

UNIVERSITY OF TARTU
FACULTY OF SCIENCE AND TECHNOLOGY
INSTITUTE OF MOLECULAR AND CELL BIOLOGY
INSTITUTE OF GENOMICS
CENTRE FOR GENOMICS, EVOLUTION AND MEDICINE

**Metabolic adaptations reflect increased dependence on an aquatic diet in the Mesolithic-
Neolithic transition of Eastern Europe**

Bachelor's thesis

12 EAP

Kaur Robin Kivirüüt

Supervisor Jon Anders Eriksson (PhD)

TARTU 2023

INFO SHEET

Metabolic adaptations reflect increased dependence on an aquatic diet in the Mesolithic-Neolithic transition of Eastern Europe

Changes in environment and lifestyle have shaped the evolution of human metabolism in unknown ways. In this thesis, genetic information from human remains (aDNA) was combined with estimates of the effect of genetic variants on selected metabolic traits from a genome-wide association study (GWAS) in order to calculate polygenic scores (PGS) that could be compared between populations in different environments and time periods. Analysis of ancient populations showed evident signs of selection among hunter-gatherers. A hypothesis is proposed about its connection with changes in lifestyle based on archaeological and biochemical evidence.

Keywords: ancient DNA, metabolites, polygenic score, Mesolithic, Neolithic, aquatic diet

CERCS: B110 Bioinformatics, medical informatics, biomathematics, biometrics

Metaboolsed kohastumused peegeldavad suurenenud toitumist veeorganismidest Ida-Euroopa mesoliitikumi lõpus

Muutused keskkonnas ja elustiilis on teadmata viisidel kujundanud inimeste ainevahetuse arengut. Käesolevas bakalaureusetöös kombineeriti inimsäilmetest leitud geneetiline informatsioon (iidne DNA) ja genoomi hõlmavate assotsiatsiooniuuringute (GWAS) põhjal saadud hinnangud valitud metaboolsete tunnuste geneetiliste variantide mõju kohta, et arvutada polügeneetilised skoorid (PGS) mida saab võrrelda populatsioonide vahel erinevates keskkondades ja ajaperioodidel. Analüüs iidsetest rahvastikest näitas selgeid märke looduslikust valikust küttide ja korilaste seas. Arheoloogiliste ja biokeemiliste tõendite põhjal püstitatakse hüpotees selle toimumise kohta seoses elustiili muutustega.

Märksõnad: iidne DNA, metaboliidid, polügeneetiline skoor, Mesoliitikum, Neoliitikum

CERCS: B110 Bioinformaatika, meditsiiniinformaatika, biomatemaatika, biomeetrika

TABLE OF CONTENT

TABLE OF CONTENT	3
ABBREVIATIONS.....	4
INTRODUCTION	5
1. LITERATURE REVIEW.....	6
1.1. Detecting adaptation to diet	6
1.2. Neolithic transition in Eastern Europe	7
2. EXPERIMENTAL PART.....	9
2.1. Aims of this thesis.....	9
2.2. Materials and methods	9
2.2.1. Polygenic scores for metabolites.....	9
2.2.2. Ancient DNA	11
2.2.3. Principal component analysis of genome-wide genetic variation	12
2.2.4. Population cluster analyses.....	12
2.2.5. Polygenic score calculation and analysis.....	13
2.2.6. Evolutionary analysis	15
2.3. Results.....	15
2.3.1. aDNA population clusters.....	15
2.3.2. Metabolite PGS clusters	17
2.3.3. aDNA PGS clusters	18
2.3.4. Temporal analysis.....	19
2.3.5. Selection analysis in Eastern Europe.....	20
2.3.6. Pleiotropic SNPs.....	25
2.4. Discussion	27
2.4.1. Pathway analysis.....	27
2.4.2. Archaeological and stable isotope evidence.....	28
2.4.3. Future work	30
SUMMARY	31
RESÛMEE	32
REFERENCES.....	33
USED WEBSITES	42
EXTENDED FIGURES.....	43

ABBREVIATIONS

aDNA – Ancient deoxyribonucleic acid (DNA)

ALA – Alpha-linolenic acid

Apo-B – Apolipoprotein B

BP – Before present

CCC – Comb Ceramic culture

EHG – Eastern Hunter-Gatherer

EPA – Eicosapentaenoic acid

GWAS – Genome-wide association study

HCA – Hierarchical cluster analysis

HDL – High-density lipoprotein

IDL – Intermediate-density lipoprotein

LC-PUFA – Long-chain polyunsaturated fatty acid

LD – Linkage disequilibrium

LDL – Low-density lipoprotein

LDL-C – Low-density lipoprotein cholesterol

LNA – Linolenic acid

MG – Metabolite group

PCA – Principal component analysis

PGS – Polygenic score

PUFA – Polyunsaturated fatty acid

SC-PUFA – Short-chain polyunsaturated fatty acid

SNP – Single-nucleotide polymorphism

SVD – Singular value decomposition

VLDL – Very-low-density lipoprotein

INTRODUCTION

In the course of human evolution, changes in culture and environment have introduced new selective pressures, that have in turn driven the adaptive evolution of specific genetic changes in response. Many of these selective events were driven by changes in diet (Hancock, Witonsky, et al., 2010) or exposure to pathogens (Fumagalli et al., 2011). Some of these past adaptations are now maladaptive (Klunk et al., 2022), while most are still undiscovered, as subtle shifts in allele frequencies in the past are challenging to detect in present-day genomes. Modern genetic diversity is therefore a reflection of past changes that have acted to adapt different populations to their environments (Balaesque et al., 2007). However, ancient DNA makes it possible to examine these past populations in a new light by observing genetic variation over time and thus witnessing selection directly. One of the most influential dietary changes in human prehistory was the adaptation of farming, which fundamentally changed subsistence strategies throughout the world (Luca et al., 2010). Radical changes in available food sources and diet undoubtedly had a direct effect on metabolism and other molecular traits, setting the stage for natural selection on genetic variation underpinning individual variation in these traits.

The aim of this thesis is to find out what kind of changes have occurred in the metabolic profiles of European populations in the last 10,000 years, and test the hypothesis that genetic influences on metabolic traits have evolved to counterbalance changes in the intake of key metabolites or their molecular precursors as the result of changes in diet.

This thesis was carried out at the Institute of Genomics, Centre for Genomics, Evolution and Medicine (cGEM).

1. LITERATURE REVIEW

1.1. Detecting adaptation to diet

Throughout human evolution, there have been many fundamental changes in lifestyle and population structure. Changes in environment and available food sources have had a direct effect on the subsistence strategies of early people, resulting in major dietary shifts (Luca et al., 2010). These changes likely created strong selective pressures and set the stage for natural selection on metabolic traits. Some notable examples of adaptation to diet include lactase persistence (Gerbault, 2013) and fatty acid metabolism (Ameur et al., 2012). The mentioned traits are controlled by variation at a single loci with strong effects. In contrast, complex traits (e.g., height, skin pigmentation, etc.) are polygenic, having thousands of alleles distributed across the genome that each contribute a small part to heritable variance (Sella & Barton, 2019; Visscher et al., 2017). Recently, genome-wide association studies have provided new insights into the genetic basis of complex traits, showing that they are mainly driven by regulatory variants and that even the most important loci only explain a limited amount of the predicted genetic variance (Boyle et al., 2017).

The standard approach (commonly called the “hard sweeps” model) for detecting selection in populations has been trying to identify mutations that spread rapidly to fixation (Bersaglieri et al., 2004). As such, this method mainly detects variants with large phenotypic effects (which are rare) and not changes in standing variation (i.e., polymorphic alleles that are already present within the population) (Pritchard et al., 2010). Variants with small phenotypic effects may not show signatures of selection on a larger scale because of the strong effect population structure has on variation in allele frequencies (Hancock, Alkorta-Aranburu, et al., 2010). New methods for identifying polygenic adaptation, dubbed “soft sweeps” (Hermisson & Pennings, 2005), are instead based on combining signals of changes in allele frequency across many loci that affect a given trait and testing whether these changes tend to affect the trait in a given direction (Berg & Coop, 2014; Sella & Barton, 2019). These methods of detecting polygenic adaptation are now also being applied to ancient DNA in order to find selection happening in the past (Irving-Pease et al., 2021).

1.2. Neolithic transition in Eastern Europe

The most recent major shift in diet is associated with the domestication of plants and animals. The adaptation of agriculture and its effects, both positive and negative, have been well studied (Childebayeva et al., 2022; I. Mathieson et al., 2018). However, the Mesolithic-Neolithic transition was not uniform across Europe (**Figure 1**). Genetic evidence has shown a distinct East-West division along a boundary zone running from the Black Sea to the Baltic (Allentoft et al., 2022; Posth et al., 2023). On the western side of this divide, the Neolithic expansion was driven by demic diffusion of Anatolian farmers (I. Mathieson et al., 2015; Omrak et al., 2016). In contrast, hunter-gatherers east of the divide prevailed with no input from Anatolian farmers (Jones et al., 2017; Mittnik et al., 2018). Here (mainly the Baltics and Eastern Ukraine), the local foraging lifestyle continued without any large changes in material culture or subsistence economy (M. Lillie & Budd, 2011). The Mesolithic and Neolithic populations of the Baltics and Ukraine have also been shown to be genetically similar (Jones et al., 2017), with varying proportions of both Western (originating from Europe) and Eastern Hunter-Gatherer (originating from regions of present-day Russia) ancestry (Mittnik et al., 2018).

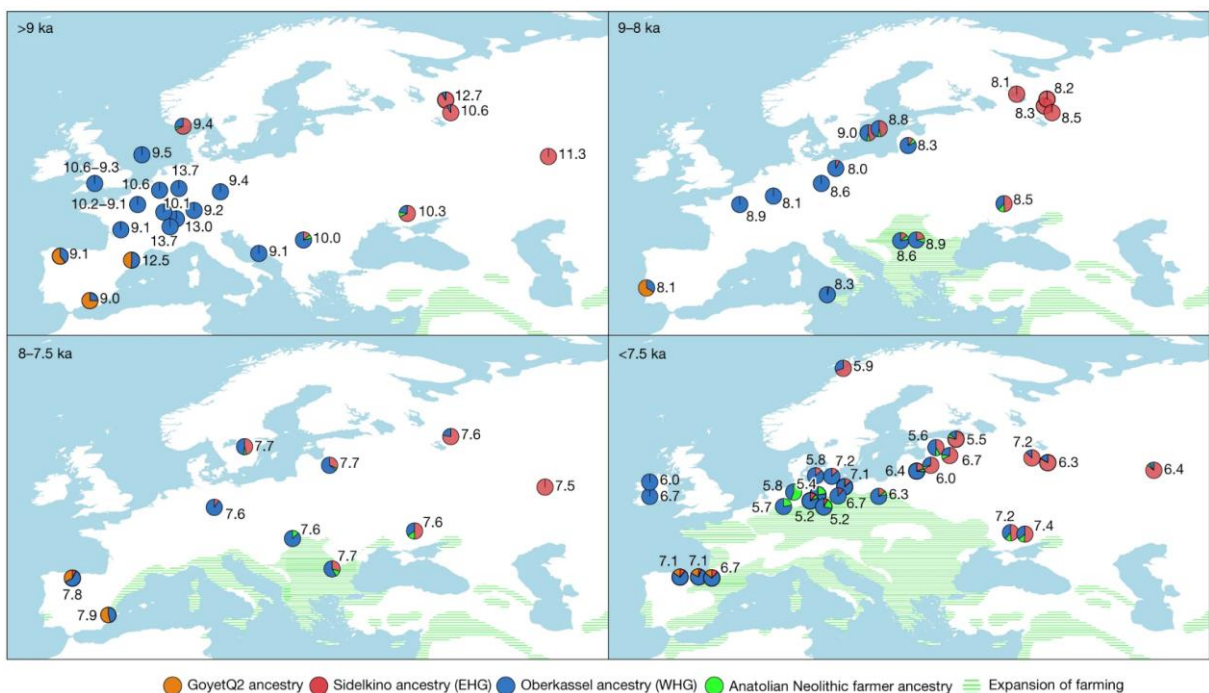


Figure 1. The genetic ancestry of hunter-gatherers dated between 14,000 BP and 5200 BP. The average calibrated date is shown, with pie charts indicating the estimated proportion of ancestry for each group or individual. The expansion of farming is shown as green shades. Figure is taken from (Posth et al., 2023).

With the absence of any major socioeconomic transformations (i.e., agriculture or domestic animal husbandry), the beginning of the Neolithic in North-Eastern Europe has historically been equated to the first appearance of pottery (Motuzaitė Matuzevičiūtė, 2013; Nordqvist & Kriiska, 2015). Pottery started spreading into North-Eastern Europe ca. 8000 before present (BP) from the northern shore of the Caspian Sea, having its origins even further east (Dolbunova et al., 2022); from there it rapidly spread north and westwards through cultural diffusion (i.e., by transmission from person to person rather than long-range movement of individuals), reaching the Baltic region around 7500-7200 BP (Kriiska et al., 2017). In fact, pottery made such a silent impact that this definition of the Neolithic has been brought into question, with the first ceramic cultures (i.e., Narva culture) being dubbed as “pottery-Mesolithic” (Kriiska, 2009). Still, it is acceptable to refer to the Neolithic as a period with pottery (Nordqvist & Kriiska, 2015), as it will be done in this thesis.

That being said, the Neolithic did not truly begin in Estonia (and neighbouring areas) until around 6000 BP (Nordqvist & Kriiska, 2015), coinciding with a sudden influx of people into the Baltic area from the north-east (Saag et al., 2017). This brought along changes in material culture and is associated with the first signs of crop cultivation and the transition of the earlier Narva culture into the Comb Ceramic culture; still, the hunter-fisher-forager lifestyle persisted and continued to be the main source of substance (Kriiska et al., 2020). Farming did not become the main source of subsistence until the Late Neolithic, with migrations from the Steppe ca. 5000 BP that brought the Corded Ware culture into this region (Kriiska et al., 2020; Saag et al., 2017). As a result of local foraging groups persisting and mixing with farmers until the Bronze Age, modern-day Eastern Baltic individuals show the highest proportion of hunter-gatherer ancestry (Lazaridis et al., 2014; Mitnik et al., 2018).

Due to the fact that farming had no clear beginning in Eastern Europe, the genetic adaptations resulting from it are understudied in evolutionary analyses when compared to Western Europe.

2. EXPERIMENTAL PART

2.1. Aims of this thesis

The overall goal of this thesis is to identify selection in loci relating to metabolic traits and, if possible, determine the driving mechanisms behind it. To this end, the objectives of the thesis are as follows:

1. Establish which ancient populations to analyse using available aDNA data.
2. Estimate polygenic scores for the ancient populations using a genome-wide association study to reflect changes in metabolic traits.
3. Analyse genome-wide and metabolic genetic signatures (principal components) for ancient populations and identify clusters with similar metabolic signatures.
4. See if significant changes occur over time, and test if these changes are a result of natural selection.
5. Determine the biological mechanisms selection has acted on and investigate possible causes of selection, if any, by comparing results to archaeological and biochemical data.

To follow these objectives, I will analyse genetic information gathered from present-day individuals and ancient human remains (aDNA). Using principal component analysis (PCA) and archaeological contexts, these individuals will be clustered together to represent distinct European and Near Eastern populations. A genome-wide association study (GWAS) is then used to estimate the effect of genetic variants on selected metabolic traits within these populations. To quantitatively measure the relative amount of expression of chosen traits, a polygenic score (PGS) is calculated. Finally, these polygenic scores will be used to compare different populations and assess changes in metabolic traits over time to find when and where selection has occurred.

2.2. Materials and methods

2.2.1. Polygenic scores for metabolites

Genome-wide association studies (GWAS) try to find connections between single-nucleotide polymorphisms (SNPs) and values for a given phenotypic trait (Marees et al., 2018). To find these connections, data for genetic markers across the genome are generated (by sequencing or genotyping using chip arrays) from DNA collected from individuals in a study

population. If a given allele (commonly denoted the risk or effect allele) occurs more frequently for certain trait values than others, then a statistical correlation can be established (using, e.g., the Pearson correlation statistic or linear regression), and it can be said that the SNP is associated with the trait (Hirschhorn & Daly, 2005; Uffelmann et al., 2021). These kinds of studies are becoming more prevalent as developments in technology allow for faster and easier whole genome sequencing and genome-wide genotyping, which increases the sample size and allows for more SNP detection (Loos, 2020; Visscher et al., 2017).

Metabolic traits are complex, resulting from the interplay of large numbers of genes, with hundreds or thousands of SNPs associated with variation in the trait across the genome. Each SNP typically explains only a small proportion of the variation in such traits, and methods to aggregate the effect of SNPs are required to predict the trait (Marees et al., 2018). In this work, I will be using polygenic score analysis, as it is relatively easy to conduct and does not require a large sample size (Dudbridge, 2013). A polygenic score (PGS) combines the effect sizes of multiple SNPs into a single aggregated value (score) that can be used as an indication of the trait level for a specific individual (Dudbridge, 2013; Purcell et al., 2009; Wray et al., 2007). As such, PGS can be used to predict an individual's genetic predisposition to a disease or complex traits (Choi et al., 2020; Dudbridge, 2016). Greater predictive power can be achieved by using a larger number of samples and by including SNPs that did not reach genome-wide significance in the GWAS but are nevertheless informative (Purcell et al., 2009).

PGS based on a GWAS in one population can have reduced predictive accuracy when applied to individuals in different populations for three main reasons: 1) populations may have different combinations of genetic variants affecting the same set of causal genes; 2) GWAS may only identify common genetic variants that happen to be co-inherited (in linkage disequilibrium, LD) with the causal variant in the discovery population, but this association may be weaker in populations with different ancestry; and 3) variation in environmental and gene-environment interactions that indirectly affect the estimated effect sizes of genetic variants (I. Mathieson, 2021; Reisberg et al., 2017). How well traits transfer between populations (and why they do not) vary a lot depending on the trait. For example, blood lipid traits transfer reasonably well between Europe and East Asia, but not so well between Europe and Africa (Kuchenbaecker et al., 2019). Individuals from the same population, but

different from the discovery population, can be compared to each other (although with less accuracy) as they draw from the same allele distribution (I. Mathieson, 2021).

The association values and effect sizes for the SNPs used in this thesis come from an already established GWAS of 123 human blood lipid and metabolite concentrations quantified by nuclear magnetic resonance spectroscopy (Kettunen et al., 2016). The GWAS cohort consists of 24,925 individuals from Northern Europe and includes up to 12,133,295 SNP positions per metabolic trait. This GWAS was chosen because it a) is one of the largest to date, b) has a discovery population from North-East Europe (and is therefore geographically close to the region of interest to this thesis, and c) offers complete summary statistics (i.e., variant, effect size, p-value, etc.) for all variable in the study, not only those with significant associations.

This data was downloaded from an open GWAS repository

(<https://gwas.mrcieu.ac.uk/datasets/>, GWAS ID: met-c) (Lyon et al., 2021).

The metabolites within the GWAS (see **Extended Table 2** for a full list) consist of lipids (fatty acids, cholesterol, phospholipids, and triglycerides), lipoproteins (chylomicrons, VLDL, LDL, HDL, and IDL particles) of various sizes and densities, some amino acids, and other small molecules (e.g., apolipoproteins, citrate, glucose, urea) that play key roles in different metabolic cycles. Also included are some derived variables, like ratios of measured metabolite concentrations, often used for characterizing metabolic functions. These metabolites are described in more detail in (Kettunen et al., 2012). A substantial portion of the measured traits are related to lipoproteins. Lipoproteins are colloidal particles comprised of various lipids and apolipoproteins that transport insoluble dietary lipids (i.e., triglycerides and cholesterol) throughout the body (German et al., 2006).

2.2.2. Ancient DNA

The human genotypes were obtained from a set compiled and curated by the Allen Ancient DNA Resource (Mallick et al., 2023). AADR offers an expanding collection of published genotypes from ancient human fossils and present-day individuals, along with additional information such as geographic location, estimated age of the fossil, sequencing quality, and archaeological contexts (e.g., designated archaeological culture). The dataset used in this thesis (<https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>; Version 50.0) consisted of 6442 ancient fossils

and 3937 present-day individuals, represented by more than one million SNPs across the genome.

2.2.3. Principal component analysis of genome-wide genetic variation

Principal component analysis (PCA) is a simple way to reduce the dimensionality of large datasets while preserving as much variability as possible and is commonly used in genetics to capture and visualize population structure (Jolliffe & Cadima, 2016; Patterson et al., 2006). Population genetic analysis of ancient genomes is challenged by large and variable amounts of missing data. Approaches such as EMU (Meisner et al., 2021), based on self-consistent estimation of principal components (PCs), have been shown to capture the main PCs with minimal bias for coverage as low as 1%. Here, the lead 20 PCs were estimated with a fixed-point method, implemented as a Matlab (v9.13) script developed by the supervisor (Dr Jon Anders Eriksson). First, PLINK (PLINK 1.9, <https://www.cog-genomics.org/plink/>) (Chang et al., 2015) was used to identify unlinked loci (with flags `--geno 0.4 --maf 0.05 --indep 50 5 2`). For the PCA, samples from AADR were restricted to Eurasia (and Arctic North America). The mean frequencies for each locus were estimated and then used to predict allele dosage at missing loci. Then, an allele dosage variance-covariance matrix between individuals was calculated. This was used to estimate the singular value decomposition (SVD) for the top 20 singular values (and thereby also the top 20 PCs). Using the SVD approximation (restricted to the interval $[0,1]$), allele frequencies at loci with missing information were again estimated, and the SVD for the top 20 singular values was re-calculated. This process was repeated until the PCs converged.

2.2.4. Population cluster analyses

Using geographic location, estimated age, archaeological contexts, and position in PC space, I manually assigned individuals to different groupings, corresponding to ancient and present-day European and Near Eastern populations. Firstly, I defined the study area of Europe and the Near East by geographic location, going as far east as Israel. For samples originating from the area of modern-day Russia, only individuals west of the 40°E line of longitude were included. Secondly, I divided the derived individuals into eight time periods based on their estimated age and assigned archaeological contexts. Then further divided them into smaller

clusters by their position in PC space. The first three PCs were used for this. For large clusters that were spread out in PC space with no clear signs of grouping, K-means clustering was used to help find the optimal point of separation. K-means is a straightforward clustering algorithm that iteratively reassigns data points to clusters in order to produce clusters with the lowest variance possible (Ikotun et al., 2023). I also took into account the composition of neighbouring time periods to make the clusters more continuous in time. Finally, I removed individuals from each cluster that were found to be outliers, either because of a divergent estimated age or position in PC space (indicating atypical ancestry). It should be noted that clusters in the same time period are not always uniform in terms of age, as the Neolithic started much earlier in the Near East. In the end, 3928 individuals were divided into 39 clusters for further analysis (**Extended Table 1**~~Error! Reference source not found.~~). K-means clustering was also done on the average PCs of population clusters, to identify further groupings (**Figure 2**).

2.2.5. Polygenic score calculation and analysis

Quality control was performed using PLINK (with flags `--mind 0.1 --maf 0.05 --geno 0.3`) on the ancient individuals to identify SNPs with sufficient coverage for analysis. After identifying overlapping SNPs between aDNA data and GWAS hits (matching physical position in the reference genome (hg19) and genetic variants), I removed linkage disequilibrium (LD) by selecting SNPs with the highest effect estimate and removing all other SNPs within 200 kbp (thousand base pairs). From the remaining SNPs (about 9800 to 9900 for any given metabolic trait), only those with an effect estimate above the pre-specified threshold ($P \leq 10^{-3}$) were selected (this threshold was determined by inspecting the Q-Q plot of the GWAS to determine the approximate P-value below which associations were systematically lower than the expected value). This brought the number of associated SNPs down to between 348 and 504 (depending on the metabolite), with some SNPs being associated with multiple metabolic traits.

From the selected SNPs, I calculated a polygenic score (PGS) for each metabolic trait in every population. Usually, a PGS is calculated by computing the sum of risk alleles that an individual has, weighted by the corresponding risk allele effect sizes on the phenotype as estimated by the GWAS (Choi et al., 2020; Dudbridge, 2013). However, because of the high level of missingness in aDNA, a PGS based on an individual genotype would be inaccurate.

More complete genomes would receive higher scores because of more alleles, not their effect sizes. Instead, the PGS was calculated from the average allele frequencies of the entire population cluster. Another thing to consider is that each trait is comprised of a different number of SNPs with varying effect sizes, meaning each trait has a different order of magnitude. In order to make the PGS for different metabolites comparable with each other, each series of PGSs (every PGS corresponding to a specific metabolite in each population) was separately standardised by subtracting the mean and dividing by the standard deviation. This way, differences in PGS capture the variations within a trait instead of the overall difference.

Since dealing with such a high number of individual metabolite measures (123), that are also biologically linked to varying degrees, it was necessary to further group them. To this end, and to further identify clustering between populations and metabolites, respectively, I conducted a PCA, along with K-means clustering, on the PGS values. This worked well for population clusters (**Figure 4**), but not for metabolites. When applying K-means cluster analysis, it is required to specify the number of clusters beforehand. To find the optimal number of population clusters, I used the Elbow method (Syakur et al., 2018). However, because of the high number of metabolites, and the fact that they showed no obvious clustering in PC space, it was difficult to determine the optimal number of clusters for them. Instead, I chose to cluster them hierarchically with Ward's method (Ward, 1963). Similar to K-means, Ward's method minimizes the total within-cluster variance. This way, the optimal number of clusters could be found later by working upwards from the branches. The PGS of a metabolite group (MG) for a given population was taken as the average of all the metabolites within that group (**Figure 3**). Hierarchical cluster analysis (HCA) was also done on population clusters, both on average PCs (**Extended Figure 8**) and PGSs (**Extended Figure 9**).

K-means clustering and HCA were done using the software R (<http://www.r-project.org/>). All plots and figures were created using R packages "ggplot2" (Wickham et al., 2023), "plotly" (Sievert et al., 2022), "heatmaply" (Galili et al., 2018), and "dendextend" (Galili, 2015).

2.2.6. Evolutionary analysis

By plotting the metabolite group PGS against increasing time periods, I was able to observe the change of PGS in genetically similar clusters (as previously determined by K-means and hierarchical clustering) through time. The aim was to find population clusters that remain similar across time for genome-wide markers (i.e., in PC) but show a significant change in genetic loci associated with metabolic traits (i.e., in PGS). Large changes in PGS, but not in PC, would indicate natural selection on those traits. To further confirm that any changes found were constant over time, I constructed a time series analysis. The population clusters were divided into segments by increments of 1000 years, and a new PGS was calculated for each segment (**Figure 5**).

In order to formalize the test for selection, I compared observed PGS changes (between two clusters) with a distribution of expected scores under a null model of no selection. Expected scores were calculated by matching trait-associated SNPs with random SNPs from across the genome with a similar frequency (± 0.01) but not associated with the trait, resampled 100 times to yield independent estimates of the null distribution. The null distributions for all expected scores were approximately normally distributed. Using the null distribution, I obtained p-values for the observed PGS. A statistically significant score (outside the null distribution) would indicate directional selection on genetic variations affecting the trait, due to selection on that trait or genetically correlated traits (Sella & Barton, 2019). Finally, by averaging the null distributions of expected scores and observed PGS changes of all traits within the same metabolite groups (**Figure 7**, **Figure 9**), I also obtained p-values for changes in average level in each metabolite group.

2.3. Results

2.3.1. aDNA population clusters

The full list of population clusters is brought out in **Extended Table 1**, and an overview of time periods is given here. Individuals older than 10,000 BP were considered to be Palaeolithic unless it was stated otherwise. The Mesolithic is comprised of western and eastern hunter-gatherers, divided by the boundary zone running from the Black Sea to the Baltics. The Neolithic shows further separation of these clusters, with the former western hunter-gatherers grouping close to the Anatolian farmers while the eastern hunter-gatherers

remain unchanged. The Chalcolithic period (sometimes referred to as the Eneolithic or Copper Age) is mostly defined by the emergence of new cultural groups (e.g., Corded Ware). Here, we see Eastern Europe start to move towards Western and Central Europe, with the latter also separating from the Near East. The unique cluster of Gotland is also brought out and shows more similarity to earlier hunter-gatherers than to other clusters from this period. In the Bronze Age, the divided eastern and western groups come together, and populations start to have PC positions close to their modern counterparts. In the Iron Age, all the clusters seem to start blending. Due to poor sample size and genotype coverage in this period, clusters are instead formed on a north-to-south gradient basis, with Scythians making a more distinct cultural group. Because of this, Northern and Eastern Europe could not be separated, as was done with neighbouring time periods. This trend also continues into the Medieval period. Due to almost no samples from that region, a Near Eastern cluster could not be brought out. In contrast, there is a very high number of samples from Northern Europe, but high admixture makes identifying smaller clusters difficult without a deeper understanding of cultural contexts. In the Modern period, individuals are clustered together based on their country of origin. Countries without sufficient samples were left out. Following historical trends and distinct ancestries, the regions of Basque and Sardinia are brought out separately from their main countries. Basque shows a close affinity to Iron Age Central Europe, while Sardinia remains largely unchanged since the Neolithic. The mean PCs for all population clusters are shown in **Figure 2**; for a complete view of individuals within clusters and the distribution of clusters in every time period see **Extended Figures 1-7**.

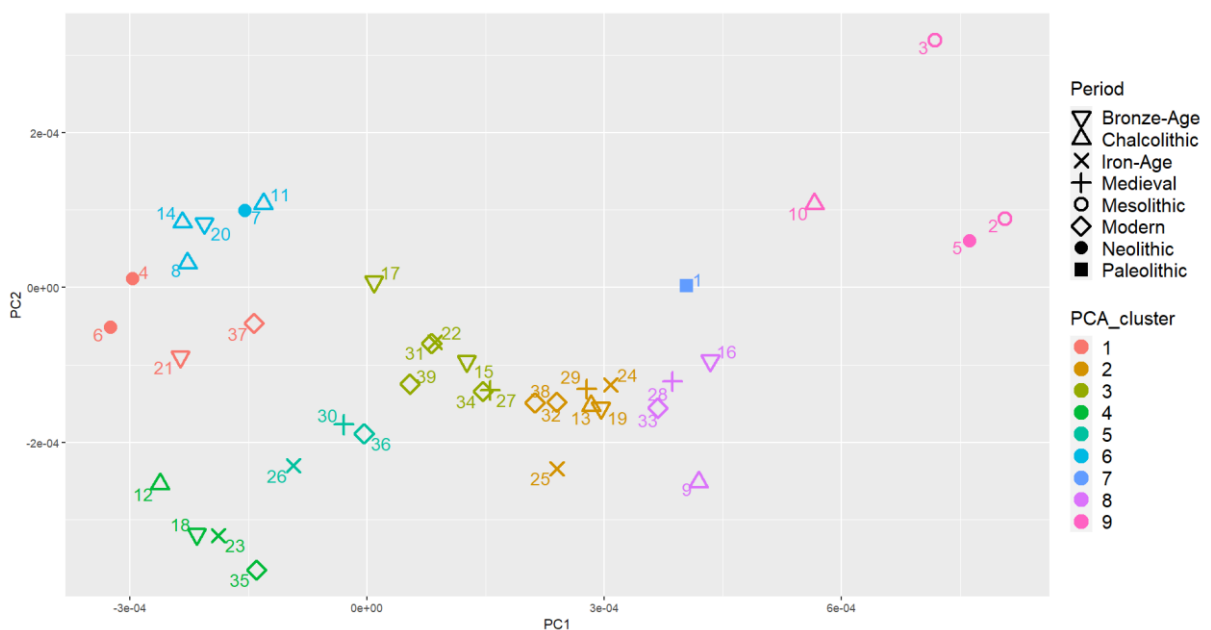


Figure 2. Average PC1 and PC2 for all population clusters. Populations are coloured by clusters determined by K-means clustering. Symbols represent time periods. Numbers refer to the names of population clusters in **Extended Table 1**.

Based on the average PCs of population clusters, the populations were further grouped into similar clusters by using K-means clustering (**Figure 2**). The optimal number of clusters was determined to be 9. Hunter-gatherer populations (PCA_cluster 9), including Chalcolithic_Gotland (nr. 10), are more separated from other post-agrarian populations. Interestingly, while PC1 and PC2 placed the Palaeolithic cluster (nr. 1) closer to the later populations than hunter-gatherers, it was very distinct in PC3 from any other cluster. Also separately from others, but closely together, grouped the post-Neolithic Near Eastern populations (PCA_cluster 4), in almost a straight and continuous line. Neolithic_Near-East (nr. 6) instead grouped towards earlier (Neolithic to Bronze Age) Southern European populations (PCA_cluster 1 and 6), along with Modern Sardinia (nr. 37). Iron-Age_South, Medieval_South, and Modern_Italy grouped together (PCA_cluster 5) and seem to be halfway between the Near Eastern group and newer (Bronze Age to Modern) Southern European populations (PCA_cluster 3). Northern Europe (PCA_cluster 2) changes little from the Bronze Age onward but shows a clear divergence of the Eastern clusters (PCA_cluster 8). HCA showed mostly analogous results to K-means (**Extended Figure 8**). A noticeable difference was that Modern_Finland (nr. 33) and Iron-Age_Scythians (nr. 25) grouped together strongly in HCA but not in K-means.

2.3.2. Metabolite PGS clusters

By applying hierarchical cluster analysis (Ward's method), metabolites with a correlating PGS distribution were clustered together into 11 metabolite groups (MG) (see **Extended Table 2**). Traits relating to one or more similar metabolites (e.g., lipoproteins) seem to have a similar distribution across populations and are often grouped together. This list of strong groupings includes: medium HDL (MG-3); large HDL (MG-9); very large HDL (MG-5); IDL (MG-2); medium and large LDL (MG-10); small and medium VLDL (MG-11); large VLDL (MG-8); chylomicrons and some fatty acids (MG-3). Overall characteristics of lipoproteins (i.e. triglycerides, cholesterol esters, etc.) did not group together.

MG PGS values tend to be higher in Near Eastern populations, and lower in Northern European (**Figure 3**). This could be a reflection of the increasingly different ancestry from the GWAS cohort.

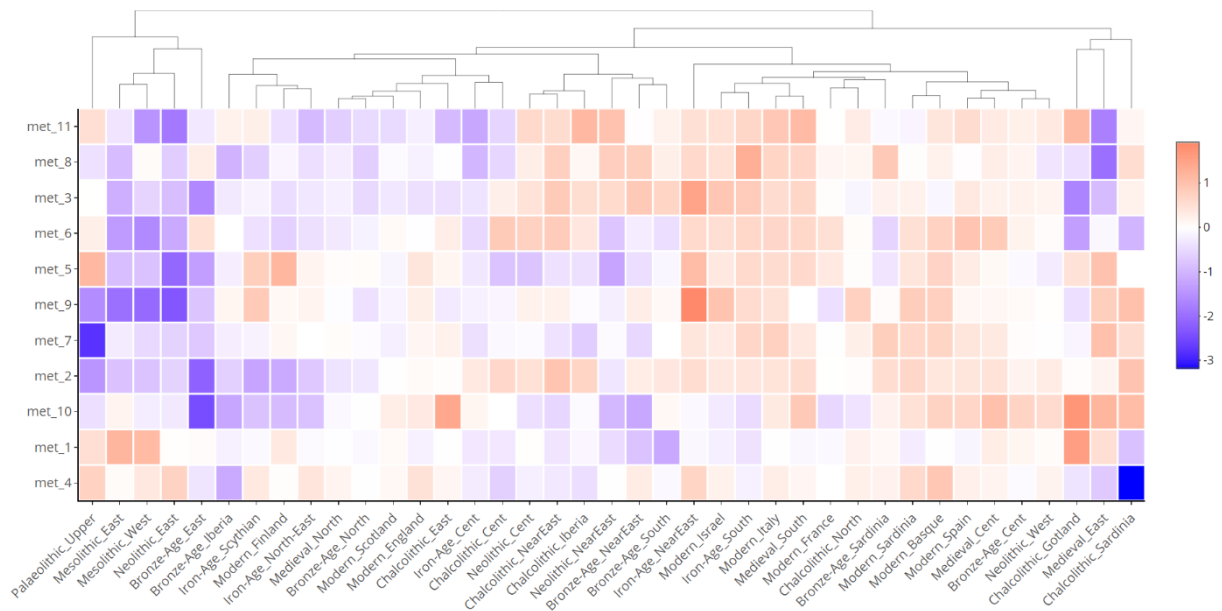


Figure 3. Heatmap of PGS values for metabolite groups across populations. Dendrogram of populations was done automatically and was left in strictly for visual aid (results may not reflect hierarchical clustering based on individual metabolites).

2.3.3. aDNA PGS clusters

Based on PGS, populations show similar, but not identical, K-means clustering to that as genetically (in PCA) (**Figure 4**). In this case, the optimal number of clusters was determined to be 11. Hunter-gatherers again grouped most separately (PGS_cluster 1), but this time Chalcolithic_Gotland (nr. 10) grouped alone. Chalcolithic_Sardinia (nr. 14), Bronze-Age_East (nr. 16), and Medieval_East (nr. 28) also grouped apart. These clusters also stood apart in HCA. HCA of PGS (**Extended Figure 9**) shows a distinct division between north and south, with almost all Northern European populations grouping separately from Southern Europe and the Near East. This can also be seen in the PC with Northern Europe (PGS_cluster 5) and Southern Europe (PGS_cluster 2) being split near the middle of the plot. Indeed, Southern Europe seems more connected with the Near East as Iron-Age_South (nr. 26) is now clustered with Near Eastern populations (PGS_cluster 4), and Bronze-Age_South (nr. 21) is placed close by. This is likely due to population migrations between the Mediterranean and the Near East during that time (Antonio et al., 2019). Interestingly, Bronze-Age_Iberia (nr. 17) grouped with Finland and the Scythians (PGS_cluster 6), as well as Iron-Age_North-East

(nr. 24), suggesting an eastern influence. The populations of Eastern Europe (PCA_cluster 8) are now very separate from each other.

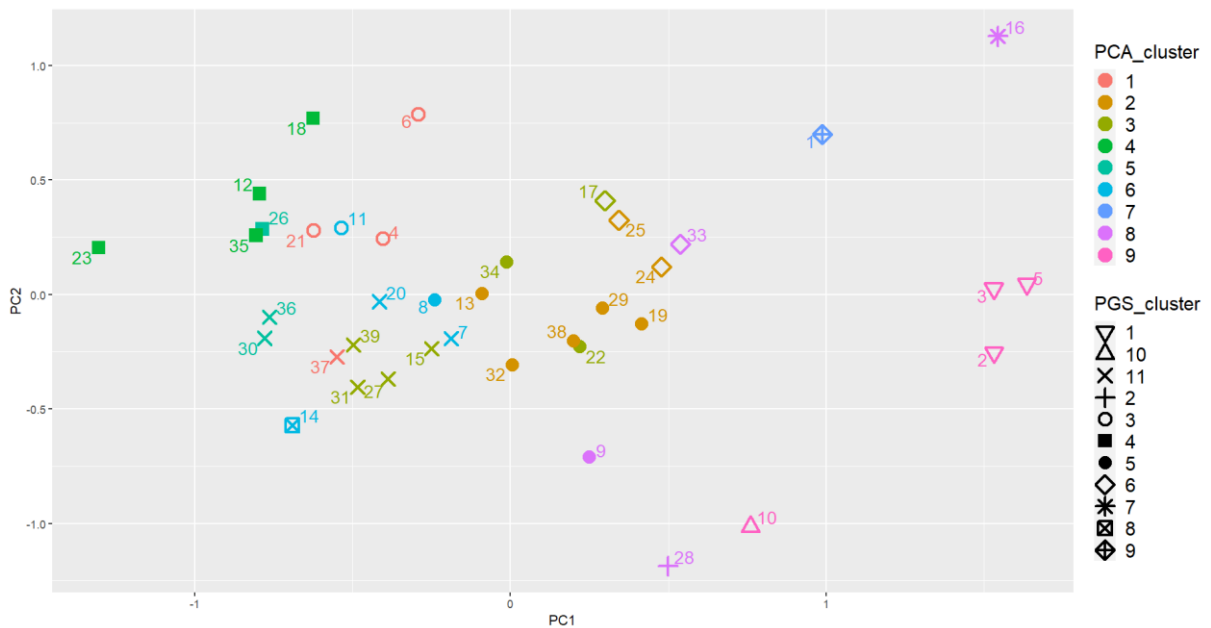


Figure 4. PC1 and PC2 for population clusters based on PCA of PGS. Populations are coloured by PCA_clusters (as seen in **Figure 2**). Symbols represent PGS_clusters determined by K-means clustering. Numbers refer to the names of population clusters in **Extended Table 1**.

2.3.4. Temporal analysis

The Mesolithic-Neolithic transition shows evidence of large metabolic changes in Eastern European populations. The clusters Mesolithic_East and Neolithic_East remained virtually unchanged in PC yet showed some of the biggest differences in PGS (**Figure 5**). A significant and consistent decline in PGS was detected in MG-11. These populations are mainly represented by individuals from the Baltics and Ukraine, with smaller input from Scandinavia and Russia and single samples from Poland and Romania. Ukrainian samples make up a significant portion of the Neolithic group but contributed little to the Mesolithic group. To rule out the possibility that the change in PGS was caused by allele frequencies changing from different ancestries instead of selection, which could confound the inference of changes in PGS due to natural selection, the temporal analysis was re-run with only samples from Ukraine and again without those samples. Both times, the decline in PGS remained visible, indicating that the change is indeed temporal and not due to a shift in ancestry.

Interestingly, the youngest individuals in the Neolithic group (ranging from 6000-5000 BP), attributed to the Baltic Comb Ceramic culture (CCC), showed a stark contrast to earlier

individuals with a sudden and large increase in PGS. This group is associated with migrations from the east and higher EHG ancestry (Saag et al., 2017). This shift in ancestry could also be detected in PC space, with CCC individuals being closer to Russians. Additionally, Ukrainian samples tended to cluster separately from the Baltic ones. To again test for possible interference from different ancestries, the population was divided into three groups: Baltic, Ukrainian, and Russian. New PGSs were calculated by first weighing the frequencies of each group with its sample size. The new PGSs remained mainly consistent with the original ones.



Figure 5. Time series analysis for Western Europe (represented by clusters Mesolithic_West and Neolithic_West) compared to Eastern Europe (represented by clusters Mesolithic_East and Neolithic_East) during the Mesolithic and Neolithic periods. The upper plots show changes in PCs (PC1 – red, PC2 – green, PC3 – blue) over time. The lower plots show changes in PGSs over time. Western Europe shows a major shift in PC1 indicating a change in ancestry (in this case admixture with Anatolian farmers) while no recognisable pattern is seen in PGS. Eastern Europe remains stable with no major shifts in PC and shows a continuous cline in PGS with a sudden shift in the end.

2.3.5. Selection analysis in Eastern Europe

The Mesolithic and Neolithic clusters were joined into one with the assumption that this new cluster represents a continuous population with changes in trait-associated allele frequencies occurring over time. This new cluster (**Figure 6**) was then divided into older and newer samples (respectively referred to as the Mesolithic group and the Neolithic group). These new groups were defined strictly by estimated age, not local archaeological contexts like the original clusters. The borderline between these groups was set at 7500 BP, as this

was approximately in the middle of the detected cline, and also roughly corresponds with the estimated time that pottery first spread into this region. As the arrival of the Neolithic was not uniform throughout this region, a few samples were reclassified. Some outliers were also removed from further analysis. Even so, these groups roughly correspond to the original clusters, with there being 35 individuals within the Mesolithic group and 57 in the Neolithic one (as opposed to 41 and 55 in the original clusters).

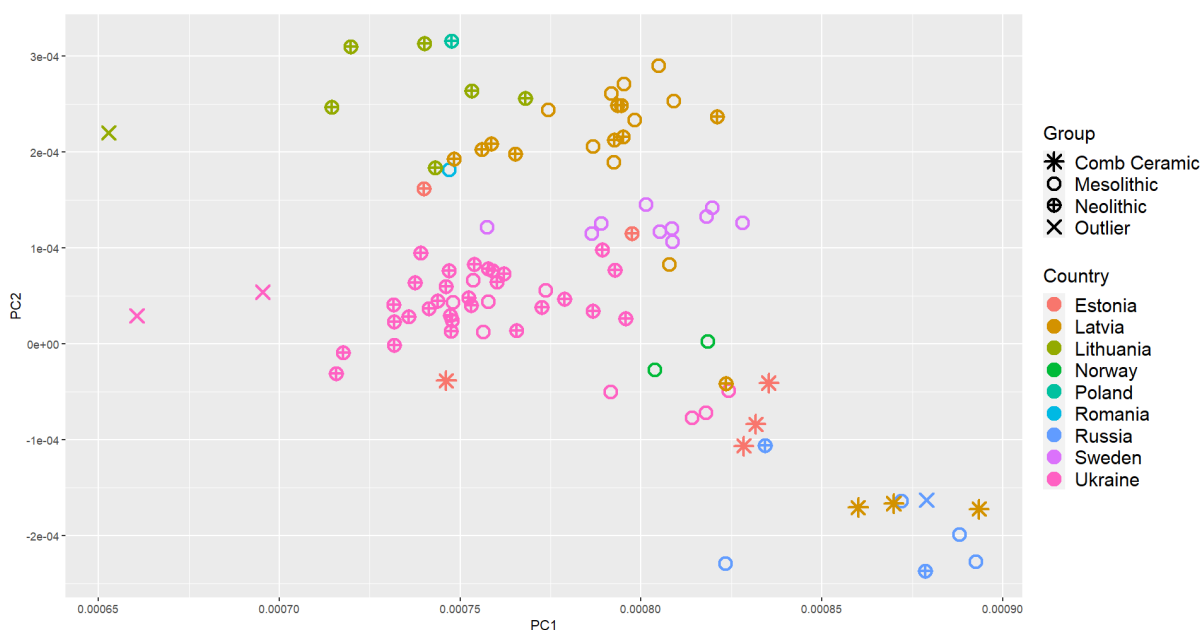


Figure 6. Mesolithic and Neolithic populations of Eastern Europe in PC space. Individuals are coloured by country of origin. Samples with an estimated age older than 7500 BP are designated as Mesolithic. Newer samples are considered Neolithic, with the newest samples (ranging from 6000 to 5000 BP) from the Baltic region labelled as Comb Ceramic.

The selection analysis was carried out twice: with and without individuals from Russia and CCC individuals. The strongest signal was found to be a reduction in traits regarding small and medium VLDL particles (MG-11) (**Figure 7**). This was also supported by increased overall VLDL diameter. A weaker signal could also be detected for a decrease in total and phospholipids in very large HDL particles (MG-5), correlated with parallel increases for medium HDL particles. Results also showed an increase in LDL cholesterol (LDL-C), particularly free cholesterol, and particle size. Furthermore, an overall trend was seen for increased phospholipid levels (except for the mentioned HDL and VLDL particles), and decreased triglyceride levels. A PGS change lower than would be expected by the null distribution (over one standard deviation) was seen for apolipoprotein-B (Apo-B), polyunsaturated fatty acids (PUFAs), and omega-3 fatty acids, as well as a significantly lower

change (over two standard deviations) for Docosahexaenoic acid (DHA), Glucose, Phenylalanine, Glycoprotein acetyls, and Ratio of bis-allylic groups to double bonds.

It is crucial to keep in mind that alternating allele frequencies are what cause PGS variations in homogeneous populations when interpreting these results. A higher score would indicate increased activity in those alleles. A larger PGS would suggest a generally higher concentration of metabolites like triglycerides and apolipoproteins, which are synthesized inside the body (and assembled into lipoproteins). In contrast, metabolites like fatty acids are obtained directly from diet. Under the assumption that natural selection acts to maintain approximate homeostasis of levels of required fatty acids, a greater PGS would instead suggest a lower intake from diet (and vice versa), as more work is required to convert readily accessible fatty acids into those that are required. It should also be kept in mind that multiple closely related traits, giving off a comparable signal, might indicate a valid signal when combined even if a single trait does not reach significance.

The p-value for average PGS change (from the null distribution) in MG-11 reached a significance of $P = 10^{-17}$, with the highest individual trait (met-c-931) reaching a significance of $P = 4.9 \times 10^{-4}$.

When including the samples with higher EHG ancestry, the same signal for a reduction in VLDL traits could still be detected, although it was weaker. The signals for increased LDL-C and decreased triglycerides could no longer be seen, however, the signal regarding HDL particles was amplified. There was also a significant decrease in PGS for monounsaturated fatty acids.

The selection analysis was carried out a third time in order to compare Early Mesolithic samples (older than 8000 BP), which also have a higher degree of EHG ancestry, directly to the CCC group (Russian samples were also included). Results showed a significant increase in all VLDL-related traits and triglycerides. Although the sample size was small (with only 8 for the CCC group), this would suggest that the differences in PGS regarding individuals from the Comb Ceramic culture are a result of changes in diet.

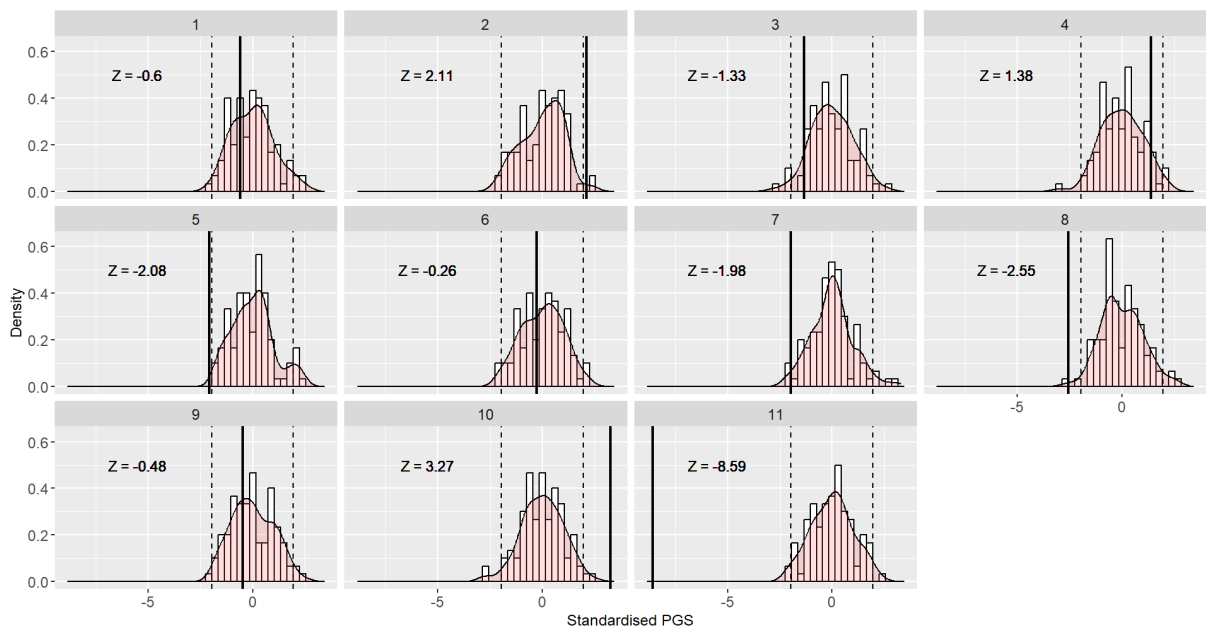


Figure 7. Visual representation of selection analysis for the Mesolithic to Neolithic transition (without Russian and CCC individuals). The standardised null distribution of each metabolite group is shown as a histogram. The average of observed PGS for each metabolite within a group is represented by a black line. The dotted lines represent critical values (± 1.96) for a 95% confidence interval.

The selection analysis was also carried out for the clusters Bronze-Age_East and Medieval_East, as they were nearly identical in PC space (**Figure 8**) and showed similar PGS changes in the temporal analysis. In fact, these clusters could largely be considered a continuation of the earlier Mesolithic and Neolithic populations, with the Bronze Age cluster consisting of 35 individuals, mainly from Estonia and Latvia. The Medieval cluster contains 36 individuals but breaks from the previous pattern with 24 of them having a Swedish origin. Unfortunately, the relevant Iron Age populations could not be included due to poor sample size and coverage in that period.

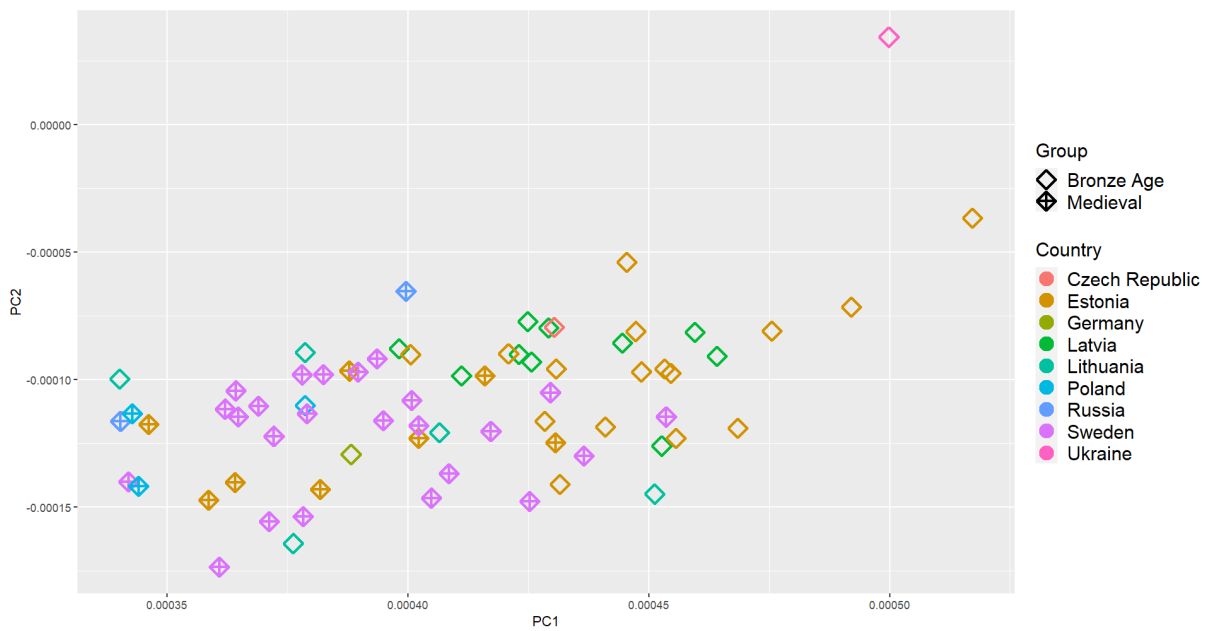


Figure 8. Bronze Age and Medieval populations of Eastern Europe in PC space. Individuals are coloured by country of origin. Samples are grouped by population cluster.

Results of selection analysis (**Figure 9**) showed strong increases for all traits relating to IDL, LDL, and HDL particles, especially cholesterol esters. The driving force behind these changes appears to be MG-10 (relating to traits for medium and large LDL), which reached a combined significance of $P = 10^{-26}$. MG-2 also reached a strong significance of $P = 10^{-10}$. Contrary to hunter-gatherers, an increase in allele frequency for Apo-B and polyunsaturated fatty acids was detected, as well as a decrease for monounsaturated fatty acids. VLDL particles showed no significant change outside the null distribution, except for a reduction in large VLDL cholesterol. Interestingly, Glucose, which is grouped with traits relating to large VLDL particles (MG-8), showed an even more significant reduction.

It must be noted that for this analysis, no attempts were made to adjust for geographic location or possible differences in ancestry, and as such, stronger results may not directly indicate stronger selection but differences in allele frequencies among populations.

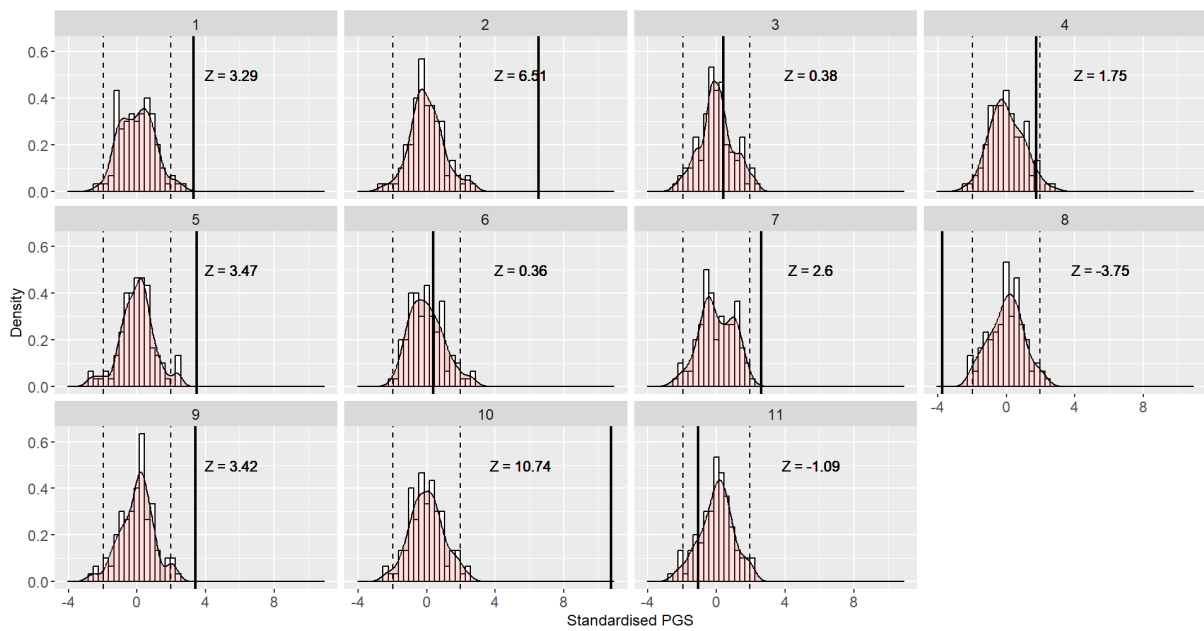


Figure 9. Visual representation of selection analysis for the Bronze Age to Medieval transition. The standardised null distribution of each metabolite group is shown as a histogram. The average of observed PGS for each metabolite within a group is represented by a black line. The dotted lines represent critical values (± 1.96) for a 95% confidence interval.

2.3.6. Pleiotropic SNPs

For the Mesolithic-Neolithic transition in Eastern Europe, two SNPs showed a significant combination of change in allele frequency and effect size in the selection analysis. 1) rs964184 is a 3'-UTR variant in *ZPR1*, which encodes for a zinc finger protein. rs964184 is associated with increased serum concentrations of triglycerides, LDL-C, and risk of Metabolic Syndrome (Mirhafez et al., 2016; Van De Woestijne et al., 2014). The risk allele (G) decreased in frequency. 2) rs429358 is a missense variant in *APOE* associated with higher LDL-C and Apo-B levels (Wu et al., 2020), and risk of Alzheimer's disease (in Chinese populations) (Yang et al., 2003). The risk allele (C) also decreased in frequency.

For the Bronze Age to Medieval transition, the four most significant SNPs were rs1532085, rs13040718, rs629301, and rs247617. Most notably, rs247617 is associated with the *CETP* gene, which encodes cholesteryl ester transfer protein that exchanges cholesterol ester and triglyceride between lipoproteins. rs629301 gave the biggest effect for MG-10 and MG-2. rs629301 is a 3'-UTR variant in *CELSR2*, which has previously been linked to LDL-C (Sandhu et al., 2008) and coronary artery disease (Noto et al., 2021). Among the 30 most significant SNPs were also rs174583 and rs174547, which are intron variants for *FADS1* and *FADS2*.

It is also notable that, among the 30 most significant SNPs from the Bronze Age to Medieval selection analysis, and the 30 most significant SNPs from the Mesolithic to Neolithic selection analysis, nine were in common: rs1532085, rs13040718, rs247617, rs964184, rs2029886, rs6065904, rs11081037, rs4296389, rs599132. Many of these SNPs are associated with a variety of metabolism-related genes in a regulatory function. Information about SNPs was gathered from Open Targets (<https://genetics.opentargets.org/>) (Ghoussaini et al., 2021) and Ensembl (<http://www.ensembl.org/index.html>, release 109) (Cunningham et al., 2022).

2.4. Discussion

2.4.1. Pathway analysis

The selection analysis for Eastern European hunter-gatherers showed changes in metabolite-related allele frequencies, that reflect significantly lower concentrations of small and medium VLDL particles, lower levels of triglycerides, and higher levels of phospholipids and LDL-C. As well as decreased frequencies for traits such as Apo-B, PUFAs and omega-3 fatty acids (especially DHA). The overall changes in metabolic traits would suggest an increased intake of marine omega-3 fatty acids (e.i., fish oil) as indicated by lower levels of VLDL and triglycerides (Bjerregaard et al., 2000; DiNicolantonio & O'Keefe, 2018). Omega-3 fatty acids are a subset of PUFAs (DeFilippis & Sperling, 2006), and as such, a lower PGS in traits relating to PUFAs would suggest less activity in synthesis pathways and thereby also indicate increased consumption from diet.

The triglyceride-lowering effects of fish oil (mainly EPA and DHA) are well known (Abbey et al., 1990; Calabresi et al., 2000; Harris et al., 1983; Von Lossonczy et al., 1978). This effect is thought to occur through reducing non-esterified fatty acid (the major component of triglycerides) delivery to the liver and subsequently hepatic VLDL production (Bornfeldt, 2021; Shearer et al., 2012). VLDL are the primary particles responsible for transporting TGs to tissues throughout the body (German et al., 2006). In turn, the rate of VLDL particle formation and the size of formed particles are determined by the availability of triglycerides (Feingold, 2022). Fish oil supplementation (in hyper-triglyceridemic adults) has been shown to increase LDL-C (Oelrich et al., 2013), and decrease the concentration of smaller, denser LDL subfractions (Calabresi et al., 2000; Wilkinson et al., 2005). However, the exact effects on LDL composition and concentration depend on other factors, such as dietary saturated fat intake (Harris, 1989). Fish oil has also been shown to enhance the conversion of VLDL to LDL particles (Lu et al., 1999), including the conversion of VLDL Apo-B to LDL Apo-B and thereby reducing overall Apo-B synthesis (Chan et al., 2003; Huff & Telford, 1989).

There is a fundamental division of PUFAs between aquatic and terrestrial ecosystems, with aquatic organisms (both marine and freshwater) having a proportionally higher content of eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), whereas terrestrial organisms have higher levels of linoleic acid (LNA) and alpha-linolenic acid (ALA) (Colombo et al., 2017; Hixson et al., 2015). This is because long-chain polyunsaturated fatty acids (LC-PUFAs) (i.e., EPA and DHA) are primarily synthesized by aquatic primary producers and are selectively

retained throughout the food web (Twining et al., 2016), with their total proportional content increasing with trophic levels (Colombo et al., 2017). Conversely, short-chain polyunsaturated fatty acids (SC-PUFAs) (i.e., LNA and ALA) decrease with increasing trophic level, suggesting they are utilized as precursors (Hixson et al., 2015). Consumers without access to aquatic resources have undoubtedly faced strong selective pressure to modify ALA and LNA into EPA and DHA. This is because the extent to which various animals rely on essential omega-3s from their diet is inversely correlated with their ability to convert SC-PUFAs into LC-PUFAs, as DHA remains minimally modified from ingestion to assimilation (Twining et al., 2016). As such, aquatic organisms constitute an important (and largely irreplaceable) source of essential omega-3s for many omnivorous terrestrial animals, including humans (Gladyshev et al., 2013).

Adaptation to a diet rich in LC-PUFAs has previously been shown in populations from Greenland (Fumagalli et al., 2015). The authors of this study concluded that *FADS* genes, which encode desaturases that catalyse the conversion of SC-PUFAs (i.e., ALA) into LC-PUFAs (i.e., EPA and DHA), faced downwards selection to compensate for a high dietary intake of EPA. Selection in the *FADS* region has also been found in European populations, occurring sometime between the Bronze Age and the present day, but in the opposite direction (S. Mathieson & Mathieson, 2018). It is thought that this increased expression of *FADS1* represents an adaptation to a high ratio of SC- to LC-PUFAs (i.e., a plant-based diet) (Ameur et al., 2012; S. Mathieson & Mathieson, 2018).

Overall, the evolutionary changes seen in Eastern European hunter-gatherers are compatible with increased consumption of marine omega-3 fatty acids (e.i., an aquatic diet). Results from the Bronze Age would be consistent with decreased aquatic consumption, but more research into this matter is needed.

2.4.2. Archaeological and stable isotope evidence

The application of polygenic scores to ancient populations could be considered strictly theoretical, and as such, more tangible evidence is needed to corroborate these findings. In order to confirm that my interpretations of genetic evidence are consistent with real-life events, I compared them to known archaeological isotopic evidence from food residues in pottery and human bones.

It has been suggested that the Baltic and Dnieper Rapids regions were ideally suited for the persistence of hunter-forager lifeways due to the network of highly productive marine coastlines and resource-rich freshwater waterways (M. Lillie & Budd, 2011). The diets of these populations were likely heavily dependent on the surrounding environment. Indeed, archaeological evidence from Estonia shows a cultural division between mainland inhabitants that were reliant on terrestrial animals and freshwater fish, and coastline and island communities, more reliant on marine mammals and fish (Kriiska et al., 2020). This is also supported by stable isotope analysis of human bones, which demonstrates the continuous importance of aquatic resources throughout the Mesolithic (including the so-called pottery-Mesolithic) with no crucial changes in dietary preferences (Törnv, 2016). Stable isotope analysis provides a direct measure of the nature of past human diets as food consumed is stored in the individual's tissues (e.g., bone collagen), with carbon stable isotope values ($\delta^{13}\text{C}$) indicating the amount of marine protein in the diet (compared to terrestrial protein), and nitrogen stable isotope ratios ($\delta^{15}\text{N}$) used to establish the trophic level of an organism in the food web (M. Lillie & Budd, 2011). Stable isotope evidence across Europe has shown that the uptake of aquatic protein started to increase in the Late Upper Palaeolithic (Richards, 2009). Stable isotope analysis at Zvejnieki (Latvia) has shown that Mesolithic and Early Neolithic people were heavily reliant on freshwater fish, with a diversification of diets occurring around 6000 BP (Eriksson et al., 2003; Meadows et al., 2018). Starting around 5000 BP, there is a distinct transition in diet towards an almost complete reliance on animal protein, which is associated with the Corded Ware culture and animal husbandry (Eriksson et al., 2003; Meadows et al., 2018). Corresponding results come from Lithuania (Jacobs et al., 2009). Similarly, in the Dnieper Rapids region (Ukraine), fishing was a pivotal subsistence strategy from the beginning of the Mesolithic to the end of the Neolithic (Budd et al., 2020; M. C. Lillie & Richards, 2000). Freshwater fish made up an important part of the diet, with overall consumption seemingly increasing in the Neolithic (M. Lillie et al., 2009), but then giving away to a greater variety in diet from ca. 5000 BP (M. Lillie & Jacobs, 2006).

Another compelling piece of archaeological evidence is the use of pottery. Hunter-gatherer pottery (both in the Baltic and East European forest-steppe) use was under strong cultural control, influenced by the surrounding foodscape and pre-existing culinary practices (Courel et al., 2020, 2021). In most areas of Eastern Europe and Western Russia, pottery was

primarily used for the processing of aquatic resources (Bondetti et al., 2020, 2021; Oras et al., 2017; Robson et al., 2019). Although the theory that pottery was adopted in response to more intensive fishing is no longer supported (Dolbunova et al., 2022), it is still possible that pottery helped facilitate the intensive exploitation of seasonal migratory fish, leading to increased sedentism and population growth (Bondetti et al., 2021; Oras et al., 2017).

The overall archaeological evidence clearly shows that fish and other aquatic organisms formed an important part of the diet of Eastern European hunter-gatherers throughout the Mesolithic and well into the Neolithic. It should be no surprise that such a vital food source acted as a driving force for natural selection over the course of thousands of years.

2.4.3. Future work

Large ancient DNA datasets have provided a new opportunity to directly study the history of human population structure and reveal past signals selection by tracking allele frequency changes over time. These signals of selection, interpreted in light of archaeological and biochemical evidence, offer new insights into gene-culture co-evolution and how changes in lifestyle have resulted in genetic adaptations (Le et al., 2022). As more ancient genomes become available, this analysis can be expanded to more time periods and geographic regions with increased accuracy.

To test the accuracy of results gathered from this thesis, this study should be replicated using another GWAS, preferably one based on a genetically distant group with homogeneous ancestry, and (if possible) with more samples, to see if the identified patterns of selection still hold up (Refoyo-Martínez et al., 2020). Parallel signs of selection should also be looked for in other parts of Europe, such as the United Kingdom, which seemingly abandoned its reliance on aquatic resources during the Neolithic expansion (Cramp et al., 2014; Pickard & Bonsall, 2022); but, as it was one of the last places for the expansion to reach, there should have been ample time for selection to occur.

SUMMARY

In this thesis, I analysed genetic information gathered from present-day individuals and ancient human remains (aDNA) in order to identify polygenic selection relating to metabolic traits in ancient European populations. To that end, these individuals were clustered together to represent distinct European and Near Eastern population clusters using principal component analysis (PCA). A genome-wide association study (GWAS) was then used to estimate the effect of genetic variants on selected metabolic traits within these populations. Finally, to quantitatively measure the relative amount of expression of the chosen traits, a polygenic score (PGS) was calculated by weighing the GWAS effect size with average population allele frequencies. These scores were used to compare different populations in space and time, and to assess changes in metabolic traits to find when and where selection had occurred.

A strong signal for selection was found among the hunter-gatherers of Eastern Europe during the Late Mesolithic and Early Neolithic period of that region. This population stayed homogeneous in genome-wide markers but showed large shifts in frequency for alleles associated with metabolite concentrations in blood plasma. Particularly, results showed decreased levels of small and medium VLDL particles and triglycerides. These results were then compared with known mechanisms of lipoprotein metabolism, effects of dietary lipid intake, and archaeological evidence of that period. A concordance of evidence strongly indicates that polygenic selection did occur among Eastern European hunter-gatherers in the Late Mesolithic, driven by an increased reliance on an aquatic diet and intake of essential omega-3 fatty acids. Later Eastern Europeans showed a strong signal for increased concentration of HDL, LDL and HDL particles, especially levels of LDL cholesterol, selected for some time between the Bronze Age and the Medieval period. More work is needed to conclusively determine if this selection was driven by the same factors.

RESÜMEE

Metaboolsed kohastumused peegeldavad suurenenud toitumist veeorganismidest Ida-Euroopa mesoliitikumi lõpus

Kaur Robin Kivirüüt

Resümee

Läbi inimajaloo on toimunud mitmeid suuri muutusi kultuuris ja keskkonnas, mis on toonud kaasa uusi selektiivseid surveid ning põhjustanud geenide evolutsioonilist adapteerumist. Suurimaks muutuseks toitumisharjumuste seas võib pidada põllumajanduse algust, mis muutis toitainete kättesaadavust üle maailma. Antud bakalaureusetöö eesmärk oli välja selgitada millised muutused on toimunud Euroopa populatsioonide metaboolsete profiilide seas viimase 10 000 aasta jooksul, ning testida hüpoteesi, et metabolismi mõjutavad geneetilised muutused on välja kujunenud, et tasakaalustada muutusi toitumisharjumustes.

Selle eesmärgi täitmiseks, analüüsiti inimsäilmetest leitud geneetilist informatsiooni (iidne DNA), et koostada iidseid Euroopa ja Lähis-Ida rahvastike esindavad populatsiooni kogumid. Seejärel kasutati genoomi hõlmava assotsiatsiooniuuringu (GWAS) põhjal saadud hinnangud valitud metaboolsete tunnuste geneetiliste variantide mõju kohta, et arvutada neile populatsioonidele polügeneetilised skoorid (PGS). PGS-i võrreldi populatsioonide vahel erinevates ajaperioodides, et tuvastada loodusliku valiku toimumist läbi valitud tunnustega seotud alleelide sageduse muutuse.

Tugev signaal looduslikule valikule leiti Ida-Euroopa küttide-korilaste seas. Need kiviaja rahvad jäid genoomi hõlmavate markerite suhtes homogeenseks kuid näitasid suuri muutusi metabolismiga seotud alleelide sageduses. Eelkõige näitasid tulemused väikesete väga väikese tihedusega lipoproteiinid (VLDL) osakeste ja triglütseriidide taseme langust veres. Neid tulemusi võrreldi arheoloogiliste allikatega, sealhulgas isotoopide analüüs savinõudest ja inimluudest, ning biokeemilise analüüsiga lipoproteiinide ainevahetusest. Sellest järeldati, et looduslikule valikule andis aluse suurenenud toitumine veeorganismidest ning sellest tulenev kõrge oomega-3 rasvhapete tarbimine. Selektiooni analüüs viidi läbi ka pronksiajast ja keskajast pärnivaste rahvaste peal, ning leiti, et nende veres on tõusnud madaltihedusega lipoproteiinide (LDL) ja kõrgtihedusega lipoproteiinide (HDL) kontsentratsioon.

REFERENCES

- Abbey, M., Clifton, P., Kestin, M., Belling, B., & Nestel, P. (1990). Effect of fish oil on lipoproteins, lecithin:cholesterol acyltransferase, and lipid transfer protein activity in humans. *Arteriosclerosis: An Official Journal of the American Heart Association, Inc.*, *10*(1), 85–94. <https://doi.org/10.1161/01.ATV.10.1.85>
- Allentoft, M. E., Sikora, M., Refoyo-Martínez, A., ... Willerslev, E. (2022). *Population Genomics of Stone Age Eurasia* [Preprint]. *Evolutionary Biology*. <https://doi.org/10.1101/2022.05.04.490594>
- Ameur, A., Enroth, S., Johansson, Å., ... Gyllensten, U. (2012). Genetic Adaptation of Fatty-Acid Metabolism: A Human-Specific Haplotype Increasing the Biosynthesis of Long-Chain Omega-3 and Omega-6 Fatty Acids. *The American Journal of Human Genetics*, *90*(5), 809–820. <https://doi.org/10.1016/j.ajhg.2012.03.014>
- Antonio, M. L., Gao, Z., Moots, H. M., ... Pritchard, J. K. (2019). Ancient Rome: A genetic crossroads of Europe and the Mediterranean. *Science*, *366*(6466), 708–714. <https://doi.org/10.1126/science.aay6826>
- Balaresque, P. L., Ballereau, S. J., & Jobling, M. A. (2007). Challenges in human genetic diversity: Demographic history and adaptation. *Human Molecular Genetics*, *16*(R2), R134–R139. <https://doi.org/10.1093/hmg/ddm242>
- Berg, J. J., & Coop, G. (2014). A Population Genetic Signal of Polygenic Adaptation. *PLoS Genetics*, *10*(8), e1004412. <https://doi.org/10.1371/journal.pgen.1004412>
- Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., Rhodes, M., Reich, D. E., & Hirschhorn, J. N. (2004). Genetic signatures of strong recent positive selection at the lactase gene. *American Journal of Human Genetics*, *74*(6), 1111–1120. <https://doi.org/10.1086/421051>
- Bjerregaard, P., Pedersen, H., & Mulvad, G. (2000). The associations of a marine diet with plasma lipids, blood glucose, blood pressure and obesity among the Inuit in Greenland. *European Journal of Clinical Nutrition*, *54*(9), 732–737. <https://doi.org/10.1038/sj.ejcn.1601088>
- Bondetti, M., González Carretero, L., Dolbunova, E., ... Craig, O. E. (2021). Neolithic farmers or Neolithic foragers? Organic residue analysis of early pottery from Rakushechny Yar on the Lower Don (Russia). *Archaeological and Anthropological Sciences*, *13*(8), 141. <https://doi.org/10.1007/s12520-021-01412-2>
- Bondetti, M., Scott, S., Lucquin, A., Meadows, J., Lozovskaya, O., Dolbunova, E., Jordan, P., & Craig, O. E. (2020). Fruits, fish and the introduction of pottery in the Eastern European plain: Lipid residue analysis of ceramic vessels from Zamostje 2. *Quaternary International*, *541*, 104–114. <https://doi.org/10.1016/j.quaint.2019.05.008>
- Bornfeldt, K. E. (2021). Triglyceride lowering by omega-3 fatty acids: A mechanism mediated by N-acyl taurines. *Journal of Clinical Investigation*, *131*(6), e147558. <https://doi.org/10.1172/JCI147558>
- Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*, *169*(7), 1177–1186. <https://doi.org/10.1016/j.cell.2017.05.038>

- Budd, C., Potekhina, I., & Lillie, M. (2020). Continuation of fishing subsistence in the Ukrainian Neolithic: Diet isotope studies at Yasinovatka, Dnieper Rapids. *Archaeological and Anthropological Sciences*, 12(2), 64. <https://doi.org/10.1007/s12520-020-01014-4>
- Calabresi, L., Donati, D., Pazzucconi, F., Sirtori, C. R., & Franceschini, G. (2000). Omacor in familial combined hyperlipidemia: Effects on lipids and low density lipoprotein subclasses. *Atherosclerosis*, 148(2), 387–396. [https://doi.org/10.1016/S0021-9150\(99\)00267-1](https://doi.org/10.1016/S0021-9150(99)00267-1)
- Chan, D. C., Watts, G. F., Mori, T. A., Barrett, P. H. R., Redgrave, T. G., & Beilin, L. J. (2003). Randomized controlled trial of the effect of n–3 fatty acid supplementation on the metabolism of apolipoprotein B-100 and chylomicron remnants in men with visceral obesity. *The American Journal of Clinical Nutrition*, 77(2), 300–307. <https://doi.org/10.1093/ajcn/77.2.300>
- Childebayeva, A., Rohrlach, A. B., Barquera, R., ... Haak, W. (2022). Population Genetics and Signatures of Selection in Early Neolithic European Farmers. *Molecular Biology and Evolution*, 39(6), msac108. <https://doi.org/10.1093/molbev/msac108>
- Choi, S. W., Mak, T. S.-H., & O'Reilly, P. F. (2020). Tutorial: A guide to performing polygenic risk score analyses. *Nature Protocols*, 15(9), 2759–2772. <https://doi.org/10.1038/s41596-020-0353-1>
- Colombo, S. M., Wacker, A., Parrish, C. C., Kainz, M. J., & Arts, M. T. (2017). A fundamental dichotomy in long-chain polyunsaturated fatty acid abundance between and within marine and terrestrial ecosystems. *Environmental Reviews*, 25(2), 163–174. <https://doi.org/10.1139/er-2016-0062>
- Courel, B., Meadows, J., Carretero, L. G., ... Craig, O. E. (2021). The use of early pottery by hunter-gatherers of the Eastern European forest-steppe. *Quaternary Science Reviews*, 269, 107143. <https://doi.org/10.1016/j.quascirev.2021.107143>
- Courel, B., Robson, H. K., Lucquin, A., ... Craig, O. E. (2020). Organic residue analysis shows sub-regional patterns in the use of pottery by Northern European hunter-gatherers. *Royal Society Open Science*, 7(4), 192016. <https://doi.org/10.1098/rsos.192016>
- Cramp, L. J. E., Jones, J., Sheridan, A., Smyth, J., Whelton, H., Mulville, J., Sharples, N., & Evershed, R. P. (2014). Immediate replacement of fishing with dairying by the earliest farmers of the northeast Atlantic archipelagos. *Proceedings of the Royal Society B: Biological Sciences*, 281(1780), 20132372. <https://doi.org/10.1098/rspb.2013.2372>
- Cunningham, F., Allen, J. E., Allen, J., ... Flicek, P. (2022). Ensembl 2022. *Nucleic Acids Research*, 50(D1), D988–D995. <https://doi.org/10.1093/nar/gkab1049>
- DeFilippis, A. P., & Sperling, L. S. (2006). Understanding omega-3's. *American Heart Journal*, 151(3), 564–570. <https://doi.org/10.1016/j.ahj.2005.03.051>
- DiNicolantonio, J. J., & O'Keefe, J. H. (2018). Effects of dietary fats on blood lipids: A review of direct comparison trials. *Open Heart*, 5(2), e000871. <https://doi.org/10.1136/openhrt-2018-000871>
- Dolbunova, E., Lucquin, A., McLaughlin, T. R., ... Craig, O. E. (2022). The transmission of pottery technology among prehistoric European hunter-gatherers. *Nature Human Behaviour*, 7(2), 171–183. <https://doi.org/10.1038/s41562-022-01491-8>

- Dudbridge, F. (2013). Power and Predictive Accuracy of Polygenic Risk Scores. *PLoS Genetics*, 9(3), e1003348. <https://doi.org/10.1371/journal.pgen.1003348>
- Dudbridge, F. (2016). Polygenic Epidemiology. *Genetic Epidemiology*, 40(4), 268–272. <https://doi.org/10.1002/gepi.21966>
- Eriksson, G., Löugas, L., & Zagorska, I. (2003). Stone Age hunter–fisher–gatherers at Zvejnieki, northern Latvia: Radiocarbon, stable isotope and archaeozoology data. *Before Farming*, 2003(1), 1–25. <https://doi.org/10.3828/bfarm.2003.1.2>
- Feingold, K. R. (2022). Lipid and Lipoprotein Metabolism. *Endocrinology and Metabolism Clinics of North America*, 51(3), 437–458. <https://doi.org/10.1016/j.ecl.2022.02.008>
- Fumagalli, M., Moltke, I., Grarup, N., ... Nielsen, R. (2015). Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science*, 349(6254), 1343–1347. <https://doi.org/10.1126/science.aab2319>
- Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admettla, A., Pattini, L., & Nielsen, R. (2011). Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution. *PLoS Genetics*, 7(11), e1002355. <https://doi.org/10.1371/journal.pgen.1002355>
- Galili, T. (2015). dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, 31(22), 3718–3720. <https://doi.org/10.1093/bioinformatics/btv428>
- Galili, T., O’Callaghan, A., Sidi, J., & Sievert, C. (2018). heatmaply: An R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics*, 34(9), 1600–1602. <https://doi.org/10.1093/bioinformatics/btx657>
- German, J. B., Smilowitz, J. T., & Zivkovic, A. M. (2006). Lipoproteins: When size really matters. *Current Opinion in Colloid & Interface Science*, 11(2–3), 171–183. <https://doi.org/10.1016/j.cocis.2005.11.006>
- Ghousaini, M., Mountjoy, E., Carmona, M., ... Dunham, I. (2021). Open Targets Genetics: Systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Research*, 49(D1), D1311–D1320. <https://doi.org/10.1093/nar/gkaa840>
- Gladyshev, M. I., Sushchik, N. N., & Makhutova, O. N. (2013). Production of EPA and DHA in aquatic ecosystems and their transfer to the land. *Prostaglandins & Other Lipid Mediators*, 107, 117–126. <https://doi.org/10.1016/j.prostaglandins.2013.03.002>
- Hancock, A. M., Alkorta-Aranburu, G., Witonsky, D. B., & Di Rienzo, A. (2010). Adaptations to new environments in humans: The role of subtle allele frequency shifts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1552), 2459–2468. <https://doi.org/10.1098/rstb.2010.0032>
- Hancock, A. M., Witonsky, D. B., Ehler, E., ... Di Rienzo, A. (2010). Human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. *Proceedings of the National Academy of Sciences*, 107(supplement_2), 8924–8930. <https://doi.org/10.1073/pnas.0914625107>
- Harris, W. S. (1989). Fish oils and plasma lipid and lipoprotein metabolism in humans: A critical review. *Journal of Lipid Research*, 30(6), 785–807.

- Harris, W. S., Connor, W. E., & McMurry, M. P. (1983). The comparative reductions of the plasma lipids and lipoproteins by dietary polyunsaturated fats: Salmon oil versus vegetable oils. *Metabolism*, 32(2), 179–184. [https://doi.org/10.1016/0026-0495\(83\)90226-3](https://doi.org/10.1016/0026-0495(83)90226-3)
- Hermisson, J., & Pennings, P. S. (2005). Soft Sweeps. *Genetics*, 169(4), 2335–2352. <https://doi.org/10.1534/genetics.104.036947>
- Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, 6(2), 95–108. <https://doi.org/10.1038/nrg1521>
- Hixson, S. M., Sharma, B., Kainz, M. J., Wacker, A., & Arts, M. T. (2015). Production, distribution, and abundance of long-chain omega-3 polyunsaturated fatty acids: A fundamental dichotomy between freshwater and terrestrial ecosystems. *Environmental Reviews*, 23(4), 414–424. <https://doi.org/10.1139/er-2015-0029>
- Huff, M. W., & Telford, D. E. (1989). Dietary fish oil increases conversion of very low density lipoprotein apoprotein B to low density lipoprotein. *Arteriosclerosis: An Official Journal of the American Heart Association, Inc.*, 9(1), 58–66. <https://doi.org/10.1161/01.ATV.9.1.58>
- Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178–210. <https://doi.org/10.1016/j.ins.2022.11.139>
- Irving-Pease, E. K., Muktopavela, R., Dannemann, M., & Racimo, F. (2021). Quantitative Human Paleogenetics: What can Ancient DNA Tell us About Complex Trait Evolution? *Frontiers in Genetics*, 12, 703541. <https://doi.org/10.3389/fgene.2021.703541>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Jones, E. R., Zarina, G., Moiseyev, V., Lightfoot, E., Nigst, P. R., Manica, A., Pinhasi, R., & Bradley, D. G. (2017). The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early European Farmers. *Current Biology*, 27(4), 576–582. <https://doi.org/10.1016/j.cub.2016.12.060>
- Kettunen, J., Demirkan, A., Würtz, P., ... Ala-Korpela, M. (2016). Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nature Communications*, 7(1), 11122. <https://doi.org/10.1038/ncomms11122>
- Kettunen, J., Tukiainen, T., Sarin, A.-P., ... Ripatti, S. (2012). Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nature Genetics*, 44(3), 269–276. <https://doi.org/10.1038/ng.1073>
- Klunk, J., Vilgalys, T. P., Demeure, C. E., ... Barreiro, L. B. (2022). Evolution of immune genes is associated with the Black Death. *Nature*, 611(7935), 312–319. <https://doi.org/10.1038/s41586-022-05349-x>
- Kriiska, A. (2009). The beginning of farming in the Eastern Baltic area. In: Dolukhanov P. M., Sarson G. R. and Shukurov, A. M. (Ed.). *East European Plain on the Eve of Agriculture*. British Archaeological Reports International. (159–179). Oxford: Archaeopress. (1964).

- Kriiska, A., Lang, V., Mäesalu, A., Tvauri, A., & Valk, H. (2020). *Eesti ajalugu. 1: Eesti esiaeg / Aivar Kriiska, Valter Lang, Ain Mäesalu, Andres Tvauri ja Heiki Valk ; toimetanud Valter Lang* (J. Ratas, Ed.). Tartu Ülikooli ajaloo ja arheoloogia instituut.
- Kriiska, A., Oras, E., Lõugas, L., Meadows, J., Lucquin, A., & Craig, O. E. (2017). LATE MESOLITHIC NARVA STAGE IN ESTONIA: POTTERY, SETTLEMENT TYPES AND CHRONOLOGY. *Estonian Journal of Archaeology*, 21(1), 52. <https://doi.org/10.3176/arch.2017.1.03>
- Kuchenbaecker, K., Telkar, N., Reiker, T., ... Wolke, D. (2019). The transferability of lipid loci across African, Asian and European cohorts. *Nature Communications*, 10(1), 4330. <https://doi.org/10.1038/s41467-019-12026-7>
- Lazaridis, I., Patterson, N., Mittnik, A., ... Krause, J. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, 513(7518), 409–413. <https://doi.org/10.1038/nature13673>
- Le, M. K., Smith, O. S., Akbari, A., Harpak, A., Reich, D., & Narasimhan, V. M. (2022). *1,000 ancient genomes uncover 10,000 years of natural selection in Europe* (p. 2022.08.24.505188). bioRxiv. <https://doi.org/10.1101/2022.08.24.505188>
- Lillie, M., & Budd, C. (2011). The Mesolithic-Neolithic Transition in Eastern Europe: Integrating Stable Isotope Studies of Diet with Palaeopathology to Identify Subsistence Strategies and Economy. In R. Pinhasi & J. T. Stock (Eds.), *Human Bioarchaeology of the Transition to Agriculture* (1st ed., pp. 43–62). Wiley. <https://doi.org/10.1002/9780470670170.ch3>
- Lillie, M., Budd, C., Potekhina, I., & Hedges, R. (2009). The radiocarbon reservoir effect: New evidence from the cemeteries of the middle and lower Dnieper basin, Ukraine. *Journal of Archaeological Science*, 36(2), 256–264. <https://doi.org/10.1016/j.jas.2008.09.005>
- Lillie, M. C., & Richards, M. (2000). Stable Isotope Analysis and Dental Evidence of Diet at the Mesolithic-Neolithic Transition in Ukraine. *Journal of Archaeological Science*, 27(10), 965–972. <https://doi.org/10.1006/jasc.1999.0544>
- Lillie, M., & Jacobs, K. (2006). Stable isotope analysis of 14 individuals from the Mesolithic cemetery of Vasilyevka II, Dnieper Rapids region, Ukraine. *Journal of Archaeological Science*, 33(6), 880–886. <https://doi.org/10.1016/j.jas.2005.11.003>
- Loos, R. J. F. (2020). 15 years of genome-wide association studies and no signs of slowing down. *Nature Communications*, 11(1), 5900. <https://doi.org/10.1038/s41467-020-19653-5>
- Lu, G., Windsor, S. L., & Harris, W. S. (1999). Omega-3 fatty acids alter lipoprotein subfraction distributions and the in vitro conversion of very low density lipoproteins to low density lipoproteins. *The Journal of Nutritional Biochemistry*, 10(3), 151–158. [https://doi.org/10.1016/S0955-2863\(98\)00094-1](https://doi.org/10.1016/S0955-2863(98)00094-1)
- Luca, F., Perry, G. H., & Di Rienzo, A. (2010). Evolutionary Adaptations to Dietary Changes. *Annual Review of Nutrition*, 30(1), 291–314. <https://doi.org/10.1146/annurev-nutr-080508-141048>
- Lyon, M. S., Andrews, S. J., Elsworth, B., Gaunt, T. R., Hemani, G., & Marcora, E. (2021). The variant call format provides efficient and robust storage of GWAS summary statistics. *Genome Biology*, 22(1), 32. <https://doi.org/10.1186/s13059-020-02248-0>

- Mallick, S., Micco, A., Mah, M., Ringbauer, H., Lazaridis, I., Olalde, I., Patterson, N., & Reich, D. (2023). *The Allen Ancient DNA Resource (AADR): A curated compendium of ancient human genomes* [Preprint]. Genomics. <https://doi.org/10.1101/2023.04.06.535797>
- Marees, A. T., De Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., & Derks, E. M. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *International Journal of Methods in Psychiatric Research*, 27(2), e1608. <https://doi.org/10.1002/mpr.1608>
- Mathieson, I. (2021). The omnigenic model and polygenic prediction of complex traits. *American Journal of Human Genetics*, 108(9), 1558–1563. <https://doi.org/10.1016/j.ajhg.2021.07.003>
- Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., ... Reich, D. (2018). The genomic history of southeastern Europe. *Nature*, 555(7695), 197–203. <https://doi.org/10.1038/nature25778>
- Mathieson, I., Lazaridis, I., Rohland, N., ... Reich, D. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, 528(7583), 499–503. <https://doi.org/10.1038/nature16152>
- Mathieson, S., & Mathieson, I. (2018). FADS1 and the Timing of Human Adaptation to Agriculture. *Molecular Biology and Evolution*, 35(12), 2957–2970. <https://doi.org/10.1093/molbev/msy180>
- Meadows, J., Bērziņš, V., Legzdīna, D., Lübke, H., Schmölcke, U., Zagorska, I., & Zariņa, G. (2018). Stone-age subsistence strategies at Lake Burtnieks, Latvia. *Journal of Archaeological Science: Reports*, 17, 992–1006. <https://doi.org/10.1016/j.jasrep.2016.03.042>
- Meisner, J., Liu, S., Huang, M., & Albrechtsen, A. (2021). Large-scale inference of population structure in presence of missingness using PCA. *Bioinformatics*, 37(13), 1868–1875. <https://doi.org/10.1093/bioinformatics/btab027>
- Mirhafez, S. R., Avan, A., Pasdar, A., ... Ghayour-Mobarhan, M. (2016). Zinc Finger 259 Gene Polymorphism rs964184 is Associated with Serum Triglyceride Levels and Metabolic Syndrome. *International Journal of Molecular and Cellular Medicine*, 5(1), 8–18.
- Mitnik, A., Wang, C.-C., Pfrengle, S., ... Krause, J. (2018). The genetic prehistory of the Baltic Sea region. *Nature Communications*, 9(1), 442. <https://doi.org/10.1038/s41467-018-02825-9>
- Motuzaitė Matuzevičiūtė, G. (2013). Neolithic Ukraine: A Review of Theoretical and Chronological Interpretations. *Archaeologia Baltica*, 20, 136–149. <https://doi.org/10.15181/ab.v20i0.812>
- Nordqvist, K., & Kriiska, A. (2015). *Towards Neolithisation: The Mesolithic-Neolithic transition in the central area of the eastern part of the Baltic Sea (with German abstract)* (pp. 537–556).
- Noto, D., Cefalù, A. B., Martinelli, N., ... Averna, M. R. (2021). Rs629301 CELSR2 polymorphism confers a ten-year equivalent risk of critical stenosis assessed by coronary angiography. *Nutrition, Metabolism and Cardiovascular Diseases*, 31(5), 1542–1547. <https://doi.org/10.1016/j.numecd.2021.01.018>
- Oelrich, B., Dewell, A., & Gardner, C. D. (2013). Effect of fish oil supplementation on serum triglycerides, LDL cholesterol and LDL subfractions in hypertriglyceridemic adults. *Nutrition*,

Metabolism and Cardiovascular Diseases, 23(4), 350–357.

<https://doi.org/10.1016/j.numecd.2011.06.003>

Omrak, A., Günther, T., Valdiosera, C., Svensson, E. M., Malmström, H., Kiesewetter, H., Aylward, W., Storå, J., Jakobsson, M., & Götherström, A. (2016). Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool. *Current Biology*, 26(2), 270–275. <https://doi.org/10.1016/j.cub.2015.12.019>

Oras, E., Lucquin, A., Lõugas, L., Tõrv, M., Kriiska, A., & Craig, O. E. (2017). The adoption of pottery by north-east European hunter-gatherers: Evidence from lipid residue analysis. *Journal of Archaeological Science*, 78, 112–119. <https://doi.org/10.1016/j.jas.2016.11.010>

Patterson, N., Price, A. L., & Reich, D. (2006). Population Structure and Eigenanalysis. *PLoS Genetics*, 2(12), e190. <https://doi.org/10.1371/journal.pgen.0020190>

Pickard, C., & Bonsall, C. (2022). Reassessing Neolithic Diets in Western Scotland. *Humans*, 2(4), 226–250. <https://doi.org/10.3390/humans2040015>

Posth, C., Yu, H., Ghalichi, A., ... Krause, J. (2023). Palaeogenomics of Upper Palaeolithic to Neolithic European hunter-gatherers. *Nature*, 615(7950), 117–126. <https://doi.org/10.1038/s41586-023-05726-0>

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. *Current Biology*, 20(4), R208–R215. <https://doi.org/10.1016/j.cub.2009.11.055>

Purcell, S. M., Wray, N. R., Stone, J. L., ... Stanley Center for Psychiatric Research and Broad Institute of MIT and Harvard. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, 460(7256), 748–752. <https://doi.org/10.1038/nature08185>

Refoyo-Martínez, A., Liu, S., Jørgensen, A. M., Jin, X., Albrechtsen, A., Martin, A. R., & Racimo, F. (2020). *How robust are cross-population signatures of polygenic adaptation in humans?* [Preprint]. *Evolutionary Biology*. <https://doi.org/10.1101/2020.07.13.200030>

Reisberg, S., Iljasenko, T., Läll, K., Fischer, K., & Vilo, J. (2017). Comparing distributions of polygenic risk scores of type 2 diabetes and coronary heart disease within different populations. *PLOS ONE*, 12(7), e0179238. <https://doi.org/10.1371/journal.pone.0179238>

Richards, M. P. (2009). Stable Isotope Evidence for European Upper Paleolithic Human Diets. In J.-J. Hublin & M. P. Richards (Eds.), *The Evolution of Hominin Diets* (pp. 251–257). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9699-0_20

Robson, H. K., Skipitytė, R., Piličiauskienė, G., Lucquin, A., Heron, C., Craig, O. E., & Piličiauskas, G. (2019). Diet, cuisine and consumption practices of the first farmers in the southeastern Baltic. *Archaeological and Anthropological Sciences*, 11(8), 4011–4024. <https://doi.org/10.1007/s12520-019-00804-9>

Saag, L., Varul, L., Scheib, C. L., ... Metspalu, M. (2017). Extensive Farming in Estonia Started through a Sex-Biased Migration from the Steppe. *Current Biology*, 27(14), 2185–2193.e6. <https://doi.org/10.1016/j.cub.2017.06.022>

Sandhu, M. S., Waterworth, D. M., Debenham, S. L., ... Mooser, V. (2008). LDL-cholesterol concentrations: A genome-wide association study. *The Lancet*, 371(9611), 483–491. [https://doi.org/10.1016/S0140-6736\(08\)60208-1](https://doi.org/10.1016/S0140-6736(08)60208-1)

- Sella, G., & Barton, N. H. (2019). Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. *Annual Review of Genomics and Human Genetics*, 20(1), 461–493. <https://doi.org/10.1146/annurev-genom-083115-022316>
- Shearer, G. C., Savinova, O. V., & Harris, W. S. (2012). Fish oil—How does it reduce plasma triglycerides? *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1821(5), 843–851. <https://doi.org/10.1016/j.bbalip.2011.10.011>
- Sievert, C., Parmer, C., Hocking, T., Chamberlain, S., Ram, K., Corvellec, M., Despouy, P., Brüggemann, S., & Inc, P. T. (2022). *plotly: Create Interactive Web Graphics via 'plotly.js'* (4.10.1). <https://cran.r-project.org/web/packages/plotly/index.html>
- Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*, 336, 012017. <https://doi.org/10.1088/1757-899X/336/1/012017>
- Tõrv, M. (2016). *Persistent Practices. A Multi-Disciplinary Study of Hunter-Gatherer Mortuary Remains from c. 6500–2600 cal. BC, Estonia* [Thesis]. <https://dspace.ut.ee/handle/10062/51352>
- Twining, C. W., Brenna, J. T., Hairston, N. G., & Flecker, A. S. (2016). Highly unsaturated fatty acids in nature: What we know and what we need to learn. *Oikos*, 125(6), 749–760. <https://doi.org/10.1111/oik.02910>
- Uffelmann, E., Huang, Q. Q., Munung, N. S., De Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, 1(1), 59. <https://doi.org/10.1038/s43586-021-00056-9>
- Van De Woestijne, A. P., Van Der Graaf, Y., De Bakker, P. I. W., Asselbergs, F. W., Spiering, W., Visseren, F. L. J., & for the SMART Study Group. (2014). Rs964184 (APOA5-A4-C3-A1) Is Related to Elevated Plasma Triglyceride Levels, but Not to an Increased Risk for Vascular Events in Patients with Clinically Manifest Vascular Disease. *PLoS ONE*, 9(6), e101082. <https://doi.org/10.1371/journal.pone.0101082>
- Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics*, 101(1), 5–22. <https://doi.org/10.1016/j.ajhg.2017.06.005>
- Von Lossonczy, T., Ruiter, A., Bronsgeest-Schoute, H., Van Gent, C., & Hermus, R. (1978). The effect of a fish diet on serum lipids in healthy human subjects. *The American Journal of Clinical Nutrition*, 31(8), 1340–1346. <https://doi.org/10.1093/ajcn/31.8.1340>
- Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301), 236–244. <https://doi.org/10.1080/01621459.1963.10500845>
- Wickham, H., Chang, W., Henry, L., ... PBC. (2023). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics* (3.4.2). <https://cran.r-project.org/web/packages/ggplot2/index.html>
- Wilkinson, P., Leach, C., Ah-Sing, E. E., Hussain, N., Miller, G. J., Millward, D. J., & Griffin, B. A. (2005). Influence of α -linolenic acid and fish-oil on markers of cardiovascular risk in subjects

with an atherogenic lipoprotein phenotype. *Atherosclerosis*, *181*(1), 115–124.
<https://doi.org/10.1016/j.atherosclerosis.2004.12.029>

Wray, N. R., Goddard, M. E., & Visscher, P. M. (2007). Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Research*, *17*(10), 1520–1528.
<https://doi.org/10.1101/gr.6665407>

Wu, H., Huang, Q., Yu, Z., Wu, H., & Zhong, Z. (2020). The SNPs rs429358 and rs7412 of APOE gene are association with cerebral infarction but not SNPs rs2306283 and rs4149056 of SLCO1B1 gene in southern Chinese Hakka population. *Lipids in Health and Disease*, *19*(1), 202. <https://doi.org/10.1186/s12944-020-01379-4>

Yang, J. D., Feng, G. Y., Zhang, J., Cheung, J., St. Clair, D., & He, L. (2003). Apolipoprotein E –491 promoter polymorphism is an independent risk factor for Alzheimer’s disease in the Chinese population. *Neuroscience Letters*, *350*(1), 25–28. [https://doi.org/10.1016/S0304-3940\(03\)00815-2](https://doi.org/10.1016/S0304-3940(03)00815-2)

USED WEBSITES

<https://www.cog-genomics.org/plink/>, visited 07.2022.

<http://www.ensembl.org/index.html>, visited 05.2023.

<https://genetics.opentargets.org/>, visited 05.2023.

<https://gwas.mrcieu.ac.uk/datasets/>, visited 10.2022.

<http://www.r-project.org/>, visited 05.2023.

<https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>, visited 07.2022.

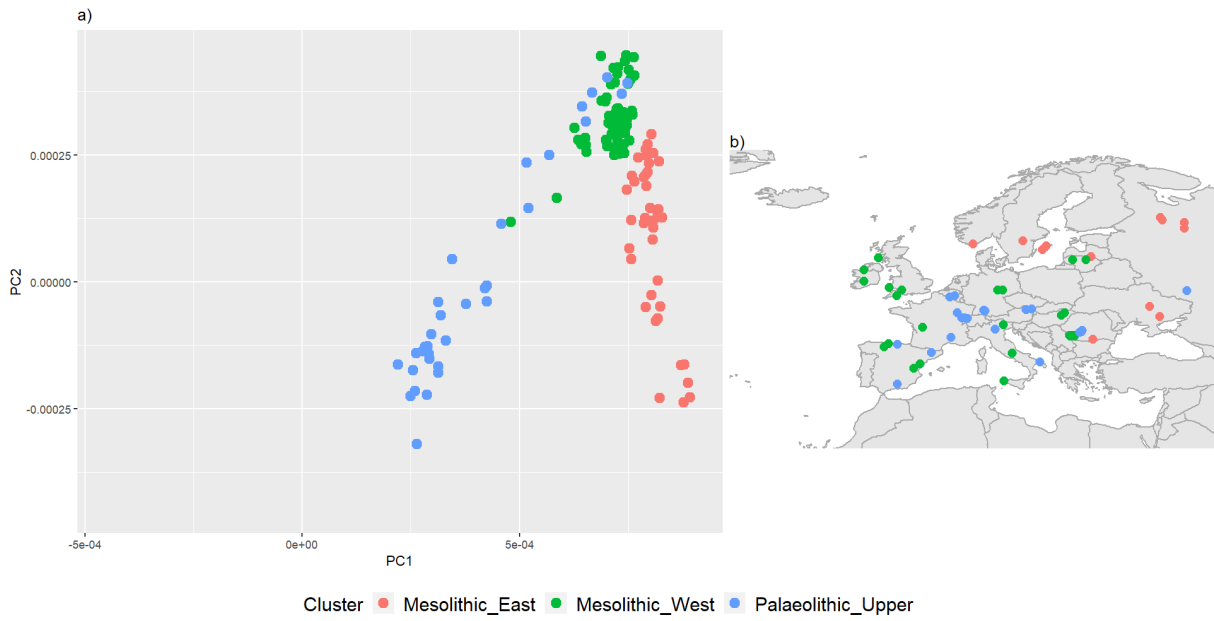
EXTENDED FIGURES

Extended Table 1. List of population clusters along with a reference number (Ref. nr.). Also shown for each cluster are its sample size (number of individuals), average age (average estimated age of all samples), and age range (ages of youngest and oldest samples). PCA and PGS clusters were determined by K-means clustering of population positions in corresponding PCA spaces. PCA and PGS clusters are numbered randomly and share no relationship with each other.

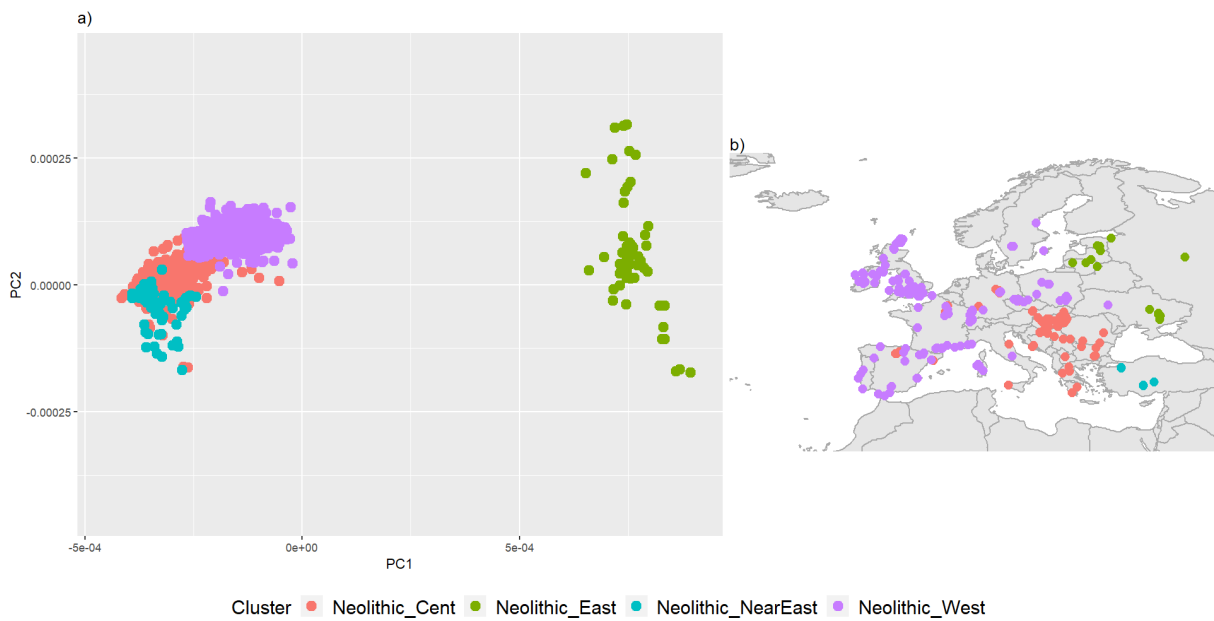
Ref. nr.	Population cluster	Sample size	Average age (cal. BP)	Age range (cal. BP)	PCA cluster	PGS cluster
1	Palaeolithic_Upper	33	23748	10134-38052	7	9
2	Mesolithic_East	41	8320	6718-12711	9	1
3	Mesolithic_West	73	8952	6001-14021	9	1
4	Neolithic_Cent	238	7012	4932-8316	1	3
5	Neolithic_East	55	6824	5407-8110	9	1
6	Neolithic_NearEast	53	8563	7870-10220	1	3
7	Neolithic_West	424	5443	4158-7363	6	11
8	Chalcolithic_Cent	119	5842	4623-7293	6	5
9	Chalcolithic_East	103	4490	3875-5394	8	5
10	Chalcolithic_Gotland	24	4809	4469-5178	9	10
11	Chalcolithic_Iberia	110	4610	3840-5397	6	3
12	Chalcolithic_NearEast	71	5781	5159-7661	4	4
13	Chalcolithic_North	65	4081	3484-4895	2	5
14	Chalcolithic_Sardinia	19	4343	3868-5198	6	8
15	Bronze-Age_Cent	240	3976	2825-5153	3	11
16	Bronze-Age_East	35	2900	2298-4735	8	7
17	Bronze-Age_Iberia	64	3557	2809-4333	3	6
18	Bronze-Age_NearEast	124	3712	3170-4700	4	4
19	Bronze-Age_North	405	3949	2612-4698	2	5
20	Bronze-Age_Sardinia	49	3540	2925-4347	6	11
21	Bronze-Age_South	37	3780	2928-4250	1	3
22	Iron-Age_Cent	70	2235	1287-3050	3	5
23	Iron-Age_NearEast	33	2455	1406-3300	4	4
24	Iron-Age_North-East	41	1938	1405-2609	2	6
25	Iron-Age_Scythian	44	2355	1537-3136	2	6
26	Iron-Age_South	87	1858	1387-3900	5	4
27	Medieval_Cent	126	1093	465-1529	3	11
28	Medieval_East	36	886	365-1120	8	2
29	Medieval_North	396	1029	236-1589	2	5
30	Medieval_South	79	987	300-1503	5	11
31	Modern_Basque	21	0	0	3	11
32	Modern_England	64	0	0	2	5
33	Modern_Finland	98	0	0	8	6
34	Modern_France	26	0	0	3	5
35	Modern_Israel	122	0	0	4	4
36	Modern_Italy	127	0	0	5	11
37	Modern_Sardinia	25	0	0	1	11
38	Modern_Scotland	43	0	0	2	5
39	Modern_Spain	107	0	0	3	11

Extended Table 2. List of metabolic traits examined in this thesis, along with their corresponding GWAS ID. Metabolites are grouped together into metabolite groups (MG) based on hierarchical clustering (Ward's method) of their PGS distribution across population clusters.

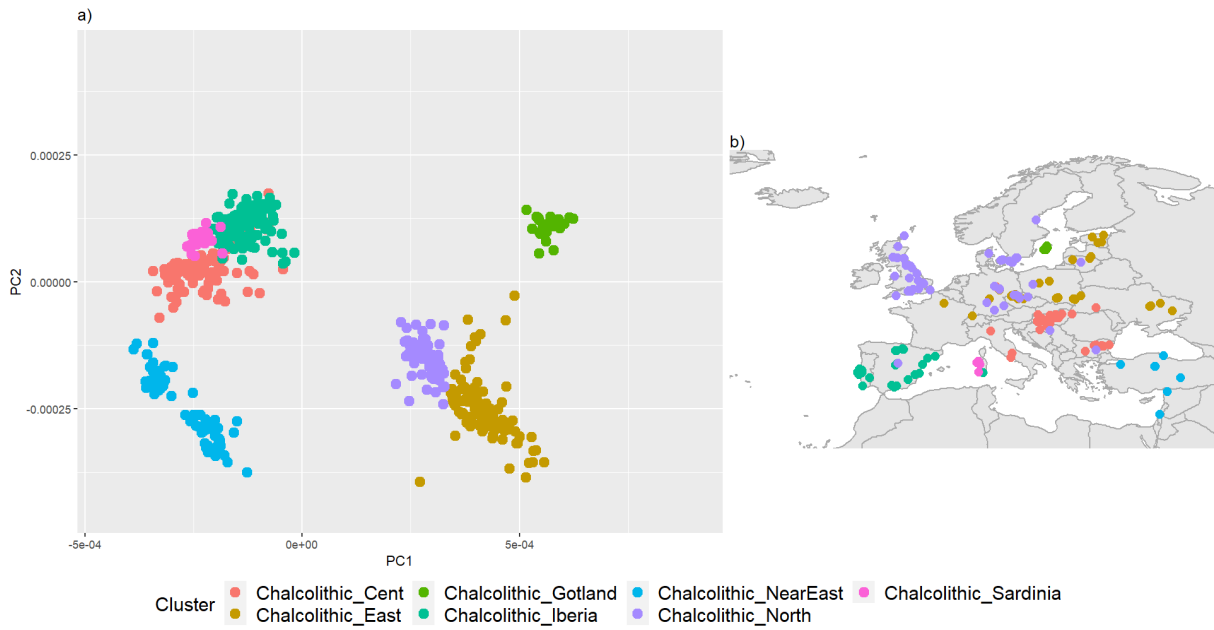
ID	Trait	MG	ID	Trait	MG	ID	Trait	MG
met-c-838	Acetoacetate	1	met-c-879	Phospholipids in large HDL	9	met-c-920	Pyruvate	8
met-c-839	Acetate	1	met-c-880	Total cholesterol in large LDL	10	met-c-921	Total lipids in small HDL	7
met-c-840	Alanine	2	met-c-881	Cholesterol esters in large LDL	10	met-c-922	Concentration of small HDL particles	7
met-c-841	Albumin	3	met-c-882	Free cholesterol in large LDL	10	met-c-923	Triglycerides in small HDL	2
met-c-842	Apolipoprotein A-I	2	met-c-883	Total lipids in large LDL	10	met-c-924	Total cholesterol in small LDL	10
met-c-843	Apolipoprotein B	1	met-c-884	Concentration of large LDL particles	10	met-c-925	Total lipids in small LDL	2
met-c-844	Ratio of bisallylic groups to double bonds	3	met-c-885	Phospholipids in large LDL	10	met-c-926	Concentration of small LDL particles	2
met-c-845	Ratio of bisallylic groups to total fatty acids	3	met-c-886	Total cholesterol in large VLDL	8	met-c-927	Total cholesterol in small VLDL	11
met-c-846	3-hydroxybutyrate	1	met-c-887	Cholesterol esters in large VLDL	8	met-c-928	Free cholesterol in small VLDL	11
met-c-847	Average number of methylene groups per double bond	1	met-c-888	Free cholesterol in large VLDL	8	met-c-929	Total lipids in small VLDL	11
met-c-848	Average number of methylene groups in a fatty acid chain	1	met-c-889	Total lipids in large VLDL	8	met-c-930	Concentration of small VLDL particles	11
met-c-849	Citrate	1	met-c-890	Concentration of large VLDL particles	8	met-c-931	Phospholipids in small VLDL	11
met-c-850	Creatinine	4	met-c-891	Phospholipids in large VLDL	8	met-c-932	Triglycerides in small VLDL	11
met-c-851	Average number of double bonds in a fatty acid chain	3	met-c-892	Triglycerides in large VLDL	8	met-c-933	Serum total cholesterol	7
met-c-852	22:6, docosahexaenoic acid	5	met-c-893	18:2, linoleic acid (LA)	3	met-c-934	Serum total triglycerides	11
met-c-853	Free cholesterol to esterified cholesterol ratio	2	met-c-894	Lactate	2	met-c-935	Sphingomyelins	6
met-c-854	Description of average fatty acid chain length, not actual carbon number	3	met-c-895	Total cholesterol in LDL	10	met-c-936	Total fatty acids	6
met-c-855	Omega-3 fatty acids	3	met-c-896	Mean diameter for LDL particles	6	met-c-937	Total phosphoglycerides	7
met-c-856	Omega-6 fatty acids	3	met-c-897	Leucine	9	met-c-938	Tyrosine	4
met-c-857	Omega-7, omega-9 and saturated fatty acids	6	met-c-898	Total cholesterol in medium HDL	4	met-c-939	Urea	2
met-c-858	Free cholesterol	7	met-c-899	Cholesterol esters in medium HDL	1	met-c-940	Valine	7
met-c-859	Glucose	8	met-c-900	Free cholesterol in medium HDL	4	met-c-941	Mean diameter for VLDL particles	8
met-c-860	Glutamine	1	met-c-901	Total lipids in medium HDL	4	met-c-942	Total cholesterol in very large HDL	5
met-c-861	Glycerol	1	met-c-902	Concentration of medium HDL particles	4	met-c-943	Cholesterol esters in very large HDL	5
met-c-862	Glycoproteins	4	met-c-903	Phospholipids in medium HDL	4	met-c-944	Free cholesterol in very large HDL	5
met-c-863	Glycoprotein acetyls	1	met-c-904	Total cholesterol in medium LDL	10	met-c-945	Total lipids in very large HDL	5
met-c-864	Total cholesterol in HDL	7	met-c-905	Concentration of medium LDL particles	10	met-c-946	Concentration of very large HDL particles	9
met-c-865	Mean diameter for HDL particles	9	met-c-906	Total lipids in medium LDL	10	met-c-947	Phospholipids in very large HDL	9
met-c-866	Histidine	5	met-c-907	Concentration of medium LDL particles	10	met-c-948	Triglycerides in very large HDL	7
met-c-867	Total cholesterol in IDL	2	met-c-908	Phospholipids in medium LDL	10	met-c-949	Total lipids in very large VLDL	3
met-c-868	Free cholesterol in IDL	2	met-c-909	Concentration of medium LDL particles	11	met-c-950	Concentration of very large VLDL particles	8
met-c-869	Total lipids in IDL	2	met-c-910	Total cholesterol in medium VLDL	11	met-c-951	Phospholipids in very large VLDL	3
met-c-870	Concentration of IDL particles	2	met-c-911	Cholesterol esters in medium VLDL	11	met-c-952	Free cholesterol in very large VLDL	3
met-c-871	Phospholipids in IDL	2	met-c-912	Free cholesterol in medium VLDL	11	met-c-953	Triglycerides in very large VLDL	6
met-c-872	Triglycerides in IDL	3	met-c-913	Total lipids in medium VLDL	11	met-c-954	Concentration of very small VLDL particles	6
met-c-873	Isoleucine	3	met-c-914	Concentration of medium VLDL particles	3	met-c-955	Concentration of very small VLDL particles	6
met-c-874	Total cholesterol in large HDL	9	met-c-915	Phospholipids in medium VLDL	11	met-c-956	Phospholipids in very small VLDL	1
met-c-875	Cholesterol esters in large HDL	9	met-c-916	Triglycerides in medium VLDL	11	met-c-956	Triglycerides in very small VLDL	1
met-c-876	Free cholesterol in large HDL	9	met-c-917	Total lipids in chylomicrons and largest VLDL particles	6	met-c-957	Total lipids in chylomicrons and largest VLDL particles	3
met-c-877	Total lipids in large HDL	9	met-c-918	Mono-unsaturated fatty acids	2	met-c-958	Other polyunsaturated fatty acids than 18:2	3
met-c-878	Concentration of large HDL particles	9	met-c-919	Other polyunsaturated fatty acids than 18:2	2	met-c-958	Concentration of chylomicrons and largest VLDL particles	3
				Phosphatidylcholine and other cholines	7	met-c-959	Phospholipids in chylomicrons and largest VLDL particles	3
				Phenylalanine	5	met-c-960	Triglycerides in chylomicrons and largest VLDL particles	3



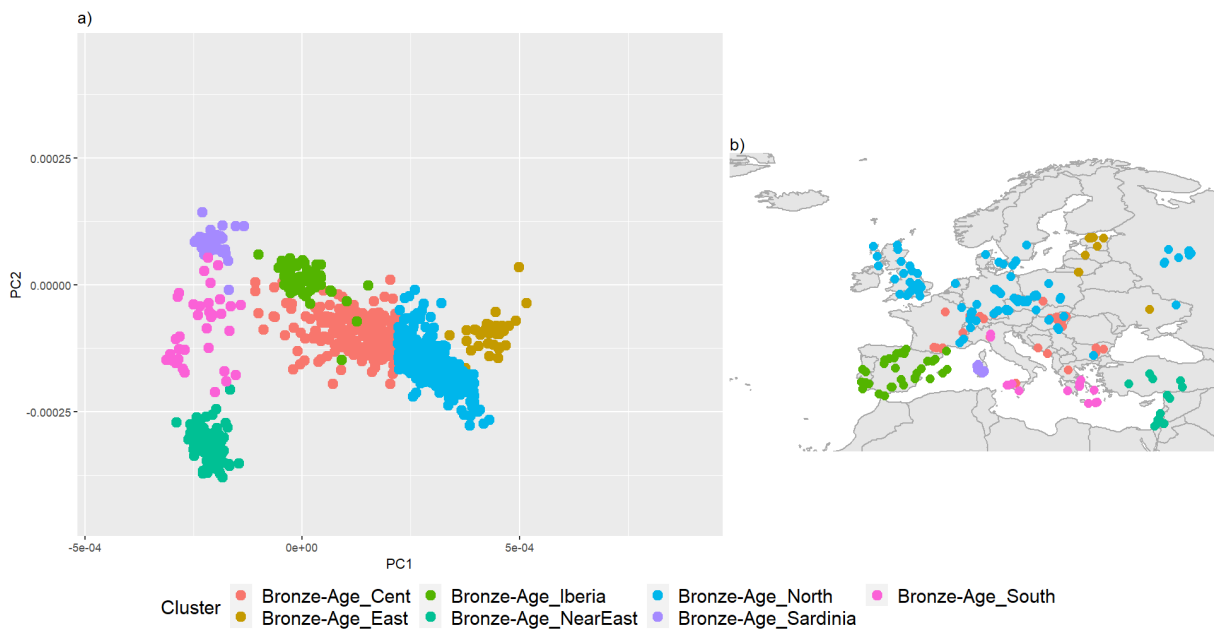
Extended Figure 1. Palaeolithic and Mesolithic population clusters in a) fixed PC space, with b) corresponding geographic locations.



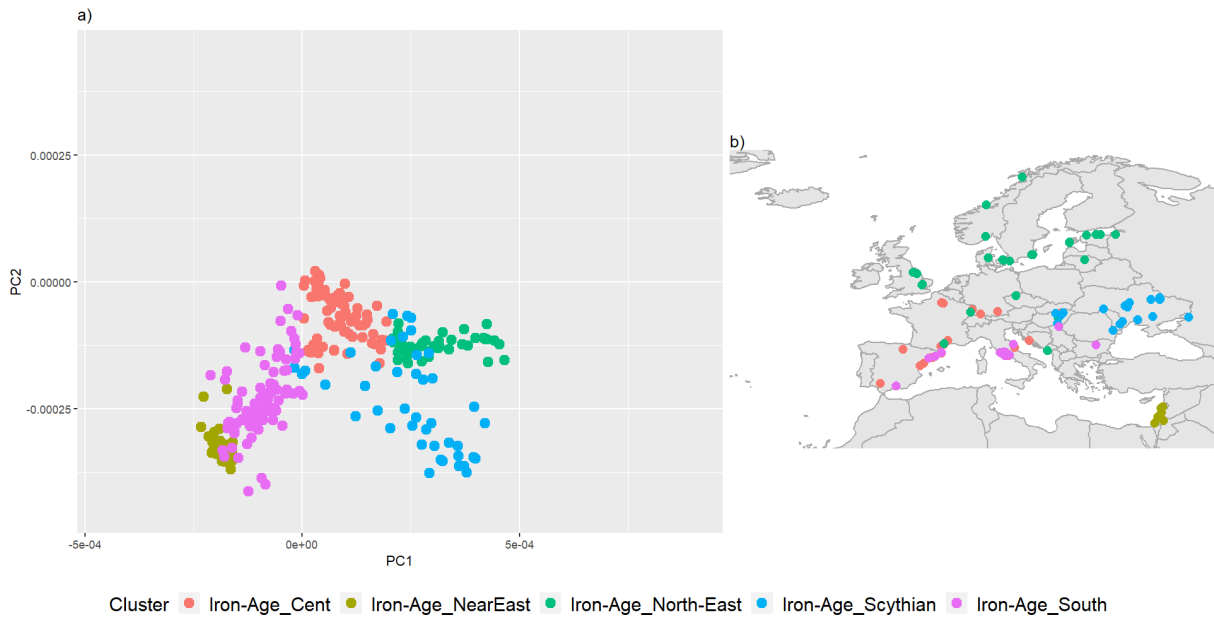
Extended Figure 2. Neolithic population clusters in a) fixed PC space with b) corresponding geographic locations.



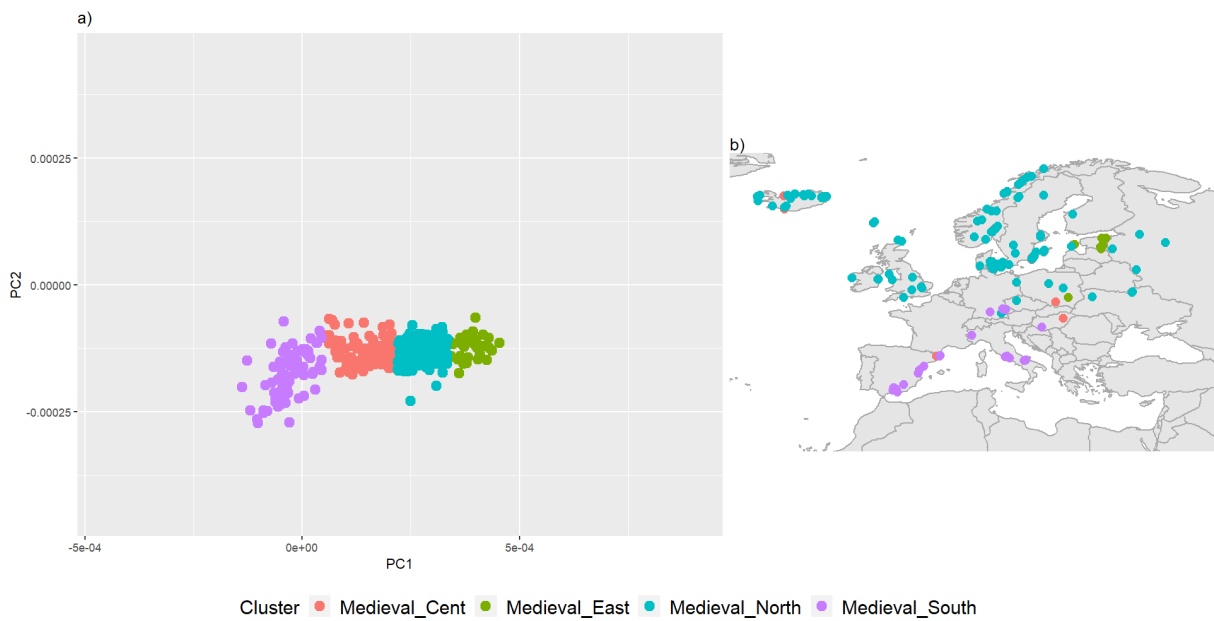
Extended Figure 3. Chalcolithic population clusters in a) fixed PC space with b) corresponding geographic locations.



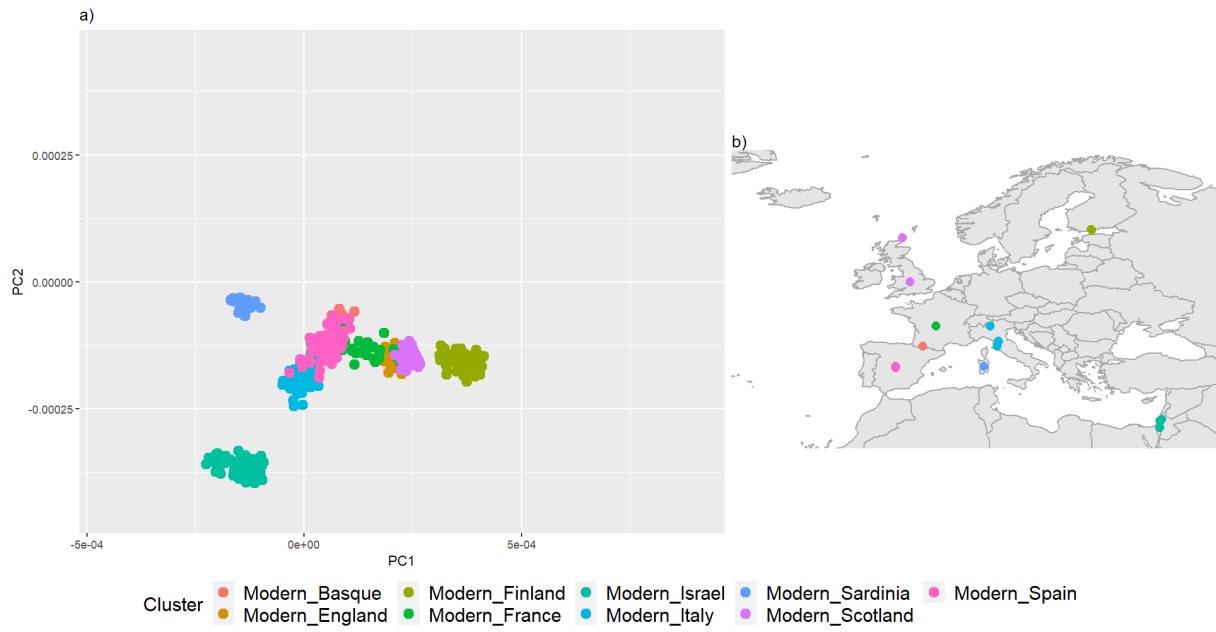
Extended Figure 4. Bronze Age population clusters in a) fixed PC space with b) corresponding geographic locations.



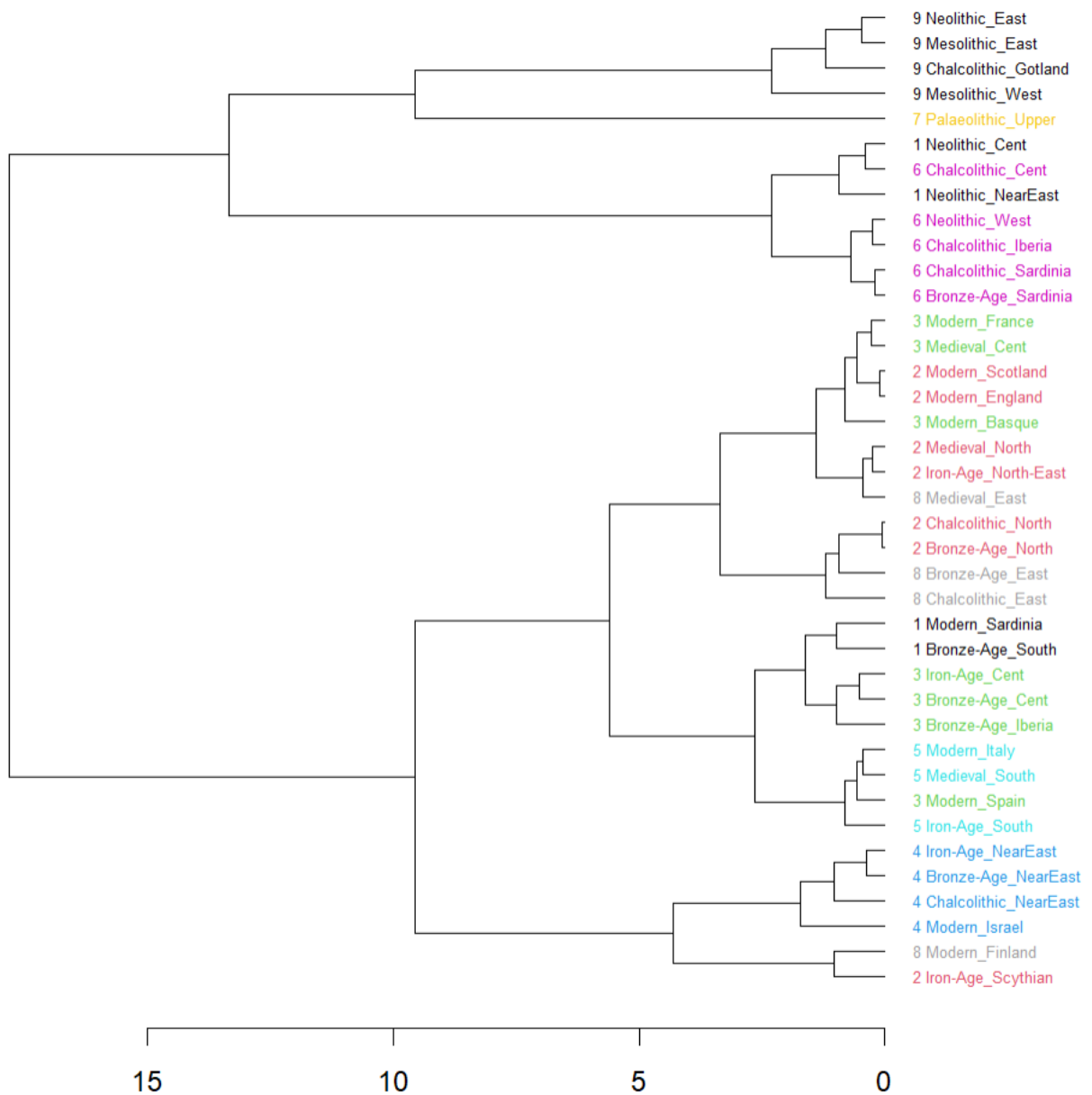
Extended Figure 5. Iron Age population clusters in a) fixed PC space with b) corresponding geographic locations.



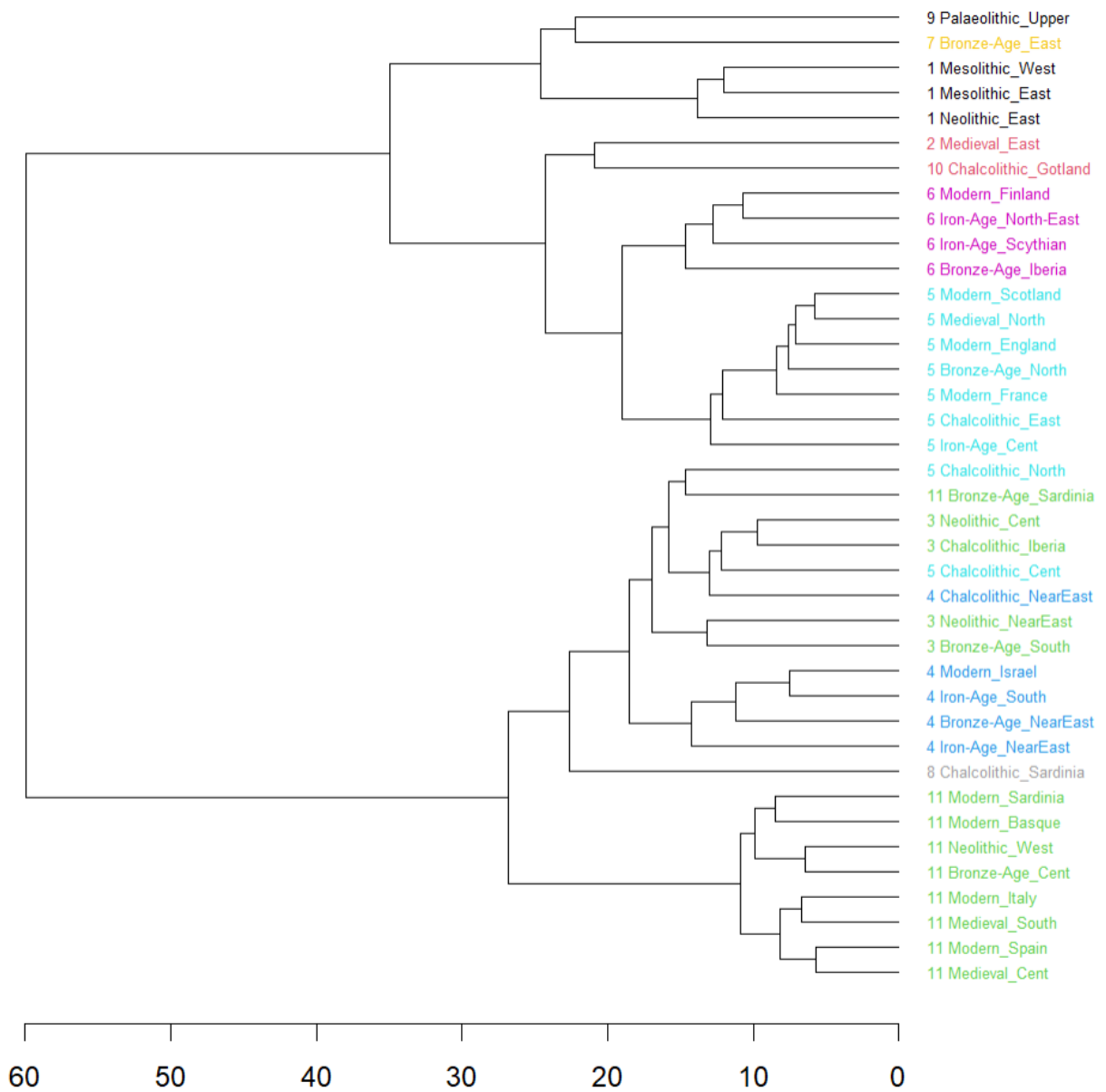
Extended Figure 6. Medieval population clusters in a) fixed PC space with b) corresponding geographic locations.



Extended Figure 7. Modern population clusters in a) fixed PC space with b) corresponding geographic locations.



Extended Figure 8. Dendrogram of population clusters based on HCA (Wards' method) of average PCs. Numbers before population names show clusters determined by K-means clustering (PCA_clusters). Colours reflect K-means clusters but are limited.



Extended Figure 9. Dendrogram of population clusters based on HCA (Wards' method) of PCA of PGS. Numbers before population names show clusters determined by K-means clustering (PGS_clusters). Colours reflect K-means clusters but are limited.

Non-exclusive licence to reproduce the thesis and make the thesis public

I, Kaur Robin Kivirüüt

grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, my thesis Metabolic adaptations reflect increased dependence on an aquatic diet in the Mesolithic-Neolithic transition of Eastern Europe, supervised by Jon Anders Eriksson.

1. I grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.
2. I am aware of the fact that the author retains the rights specified in points 1 and 2.
3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Kaur Robin Kivirüüt

29/05/2023