot be exhaustively explained by the introduction of farming alone, whether it happened through cultural or demic diffusion [3]. This is because archaeological evidences show remarkable diversity of social practises and material expression among Neolithic communities, suggesting that the spread of agriculture in Europe occurred through a diverse and intricate set of processes, with different regions probably experiencing different combinations of cultural exchanges and migratory movements [3].

While the manner of spread is the subject of long-lasting debates, the routes through which Neolithic ways of life spread across Europe are quite undisputed [1]. The diffusion of the Neolithic transition started in the Near East towards Europe, following two main routes: an inland route over the Balkans and central Europe, and a sea route along the northern Mediterranean coast [2].

Around 6500 BC, the Aegean shores were reached by Neolithic settlers probably arrived from Anatolia—central Anatolia seems to be a region of secondary Neolithization itself, where sedentary settlements appeared around 8500 BC. Around 6000 to 5800 BC a new wave of diffusion extended from the Balkans northward to the southern Carpathian Basin, and later it continued into central Europe [1]. This movement was followed by a long stasis after 5200 BC: northwestern France was not reached until around 4500 BC, and the British Isles and southern Scandinavia not until around 4000 BC [1]. The second route of expansion followed a maritime route along the northern Mediterranean coast, following the Italian and Iberian peninsula coasts [1].

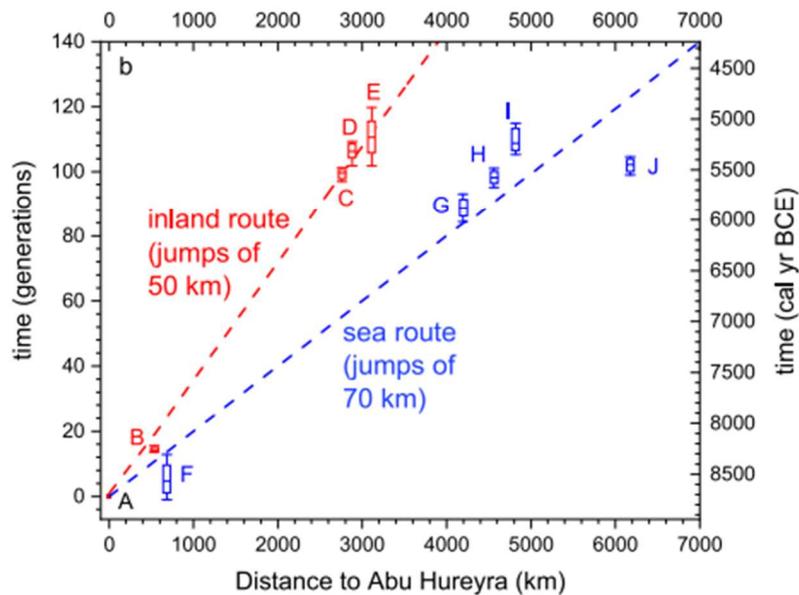
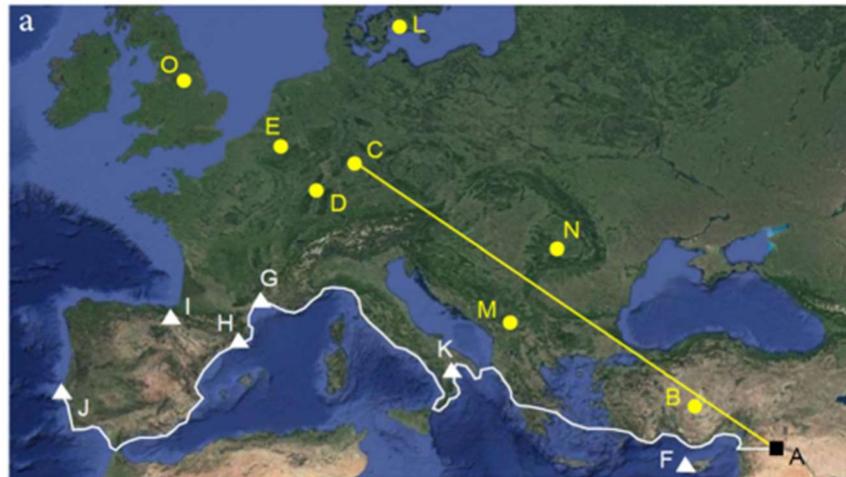


Fig1. The two routes of Neolithic spread in Europe. **a** Visualization of the two routes. Each yellow circle is the oldest Neolithic site in a region reached through the inland route. Each white triangle is the oldest Neolithic site in a region reached mainly along the sea route. Distances are estimated using great circles for the inland route (e.g., yellow line) and along the coast for the sea route (e.g., white line). The black square A is Abu Hureyra (Syria), the presumed origin of the Neolithic wave of advance. **b** Arrival times of Neolithic farmers in several regions according to the radiocarbon dates of sites in **a** and simulations (lines) for the inland route (red, jumps of 50 km per generation) and the sea route (blue, jumps of 70 km per generation). Each symbol is the radiocarbon date of the oldest Neolithic site in the region considered. The regions are A northern Mesopotamia, B Anatolia (present-day Turkey), C Germany, D northern France, E Belgium, F Cyprus, G southern France,

H Catalonia, I Navarre and J central Portugal. Some regions (K Italy, L Sweden, M Serbia, N Romania, and O United Kingdom) are not included in **b** due to the existence of delays in the arrival of the Neolithic⁵ , but more complicated models accounting for them would yield the same conclusions [Joaquim Fort & Joaquim Pérez-Losada, 2024]

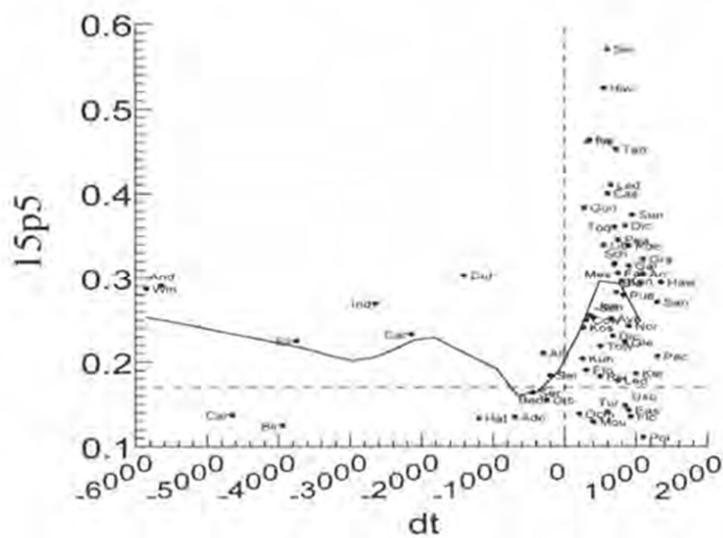


Fig. 3 Profile of variation in the proportion of immature skeletons in cemeteries (vertical axis, labelled 15p5) relative to local emergence of the Neolithic (horizontal axis, dt) (60 North American cemeteries) [Bocquet-Appel and Naji 2006]

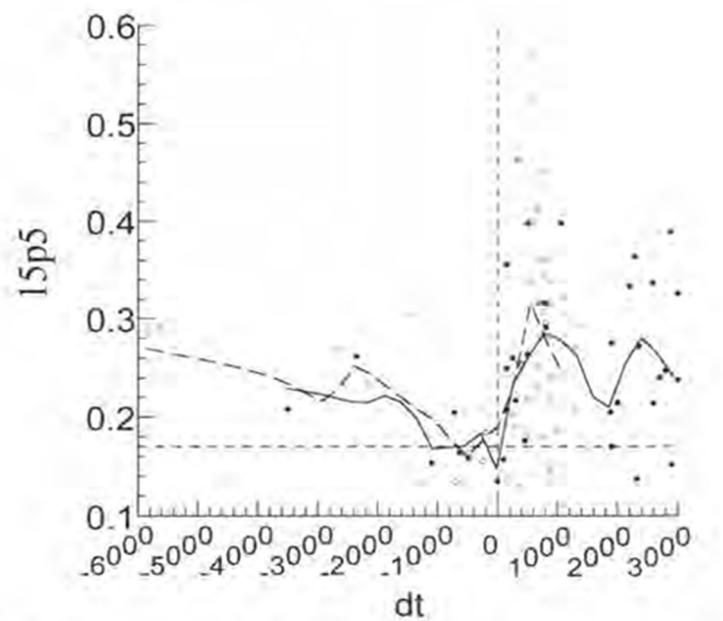


Fig. 4 Signatures of the two Neolithic demographic transitions, in Europe/North Africa and North America, relative to local emergence of the Neolithic [Bocquet-Appel and Naji 2006]

dt represents the relative chronology, specifically dt=0 corresponds to the moment when the neolithic agricultural production system was

introduced. **Figure 3** shows how the economic change, i.e. the new subsistence system, and the demographic change do not precede nor succeed each other, but they strictly coincide. Both farming and fertility explosion, however, are associated with sedentarism, which in some regions preceded the farming system while in others occurred simultaneously [5].

This leads us to conclude that population was at once the cause and the effect of this shift in civilization. The cause, because by exerting pressure on the carrying capacity of the hunter-gatherer production system, population also increased the probability of a system shift; and the consequence, because as soon as the new economic system was introduced, the population tended to grow towards the new carrying capacity of the horticulturist-farmer system, as attested by the fertility explosion. [5]

3.1 The disadvantages of sedentism

Our understanding of Palaeolithic ancestors has often been misleading, as they were commonly depicted as people living in hardship and teetering on the brink of starvation. However, more recent studies have allowed to create a clearer and more close-to-reality picture.

For many years, Thomas Hobbes' characterization of life in the Palaeolithic as «solitary, poor, nasty, brutish and short» was accepted as fact. This perception changed with studies and characterizations of gatherer-hunters suggested that Palaeolithic populations had abundant resources and could be considered the earliest affluent society. [6] On the other hand, sedentism and domestication of plants and animals arose problems in the Neolithic populations that nomad bands were not familiar with. «The incidence of infectious diseases and of disease-related mortality is consistently found to be less in the Palaeolithic than formerly assumed, but to have risen markedly in

sedentary Neolithic groups» [6]. In fact, the increase in population size provided an adequate host population for specialized pathogen, while before hunter-gatherers bands, that generally averaged 25 to 50 people [6], were not big enough to retain an endemic infectious pathogen.

Furthermore, there are several factors other than population size that influence the epidemiological profile of a population, such as ecosystem complexity, and that have major implications for the kind of pathogen that human population would be exposed to. Band societies rarely achieved sufficient size or density to become the sole focus of a specializing pathogen [6], but their epidemiology was strictly dependent on the relationship with other animals, being part of a zoonotic system characterized by generalist pathogens. Hunting-gathering ancestors would, then, be infected by insect bites, wounds inflicted by animals or by eating contaminated flesh, on the other hand they would be less exposed to viral diseases such as influenza, smallpox or measles [6].

On the contrary, the adoption of an agricultural-sedentary subsistence system led to a simpler ecosystem due to hunting and destruction of ecological niches in association with human activities. Also, the change in human group structures would make the early farmers much more vulnerable to more specific and systemic threats in terms of pathogens. Moreover, sedentism had as an obligate consequence the proximity of habitations to human and other animals waste deposits [6], often leading to water supply contamination and therefore to the spread of several diseases. In general, domesticated animals and their derivatives represented a steady supply of disease vectors.

The agricultural revolution also led to significant changes in the dietary habits of those who became producers. Unlike how they are commonly and quite unjustifiably portrayed, hunter-gatherers were rather well provided for by their surrounding resources for most of the year [6], and were able to successfully go through lean periods without any

major risk of starvation. On the contrary, agricultural groups that relied on a few species of plants and that simplified the surrounding environment were much more susceptible to nutritional problems.

Also, «a common practice among agricultural populations is that of early weaning, possibly so that the mother will return to a state of reproductive readiness» [6]. Newly-weaned infants placed in inadequate diet and in close contact with possible pathogens vectors and sources of viral disease were, as a consequence, more at risk compared to those of nomad bands.

3.2 Reproductive trade-offs

Despite the disadvantages associated with sedentism discussed before, the Neolithic transition allowed food storage and wealth accumulation: the increased calories availability combined with a reduction in energy expenditure due to a less mobile life style brought increasing energy availability for reproduction [7].

If, during the transition towards higher fertility values, the mortality rate had remained unchanged, the population size of the early farmers would have reached unbearable dimensions [7]. However, it is clear and undoubted that the agricultural revolution was characterized by deteriorating health and increasing morbidity and mortality rates, especially among younger age groups compared to mobile populations.

Studying extant hunter-gatherer societies, such as Agta foragers in the Philippines, it has been proposed a solution to the paradox of higher rates of population growth associated with increased mortality rates. The conflict existing between increasing disease burden and increasing reproductive success might be reconciled by a life history trade-off linked to settled agriculture, where mothers invest in offspring quantity rather than quality, and doing so they increase their reproductive fitness [7].

A quantity–quality trade-off is an adaptive process based on a principle of optimal distribution of limited resources: Parents must allocate finite resources either to a few higher-quality offspring or to multiple lower quality ones. [7]

A. E. Page *et al.* [7] provided empirical evidence of this adaptive mechanism underlying the agricultural revolution. Their study conducted on Agta foragers shows that settled mothers had significantly higher fertility compared to those that moved camp at least once, which might be due to the possibility for settled women to accumulate more somatic resources for reproduction (settlement was, indeed, a significant predictor of maternal body mass index (BMI)) [7].

Also, A. E. Page *et al.* used three types of white blood cells—lymphocytes, neutrophils, and eosinophils— as proxies for different immunological attacks: viral, bacterial, and helminthic, respectively. The sample, which included both adults and children, showed that people residing in camps with permanent and semipermanent houses were much more exposed to helminth infestations (associated with higher eosinophil concentrations) compared to those living in temporary camps. They also had a higher probability of having lymphocytosis but lower odds of neutrophilia. These negative effects of settlements extend to child mortality: settled mothers experienced significantly higher childhood mortality rates.

These results «show that women who settle down trade child survival for fertility, increasing their overall reproductive success [...] Although these settled women faced higher child mortality rates, they also experienced extremely high fertility» [7].

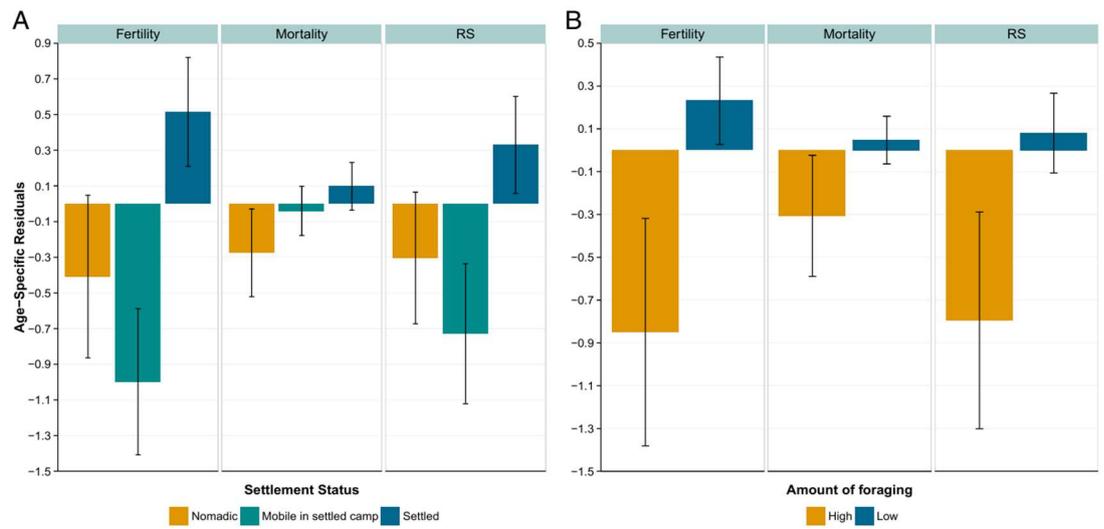


Fig. 5 Relationship between measures of transition and age-specific fertility, mortality, and reproductive success (RS) separated by individual settlement status (A) and degree of foraging (B) [A. E. Page *et al.* 2016]

4. THE SELECTION LANDSCAPE

The dramatic changes in the ways of life of Neolithic populations are to be held responsible for the shifts in the evolutionary forces acting on humans' gene pool and for imposing a series of heterogeneous selective pressures [8]. E. K. Irving-Pease *et al.* [8] sought to identify phenotype associated variants that have evidence for directional selection over the last 13,000 years. To estimate allele frequency trajectories and selection coefficients of trait-associated variants through time, they used the software *CLUES* which can perform inference of allele frequency trajectories using marginal trees sampled from a reconstruction of an ancestral recombination graph (ARG) for a set of genomic sequences, in combination with genotype likelihoods from serially sampled ancient DNA (aDNA). They used genomes from the 1,000 Genomes Project project and inferred allele trajectories and selection coefficients for 32,079 phenotype-associated variants, along with an equal number of putatively neutral "control" variants. These analysis identified no genome-wide significant selective sweeps ($p < 5e-8$) using present-day data alone. However, the trait-associated variants were significantly enriched for evidence of selection when compared to the control group. By contrast, when using imputed aDNA genotype probabilities, they identified 11 genome-wide significant selective sweeps in the GWAS variants, and none in the control groups. These results are consistent with selection preferentially acting on trait-associated variants. This analysis confirms many of the previously reported selection loci in West Eurasians, identified from present-day and ancient DNA, and reveals novel selective sweeps, while refining the temporal dynamics of the selected alleles.

Strong signs of selection can be detected in different parts of the genome, involving processes of adaptation to changes in diet, famine and pathogen exposure. In this work MCM6/LC locus, FADS cluster and Human Leukocyte Antigen (HLA) region will be discussed.

4.1 Signatures of selection

Mutation rates, recombination events, natural selection and gene flow are some of the driving forces of evolution responsible for the vast diversity we observe among human populations, and which is reflected in different phenotypes, including healthy and diseased ones. Furthermore, the effects of these forces on genetic diversity depends on the amount of genetic variations that already exist in a population, which remains unchanged in absence of selection, mutation, migration, genetic drift [9].

Negative selection, also called purifying selection, is responsible for the elimination of deleterious mutations. It can act more strongly or mildly depending on how disadvantageous the mutation is, with an equally strong or wick effect on the diversity in the population gene pool. On the other hand, positive selection favours those mutations that bring an advantage to their carrier, i.e. they increase their fitness. As for the decrease, also the increase in frequency of variants will affect the genetic diversity within a population, also by increasing that of genetically linked variants through the so called hitchhiking process [9]. Another form of natural selection is the balancing selection, where multiple alleles can persist at a given locus because they offer a benefit either on their own or in combination.

These forms of selection create characteristic molecular fingerprints [9] that can be identified using patterns of diversity [8]. However, identifying signatures of selection might be difficult for species that have been exposed to a wide range of changing environments through time and space, like *Homo sapiens*.

Yet, genome-wide association studies (GWAS)— which have identified many genetic variants associated with phenotypes of interest— and ancient DNA— by directly observing changes in trait-associated allele frequencies over time— may help us infer patterns of selection across Eurasia. Indeed, ancient DNA considerably improves our power to detect variants under selection, compared to a

method that only uses ARG inferred from present day data alone. Furthermore, ancient DNA provides a genetic snapshot of a time before demographic events, such as bottlenecks and isolation, that exacerbated drift and because of which the genetic variation that was present in past populations does no longer exist, and allows us to better understand the complex history of human populations, involving split, admixture, migration events. In particular, local ancestry inference (LAI) can be used to identify from which ancestral population chromosomal segments in admixed individuals are inherited.

4.2 Evolution of lactase persistence in Europe

Lactase persistence (LP) is an autosomal dominant trait that allows the production of the enzyme lactase through adult life. For most mammals this enzyme is crucial in early life stages, as it enables infants to live on mother's milk, but its expression decreases after weaning [10]. This would also be the ancestral state for humans, as continuing to produce an enzyme that is no longer needed would be a waste of energy.

The selection of LP has often been associated with the consumption of prehistoric milk. However, although these two phenomena must be linked to some extent, the reasons for positive LP selection are still quite uncertain [11].

The strongest overall signal of selection is observed at MCM6/LCT locus (rs4988235-A, also known as 13,910*T; ref. 30) where the derived allele gives lactase persistence [8]. This signal is driven primarily by sweeps in two of the ancestral backgrounds, associated with Eastern hunter-gatherers and Caucasus hunter-gatherers related ancestries¹.

¹ Eurasia during the post-glacial Holocene was inhabited by different lineages of hunter-gatherers, whose distribution reflects a complex mosaic of genetically and culturally distinct populations shaped by geography, climate, and waves of migration.

The distribution of this allele in modern Europe appears to be geographically highly structured, which might reflect different strengths of selection but also different population history [11].

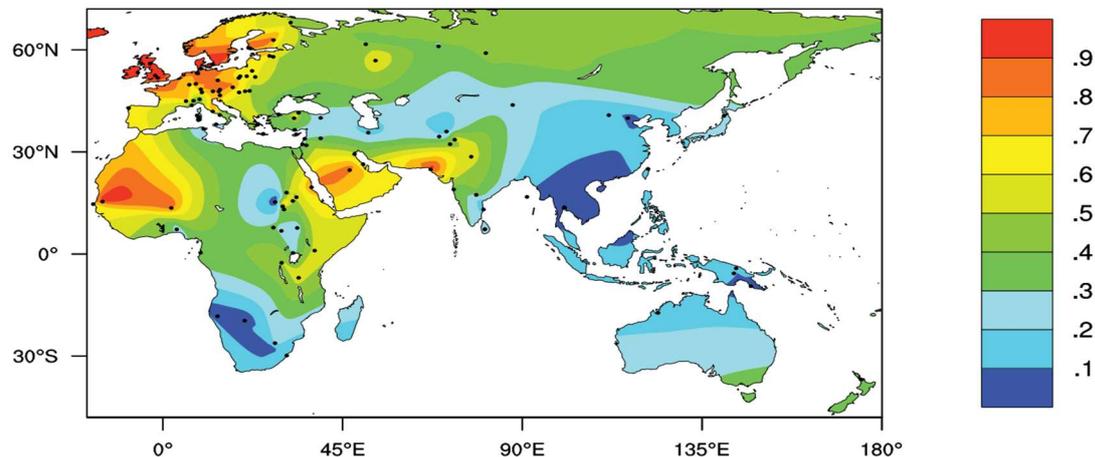


Fig. 6 Interpolated map of Old World LP phenotype frequencies. Dots represent collection locations. Colours and colour key show the frequencies of the LP phenotype estimated by surface interpolation [Yuval Itan *et al.* 2010].

From the Neolithic period onwards, Europe witnessed the widespread of milk consumption, characterized by spatial and temporal variability, also suggested by faunal and organic residue analysis [11]. The earliest organic residue which have been found that could witness the onset of dairying practises locates to seventh millennium BC northwest Anatolia. Afterwards, dairy husbandry was brought beyond the borders of sub-Mediterranean zone of Europe, through the inland and costal routes into central Europe during the sixth millennium BC [2], although milk consumption remains undetected in northern Greece neolithic sites [11].

However, selection studies on ancient DNA (aDNA) samples showed that in Europe LP only reached appreciable frequencies in Bronze and Iron Ages while it was nearly absent in Early Neolithic [10] and still low

The three major hunter-gatherer lineages are: Western Hunter-Gatherers (WHG) occupying Western and Central Europe, Eastern Hunter-Gatherers (EHG) occupying Eastern Europe, and Caucasus Hunter-Gatherers (CHG) occupying the Caucasus Mountains.

during Late Neolithic [10]. These data might challenge the commonly accepted notion that milk consumption drove selection on LP.

R. P. Evershed *et al* [11] analysed contemporary genotypic and phenotypic data from UK Biobank from about 500,000 people aged between 37 and 73 and saw that LP genotypes were out of Hardy Weinberg equilibrium. Among LNP participants the majority use cows' milk rather than some kind of substitute and only a small percentage of milk consumers reported following a lactose-free diet. This implies that lactase persistence has only a minor influence on milk consumption and it is not due to the use of lactose-free products. Furthermore, «consistent with these results, some non-European countries with very low levels of LP have been importing milk in vast quantities in recent years as part of a more general adoption of Western diets» [11].

Aside from dietary changes, there must have been a number of other factors influencing LP allele trajectories, probably associated with cultural and settlement-related changes. The new conditions of humans societies following the establishment of farming communities, i.e. proximity to domesticated animals, increased population density, famine due to crop failure, increased infectious disease load, might be – at least partially – responsible for LP selection [11].

It has been suggested that LP selection pressures would have been of major impact during times of subsistence instability: detrimental health consequences of lactose consumption by LNP individuals «would be acutely manifested during famines, leading to high but episodic selection favouring LP» [11]. This is because milk products were more likely to be consumed in absence of other food sources, and also lactose-induces diarrhoea could become fatal in severely malnourished individuals rather than healthy ones [11].

Another explanation might be linked to the increased pathogen loads associated with farming and higher population densities: the negative

effects on LNP people consuming milk would increase mortality and morbidity in a context of already high pathogen exposure, due to fluid loss and other gut disturbances. Yet, in a healthier environment milk consumption would not be as much detrimental and would not be able to drive LP positive selection [11].

This hypothesis seems to be sustained by the high linkage disequilibrium between rs4988235 and rs1438307 alleles in present-day individuals. Many SNPs within the MCM6/ LCT locus showed earlier evidence of selection than rs4988235, suggesting that selection at this locus is more complex than previously thought. LP allele began rising in frequency about 6,000 years ago, while rs1438307 about 12,000 years ago [8]. «This allele has been shown to regulate energy expenditure and contribute to metabolic disease and it has been suggested to be an ancient adaptation to famine» so it might explain the observed correlation between increased frequency of LP and the archaeological proxies for famine and increased pathogen exposure [8].

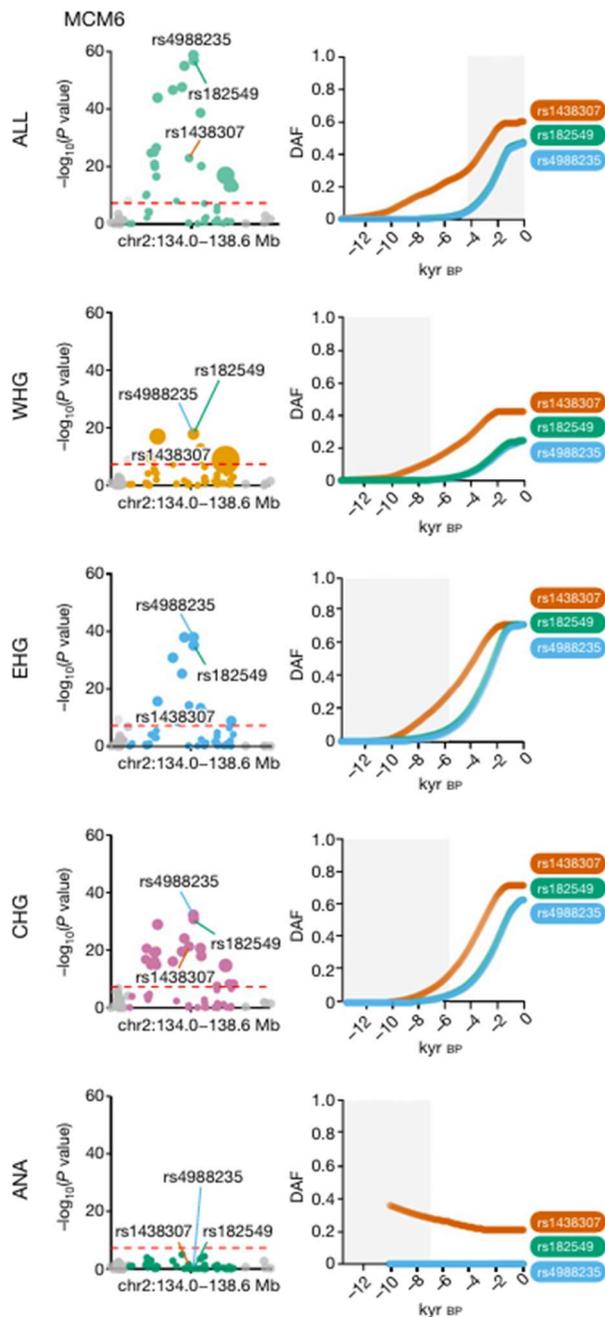


Fig. 7 Detailed plot for genome-wide significant sweep locus MCM6. Rows show results for the pan-ancestry analysis (ALL) plus the four marginal ancestries: WHG, EHG, CHG and ANA. The first column shows zoomed Manhattan plots of the P values for each ancestry and column two shows allele frequency trajectories for the top SNPs across all ancestries (grey shading for the marginal ancestries indicates approximate temporal extent of the pre-admixture population). [A. E. Page *et al.* 2016]

4.3 Signs of selection on the FADS gene cluster

Long-chain polyunsaturated fatty acids (LC-PUFAs) are crucial components of mammalian tissues and play an essential role in a variety of biological processes, such as neuronal membrane development, central nervous system cell signalling, inflammation control [12].

Humans can obtain LC-PUFAs directly by consuming meat and fish, or they can be synthesized endogenously from short chain polyunsaturated fatty acids (SC-PUFAs), linoleic acid (LA) or alpha-linolenic acid (ALA) [12]. SC-PUFAs are obtained mainly through the consumption of vegetable oils, and are used to synthesize LC-PUFAs in a pathway catalysed by two fatty acid desaturases: delta-5 desaturase (D5D) and delta-6 desaturase (D6D), encoded by *fatty acid desaturase 1* (FADS1) and *fatty acid desaturase 2* (FADS2) [12].

Different diets contain different ratios of SC- and LC-PUFAs and thus require different levels desaturases: because of the interaction between the environment and FADS gene cluster, the two genes (FADS1 and FADS2) have been repeatedly targeted by natural selection and have often been hypothesized to reflect adaptation to a more vegetarian diet associated with the onset of agriculture [13].

The pan-ancestry analysis conducted by E. K. Irving-Pease *et al.* [8] shows that the strong signal of selection is driven primarily by a sweep occurring across the Eastern hunter-gatherers (EHG), Western hunter-gatherers (WHG) and Anatolian farmers (ANA) haplotypic groups, but not statistically significant evidence of selection in the CHG background. However, the increase in frequency of the derived allele in the EHG background occurs around 8,000 years ago, after the admixture with Caucasus hunter-gatherers (CHG), where the derive allele was already close to present-day frequency.

These data suggest that much of the selection on the allele might have occurred in Neolithic populations before they arrived in Europe, it

probably already existed at significant frequencies in early farmer populations in the Middle East and Caucasus hunter-gatherers. Then the selection process continued during the Neolithic and Bronze Age when eastern groups moved northwards and westwards [8].

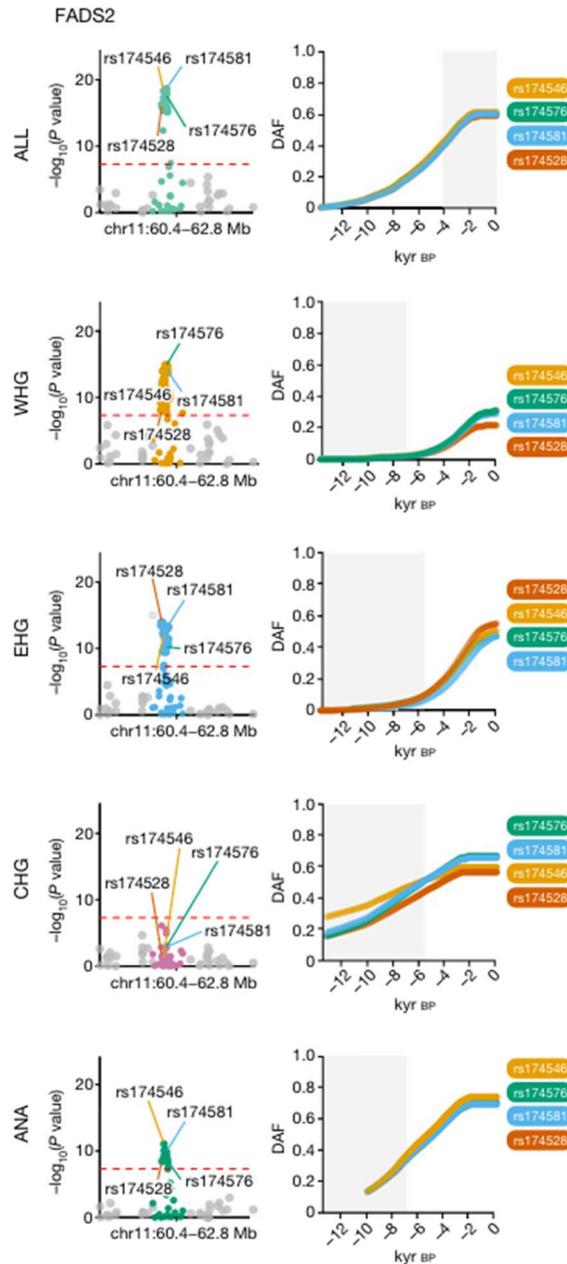


Fig. 8 Detailed plot for genome-wide significant sweep locus FADS2. Rows show results for the pan-ancestry analysis (ALL) plus the four marginal ancestries: WHG, EHC, CHG and ANA. The first column shows zoomed Manhattan plots of the P values for each ancestry and column two shows allele frequency trajectories for the top SNPs across all ancestries (grey shading for the marginal ancestries indicates

approximate temporal extent of the pre-admixture population). [A. E. Page *et al.* 2016]

4.4 Selection on immune-associated loci

Changes in diet and closer contact with domestic animals, combined with new social structures and increasing population density, forced humans to confront new challenges due to increased exposure to pathogens, viruses and deadly microbes. In particular, infectious diseases represented one of the major selective forces in human populations [14], in accordance with the Red Queen Hypothesis. Also, it has been suggested that human populations have become so well adapted to parasites that the reduced exposure to infectious agents in present-day societies may have disrupt the immune balances [14]. These putative selection events, then, might help explaining today prevalence of autoimmune diseases [8].

When comparing signatures of selection differentiating ancient hunter-gatherer and farmer populations we observe that several loci associated with immunity and autoimmune diseases have been under strong selection [8]. E. K. Irving-Paese *et al.* [8] detected an 8Mb-wide selection sweep signal in chromosome 6, spanning the full length of the HLA region. This study shows that «the strongest signal of selection at this locus in the pan-ancestry analysis is at an intergenic variant, located between HLA-A and HLA-W (rs7747253: A; $P = 7.56 \times 10^{-32}$; $s = -0.0178$), associated with protection against chicken pox [...], increased risk of intestinal infections [...] and decreased heel bone mineral density» [8]. It has been discovered that this allele rapidly decreased in frequency around 8,000 years ago, reducing the risk of intestinal infections at the cost of increasing risk of chickenpox [8], suggesting an adaptive response and a genetic trade-off following the increased exposure to infectious diseases associated with sedentism.

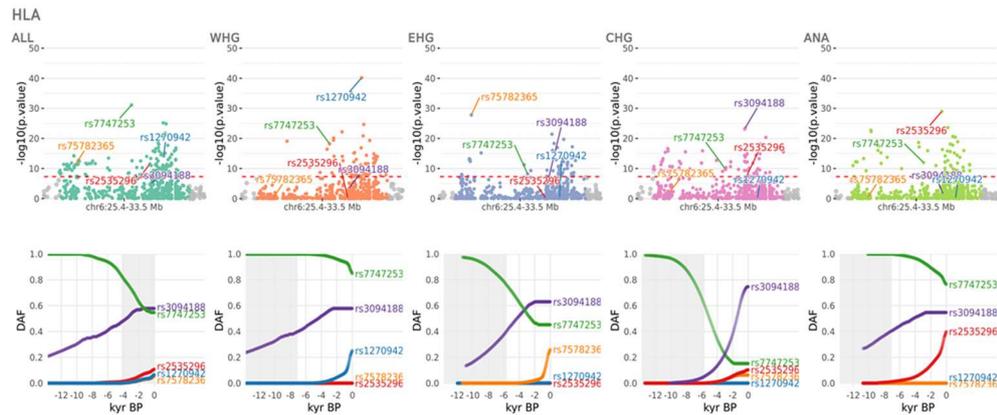


Fig. 9 CLUES selection results for the HLA locus, showing the pan-ancestry analysis (ALL) plus the four marginal ancestries: Western hunter-gatherers (WHG), Eastern hunter-gatherers (EHG), Caucasus hunter-gatherers (CHG) and Anatolian farmers (ANA). Row one shows zoomed Manhattan plots of the p values for each ancestry and row two shows allele trajectories for the top SNPs across all ancestries (grey shading for the marginal ancestries indicates approximate temporal extent of the pre-admixture population). [A. E. Page *et al.* 2016]

Also, within this sweep sign of selection at C2 (rs9267677: C; $P = 6.60 \times 10^{-26}$; $s = 0.0441$) shows that the allele associated with protection against sexually transmitted diseases and with increase psoriasis risk started to increase about 4,000 years ago, and then it began rising more rapidly around 1,000 years ago. This locus serves as a clear example of how the high rates of autoimmune diseases seen today could be partly explained by genetic trade-offs shaped by evolutionary pressures [8].

Because HLA region is highly pleiotropic, it is challenging to identify the specific selective forces responsible for the frequency increases across different historical periods [8]. Nonetheless, «profound shifts in lifestyle in Eurasian populations during the Holocene have been suggested to be drivers for strong selection on loci involved in immune response» [8].

5. DISTRIBUTION OF ANCIESTRIES IN PRESENT-DAY EUROPEANS

The Holocene has been a moment of great changes in the evolutionary history of Humans, marked by a number of selection events, migrations and cultural revolutions. The spread of the first farmers out of southwest Asia was responsible for major changes in the genetic pool of European populations, with almost complete replacement in the south and increased mixing further north with local Eastern and Western hunter-gatherers (EHG and WHG) [15]. Other contributions to Europe genetic pool came from Caucasus hunter-gatherers (CHG) [8] and Yamnaya populations from the Pontic steppe that in Early bronze Age carried partial ancestry from ancient North Eurasians (ANE) [16].

Today, the legacy of these populations can still be seen in present-day Europeans genetic background, which carry different ancestries and admixture proportions. However, although signatures of selection can be detected from patterns of genetic diversity among extant populations, «in the complex mosaic of genetic affinities that constitute a present-day human genome, any putative signatures of selection may misrepresent the timing and magnitude of the selective process. For example, episodes of admixture between ancestral populations can result in present-day haplotypes that contain no evidence of selective processes occurring further back in time» [8].

The characterization of the genetic legacy of Mesolithic, Neolithic and Bronze Age populations not only may help us to better understand the phenotypic heterogeneity of present populations and its distribution, but also to calculate ancestry-specific polygenic risk score for complex traits [8].

5.1 Phenotypic legacy of ancient Eurasians

E. K. Irving-Pease *et al.* [8] inferred the local ancestry and admixture proportions in present-day populations by chromosome ‘painting’ using ChromoPainter. The process of painting consists of creating a reference/donor panel consisting of ancient individuals of as pure ancestry as possible, and a target/recipient panel of which we want to estimate the admixture proportions. The target/recipient panel and reference/donor panel are filtered for variants, merged, and the target panel is painted using the reference panel as donors. In particular, the SNPs used were chosen to capture genome-wide variation, rare and coding variants, and variants relevant to specific phenotypes or regions of interest. ChromoPainter uses an approach based on the observation that markers on the same chromosome are inherited together unless separated by recombination; at the population level, this results in linkage disequilibrium between markers that reflect a shared history of descent. The haplotype-based algorithm of ChromoPainter aims to use this information, detecting shared haplotypes to reconstruct phased recipient genomes as chunks ‘copied’ from donors. Considering the genealogy of a single locus, we can identify one or more closest relatives to that locus, called ‘nearest neighbours’. Doing so the haplotypes of a recipient/target individuals can be reconstruct as a series of chunks of genetic material donated by donor/reference individuals.

E. K. Irving-Pease *et al.* results show that Mesolithic, Neolithic and Bronze Age ancestries are heterogeneously distributed among Eurasian populations. Specifically:

Western hunter-gatherer (WHG)-related ancestries are highest in present-day individuals from the Baltic States, Belarus, Poland and Russia; Eastern hunter-gatherer (EHG)-related ancestries are highest in Mongolia, Finland, Estonia and Central Asia; and Caucasus hunter-gatherer (CHG)-related ancestries are highest in countries east of the Caucasus, in Pakistan, India, Afghanistan and Iran [...]. The CHG-

related ancestries probably reflect affinities to both CHG and Iranian Neolithic individuals [...]. Consistent with expectations, Neolithic Anatolian-related farmer ancestries are concentrated around the Mediterranean basin, with high levels in southern Europe, the Near East and North Africa, including the Horn of Africa, but are less frequent in Northern Europe. This is in direct contrast to the Steppe-related ancestries, which are found in high levels in northern Europe, peaking in Ireland, Iceland, Norway and Sweden and decreasing further south. [8]

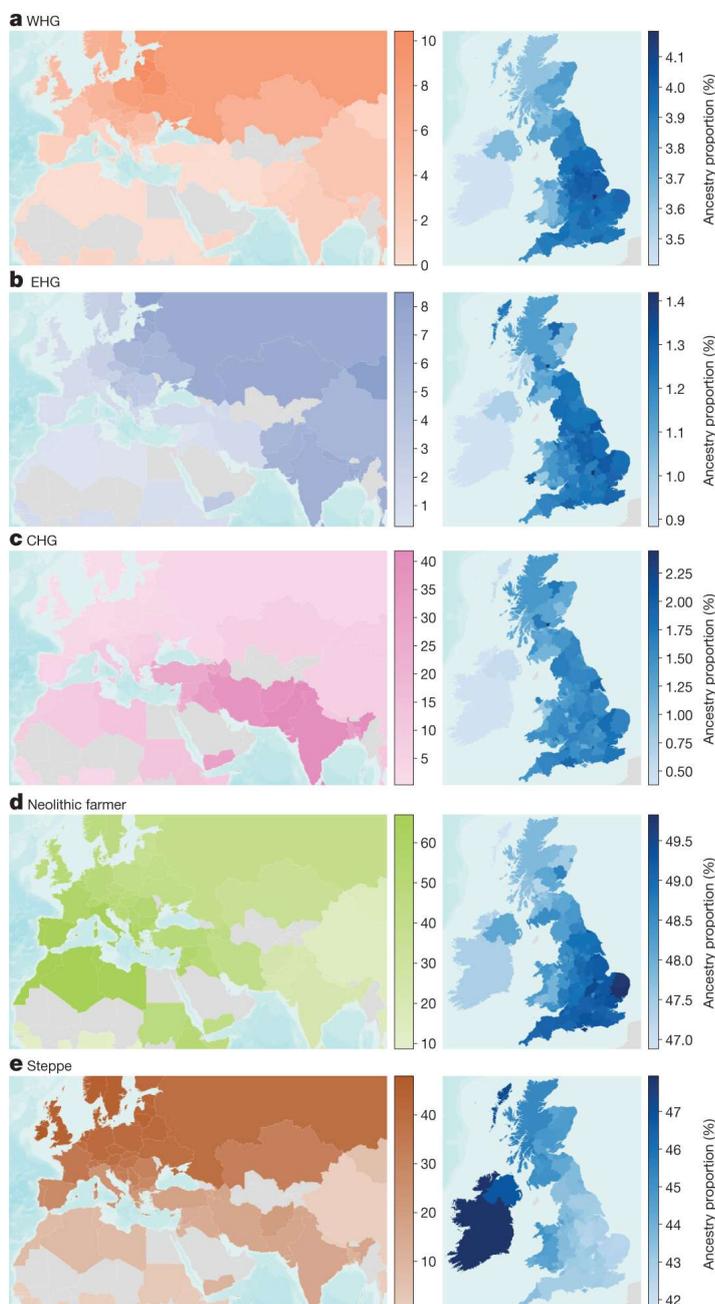


Fig 9 The genetic legacy of ancient Eurasian ancestries in present-day populations. a–e, Maps showing the average ancestry of: WHG (a); EHG (b); CHG (c); Neolithic farmer (d); and Steppe pastoralist (e) ancestry components per country (left) and per county or unitary authority within Great Britain and per country for the Republic of Ireland and Northern Ireland (right). Estimation was performed using ChromoPainter and NNLS², on samples of a ‘typical ancestral background’ for each non-UK country (n = 24,511) and Northern Ireland. For Great Britain, an average of self-identified ‘white British’ samples was used to represent each UK county and unitary authority, based on place of birth (n = 408,884). Countries with less than 4 and counties with less than 15 samples are shown in grey. Map uses ArcGIS layers World Countries Generalized and World Terrain. [A. E. Page *et al.* 2016]

Different admixture proportions and ancestries distribution are responsible for different phenotypes among European extant populations. As regards for phenotypes not linked to risk of diseases: for many anthropometric traits the ARS for Steppe ancestry was the highest, followed by EHG and CHG/WHG; scores for hair and skin pigmentation were higher for WHG, EHG and CHG compared to Neolithic farmer and Steppe-associated ancestries [8]. In terms of genetic contributions to risk for diseases the WHG ancestral component had strikingly high ancestral risk scores (ARS) for traits related to cholesterol, blood pressure and diabetes [8].

Also, it was found that ApoE4 allele (which increases risk of Alzheimer's disease) was primarily associated with WHG/EHG haplotypic background, suggesting it was brought into Western Eurasia by early hunter-gatherers. These data are «in line with the present-day European distribution of this allele, which is highest in northeastern Europe, where the proportion of these ancestries is larger than in other regions of the continent» [8]. On the other hand, ApoE2 (which decreases the risk of Alzheimer's disease) was not detected in WHG and EHG: the decreased frequency of this allele might be the result of genetic drift, but it is unlikely that both WHG and

² Non-Negative Least Squares (NNLS) is a mathematical optimization method that only allows the coefficients to be zero or positive. It is used in populations genetics to estimate genetic ancestry proportions in a population without allowing negative values in the estimates – which is essential, as a proportion cannot be negative.

EHG were independently affected by the same selective pressures, while it is much more likely that ApoE4 was more advantageous for these populations [17]. ApoE2 allele seems to have been introduced to the European gene pool after the arrival of early farmers and Steppe populations [17]. Indeed, this allele has been found on an haplotypic background with affinities to Steppe pastoralists, and probably the selective forces favouring ApoE2 in these populations are associated with protective immune responses against infectious disease [8].

6. METHODS

This work has been redacted starting from reading the article The selection landscape and genetic legacy of ancient Eurasians (E. K. Irving-Pease *et al.*, 2024). Once I decided which were the topic I wanted to explore further, I thought of a main thread that could confer cohesiveness and consistency to my research.

I decided to focus on the onset of Neolithic cultural and economic transition, which were the immediate consequences and which the traces still visible nowadays. In particular, I tried to understand how these changes impacted the selective pressures on diet- and immune associated loci and for what reasons.

First of all, I thought to be essential to give the context of the studied selection events, which led me to search for articles with archaeological and genetic records explaining when the agricultural revolution appeared and how it spread across Europe. So, I consulted databases such as NCBI and Google Scholar to find some of the references cited in the starting article and some articles of my discretion. Doing so I found M. Furholt (2021), J. Fort and J. Pérez-Losada (2024), M. F. Deguilloux, R. Leahy, M. H. Pemonge, and S. Rottier (2012) which I thought to be adequate and exhaustive for my purposes.

Then, I tried to understand what have been the first consequences of the onset of agriculture on the first farmers in terms of demographic and mortality responses. In this regard J.-P. Bocquet-Appel (2008), S. Marciniak *et al.* (2025), G. J. Armelagos, A. H. Goodman, and K. H. Jacobs (1991) and A. E. Page *et al.* (2016) works seemed appropriate.

Once I was satisfied with the found contextual information, I focused on three main signs of selection found in the human genome, which are: selection of LP allele at MCM6/LCT locus, on FADS cluster and on HLA region. These regions show adaptation to changes in diet, exposure to pathogen and to famine, three among the several

consequences of Neolithization. The articles used to collect the reported information were once again chosen consulting NCBI and Google Scholar. I selected S. Mathieson and I. Mathieson (2018), E. K. Irving-Pease *et al.* (2024), J. Fort and J. Pérez-Losada (2024), L. Saag (2020), Y. Itan, A. Powell, M. A. Beaumont, J. Burger, and M. G. Thomas (2009), R. P. Evershed *et al.* (2022), M. T. Buckley *et al.* (2017), M. Fumagalli *et al.* (2009) and A. T. M. Saeb and D. Al-Naqeb (2016) that gave me a general yet thorough picture of the selective processes behind these adaptations.

Finally, I tried to understand how Mesolithic and Neolithic ancestries are distributed in extant populations and if the local ancestry and proportion of admixture contribute to risk of disease, in particular I focused on ApoE4 and ApoE2 alleles as example. To do so I used I. Lazaridis *et al.* (2014), E. R. Jones *et al.* (2015), E. K. Irving-Pease *et al.* (2024) and D. Kolbe *et al.* (2023) as sources of information.

7. CONCLUSIONS

The agricultural revolution has been a watershed and a crucial chapter of human history, whose outcomes had far-reaching consequences that persist in modern populations. The novelty of Neolithization slowly spread across Europe through several processes of migration and cultural exchange that led to the appearance of socially and culturally highly diversified populations. The gradual yet profound upheaval of eating habits, social and settlement structures, of hygienic conditions, of relationships with animals and the environment have led to a cascade of adaptive responses as the consequence of new and never-before-challenged selective pressures.

Sedentism allowed food storage and diminished energy expenditure compared to a mobile lifestyle that led to a demographic increase with the so-called Neolithic Demographic Transition. Yet the higher exposure to pathogens, crop failure and close contact to cattle brought rising mortality. Reproductive trade-offs seem to have reconciled the paradox of both increasing mortality and population growth, with parents allocating resources to multiple lower quality offspring rather than a few higher-quality ones.

These profound changes can be detected at genomic level thanks to signs of selection associated to adaptive responses to new environmental pressures. In particular, the increase in frequency of LP phenotype, despite what was commonly accepted as true, would not be associated with the consumption of milk throughout adulthood per se but rather with the detrimental consequences of its consumption in association with famine and pathogen exposure. On the other hand, changes in the diet following the agricultural revolution seem to be responsible for adaptation at the FADS cluster level. This would be linked to a more vegetarian diet and so to different ratio of SC-PUFAs and LC-PUFAs. As regards for increased exposure to pathogens, it also led to a number of adaptation events in the HLA region, often associated with genetic trade-offs. Yet, HLA region is highly pleiotropic

thus it is challenging to identify the specific selective forces responsible for the frequency increases across different historical periods. Also, adaptation to pathogen might be considered responsible for today prevalence of autoimmune disease.

Extant populations distribution of phenotypes is, indeed, strictly linked to ancestries and their genotypic contribute. Mesolithic, neolithic and Bronze Age ancient populations have contributed heterogeneously to the genetic identity of present-day Europe, of which it has been possible to infer not only the local ancestry and admixture proportions, but also how they contribute to today risk of disease and if they represent a valid explanation for the distribution of alleles associated with diseases such as Alzheimer.

These data suggest that studying the ancient legacy, through molecular analysis of modern and ancient DNA and using tools of populations genetics, might be a crucial part of the complex and multilayered process of understanding who we are today.

BIBLIOGRAPHY

- [1] M. Furholt, "Mobility and Social Change: Understanding the European Neolithic Period after the Archaeogenetic Revolution," *Journal of Archaeological Research*, vol. 29, no. 4, pp. 481–535, Dec. 2021, doi: 10.1007/s10814-020-09153-x.
- [2] J. Fort and J. Pérez-Losada, "Interbreeding between farmers and hunter-gatherers along the inland and Mediterranean routes of Neolithic spread in Europe," *Nat Commun*, vol. 15, no. 1, Dec. 2024, doi: 10.1038/s41467-024-51335-4.
- [3] M. F. Deguilloux, R. Leahy, M. H. Pemonge, and S. Rottier, "European Neolithization and Ancient DNA: An Assessment," *Evol Anthropol*, vol. 21, no. 1, pp. 24–37, Jan. 2012, doi: 10.1002/evan.20341.
- [4] S. Marciniak *et al.*, "An integrative skeletal and paleogenomic analysis of stature variation suggests relatively reduced health for early European farmers," 2025, doi: 10.1073/pnas.
- [5] J.-P. Bocquet-Appel, "The Neolithic Demographic Transition, Population Pressure and Cultural Change," 2008. [Online]. Available: <https://scholarsarchive.byu.edu/ccr> Available at: <https://scholarsarchive.byu.edu/ccr/vol58/iss58/6>
- [6] G. J. Armelagos, A. H. Goodman, and K. H. Jacobs, "The Origins of Agriculture: Population Growth During a Period of Declining Health," 1991.
- [7] A. E. Page *et al.*, "Reproductive trade-offs in extant hunter-gatherers suggest adaptive mechanism for the Neolithic expansion," *Proc Natl Acad Sci U S A*, vol. 113, no. 17, pp. 4694–4699, Apr. 2016, doi: 10.1073/pnas.1524031113.

- [8] E. K. Irving-Pease *et al.*, “The selection landscape and genetic legacy of ancient Eurasians,” *Nature*, vol. 625, no. 7994, pp. 312–320, Jan. 2024, doi: 10.1038/s41586-023-06705-1.
- [9] A. T. M. Saeb and D. Al-Naqeb, “The Impact of Evolutionary Driving Forces on Human Complex Diseases: A Population Genetics Approach,” 2016, *Hindawi Limited*. doi: 10.1155/2016/2079704.
- [10] L. Saag, “Human Genetics: Lactase Persistence in a Battlefield,” Nov. 02, 2020, *Cell Press*. doi: 10.1016/j.cub.2020.08.087.
- [11] R. P. Evershed *et al.*, “Dairying, diseases and the evolution of lactase persistence in Europe,” *Nature*, vol. 608, no. 7922, pp. 336–345, Aug. 2022, doi: 10.1038/s41586-022-05010-7.
- [12] M. T. Buckley *et al.*, “Selection in Europeans on fatty acid desaturases associated with dietary changes,” *Mol Biol Evol*, vol. 34, no. 6, pp. 1307–1318, Jun. 2017, doi: 10.1093/molbev/msx103.
- [13] S. Mathieson and I. Mathieson, “FADS1 and the timing of human adaptation to agriculture,” *Mol Biol Evol*, vol. 35, no. 12, pp. 2957–2970, Dec. 2018, doi: 10.1093/molbev/msy180.
- [14] M. Fumagalli *et al.*, “Parasites represent a major selective force for interleukin genes and shape the genetic predisposition to autoimmune conditions,” *Journal of Experimental Medicine*, vol. 206, no. 6, pp. 1395–1408, Jun. 2009, doi: 10.1084/jem.20082779.
- [15] I. Lazaridis *et al.*, “Ancient human genomes suggest three ancestral populations for present-day Europeans,” *Nature*, vol. 513, no. 7518, pp. 409–413, Sep. 2014, doi: 10.1038/nature13673.

- [16] E. R. Jones *et al.*, "Upper Palaeolithic genomes reveal deep roots of modern Eurasians," *Nat Commun*, vol. 6, Nov. 2015, doi: 10.1038/ncomms9912.
- [17] D. Kolbe *et al.*, "Current allele distribution of the human longevity gene APOE in Europe can mainly be explained by ancient admixture," *Aging Cell*, vol. 22, no. 5, May 2023, doi: 10.1111/accel.13819.