Dissertation

Genetic variation related to the adaptation of humans to an agriculturalist lifestyle

submitted in fulfillment of the requirements for the degree

Doctorate of natural science

doctor rerum naturalium

at the Faculty of Biology

Johannes Gutenberg University Mainz

by

Jens Blöcher

born 22.04.1984

Marburg, Germany

Mainz, 2019

Abstract

This study investigates the influence of climate and subsistence on the genetic variation of prehistoric individuals of Europe. Partial genomes from more than 100 individuals were analyzed, originating from a region spanning from north-western Anatolia, over Germany to Lithuania. The data is spread over a period of more than 10,000 years and thus covers the phases before and after people started to become sedentary. To fill temporal and geographical gaps, 27 individuals were sequenced specifically for this study. For this purpose, a hybridization-based enrichment process was used in combination with next-generation sequencing to obtain predefined genomic segments from ancient DNA (aDNA). These regions consist of over 400 functional markers, 68 non-recombinant Y markers, and more than 5,000 neutral regions of 1kb in length. The data set was completed by genomes of three modern reference populations. These made it possible to calculate haplotype- and frequency-based neutrality tests such as the integrated haplotype score or Tajima's D and thus to be able to detect signals of past selection events. Using the allele frequencies of the ancient genomes, the timing of possible selection processes could be narrowed down.

The analyzes focused on three main topics: 1.) the development of skin, eye and hair color phenotypes in Europe; 2.) The development of Europe's predisposition to type 2 diabetes and the metabolic syndrome; and 3.) The development of the immune system with respect to autoimmune diseases. The historical focus was on the influence of subsistence and the transition to farming on the evolutionary processes.

The results suggest that a high proportion of the genetic variation in genes associated with the energy and fat metabolism has been affected by adaptations to the local climate in Central Europe. Furthermore, clear indications were found that due to the reduced UV radiation in Central and Northern Europe, depigmentation of the skin had already begun before the onset of the Mesolithic. On the basis of the available data, it can be assumed that the the shift in diet and lifestyle that accompanied the transition to farming partially reinforced already existing selection pressures. Thus, with the Neolithic, a significant increase in allele frequencies associated with lighter pigmentation phenotypes was found. This is probably due to the reduced vitamin D content in the diet. Furthermore, evidence was found that indicates a metabolic adaptation to a plant-rich diet. For the later Neolithic and Bronze Age periods, an increase in allele frequencies has been found in genes that enhance protection against infectious diseases. A connection with the increasing population density and a sedentary lifestyle is very likely here.

Contents

Abstract i

Acknowledgement ii

1 Demography and Sample grouping 1

1.1 Introduction . . . 1
1.2 Archaeological background . . . 9
1.3 Methods . . . 14
1.3.1 Sawing and Milling . . . 14
1.3.2 DNA Extraction . . . 14
1.3.3 Library preparation . . . 14
1.3.4 Quantification and Quality control . . . 14
1.3.5 Nuclear Capture enrichment . . . 15
1.3.6 Read processing . . . 15
1.3.7 Estimating contamination . . . 15
1.3.8 Variant Detection . . . 16
1.3.9 Mitochondrial haplogroups . . . 16
1.3.10 Sex determination . . . 16
1.3.11 Principal Component Analysis and Clustering . . . 17
1.3.12 Model based clustering . . . 17
1.4 Results . . . 18
1.4.1 Coverage . . . 18
1.4.2 Contamination . . . 26
1.4.3 Sex determination . . . 26
1.4.4 Mitochondrial haplogroups . . . 26
1.4.5 Model-based clustering . . . 27
1.4.6 Principal component analysis . . . 29
1.5 Discussion . . . 31

2 Methods used for the analyses of the functional markers 34

2.1 Introduction . . . 34
2.2 Capture design . . . 38
2.3 Reference populations . . . 38
2.4 Population genetic estimations . . . 38
2.4.1 Allele frequencies . . . 38
2.4.2 Mean difference of allele frequencies . . . 39
2.4.3 F_{st} . . . 39
2.4.4 Pairwise F_{st} comparisons . . . 39
2.4.5 Hardy-Weinberg equilibrium . . . 40
2.4.6 Neutrality tests . . . 40

2.4.7 Averaged expected heterozygosity or gene diversity D_e . . . 40
2.5 Phenotypic predictions . . . 41
2.5.1 Predictions of N-acetyltransferase 2 phenotpyes . . . 41
2.5.2 Predictions of hair and eye color with the HirisPlex webtool . . . 41
2.6 Tracking changes in allele frequencies over multiple loci . . . 41
2.6.1 Genetic risk scores . . . 41
2.6.2 Pigmentation score . . . 42
2.7 Beta-binomial forward simulation . . . 43

3 The history of the Central European pigmentation phenotypes 46
3.1 Introduction . . . 46
3.2 Hypotheses . . . 52
3.3 Results . . . 53
3.3.1 Allele frequencies . . . 53
3.3.2 Neutrality tests . . . 61
3.3.3 Average expected heterozygosity D_e . . . 67
3.3.4 Pairwise F_{st} comparison . . . 71
3.3.5 Pigmentation Score . . . 75
3.3.6 Phenotypic predictions with the Hirisplex webtool . . . 78
3.3.7 Beta-binomial simulation . . . 80
3.4 Discussion . . . 82

4 The influence of subsistence and climate on the energy metabolism 93
4.1 Introduction . . . 93
4.2 Hypotheses . . . 100
4.3 Results . . . 101
4.3.1 Allele frequencies . . . 101
4.3.2 Neutrality tests . . . 114
4.3.3 Average expected heterozygosity D_e . . . 122
4.3.4 Pairwise F_{st} comparisons . . . 124
4.3.5 Mean difference of risk allele frequencies . . . 128
4.3.6 Genetic risk score determination . . . 131
4.3.7 Beta-binomial simulations . . . 135
4.4 Discussion . . . 136

5 The influence of subsistence and climate on the immune system 144
5.1 Introduction . . . 144
5.2 Hypotheses . . . 148
5.3 Results . . . 149
5.3.1 Frequencies . . . 149
5.3.2 Neutrality tests . . . 155
5.3.3 Average expected heterozygosity D_e . . . 158

5.3.4 Pairwise F_{st} comparisons . . . 159
5.3.5 NAT2 acetylation-phenotype prediction as described in Sabbagh et al. [2009] 161
5.3.6 Beta-binomial simulations . . . 161
5.4 Discussion . . . 164

6 Synthesis discussion and conclusion **174**

References **180**

CV **207**

1 Demography and Sample grouping

1.1 Introduction

The aim of this study is to investigate the influence of subsistence and eco-region on the genomes of European populations. Studies have shown that genetic variation in European populations is the product of migration and admixture of several ancient populations. There is also evidence suggesting that selection shaped regions of the genome in response to new environments and living conditions. To shed further light on the development of the genomes of Central European populations, a diachronic data set was used, consisting of genome-wide data of more than 100 ancient samples. This data set spans from the early Mesolithic to the Bronze Age, and covers an area from north-western Turkey, to the Baltic and western Russia. A pre-selected set of markers associated with distinct phenotypic traits was analyzed, accompanied by genome-wide data from an extensive reference set of modern-day populations. The set of markers was selected to address long-standing questions related to the population history of Europe, and the transition to farming. This first chapter addresses the demographic background of the samples and explains the reasoning for the sample groupings, which provides the foundation of this study. The following chapters will discuss the methods that were used in this study and subsequently the specific questions relating to the selected markers.

The first inhabitants of Europe

The emergence of the first anatomically modern humans (AMH) is estimated to have happened over 300,000 years ago in Africa [Schlebusch et al., 2017]. This is more than 250,000 years before AMH settled in Europe [Haber et al., 2016] and more than 280,000 years before the migration to the Americas [Slatkin and Racimo, 2016]. The specifics of how and when exactly AMH migrated out of Africa are still part of debate, but it is estimated to have happened between 80,000 - 50,000 years ago [Seguin-Orlando et al., 2014]. Early on, during the migration, AMH met other hominins, such as Neanderthals that already lived outside of Africa for thousands of years. Neanderthal remains were found all over western Eurasia, from France to southern Siberia. The oldest confirmed Neanderthal bones were found in Croatia and Italy, with an estimated age of 130,000 and 120,000 years respectively [Hublin, 2009]. The interactions between AMH and Neanderthals led to interbreeding, with the result that Neanderthal ancestry still can be traced in the genomes of modern populations [Green et al., 2010, Reich et al., 2010]. The fact that this ancestry can be found in populations from Central Europe as well as South-east-Asia it is assumed that these interactions occurred during a time before the initial split of the populations, which became ancestral to Central Europeans and South-east Asians [Fu et al., 2015].

Approximately 45,000 to 40,000 years ago, modern humans started to arrive in Central Europe. The population density is estimated to have been low and people are assumed to have lived in small bands as hunter-gatherers. However, some form of cultural complexity can be assumed from the cave paintings and figurines that can be found all over western Eurasia. From the depicted scenes in those paintings, as well as from bones found at Palaeolithic sites, it is assumed that early Europeans were big game hunters that relied on aurochs, mammoth, horses, and other mega-fauna of that era

in addition to plant-based foodstuff that could be gathered [Weiss et al., 2004, Jones, 2015].

Climatic and geological data show that there was a cooling period starting around 27,000 years ago, with its maximum around 19,000 years ago [Clark et al., 2009]. This time is known as the Last Glacial Maximum (LGM). During that time, vast parts of northern Europe were trapped under ice sheets and therefore uninhabitable. With the incoming cold it is assumed that the local hunter-gatherers retreated further south to more suitable climate zones [Tallavaara et al., 2015]. When the climate started to warm again, re-population began and people moved back northwards, following the melting ice-sheets. This period was interrupted around 14,500 years ago by a cooling period, the Younger Dryas, which peaked between 13,000 and 11,700 years ago, before the warming of the climate continued [Yang and Fu, 2018].

Hunter-gatherers of Mesolithic Europe

After the retreat of the ice sheets from central Europe, the tundra vegetation that was predominant before the LGM, populated with herds of big game such as reindeer and horse, decreased and forestation of the land took place [Shennan, 2009]. In general it is assumed that the majority of Mesolithic hunter-gatherers were nomadic people who depended on seasonal movements of game and the general availability of food in their habitat [Wellington, 2004]. As the environment changed, people had to adapt to new conditions. This can be seen in the middens of late Palaeolithic/early Mesolithic settlements, where a diversification of the animal remains can be seen [Dobrovolskaya, 2005]. While the reliance on seafood appears to increase in coastal regions [Richards et al., 2005], freshwater fish and sea fowl were also increasingly included in their diets [Eriksson et al., 2003a, Lillie and Jacobs, 2006, Fischer et al., 2007]. For some parts of Europe sea mammals such as dolphins, whales and seals provided a food-source [Fischer et al., 2007, Richards and Hedges, 1999]. In addition, plant components were used such as fruits, berries, roots and tubers, as well as nuts. There is evidence that hunter-gatherers of the Mesolithic had settlements specifically for harvesting hazelnuts, which likely were roasted to be made durable [Holst, 2010]. Further evidence indicated that starchy plants were ground into flour for further preparation since 30,000 years ago [Revedin et al., 2010]. However, none of these resources were managed in a way that led to domestication. The only domesticate that is associated with hunter-gatherers is the dog [Driscoll and Macdonald, 2010, Botigué et al., 2017].

Data from charcoal and pollen suggest that forests were deliberately burnt, either to drive out game so it could be slaughtered in large numbers, or to stimulate growth of desired plants in the following season [Scott, 2017]. From people in marine habitats, elaborate fish traps that use the tides to catch fish are known from the late Mesolithic in the Netherlands and Denmark [Wellington, 2004]. The broadening of the dietary spectrum as well as the increasing warmth is believed to have caused an increase in population size [Flannery et al., 1969, Cochran and Harpending, 2009, Crombé and Robinson, 2014]. Compared to population sizes today, the numbers of hunter-gatherers in Europe were still low. Estimations for the British Isles suggest that there were approximately 5,500 or less residential hunter-gatherers at a time [Wellington, 2004]. However, since these estimates of a

population size are based on artefact distributions and hunter-gatherers are thought to have been nomadic, with few settlements in continuous use and low numbers of cemeteries, the data has to be assessed with caution [Orschiedt et al.].

The general consensus is that hunter-gatherers lived in small bands with a territory of at least 80-100km in diameter with a loose connection to a larger group. The size of the territory assumed for the entire group is estimated between 50,000km^2 to 230,000 km^2 [Crombé and Robinson, 2014]. Mesolithic settlements are often found around lakes and rivers, suggesting the use of water transportation with dugout canoes [Crombé and Robinson, 2014] while the dispersal of raw materials among vast territories indicates elaborate trading networks [Rigaud et al., 2015]. From the upper Palaeolithic towards the later Mesolithic a decline in the number of cultural artefacts can be noted for Central Europe, even though the complexity and specialization of tools increased [Crombé and Robinson, 2014].

The Neolithic Transition

The overall mobility of hunter-gatherers is assumed to have been high, but towards the late Mesolithic an increase in permanent settlements can be noted, particularly when relying on aquatic resources [Eriksson et al., 2003b, Boroneanţ and Dinu, 2006]. When the abundance of food was high enough, people gave up the nomadic lifestyle. This trend started even earlier in the Levant and the Fertile Crescent [Cochran and Harpending, 2009]. Over 12,000 years ago, people of that region started to experiment with new forms of subsistence. Although several plant management techniques were already used in earlier periods [Revedin et al., 2010], the increasing reliance on these techniques was new. Also noticeable was that people started breeding animals in captivity. This was done on multiple occasion in various areas of the Fertile Crescent, as zooarchaeological investigations have shown [Rigaud et al., 2015]. It is assumed that these techniques in combination with the fertile soil of the area and the mild climate were some of the driving factors during this process, although specifics are still being debated [Scott, 2017].

Many of the animals and plants that became domesticated were already used as prey or were harvested in their wild forms. It is therefore unclear if the increasing population size influenced the decision to invest in new forms of subsistence, or if the disappearance of the great game herds form the region was the driving factor behind the development. In contrast, there is a consensus regarding the geographic origin of the first domesticated animals and plants. The Fertile Crescent offered an environment where several major factors came together that made the domestication possible: a favorable climate, fertile soils and a wild populations of crops that could be managed. Among the first crops that were domesticated were emmer, einkorn, barley, lentils, peas, chickpeas and bitter vetch as well as flax [Scott, 2017]. All these plants grew as wild forms in the area, yet not in central Europe. The same appears true for animals. The aurochs and the boar were native to Central Europe, while sheep and goat were not. Several studies showed that the founder populations that are ancestral to first European domesticates, originate in the Fertile Crescent [Scheu et al., 2015, Zeder, 2011, Ottoni et al., 2012, Pedrosa et al., 2005].

Figure (1) *Map showing the spread of the Neolithic in Europe, incl. important sites from which genomic data is published. Colors indicate isochrones, with numbers giving approximate arrival times of the Neolithic culture (in years BCE). Map taken from [Broushaki et al., 2016]*

Thus, the abundant availability of these animals and plants was likely the main reason why the domestication of animals and plants developed in this particular geographic area. It was probably a process of trial and error that took place several times, in several locations, with varying results. Not all attempts were successful and in some sites the sedentary lifestyle was abandoned again after a few generations [Weiss et al., 2006, Scott, 2017]. However, some people managed to develop a functioning strategy that led to a permanent sessile lifestyle with a farming-based ecology [Redding, 1988]. The sedentary lifestyle was characterized by the construction of permanent shelters in the form of stone houses where people were living off the crops and animals they were managing. Another key innovation of that time was the development of pottery. Over 1,000 years after the emergence of farming, pottery started to appear in the core region of the Neolithic. This advancement was likely based on clay coated baskets that were used for storage and transportation and became widespread around 6,700 BC [Gheorghiu, 2009]. The invention of pottery and its widespread use had a large impact on the early farmers. It enabled direct cooking, storage and transport as well as the fermentation of food and beverages [Scott, 2017].

The onset of the Neolithic is assumed to be accompanied by an increase in population size [Bellwood and Oxenham, 2008] and several changes in lifestyle. Child birth intervals decreased, possibly due to the necessity to carry from one place to another being removed [Bocquet-Appel, 2011]. Some scholars assume that the investment in offspring was lower in farming populations so that children became useful or productive at an earlier age compared to nomadic groups [Shennan, 2009]. When

skeletal remains of early farmers were compared to remains of contemporary hunter-gatherers an overall decline in health and stature could be noticed in the farmers [Eshed et al., 2010, Ash et al., 2016, May and Ruff, 2016, Latham, 2013]. It is assumed that denser settlements led to a higher pathogenic load and germs were able to manifest themselves, since the number of hosts was now high enough compared to the smaller hunter-gatherer communities. The close proximity of domesticated animals further increased the pathogen exposure. Many diseases are assumed to have originated in animals, and with the domestication the opportunity to transfer to new hosts was created. In addition, a denser settlement structure introduced the problem of waste disposal, which had only been a minor challenge in nomadic cultures where sites were changed on a more or less regular basis [Harper and Armelagos, 2010].

New sources of food were another major influence on the decreasing health. By investigating zooarchaeological remains as well as isotopic data, the vast contrast between the diets of early hunter-gatherers and the diet of early farmers becomes obvious [Richards et al., 2003, Weiss et al., 2004]. The reliance on seasonal food rich in game, fish, as well as roots and tubers, shifted towards an agrarian diet rich in grain and high in carbohydrates. Micro-nutrients such as iron and vitamin D have an important role in body maintenance and the immune system. Estimations show that the diet of early settlers had a lower amount of iron, zinc and vitamin A compared to the hunter-gatherer equivalent [Larsen, 2006, Carrera-Bastos et al., 2011]. Yet, despite the declining health and the increased mortality, the farming communities grew continuously and farming started to emerge in regions outside the core zone.

Around 8,200 BC Neolithic settlements started to spread further west and finally reached the Aegean coast of Anatolia and Greece [Özdoğan, 2011b]. From there it started to further spread into the Mediterranean along the coastlines and northwards towards the Balkans, following the Danube. By the end of the seventh millennium BC, settlements appeared in northern Macedonia, and by 6,000 BC, in the southern Hungarian plain [Orton, 2012]. The settlements between modern-day Serbia and Hungary dating to this period are assumed to belong to the same cultural complex, based on the similarities of the found artefacts. It is named Starĉevo-Körös-Cris complex, after its eponymous sites at the village Starĉevo near Belgrade and the sites at the Körös river in Hungary as well as the Cris-culture of Romania [Luca et al., 2010b].

The first farmers of Central Europe appeared around 5,700 BC in the Hungarian Plain as a successor of the Starĉevo-Körös-Cris culture [Price et al., 2001]. Until 5,000 BC, settlements with similar pottery emerged in areas further west. Based on the similarities this cultural complex is termed *Linearbandkeramik* (LBK) culture. They were equipped with a basic set of tools that allowed them to farm on the fertile loess soils along the main rivers [Vanmontfort, 2008]. Isotope data suggests that manuring with the waste of their cattle to increase the fertility of the soil was widespread [Sjögren et al., 2016, Bogaard et al., 2013]. The main domesticated animal of these early farmers was cattle, yet sheep, goat and pigs were held as well [Oelze et al., 2011]. Likely the animals lived among the farmers in their wooden long-houses. It took time and the refinement of the farming techniques until

farming spread further north. During the early time settlement sizes increased constantly up until 5,000 BC when several settlements were abandoned and a decline in overall population size in the farming communities is assumed [Shennan et al., 2013]. This decline in settlement size goes along with a regional diversification of the LBK culture [Bentley et al., 2012].

Population genetic history of Europe

How exactly farming arrived in Europe was part of a longstanding debate [Svizzero, 2017]. For a long time it was unclear, to what extend the change in subsistence was mediated by the arrival of migrants from the core-zones of the Neolithic transition, and what role local populations had. In 2009, Bramanti et al. [2009] could show, based on mitochondrial genomes from European hunter-gatherers and early farmers, that the early farmers were no descendants of local hunter-gatherers, but rather newcomers to the area. This was later confirmed by Skoglund et al. [2012] using genome-wide data, as well as in several subsequent studies [Lazaridis et al., 2014, Hofmanová et al., 2016, Lipson et al., 2017]. Genomic studies of early farmers from the Fertile Crescent and the adjacent areas revealed that the early farmers were not a homogeneous population [Broushaki et al., 2016, Lazaridis et al., 2016]. It became evident that in several regions local hunter-gatherers had independently converted to a sessile lifestyle, despite widespread cultural similarities in the Pre-Pottery Neolithic core zone. The source population from which Central European farmers derive most of their ancestry could be placed in the broader Aegean region including North-western Anatolia and the Greek mainland [Hofmanová et al., 2016, Omrak et al., 2016]. From there, people started to migrate towards Central Europe along a mainland route following rivers such as the Danube as well as along the Mediterranean coastlines [Olalde et al., 2015]. Although they must have encountered local hunter-gatherers during their migration, hunter-gatherer ancestry in early farmers increases only very little during the early Neolithic in Central Europe [Hofmanová et al., 2016]. When analyzed jointly in a principal component analysis, early Neolithic farmers of Central Europe cluster tightly with early Neolithic farmers from the broader Aegean region on the lower right of the plot, in contrast to Central European hunter-gatherers, which are distributed at the opposing end of the PCA plot [Haak et al., 2015, Lipson et al., 2017].

Studies of the genomes of European hunter-gathers show that there was a potential reduction in genetic diversity after the LGM [Fu et al., 2016]. There is evidence indicating that, after the ice-sheets retreated, Europe was repopulated from different refugia [Villalba-Mouco et al., 2019]. Potentially due to local differences in admixture between people from different refugia and in addition to varying drift and isolation by distance, hunter-gatherers of Europe also developed some form of population structure. It seems that individuals from an area between the Iberian peninsula in the southwest, the Baltic region in the northeast, and the Balkans in the south-east share more common ancestry compared to individuals from east of this region [Mathieson et al., 2018]. Therefore, individuals that fall on the extreme sides of this spectrum are often referred to as Western Hunter-Gatherers (WHG) and Easter Hunter-Gatherers (EHG) [Haak et al., 2015]. Local ancestry is variable and is often described as tending more to one side or another, often in correlation with the individuals place of origin [Mathieson et al., 2018, Mittnik et al., 2018]. While hunter-gatherers of Latvia were found to

share more ancestry with WHG samples compared to EHG samples, Scandinavian hunter-gatherers could be placed on the spectrum in between EHG and WHG [Günther et al., 2018].

Even though farming spread throughout most of Central Europe, the hunter-gatherers did not disappear completely. They coexisted with the incoming farmers for at least 2,000 years [Bollongino et al., 2013] and during that time started to admix [Lipson et al., 2017]. This can be seen in genomes from the middle and later Neolithic period, where alongside ancestry of early Neolithic farmers, hunter-gatherer ancestry started to re-appear. When ancient ancestry is modeled in modern-day populations, a south to north cline of increasing hunter-gatherer ancestry becomes visible [Lazaridis et al., 2014]. This cline is, to some extent, negatively correlated with the arrival of farming in the respective areas. The highest degree of ancient farmer ancestry is estimated for the people of Sardinia, while the highest degree of hunter-gatherer ancestry was estimated for modern-day Latvians. During these modeling approaches it became evident that modern-day European populations can not be described as descendants of only these two ancient populations [Lazaridis et al., 2014]. A third component was identified as ancestry related to people originating from the Pontic Caspian steppe region [Allentoft et al., 2015, Haak et al., 2015]. Based on genomic data, it could be shown that this ancestry component was introduced to Central Europe during the late Neolithic and Early Bronze Age and it increased throughout Europe and subsequently reached the Iberian peninsula and the British Isles [Brace et al., 2018, Olalde et al., 2018].

The steppe ancestry component was associated with the "Yamnaya" or "pit-grave culture". This cultural horizon is mainly defined by its unifying burial culture, where an ochre-smothered body was placed in a pit, which was covered by a burial mound, often referred to as a kurgan [Heyd, 2012]. How this steppe ancestry component was introduced into Europe is still being debated [Kristiansen et al., 2017, Heyd, 2017]. Genetic studies indicated a substantial migration [Haak et al., 2015] that was potentially biased towards migrating males [Goldberg et al., 2017]. However, the majority of individuals associated with cultural complexes dating to the Late Neolithic or later periods, such as Corded Ware or Bell Beaker, show signs of ancestry related to Neolithic farmers from the broader Aegean region, local European hunter-gatherers, and people from the Pontic Caspian steppe [Olalde et al., 2018].

Migrations and admixture formed the genomes of European populations in a way that is still visible today [Lazaridis et al., 2014, Haak et al., 2015]. Nevertheless, there is strong evidence indicating that selection also played a role. Migrations into new environments as well as changes in subsistence are assumed to have created selective pressures, which led to adaptations that shaped the phenotypes seen in recent populations [Bersaglieri et al., 2004a, Alonso et al., 2008, Pickrell et al., 2009, Hudjashov et al., 2013, Wang and Speakman, 2016]. The plethora of evidence in relation to selection is based on inferences from modern-day data, while only a few studies have used ancient DNA to investigate potential adaptations [Wilde et al., 2014, Mathieson et al., 2015, Loog et al., 2017].
To shed further light on the effect of new living conditions on the genomes of early Europeans, either caused by migration or the transition to farming, a capture array was composed. This array contains

a set of markers, related to genes that are associated with visible phenotypes, the immune system and detoxification of xenobiotics as well as the energy metabolism. While some of these markers can be used to reconstruct the appearance of ancient individuals, several other SNPs are associated with medical conditions such as auto-immune diseases, drug resistance and the metabolic syndrome.

This capture array was applied to a set of ancient samples that was accompanied by additional literature data, to create a data set that can be used to investigate for signs of selection by tracing changes in allele frequency data over time. To be able to estimate allele frequencies, the grouping of the sample is crucial. To ensure that only samples that originate from the same population background were grouped together, the archaeological data was compared to results from population genetic analysis.

1.2 Archaeological background

Ukraine

The site of **Lesnik cave** is located near Odessa on Crimea, Ukraine. The sample (Lec2) was found in a cave, known as "bear cave", because remains of Palaeolithic brown bears were found [Ridush, 2009]. It was not directly dated, but based on contextual finds it was placed in the early Mesolithic or late Palaeolithic of the region.

Latvia

The site of **Zvejnieki** is located in northern Latvia some kilometers inland of the coast. The sampling site is a cemetery that was in use for several thousand years, from the middle Mesolithic until Neolithic times. In addition to the cemetery, settlement sites are known. The oldest samples date to 7182±107 cal BC. Isotopic data from the remains found suggest a diet rich in protein from mammals and freshwater fish. Even though the site is located on the mainland, seal remains were found, which suggest either the use of coastal resources or at least contact with other hunter-gatherers from coastal regions [Eriksson et al., 2003a]. The samples that were used in this study span over a time between 3890±67 cal BC to 7182±107 BC. In total there were six samples that yielded enough DNA to be sequenced (see table: 1).

Russia

The sites of **Minino** are located in north-west Russia in the upper Volga region near the shores of lake Kubenskoye (see figure: 2). They consist of a burial complex that was used during the late Palaeolithic, throughout the Mesolithic, and into Neolithic times. The main part of the site is a Mesolithic cemetery which was discovered in 1984. Based on isotopic data it was suggested that the people buried at the sites relied on a diet rich in meat with additional plant food [Dobrovolskaya, 2005], but also rich in freshwater resources [Wood et al., 2013]. Five samples from those sites were sequenced for this study. The ages range from 5650-4600 cal BC to 8671±48 cal BC (see table: 1 for more detail).

Austria

Seven samples were taken from **Asparn-Schletz** in Austria. The site, which is associated with the later stages of the *Linearbandkeramik* culture, dates are estimated between 5,150 and 5,050 BC. It is located in a loess-soil region and was enclosed by a 2m deep ditch, probably for protection. Remains of 67 people were uncovered, but it is believed that up to 300 people were buried there. The site is assumed to be the scene of a massacre, where a whole village was destroyed. Analysis of strontium isotopes further suggest that all the individuals found at the site, were of local origin. The remains include adults and children and display signs of blunt force trauma. Based on the sex and age distribution of the remains it is assumed that young women were kidnapped, as this group is under-represented. The bones of the adults display signs of work, associated with plowing and general farming [Teschler-Nicola et al., 1999, Christensen, 2004, Lidke, 2005].

Kleinhadersdorf in Austria is a cemetery that is also associated with the *Linearbandkeramik*- culture. The next settlement is only 3km away and the general area was densely populated, because of its fertile loess-soil. However it is not possible to associate the cemetery with one of the settlements. The first graves were found in 1931 and after several excavations, 62 individuals could be excavated. All bodies were buried in a squatting position, which is a feature typically associated with the LBK culture. It is assumed that there were up to 200 burials at the site, which was in use during the earlier phase (*Flomborn* phase) of the LBK culture. Nearly half of the individuals found, displayed sings of caries. Signs for other diseases, often associated with malnutrition, were also found [Lenneis and Neugebauer-Maresch, 2015].

Germany

Two samples originate from the northern German site of **Criewen**, northeast of Berlin at the border to Poland. The samples were found in separate graves, associated with an non-agricultural context, and were dated to 4770±40 cal BC (Gr1) and 4600±60 cal BC (Gr2). Based on the anthropological analysis, both samples were assumed to be male. One of the samples was buried with ca. 3,000 perforated shells of the river nerite *Theodoxus fluviatilis* [Street et al., 2001].

Another site from northern Germany is **Groß Fredenwalde**. On this site, nine individuals were found in at least three distinct graves. The site was used in a period between 6,400 BC to 4,900 BC, which overlaps with the arrival of agriculture in the area ca. 5,200 BC. The sample that was prepared for this study dates to 6200 – 5400 cal BC and is associated with a pre-agricultural phase [Terberger et al., 2015].

One sample originates from the **Große Ofnet Höhle**, Germany. This, too, is a unique site compared to other Mesolithic sites from central Europe. The sample was found among 33 others in nests of heads that were separated from the bodies, all facing west. Some assumed female skulls had grave goods in the form animal teeth and mollusk shells. There are clear signs of trauma and theories range from cannibalism to a war-like massacre. This was considered a novelty, since not many signs of intergroup violence are known form the Mesolithic. The site dates to 7720±80 calBP [im Walde et al., 1986, Orschiedt, 1998].

The **Blätterhöhle** in Hagen, Germany, shows signs of sporadic occupation since the Mesolithic. Several fireplaces that date to the Mesolithic can be found around the cave and in front of its main entry. The cave was not used as a permanent settlement, but rather as temporary shelter and as a secondary burial site. For this purpose, remains were relocated from elsewhere into the cave. Interestingly, it was not only used by hunter-gatherers and farmers in different periods, but also at the same time for around 2,000 years. If whole bodies, or just skulls and certain bones were brought into the cave, is difficult to say, because of the state of preservation of the remains in the cave. The cave was filled with debris and the presence of badgers and smaller rodents led to disturbance of the layers and the bones therein. The 11 samples included in this study fall between 8652±58 cal BC and 3020±61 cal BC [Orschiedt et al., Bollongino et al., 2013].

In Germany near the town Landshut lies the burial site of **Essenbach-Ammerbreite**. This site, too, is associated with the LBK culture. The site is not completely excavated, but 29 individuals were found so far, with the youngest being approximately six years of age. [Bickle and Whittle, 2013]. Another burial site in close proximity is the site of **Dillingen** near Ulm. This site was used during the middle and later LBK phases and 27 individuals could be excavated. Strontium isotope analysis of the site showed that all females buried were non-local, in contrast to only half of the males. Only 300m south of the burial site, houses surrounded by palisades were uncovered. Alongside these houses which seem to date to the same time as the burial ground, another 20 skeletons were found, some of which displayed signs of violence[Bickle and Whittle, 2013, Bentley et al., 2002].

One of the most interesting early neolithic sites in Germany is **Herxheim** in Rhineland-Palatine. The site dates to the earlier LBK phase and was occupied until the later LBK phase. It was enclosed by a mound with palisades and a surrounding ditch. The remains of up to 450 individuals were uncovered and most of them displayed signs of violence. The sites purpose is still debated, but a ritualistic character is generally assumed [Orschiedt and Haidle, 2006]. The majority of the remains showed signs of alteration. People were decapitated, limbs were cut off and bones were opened. There is a discussion as to whether the abundance of cut marks found can be interpreted as signs of cannibalism. The fact that long bones were opened or shattered point in the same direction, since it is assumed that this was done to get access to the bone marrow. All individuals found at the site have died in a relatively short period of time during the later LBK phase. Differences in pottery styles found in the ditches, as well as Strontium isotope analysis, suggest that the people found at the site came from a broader area [Orschiedt and Haidle, 2006, Zeeb-Lanz et al., 2008, Turck et al., 2012, Haack, 2016].

The youngest site is located in northern Germany in the **Tollense valley** near the river Weltzin. It dates to the later Bronze Age period, 1129 – 924 cal BC. The remains found were scattered over a broad area and are assumed to be the result of a war-like conflict between larger groups. The bones have various marks that are believed to be the results of extreme violence and in many cases the likely cause of death. The marks consist of fractures, blunt force trauma or even arrow tips that stayed stuck in the bone. Isotopic data of the remains suggest that the people relied on C4 crop plants such as millet as main food resource [Jantzen et al., 2011].

Serbia

Eight samples with a hunter-gatherer background originate from the Iron Gates area in Serbia, near the Romanian border. The samples come from two different sites, **Vlasac and Lepenski Vir**, that were 2km apart from each other and were part of the Phd Thesis of Zuzana Hofmanova [Hofmanová, 2016]. The Iron Gates area was densely populated by hunter-gatherers that lived in settlements nested along the Danube river that are assumed to have been permanent. It is thought that the area surrounding the settlements provided its inhabitants with enough food and therefore allowed for a fully sedentary lifestyle. Inhabitants relied heavily on fish from the river, but also

on game and plants from the nearby forests [Bonsall et al., 2004]. The site of Vlasac is older then Lepenski Vir, with the six samples from the site being estimated to date 7400 and 6200 BC. The majority of samples from Lepenski Vir were estimated to be younger, with the exception of Lepe51, which dates to 7940 – 7571 cal BC. Therefore, they fall into the transitional period where agriculture arrived in Serbia. There is evidence of contact between the hunter-gatherers and incoming farmers, since remains from that period include not only hunter-gatherers, but also farmers and individuals of mixed parents [Mathieson et al., 2018, Hofmanová, 2016]. Furthermore, there are at least 15 sites in close proximity that show cultural similarities. However, several differences regarding burial practices were confined to specific sites only [Boroneanţ et al., 2012].

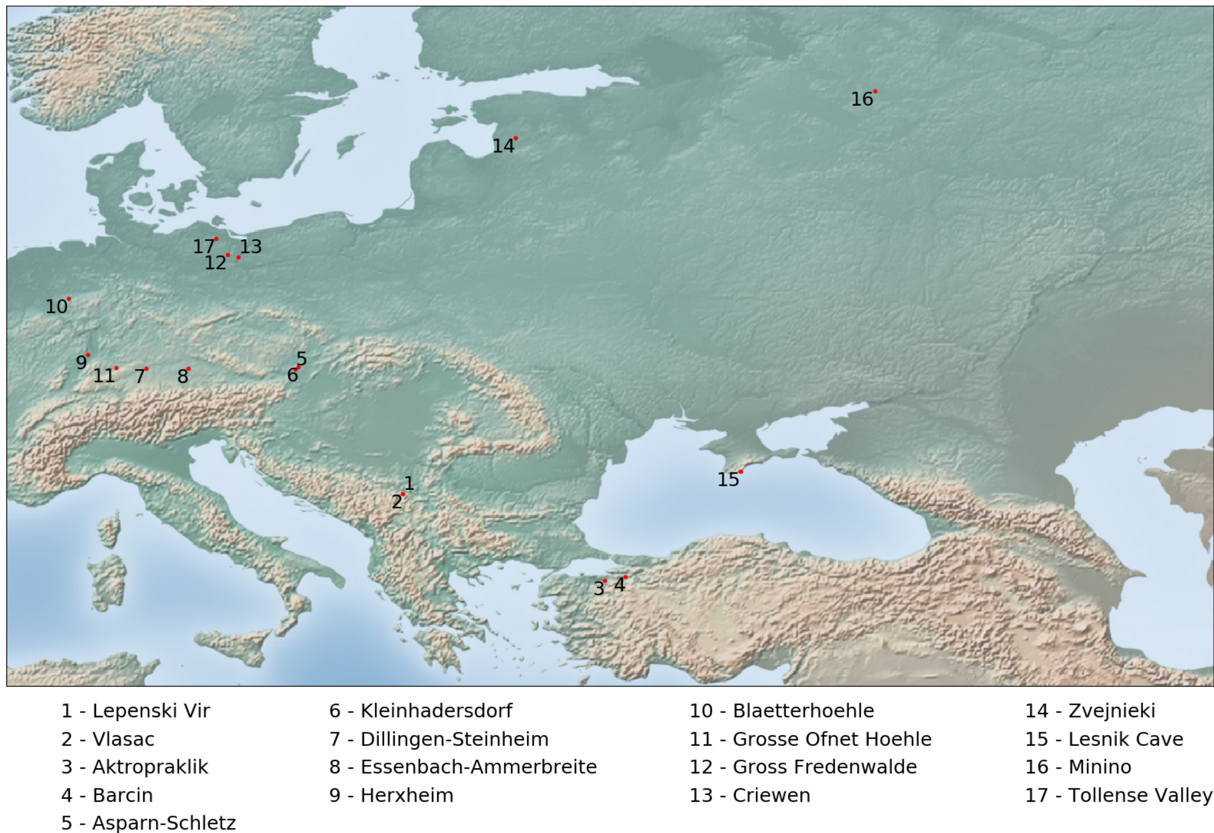

Figure (2) *Sampling sites of the samples that were prepared in ancient DNA laboratory in Mainz*

Turkey

Aktopraklik is an early farming site in the Marmara region of north-western Anatolia. The site is located near lake Ulubat, west of Bursa, and was occupied since approximately 6,500 BC. The full 'Neolithic package' can be found at the site, with domesticated animals and plants as well as pottery. Isotopic studies showed that the people relied on a terrestrial based diet, that consisted mainly of C3 plants and the animals that consumed them [Budd et al., 2013]. The people lived in oval huts with sunken floors and buried their dead in cemeteries, instead of below the houses [Özdoğan, 2011a].

The site of **Barçin** (*Barçin Höyük*) is also located in the Marmara region of north-western Anatolia. Possibly, the occupation started a bit earlier compared to Aktopraklik, at approximately 6,600 BC. It is located on a plain further inland, near the shores of lake Iznik, north-east of Bursa. The people

of Barçin were farmers and lived in rectangular houses. Most of their dead were buried in-between houses in courtyard areas or in the fill of abandoned houses. The occupation is assumed to have ended after 600 years, but re-occupations frequently occurred throughout later periods [Gerritsen et al., 2013, Weninger et al., 2014].

1.3 Methods

For this study, 27 samples were newly sequenced, while the rest of the data was collected in the context of other projects in the Palaeogenetics group, or added from literature sources. The wet-lab methods described in the following paragraphs only concern the newly sequenced samples. From the read-processing onwards,, all described methods apply to all samples, since they were all newly aligned for this work.

1.3.1 Sawing and Milling

After UV radiation of 30 min from each side, the surface of the samples was removed via rotating blade (Electer Emax IH-300, MAFRA), or in case of teeth, via sandblasting (P-G 400, Harnisch & Rieth, Winterbach, Germany). Afterwards samples were cut into smaller pieces for the milling process. Pieces were again UV-radiated for 30 min before pulverization by milling was performed in a ceramic shakers (MM200, Retsch) with ceramic balls until the sampling material was completely pulverized. After milling, a blank control was processed alongside the Samples in each step as described in [Scheu et al., 2015].

1.3.2 DNA Extraction

DNA extraction was performed as described in [Scheu et al., 2015], [Hofmanová et al., 2016] and [Broushaki et al., 2016]. The powdered sample was mixed with EDTA (5ml-15ml, 0.5M, Ambion/Applied Biosystems, Life technologies, Darmstadt, Germany), N-Laurylsarcosine (250µl, 0.5%; Merck Millipore, Darmstadt, Germany) and Proteinase K (30µl, 18Uµl; Roche, Mannheim, Germany) and incubated on a shaker at 37℃ for 48h in order to dissolve the powder. DNA was isolated with a phenol-chloroform (Roth, Karlsruhe, Germany) extraction step and washed and concentrated to a volume between 150µl to 300µl using an Amicon Ultra-15 Centrifugal Filter Unit (Merck Millipore, Darmstadt, Germany).

1.3.3 Library preparation

Libraries for all Samples were prepared after [Kircher et al., 2011] with slight modifications as mentioned in [Broushaki et al., 2016] and [Hofmanová et al., 2016]. During the adapter ligation, the amount of adapter mix used was adjusted according to the sample preservation to prevent adapter-dimer formation during the library PCR. The amount of primers, and therefore indexed adapters, used during the PCR-step was also adjusted accordingly as well.

1.3.4 Quantification and Quality control

For quality assessment, the amount of DNA for each library was measured with the QuBit flurometric system (dsDNA HS assay, Invitrogen). The fragment-length distribution of each library was

measured with the BioAnalyzer High Sensitivity Kit (HS, Agilent Technologies) and the amount of unique molecules was determined via quantitative real time PCR (qPCR) as described in [Hofmanová et al., 2016] on a Step One Plus^{TM} Real Time PCR-System (Applied Bisosystems, Thermo Fisher Scientific). In addition to the quality assessment, a library from each sample and extraction was sequenced on a Illumina MiSeq at StarSEQ GmbH (Mainz, Germany) in a shallow 50bp shotgun approach to determine its endogenous DNA-content, by aligning the resulting reads against a reference Genome and calculating the ratio of mapped reads vs. total reads sequenced.

1.3.5 Nuclear Capture enrichment

The in-solution capture enrichment [Gnirke et al., 2009] of the pre-selected nuclear markers [Veeramah et al., 2018] was performed according to the manufacturers protocol with some alterations as described in [Veeramah et al., 2018]. Sequencing was performed on an Illumina HiSeq 2000 with 100 base pair single end runs in Rapid mode at the sequencing facilities of the University of Mainz.

1.3.6 Read processing

Sequenced reads were demultiplexed at the sequencing facilities and were provided in the fastq format. This is a text-based format which stores the bases of each read with the corresponding quality values as a phred-score. Processing of the reads was conducted as described in [Broushaki et al., 2016], [Hofmanová, 2016] and in [Sell, 2017].
Residual adapter bases at the 3' end of the reads were removed as described in [Kircher, 2012]. Paired-end runs, obtained from different projects of the Palaeogenetics group were collapsed into one read, prior to further steps, by using the ea-utils package [Aronesty, 2013]. Afterwards, merged paired-end reads could be treated as single-end reads. Each library was aligned against the human reference build GRCh37/hg19 by using the burrows-wheeler aligner (BWA) [Li and Durbin, 2009] with the *aln* option with default parameters. The samtools package [Li et al., 2009] was used to transfer the alignment into a smaller binary format and also to apply a quality filter of a minimum mapping-quality of 25. After the removal of duplicate reads with MarkDuplicates from the picard tools package (picardtools, http:// broadinstitute.github.io/picard), reads shorter than 30bp were removed from the alignment, using a custom python script. Alignments of libraries belonging to one sample were merged with MergeSam of the picard tools package. The merged files were re-aligned around known indels using GATK [DePristo et al., 2011]. Depth of coverage was calculated using a custom python script.

1.3.7 Estimating contamination

Estimating contamination in ancient samples is a crucial point to prevent a bias in downstream analysis. To determine contamination rates in the samples, *contammix* version 1.0-10 [Fu et al., 2013] was used. Contamination is estimated in a likelihood-based approach using all reads that align

against the mtiochondrial genome. First, all of the reads in the sample were re-aligned against the revised Cambridge Reference Sequence. In a next step, the consensus sequence for each sample was built. This sequence was then used, in combination with a set of reference sequences, to determine the amount of reads that align against the consensus sequence vs. the amount of reads that align to any of the reference sequences.

1.3.8 Variant Detection

Determination of diploid genotypes was performed in a similar manner as described in [Hofmanová et al., 2016] and [Broushaki et al., 2016] using the software package ATLAS [Link et al., 2017]. The software incorporates sequencing errors and postmortem damage (PMD) during the variant detection and is able to correct for the false positive identification of heterozygous sites [Kousathanas et al., 2016]. First, PMD was estimated for each sample separately. The sequencing error rate was determined using the BQSR option since not all samples were male which prevented the RECAL approach on the male X-chromosome. If no sufficient amount of reads was present for one sample, samples that were sequenced together on the same lane or flow-cell were pooled together. Since they were processed in the same run, the sequencing error should not differ between samples. BQSR-tables were converted for each sample and genotypes were called separately for every possible position. The VCF-files were filtered for minimum depth of two reads per base and a minimum genotype quality of 30.

1.3.9 Mitochondrial haplogroups

In order to determine the mitochondrial haplotypes of each sample, the reads that aligned against the mitochondrial chromosome were realigned against the revised Cambridge Reference Sequence (rCRS). Therefore all reads that mapped against the mt-chromosome were extracted from the alignment, while *bamToFastq* was used to transform the reads into the fastq format. With the *samse* option of *bwa*, the reads were aligned against the rCRS. The variant detection was performed on the new alignment with GATKs Unified Genotyper [McKenna et al., 2010] with the ploidy set to one, since the mitochondrial chromosome is a uniparental marker. Variant calls were filtered for a minimum depth of two and a minimum genotype quality of 30. Haplotypes were determined by using the HaploGrep 2.0 Website [Weissensteiner et al., 2016].

1.3.10 Sex determination

In order to determine the sex of each sample, the approach of [Skoglund et al., 2013] was applied to the aligned genomes. In this approach the ratio of reads aligned to the Y-chromosome compared to the reads aligned against all sex-chromosomes is calculated. If the lower bound of the confidence intervals lies above 0.075, the sample is considered male, if the upper bound of the confidence interval

falls below 0.016, the sample is considered female.

1.3.11 Principal Component Analysis and Clustering

To visualize the affinities of the samples ,based on their genetic data, a principal component analysis (PCA) was performed for all ancient samples in addition to a reference data set with modern-day samples. This allows clustering of the ancient samples in reference to the recent individuals. The PCA was conducted using *LASER* version 2.02 [Wang et al., 2014] with modern data from [Hellenthal et al., 2014] as reference. The reference data set was sub-sampled to only include individuals from western Eurasia. The reference space is generated by a standard PCA, based on the genotypes of the modern data. In a second step for each ancient individual, a PCA is performed where the reference data is sub-sampled to resemble the coverage pattern of the ancient sample. This data is then used in a Procrustes analysis with respect to the reference data to project the ancient samples onto the contemporary data. The input is generated from BAM-files by performing a pileup over every position of the reference data.

1.3.12 Model based clustering

The model based clustering approach of ADMIXTURE [Alexander et al., 2009] was used to determine ancestral components in the ancient samples. For each sample the bases of the reference data set were called using the majority rule where at each position the bases with the highest occurrence is picked. If two bases are present at equal frequencies, a base is picked at random. Each base was duplicated, producing a pseudo-diploid call. This data set was merged with selected reference samples from [Lipson et al., 2017] and [Mathieson et al., 2015] using *plink* version 1.9 [Purcell et al., 2007]. During merging, the data set was filtered for linkage disequilibrium in a sliding window approach with a window size of 200 and r^2 set to 0.5 . The whole data set was used in 100 consecutive unsupervised runs with components K set to three and a random seed. The best fit was determined using *pong* [Behr et al., 2016].

1.4 Results

1.4.1 Coverage

Coverage and depth varied among samples depending on the bone element they were taken from. Samples taken from petrous bone had a significant higher coverage compared to samples taken from other bones (Ø 98.13%±2.10, Ø 71.12 ± 40.09 respectively, $p \leq 0.0010$,). Similarly, the average depth was higher for samples taken from petrous bone (Ø 36.29 ± 20.25) compared to samples taken from other elements (Ø 11.05 ± 11.20, $p \leq 0.001$). The sample with the highest average depth with 94.14±64.47 was VLASA4. It also had a coverage of 99.6%. The sample with the lowest coverage and depth was Fre3. This Mesolithic sample had a coverage of 1.05% and an average depth of 1.17±0.48.

Table (1) *Results for sex determination as described in Skoglund et al. [2013], mitochondrial Haplogroup with CI as assigned by HAPLOGREP 2.0, sequencing depth and coverage for the capture array, contamination rate estimated with contammix [Fu et al., 2016], bone element the sample was taken from, as well as sample date [calBC] for each sample processed in the Laboratory in Mainz. The ∼ indicates if a date is the result of indirect or contextual dating. Samples are grouped by the groups they were assigned as described in the discussion.*

Sample	**Site**	**R_y**	**CI**	**Sex**	**MT-Haplo**	**Conf.**	**Depth**	**Cov [%]**	**Cont.**	**Source**	**calBC**
EHG											
Lec2	**Crimea**	0.1057	0.1018-0.1096	XY	U5a2	0.94	10.39±4.85	99.43	4.47	O	Mesolithic
Min10	**Minino**	0.0058	0.0055-0.0061	XX	U4d	0.86	11.57±18.51	17.64	4.68	O	5650 – 4600
Min11	**Minino**	0.1333	0.1315-0.1350	XY	U4a1	0.93	46.70±62.01	98.67	0.96	O	8671 ± 48 8092 ± 94
Min2	**Minino**	0.0079	0.0073-0.0086	XX	U4a2	0.89	15.32±11.46	96.28	2.07	O	NA
Min5	**Minino**	0.0927	0.0913-0.0941	XY	U4a2	0.89	3.79±2.25	94.00	0.82	O	8740 – 8420
Min8	**Minino**	0.0921	0.0876-0.0966	XY	U4a1	0.85	2.83±2.30	5.63	1.55	O	6450 – 5800
Min3	**Minino**	0.0980	0.0969-0.099	XY	U4a1	0.71	62.50±0.27.46	99.88	1.24	O	7472±52
Zv317	**Zvejnieki**	0.1525	0.1495-0.1555	XY	U4a1	0.71	33.93±37.04	94.93	7.76	O	3890 ± 67
WHG											
Bla20	**Blaetterh.**	0.0056	0.0052-0.0060	XX	U5a2c3	0.85	7.81±5.08	96.06	2.46	O	8652 ± 58
Bla45	**Blaetterh.**	0.0895	0.0878-0.0911	XY	U5b2b2	0.96	23.94±9.84	99.64	0.44	O	3616 ± 56 3922 ± 60
Gr1	**Criewen**	0.1148	0.1080-0.1216	XY	5b2a +@16192	0.79	6.10±4.11	92.56	5.30	O	4770 ± 40
Gr2	**Criewen**	0.0064	0.0061-0.0067	XX	U4b1b1	0.79	23.85±22.43	98.36	1.93	O	4600 ± 60

Table 1 continued from previous page

Sample	**Site**	**R_y**	**CI**	**Sex**	**MT-Haplo**	**Conf.**	**Depth**	**Cov [%]**	**Cont.**	**Source**	**calBC**
GrO1	**Gr. Ofnet**	0.0956	0.0925-0.0986	XY	U5b1d1	0.84	12.92±9.73	95.22	1.31	P	late Mesolithic
LEPE18	**Lepenski Vir**	0.0970	0.0960-0.0980	XY	T2e	0.96	87.01±72.18	99.23	1.20	P	6200 – 5950
LEPE45	**Lepenski Vir**	0.0822	0.0820-0.0825	XY	U5b2c1	0.96	5.15±2.42	98.97	1.17	a	6588 – 6395
LEPE51	**Lepenski Vir**	0.0026	0.0026-0.0026	XX	U5	0.80	4.75±2.45	97.93	0.42	a	7940 – 7571
LEPE53	**Lepenski Vir**	0.0059	0.0056-0.0061	XX	H	0.96	40.26±20.79	99.48	0.26	P	∼6,200 – 6,000
VLASA10	**Vlasac**	0.0929	0.0919-0.0939	XY	K1+16362	0.96	42.13±26.56	99.61	0.04	P	∼7,400 – 6,200
VLASA32	**Vlasac**	0.0949	0.0941-0.0958	XY	U5a2 +16294	0.88	47.85±30.00	99.28	0.05	P	∼7,400 – 6,200
VLASA37	**Vlasac**	0.0832	0.0830-0.0834	XY	K1f	0.92	6.24±2.74	99.58	0.56	a	6,767 – 6,461
VLASA41	**Vlasac**	0.0066	0.0064-0.0069	XX	U5b2b	0.93	62.29±43.90	99.36	0.07	P	∼7,400 – 6,200
VLASA44	**Vlasac**	0.0977	0.0966-0.0987	XY	U5b2b	0.97	71.73±49.55	99.40	0.10	P	∼7,400 – 6,200
VLASA4	**Vlasac**	0.0959	0.0950-0.0968	XY	U5a1c	0.94	94.14±64.47	99.60	1.38	P	∼7,400 – 6,200
Zv122	**Zvejnieki**	0.1068	0.1030-0.1105	XY	U5a2d	0.92	3.93±3.38	7.55	0.72	O	5383 ± 68
Zv162	**Zvejnieki**	0.1257	0.1236-0.1279	XY	U4a1	0.89	23.14±25.29	92.21	2.62	O	4470 ± 72
Zv170	**Zvejnieki**	0.1018	0.0972-0.1063	XY	U5a2d	0.93	1.55±1.05	4.06	2.73	O	7182 ± 107
Zv39	**Zvejnieki**	0.0967	0.0920-0.1015	XY	U5a2c	0.93	4.38±2.87	6.00	3.53	O	5681 ± 36
Zv76	**Zvejnieki**	0.0060	0.0056-0.0065	XX	U5a2	0.92	4.24±5.14	15.05	0.26	O	5802 ± 73
SEF											

Table 1 continued from previous page

Sample	**Site**	R_y	**CI**	**Sex**	**MT-Haplo**	**Conf.**	**Depth**	**Cov [%]**	**Cont.**	**Source**	**calBC**
AKT16	**Aktop raklik**	0.0096	0.0093-0.0099	XX	K1a3	0.96	40.89±40.68	97.00	0.08	P	6683 – 6533
AKT18g	**Aktopraklik**	0.0104	0.0098-0.0111	XX	H2a2a	1.00	8.77±8.52	86.75	0.33	P	6469 – 6393
AKT20g	**Aktopraklik**	0.0904	0.0889-0.0919	XY	J2b1	0.97	23.63±16.56	98.64	3.01	P	6493 – 6418
AKT26g	**Aktopraklik**	0.0075	0.0071-0.0079	XX	J2b	0.78	22.16±23.14	95.10	3.13	P	~6,500-6,000
AKT6g	**Aktopraklik**	0.0993	0.0978-0.1007	XY	T2b	0.88	36.40±30.83	98.29	3.95	P	5633 – 5535
BAR11	**Barçin**	0.0966	0.0947-0.0984	XY	X	0.77	27.40±23.45	98.27	4.96	P	~6,600-6,000
BAR15	**Barçin**	0.0114	0.0111-0.0118	XX	K1a	0.90	29.10±26.64	97.75	0.19	P	6213 – 6049
BAR16	**Barçin**	0.0096	0.0093-0.0100	XX	K1a1	0.72	26.33±28.84	95.14	1.15	P	6233 – 6084
BAR20	**Barçin**	0.0096	0.0093-0.0100	XX	W	0.77	15.30±10.37	98.05	0.21	P	6438 – 6258
BAR31	**Barçin**	0.0907	0.0904-0.0909	XY	K1a2	0.92	4.67±2.77	96.51	0.04	P	6,419 – 6,238
BAR32	**Barçin**	0.0079	0.0076-0.0082	XX	X2m2	0.88	14.99±10.08	98.01	1.38	P	6396 – 6241
BAR8	**Barçin**	0.0066	0.0066-0.0066	XX	K1a2	0.91	8.61±4.24	99.46	1.15	P	6,212-6,030
LEPE39	**Lepenski Vir**	0.0914	0.0899-0.0929	XY	U5a2d	0.91	26.09±16.16	99.17	1.71	P	6200 – 5950
LEPE52	**Lepenski Vir**	0.0989	0.0978-0.0999	XY	U4a2	0.88	67.70±47.78	99.38	0.25	P	6005 – 5845
CEF											
281-19-6	**Herxheiim**	0.0062	0.0058-0.0066	XX	K1a4a1i	0.96	10.78±7.49	94.94	1.69	P	~5000
282-104-4	**Herxheiim**	0.0909	0.0895-0.0924	XY	K1a1a	0.93	38.08±24.77	99.18	0.11	P	~5000

Table 1 continued from previous page

Sample	Site	R_y	CI	Sex	MT-Haplo	Conf.	Depth	Cov [%]	Cont.	Source	calBC
282-126-16	**Herxheiim**	0.0059	0.0057-0.0062	XX	T2e	0.95	31.92±17.34	99.10	1.28	P	~5000
282-126-7	**Herxheiim**	0.1398	0.1375-0.1420	XY	J2b1	1.00	62.21±46.76	99.19	1.47	P	~5000
282-13-7	**Herxheiim**	0.0054	0.0050-0.0058	XX	W1+119	0.90	18.23±11.73	97.15	2.27	P	~5000
282-23-1	**Herxheiim**	0.0981	0.0968-0.0994	XY	K1a	0.92	69.96±42.88	99.31	1.41	P	~5000
282-88-2	**Herxheiim**	0.0043	0.0041-0.0046	XX	K1a+150	0.93	63.09±42.01	99.30	0.21	P	~5000
282-94-11	**Herxheiim**	0.1001	0.0989-0.1014	XY	HV+16311	0.91	33.36±20.85	98.90	0.62	P	~5000
7034	**Herxheiim**	0.0052	0.0050-0.0054	XX	U3	0.92	40.40±21.38	99.44	0.10	P	~5000
Asp10	**Asparn-Schlez**	0.0987	0.0976-0.0998	XY	K1a	0.98	38.29±22.20	99.08	1.29	P	5500 – 5000
Asp1	**Asparn-Schlez**	0.0915	0.0901-0.0928	XY	K1a	0.95	20.26±11.12	98.98	0.70	P	5500 – 5000
Asp2	**Asparn-Schlez**	0.0972	0.0955-0.0988	XY	N1a1a1a2	0.88	16.42±9.58	98.80	3.91	P	5500 – 5000
Asp3	**Asparn-Schlez**	0.0040	0.0038-0.0042	XX	J1c17	0.91	30.44±18.06	98.57	0.20	P	5500 – 5000
Asp4	**Asparn-Schlez**	0.0045	0.0043-0.0047	XX	T2b	0.87	26.66±13.38	99.19	2.34	P	5500 – 5000
Asp6	**Asparn-Schlez**	0.0884	0.0872-0.0897	XY	U5a1c1	0.94	22.24±11.63	99.22	2.19	P	5500 – 5000
Asp8	**Asparn-Schlez**	0.0046	0.0043-0.0049	XX	X2b	0.95	55.85±31.50	99.18	0.11	P	5500 – 5000
Dil15	**Dillingen**	0.1035	0.1019-0.1051	XY	J1c	0.95	32.56±22.07	98.53	1.16	P	5500 – 5000
Dil16	**Dillingen**	0.1094	0.1076-0.1113	XY	J1c	0.92	28.57±19.37	98.27	1.73	P	5500 – 5000
Ess7	**Essenbach-Am.**	0.0917	0.0904-0.0930	XY	U5b2c1	0.94	14.44±10.07	95.72	0.23	P	5500 – 5000

Table 1 continued from previous page

Sample	**Site**	R_y	**CI**	**Sex**	**MT-Haplo**	**Conf.**	**Depth**	**Cov [%]**	**Cont.**	**Source**	**calBC**
Klein10	**Kleinhadersd.**	0.0053	0.0052-0.0055	XX	T2b	0.95	37.93±26.13	97.95	2.00	P	~5950 – 5150
Klein1	**Kleinhadersd.**	0.0041	0.0039-0.0043	XX	N1a1a1	0.88	54.67±32.56	98.88	2.43	P	~5950 – 5150
Klein2	**Kleinhadersd.**	0.0046	0.0043-0.0048	XX	H2a	0.77	37.31±20.91	98.87	1.03	P	~5950 – 5150
Klein3	**Kleinhadersd-**	0.0046	0.0044-0.0048	XX	J1c2	0.96	27.08±13.46	99.33	0.09	P	~5950 – 5150
Klein4	**Kleinhadersd.**	0.0035	0.0034-0.0037	XX	T2b	0.95	39.94±26.36	98.10	1.79	P	~5950 – 5150
Klein5	**Kleinhadersd.**	0.0937	0.0925-0.0948	XY	N1a1a1a3	0.92	35.68±25.48	97.14	1.40	P	~5950 – 5150
Klein8	**Kleinhadersd.**	0.0044	0.0042-0.0045	XX	U5b	0.92	45.06±24.94	99.29	0.24	P	~5950 – 5150
MN											
Bla1	**Blaetterh.**	0.1136	0.1006-0.1266	XY	H1bm	0.66	2.21±1.59	3.94	0.49	O	3508 ± 102
Bla10	**Blaetterh.**	0.0914	0.0874-0.0954	XY	J1c3c	0.51	4.41±2.97	90.73	3.55	O	3418 ± 63
Bla13	**Blaetterh.**	0.0057	0.0044-0.0071	XX	U5b2a +@16192	0.68	6.78±4.53	5.09	0.47	O	3513 ± 102
Bla15	**Blaetterh.**	0.0982	0.0863-0.1102	XY	H5	0.61	1.71±1.15	2.99	0.72	O	3571 ± 47
Bla17	**Blaetterh.**	0.1164	0.1097-0.1231	XY	H1ba	0.59	4.73±4.16	5.33	5.35	O	3681 ± 19
Bla28	**Blaetterh.**	0.1005	0.0910-0.1100	XY	J1c8	0.74	1.49±0.89	3.62	2.51	O	3196 ± 103
Bla29	**Blaetterh.**	0.0067	0.0049-0.0085	XX	H2a2	0.82	2.18±1.51	66.60	2.96	O	3020 ± 61
Bla32	**Blaetterh.**	0.0052	0.0049-0.0055	XX	H5	0.94	52.61±29.45	99.52	1.22	O	
Bla59	**Blaetterh.**	0.1071	0.1036-0.1106	XY	U5b2a2	0.88	12.11±7.41	98.38	2.49	O	3869 ± 59

Table 1 continued from previous page

Sample	Site	R_y	CI	Sex	MT-Haplo	Conf.	Depth	Cov [%]	Cont.	Source	calBC
WEZ16	**Weltzin**	0.0040	0.0031-0.0049	XX	T2b	0.97	1.32±0.62	39.42	0.55	O	3026 – 2893
WEZ											
WEZ15	**Weltzin**	0.0992	0.0974-0.1010	XY	U2e1a1	0.96	3.39±1.85	94.50	0.30	O	1129 – 924
WEZ24	**Weltzin**	0.0795	0.0765-0.0825	XY	H27	0.86	6.94±3.26	98.64	0.11	O	~1129 – 924
WEZ35	**Weltzin**	0.0901	0.0876-0.0925	XY	K1c1	0.96	13.31±5.66	99.66	0.08	O	~1129 – 924
WEZ39	**Weltzin**	0.0839	0.0803-0.0876	XY	J2b1a1	0.97	4.79±2.48	97.03	1.35	O	~1129 – 924
WEZ40	**Weltzin**	0.0996	0.0968-0.1024	XY	T1a1	0.92	11.62±5.36	99.30	2.36	O	~1129 – 924
WEZ48	**Weltzin**	0.0836	0.0787-0.0885	XY	J1c	0.88	2.96±1.76	88.22	7.91	O	~1129 – 924
Wez51	**Weltzin**	0.0927	0.0906-0.0949	XY	H1c	1.00	22.99±9.69	99.73	0.13	O	~1129 – 924
WEZ53	**Weltzin**	0.0999	0.0964-0.1034	XY	U2e2a1a2	0.93	3.29±2.02	89.94	0.17	O	~1129 – 924
WEZ54	**Weltzin**	0.0902	0.0878-0.0926	XY	V3a	0.92	13.05±5.36	99.42	3.50	O	~1129 – 924
WEZ56	**Weltzin**	0.0881	0.0859-0.0903	XY	T2b	0.97	25.03±10.43	99.64	0.15	O	~1129 – 924
WEZ57	**Weltzin**	0.0876	0.0854-0.0898	XY	H2a1	1.00	20.43±8.22	99.59	0.49	O	~1129 – 924
WEZ58	**Weltzin**	0.0796	0.0772-0.0819	XY	T2b	1.00	8.54±3.68	99.36	0.24	O	~1129 – 924
WEZ59	**Weltzin**	0.0948	0.0923-0.0972	XY	U5a2b1a	0.96	7.73±3.41	99.35	0.74	O	~1129 – 924
WEZ61	**Weltzin**	0.0031	0.0028-0.0034	XX	U4b1b1	0.94	17.21±6.53	99.58	1.02	O	~1129 – 924
WEZ64	**Weltzin**	0.0993	0.0965-0.1022	XY	I1a1a	0.96	2.54±1.49	86.99	0.08	O	~1129 – 924

Table 1 continued from previous page

Sample	Site	R_y	CI	Sex	MT-Haplo	Conf.	Depth	Cov [%]	Cont.	Source	calBC
WEZ71	**Weltzin**	0.0902	0.0869-0.0935	XY	J1c	0.90	20.57±9.50	99.30	0.17	O	~1129 – 924
WEZ74	**Weltzin**	0.0022	0.0017-0.0027	XX	T2b	0.98	9.79±4.42	98.90	1.66	O	~1129 – 924
WEZ83	**Weltzin**	0.0962	0.0938-0.0987	XY	I4a	0.94	15.78±7.43	98.99	0.68	O	~1129 – 924
OUT											
Bla75	**Blaetterh.**						1.20±0.56	5.92	46.77	O	
Fre3	**Gr. Fredenw.**	0.0895	0.0758-0.1031	XY	R0	0.75	1.17±0.48	1.05	0.50	O	6200 – 5400

1.4.2 Contamination

The overall rate of contamination in the samples, estimated from mitochondrial DNA, was low. The estimates varied between 0.038% and 46.77%, where the sample with the highest contamination rate (Bla75) was a clear outlier. The next highest contamination rate was found in the sample WEZ48 with 7.91 %. The average contamination rate for all samples was 2.06 %. When the outlier was removed, the average contamination rate went down to 1.56 %. When samples were grouped according to their bone elements, it became evident that the average contamination rate in samples taken from petrous bone was significantly lower compared to samples taken from other bone parts (1.2%, 2.4% respectively, $p=0.001$).

1.4.3 Sex determination

Males and females were present at all sites were multiple samples were taken from. In total, 61.7% of the samples were male, while 38.3% were female. Two sites show a particularly biased sex ration: at the early Neolithic site of Kleinhadersfor in Austria one male was found among seven sequenced samples, while at the Bronze Age site of the Tollense valley only two females were found among the 18 sequenced samples .Two sites show a particularly biased sex ration: at the early Neolithic site of Kleinhadersfor in Austria one male was found among seven sequenced samples, while at the Bronze Age site of the Tollense valley only two females were found among the 18 sequenced samples. The sample from the site in Criewen (Gr2), which was found buried with a rich decoration of mollusk shells and teeth had in earlier studies been described as male [Street et al., 2001], but was identified as female in this analysis.

1.4.4 Mitochondrial haplogroups

The majority of the samples with a hunter-gatherer background, were assigned to a U haplogroup (83 %). Of those U haplogroups, 62 % were assigned to U5, while 38 % were assigned to U4. The only other haplogroups found were H, K1, T2 and R0. All samples from the Minino were assigned to haplogroup U4.

Among the early Neolithic samples, 13 major haplogroups were found. K1a was assigned to 28.21 % of the samples, while T2 and U5 were assigned to 12.82 % each. 10.26 % were assigned to haplogroup Jc1, while the rest had either H2 HV1, J2b, N1a1, U3, U4, W, W1, X or X2. Of the 11 samples that date to the Middle/Late Neolithic period, 10 samples yielded sufficient data to determine mitochondrial haplogroups. The sample Bla75 was excluded, due to high contamination rates and low coverage. For those 10 samples, the haplogroups H1b, H5, J12c and U5 were each found twice, while T2ba and H2a were found in one sample each. The samples from the Bronze Age battlefield displayed a high heterogeneity in their distribution of haplogroups. T2b was present at a frequency of 16.67 %, while U2e and J1c were found at 11.11 %. All other haplogruops, namely H1c, H27, H2a1, I1a1a, I4a, J2b1a1, K1c1, T1a1, U4b, U5a and V3a were assigned once each (see table: 1 for details).

1.4.5 Model-based clustering

In samples that were found in Central European hunter-gatherer context, the blue component was maximized, while in the hunter-gatherer samples from the Russian site of Minino, as well as the Ukrainian sample and one sample from Zvejnieki in Latvia (Zv317), an additional orange component was present at intermediate levels. In all but one case (Lec2), the orange component was bigger compared to the blue component. All samples originating from an early farmer context, were dominated by a green component, but the majority of these samples also displayed small amounts of the blue component that was maximized in Central European hunter-gatherers. The samples from Barçin had the lowest amount of blue components, compared to the other early farmer samples. The samples dating to the Middle Neolithic mainly originated from the Blätterhöhle. Only one sample from the Tollense Valley fell into that period. All of those samples displayed a blue component. A further seven samples also showed amounts of the green component that was maximized in early farmers. One sample also displayed low to medium amounts of the orange component that was also present in the hunter-gatherers from Russia. Two samples displayed no other component. The Bronze Age samples from the Tollense Valley displayed all three components, to almost equal amounts.

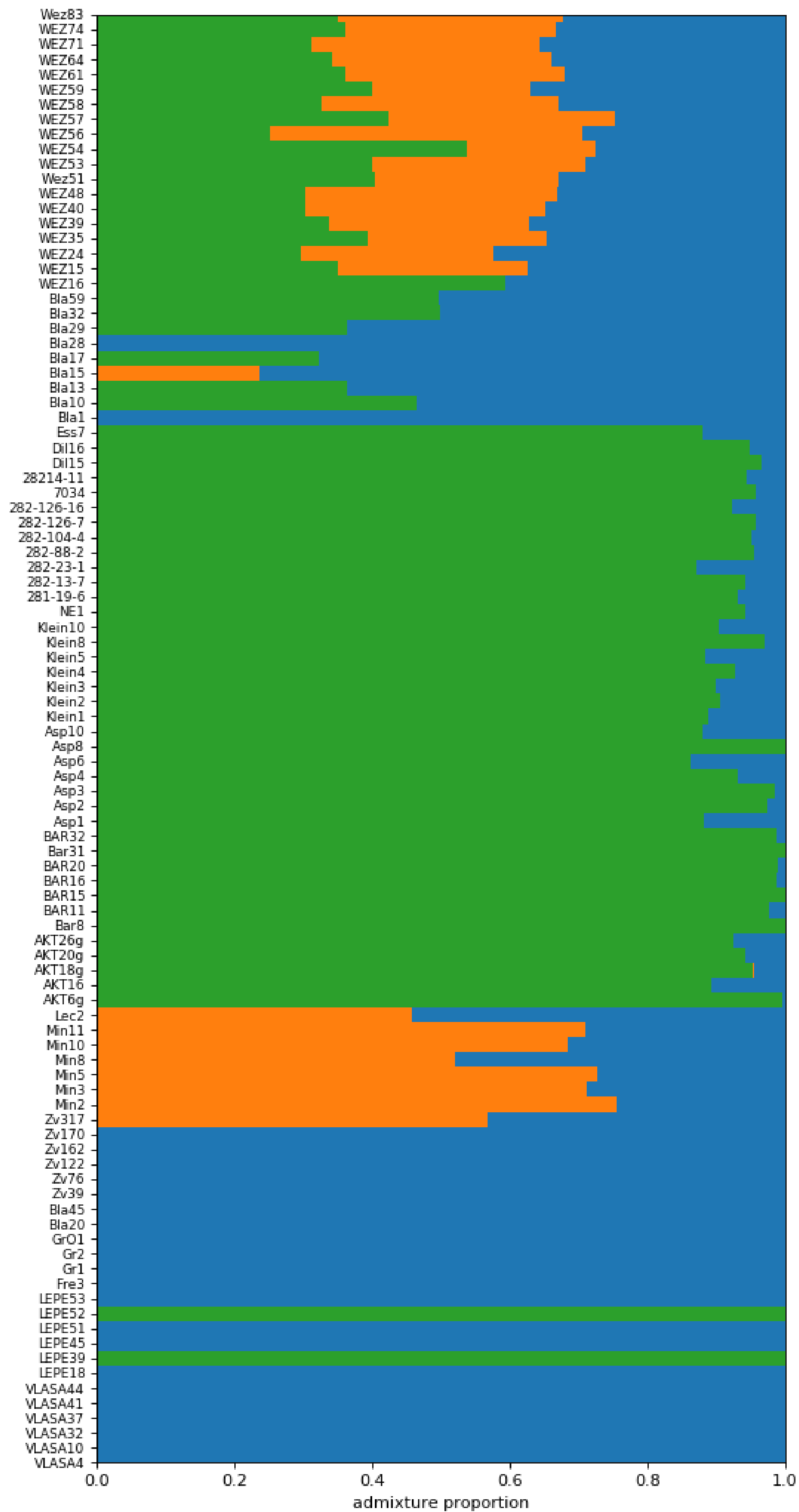

Figure (3) *Results from the ADMIXTURE run with components K set to three. Samples are displayed without the reference data set and are grouped by location.*

1.4.6 Principal component analysis

All samples that date prior to the bronze age clustered outside of the variation of the modern Central European reference data. Samples that were associated with early European farmers, clustered to the lower left of the plot. This cluster is located in close proximity to modern-day individuals from Sardinia and contains the samples from Asparn-Schletz (Austria), Kleinhadersdorf (Austria), Dillingen-Steinheim, Essenbach-Ammerbreite and Herxhheim (all Germany) who are all associated with the Linearbandkeramik culture of central Europe. Samples from Barçin and Aktopraktlik (North-western Turkey) were also part of the same cluster to the lower right of the plot. Two samples from Lepenski Vir (LEPE39 & LEPE53) also fell within this cluster, associated with early farmers.

Samples associated with a hunter-gatherer context from Southern and Central Europe, as well as from the Baltic, clustered left of the modern-day European reference. However, there were two exceptions. The sample Fre3 from Groß Fredenwalde in Germany was placed in the middle of the plot, above the Central European Reference. Sample Zv317 from the Baltic site in Latvia also fell outside the other Central European hunter-gatherer samples.

The hunter-gatherers found in Minino, Russia, clustered above the European reference data set. The sample from Lesnik Cave also fell within this cluster, as well as the Zv317 sample from Zvejnieki, Latvia.

One smaller cluster could be determined below the Central European reference data set, between the early farmer cluster and the European hunter-gatherer cluster. The samples in this cluster were mainly from the Blätterhöhle in Germany and date to the Middle Neolithic. Two samples from the Tollense Valley also fell into that cluster (WEZ16 and WEZ56). One of the samples dates also to the late Neolithic (WEZ16), while the other dates to the Bronze Age. Another cluster could be identified, on top of the Central European Reference data set. All samples in this cluster originate from the Bronze Age sampling site int the Tollense valley.

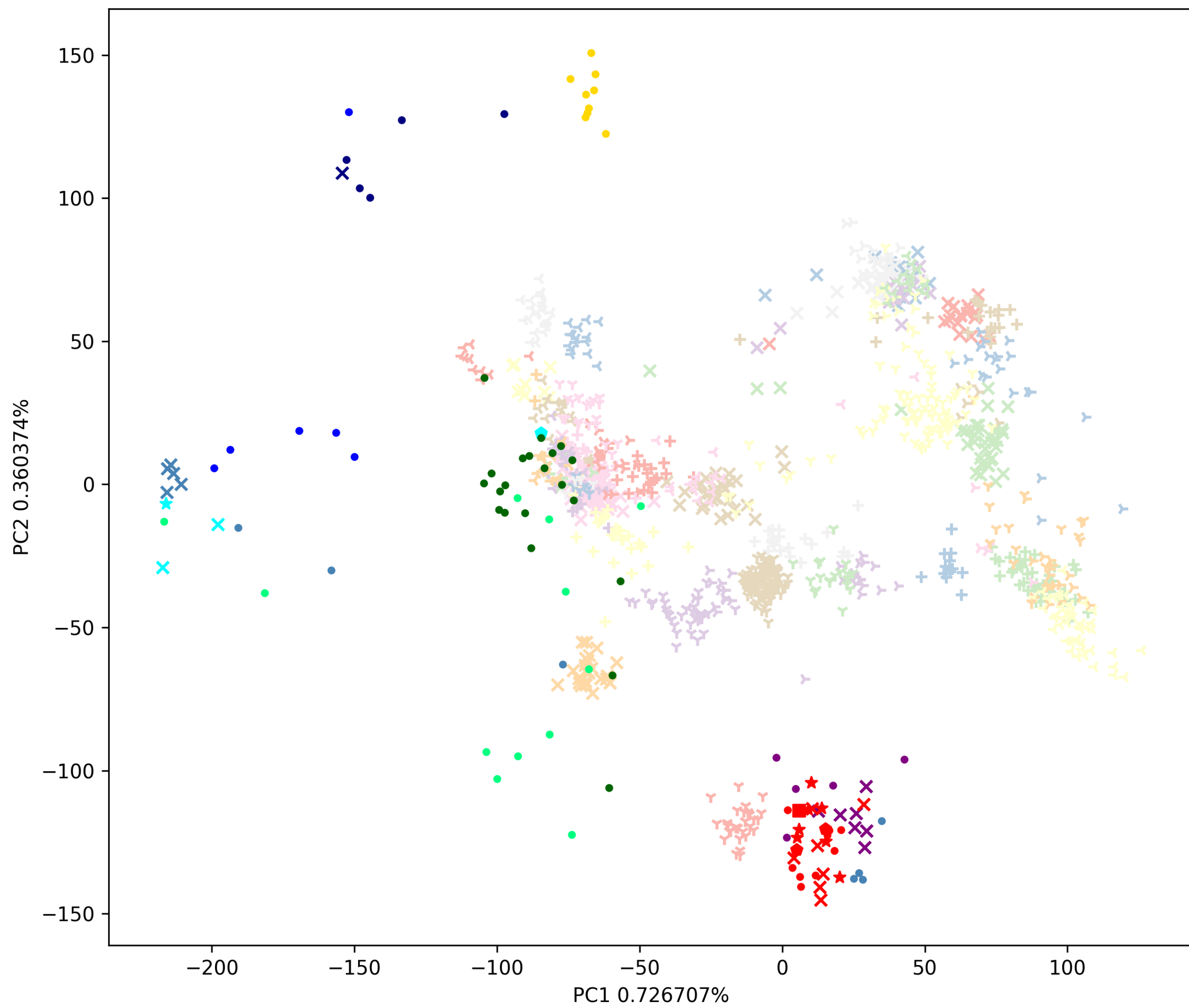

Figure (4) *Plot of the first two principal components from a PCA of the ancient samples prepared in the laboratory in Maniz with additional data for individuals with steppe ancestry labeled 'Yamnaya' from Mathieson et al. 2015, projected onto contemporary reference data from Hellenthal et al. 2014.*

1.5 Discussion

Grouping of the samples

The aim of this study was to identify the influence of subsistence and eco-region on the genomes of early Europeans. To be able to track changes in allele frequencies over time, the samples had to be joined in groups. This was done based on the archaeological context the samples were found in, as well as population genetic analysis.

Based on the high contamination rate of over 45% and the very low coverage and sequencing depth, the sample Bla75 from the Blätterhöhle was excluded from further analysis. The sample Fre3 from Groß Fredenwalde in Germany was also excluded. The sample was found in a Mesolithic hunter-gatherer context, yet due to the very low coverage and sequencing depth, this could not be confirmed on the basis of PCA. The sample fell within the variation of the European reference population, when projected onto the PCA-plot. Since this could be caused by either the low coverage or contamination, the sample was excluded.

The division of early farmers and hunter-gatherers based on the archaeological context was mirrored by the clustering approaches. The two sample groups were clearly distinguishable on the PCA plot, where they appeared on opposite sides of the plot. Confirming previous studies [Lazaridis et al., 2014, Haak et al., 2015], the early farmers clustered on the lower left of the plot, while samples from a hunter-gatherer context clustered to the upper and middle left. Separation of the samples was furthermore reinforced by the ADMIXTURE results. The mitochondrial haplogroups determined for the samples were also in accordance with previous findings [Bramanti et al., 2009].

Two groups of hunter-gatherers

Based on the PCA and ADMIXTURE analysis, the hunter-gatherers could be subdivided in two groups. On the PCA plot the hunter-gatherer samples appeared in two clusters. One to the middle left of the European reference data and one on the upper left. These two clusters were also found in the ADMIXTURE analysis, where the samples from the lower PCA-cluster were described by the blue component, while the upper PCA-cluster was described as a mixture between the blue and the orange component. The blue-orange cluster consisted of all samples from the Russian site of Minino, as well as the Ukrainian sample from the Lesnik Cave, Crimea (Lec2), and one sample from the Latvia site of Zvejnieki (Zv317). Because of their eastern European origin and in accordance with literature [Haak et al., 2015, Mathieson et al., 2018], the group was termed **Eastern European hunter-gatherer (EHG)**. The blue cluster consisted of samples from Central and Southern Europe. Since the hunter-gatherer sample from the Loschbour in Luxembourg Lazaridis et al. [2014] also clustered among these samples, the group was termed **Western European hunter-gatherers (WHG)**. In addition to the Loschbour individual, two additional sample from the literature were added to the group. A Swiss hunter-gatherer from the "Grotte du Bichon" [Jones et al., 2015] and one individual from a site in Spain [Olalde et al., 2014].

The northernmost sampling site with samples clustering with the WHG, is located in Latvia (img.: 2). All samples but the before-mentioned Zv317 sample can be considered WHG, based on the PCA and the ADMIXTURE plot. Five samples in the WHG cluster originate from Germany. Two are from the northern German sampling site of GroßFredenwalde (Gr1, Gr2), one is from the Central German site Große Ofnet Höhle and another two are from the Blätter Höhle in Hagen (Bla20, Bla45). All of those samples date to the Mesolithic, except the Bla45 sample. This sample is from the Middle Neolithic period, when the Blätter Höhle was used simultaneously by hunter-gatherers and farmers [Orschiedt et al., Bollongino et al., 2013]. Interestingly, the sample showed no early farmer related components in the ADMIXTURE plot. This may be due to the fact that the used capture approach has not enough resolution, or the samples ancestors didn't experience any admixture from early European farmers.

Two groups of early farmers

The early farmers samples were also subdivided into two groups. However, no division was possible based on the PCA and ADMIXTURE analysis alone. This is due to the fact, that early farmers of Central Europe were direct descendants of early farmers from the broader Aegean region in north-western Anatolia and mainland Greece [Hofmanová et al., 2016]. Therefore, samples were grouped according to their dating and location. For the first group, samples from the two Turkish sites in the eastern Marmara region, Aktopraklik and Barçin [Hofmanová, 2016, Hofmanová et al., 2016] were combined with the Neolithic samples from Serbian site Lepenski Vir, based on their close proximity in age and location. For their south-eastern European origin the group was termed **South-eastern European Farmers (SEF)**. The samples span over an age of approximately 1,000 years, from 5633 – 5535 cal BC to 6683 – 6533 cal BC. But all except one sample from Aktropraklik were older than 6000 BC and thus pre-date the central European Neolithic samples by 500 years.

For the second early farmer group, the samples from Austria and Germany were grouped together. The samples from the sites in Asparn-Schletz (Austria), Kleinhadersdorf (Austria), Herxheim (Germany), Essenbach-Ammerbreite (Germany) and Dillingen (Germany) were all associated with a LBK context. In addition to the samples produced in the ancient DNA facilities in Mainz, two additional samples from the literature were added to the group. Both samples were also associated with the LBK culture of central Europe. One originates from Stuttgart in Germany [Lazaridis et al., 2014] and the other one from a site in Hungary [Gamba et al., 2014]. Due to the Central European origin, the group was termed **Central European Farmers (CEF)**. Including the literature data, the group size reached 29 samples, making it the biggest group in the study.

The Middle Neolithic samples **(MN)** were grouped together, mainly based on the dating. Although they formed a loose cluster on the PCA plot, the samples could not be grouped on a genetic basis alone. All but one sample from that period originated from the Blätterhöhle in Germany. The majority of those samples was characterized by a poor preservation, leading to poor sequencing results. The low coverage and sequencing depth of those samples made their placement on the PCA plot. The low resolution also affected the ADMIXTURE run. However, all samples were dated between

3,800 BC and 3,000 and were found mainly at the Blätterhöhle in Hagen. They therefore fall into the Middle to Late Neolithic period in Germany, but predate the arrival of steppe-related ancestry in the area [Haak et al., 2015]. The only sample not from the same site was a female (WEZ16) from a burial site in the Tollense Valley in northern Germany. On the PCA plot this sample also fell in the loose cluster between WHG and early farmers and dated to the same period and was therefore included in the group.

The classification of the last group was based mainly on the archaeological context. All samples from this group were found at the Bronze Age battlefield at the Tollense Valley, in the Weltzin river [Brinker et al., 2013]. Therefore, the group was termed **Weltzin (WEZ)**. The results from the PCA as well as the ADMIXTURE analysis further emphasize the grouping. All samples were composed of three components and plotted on top of the Central European reference population, which is consistent with other samples that post-date the arrival of steppe-related ancestry in Central Europe [Allentoft et al., 2015, Olalde et al., 2018]. This also suggests that some population continuity between the people of the Tollense Valley and the populations of Central Europe can be assumed.

Conclusion

A total of 101 samples was divided into six groups, based on the archaeological background, the dating of the sample, and the population genetic analysis. The hunter-gatherers could be divided mainly based on the PCA and ADMIXTURE analysis. The early farmers in this data set were divided mainly based on the archaeological context and the dating of the samples. One group mainly consisted of Central European farmers, who all have been dated between 5,500 BC and 5,000 BC. Another group of early farmers was composed of Southern European farmers, from the broader Aegean region and Serbia. Two more groups were formed, one consisting of all samples dating to the Middle Neolithic, while the other group consisted entirely of remains from a Bronze Age battlefield from northern Germany. All subsequent analyses, described in the following chapters are based on the groups, identified here.

2 Methods used for the analyses of the functional markers

2.1 Introduction

Detecting selection in population genetic data is challenging, due to the variety of influences that can impact allele frequencies. To disentangle these various signals and identify markers that were affected by selection, a special set of methods is required. This chapter describes the methods that were applied to this dataset and how they are equipped to the task.

Natural selection is often described as the non-random development of an allele in a population, in consequence to its phenotypic effect. If the effect of the phenotype is positive on the organisms fitness, in terms of survival or reproductive success, the frequency of the associated allele increases. If the effect is detrimental, the associated allele frequencies will decrease over time [Gillespie, 2004]. Any locus in the genome could be affected by selection, if the locus is polymorphic and at least one allele has a phenotypic non-neutral effect. Variation in a population at a specific locus can arise from novel mutation or via introduction by admixture. The majority of new mutations is assumed to be deleterious and is therefore purged from the population gene pool. This process is often referred to as 'background selection'. If a new mutation increases to high frequencies or reaches fixation in a population due to positive selection, it is described as a 'hard sweep'. In contrast, an incomplete or partial sweep that leads to elevated or intermediate frequencies is often described as 'soft sweep' [Pritchard et al., 2010].

It is challenging to detect selection and the methods that can be applied vary with the available data. Before the availability of genome-wide data, measures of population differentiation, such as F_{st}, were often used to identify loci that deviate from neutral expectations. Neutral expectations describe the null-model which assumes a large but finite population size, non-overlapping generations, and drift as the only factors acting on allele frequencies. Drift is the effect of the 'random sampling' of alleles during the formation of a new generation [Hamilton, 2011]. If the assumption of random mating is not violated, the strength of drift is only dependent on the effective population size (N_e). The time to fixation in generations (T_{fixed}) of an allele at the frequency p due to drift can be estimated by the formula from Kimura and Ohta [1969]:

$$T_{\text{fixed}} = \frac{-4N_{\text{e}}(1-p)ln(1-p)}{p} \tag{1}$$

According to this formula, a small population size can increase the effect of drift, while a larger population size will reduce the effect. Over time, drift leads to the decrease in diversity in an isolated population. In addition to drift, other processes can also affect the genetic diversity in a population [Felsenstein, 2005]. For example, during migration, diversity is often reduced. When a sub-population branches of and settles in a new environment, only a fraction of the original diversity will be present. This is known as 'Founder effect' and can be observed in populations worldwide. Estimates could show, that genetic diversity is decreasing with increasing distance to Addis Abeba [Li et al., 2008]. This is believed to be a direct result from serial founder effects after the Out-of-Africa migration. Furthermore, it was found that after the split of the first populations to migrate out

of Africa, the sub-populations that became ancestral to modern-day South-East Asians experienced more drift, than Central European populations [Coop et al., 2009].

If a population increases in size after an initial settlement, allele frequencies can rise rapidly in frequency, while others will decline. In contrast, a rapid decline in population size or 'bottleneck', also leads to a loss in diversity, similar to the founder effect. If two populations are isolated, due to distance or a physical barrier, drift and the differential demographic history will increase their genetic distance [Hamilton, 2011]. This phenomenon can be visualized in a PCA plot, where the Euclidean distance between the principle components of the samples correlate with the geographic distance [Novembre et al., 2008].

Neutrality tests

To be able to distinguish between signs originating from demography and genuine signs of selection, a multitude of tests was developed, many of which are based on the idea that selection distorts patterns of neutral variation [Barreiro et al., 2008]. If an allele is under selection it will rise in frequency, but because of varying recombination rates, other frequencies of variants that are located in close proximity on the same chromosome will also increase. This can be measured in the site frequency spectrum (SFS) which sums up the number of derived alleles at each frequency for a specific site. Selection at at locus in combination with hitch-hiking of linked alleles will shift the SFS. Under neutral conditions, new mutations can arise, increasing the numbers of low-frequency-alleles at a certain locus, while selection will decrease this number. The SFS is often also used in simulation and modeling approaches, where an explicit scenario is simulated and the generated data is compared to observed variation. Thereby, parameters that shaped a the populations SFS can be estimated [Gutenkunst et al., 2009].

Another method relying on the SFS is Tajima's D [Tajima, 1989]. This estimator compares the mean number of pairwise differences, with the number of segregating sites for a defined window at the chromosome. If D is close to 0, neutrality can be expected as both measurements are equal. A D-value smaller than 0 indicates the presence of rare alleles at elevated frequencies. This can be caused by a selective sweep, or bottleneck, followed by population growth. In general, this suggests a loss of variation. D-values higher than 0 indicate the presence of rare alleles at low frequencies and a general increase in variation. This can be a sign for balancing selection, but also for a decline in population size. Tajima's D is often used in a sliding window approach. Windows of a fixed size are shifted across the chromosome in small steps. The distribution of values is then used to determine outliers in the extreme ends. By comparing the results to each other it is possible to distinguish selection that only affected certain loci, from demography, which often has a broader impact on the whole genome.

To be able to calculate Tajima's D, the phase of each genotype has to be known. Phasing genotypes describes the process of determining which allele originates from which parent. This also allows the assignment of haplotypes. Haplotypes are neighboring alleles that are inherited jointly due to link-

age. Strong selection in a population on a certain allele leads to the development of long haplotypes. Several statistics that try to detect selection, rely on the identification of haplotypes. Garud et al. [2015] developed a method that measures the haplotype homozygosity at a certain locus (H). In addition, this methods can also be used to calculate the combined haplotype homozygosity of the most frequent two (H12) and three (H123) haplotypes. By dividing the frequencies of the second by the first most frequent haplotype (H2_H1), it is possible to estimate if balancing or directional selection affected a locus. A high H2_H1 value indicates, that more than one haplotype is present at high frequency. This, in combination with a high H12 value, suggests balancing selection. In contrast, a low H2_H1 value in combination with a high H12 value indicates, that only one haplotype is present at high frequency, suggesting directional selection. Similar to Tajima's D, this approach is done in a sliding window approach, so outliers can be determined.

In addition to haplotype frequencies, their length is also used to measure neutrality in populations. Selection of a beneficial allele causes alleles in linkage disequilibrium to rise along in frequency, thus creating long homozygous haplotypes. Extended haplotype homozygosity (EHH) is defined by measuring the distance outwards from the core (selected) allele in both directions. The distance is defined by the probability that any two randomly chosen chromosomes in a population, carrying the core region, are identical by descent for the whole region. The long-range haplotype test (LRH) can for example be used to compare the frequency of a haplotypes with the EHH, and thus identify haplotypes that are long as well as common [Sabeti et al., 2002]. When a sweep is complete, recombination will start to break down the haplotype structure and mutations will start to appear. This allows to estimate the time that has passed since a potential sweep. Another method based on the EHH is the integrated haplotype score (iHS) [Voight et al., 2006]. The iHS calculates the area under the curve, defined by EHH for a core SNP in a population. As the score incorporates the EHH for the derived and the ancestral allele, knowledge of the alleles state is required. This can be gained by comparing the genotype data to an ancestral outgroup, such as the genome of a chimpanzee or an ancient hominin. In a similar approach, Ferrer-Admetlla et al. [2014] computed the number of segregating sides by length (nS_L), which also compares haplotypes for the ancestral and derived alleles and came to similar results as when using the iHS. Other methods compare haplotype length in different populations (cross-population extended haplotype homozygosity or XP-EHH) which can lead to a deeper understanding of the haplotypes age and history [Zhong et al., 2011].

With the wide availability of genome-wide data in combination with increasing computational power, new methods are constantly being developed, that can trace more nuanced forms of selection [Field et al., 2016]. The majority of adaptations is assumed to originate from subtle shifts in frequencies, at quantitative trait loci [Schrider and Kern, 2017]. The majority of phenotypes is affected by variation at several loci, thereby extending the potential 'target' of selection. This also allows for a much faster adaptation on a population level, compared to phenotypes resulting from single loci. While the vast majority of selection tests is done using modern-day data, some studies have relied on ancient DNA. Here, a different approach is needed. While approaches relying on modern data try to detect signs resulting from past or ongoing selection, ancient DNA can provide a window back in

time. Wilde et al. [2014] used ancient samples from the Pontic-Caspian steppe region and analyzed markers related to pigmentation phenotypes, that were indicated to be under selection in European populations. Under the assumption of population continuity, they modeled the development of the allele frequencies for the markers over time. This allowed for a comparison between the estimated frequencies and frequencies obtained from a modern-day population of the area. By using binomial sampling, drift and population growth could be simulated over time. Potential uncertainties of the sampling were incorporated, as the initial frequencies for each SNP were estimated by drawing a random number from a beta-distribution, that was based on the allele counts. Based on the results, drift as the only explanation had to be rejected, thus reinforcing the evidence for selection at pigmentation related loci.

A different approach was used by Mathieson et al. [2015]. Based on previous publications [Lazaridis et al., 2014, Haak et al., 2015], they developed an explicit demographic model. Under the assumption, that Central European populations can be modeled as a mixture of three ancient source groups, namely Central European hunter-gatherers, early Neolithic farmers, and late Neolithic/early Bronze Age people from the Pontic-Caspian steppe region, they inferred the admixture proportions for four Central European populations. Frequencies for genome-wide SNPs were estimated, based on genotype-likelihoods, obtained from ancient samples originating from the three source groups. By relying on the estimated admixture proportions, expected frequencies for the modern populations were modeled as a linear combination of the frequencies in the three ancient groups. By comparing observed and expected frequencies, outliers could be determined. This led to the identification of a set of SNPs that presumably was under selection during the past 8,000 years. These SNPs were related to phenotypic traits associated with food processing, depigmentation, and immunity, among others.

For this study, a combination of different approaches was chosen. As the data set does not include samples from the Pontic-Caspian steppe region, an approach as described in Mathieson et al. [2015] could not be applied. However, the Bronze Age samples from the Tollense Valley provided the opportunity for a forward-simulation approach as described in Wilde et al. [2014]. Based on previous studies, continuity between populations from the Bronze Age to modern-day populations of Central Europe can be assumed to some extent. This allows to investigate potential selection over the last 3,000 years [Haak et al., 2015]. In addition, four neutrality test were applied to the reference data (iHS, nS_L, Garud's H, Tajima's D), to analyze if the SNPs of the capture array fall in regions, that deviate from neutral expectations. While iHS and nSL are well suited to identify variants that rose from a very low to a very high frequency, they can be unsuitable for detecting selection on standing variation [Hancock et al., 2010]. Therefore Tajima's D and Garud's H-statistics were also applied, as they are equipped to identify diversifying selection as well [Garud et al., 2015], while relying on a fixed window size for computation. The frequencies obtained from the ancient samples in combination with additional summary statistics then were used to determine the time during which the changes in frequency may have occurred.

2.2 Capture design

The capture array, described in Veeramah et al. [2018], was designed to allow demographic inferences and the investigation of phenotypic traits. The array consists of 5,000 regions of 1,000 base pairs that are located in functionally neutral regions of the genome (hence named: neutralome), in addition to over 480 regions associated with phenotypic markers and Y-chromosomal SNPs. The phenotypic markers can be grouped into three sets. A set of over 100 markers associated with pigmentation of skin, hair and eyes and some other visible phenotypes. A set of over 180 markers, associated with the energy metabolism and processing of foodstuff. The majority of these markers was chosen for their known association with the metabolic syndrome and type-2-diabetes. A third set consists of over 80 markers associated with the immune system and the detoxification of xenobiotics. Several of these markers are furthermore linked to autoimmune diseases, such as asthma, Chron's disease, Coeliac's disease or multiple sclerosis. In contrast to other capture arrays, this array was deliberately chosen to include lesser SNPs to ensure a higher sequencing depth at similar sequencing costs.

2.3 Reference populations

Genotype calls from the latest phase of the 1000 Genomes project [Consortium et al., 2015] were used as modern reference data. As a representation of a modern-day Central European population the Utah Residents with Northern and Western European Ancestry (CEU) were chosen. This group consists of 99 individuals. In order to be able to detect differences between modern-day populations, two additional data sets were added. As a population from Sub-Saharan Africa, the Yoruba from Ibadan in Nigeria (YRI - 101 individuals) were used and as a representative of an East Asian population, the Southern Han Chinese (CHS - 105 individuals) group was used.

2.4 Population genetic estimations

2.4.1 Allele frequencies

Allele frequencies (p) were estimated based on allele-counts (n_p, n_q), obtained from the genotypes called by the ATLAS software as follows, given the alleles p,q:

$$p = \frac{n_\mathrm{p}}{n_\mathrm{p} + n_\mathrm{q}} \tag{2}$$

A 95% confidence interval (CI) was calculated using:

$$CI = p \pm 1.96 * \sqrt{p * ((1-p)/n)} \tag{3}$$

If an allele reached fixation in a group, equation 3 would result in a CI of 0. In these cases the CI was defined by the 2.5^{th} quantile of the beta-distribution, determined by the counted alleles (beta(n_p+1, n_q+1)). Differences between allele frequencies were detected by F_{st} in combination with Fisher's exact test, based on the allele counts for each group. Differences were deemed significant, when the result of the Fisher test was $p \leq 0.05$ and the F_{st} was ≥ 0.

2.4.2 Mean difference of allele frequencies

To determine the mean difference in allele frequencies for a set of markers between groups, an iterative approach was used. For each SNP, a random number was drawn from a uniform distribution, defined by the SNPs confidence intervals in each group, before the average difference between groups was determined. This was repeated for 10,000 times and the mean and standard deviation were calculated based on the computed results of each group. Differences between groups were determined by absolute distance between the group means, as a function of the combined standard errors. The resulting Z-score was transferred into a p-value.

2.4.3 F_{st}

The fixation index F_{st} was estimated as a measure of genetic distance. The index was estimated following the argumentation in Bhatia et al. [2013] as this interpretation (equation 4) allows more accurate estimations even if sample size differs between populations. This is important since sample size varies across ancient sample groups and fluctuations in coverage reinforce these differences. Estimations were conducted using the a python scrip, relying on the *scikit-allel* library (https://github.com/cggh/scikit-allel/tree/v1.1.8).

$$F_{\mathrm{st}} = \frac{(\tilde{p}_1 - \tilde{p}_2)^2 - \frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1-1} - \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2-1}}{\tilde{p}_1(1-\tilde{p}_2) + \tilde{p}_2(1-\tilde{p}_1)} \tag{4}$$

F_{st} is calculated based on the estimated allele frequencies (p1, p2) and the numbers of samples (n1, n2) in each group.

2.4.4 Pairwise F_{st} comparisons

To investigate if differences between two groups at certain loci were larger than expected under a neutral model, a pairwise comparison of F_{st} values was used. First, the average F_{st} was calculated over all functional loci under investigations (f-F_{st}). For comparison, the same number of loci was drawn at random from the neutralome, and the average F_{st} was calculated in an iterative approach 10,000 times (n-F_{st}). Then the distance of the f-F_{st} to the mean of the n-F_{st} was calculated as a function of the standard error of the n-F_{st} distribution. The resulting Z-score was transferred into a p-value.

Since the F_{st} was calculated based on frequencies, groups were filtered for a minimum of eight alleles. Only pairings with the majority of alleles present were included in the calculations. The SNPs in the neutralome were filtered to only include bi-alleleic SNPs that were polymorphic in all reference populations. Insertions as well as deletions were excluded from all calculations.

2.4.5 Hardy-Weinberg equilibrium

The Hardy-Weinberg equilibrium was calculated based on the genotype data, to determine if a genotype was present in excess. Calculations were performed using the scikit-allel python library (https://github.com/cggh/scikit-allel/tree/v1.1.8).

2.4.6 Neutrality tests

All neutrality tests were calculated based on the genomes of the reference groups. To be able to calculate iHS and nSL-scores, the state (ancestral/derived) of each allele must be determined. To determine states, the human ancestor alignment from the 1000Genomes project was used [Consortium et al., 2015]. For each allele where information was available, the state was assigned using PLINK 1.9 [Purcell et al., 2007]. Since information was not available for all alleles, some regions of the genome were excluded from the analysis. The standardized integrated haplotype score (iHS) was calculated as described in [Voight et al., 2006], using a python script relying on the scikit-allel library (https://github.com/cggh/scikit-allel/tree/v1.1.8). The same script also calculated the standardized number of segregating sites per length (nSL) as described in [Ferrer-Admetlla et al., 2014]. Both iHS and nSL were calculated for each SNP that was polymorphic in the reference data and for which a state could be determined.

Tajima's D [Tajima, 1989] and Garud's H-statistics [Garud et al., 2015] were calculated on the standard 1000Genomes data, without determining ancestral and derived states. Both tests were calculated in a sliding window approach with a window size of 400 SNPs, similar to [Garud et al., 2015]. The windows were shifted in steps of 50 SNPs. For each SNP in the capture array the highest test result was determined for each population.

For all four tests, extreme high as well as extreme low outliers indicate deviations from neutrality. Values higher then 2 or lower then -2 for Tajima's D, iHS and nSL are known to violate neutrality under several demographic scenarios. Here, significance was determined per chromosome if values were in the highest or lowest 2.5 percent of distribution of results and above 2 or below -2 for Tajima's D, iHS and nSL. The scripts used for these calculations can be found on the supplementary DVD.

2.4.7 Averaged expected heterozygosity or gene diversity D_e

The averaged expected heterozygosity or gene diversity D_e for a given locus or gene, was calculated based on Nei [1973], using formula 5, were p_i is the allele frequency of the ith of n alleles, at the lth of m loci.

$$D_e = 1 - \frac{1}{m} * \sum_{l=1}^{m} \sum_{i=1}^{n} * {p_i}^2 \tag{5}$$

To prevent confusion with Tajima's D, the gene diversity is referred to as D_e hereafter. In order

to acknowledge possible uncertainties in frequencies resulting from sampling, D_e was computed in an iterative approach. A random value was drawn from a uniform distribution, defined by the 95% confidence interval for each SNP in each group, for the locus under investigation. The mean and standard deviation were calculated for a set of 10,000 calculations for each group. This also allowed for the calculation of a Z-score, based on the distance between two means, as a measure of their combined standard deviations. Subsequently, the Z-score was translated into a two-sided p-value.

2.5 Phenotypic predictions

2.5.1 Predictions of N-acetyltransferase 2 phenotpyes

Acetylation phenotypes for the N-acetyltransferase 2 gene were predicted using the NAT2PRED webtool (http://nat2pred.rit.albany.edu/) described in Kuznetsov et al. [2009]. Genotype information for each sample was obtained from the variant calls produced by the ATLAS software and transformed into the input format, needed for submission to the webtool.

2.5.2 Predictions of hair and eye color with the HirisPlex webtool

Hair and eye color phenotypes for the ancient samples were predicted using the HirisPlex webtool (https://hirisplex.erasmusmc.nl/), described by [Walsh et al., 2014]. Genotype information for each sample was obtained from the variant calls produced by the ATLAS software and transformed into the input format, needed for submission to the webtool.

2.6 Tracking changes in allele frequencies over multiple loci

To take the varying coverage of the ancient samples into account, score-calculations were not preformed based on the genotype data. Instead, a single allele, determined as the most common allele at the locus in the alignment, was used. To allow comparison, the reference data was treated similar. For the reference data, one allele was sampled for each locus for each individual from the VCF. If the individual was homozygous at a locus, one allele was chosen at random.

2.6.1 Genetic risk scores

To compare individual type 2 diabetes risks between sample groups, three different genetic rsik scores were calculated according to Meigs et al. [2008], Cornelis et al. [2009], and Hivert et al. [2011]. The unweighted genetic risk score sums up the number of risk alleles present at each locus in each individuals. To account for missing data, the generalized risk score (GRS) was calculated, following Cornelis et al. [2009]. Here, the sum of risk alleles is divided by the number of loci for which data was present, and is then multiplied by the total number of SNPs for each score (see equation 6, where m is the number of covered SNPs). Only samples where more than half of the SNPs were covered were considered.

$$GRS = n_{\text{total}} * (\frac{1}{m}\sum_{i=1}^{m} a_{\text{risk}}) \tag{6}$$

2.6.2 Pigmentation score

Similar to the GRS, the pigmentation score sums up the number of derived alleles for a set of SNPs. A total of 42 loci was chosen, based on their high difference in derived allele frequencies in the CEU and YRI reference populations (see table 2). The score is calculated as described in formula 7, where n is the total number of sites, while m denotes the sites that are covered by data in the sample, and a_{der} is the number of derived alleles.

$$Pig_{\text{sc}} = n * (\frac{1}{m}\sum_{i=1}^{m} a_{\text{der}}) \tag{7}$$

Only scores for samples with more than half of the SNPs covered (at least 25) were taken into account.

Table (2) *SNPs used to calculate Pigmentation score with chromosome, position, ID, gene and derived allele.*

Chr	Position	rs ID	Gene	derived allele
5	33948589	rs35395	SLC45A2	C
5	33951693	rs16891982	SLC45A2	G
5	33958959	rs28777	SLC45A2	A
9	12675264	rs10960751	TYRP1	C
9	12675284	rs10960752	TYRP1	A
9	12703484	rs2733831	TYRP1	G
9	12704725	rs2733832	TYRP1	T
11	88557991	rs10831496	GRM5	A
12	89299746	rs642742	KITLG	C
14	92773663	rs12896399	SLC24A4	T
15	28131633	rs2703969	OCA2	A
15	28226541	rs12914687	OCA2	T
15	28235773	rs1800404	OCA2	T
15	28288748	rs7170869	OCA2	A
15	28335820	rs4778138	OCA2	A
15	28356859	rs1129038	HERC2	T
15	28359258	rs12593929	HERC2	A
15	28364059	rs7494942	HERC2	G
15	28365618	rs12913832	HERC2	G
15	28374012	rs3935591	HERC2	C
15	28412447	rs8043281	HERC2	G
15	28422026	rs6497284	HERC2	T

Table 2 continued from previous page

Chr	Position	rs ID	Gene	derived allele
15	28427986	rs7170852	HERC2	A
15	28468723	rs3940272	HERC2	G
15	28494202	rs2240203	HERC2	T
15	28513364	rs916977	HERC2	C
15	28516084	rs8039195	HERC2	T
15	28526228	rs16950987	HERC2	G
15	28530182	rs1667394	HERC2	T
15	29261716	rs4424881	APBA2	C
15	48392165	rs1834640	SLC24A5	A
15	48400199	rs2675345	SLC24A5	A
15	48414969	rs2433354	SLC24A5	C
15	48418645	rs2675347	SLC24A5	A
15	48419386	rs2555364	SLC24A5	G
15	48420744	rs2675348	SLC24A5	A
15	48426484	rs1426654	SLC24A5	A
15	48433494	rs2470102	SLC24A5	A
15	48445387	rs2675349	SLC24A5	A
15	48461146	rs3817315	SLC24A5	C
20	32785212	rs6119471	ASIP	C
20	32856998	rs6058017	ASIP	A

2.7 Beta-binomial forward simulation

The beta-binomial simulation approach as described in Wilde et al. [2014] tests if changes in allele frequencies over time can be the result of drift, or if selection must be invoked. Under the assumption of population continuity and exponential growth, frequencies are modeled over time to be compared to observations from a modern-day population. Frequencies are modeled by binomial sampling from the previous generation. The initial frequency at the start of the simulation is drawn at random from a beta-distribution that is defined by the allele counts in the ancient population, plus one, to account for possible sampling uncertainties (beta(n_p+1, n_q+1)).

Here, the Bronze Age samples from the Tollense valley were used as proxy of an population ancestral to the Central Europeans from Utah (CEU) from the 1000Genomes project. A generation time of 25 was assumed and simulated over a time of 3150 years. All allele counts were taken from the variant calls produced with the ATLAS software [Link et al., 2017]. The frequencies for CEU were taken directly from the latest VCF-files.

Since the initial population size of the Bronze Age population, from which the Tollense samples originate, is unknown, 50 even-spaced loq_{10} values were used. For the final population size, 10% of

the census population of Germany were assumed (rounded to 8,000,000 Individuals). To estimate the effect of drift and selection, 50 even-spaced values between 0.1 and -0.1, plus 0, were used as selection coefficients (S). For each combination of initial population size and selection coefficient, 10,000 simulations were run, with the final allele frequency being stored for each locus. To analyze if the frequency estimated for the Central European reference group could originate from the distribution of frequencies estimated during simulations, a Z-score was calculated. This score described the absolute distance between the mean of the simulated frequencies and the frequency estimated for the reference group as a measure of the standard deviation of the simulated frequencies (equation 9). This Z-score was then translated into a two-tailed p-value (see. equation 10).

$$P(N) = \binom{n}{N} * p^N (1-p)^{n-N} \tag{8}$$

$$Z = \frac{|\overline{X} - f|}{\delta} \tag{9}$$

$$p = 2 * cdf(Normal, Z) \tag{10}$$

The simulation model was tested with allele counts from the neutralome portion of the capture array. Under the assumption of neutrality, variants for which drift was rejected during simulations can be seen as false positive results. In combination with the results of the functional markers, a false discovery rate (FDR) could be calculated by dividing the percentage of SNPs where drift was rejected for the functional markers, by the same percentage of SNPs for the neutral markers. P-value thresholds of p=0.05, p=0.01 and p=0.001 to determine the significance of the differences in frequencies were tested (see figure: 5). Since FDR was lowest at a p-value threshold of p≤0.001, this value was chosen to determine significance. Since the number of SNPs indicated as selected during simulations did not differ above an initial population size of 625, the largest number with the lowest FDR was chosen. Consequently, all selection coefficients were estimated for an initial population size of 1456 individuals.

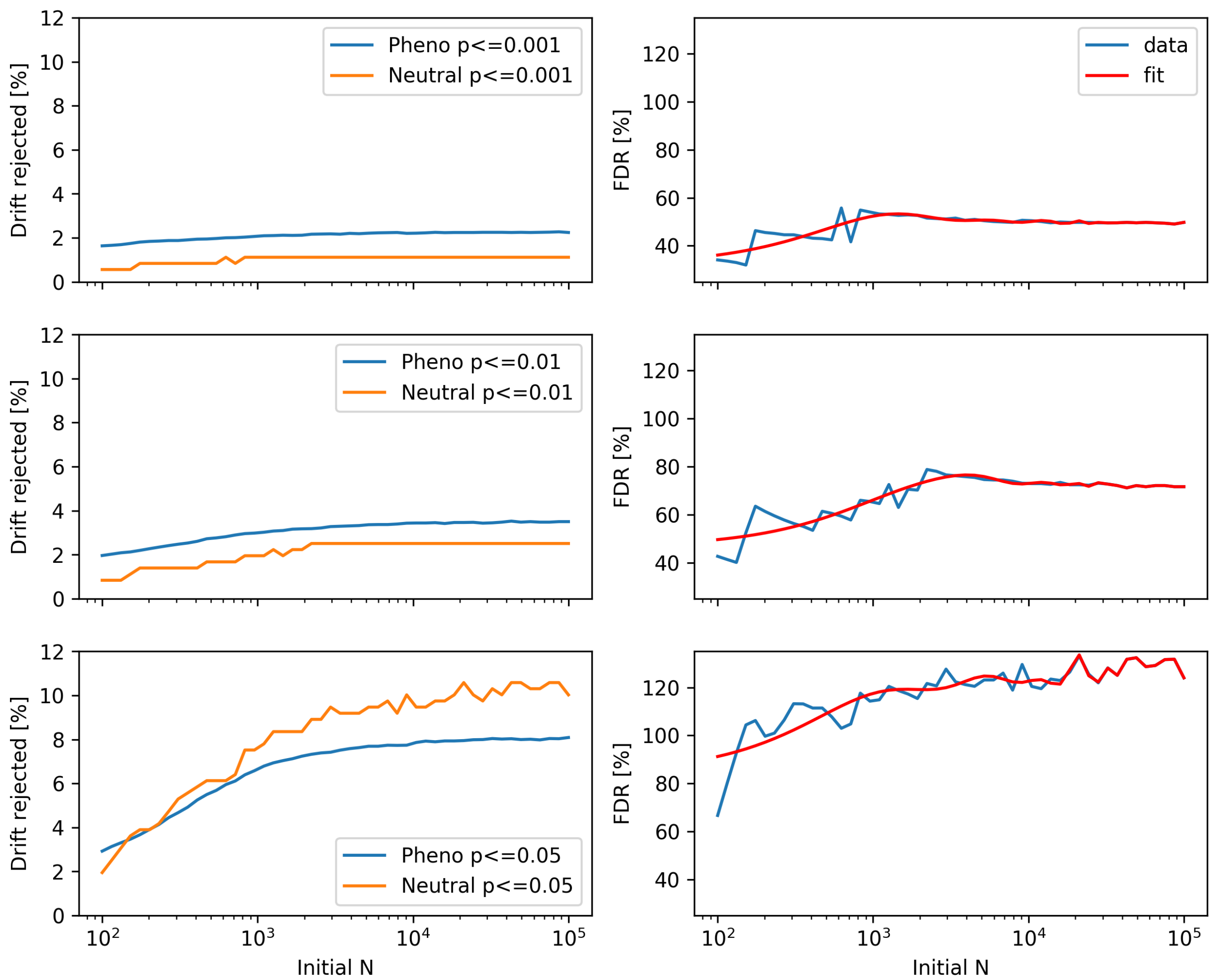

Figure (5) *Percentage of SNPs where drift was rejected during beta-binomial simulations for each initial population size for neutral and functional markers. Results are shown for three different p-value thresholds, incl. false discovery rates (FDR).*

3 The history of the Central European pigmentation phenotypes

3.1 Introduction

This chapter investigates the development of the visible phenotype of European populations. It is assumed that a lightening in pigmentation has occurred early after humans migrated out of Africa, in response to the lower UV radiation at higher latitudes and was further intensified by changes in diet during the Neolithic [Yang et al., 2018].
Pigmentation in humans is a complex phenotype that shows great variation among populations. Dark pigmentation is thought to be the ancestral phenotype and under functional constraint in Africa. It probably arose when early hominins became hairless in favor for a better thermo-regulation [Jablonski and Chaplin, 2000]. Without a protective hair-coat the skin became exposed to UV radiation. A dark complexion is thought to be protective, not only against sunburn but also against DNA-damage and photo-degradation of micro-nutrients such as flavins, carotenoids, tocopherol and folate. The photolysis of folate can have a direct influence on reproduction by influencing male fertility and causing embryonic defects such as *spina bifida*, therefore maintaining a dark skin was likely beneficial [Jablonski, 2012].

Most skin and hair pigmentation in mammals is caused by melanin. It occurs in two forms: the dark black/brown eumelanin and the reddish/yellow pheomelanin. [Sturm, 2006]. Melanin is synthesized in melanosomes inside the melanocytes in the skin, at the boundary between the dermis and the epidermis, as well as in hair follicles and the iris. In the skin the melanosomes containing the melanin are transported and stored in neighboring keratinocytes. Variation in pigmentation is caused by the differential distribution and size of the keratinocytes and the amount of melanin stored in them, while the number of melanin producing melanocytes is roughly the same between individuals regardless of the degree of pigmentation [Lin and Fisher, 2007].

Genetics of pigmentation

Pigmentation is a polygenic trait that is regulated by a number genes [Branicki et al., 2009]. Effects of variations in the genes involved are often assumed to be additive. Almost all genes involved in creating the pigmentation phenotype are polymorphic. Variation can therefore affect the process during each stage with varying severity. Most of the alleles associated with depigmentation are derived, since the ancestral state is assumed to be a dark-skinned phenotype [Crawford et al., 2017]

The melanocortin-1 receptor gene (MC1R) encodes a trans-membrane receptor that plays an important role in the activation of many transcription factors. That transcription factor in turn activate the transcription of various genes involved in the synthesis and distribution of melanin. A variety of variants is known in the MC1R locus that impair the function of the receptor [Ibarrola-Villava et al., 2010].

Antagonists of MC1R regulate the gene's activity. While the α-melanocyte-stimulating hormone leads to an increased production of eumelanin [D'orazio et al., 2006, Barsh et al., 2000] the agouti

signaling gene (ASIP) encodes an antagonist that leads to an increased synthesis of pheomelanin [Furumura et al., 1996]. Mutations in the ASIP gene can lead to its inactivation resulting in dark hair and brown eyes [Kanetsky et al., 2002]. Polymorphisms in MC1R are also associated with hair color variation. Several variants in the coding region are associated with red hair, fair skin, freckles and an impaired tanning response [D'orazio et al., 2006].
Transcription factors activated by MC1R lead to the expression of the TYR gene. This gene encodes tyrosinase, a protein that is involved in the hydroxylation of tyrosine to dopaquinone, which is a precursor to melanin [Land and Riley, 2000]. The TYR gene is also highly polymorphic and variants can have a direct effect on the amount of melanin produced. The tyrosinase related proteins 1 and 2 (TYRP1 and TYRP2 or dopachrome tautomerase DCT) are assumed to help stabilize the tyrosinase, while also regulating melanin synthesis in favor of the darker eumelanin [Land and Riley, 2000].

Tyrosinase activity is affected by the OCA2 gene, named after the abnormal pigmentation phenotype - oculocutaneous albinism type II. It encodes for a P-protein that is among other things involved in the transport of the tyrosinase as well as regulating the melanosomal pH. Melanogenesis is a pH dependent process in which a shift in the pH-value can affect the amount and form of melanin produced [Ibarrola-Villava et al., 2010]. The OCA2 gene is assumed to be regulated by HERC2, a gene located in close proximity. Polymorphisms in both genes have been shown to be tightly linked to eye and hair color phenotypes [Donnelly et al., 2012]. A haplotype, including SNPs from both genes and associated with blue eyes, occurs at high frequency in Europe, but is almost absent in other parts of the world [Duffy et al., 2007, Eiberg et al., 2008, Donnelly et al., 2012]. The OCA2 gene is highly polymorphic in populations outside of Africa, with different variants known to affect pigmentation [Candille et al., 2012].

The acquisition of melanogenic enzymes as substrate for melanogenesis is regulated by membrane-associated transporters. Two of those are members of the solute-carrier family, SLC24A5 and SLC45A2. Variants within those two genes (rs1426654, rs16891982) have the strongest association with pigmentation variation between African and European populations with the derived alleles being almost fixed in Europe, but at low frequencies in other parts of the world [Ainger et al., 2017, Deng and Xu, 2018]. It is assumed that the effect on pigmentation is caused by a down-regulation of the available substrate.

Additional genes as RAB27A, KITLG and MYO5A play a role in the transport and binding of the maturated melanosomes in the keratinocytes. Variations in these genes can lead to a disrupted melanin transport and impaired binding [McEvoy et al., 2006, Ainger et al., 2017]. Several other genes are associated by regulating the expression of other pigmentation related genes or through still unknown associations [Alonso et al., 2008].

The vitamin D hypothesis

The variation in pigmentation is strongly correlated with UV radiation and thereby with latitude

[Chaplin, 2004]. Dark skinned phenotypes are more common near the equator, while pigmentation levels decrease in regions further north. According to a widely accepted theory, vitamin D deficiency due to lower UV radiation at higher latitudes caused a selection pressure that has led to the development of a depigmented phenotype [Jablonski and Chaplin, 2000, Yuen and Jablonski, 2010]. Vitamin D is synthesized in the skin at spectral range from 255-330nm, with an optimum at 295nm (UVB). During synthesis a precursor is converted into previtamin D_3, which in turn is further converted into vitamin D. After exposure to sunlight the serum vitamin D levels peak with a 24-48h delay. As a lipid-soluble molecule, vitamin D can accumulate in adipocytes, where it can last for several month [Mostafa and Hegazy, 2015].

While the high UVB radiation necessitates a protective function of the skin to prevent DNA damage and folate photolysis near the equator, the radiation that reaches the dermal layers is still high enough to ensure vitamin D synthesis in these regions [Chaplin, 2004]. At higher latitudes, however, UVB radiation is reduced due to the inclination of the earth's axis to the sun, increasing exposure times necessary to maintain healthy vitamin D levels. Only during a period between March and October the UV radiation is high enough for a sufficient vitamin D synthesis in Boston [Holick et al., 1987]. Comparative studies between women of African and European descent in the US have shown that dark skin is associated with a lower vitamin D level in the blood when exposed to the same dose of UVB radiation [O'Connor et al., 2013].

Vitamin D plays an important role in the human organism by influencing the calcium level in the blood, as it is an important cofactor in the intestinal calcium absorption. Consequently, a vitamin D deficiency can have severe impact on bone health. Insufficient amounts of vitamin D can cause rickets in children as well as osteomalcia in adults, resulting in an insufficient calcium dosage leading to softening and subsequent deformation of the bones [Holick, 2004]. Vitamin D deficiency is in addition linked to heart disease, with low levels of vitamin D increasing the risk of myocardial infarction by a factor of 2.4 [Wagner et al., 2008, Holick, 2007, Bikle, 2008, Aranow, 2011, Hewison, 2012]. Furthermore, there are indications that vitamin D can promote an adequate response to infection [Holick, 2004]. Even though vitamin D synthesis as a driving force for the depigmentation in humans outside of Africa is a widely accepted theory, it is still part of an ongoing debate [Nessvi et al., 2011, Ahn et al., 2010, Wang et al., 2010, Robins, 2009, Yuen and Jablonski, 2010].

Selection on pigmentation phenotypes

It is assumed that the pigmentation phenotype is affected by selection on a worldwide scale. For regions near the equator, a functional constraint is assumed that is affected by purifying selection, thereby preventing light skin associated variants from segregating. Outside of Africa the pigmentation phenotype is assumed to have been affected by at least three major selective sweeps [Deng and Xu, 2018, Sturm and Duffy, 2012, Beleza et al., 2012, Pickrell et al., 2009, Coop et al., 2009, Alonso et al., 2008].

Comparative studies have identified variants in genes that occur at elevated frequencies in both East Asian and European populations. SNPs in the KITLG gene are believed to be affected by a selective sweep in all non-African populations. This is likely to have happened before the initial split of the ancestral populations moving further apart after their initial migration out-of-Africa [Lao et al., 2007, Coop et al., 2009, Pickrell et al., 2009]. Other genes also affected by such a theorized early sweep are ASIP and BNC2. In both genes variants have increased in frequency in East Asian and European populations while being present at low frequencies on the African continent [McEvoy et al., 2006].

In European populations, SLC24A5 and SLC45A2 are referred to as the golden genes associated with skin depigmentation [Deng and Xu, 2018]. The derived allele at the SNP rs1426654 in SLC24A5 is almost fixed in European populations but nearly absent in populations outside of Europe [Soejima and Koda, 2007]. Furthermore, haplotype diversity surrounding SLC24A5 is low in populations of European ancestry [Canfield et al., 2013, Giardina et al., 2008]. The diversity of SNPs in SLC45A2 shows a similar pattern in Europe. Several SNPs can be found at high frequencies almost exclusively in European populations [Nakayama et al., 2002]. This pattern is assumed to have been shaped by selection since the last glacial maximum, approximately 19,000 years ago [Beleza et al., 2012]. Other loci with a Europe-specific variation include the genes OCA2/HERC2, TYR, TYRP1, and MC1R. Some of these variants contribute also to the variation in hair and eye color that is highest among European populations [Branicki et al., 2009, Donnelly et al., 2012, Duffy et al., 2007, Alonso et al., 2008].

In East Asian populations different variants in the same genes were indicated as potential targets of selection. Variants associated with a light skinned phenotype in TYR, MC1R, TYRP1 and OCA2 have increased in frequency in East Asian populations, leading to population specific distribution [Beleza et al., 2012, Norton et al., 2006]. This can be interpreted as a sign for the convergent evolution of a light skinned phenotype in East Asia and Europe.

Research suggests that selective pressure has affected variants in different genes at different times since the last glacial maximum in Europe. Age estimations for variants in SLC24A5 and SLC45A2 vary, but 20,000 years or less seems most likely [Beleza et al., 2012, Soejima and Koda, 2007, Norton et al., 2006]. Timing estimates for the selective sweeps suggest that variants in both genes started to be affected by positive selection between 19,000 and 11,000 years ago [Beleza et al., 2012]. These estimation could suggests that selection on these variants was at least partially influenced by the onset of the Neolithic.

Ancient DNA and the pigmentation phenotype

The gradual lightening of pigmentation in Europe could be directly studied in the last decade for the first time. The reason for this is the advance in sequencing technologies that are giving rise to a vast amount of new data [Marciniak and Perry, 2017]. First estimates for the timing of selective

sweeps could be validated with data from the respective periods.

Consistent with the theory of an early onset of depigmentation in humans, the first people that arrived in Central Europe, approximately 40,000 years ago, were already carrying derived alleles for some of the pigmentation related SNPs. Nevertheless, they are assumed to have still maintained a relatively dark skinned phenotype, compared to modern-day populations, as they were still carrying ancestral variants for the major causative alleles in SLC24A5 and SLC45A2 [Fu et al., 2016]. First evidence of derived alleles for SLC45A2 were found in ancient genomes from Scandinavia dating to 9,000 BP while being absent or present at low frequencies in the rest of central Europe until the Late Neolithic [Gamba et al., 2014, Günther et al., 2015, Mathieson et al., 2015, Allentoft et al., 2015, González-Fortes et al., 2017]. Derived alleles in SLC24A5 were also nearly absent in Mesolithic times in central Europe, but arrived at high frequencies with the farmers from the Aegean region, 8,000 years ago [Olalde et al., 2014, Lazaridis et al., 2014, Hofmanová et al., 2016].

A study addressing selection in European populations based on ancient DNA data found evidence that variants in SLC45A2 were under selection in Europe during the Neolithic [Mathieson et al., 2015]. In contrast, they could not find evidence for a selection affecting SLC24A5, as predicted by studies using modern data [Beleza et al., 2012, Sturm and Duffy, 2012]. Therefore they concluded that selection may have affected SLC24A5 variation in early farmer populations already during their migration towards Central Europe, introducing variants already at a high frequency [Mathieson et al., 2015]. Strong signals of selection on SLC45A2, TYR and a variant in HERC2 were also found in ancient samples from the Ukraine [Wilde et al., 2014].

Frequencies of variants in the OCA2/HERC2 gene complex, associated with a blue eye color and blond hair, were found to be high in hunter-gatherers of Central Europe [Brace et al., 2018, Olalde et al., 2014, Mathieson et al., 2015]. This is consistent with estimations from recent data that placed the origin of this phenotype to Europe [Duffy et al., 2007]. Phenotypic reconstructions based on ancient DNA suggested that some of the hunter-gatherers of Central Europe had a unique appearance as a result of a dark skin tone in combination with light eyes and a light hair color [Brace et al., 2018, Olalde et al., 2014]. In contrast, hunter-gatherers from Russia were assumed to have had a lighter skin tone with almost exclusively dark hair and dark eye color [Mathieson et al., 2015]. In comparison, the pigmentation phenotype of the early farmers of Europe was assumed to be relatively light, but still darker compared to today's populations of Central Europe [Mathieson et al., 2015, Hofmanová et al., 2016].

This is especially interesting since the first farmers of Europe were descendants of populations that originated in lower latitudes compared to hunter-gatherers of Europe from the same period [Hofmanová et al., 2016, Lazaridis et al., 2016]. This difference may be explained by the differences in subsistence and diet between the groups. While hunter-gatherers probably relied on a wide variety of foodstuff, depending on season and habitat, the early farmers were producing most of their food themselves by relying on domesticated plants and animals, supplemented only occasionally with

additional hunting or fishing. Consequently, the hunter-gatherer diet was estimated to be richer in supplementary vitamin D, originating from sea- and freshwater fish as well as from game meat. This is further emphasized by the fact that vitamin D levels are higher in wild game compared to domesticated animals [Chen et al., 2007, Lu et al., 2007].

As a result the need to synthesize vitamin D may have been comparatively low in hunter-gatherers. The shift to a new diet, low in vitamin D, during the onset of farming facilitated the need for an optimization of the vitamin D metabolism [Fischer et al., 2007, Richards and Hedges, 1999, Richards et al., 2003]. This argument is reinforced by findings that show the consequences of malnutrition in early farming groups [May and Ruff, 2016, Armelagos et al., 1991]. As indicated by skeletal remains, the body size and bone health was deteriorating in some early farming communities, compared to contemporary hunter-gatherers [Garn and Leonard, 1989, Cochran and Harpending, 2009].

In contrast to skin color, the development of light hair- and eye color cannot be explained by the vitamin D hypothesis alone. One possible supplementary explanation is sexual selection [Frost, 2006], but a linkage to other phenotypes is also possible [Donnelly et al., 2012]. Moreover, some variants highly associated with light eye color or blond hair may have additional associations with traits such as skin color, which pose as additional selection criteria.

After the onset of the Neolithic in Europe, the gradual depigmentation further increased subsequently creating the now-present phenotype with its still visible latitudinal cline. Although the depigmentation was already well advanced at the end of the Bronze Age, evidence suggests that selection may have still been ongoing over the past 2,000 years [Field et al., 2016].

The data set presented in this work presents the opportunity to shed further light on the development of the pigmentation phenotypes of Central Europe. By applying neutrality tests to the genome-wide reference data signs under potential selection could be identified. The diachronic character of the ancient DNA data allowed to track changes in allele frequencies over time between groups that also varied in subsistence. This made it possible to correlate potential selection to specific period in time and form of subsistence.

3.2 Hypotheses

The following hypothesis will be tested in this chapter:

1. The light skinned phenotype developed outside of Africa, following a stepwise model. The first selective sweep happened before the split of the ancestral populations of Central Europe and East Asia.

   Derived alleles are found in all non-African groups in the data set. Frequencies for the affected alleles are similar in all groups, except for the African reference group.

2. Early Neolithic farmers were affected stronger by the decreased UV radiation outside of Africa, than Mesolithic hunter-gatherers. Thus, depigmentation was already advanced in early farming populations arriving in Europe, compared to the hunter-gatherers, leading to the development of a differing phenotype.

   Derived alleles are present at higher frequencies int the early farmer groups of the data set in comparison to the hunter-gatherers. Signs of non-neutrality can be expected for some of the alleles were differences can be found between groups.

3. Blonde-haired and blue-eyed phenotypes originated in Europe before the onset of the Neolithic and rose in frequency over time.

   Derived alleles in related genes such as TYRP1, HERC2 and OCA2 can be found in the Central European hunter-gatherers at elevated frequencies, in contrast to the Easter European/Russian hunter-gatherers. In addition, derived alleles are absent or at low frequencies in the early farmers from the Aegean/Balkan region and only slightly increased in the early farmers of Central Europe.

3.3 Results

3.3.1 Allele frequencies

The data set presented here consists of 103 SNPs in 13 genes, for which the frequencies of the derived alleles were estimated. Significant differences were found between the distributions of the derived alleles between the three modern-day reference groups (CEU - CHS: stat: 3666.0, $p \leq 0.0001$; CEU - YRI: stat: 2979.0, $p \leq 0.0001$; CHS - YRI: stat: 4362.0, $p \leq 0.0500$, see figure: 6). While in the African reference group the number of alleles that were fixed at 0 was highest, the number of derived alleles fixed at 1 was higher in the Central European reference group. In the Chinese reference group the number of derived alleles that reached fixation was lower compared to the Central European reference group, while it was higher compared to the African reference (see figure: 7). For all three reference groups, significant differences between the distribution of the functional allele frequencies could be found, when compared to the distribution of the same number of neutral allele frequencies (Mann-Whitney U $p \leq 0.0050$).

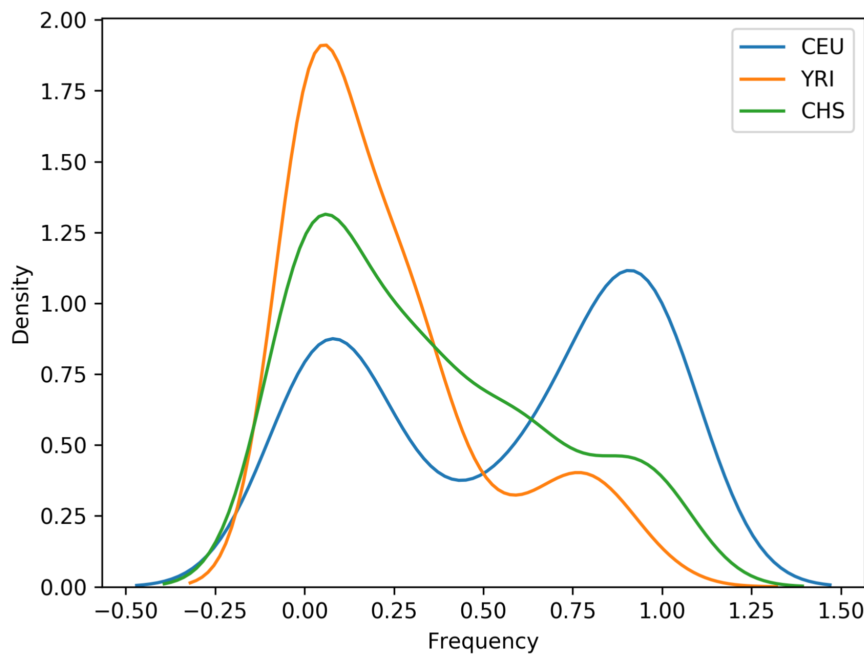

Figure (6) *Kernel-density-estimation plot for derived allele frequencies in the three reference groups. CEU: Central European reference, YRI: Afircan reference group, CHS: Chinese reference group.*

Several significant differences in allele frequencies were found between the ancient sample groups, when compared to the Central European reference group. The highest number of differences was found between the early farmers of Central Europe (CEF) in comparison to CEU (35 SNPs, 91.43% lower, 8.57% higher), followed by the Central European hunter-gatherer group (33 SNPs, 90.91% lower, 9.09% higher). Between the hunter-gatherers of Eastern Europe/Russia, 25 SNPs (84.00% lower, 16.00% higher) were at significantly different frequencies compared to CEU, while the Bronze

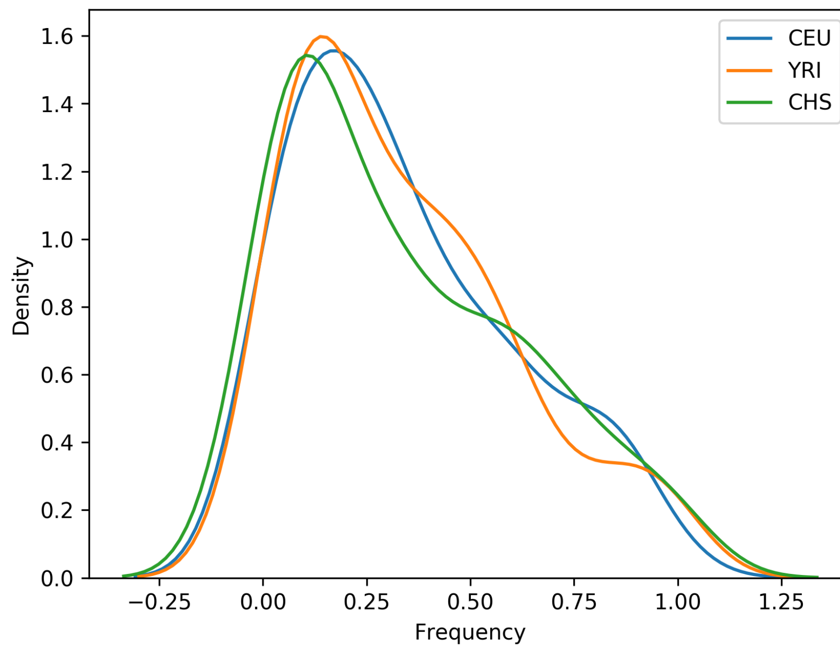

Figure (7) *Kernel-density-estimation plot for derived allele frequencies of randomly drawn neutral alleles in the three reference groups. CEU: Central European reference, YRI: Afircan reference group, CHS: Chinese reference group.*

Age group had different frequencies for 5 SNPs (80.00% lower, 20.00% higher), all located in the SLC45A2 gene (see table: 3).

In each ancient sample group except the Middle Neolithic farmers, one SNP was found to deviate from Hardy-Weinberg-Equilibrium. In the early farmer group from the Aegean/Balkan region, the SNP rs139335 in the gene TYR displayed an excess of homozygous genotypes. In all other groups an excess of heterozygous genotypes was detected: In the Eastern European/Russia hunter-gatherers for the SNP rs26722 in the gene SLC45A2; In the Central European hunter-gatherers, the Central European farmers and the Bronze Age group for the SNP rs4911414 in the gene ASIP (see table: 4).

Table (3) *Frequencies of derived alleles of pigmentation related SNPs for all groups. Colors in frequency fields indicate significant differences between the group and the CEU reference (red=higher, blue=lower). Colors in the ID column indicate SNPs that were found to deviate from neutral expectations in at least on of the four tests, while the color in the Gene column indicates SNPs identified during the beta-binomial simulations.*

ID	Gene	Allele	EHG	WHG	SEF	CEF	MN	WEZ	CEU	YRI	CHS
rs1545397	**OCA2**	**T**	0.00±0.31	0.00±0.12	0.10±0.13	0.04±0.08	0.00±0.46	0.00±0.15	0.08±0.04	0.00±0.02	0.88±0.04
rs1800401	**OCA2**	**A**	0.00±0.26	0.00±0.26		0.10±0.09	0.00±0.37	0.25±0.30	0.05±0.03	0.20±0.05	0.00±0.02
rs1800407	**OCA2**	**T**	0.00±0.26	0.00±0.26		0.00±0.23	0.17±0.30	0.00±0.60	0.08±0.04	0.00±0.02	0.00±0.02
rs1800414	**OCA2**	**C**	0.00±0.31	0.00±0.15	0.00±0.84	0.00±0.08	0.00±0.46	0.00±0.26	0.00±0.02	0.00±0.02	0.60±0.07
rs4778138	**OCA2**	**A**	0.60±0.30	0.65±0.18	0.50±0.19	0.64±0.14	0.50±0.49	0.83±0.15	0.87±0.05	0.24±0.06	0.32±0.06
rs7495174	**OCA2**	**A**	0.75±0.30	0.50±0.69	0.50±0.69	0.79±0.21	1.00±0.84		0.94±0.03	0.83±0.05	0.40±0.07
rs11074314	**OCA2**	**G**	0.25±0.30	0.25±0.17	0.14±0.14	0.13±0.23	0.50±0.49	0.13±0.13	0.17±0.05	0.54±0.07	0.00±0.02
rs12914687	**OCA2**	**T**	0.70±0.28	0.75±0.16	0.63±0.24	0.77±0.12	0.50±0.40	0.75±0.17	0.71±0.06	0.10±0.04	0.17±0.05
rs1800404	**OCA2**	**T**	0.88±0.23	0.80±0.25		0.81±0.19	0.67±0.38	0.67±0.38	0.82±0.05	0.07±0.04	0.41±0.07
rs2015343	**OCA2**	**A**	0.20±0.25	0.14±0.18		0.19±0.19	0.17±0.30	0.33±0.38	0.18±0.05	0.85±0.05	0.60±0.07
rs2703969	**OCA2**	**A**	0.75±0.30	0.86±0.14	0.73±0.19	0.60±0.30	0.67±0.38	0.83±0.17	0.71±0.06	0.18±0.05	0.41±0.07
rs4778136	**OCA2**	**T**	0.00±0.37	0.00±0.19	0.20±0.25	0.08±0.08	1.00±0.60	0.50±0.69	0.07±0.04	0.67±0.06	0.39±0.07
rs7170869	**OCA2**	**A**	0.30±0.28	0.83±0.13	0.64±0.25	0.60±0.14	1.00±0.46	0.93±0.13	0.88±0.04	0.02±0.02	0.13±0.05
rs746861	**OCA2**	**C**	0.80±0.25	0.46±0.19	0.44±0.24	0.80±0.11	0.25±0.42	0.45±0.22	0.50±0.07	0.34±0.07	0.25±0.06
rs895828	**OCA2**	**G**	1.00±0.37	0.83±0.15	0.50±0.69	0.95±0.06	0.67±0.38	0.86±0.18	0.88±0.04	0.41±0.07	0.40±0.07
rs4778241	**OCA2**	**C**	0.50±0.31	0.80±0.18	0.50±0.49	0.50±0.16	0.75±0.42	1.00±0.31	0.84±0.05	0.39±0.07	0.29±0.06
rs1667394	**HERC2**	**T**	0.33±0.27	0.59±0.17	0.70±0.20	0.63±0.13	0.50±0.40	0.78±0.19	0.83±0.05	0.05±0.03	0.34±0.06
rs7494942	**HERC2**	**G**	0.40±0.30	0.90±0.19		0.64±0.18	0.50±0.40	0.83±0.30	0.83±0.05	0.06±0.03	0.33±0.06
rs12593929	**HERC2**	**A**	0.67±0.38	0.67±0.38		0.90±0.19	1.00±0.60		0.92±0.04	0.31±0.06	0.61±0.07
rs3935591	**HERC2**	**C**	0.50±0.69	0.75±0.42		0.88±0.23			0.86±0.05	0.14±0.05	0.34±0.06
rs7170852	**HERC2**	**A**	0.40±0.30	0.79±0.16	0.50±0.35	0.64±0.13	0.75±0.30	0.88±0.16	0.85±0.05	0.13±0.05	0.35±0.06
rs2238289	**HERC2**	**G**	0.33±0.38	0.13±0.23		0.25±0.21	0.17±0.30	0.00±0.46	0.13±0.05	0.78±0.06	0.65±0.06

Table 3 continued from previous page

ID	Gene	Allele	EHG	WHG	SEF	CEF	MN	WEZ	CEU	YRI	CHS
rs3940272	**HERC2**	**G**	0.25±0.42	0.75±0.42		0.63±0.34	0.50±0.69	±	0.86±0.05	0.13±0.05	0.35±0.06
rs8028689	**HERC2**	**T**	0.70±0.28	0.70±0.16	0.83±0.15	0.82±0.10	1.00±0.46	0.95±0.10	0.93±0.04	0.57±0.07	0.63±0.07
rs2240203	**HERC2**	**T**	0.75±0.30	0.62±0.19	0.75±0.21	0.82±0.11	1.00±0.60	1.00±0.17	0.93±0.04	0.28±0.06	0.64±0.06
rs916977	**HERC2**	**C**	0.40±0.30	0.69±0.16	0.67±0.19	0.68±0.12	0.50±0.35	0.80±0.14	0.83±0.05	0.05±0.03	0.34±0.06
rs12913832	**HERC2**	**G**	0.00±0.31	0.50±0.19	0.30±0.20	0.34±0.12	0.50±0.35	0.67±0.22	0.77±0.06	0.00±0.02	0.01±0.01
rs1129038	**HERC2**	**T**	0.00±0.60	0.50±0.69		0.25±0.30	0.50±0.40		0.77±0.06	0.00±0.02	0.00±0.01
rs7183877	**HERC2**	**A**	0.30±0.28	0.08±0.10	0.00±0.46	0.13±0.11	0.17±0.30	0.13±0.16	0.06±0.03	0.11±0.04	0.27±0.06
rs11635884	**HERC2**	**C**	0.00±0.31	0.00±0.13	0.08±0.11	0.06±0.07	0.00±0.60	0.00±0.26	0.02±0.02	0.69±0.06	0.01±0.01
rs11636232	**HERC2**	**T**	0.00±0.37	0.50±0.19	0.00±0.31	0.02±0.04	0.50±0.40	0.46±0.20	0.40±0.07	0.00±0.02	0.00±0.02
rs16950960	**HERC2**	**G**	0.00±0.31	0.00±0.23	±	0.00±0.14	0.00±0.60	0.00±0.31	0.01±0.01	0.37±0.07	0.00±0.02
rs16950987	**HERC2**	**G**	0.70±0.28	0.58±0.19	0.82±0.14	0.85±0.10	1.00±0.60	1.00±0.15	0.93±0.04	0.27±0.06	0.64±0.06
rs6497284	**HERC2**	**T**	1.00±0.37	1.00±0.14	1.00±0.17	1.00±0.07	1.00±0.60	1.00±0.19	0.99±0.01	0.39±0.07	1.00±0.01
rs8039195	**HERC2**	**T**	0.40±0.30	1.00±0.46		1.00±0.31	0.75±0.42	1.00±0.31	0.86±0.05	0.13±0.05	0.35±0.06
rs8043281	**HERC2**	**G**	1.00±0.31	1.00±0.13	1.00±0.37	1.00±0.07	1.00±0.46	1.00±0.23	0.99±0.01	0.40±0.07	1.00±0.01
rs9302376	**HERC2**	**T**	0.00±0.46	0.00±0.12	0.00±0.17	0.00±0.10	0.00±0.60	0.00±0.19	0.01±0.01	0.38±0.07	0.00±0.02
rs1426654	**SLC24A5**	**A**	1.00±0.31	0.19±0.15	0.96±0.08	0.96±0.05	0.83±0.30	1.00±0.15	1.00±0.02	0.01±0.02	0.02±0.02
rs16960620	**SLC24A5**	**G**	0.00±0.37	0.00±0.15	0.00±0.31	0.00±0.08	0.00±0.84	0.00±0.21	0.00±0.02	0.00±0.02	0.13±0.05
rs16960631	**SLC24A5**	**G**	0.10±0.19	0.00±0.26		0.00±0.11	0.00±0.60	0.00±0.37	0.00±0.02	0.01±0.02	0.18±0.05
rs17426596	**SLC24A5**	**C**	0.00±0.37	0.00±0.12	0.10±0.19	0.04±0.05	0.00±0.60	0.00±0.17	0.05±0.03	0.00±0.02	0.00±0.01
rs1834640	**SLC24A5**	**A**	1.00±0.37	0.45±0.21	0.90±0.19	1.00±0.14	1.00±0.60	1.00±0.60	0.99±0.01	0.01±0.02	0.10±0.04
rs2433354	**SLC24A5**	**C**	0.80±0.25	0.50±0.17	1.00±0.15	0.96±0.05	1.00±0.84	1.00±0.17	1.00±0.02	0.27±0.06	0.55±0.07
rs2433356	**SLC24A5**	**G**	1.00±0.84	0.70±0.28	1.00±0.60	0.75±0.42	1.00±0.84	1.00±0.60	1.00±0.02	0.51±0.07	0.93±0.03
rs2459391	**SLC24A5**	**A**	0.90±0.19	0.70±0.16	1.00±0.14	0.96±0.05	1.00±0.60	1.00±0.15	1.00±0.02	0.71±0.06	0.93±0.03
rs2469592	**SLC24A5**	**A**	1.00±0.31	0.72±0.16	1.00±0.15	0.96±0.05	1.00±0.46	1.00±0.23	1.00±0.02	0.27±0.06	0.07±0.03

Table 3 continued from previous page

ID	Gene	Allele	EHG	WHG	SEF	CEF	MN	WEZ	CEU	YRI	CHS
rs2470101	**SLC24A5**	**T**	1.00±0.31	0.72±0.16	1.00±0.17	0.96±0.05	1.00±0.60	1.00±0.17	1.00±0.02	0.76±0.06	0.93±0.03
rs2470102	**SLC24A5**	**A**	0.80±0.25	0.30±0.28		0.93±0.10	1.00±0.60	1.00±0.31	0.98±0.02	0.01±0.02	0.28±0.06
rs2555364	**SLC24A5**	**G**	0.80±0.25	0.53±0.16	1.00±0.84	0.96±0.05	1.00±0.60	1.00±0.23	1.00±0.02	0.28±0.06	0.55±0.07
rs2675345	**SLC24A5**	**A**	0.90±0.19	0.21±0.15	0.96±0.08	0.98±0.04	1.00±0.37	1.00±0.19	1.00±0.02	0.01±0.02	0.26±0.06
rs2675347	**SLC24A5**	**A**	0.80±0.25	0.58±0.28		0.94±0.12	1.00±0.37	1.00±0.37	0.98±0.02	0.27±0.06	0.55±0.07
rs2675348	**SLC24A5**	**A**	1.00±0.37	0.54±0.19	1.00±0.13	0.95±0.06	1.00±0.60	1.00±0.15	1.00±0.02	0.25±0.06	0.55±0.07
rs2675349	**SLC24A5**	**A**		0.67±0.38		0.75±0.42			1.00±0.02	0.27±0.06	0.54±0.07
rs3817315	**SLC24A5**	**C**	0.90±0.19	0.45±0.22		0.91±0.12	1.00±0.60	1.00±0.60	1.00±0.02	0.27±0.06	0.55±0.07
rs7163587	**SLC24A5**	**C**	1.00±0.37	0.70±0.16	1.00±0.13	0.96±0.05	1.00±0.84	1.00±0.14	1.00±0.02	0.72±0.06	0.93±0.03
rs938505	**SLC24A5**	**T**	0.00±0.31	0.33±0.19		0.02±0.04	0.00±0.60	0.00±0.26	0.00±0.02	0.19±0.05	0.07±0.03
rs16891982	**SLC45A2**	**G**	0.25±0.25	0.13±0.16	0.50±0.31	0.27±0.13	0.00±0.46	0.83±0.30	0.98±0.02	0.00±0.02	0.00±0.01
rs26722	**SLC45A2**	**T**	0.50±0.28	0.22±0.14	0.38±0.24	0.22±0.11	0.50±0.40	0.13±0.13	0.00±0.02	0.06±0.03	0.40±0.07
rs28777	**SLC45A2**	**A**	0.20±0.25	0.30±0.28		0.38±0.19	0.00±0.46	0.75±0.30	0.98±0.02	0.16±0.05	0.11±0.04
rs35391	**SLC45A2**	**C**	0.33±0.27	0.50±0.18	0.68±0.19	0.58±0.13	0.00±0.46	0.82±0.16	0.99±0.01	0.57±0.07	0.42±0.07
rs13289	**SLC45A2**	**G**	0.63±0.34	0.09±0.12	0.40±0.30	0.25±0.13	0.33±0.38	0.36±0.25	0.64±0.07	0.25±0.06	0.72±0.06
rs35395	**SLC45A2**	**C**	0.20±0.25	0.14±0.14	0.63±0.34	0.50±0.23	0.00±0.46	0.83±0.21	0.98±0.02	0.19±0.05	0.11±0.04
rs12896399	**SLC24A4**	**T**	0.30±0.28	0.28±0.21	0.50±0.35	0.31±0.14	0.30±0.28	0.29±0.24	0.56±0.07	0.00±0.02	0.27±0.06
rs2402130	**SLC24A4**	**A**	0.92±0.16	0.70±0.16	0.75±0.19	0.86±0.09	0.67±0.38	0.82±0.16	0.79±0.06	0.34±0.07	0.94±0.03
rs1042602	**TYR**	**A**	0.00±0.26	0.00±0.19	0.00±0.84	0.11±0.10	0.33±0.38	0.33±0.28	0.40±0.07	0.00±0.02	0.00±0.01
rs1126809	**TYR**	**A**	0.00±0.26	0.00±0.17	0.00±0.84	0.04±0.05	0.33±0.30	0.25±0.25	0.25±0.06	0.00±0.02	0.00±0.02
rs1393350	**TYR**	**A**	0.00±0.37	0.00±0.12	0.09±0.12	0.06±0.06	0.00±0.42	0.25±0.21	0.24±0.06	0.00±0.02	0.00±0.02
rs1408799	**TYRP**	**C**	0.58±0.28	0.71±0.17	0.65±0.21	0.56±0.13	0.67±0.38	0.50±0.23	0.67±0.07	0.20±0.05	0.01±0.02
rs2733832	**TYRP**	**T**	0.57±0.26	0.69±0.16	0.46±0.20	0.33±0.13	0.50±0.35	0.63±0.19	0.59±0.07	0.04±0.03	0.01±0.02
rs683	**TYRP**	**C**	0.13±0.23	0.30±0.20	0.63±0.34	0.54±0.20	0.50±0.49	0.33±0.27	0.36±0.07	0.91±0.04	0.99±0.01

Table 3 continued from previous page

ID	Gene	Allele	EHG	WHG	SEF	CEF	MN	WEZ	CEU	YRI	CHS
rs10809814	**TYRP**	**A**	0.00±0.37	0.13±0.23		0.21±0.21	0.00±0.60	0.50±0.69	0.07±0.03	0.22±0.06	0.68±0.06
rs10960751	**TYRP**	**C**	0.70±0.28	0.68±0.17	0.56±0.24	0.41±0.13	0.50±0.40	0.60±0.21	0.63±0.07	0.07±0.04	0.01±0.02
rs10960752	**TYRP**	**A**	0.70±0.28	0.71±0.17	0.56±0.24	0.41±0.13	0.50±0.40	0.61±0.23	0.63±0.07	0.06±0.03	0.01±0.02
rs2733831	**TYRP**	**G**	0.50±0.40	0.65±0.18	0.43±0.26	0.38±0.15	0.50±0.40	0.67±0.27	0.61±0.07	0.04±0.03	0.01±0.02
rs4741242	**TYRP**	**G**	0.20±0.25	0.00±0.26		0.14±0.14	0.00±0.60	0.17±0.21	0.10±0.04	0.00±0.02	0.21±0.06
rs932761	**TYRP**	**G**	1.00±0.46	1.00±0.14	1.00±0.31	1.00±0.08	1.00±0.60	1.00±0.23	1.00±0.02	0.79±0.06	1.00±0.02
rs1805005	**MC1R**	**T**	0.00±0.31	0.00±0.23	0.50±0.69	0.04±0.08	0.00±0.60	0.00±0.69	0.07±0.04	0.00±0.02	0.00±0.02
rs1805006	**MC1R**	**A**	0.00±0.26	0.00±0.23		0.00±0.17	0.00±0.46	0.00±0.46	0.02±0.02	0.00±0.02	0.00±0.02
rs2228479	**MC1R**	**A**	0.00±0.26	0.00±0.23		0.00±0.21	0.00±0.37	0.00±0.30	0.07±0.03	0.00±0.02	0.33±0.06
rs11547464	**MC1R**	**A**	0.00±0.31	0.00±0.23		0.00±0.07	0.00±0.46	0.00±0.84	0.00±0.02	0.00±0.02	0.00±0.02
rs1805007	**MC1R**	**T**	0.00±0.26	0.00±0.17	0.00±0.84	0.00±0.07	0.00±0.37	0.00±0.84	0.00±0.05	0.00±0.02	0.00±0.02
rs1110400	**MC1R**	**C**	0.00±0.31	0.00±0.17		0.00±0.08	0.00±0.37	0.00±0.84	0.01±0.01	0.00±0.02	0.00±0.02
rs1805008	**MC1R**	**T**	0.00±0.31	0.00±0.21		0.00±0.08	0.00±0.46	0.00±0.84	0.11±0.04	0.00±0.02	0.00±0.02
rs885479	**MC1R**	**A**	0.00±0.46	0.00±0.21		0.00±0.10	0.00±0.37	0.00±0.84	0.08±0.04	0.00±0.02	0.60±0.07
rs1805009	**MC1R**	**C**	0.00±0.26	0.00±0.12	0.00±0.21	0.00±0.07	0.00±0.60	0.00±0.26	0.01±0.01	0.00±0.02	0.00±0.02
rs201326893	**MC1R**	**C**	0.00±0.31	0.00±0.19		0.00±0.08	0.00±0.37	0.00±0.84	0.00±	0.00±	0.00±
rs6119471	**ASIP**	**C**	1.00±0.37	1.00±0.31		1.00±0.13	1.00±0.60	1.00±0.46	1.00±0.02	0.27±0.06	1.00±0.02
rs2378249	**ASIP**	**G**	0.20±0.25	0.03±0.06	0.12±0.12	0.02±0.03	0.00±0.46	0.31±0.23	0.15±0.05	0.15±0.05	0.20±0.05
rs4911442	**ASIP**	**G**	0.00±0.31	0.06±0.12	0.33±0.38	0.06±0.11	0.00±0.46	0.13±0.23	0.12±0.05	0.00±0.02	0.00±0.02
rs4911414	**ASIP**	**G**	0.00±0.46	0.55±0.22	0.75±0.42	0.55±0.21	1.00±0.84	0.50±0.28	0.33±0.07	0.87±0.05	0.81±0.05
rs1015362	**ASIP**	**A**	0.08±0.16	0.55±0.21	0.50±0.49	0.71±0.14	0.25±0.30	0.50±0.31	0.73±0.06	0.86±0.05	0.19±0.05
rs6058017	**ASIP**	**A**	1.00±0.46	1.00±0.84		1.00±0.46	1.00±0.84		0.88±0.05	0.17±0.05	0.80±0.05
rs12877248	**DCT**	**A**	0.00±0.31	0.00±0.13	0.00±0.14	0.00±0.07	0.00±0.46	0.00±0.14			
rs9516418	**DCT**	**T**	0.90±0.19	0.61±0.23	0.50±0.69	0.55±0.15	1.00±0.60	0.64±0.25	0.61±0.07	0.66±0.07	0.89±0.04

Table 3 continued from previous page

ID	Gene	Allele	EHG	WHG	SEF	CEF	MN	WEZ	CEU	YRI	CHS
rs9584234	**DCT**	**T**	0.00±0.31	0.08±0.10	0.23±0.16	0.14±0.09	0.00±0.46	0.08±0.11	0.12±0.05	0.17±0.05	0.10±0.04
rs16964944	**MYO5A**	**K**	0.00±0.26	0.00±0.15	0.00±0.60	0.00±0.07	0.00±0.37	0.00±0.46	0.00±0.02	0.21±0.06	0.03±0.02
rs1724577	**MYO5A**	**G**	0.00±0.26	0.00±0.11	0.00±0.31	0.00±0.07	0.00±0.46	0.00±0.23	0.00±0.02	0.35±0.07	0.03±0.02
rs4776053	**MYO5A**	**C**	0.25±0.25	0.25±0.15	0.18±0.16	0.12±0.09	0.13±0.23	0.08±0.10	0.13±0.05	0.07±0.04	0.20±0.05
rs4590952	**KITLG**	**A**	0.20±0.25	0.50±0.26	0.50±0.69	0.19±0.15	0.00±0.60	0.38±0.34	0.21±0.06	0.74±0.06	0.20±0.05
rs642742	**KITLG**	**C**	0.92±0.16	0.69±0.23	1.00±0.84	0.57±0.26	0.50±0.49	0.64±0.25	0.82±0.05	0.06±0.03	0.80±0.05
rs12821256	**KITLG**	**T**	0.92±0.16	1.00±0.13	1.00±0.37	1.00±0.07	1.00±0.46	1.00±0.14	0.87±0.05	1.00±0.02	1.00±0.02
rs12203592	**IRF4**	**T**	0.58±0.28	0.58±0.16	0.00±0.17	0.05±0.06	0.67±0.38	0.29±0.18	0.16±0.05	0.00±0.02	0.00±0.02
rs1540771	**IRF4**	**T**	0.70±0.28	0.69±0.15	0.50±0.25	0.54±0.13	0.67±0.38	0.55±0.21	0.55±0.07	0.07±0.04	0.24±0.06

Table (4) *Significant results of Hardy-Weinberg equilibrium calculations for pigmentation related SNPs.*

Group	**ID**	**Gene**	$\mathbf{Ho1_{obs}}$	$\mathbf{He_{obs}}$	$\mathbf{Ho2_{obs}}$	$\mathbf{Ho1_{exp}}$	$\mathbf{He_{exp}}$	$\mathbf{Ho2_{exp}}$	χ^2	**P**	Het deficit	Het excess
EHG	rs26722	SLC45A2	0	6	0	1.50	3.00	1.50	6.00	0.04979	-1.00	1.00
WHG	rs4911414	ASIP	0	7	0	1.75	3.50	1.75	7.00	0.03020	-1.00	1.00
SEF	rs1393350	TYR	10	0	1	9.09	1.82	0.09	11.00	0.00409	1.00	-1.00
CEF	rs4911414	ASIP	0	10	1	2.27	5.45	3.27	7.64	0.02194	-0.83	0.83
WEZ	rs4911414	ASIP	0	7	0	1.75	3.50	1.75	7.00	0.03020	-1.00	1.00

3.3.2 Neutrality tests

A total of 56 SNPs in nine genes deviated from neutral expectations in the Central European reference group, as indicated by at least one of the tests. In the African reference group, 21 SNPs in seven genes were found to deviate from neutral expectations, indicated by at least one of the neutrality tests. In the Chinese reference group, 26 SNPs in eight genes were identified by at least one test to have developed under non-neutral conditions. When results for each groups were compared a certain overlap could be found. From all 56 SNPs that were indicated by one of the tests, 28 were unique to the Central European reference group, while ten also deviated from neutral expectations in the Chinese reference group, additional six in both, the Chinese and the African, as well as 12 also in the African reference group. Three SNPs deviated only in the Chinese reference group, but not in any other groups, while six SNPs were unique in the African reference group. Two SNPs were identified as non-neutral in the Chinese and the African group but not in the Central Europeans (see figure: 8 and table: 5).

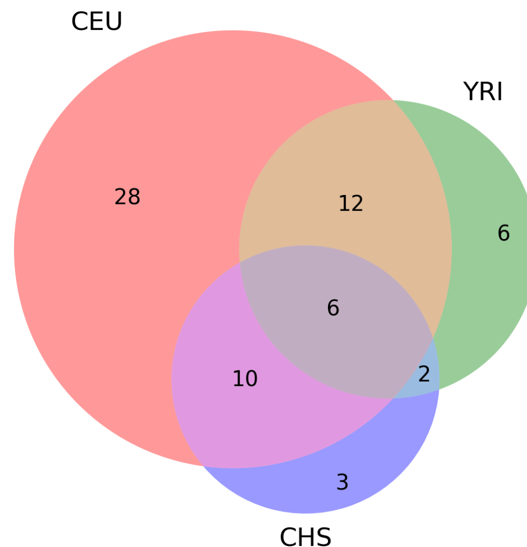

Figure (8) *Venn diagram of the number SNPs that deviated from neutral expectations in each of the three reference groups. CEU=Central Europeans, YRI=Africans, CHS=Chinese.*

Table (5) *Results of neutrality tests for reference groups. Only results were reported, were at least one test result was found in the highest or lowest 2.5^{th} percentile of all results per chromosome. For each test the resulting value and the accompanying percentile are reported.*

ID	Gene	iHS	P_i	nSl	P_i	D	P_i	H12	P_i	H2_H1	P_i	
rs1667394	**OCA2**	2.0758	98.1875	2.3762	99.2734	-2.1172	1.4355	0.5418	98.7584	0.0886	10.5866	CEU
rs1545397	**OCA2**					1.3178	90.4405	0.0143	0.7157	0.8621	99.7154	
rs1800401	**OCA2**	-0.3514	35.4905	-0.1382	43.6148	-1.8496	3.436	0.765	99.934	0.0142	0.8291	
rs7495174	**OCA2**	-0.4538	31.7502	-0.1861	41.7516	-2.1901	1.0415	0.3907	93.6353	0.4973	74.1245	
rs4778136	**OCA2**	-1.3969	8.3206	-1.5327	6.6884	-1.8009	3.8877	0.7476	99.9154	0.0149	0.9054	
rs7494942	**OCA2**	1.0413	85.6025	1.1979	88.7379	-2.2876	0.6043	0.4745	97.1023	0.1168	13.5482	
rs4778241	**HERC2**	-1.614	5.6741	-1.6645	5.2321	-2.1901	1.0415	0.3907	93.6353	0.4973	74.1245	
rs12593929	**HERC2**	-0.7549	21.771	-0.8015	21.0007	-2.0776	1.683	0.2992	86.7766	0.4106	59.3677	
rs3935591	**HERC2**	0.8132	79.4261	0.8974	81.4357	-2.2219	0.891	0.3508	91.2294	0.1412	16.3955	
rs7170852	**HERC2**	1.8109	96.6919	1.9755	97.9368	-2.2744	0.662	0.5382	98.6305	0.0739	9.0583	
rs2238289	**HERC2**	1.4543	93.1375	1.4644	93.351	-2.2761	0.6559	0.658	99.7092	0.0431	5.4346	
rs3940272	**HERC2**					-2.0939	1.5943	0.4917	97.6117	0.0676	8.3921	
rs8028689	**HERC2**					-2.1339	1.3365	0.5928	99.3431	0.0565	7.1753	
rs2240203	**HERC2**					-2.1339	1.3365	0.5928	99.3431	0.0565	7.1753	
rs12913832	**HERC2**	-1.5589	6.2705	-1.9019	3.2397	-2.2876	0.6043	0.4745	97.1023	0.1168	13.5482	
rs1129038	**HERC2**	1.3812	92.0709	1.5017	93.8964	-2.069	1.7366	0.2837	85.1029	0.3887	55.7542	
rs7183877	**HERC2**	-0.4243	32.8354	-0.3851	34.1854	-2.2876	0.6043	0.4745	97.1023	0.1168	13.5482	
rs11635884	**HERC2**					-2.2876	0.6043	0.4745	97.1023	0.1168	13.5482	
rs11636232	**HERC2**	0.7575	77.7484	1.0023	84.2718	-2.2381	0.8085	0.3417	90.5571	0.6484	92.1256	
rs16950987	**HERC2**	-0.5073	29.8023	-0.1532	42.9942	-2.1172	1.4355	0.5418	98.7584	0.0886	10.5866	
rs6497284	**HERC2**					-2.2453	0.7755	0.5237	98.3397	0.0765	9.3429	
rs8039195	**HERC2**					-2.1166	1.4417	0.5411	98.7367	0.0619	7.7528	
rs8043281	**HERC2**					-2.2738	0.6662	0.6249	99.5463	0.1035	12.0942	

rs1426654	**SLC24A5**					-2.4506	0.1959	0.8182	99.9794	0.0047	0.0536	
rs16960620	**SLC24A5**					-2.4427	0.2042	0.827	99.9825	0.001	0.0052	
rs16960631	**SLC24A5**					-2.5275	0.0784	0.6959	99.8216	0.0025	0.0309	
rs17426596	**SLC24A5**	-0.6647	24.4922	-0.8007	21.0252	-2.3252	0.4744	0.8182	99.9794	0.0047	0.0536	
rs1834640	**SLC24A5**			0.2367	58.6708	-2.5707	0.0289	0.7556	99.9216	0.0017	0.0165	
rs2433354	**SLC24A5**					-2.4427	0.2042	0.827	99.9825	0.001	0.0052	
rs2433356	**SLC24A5**					-2.4427	0.2042	0.827	99.9825	0.001	0.0052	
rs2459391	**SLC24A5**					-2.4427	0.2042	0.827	99.9825	0.001	0.0052	
rs2469592	**SLC24A5**					-2.5707	0.0289	0.827	99.9825	0.001	0.0052	
rs2470101	**SLC24A5**					-2.4908	0.1341	0.827	99.9825	0.001	0.0052	
rs2470102	**SLC24A5**					-2.5275	0.0784	0.6959	99.8216	0.0025	0.0309	
rs2555364	**SLC24A5**					-2.4346	0.2227	0.8182	99.9794	0.0047	0.0536	
rs2675345	**SLC24A5**					-2.5707	0.0289	0.827	99.9825	0.001	0.0052	
rs2675347	**SLC24A5**			0.3186	61.8672	-2.4346	0.2227	0.8182	99.9794	0.0047	0.0536	
rs2675348	**SLC24A5**					-2.3995	0.2949	0.8182	99.9794	0.0047	0.0536	
rs2675349	**SLC24A5**					-2.5223	0.0846	0.6959	99.8216	0.0025	0.0309	
rs3817315	**SLC24A5**					-2.2898	0.594	0.698	99.8371	0.009	0.2846	
rs7163587	**SLC24A5**					-2.2898	0.594	0.8003	99.966	0.0079	0.2073	
rs938505	**SLC24A5**					-2.4908	0.1341	0.827	99.9825	0.001	0.0052	
rs16891982	**SLC45A2**			3.2709	99.9351	-2.5516	0.038	0.7474	99.9848	0.0102	0.3124	
rs26722	**SLC45A2**					-2.5516	0.038	0.739	99.9772	0.0096	0.2554	
rs28777	**SLC45A2**			0.9249	82.4985	-2.5516	0.038	0.7474	99.9848	0.0102	0.3124	
rs35391	**SLC45A2**					-2.5516	0.038	0.7474	99.9848	0.0102	0.3124	
rs35395	**SLC45A2**			0.9299	82.6451	-2.5203	0.0513	0.7474	99.9848	0.0102	0.3124	
rs1042602	**TYR**	2.2912	98.8646	2.6302	99.6343	0.4256	63.336	0.2188	75.1177	0.5135	80.652	

rs1393350	**TYR**	0.4264	66.3897	0.5637	70.7637	-1.3339	7.3617	0.4956	98.1268	0.1096	12.7532	
rs1408799	**TYRP**	2.9649	99.8374	2.4093	99.3891	0.4061	64.5704	0.0851	24.2529	0.2423	33.8974	
rs10960751	**TYRP1**	-4.2692	0.0056	-3.953	0.0094	-0.5582	28.044	0.191	68.186	0.1213	15.8203	
rs10960752	**TYRP1**	3.3097	99.9437	2.7072	99.7434	-0.5582	28.044	0.191	68.186	0.1213	15.8203	
rs6119471	**ASIP**					-1.2821	9.5092	0.5391	98.811	0.0568	7.4126	
rs4590952	**KITLG**	-2.58	0.6328	-2.6268	0.5642	0.4312	65.7463	0.497	98.1909	0.0143	0.7769	
rs642742	**KITLG**	-1.01	15.2777	-1.8715	3.384	-1.8701	2.8348	0.5398	98.9691	0.0089	0.2036	
rs1540771	**IRF4**	-0.4875	30.1671	-0.3449	35.3848	-0.631	22.8027	0.0179	1.1311	0.675	95.1316	
rs1667394	**OCA2**	2.0758	98.1875	2.3762	99.2734	-1.2461	11.6673	0.0244	15.6736	0.8027	93.9044	YRI
rs1545397	**OCA2**					-1.7176	1.3571	0.2904	99.5896	0.0346	0.9446	
rs1800401	**OCA2**	-0.3514	35.4905	-0.1382	43.6148	-1.8934	0.429	0.319	99.8103	0.0754	5.6016	
rs1800414	**OCA2**	nan	nan	nan	nan	-1.7182	1.355	0.2904	99.5896	0.0346	0.9446	
rs4778136	**OCA2**	-1.3969	8.3206	-1.5327	6.6884	-1.7958	0.8229	0.2246	98.2562	0.0814	6.5576	
rs16960631	**SLC24A5**					-0.9444	25.6713	0.2372	98.7048	0.5173	69.5273	
rs2470102	**SLC24A5**					-0.9867	23.2541	0.2372	98.7048	0.5173	69.5273	
rs2675349	**SLC24A5**					-0.9267	26.6634	0.2372	98.7048	0.5173	69.5273	
rs16891982	**SLC45A2**	nan	nan	3.2709	99.9351	-0.887	31.4411	0.0234	9.0691	0.6156	85.0149	
rs1042602	**TYR**	2.2912	98.8646	2.6302	99.6343	-0.2091	70.7336	0.0403	21.9999	0.5659	79.8813	
rs1408799	**TYRP**	2.9649	99.8374	2.4093	99.3891	-0.0192	79.541	0.0473	30.5023	0.7607	93.5216	
rs2733832	**TYRP**	1.4837	93.5194	1.9721	97.9849	-2.0573	0.3469	0.1944	94.7518	0.0952	11.7034	
rs10960751	**TYRP1**	-4.2692	0.0056	-3.953	0.0094	-0.3442	63.5705	0.113	73.8526	0.1108	14.8084	
rs10960752	**TYRP1**	3.3097	99.9437	2.7072	99.7434	-0.3442	63.5705	0.113	73.8526	0.1108	14.8084	
rs2733831	**TYRP1**	1.1705	88.2597	1.7217	96.2536	-2.0573	0.3469	0.2023	95.539	0.1156	15.7936	
rs4911442	**ASIP**	-0.512	29.7137	-0.7953	21.3689	-1.0885	16.9494	0.2422	97.9903	0.1452	18.6488	

rs4590952	**KITLG**	-2.58	0.6328	-2.6268	0.5642	0.3257	88.9415	0.1034	68.7575	0.5278	74.7179	
rs642742	**KITLG**	-1.01	15.2777	-1.8715	3.384	-1.7784	1.2603	0.2409	98.072	0.0476	3.1967	
rs1540771	**IRF4**	-0.4875	30.1671	-0.3449	35.3848	0.2589	85.1316	0.0057	0.0891	0.9259	99.4178	
rs2378249	**ASIP**	1.1064	87.1586	0.9392	82.7759	-1.3893	6.7588	0.2075	96.0633	0.0416	1.9159	
rs7170852	**HERC2**	1.8109	96.6919	1.9755	97.9368	-0.1502	73.1036	0.1235	82.7332	0.1991	26.8593	
rs1800401	**OCA2**					-1.918	1.9284	0.6928	99.7092	0.2212	25.3331	CHS
rs1800407	**OCA2**					2.777	98.9976	0.0931	24.4411	0.4204	57.0917	
rs1800414	**OCA2**	-1.2359	11.1207	-1.5812	6.0993	-0.9684	15.7674	0.4455	94.5758	0.2553	30.384	
rs1800404	**OCA2**	-1.2601	10.7168	-0.9864	16.4732	3.4121	99.7339	0.255	77.9844	0.436	59.9905	
rs2015343	**OCA2**	-1.3006	10.0198	-0.9675	16.9152	3.4121	99.7339	0.2696	80.2242	0.3831	50.527	
rs4778136	**OCA2**	0.7518	77.6732	0.6095	72.1139	3.1519	99.5318	0.6088	99.1111	0.2497	29.6085	
rs7494942	**OCA2**	-0.0172	48.0336	-0.3198	36.4988	-2.1429	0.8869	0.25	77.0387	0.2475	29.261	
rs12593929	**HERC2**	0.0117	49.1917	-0.1824	41.6519	-2.0651	1.1343	0.1156	33.8417	0.1611	17.2699	
rs12913832	**HERC2**	nan	nan	-1.1923	12.1278	-2.1429	0.8869	0.25	77.0387	0.2475	29.261	
rs1129038	**HERC2**					-2.0651	1.1343	0.1156	33.8417	0.1611	17.2699	
rs7183877	**HERC2**	-0.5894	26.8963	-0.5019	30.1368	-2.1429	0.8869	0.25	77.0387	0.2475	29.261	
rs11635884	**HERC2**					-2.1429	0.8869	0.25	77.0387	0.2475	29.261	
rs11636232	**HERC2**					-2.1657	0.8209	0.4307	93.9096	0.1107	11.6322	
rs8043281	**HERC2**					-2.2005	0.728	0.3945	91.9234	0.2524	29.9489	
rs16891982	**SLC45A2**					0.5262	66.918	0.0256	1.2976	0.8337	99.7037	
rs26722	**SLC45A2**	1.1669	88.4878	1.1844	88.6519	0.5686	68.2816	0.0274	1.4918	0.5831	87.9876	
rs35391	**SLC45A2**	0.2093	57.2645	0.328	61.5452	0.5262	66.918	0.0261	1.3465	0.8375	99.737	
rs35395	**SLC45A2**	0.3473	62.7421	0.7754	77.5619	0.5262	66.918	0.0256	1.2976	0.8337	99.7037	
rs2402130	**SLC24A4**	1.0045	84.7855	0.6529	73.5718	-2.2714	0.5443	0.4676	95.6784	0.0673	6.6768	

rs1393350	**TYR**					-1.1591	11.6878	0.4557	94.4472	0.0211	1.9319	
rs683	**TYRP1**					-2.0408	1.9844	0.2868	81.6469	0.375	51.6853	
rs2733831	**TYRP1**					-2.0343	2.0335	0.364	89.1828	0.1282	14.1294	
rs4590952	**KITLG**	-1.2463	10.5147	-0.9614	16.4818	0.658	68.7077	0.5839	97.873	0.0408	5.6062	
rs642742	**KITLG**	-2.3996	1.1231	-2.9924	0.3294	-1.6422	6.8407	0.5424	96.7658	0.0124	1.314	
rs12203592	**IRF4**					1.0031	78.7529	0.032	2.7659	0.6526	93.2107	
rs1540771	**IRF4**	-0.2864	37.2433	-0.319	36.2424	1.7356	92.2697	0.0063	0.0781	0.937	99.9224	

3.3.3 Average expected heterozygosity D_e

SNPs located in the ASIP gene

The gene diversity D_e calculated for all SNPs in the ASIP gene revealed that diversity was elevated or high in all groups. The highest D_e in the reference groups was estimated for the Central European group (D_e: 0.2538±0.0142), followed by the African reference (D_e: 0.2347±0.0147) and the Chinese reference group (D_e: 0.2147±0.0122). The highest value for the ancient sample groups was found for the Bronze Age group (D_e: 0.3736±0.0473), followed by the Middle Neolithic group (D_e: 0.3479±0.0580). The lowest value was estimated for the Central European farmers (D_e: 0.2503±0.0330). For the farmers from the Aegean/Balkan region, a higher value was estimated (D_e: 0.3207±0.0675), while the two hunter-gatherer groups were found to have a diversity intermediate between the two early farmer groups (EHG - D_e: 0.2796±0.0554; WHG - D_e: 0.2982±0.0352). Significant differences were found between the Chinese reference group and the Bronze Age group ($p \leq 0.0010$) as well as the African reference group ($p \leq 0.0050$, see figure: 9).

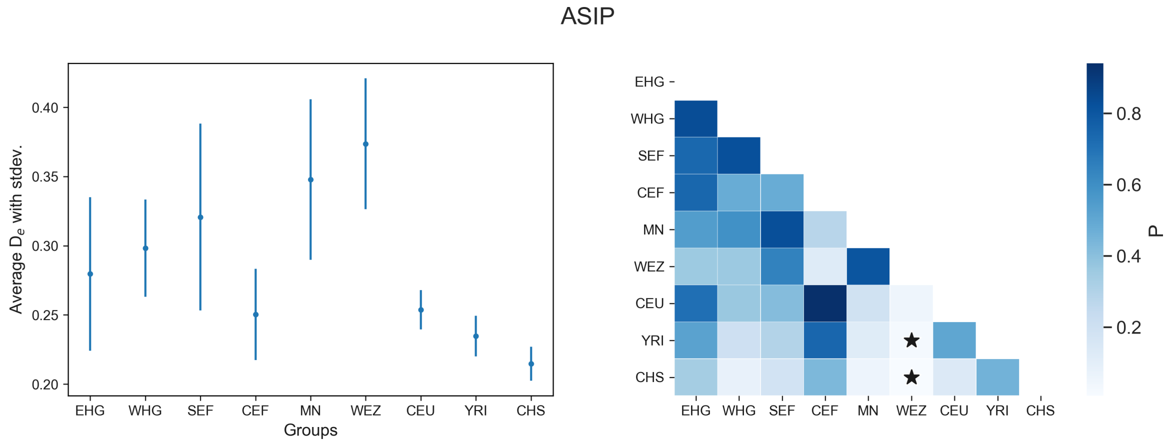

Figure (9) *Average expected heterozygosity D_e for all groups for the SNPs in the ASIP gene. P-values for differences between groups are displayed in the heatmap. Significant values are highlighted by black stars.*

SNPs located in the OCA2 gene

The average diversity was lowest in the Central European reference (D_e: 0.2340±0.0084) group for the OCA2 SNPs, followed by the African reference group (D_e: 0.2654±0.0069). The Chinese reference group had the highest value from the three references (D_e: 0.3332±0.0061). In the ancient sample groups, the highest diversity was found for the farmers from the Aegean/Balkan region (D_e: 0.3549±0.0334), followed by the Middle Neolithic group (D_e: 0.3465±0.0354) and the hunter-gatherers from Eastern Europe/Russia (D_e: 0.3135±0.0323). The hunter-gatherers from Central Europe had a similarly high diversity (D_e: 0.2768±0.0262) compared to the Central European farmers (D_e: 0.2761±.0227), with a slightly higher diversity found in the Bronze Age group (D_e: 0.2916±0.0309). For the European groups, significant differences were found between the Central European reference group and the farmers from the Aegean/Balkan region ($p \leq 0.0100$) and the Middle Neolithic group ($p \leq 0.0500$). Additionally, the Central European reference group had a significantly different diversity compared to the African reference ($p \leq 0.0500$) and the Chinese reference group

($p \leq 0.0001$, for details see figure 10).

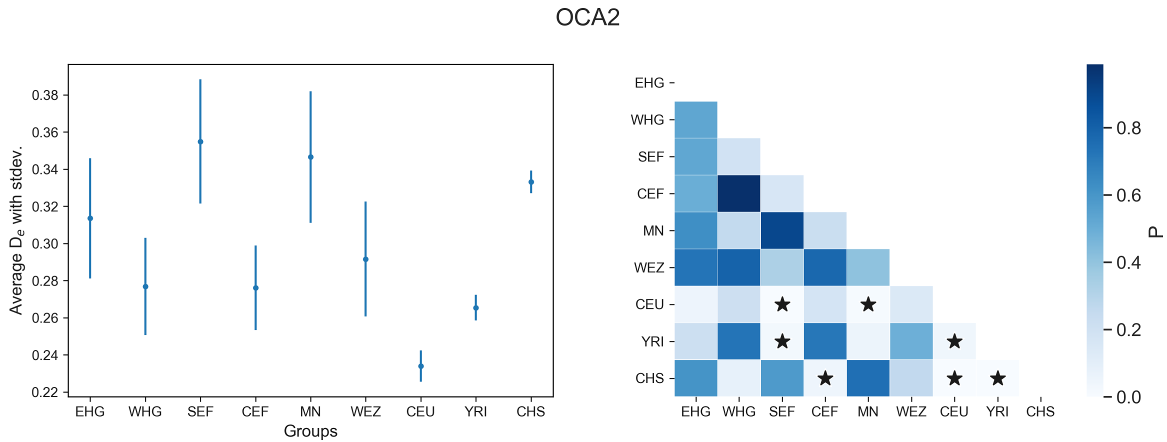

Figure (10) *Average expected heterozygosity D_e for all groups for the SNPs in the OCA2 gene. P-values for differences between groups are displayed in the heatmap. Significant values are highlighted by black stars.*

SNPs located in the HERC2 gene

For the SNPs in the HERC2 gene, the average diversity was lowest in the Central European reference group (D_e: 0.1896±0.0074), while highest in the hunter-gatherers from Eastern Europe (D_e: 0.3431±0.0280). In the two non-European reference groups, the diversity was similar (YRI - D_e: 0.2771±0.0063, CHS - D_e: 0.2840±0.0040). In the remaining ancient sample groups, the diversity was highest in the Middle Neolithic group (D_e: 0.3633±0.0299) followed by the hunter-gatherers of Central Europe (D_e: 0.3121±0.0216). The early farmers from the Aegean Balkan region had a slightly higher diversity (D_e: 0.2940±0.0293) compared to the Central European farmers (D_e: 0.2673±0.0190), while diversity was lowest in the Bronze Age group (D_e: 0.2507±0.0258). The diversity in the Central European reference group was found to be significantly lower, compared to all ancient sample groups, except the Bronze Age group (EHG: $p \leq 0.0001$, WHG: $p \leq 0.0001$, SEF: $p \leq 0.0100$, CEF: $p \leq 0.0100$, MN: $p \leq 0.001$), as well as both other reference groups (CHS: $p \leq 0.0001$, YRI: $p \leq 0.0001$). Between the ancient sample groups, the differences between the Middle Neolithic group and the Central European farmers ($p \leq 0.0500$), as well as the Bronze Age group ($p \leq 0.0500$) were also significant (for detail see figure: 11).

SNPs located in the TYRP1 gene

Diversity for the TYRP1 gene was lowest in the Chinese reference group, followed by the African reference. For all groups of European origin, an elevated D_e was estimated (EHG - D_e: 0.3551±0.0408, WHG - D_e: 0.3392±0.0300, SEF - D_e: 0.4125±0.0303, CEF - D_e: 0.3767±0.0231, MN - D_e: 0.3778±0.0411, WEZ - D_e: 0.3777±0.0336, CEU - D_e: 0.3447±0.0080, YRI - D_e: 0.1739±0.0114, CHS - D_e: 0.1054±0.0070). Significant differences between the Chinese reference group, as well as the African reference group and every other group, were found ($p \leq 0.0500$, for details see figure 12).

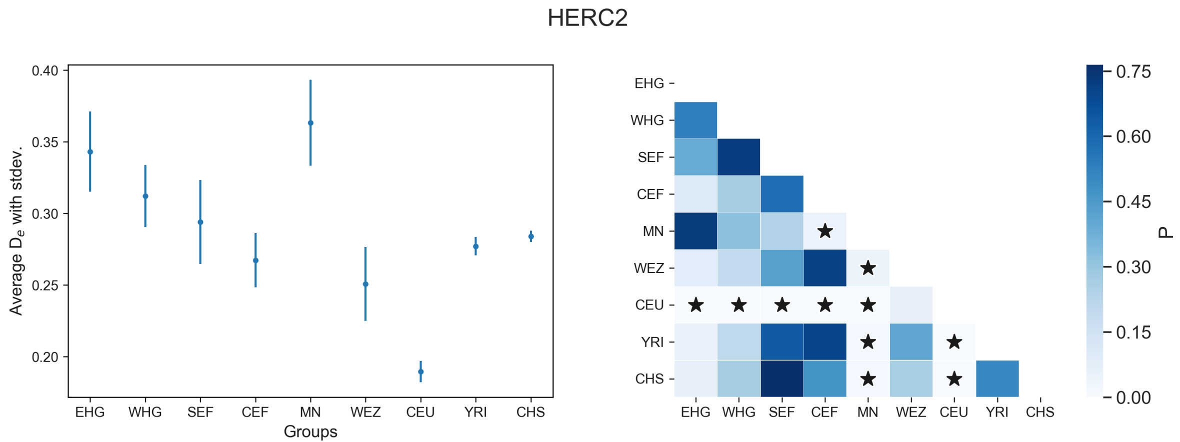

Figure (11) *Average expected heterozygosity D_e for all groups for the SNPs in the HERC2 gene. P-values for differences between groups are displayed in the heatmap. Significant values are highlighted by black stars.*

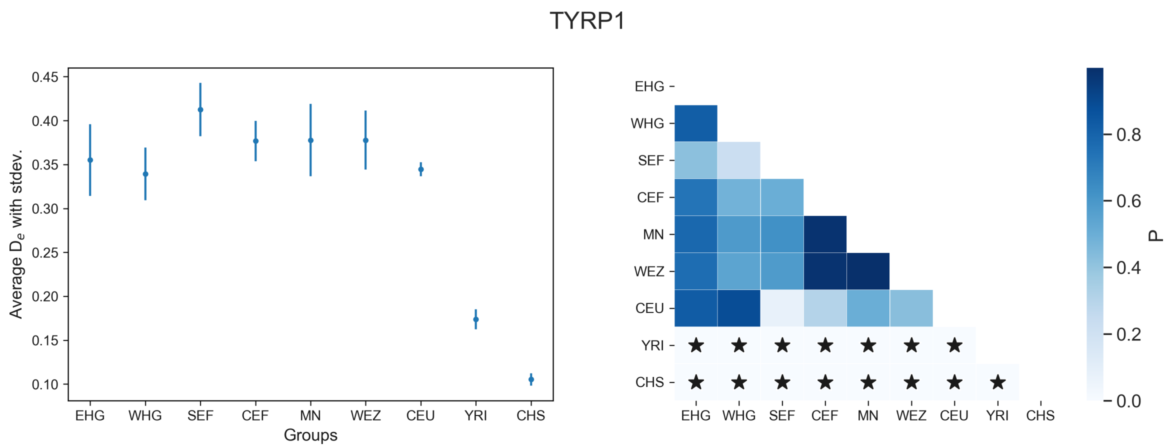

Figure (12) *Average expected heterozygosity D_e for all groups for the SNPs in the TYRP1 gene. P-values for differences between groups are displayed in the heatmap. Significant values are highlighted by black stars.*

SNPs located in the SLC245 gene

For the SLC24A5 gene, diversity was lowest in the Central European reference group while highest in the Central European hunter-gatherer group. In the ancient sample groups, the lowest diversity was estimated for the Central European farmers, followed by the farmer group from the Aegean/Balkan region (EHG - D_e: 0.2744±0.0310, WHG - D_e: 0.3674±0.0196, SEF - D_e: 0.1882±0.0273, CEF - D_e: 0.1277±0.0163, MN - D_e: 0.3487±0.0337, WEZ - D_e: 0.2122±0.0253, CEU - D_e: 0.0257±0.0033, YRI - D_e: 0.2577±0.0061, CHS - D_e: 0.2768±0.0059). The majority of comparisons led to significant results (for details see figure: 13).

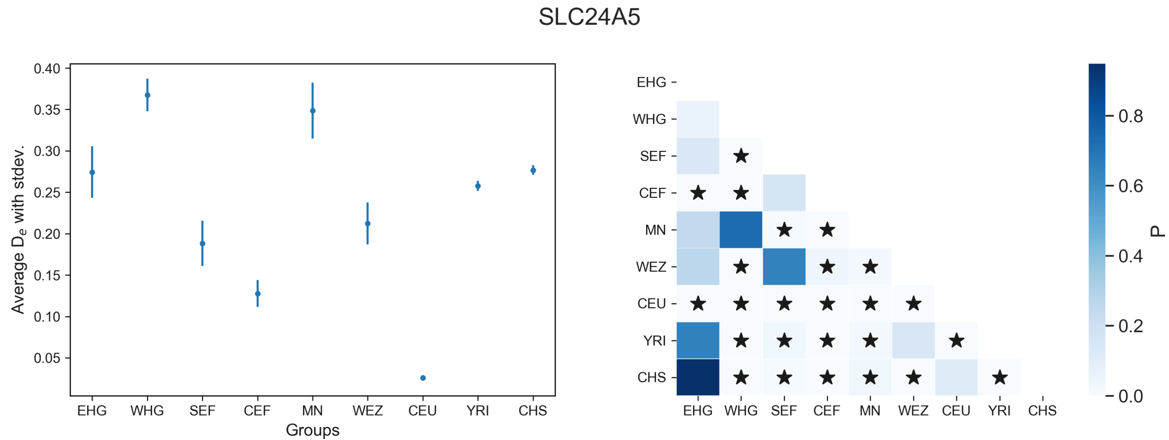

Figure (13) *Average expected heterozygosity D_e for all groups for the SNPs in the SLC24A5 gene. P-values for differences between groups are displayed in the heatmap. Significant values are highlighted by black stars.*

SNPs located in the SLC45A2 gene

The lowest gene diversity D_e was estimated for the Central European reference group (D_e: 0.1024±0.0079) for the locus SLC45A2. The highest diversity was found for the Chinese reference group (D_e: 0.2935±0.0105) with a slightly lower D_e found for the African group (D_e: 0.2608±0.0121). In the ancient sample groups, higher diversity was estimated for all groups, compared to the modern reference groups (EHG - D_e: 0.3655±0.0514; WHG - D_e: 0.3029±0.0422; SEF - D_e: 0.4178±0.0390; CEF - D_e: 0.4115±0.0239; MN - D_e: 0.3403±0.0565; WEZ - D_e: 0.3111±0.0515). In comparison, the diversity estimated for the Central European reference group was found to be significantly lower, compared to all other groups (see figure 14).

SNPs located in the MC1R gene

The diversity across the SNPs in the MC1R gene were lowest in the African reference group (0.0198±0.0038), while being similarly low in the Central European (0.1095±0.0099) and Chinese reference group (0.1172±0.0046). The lowest diversity in the ancient sample groups was found in the Central European farmers (0.0964±0.0179), followed by the Central European hunter-gatherers (0.1716±0.0294). In the Eastern European/Russian hunter-gatherers an elevated diversity was found (0.2411±0.0388) that was even higher in the early farmers from the Aegean/Balkan region (0.2953±0.0729) the Middle Neolithic group (0.3071±0.0442), and the Bronze Age group

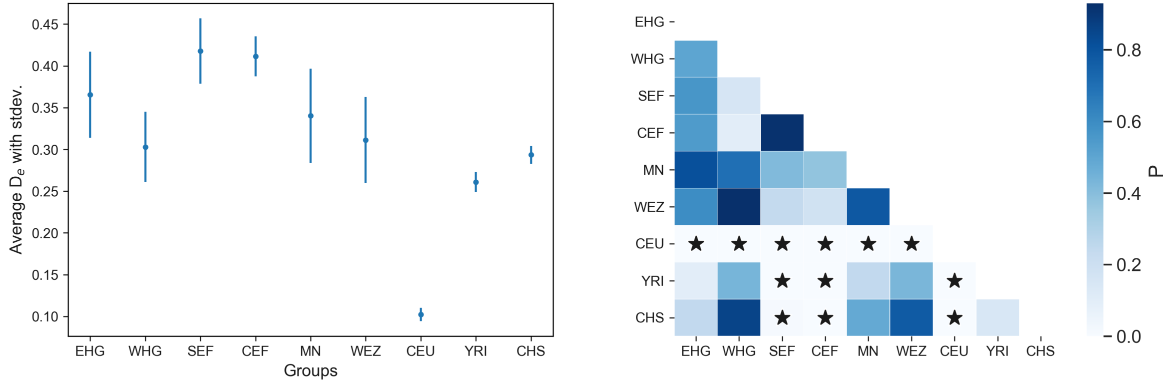

Figure (14) *Average expected heterozygosity D_e for all groups for the SNPs in the SLC45A2 gene. P-values for differences between groups are displayed in the heatmap. Significant values are highlighted by black stars.*

(0.3364±0.0417). Since the diversity varied among the groups of the data set, several of the estimated diversities differed significantly (for details see figure 15).

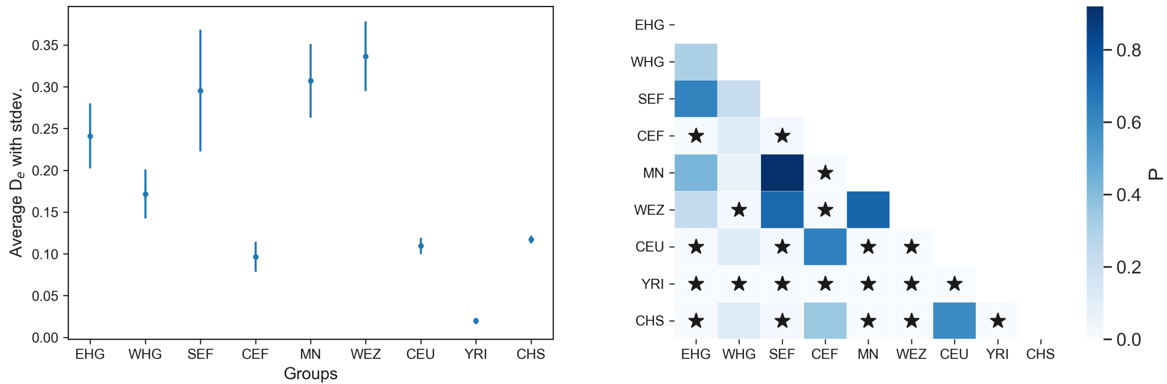

Figure (15) *Average expected heterozygosity D_e for all groups for the SNPs in the MC1R gene. P-values for differences between groups are displayed in the heatmap. Significant values are highlighted by black stars.*

3.3.4 Pairwise F_{st} comparison

SNPs located in the ASIP gene

Significant differences could be found when the Central European group was compared to the Eastern European/Russian hunter-gatherers (Diff: -0.1409 $p \leq 0.0005$) and the African reference group (Diff: -0.2343, $p \leq 0.0001$) for the average F_{st} values for the SNPs in the ASIP gene (see figure: 16).

SNPs located in the OCA2 gene

For the SNPs in the OCA2 gene, three pairings were found, where the average F_{st} was higher, compared to neutral loci when compared to the Central European reference group (CEF - Diff: -0.0202,

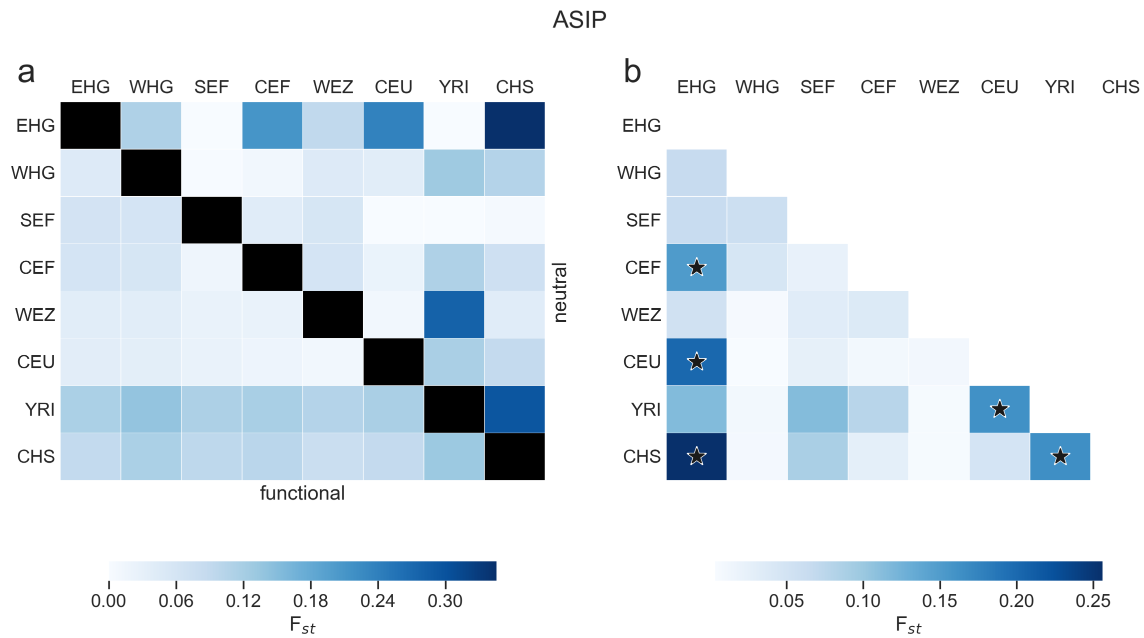

Figure (16) *a) Pairwise F_{st} comparison between SNPs in the ASIP gene and SNPs randomly drawn from the neutralome. b) Absolute differences between F_{st} values. Significant values are highlighted by a black star.*

p≤0.0500, CHS - Diff: -0.2918, p≤0.0001, YRI - Diff: -0.2104, p≤0.0001, see. figure: 17).

SNPs located in the HERC2 gene

For the SNPs located in the HERC2 gene, the average F_{st} was higher than across neutral variants for all but the Bronze Age group and the Central European hunter-gatherers when compared to the Central European reference group (EHG - Diff: -0.1648, p≤0.0001, SEF - Diff: -0.0691 p≤0.0001, CEF - Diff: -0.0522, p≤0.0001, CHS - Diff: -0.2081, p≤0.0001, YRI - Diff: -0.4612, p≤0.0001, see figure: 18).

SNPs located in the TYRP1 gene

When the average F_{st} across the nine SNPs located in the TYRP1 gene was compared to average neutral F_{st} values, significant differences between the Central European reference and the Central European farmer group (-0.0315, p≤0.0100), the African reference group (-0.2466, p≤0.0001) and the Chinese reference group (Diff: -0.3813, p≤0.0001) could be found (see figure: 19.

SNPs located in the SLC24A5 gene

Comparison of average F_{st} values revealed that the average F_{st} values across all loci in the SLC24A5 gene were significantly higher compared to average F_{st} values drawn at random from the neutralome regions between the Central European reference group and the Central European hunter-gatherers (Diff: -0.3290, p≤0.0001), the African reference group (Diff: -0.4102, p≤0.0001), and the Chinese reference group (Diff: -0.2619, p≤0.0001, see figure: 20).

SNPs located in the SLC45A2 gene

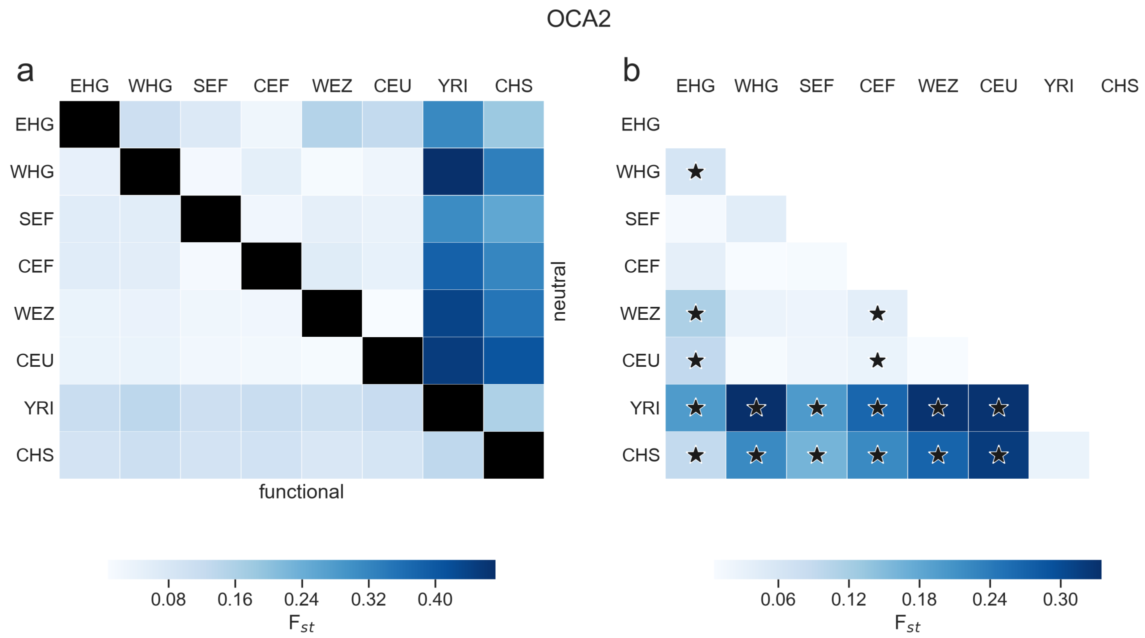

Figure (17) *a) Pairwise F_{st} comparison between SNPs in the OCA2 gene and SNPs randomly drawn from the neutralome. b) Absolute differences between F_{st} values. Significant values are highlighted by a black star.*

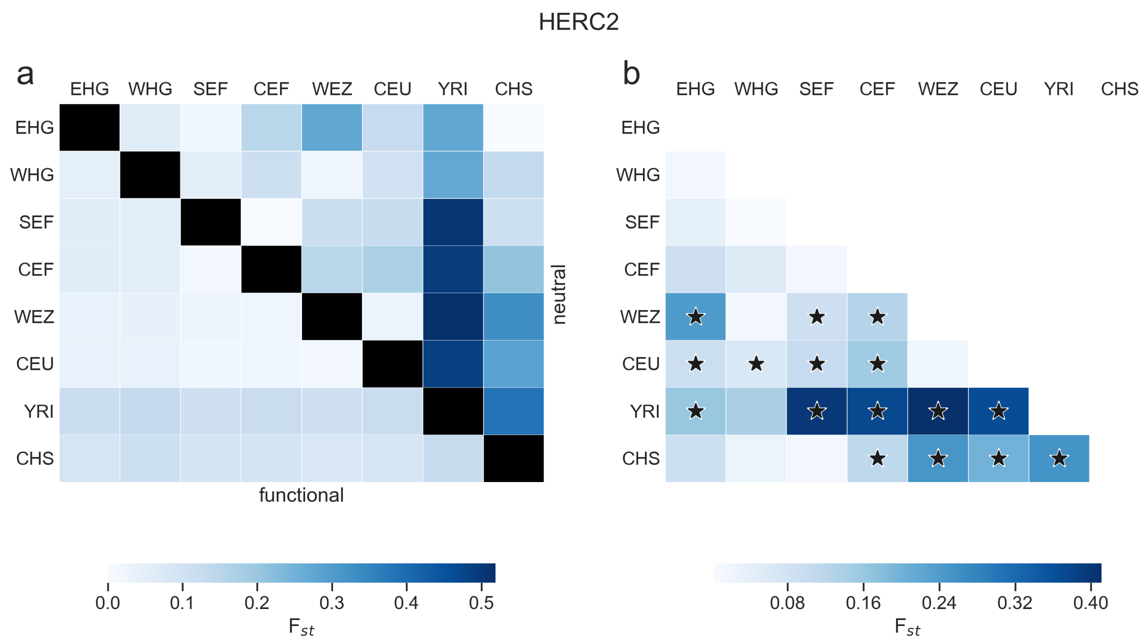

Figure (18) *a) Pairwise F_{st} comparison between SNPs in the HERC2 gene and SNPs randomly drawn from the neutralome. b) Absolute differences between F_{st} values. Significant values are highlighted by a black star.*

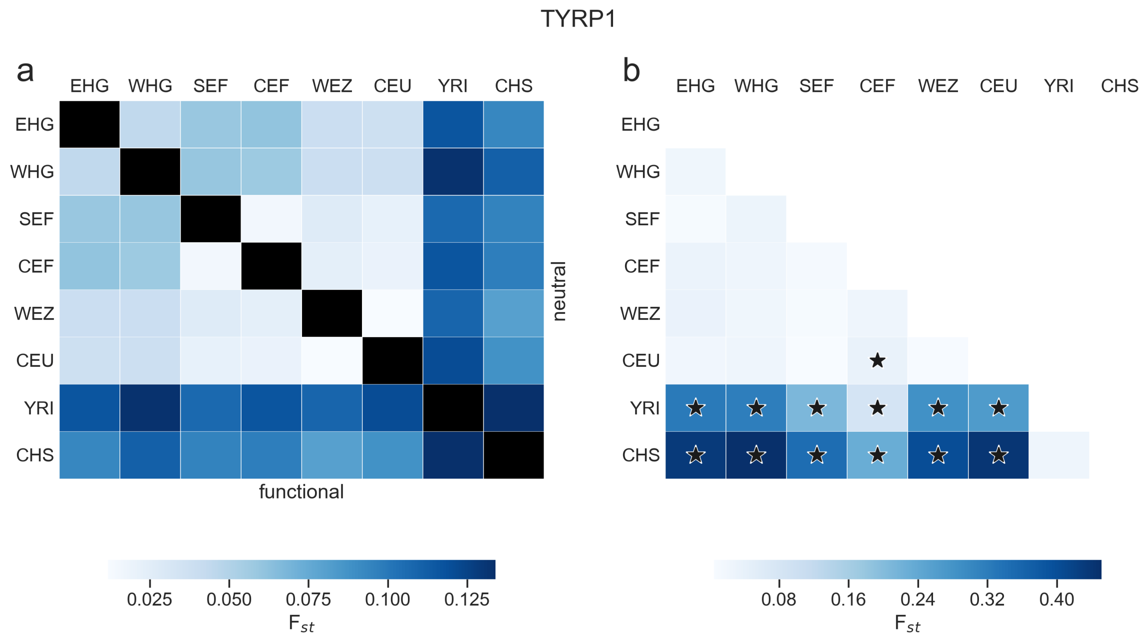

Figure (19) *a) Pairwise F_{st} comparison between SNPs in the TYRP1 gene and SNPs randomly drawn from the neutralome. b) Absolute differences between F_{st} values. Significant values are highlighted by a black star.*

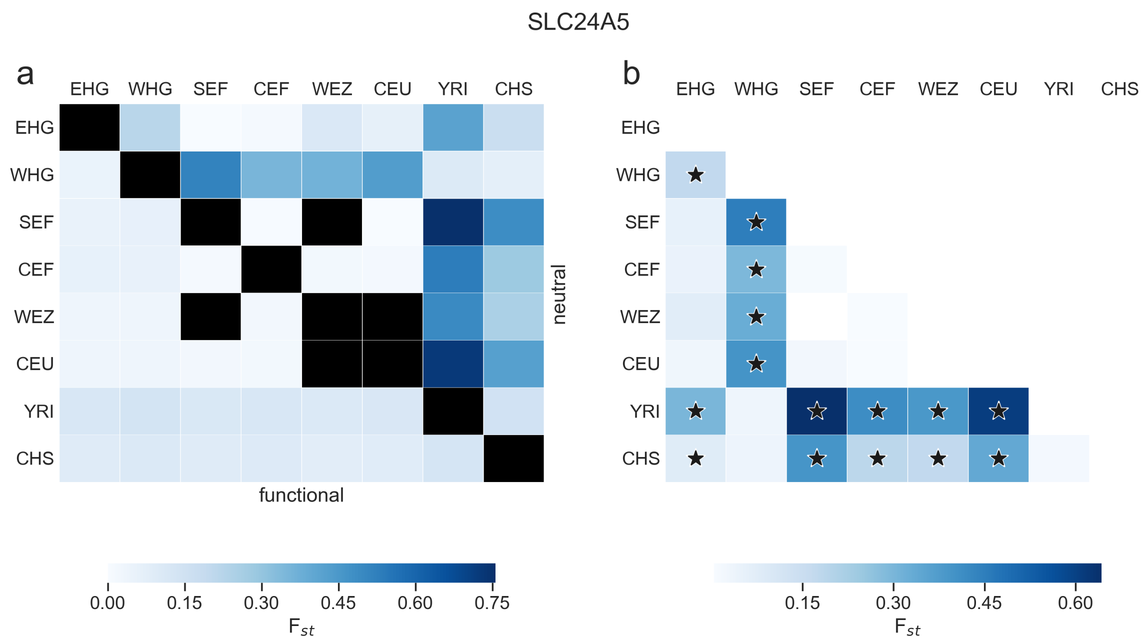

Figure (20) *a) Pairwise F_{st} comparison between SNPs in the SLC24A5 gene and SNPs randomly drawn from the neutralome. b) Absolute differences between F_{st} values. Significant values are highlighted by a black star.*

For the SLC45A2 gene significant differences were found between the Central European reference and every other group in the data set (EHG - Diff: -0.5116, $p \leq 0.0001$, WHG - Diff: -0.5416, $p \leq 0.0001$, SEF - Diff: -0.2424, $p \leq 0.0001$, CEF - Diff: -0.4093, $p \leq 0.0001$, WEZ - Diff: -0.0878, $p \leq 0.0001$, YRI - Diff: -0.4314, $p \leq 0.0001$, CHS - Diff: -0.5256, $p \leq 0.0001$, see figure: 21).

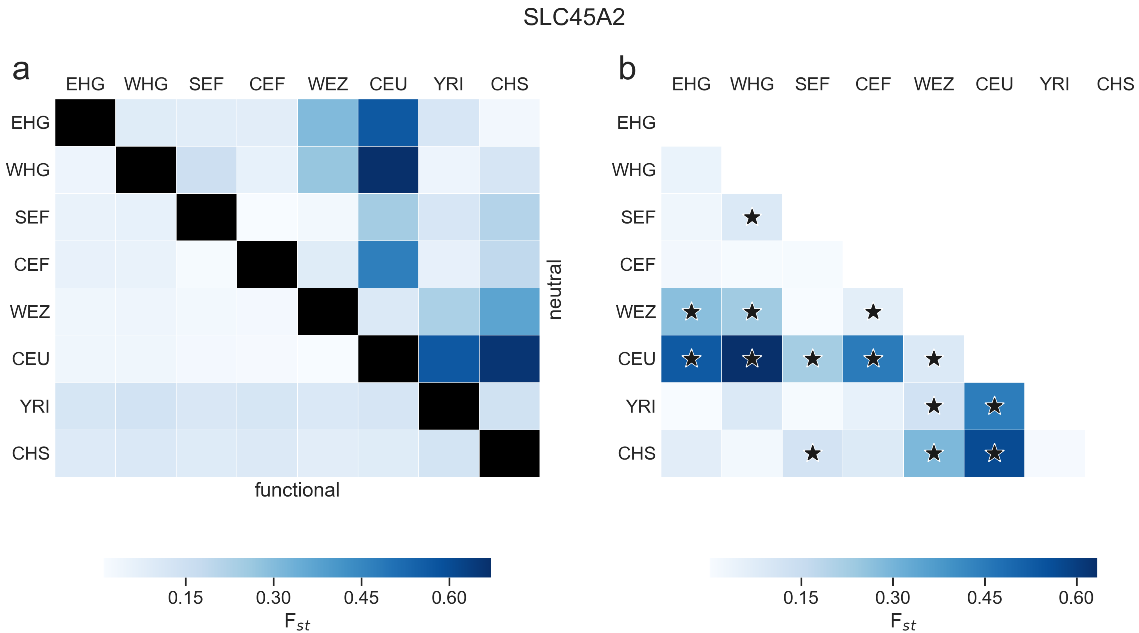

Figure (21) *a) Pairwise F_{st} comparison between SNPs in the SLC45A2 gene and SNPs randomly drawn from the neutralome. b) Absolute differences between F_{st} values. Significant values are highlighted by a black star.*

SNPs located in the MC1R gene

During pairwise F_{st} comparisons of the MC1R SNPs the early farmers from the Aegean/Balkan region and the Bronze Age group had to be excluded from the analysis, since only one SNP could be genotyped in a sufficient number of samples. Significantly higher neutral F_{st} values were found four comparisons; between the African reference group and the early farmers from Central Europe (diff: 0.1187, $p \leq 0.0100$), the hunter-gatherers of Central Europe (diff: 0.1366, $p \leq 0.0085$) and the Eastern European/Russian hunter-gatherers (diff: 0.1175, $p \leq 0.0500$). A significantly higher neutral F_{st} was also found between the early farmers and the hunter-gatherers of Central Europe (diff: 0.0577, $p \leq 0.0500$, see figure: 22).

3.3.5 Pigmentation Score

Among all groups investigated here, the African reference group led to the by far lowest pigmentation scores (6.02±3.73). In the groups of European origin, the lowest value for the mean score was calculated for the Iberians from Spain (IBS) with 32.02±4.58, followed by the Toscani from Italy (TSI) with 32.79±4.79. The mean values for the scores of the Central Europeans (CEU), British (GBR) and Finish (FIN) were all close with 36.45±3.59, 37.16±3.04 and 37.82±2.59 respectively.

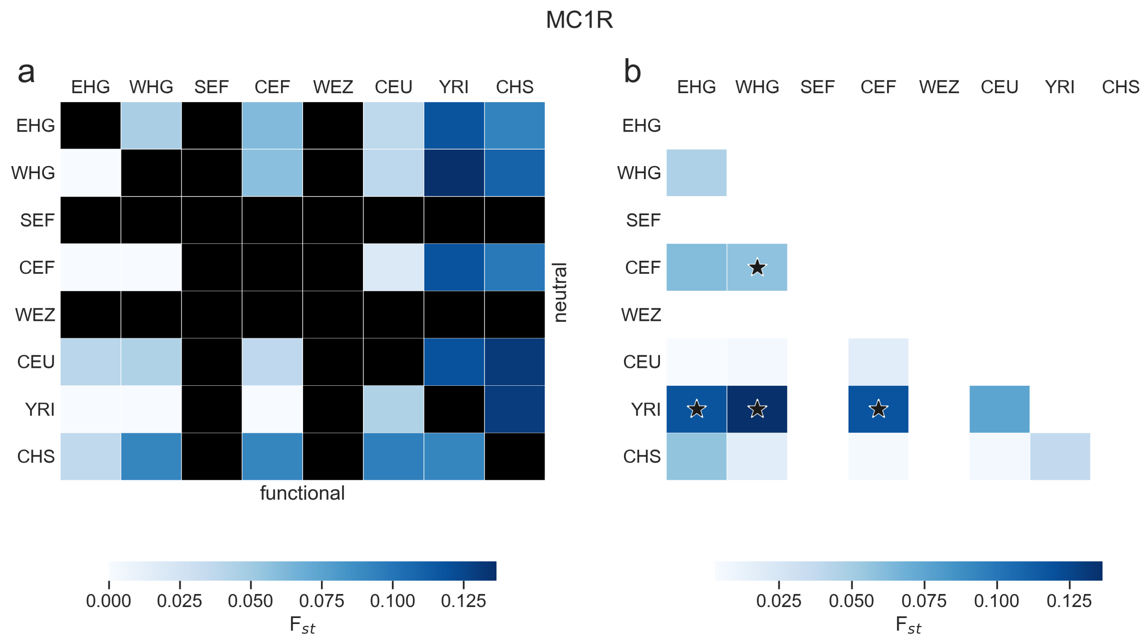

Figure (22) *a) Pairwise F_{st} comparison between SNPs in the ASIP gene and SNPs randomly drawn from the neutralome. b) Absolute differences between F_{st} values. Significant values are highlighted by a black star.*

When coordinates were assumed for each groups location of origin (as specified by the 1000Genomes project), the latitude corresponded almost linearly with the mean values of the pigmentations scores of each group. The score increases with distance to the equator (R=0.942, $p<0.05$, table: 6). The opposite could be observed for the variance within each group, with a negative Pearson's R of -0.939 ($p<0.05$).

In the ancient sample groups, the Central European hunter-gatherers produced the lowest mean score (23.37±7.06), followed closely by the Eastern European/Russian group (23.84±5.33). Both early farmer groups had higher mean scores (SEF: 30.50±3.30, CEF: 27.57±5.20) compared with the hunter-gatherer groups, with the Central European farmers having a lower mean score compared to the farmers from the Aegean/Balkan region (see table 6). In the Middle Neolithic group, the scores were increased (29.96±4.01) compared to the Central European farmer group, but lower compared to the mean values calculated for farmers from the Aegean/Balkan region. The highest scores for any ancient sample group was calculated for the Bronze Age group (35.02±2.33), with a mean score falling between the range of reference samples IBS, TSI on the lower end, and CEU, GBR and FIN on the higher end (table: 6, figure: 23).

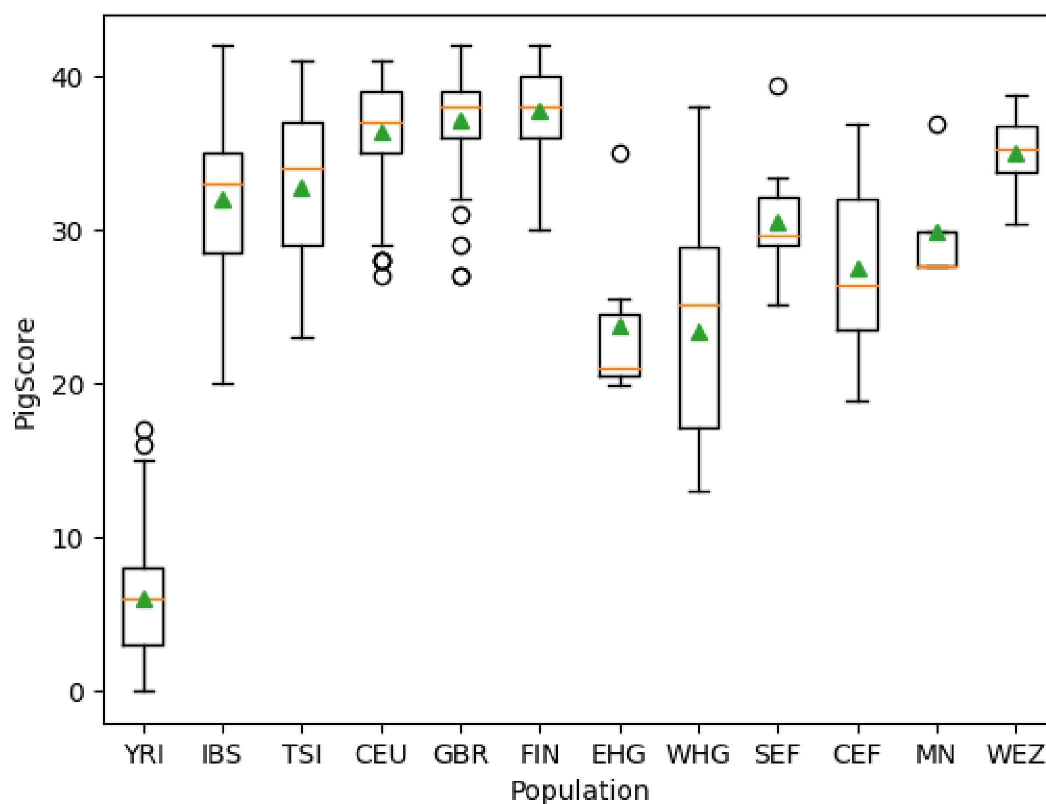

Figure (23) *Pigmentation score for ancient samples and reference samples from the 1000Genomes project. Orange line = median score, green triangle = mean score.*

Table (6) *Mean and median Pigmentation score results with variance for ancient samples and reference.*

Group ID	Median	Mean	Var	Lat.	Long.	
EHG	21.000	23.841	28.451	-	-	Ancient
WHG	25.200	23.371	49.892	-	-	
SEF	29.636	30.497	10.870	-	-	
CEF	26.442	27.568	27.016	-	-	
MN	27.659	29.964	16.076	-	-	
WEZ	35.318	35.024	5.436	-	-	
YRI	6.000	6.020	13.920	-	-	Africa
IBS	33.000	32.019	20.972	40.291	-3.193	Europe
TSI	34.000	32.794	22.911	43.458	11.207	
CEU	37.000	36.455	12.915	52.364	6.211	
GBR	38.000	37.165	9.237	53.475	-1.389	
FIN	38.000	37.818	6.714	63.558	26.516	
R	0.914	0.942	-0.939			Correlation
P	0.030	0.017	0.018			w. latitude

3.3.6 Phenotypic predictions with the Hirisplex webtool

From the eight samples in the Eastern-European hunter-gatherer group, two did not yield enough data for a prediction (Min5 and Min8). Results for the other samples were uniform: Dark hair-color in combination with brown eyes was estimated to be the most possible phenotype for all samples. This is in contrast to the Central European hunter-gatherers. In this group, 14 of the 25 samples had sufficient data for a prediction of eye-color and hair-shade. For six of those samples, a hair-color estimation was possible. Five out of 14 samples were most likely blue-eyed and had a dark hair-shade. Three samples were estimated to have had a light/blond hair-shade with most likely brown eyes. The rest of the samples was most likely brown eyed and dark haired.

In the early farmer group from the Aegean/Balkan region, hair-color could only be predicted for one sample out of 14. Four samples did not yield enough data for any prediction while the rest had sufficient data for eye-color and hair-shade predictions. All of these ten samples were most likely brown-eyed. Three of the ten samples had a high probability of being light haired while the rest was most likely dark haired.

Eye color could be predicted for all samples in the Central European farmer group, while hair-color could be predicted for all but three samples. Only four of the 28 samples were likely to have had a lighter hair shade, all other samples were estimated to have been dark haired. All but four samples were have probably been brown-eyed. The four samples that were probably blue-eyed had also high probabilities for a light hair shade and/or color. All other samples were most likely dark-haired.

From the Middle Neolithic group, only four samples yielded enough data for eye-color and hair shade predictions. For two samples, a hair-color could also be estimated. All samples had high probabilities of having been brown eyed. Two samples were estimated to have been light haired, while the other two had most likely black hair.

For the Bronze Age group, predictions for nine samples were possible. Only one sample was assumed to have had dark hair, while the rest had higher probabilities for being light haired. Three samples were predicted to have been blue eyed, while the rest had higher probabilities for brown eyes. The sample that was most likely dark haired was also estimated to have had brown eyes as well (see table 7).

Table (7) *Results for phenotypic reconstructions from the HirisPlex webtool. Values in cells report the probabilities for each phenotype. Only the highest probabilities were reported. The color scheme in the cells corresponds to the predicted phenotypes.*

Sample ID	Eye color	Hair color	Hair shade	Group
Zv317	1.00	0.88	0.99	EHG
Min10	0.98	0.70	0.99	EHG
Min11	0.98	0.69	1.00	EHG
Min2	0.99	0.90	1.00	EHG

Table 7 continued from previous page

Sample ID	Eye color	Hair color	Hair shade	Group
Min3	0.99	0.85	1.00	EHG
Lec2	NA	0.49	0.95	EHG
Bichon	0.78	0.44	0.75	WHG
Bla20	0.72	NA	0.97	WHG
Gr1	0.59	NA	0.66	WHG
Gr2	0.67	0.55	0.94	WHG
GrO1	0.80	0.60	0.92	WHG
LEPE18	0.74	NA	0.83	WHG
LEPE52	0.74	NA	0.82	WHG
Losch	0.71	0.61	0.97	WHG
VLASA10	0.95	NA	0.95	WHG
VLASA32	0.95	NA	0.91	WHG
VLASA41	0.98	0.71	0.99	WHG
VLASA44	0.95	NA	0.91	WHG
VLASA4	0.96	NA	0.95	WHG
Zv162	0.89	0.77	1.00	WHG
AKT16	0.89	NA	0.60	SEF
AKT20g	0.74	NA	0.82	SEF
AKT26g	0.97	NA	0.79	SEF
AKT6g	0.96	NA	0.79	SEF
BAR11	0.89	NA	0.68	SEF
BAR15	0.99	NA	0.91	SEF
BAR20	0.74	NA	0.74	SEF
BAR32	0.71	NA	0.66	SEF
Bar8	0.99	NA	0.91	SEF
LEPE53	0.60	0.68	0.57	SEF
281-19-6	0.99	0.63	0.93	CEF
282-104-4	1.00	0.73	0.98	CEF
282-126-16	1.00	0.75	0.98	CEF
282-126-7	0.94	0.52	0.83	CEF
282-13-7	0.99	NA	0.99	CEF
282-23-1	0.92	0.45	0.74	CEF
282-88-2	0.85	0.76	0.97	CEF
282-94-11	1.00	0.83	0.99	CEF
7034	0.99	0.62	0.94	CEF
Asp10	0.61	NA	0.61	CEF

Table 7 continued from previous page

Sample ID	Eye color	Hair color	Hair shade	Group
Asp1	0.97	0.45	0.66	CEF
Asp2	0.99	0.54	0.96	CEF
Asp3	0.89	0.49	0.62	CEF
Asp4	0.81	0.43	0.74	CEF
Asp6	1.00	0.78	0.98	CEF
Asp8	0.88	0.49	0.58	CEF
Dil15	0.68	0.53	0.84	CEF
Dil16	0.94	0.47	0.73	CEF
Ess7	1.00	0.88	0.99	CEF
Klein10	0.92	0.44	0.61	CEF
Klein1	0.74	0.43	0.56	CEF
Klein2	0.91	NA	0.96	CEF
Klein3	0.97	0.60	0.84	CEF
Klein4	1.00	0.83	0.99	CEF
Klein5	0.90	0.70	0.99	CEF
Klein8	0.92	0.52	0.77	CEF
LBK	0.99	0.75	0.99	CEF
NE1	0.90	NA	0.96	CEF
Bla32	0.95	0.64	0.99	MN
Bla59	0.87	0.72	1.00	MN
Bla10	0.67	NA	0.63	MN
Bla13	0.73	NA	0.68	MN
Wez51	0.86	0.74	0.98	WEZ
WEZ53	0.68	NA	0.65	WEZ
WEZ54	0.60	NA	0.81	WEZ
WEZ56	0.45	NA	0.89	WEZ
WEZ57	0.68	NA	0.72	WEZ
WEZ64	0.68	NA	0.62	WEZ
WEZ71	0.91	NA	0.96	WEZ
WEZ74	0.75	NA	0.69	WEZ
Wez83	0.79	NA	0.56	WEZ

3.3.7 Beta-binomial simulation

All changes in frequencies for the pigmentation related SNPs could be explained by drift during the simulations (see figure: 24).

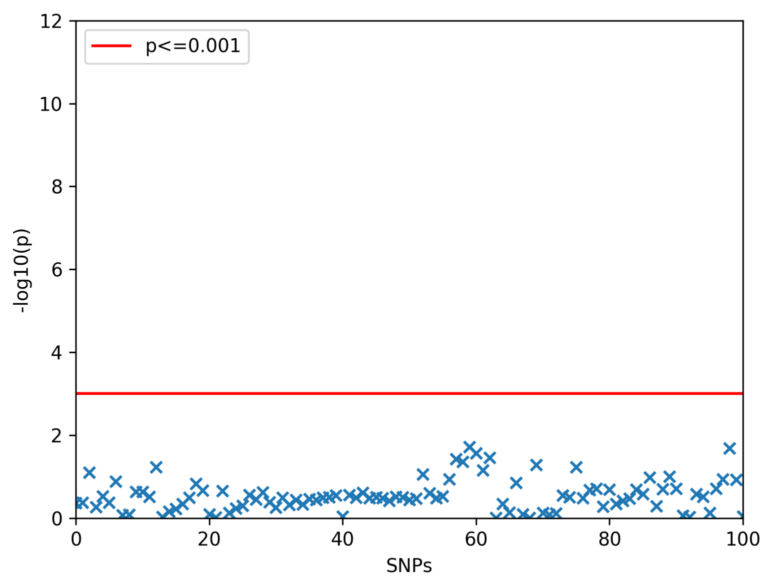

Figure (24) *Negative log_{10} p-values for beta-binomial simulation results for the pigmentation related SNPs at initial populations size of 1456 individuals for a selection coefficient of S=0. Threshold of $p \leq 0.001$ is indicated by red line.*

3.4 Discussion

Pigmentation in humans is a highly variable trait that differs between populations. The highest pigmentation rate is found in populations living near the equator, from where it declines with increasing latitude [Canfield et al., 2013]. Differences in pigmentation are mainly caused by the ratio of eumelanin to pheomelanin and their cumulative amount. From the genes regulating the synthesis and distribution of melanin, several SNPs are known to have an effect on the pigmentation phenotype. The impact of selection on these genes during human evolution, likely contributed to the differentiation between populations [Sturm and Duffy, 2012]. This chapter investigates the development of the human pigmentation phenotype in European, focusing on skin, hair and eye pigmentation. Neutrality tests were applied to a pre-selected set of relevant markers using an extensive data set of three modern reference populations from Africa, Central Europe and East Asia. Allele frequency data from six ancient European sample groups, dating from the Mesolithic to the Bronze Age, was used to track changes in frequencies through time. In addition, genotype data from the ancient samples was used for phenotypic reconstructions. By applying these methods to this hetero-chronic data set, strong evidence for selection on different genes associated with pigmentation, could be found.

The results show that the pigmentation lightening in Central Europe was a gradual process. Over the course of time several alleles in associated genes experienced changes in frequencies. When alleles in a gene had reached high frequencies, derived alleles in other genes began to increase as well. Since anatomical modern humans are assumed to originate in Africa, their ancestral pigmentation phenotype is believed to be dark, similar to modern day populations living near the equator [Jablonski and Chaplin, 2000]. The true state of pigmentation can only be assumed, since even for sub-Saharan African populations variation in pigmentation phenotypes are known [Crawford et al., 2017].

The results indicate that depigmentation was already advanced during the Mesolithic in Europe. The pigmentation-scores show that an increased number of derived alleles were already present in both hunter-gatherer groups of the data set (EHG: 23.84±5.33, WHG: 23.37±7.06) compared to the Africa reference group (6.02±3.73). The scores further show, that derived alleles were present at an even higher level in the early Neolithic (SEF: 30.50±3.30, CEF: 27.57±5.20) and later groups (MN: 29.96±4.01, WEZ: 35.02v2.33). In addition, the variance in the groups decreased over time, indicating that the range of pigmentation phenotypes became more homogeneous. The mean pigmentation score in the Bronze Age group was similar to the European reference groups (36.45±3.59), suggesting that the level of depigmentation reached during the Bronze Age, was probably very similar to modern-day Central European populations. However, while the pigmentation-score results cannot be translated to a direct phenotype, a correlation with latitude can be found for the modern-day European populations (R:0.9420, $p \leq 0.0500$), similar to the correlation found for pigmentation in Europe [Deng and Xu, 2018].

Data indicated that a substantial amount of variation at pigmentation-related loci was and still is

influenced by selection in all three reference populations and their ancestors [Lao et al., 2007, Rees and Harding, 2012, Yang et al., 2018]. Comparisons of the distributions of allele frequencies in the three reference groups showed that the frequencies of the pigmentation-related SNPs were differently distributed for all groups compared to frequencies of functionally neutral SNPs (Mann-Whitney-U: CEU - $p \leq 0.0005$, YRI - $p \leq 0.0001$, CHS - $p \leq 0.0500$). These differences are unexpected under the assumption of neutrality for the pigmentation related SNPs. Furthermore, the generally lower number of derived alleles found in the African population indicates that a functional constraint prevents new mutations from segregating, confirming studies by [Hudjashov et al., 2013]. Neutrality tests applied to the Central European reference group identified SNPs that deviated from neutral expectations in 10 out of 14 pigmentation-related genes. In comparison, the number of SNPs identified as non-neutral was lower in both non-European reference groups. This result can be explained by the design of the capture array, which was specifically designed to investigate SNPs under potential selection in European populations. In populations with different ethnic backgrounds the genetic variation that influences pigmentation phenotypes can differ [Deng and Xu, 2018].

Early selective sweeps affecting the pigmentation phenotype

The combination of results for SNPs in the genes ASIP, IRF4, and KITLG show that derived alleles at these loci increased in frequency before the Mesolithic in Europe. Changes in ASIP probably occurred earlier, compared to the SNPs in IRF4 and KITLG, which display a unique frequency distribution in Europe.

For the SNP rs6119471 in the ASIP gene, which has a strong effect on pigmentation, changes in frequency likely occurred even before the ancestors of East Asians and Central Europeans became separate populations over 40,000 years ago. Non-neutrality was found for the SNP by a high H12 value (0.5391 $P_{98.8110}$) in the Central European reference group. Derived alleles occurred at low to intermediate frequencies in the African reference group (0.27 ± 0.06), while they were fixed in both the Central European and the Chinese reference groups. No copies of the ancestral allele could be found in any of the ancient sample groups. This suggests that the derived allele reached fixation early, probably due to positive selection. The ASIP gene functions as an antagonist for the MC1R gene and is involved in eumelanogenesis by down-regulating the expression of the tyrosinase gene [Scherer and Kumar, 2010]. The SNP is used in skin color predictions to discriminate between a dark and non-dark pigmentation phenotype [Walsh et al., 2017, Spichenok et al., 2011]. For individuals who are homozygous for the G allele a dark skin tone is assumed [Hart et al., 2013]. Even though a certain variance in pigmentation phenotypes is expected for the ancient samples based on the allele frequency data and the Pigmentation score results, the absence of the G allele in all ancient samples groups indicates that none of the individuals had a very dark skin tone, similar to modern-day sub-Saharan populations.

Since the derived allele was fixed in all non-African groups, an early onset of depigmentation can be assumed. The fact, that the SNP is polymorphic in the African reference group, could indicate that selection affected a standing variation after anatomically modern humans migrated out of Africa,

rather than on a de-novo mutation. Although the derived allele also reached fixation in the Chinese population, no deviations from neutrality were found. If the variant reached fixation during an early time, as the data suggests, the time may have been long enough to reintroduce diversity around the locus in the Chinese group.

Results similar to rs6119471 were found for SNPs in the KITLG gene. Two of the three SNPs in the gene that is associated with skin and hair pigmentation, were found to deviate from neutrality in all three reference groups. While rs4590952 occurred at high frequency in the African reference group (0.74±0.06), it was present at lower frequencies in both non-African reference groups (CEU: 0.21±0.06 , CHS: 0.20±0.05). For rs642742 an inverted pattern was found, with a lower frequency in the African group (0.06±0.03) and a higher frequency in both non-African groups (CEU: 0.82±0.05, CHS: 0.80±0.05). These results indicate that different forms of selection affected the alleles in the three populations. Previous work suggests, that a selective sweep affected alleles associated with depigmentation in the KITLG gene in European populations, starting approximately 32kya [Beleza et al., 2012]. Estimations place the split between East Asian and Central European populations between 40kya to 80kya [Fu et al., 2016]. Based on this data, the results could originate from a convergent evolution in the Central European and Chinese reference groups. This is further emphasized by the differences found in frequencies between two of the ancient sample groups and the Central European reference. For rs4590952, a higher frequency was found in the Central European hunter-gatherers (0.50±0.26). In contrast, a lower frequency for rs642742 was found in Central European farmers (0.57±0.15). These results show, that even though the frequencies between the Chinese and the Central European reference population are highly similar, they varied in previous populations over the last 8,000 years in Europe. Population stratification, caused by migration or population growth or decline, can have a strong effect on allele frequencies that are not fixed. In addition, the sample size of the groups (16.67 sample on average) also has to be taken into account. Interestingly, an association with thermogenesis via brown adipose tissue is also known for variants in the KITLG gene. Previous research suggests that these variants were also under selection approximately 30-58ky ago [Yang and Fu, 2018]. It is therefore possible that selection affected variation in KITLG at different times for based on different pressures.

Another SNP that did not display any significant differences in frequency towards the Central European reference group since the Mesolithic was rs1540771. For this SNP, extremely low haplotype frequencies in the gene interferon regulatory factor 4 (IRF4) were found, as indicated by H12 values (0.0179, $P_{1.1311}$). The derived allele was found at low frequency in the African reference group (0.07±0.04), at intermediate frequency in the Chinese (0.24±0.06), and at elevated frequency in the Central European reference group (0.55±0.07). The derived allele at the SNP is associated with a decreased tanning response as well as the presence of freckles in people of European descent [Han et al., 2008]. Previous research suggested that the SNP may have been under positive selection in European populations [Sulem et al., 2007]. This could explain the results found in the data set, as extreme low H12 values are the result of high haplotype diversity [Ferrer-Admetlla et al., 2014]. This high diversity could be explained by diversifying or balancing selection at the locus. In combination,

these results indicate that selection started to affected pigmentation related loci soon after humans arrived at higher latitudes. Selection apparently intensified as humans migrated further north. The need to synthesize vitamin D was probably further increased after humans started to live as farmers [Beleza et al., 2012].

Selection on depigmentation in Europe since the Mesolithic

For the majority of SNPs identified as non-neutral in the Central European reference group, differences in allele frequencies could be found among the ancient sample groups. Changes in diet that are thought to have accompanied the Neolithic transition likely intensified the selective pressure on the pigmentation phenotype [Beleza et al., 2012]. Research suggests that the diet of early farmers was substantially depleted in micro-nutrients compared to contemporary hunter-gatherers. The increased amount of vitamin D consumed by hunter-gatherers is believed to have lessened the selective pressure caused by the lowered UV radiation [Sturm and Duffy, 2012, Hlusko et al., 2018]. Therefore, hunter-gatherers could maintain a darker complexion even at higher latitudes in comparison to the early farmers [Marciniak and Perry, 2017].

One of the genes that was probably affected early after the adoption of agriculture is SLC24A5. Differences between the hunter-gatherers of Central Europe and early Neolithic and later groups were especially prominent in this gene. For the majority of variants under investigation in this data set, the derived allele had reached fixation in the Central European reference group, which in turn resulted in a low gene diversity (D_e: 0.0257±0.0034). Even though diversity was significantly higher in all ancient sample groups ($p \leq 0.0500$), only the Central European hunter-gatherers had a functional F_{st} that was significantly higher compared to neutral expectations compared to the Central European reference (diff: -0.3977, $p \leq 0.0001$). Deviations from neutrality were detected for all variants in the SLC24A5 gene in the Central European reference group (between H12: 0.6959, $P_{99.8216}$ - H12: 0.827, $P_{99.9825}$), confirming previous evidence of selection at this locus [Hancock et al., 2011, Sturm and Duffy, 2012, Beleza et al., 2012]. It is assumed that while each variant adds variable amounts to the overall impact, the SNP rs1426654 has the strongest effect and can explain 25-38% of variation in eumelanin levels between populations of European and African descent [Lamason et al., 2005]. Frequencies for this variant were already at modern-day levels in both early farmer groups in the data set (SEF: 0.96±0.08, CEF: 0.96±0.05, CEU: 1.00±0.02), but low in the Central European hunter-gatherers (0.19±0.15), confirming previous findings [Gamba et al., 2014, Lazaridis et al., 2014, Hofmanová et al., 2016]. The adoption of agriculture is assumed to be accompanied by a diet depleted in micro-nutrients and vitamin D [Larsen, 2006, Carrera-Bastos et al., 2011]. It is highly likely that this increased the selective pressure already present for the pigmentation phenotype. Selection of rs1426654 is estimated to have started between 19,000 and 11,000 years ago [Beleza et al., 2012]. It is therefore possible that selection predates the onset of farming approximately 12,000 years ago. This could be further investigated if allele frequency data from hunter-gatherers ancestral to the early farmer populations were available.

However, the frequencies of the derived allele at rs1426654 were also high in the hunter-gatherers

from Eastern Europe/Russia. This might be explained by their assumed habitat being located at higher latitude than that of the Central European hunter-gatherers. It is possible that their diet's vitamin D content was not high enough to compensate the effect of the lowered UV radiation. In addition, demographic factors need to be taken into account. The age of the derived alleles in SLC24A5 is assumed to be old. For several variants in SLC24A5, derived alleles are also present in the African reference group and are known to be polymorphic in other African populations [Crawford et al., 2017]. A decline in population sizes was estimated for the last glacial maximum, followed by a re-expansion during the Mesolithic and later times [Tallavaara et al., 2015]. This could have led to a differential distribution of alleles in the population ancestral to the hunter-gatherers of Central and Eastern Europe. If the derived alleles were already at elevated or high frequencies in the Eastern European/Russian hunter-gatherers, drift could have increased them further in a short time span, depending on the population size. This process could furthermore have been accelerated by selection.

Even though depigmentation was probably already advanced in early farmers arriving in Central Europe, the allele frequencies found in this study show that a darker phenotype compared to modern-day populations of the area can be assumed. Between the groups of the Early Neolithic and the Bronze Age several differences in frequencies were found. These differences included SNPs in the genes SLC45A2 and TYR. Both genes are assumed to have undergone selection during the course of the Neolithic in Europe [Wilde et al., 2014].

All of the six SLC45A2 SNPs in the Central European reference data set occurred at high frequency, close to fixation, which resulted in a low diversity (D_e:0.1023±0.0078). In comparison, the diversity was significantly higher in all ancient sample groups (between WHG - D_e: 0.3029±0.0422 and SEF - D_e: 0.4178±0.0390, $p \leq 0.0005$). The SNPs in the SLC45A2 gene were the only SNPs where the frequencies in the Bronze Age group differed significantly from the Central European reference. F_{st} comparisons further show that the functional F_{st} at the SLC45A2 locus was higher in all ancient sample groups compared to neutral F_{st} values in comparison to the Central European reference group (between WHG - diff: -0.6341, $p \leq 0.0001$ and WEZ - diff: -0.0910, $p \leq 0.0001$). In addition to these differences, deviations from neutrality were found for all six SNPs in the Central European reference group (D: -2.5516, $P_{0.0380}$, H12: 0.7474, $P_{99.9848}$, H1_H2: 0.0102, $P_{0.3124}$). These findings confirm previous reports of strong positive selection on the gene in Europe [Hancock et al., 2011, Wilde et al., 2014, Mathieson et al., 2015]. The differences in frequencies and diversity between the Bronze Age group and the Central European reference indicate that selection may have been ongoing at these loci beyond the Bronze Age, as suggested for pigmentation in previous studies [Field et al., 2016]. However, this could not be reinforced by the beta-binomial simulations. The strength of selection on rs16891982 in SLC45A2 is assumed to have been similar compared to rs1426654 in the SLC24A5 [Hart et al., 2013].

In contrast to variants in the SLC24A5 gene, the haplotype structure around rs16891982 suggests an European origin with subsequent selection [Beleza et al., 2012]. Thus, the data in this study could confirm the assumption, that the timing of selection was later than for variants in SLC24A5 [Sturm

and Duffy, 2012, Gamba et al., 2014]. This stepwise process of selection on pigmentation genes could result from a persistent selective pressure, originating from the combination of an agriculturalists diet and the lowered levels of UV radiation in Central Europe. When pigmentation related variants in genes such as SLC24A5 reached high frequencies, selection apparently started to affect variants in other genes, such as SLC45A2 or TYR.

Another variant that increased in frequency over the course of the Neolithic was rs1042602 in the TYR gene. The derived allele is associated with light pigmentation and the presence of freckles [Sulem et al., 2007]. In the data set the derived allele first appeared in the early farmers of Central Europe (0.11 ± 0.10) and increased in frequency towards the Bronze Age group (0.33 ± 0.28). It was found at an intermediate frequency in the Central European reference group (0.40 ± 0.02), but was absent in the all non-European reference groups. The absence of derived alleles in both hunter-gatherer groups suggests that the mutation first occurred in a farmer population. Estimations of the allele's age between 6,100 and 15,600 years are consistent with an origin of the derived allele in a farming population or its immediate predecessors. Subsequent selection in Europe was indicated by high iHS and nSL scores in the Central European reference group (iHS: 2.2912, $P_{98.8646}$; nSL: 2.6302, $P_{99.6343}$). Selection on variants in the TYR gene was reported previously for a time-frame similar to the presented data [Wilde et al., 2014]. In contrast to the SNPs in SLC45A2, variation in the TYR gene had already reached modern-day levels in the Bronze Age group.
Hudjashov et al. [2013] theorized that the relaxation of a functional constraint enabled mutations in the TYR gene after humans migrated out of Africa. The results for the African group where deviations from neutrality (iHS: 2.2912, $P_{98.8646}$; nSL: 2.6302, $P_{99.6343}$) in combination with the absence of derived alleles were detected furthermore indicate a functional constraint in Africa on variants in the TYR gene. This pattern could result from purifying selection, purging alleles associated with a lighter phenotype from the population.

The development of eye and hair color phenotypes in Europe

Similar to pigmentation of the skin, hair and eye color are also defined by the amount and type of melanin present in the cells [Sulem et al., 2007]. For eye color, the majority of variation in Europe can be attributed to mutations in the OCA2/HERC2 genes [Scherer and Kumar, 2010]. The HERC2 gene is located in close proximity to the OCA2 gene and contains a promoter region that influences the expression of OCA2 [Ainger et al., 2017]. The majority of variation of eye color in Europeans is explained by variants in the OCA2/HERC2 genes. In addition, variants in other genes, such as rs1408799 in the TYRP1 [Sturm and Duffy, 2012] or rs2402130 in the gene SLC24A4 [Sulem et al., 2007] are also influential. Several studies suggest that light hair and eye color were under selection in Europe over the last 10,000 years [Sturm and Duffy, 2012, Wilde et al., 2014, Günther et al., 2018]. Many of these variants are also known for having an influence on hair color too [Sturm and Duffy, 2012].

Previous work suggests that hunter-gatherers of Central Europe had a unique appearance, with a darker skin tone, and for some individuals, a light hair and eye color [Mathieson et al., 2015]. This

was confirmed for the Central European hunter-gatherers of this data set. The hirisplex results indicate that blue eye color was a common trait and while light hair shades were present, blond hair color was only found once. This is a clear contrast to the early farmers of the Aegean/Balkan region, where the data indicated only dark eyed phenotypes with few lighter hair shades. In the early farmers of Central Europe light eye and hair color both occurred; interestingly only in the same individuals. It can be argued if these phenotypes in the early farmers were introduced by admixture with contemporary hunter-gatherers during the migration towards Central Europe [Lipson et al., 2017] or if they are the result of population stratification. Selection may have also played a role in increasing these phenotypes as indicated by the data for several SNPs in the genes OCA2, HERC2 as well and TYRP1. By the time of the late Bronze Age, 3,000 years ago, light eye and hair color was fairly common as seen in the data from the Tollense Battlefield presented in this study.

The capture array contained 37 SNPs located in the OCA2/HERC2 gene complex. For the 16 SNPs in the OCA2 gene an intermediate diversity was found in the Central European reference group (D_e: 0.2340 ± 0.0084). The diversity in all ancient sample groups was similar, except for the early farmers from the Aegean/Balkan region (D_e: 0.3549 ± 0.0334, $p \leq 0.0050$) and the Middle Neolithic group (D_e: 0.346 ± 0.0354, $p \leq 0.0500$) where a higher diversity was found. For five out of 15 SNPs deviation from neutrality were found in the Central European reference group. The results indicate, that different forms of selection affected these alleles during different times. For two SNPs high H12 (rs1800401 - H12: 0.765, $P_{99.9340}$; rs4778136 - H12: 0.7476, $P_{99.9154}$) and low H2_H1 values (rs1800401 - H2_H1 : 0.0142, $P_{0.8291}$; rs4778136 - H2_H1 :0.0149, $P_{0.9054}$) indicate directional selection. The frequencies for the SNPs in the Central European reference group were high for rs4778136 (0.93 ± 0.04), and low for rs1800401 (0.05 ± 0.03). No significant differences were found for any ancient sample group in comparison to the Central European references. Therefore, it can be assumed that frequencies for the two SNPs were constant over the last 8,000 years in Europe. This indicates that selection at these loci happened before the Mesolithic. Selection in hair and eye pigmentation was previously suggested for the Mesolithic and earlier in Europe [Günther et al., 2018, Beleza et al., 2012].

Similar results were found for two different SNPs, only here the neutrality tests were consistent with diversifying of balancing selection (rs7495174 - D: -2.1901, $P_{1.0415}$; rs1545397 - H12: 0.0143, $P_{0.7157}$; H2_H1: 0.8621, $P_{99.7154}$). The frequencies for the two SNPs were also similarly distributed in the Central European reference (rs7495174 - 0.94 ± 0.03, rs1545397 - 0.08 ± 0.04) and no significant differences could be found for any ancient sample group in comparison. Apparently, two different forms of selection led to a similar outcome in allele frequencies which were maintained over the last 8,000 years.

The only SNP that deviated from neutrality and also differed in two of the ancient sample groups was rs4778241. The SNP was at high frequency in the Central European reference (0.84 ± 0.05) while at intermediate frequency in the early farmers of Central Euorpe (0.50 ± 0.21) and the Eastern European hunter-gatherers (0.50 ± 0.30). The early farmer group of Central Europea was also the only group were a F_{st} value was found for the OCA2 SNPs towards the Central European reference that

was significantly higher compared to F_{st} values across neutral regions (diff: -0.0202, $p \leq 0.0500$). The SNP is located on a 166kb long haplotype described by Eiberg et al. [2008], alongside several SNPs in the HERC2 gene. The haplotype (H1) is one of four (out of ten) haplotypes associated with light hair and eye color. For 19 of the 21 HERC2 SNPs in the capture array high negative Tajima's D values were found in the Central European reference group (between rs1129038 - D: -2.069, $P_{1.7366}$ and rs7494942 - D: -2.2876, $P_{0.6043}$). This inclueded all of the SNPs located on the H1 haplotype.

The frequencies of the HERC2 SNPs in the Central European were high for most alleles, which resulted in a lowered diversity (D_e: 0.1896±0.0074). For all ancient sample groups the diversity was significantly higher in comparison (between MN - D_e: 0.3633±0.0299 and WEZ - D_e: 0.2507±0.0258, $p \leq 0.0050$). F_{st} comparisons show that for both early farmer groups and the Eastern European hunter-gatherers the F_{st} across the functional markers was higher then expected under neutrality (EHG - diff: -0.1648, $p \leq 0.0001$; SEF - diff: -0.0691, $p \leq 0.0001$; CEF - diff: -0.0522, $p \leq 0.0001$). These results show that in contrast to the majority of the OCA2 SNPs, selection on HERC2 was more recent as frequencies were affected since the early Neolithic, over 8,000 years ago and the diversity was still higher in the late Bronze Age. Selection on HERC2 was previously reported for a similar time frame [Wilde et al., 2014]

In addition to the OCA2 and HERC2 SNPs, three SNPs in the TYRP1 gene (rs1408799, rs10960751, rs10960752) showed extreme iHS values in the Central European reference group. The SNP rs10960751 had one of the most extreme iHS results of all SNPs in the capture array (iHS: -4.2692, $P_{0.0056}$). Interestingly, a similar extreme value for the SNP was found in the African reference group (iHS: -4.2692, $P_{0.0056}$), while the SNP was at elevated frequency in the Central European reference (0.63±0.07), but at low frequency in the African reference group (0.03±0.06). The majority of the SNPs in the TYRP1 gene were at similar frequencies in the Central European reference group, which resulted in a high diversity (D_e: 0.3447±0.0080) that was matched in all ancient sample groups (between WHG - D_e: 0.3392±0.0300 and SEF - D_e: 0.4125±0.0303). These results, in combination with the iHS values indicate that balancing selection affected the SNPs. Selection on TYRP1 was previously reported [Alonso et al., 2008] with and estimated start of the sweep between 19,000 and 11,000 years ago in Europe [Beleza et al., 2012]. These estimates are consistent with the results found in this data set. No significant diffrences in allele frequencies could be found in any ancient sample groups when compared to the Central European reference group, except for the early farmers of Central Europe. This led to a significant higher F_{st} value for the TYRP1 SNPs, compared to neutral SNPs (diff: -0.0369, $p \leq 0.0018$) towards the Central European reference. This shows that the TYRP1 SNPs that seemed to be under selection, were already at frequencies similar to those found today in Europe, over 7,500 years ago, before the first farmers arrived.

Interestingly, differences were only found for the Central European farmers, but not for the early farmers of Aegan/Balkan region. The farmers of the Aegean/Balkan regions are closely related to the ancestral populations that migrated north and became the first farmers of Central Europe [Hofmanová et al., 2016]. These differences may have originated due to population stratification

during the migration. Loss of diversity during migration due to serial found effects or increased drift in smaller populations are factors that can have a severe influence on allele frequencies [Hamilton, 2011]. Apparently, due to admixture or selection in Central Europe during the Neolithic, the frequencies for the SNPs in the TYRP1 gene increased over time. All three TYRP1 SNPs indicated as selected by the results have strong associations with eye and skin color in European populations [Sulem et al., 2007, Alonso et al., 2008, Zhang et al., 2013].

Depigmentaion of the skin is widely attributed to the need to synthesize vitamin D at higher latitudes and thereby correlates with latitude [Jablonski and Chaplin, 2000]. This correlation is not equally strong for eye and hair color, although the highest phenotypic variation is found in populations with European ancestry [Sturm and Larsson, 2009]. In addition, the architecture of hair and eye color phenotypes is more complex in comparison to pigmentation of the skin. The majority of variants associated with skin depigmentaion is assumed to be additive. Several variants with varying small effects are known [Crawford et al., 2017]. For eye color, over-dominance of alleles associated with dark pigmentation phenotypes is assumed for a vast number of SNPs [Ainger et al., 2017]. No phenotypic advantageous are known for light hair and eye color, which makes the identification of potential selective pressures even harder. Different theories were proposed to explain the signs of selection found for these traits.

Since a high number of SNPs associated with eye and hair color are also known to affect skin pigmentation [Crawford et al., 2017] a correlation was assumed [Sulem et al., 2007]. It was suggested that hair and eye color were a by product of skin depigmentaion and thereby also influenced by the need for vitamin D. This is contrasted by the often spotty distribution of eye and hair color phenotypes across Europe, for which only a weaker correlation with latitude can be found [Sturm and Duffy, 2012]. Others suggested that demographic factors were the major drivers of diversification for these phenotypes [Jablonski and Chaplin, 2017]. It was assumed that the relaxation of the functional constraint on pigmentation phenotypes outside of Africa allowed variants to segregate. Since these variants were neutral they could rise in frequencies and were subsequently shaped by drift and demographic history. This is contrasted by the clear signs of selection found in this and in previous studies [Wilde et al., 2014, Mathieson et al., 2015, Sturm and Duffy, 2012].

Frost [2006] suggested that the diversity of eye and hair color in European population can be explained by sexual selection. The theory is based on the assumption that hunter-gatherer men had a higher mortality, caused by the dangers of hunting. Based on their scarcity, men could assert a higher mate preference. This led to the selection of phenotypes that had a signaling factor and stood out of the mass. The assumed scarcity of men was less severe in latitudes closer to the equator, since food sources were more abundant. In addition, signaling of group affiliation was also assumed. However, these conditions would have not persisted after the adaption of agriculture, as the need to hunt was moderated by food production. This stands in contrast to the results present here. Several alleles in HERC2 and TYRP1, indicated as selected by the neutrality tests results, increased in frequency over the last 7,500 years since the Neolithic arrived in Central Europe.

It is possible that the variants associated with light eye and hair color are related to additional phenotypes. Variants in the gene MC1R have a strong association with red hair and light eyes, in addition to increase resistance to cold-stress related pain [Liem et al., 2005]. The exact function of the HERC2 gene is still unknown [Ibarrola-Villava et al., 2010]. Even though it is closely related to pigmentation phenotypes, it was suggested that the HERC2 gene is also involved in spermatogenesis, among other functions [Jablonski and Chaplin, 2000]. For the OCA2 gene a connection to neurodevelopment via an association to its neighboring gene GABRB3 was proposed [Delahanty et al., 2016]. The GABRB3 gene is assumed to affect cognitive abilities and has association to autism and bipolar disorder [Delong, 2007]. Additional research is needed to identify the underlying mechanism that led to the distribution of pigmentation phenotypes in European populations. With a broader understanding of gene function and their interactions, it may be possible to shed further light on the development of hair and eye color related phenotypes.

Conclusion

Pigmentation in humans is a polygenic trait. Variation in many genes is known to affect skin hair and eye color phenotypes to a varying degree. Vitamin D deficiencies at higher latitudes are assumed to have led to a decreased pigmentation in Central Europe over the last 40,000 years. This chapter investigates the development of pigmentation phenotypes in Europe by applying statistical tests to a diachronic data set consisting of over 100 ancient samples ranging from the Mesolithic to the Bronze Age. By using neutrality tests on modern-day reference populations potentially selected alleles could be identified. The ancient DNA data then could be used to track changes in allele frequencies over time narrowing down a time-frame in which selection may have occurred. Thereby it could be shown, that depigmentation was a gradual process that started early after humans had migrated to habitats at higher latitudes. Signs of selection were found for a variant in the ASIP gene with a strong effect on skin pigmentation, while it was found fixed in all non-African populations of the data set, indicating selection had affected this allele over 40,000 years ago. As humans moved further north, depigmentaion also increased. The results show that derived alleles in KITLG and IRF4 increased in frequencies before the onset of the Mesolithic.

Pigmentation-score results indicate that depigmentation was already advanced in the hunter-gatherers of Mesolithic Europe, while an even lighter pigmentation phenotpye can be assumed for early farmers. The data is consistent with the assumption that a vitamin D depleted diet increased to selective pressure in early farming groups propelling depigmentaion even further. Strong evidence was found for a selection of variants in the SLC24A5 gene, after the onset of the Neolithic. These variants were introduced to Central Europe at high frequencies by migrating farmers. Additional evidence suggests selection also affected variants in TYR and SLC45A2 over the course of the Neolithic until the late Bronze Age. Based on the genetic data of the Bronze Age, samples a phenotype highly similar to today can be assumed for that time. The loss of genetic diversity over time shows, that the majority of alleles associated with skin depigmentation were affected by positive directional selection.

For the alleles associated with eye and hair color phenotypes a different picture emerged. Signs of selection could be found for variants in OCA2, HERC2 and TYRP1 associated with light eye and hair color. Comparisons of F_{st} values and measurements of genetic diversity indicate that a major proportion of these alleles was under selection even before the Mesolithic. Genetic diversity in OCA2 and TYPR1 as well as F_{st} values for HERC2 show, that higher similarities can be found between hunter-gatherers and modern populations of Central Europe, compared to the early farmer groups of the area. This could indicate that selection on light eye and hair color phenotypes was driven by a European specific factor, uninfluenced by the Neolithic.

To further investigate the development of pigmentation phenotypes in Europe, additional research is needed. The increasing availability of full genome data of ancient individuals can be used in the future to apply haplotype based test directly to ancient samples. This could help to create a more nuanced picture of potential selection events, their timings and strength. Time series data from hunter-gatherers pre-dating the onset of farming in Central Europe and the Near East could also be highly valuable to this work. With a higher resolution data set it could be determined if selection increased on variants in genes such as SLC24A5 after the adoption of farming, or if it selection was already ongoing prior to the Neolithic. Further investigations in the genes related to eye and hair color phenotypes are also needed to be able to identify the underlying mechanism that led to the distribution seen today in Europe.

4 The influence of subsistence and climate on the energy metabolism

4.1 Introduction

This chapter investigates the differential prevalence of civilization diseases associated with the energy metabolism, such as the metabolic syndrome and type 2 diabetes. Several theories have tried to explain why these diseases are more widespread in some populations compared to others. While some theories suggest that adaptations to previous life conditions have led to adaptations detrimental in a modern lifestyle [Neel, 1962], others propose that the differential distributions were the outcome of drift, influenced by the various populations histories [Speakman, 2006]. In this study, ancient DNA is used to test hypotheses related to the differential distribution of risk alleles associated with civilization diseases, with a focus on European population history.

After the Neolithic transition, the Industrialization is thought of as the other major transition since the Palaeolithic [Armelagos, 2009]. The steady refinement and improvement of tools and processes since the adoption of agriculture was suddenly accelerated in an unprecedented speed. New advancements in almost every sector transformed populations around the globe. Agricultural productivity was further increased by the development of fertilizers and their industrial production, while technological advancements led to a reduction of physical labor in almost every profession. This allowed for populations to grow, leading to increased urbanization. Advances in the medical field, in the form of vaccinations, antibiotics, or general new treatments such as radiation therapy, as well as increased sanitation, promoted an increase of the overall life expectancy [Cochran and Harpending, 2009]. As a result, the average lifespan in the west has increased to roughly 80 years since the end of the second World War [Wang et al., 2012].

However, as prosperity increased, a number of new health threats arose. While infectious diseases were steadily declining, obesity, cardiovascular disease (CVD) and type 2 diabetes (T2D) took their place. Especially in westernized countries they have increased strongly. After becoming the number-one causes of non-violent deaths, they are now seen as a global health threat. Estimates predict a doubling of the prevalence of T2D in the time-span from 2000 to 2035, rising from 171mio cases to 366mio worldwide [Carrera-Bastos et al., 2011]. T2D is a metabolic disorder that is characterized by high blood-sugar levels, insulin resistance, and a relative lack of insulin production [Gross et al., 2004]. The growing epidemic of T2D is in addition fueled by a parallel epidemic of obesity. While 500mio people were obese in 2005, this number climbed to 700mio in only 10 years [Diamond, 2003, Prentice, 2005]. Simultaneously, also the number of CVD-cases rose in Europe and the US [Cordain et al., 2005]. The contributing factors that can lead to CVD and T2D, are summarized as the Metabolic Syndrome (MetS). The metabolic syndrome is diagnosed by the presence of any three of the following five conditions: (i) obesity, defined by a Body Mass Index (BMI) above 30 kg/m^2 or a significant deviation from the population middle, (ii) elevated triglyceride levels, (iii) reduced levels of high-density lipid (HDL) cholesterol ("good" cholesterol), (iv) elevated blood pressure and (v) elevated fasting glucose levels [Brown and Walker, 2016, Milici, 2010].

Environmental factors influencing the risk of MetS developement
A variety of environmental factors contribute to this epidemic: Nutrition, next to low amounts of physical activity, is among the main contributors. The over-abundance of high-caloric food has created an environment in which the general energy intake is not met by the level of physical activity. The per capita consumption of refined sugar per year in the US has increased from 55.5kg in 1970 to 69.1kg over the next 30 years [Cordain et al., 2005]. Before that, consumption rates were considerably more stable; e.g., annual consumption rates in England in the 1820s have been estimated at 6.8kg per capita [Cleave, 1974]. Large proportions of daily meals consumed today consist of highly refined grains, added sugars, fat, and salt in quantities unknown before the industrialization. Refined grains and sugars have the ability to raise the blood sugar level at much faster rates than for example whole grain products [Greenwood et al., 2013].

To be able to categorize how various foods affect the blood-sugar level, the glycemic index (GI) was developed. This index ranks foods between 0 and 100, with 100 leading to the fastest increase of the blood-sugar level. Food with a GI below 55 is considered low-glycemic, while foods with a GI above 70 are considered high-glyecmic [Cordain et al., 2005]. Processing of grains has an increasing effect on the GI, while the additional fiber in the whole grain flour lowers the GI [Liu et al., 2000]. The reason for this is that it takes the body longer to digest fiber and access the carbohydrates, which slows down the increase in blood-sugar. To take all this into account, the glycemic load is used to describe the overall ability of food to raise the blood-sugar level (GI x total amount of carbohydrates in diet). The regular consumption of high-glycemic foods, such as soft drinks or white bread, will lead to constantly increased blood sugar levels, which in turn will trigger an increased insulin secretion by pancreatic β-cells [Greenwood et al., 2013]. If insulin levels are constantly high, cells can build up resistance over time. This will cause the pancreas to increase insulin secretion even further to meet demands created by food consumption. The resulting chronic hyperglycemia and hyperinsulinemia can trigger several hormonal and physiological changes, and can finally result in pancreatic β-cells ceasing insulin production altogether [Cordain et al., 2005, Greenwood et al., 2013].

Obesity, caused by nutrition and general energy imbalance, plays a key role in the MetS and is a major risk-factor for developing T2D. Similar to diabetes, the prevalence of obesity is increasing rapidly. With the majority of meat coming from factory farms, the amount of fat in a serving of beef has increased. Prior to this development, most cattle were held on free-range farms. Today, the majority is stable farmed and raised on a diet of corn in a much shorter amount of time. The animals grown this way have a much higher fat content, visible in the "marbled meat" that originates from fat being lined in between the muscle tissue [Cordain et al., 2005]. Besides obesity, the saturated fats from meat and dairy products can have additional negative effects on health, visible in the positive correlation between the consumption of saturated fats and coronary heart diseases [Mozaffarian et al., 2010].

Next to saturated fats, also other fats can have detrimental effects on health. Heated plant oil, widely available since the 1950 as margarine or baking fat, contains a large amount of trans-fatty acids as

a by-product of the production process. Theses trans-fatty acids are associated with increasing the risk of cardiovascular diseases by affecting the balance between low-density-lipoprotein (LDL, "bad" cholesterol) and high-density-lipoprotein (HDL, "good" cholesterol), while also increasing triglyceride levels and causing systemic inflammation [moz, Mozaffarian et al., 2010, Mensink and Katan, 1990]. Apart from inflammation, clogging of arteries through cholesterol can narrow their diameter, subsequently increasing blood pressure [Blaustein et al., 2006]. Additional factors part of the westernized lifestyle that add to the risk of developing MetS are inactivity, alcohol consumption, lack of sleep, high stress, and smoking [Vancampfort et al., 2011, Carrera-Bastos et al., 2011]. The rapid increase of civilization diseases among developing countries unaffected before is seen as a direct result of advancing globalization [Littlewood, 2004].

Genetic factors involved in the MetS and T2D developement

While the environmental causes for T2D and MetS are well known, the genetic component is far less understood. Nevertheless, population, family and twin-studies have shown that genetic factors indeed contribute to the disease risk, e.g., the heritability of T2D is estimated to range between 20% - 80% [Meigs et al., 2000]. The risk to develop T2D is increased threefold in first degree relatives of T2D patients, compared to unrelated people [Florez et al., 2003]. In general, MetS and T2D are complex diseases that are not caused by a single mutation or an alteration of a single gene. Instead, a whole cascade of variants influences different aspects that all contribute a small proportion to the increased risk [Vassy et al., 2012]. Many of the genetic factors identified today originate from genome-wide association studies (GWAS) where large datasets of patients and controls were genotyped and compared after correcting for population stratification. These studies have identified a plethora of variants and genes that are associated with different aspects of the diseases. One of the major variants associated with T2D risk is rs7903146, located in the transcription-factor 7-like 2 (TCF7L2) gene. This gene is highly polymorphic and influences several pathways and thereby regulates glucose metabolism in the pancreas and liver [Facchinello et al., 2017]. Interestingly, additional variants in the gene are known for their association with other diseases, including schizophrenia [Liu et al., 2017], multiple sclerosis [Vallee et al., 2018] and cancer [Torres et al., 2017].

Several gene variants have been found to be associated with an increased risk for insulin resistance. Variants in the krüppel-like factor 14 (KLF14) gene, for example, were identified to be linked to reduced insulin action in Europeans [Voight et al., 2010]. Furthermore, gene expression was reduced in adipose and muscle tissue in T2D patients [Yang et al., 2015]. Other variants associated with insulin resistance were found in genes such as glucokinase regulator (GCKR), insulin receptor substrate 1 (IRS1), insulin-like growth factor 1 (IGF1), melanocortin 4 receptor (MC4R) and transcription elongation regulator 1-like (TCERG1L) [Brown and Walker [2016] and the sources therein].

Obesity, as one of the greatest risk factors in MetS and T2D, has also been found to be affected by genetic variation. SNPs found in the gene that transcribes the fat mass and obesity associated protein (FTO), are linked to insulin resistance as well as abdominal fat mass [Brown and Walker, 2016]. Carriers of this risk allele have, on average, a higher body-mass-index, compared to non-

carriers. Other studies have identified SNPs associated with an increased risk for hypertension. Variants in the genes angiotensie (AGT) and cytochrome P3A5 are associated with salt retention in the kidneys.
The mutations in the cytochrome gene are causing the protein to be ineffective, allowing higher rates of salt to be excreted which in turn can lead to an Na^{2+} imbalance in a low-sodium environment [Bochud et al., 2009].

Another indicator for the genetic component in MetS and T2D is the fact that groups with different ethnic backgrounds have different susceptibilities for developing them [Congdon, 2006]. A comparison of groups in the UK and the US showed that populations of African descent have higher rates of T2D, compared to populations of European or Asian ancestry [Maty et al., 2010]. A similar case was found for inhabitants of small pacific islands, where extremely high rates of obesity and T2D were found after people were introduced to a westernized diet [Jowitt, 2014]. By comparing risk allele frequencies among worldwide populations, a pattern emerged which showes a decrease in risk alleles from Africa over Europe towards East-Asia [Chen et al., 2012].

Explanations for the worldwide differences in MetS prevalence

The genetic background leading to different prevalences in MetS in worldwide populations has been the focus of a debated since it was first noticed. In 1962, Neel released a study called "A thrifty genotype rendered detrimental by progress" [Neel, 1962]. He claimed that the widespread rise in T2D was caused by a mismatch of genetics and environment. Humans spend most of their history as nomadic hunter gatherers, which allowed plenty of time for adaptive mutations to rise in frequency. Hunter gatherers were prone to seasonal fluctuations in food-availability, which created a cycle of feast and famine. According to theory, reoccurring famine would have created a selective pressure high enough to favor energy efficient variants, which led to the now predominantly present phenotype. The resulting faster increase in body mass would have been especially beneficial for childbearing women, allowing them to sustain enough energy even during food shortages, and thus to maintain the health of the child. These "thrifty genes" that allowed for the storage of energy for times of need would become detrimental after the onset of the Neolithic, when food production lead to a more stable environment. Neel interpreted the time since hunter-gatherers became sessile farmers as to short for a reversing the genetic effects on those variants. The westernized lifestyle is thought to have further increased the bias between genes and environment.

His theory was met with interest, but also criticism and before the availability of genetic data related to the subject, the hypothesis was hard to test. Critics of Neel's theory argued that there is no evidence for periodic feast and famine in hunter-gatherer societies and by comparing recent hunter-gatherers with pre-industrialized farmers, no significant differences in food availability could be determined [Benyshek and Watson, 2006]. Others assumed that the opposite to Neel's theory was more likely, with famine being more prominent and more severe in early agricultural societies [Speakman, 2006]. It is easy to imagine that a crop failure, caused by a drought or a pest, could have a harsh impact on early farmers, while hunter-gatherers were more flexible due to their mobility.

A similar by theory Colagiuri and Miller [2002] states that variants linked to insulin resistance were beneficial as long as humans lived as hunter-gatherers, because they were an adaptation to a diet high in protein, but low in carbohydrates. Even after the onset of the Neolithic, theses variants were not detrimental, since the carbohydrates consumed until the Industrialization had a low glycemic index. Only after the Industrialization these mal-adaptations became visible.

In contrast to Neel, Cochran and Harpending [2009] suggest that protective variants increased during the Neolithic as an adaptation to the new carbohydrate-rich diet. As a result, populations that never, or only recently, adopted farming were more prone to T2D. This would also explain the high rates of T2D in aboriginal people of Australia, where agriculture was absent before the colonization of the continent by Europeans.

As one of Neel's biggest critics, Speakman formulated his own theory [Speakman, 2006]. He argued that there was no feast or famine present in hunter-gatherer societies and thus no selective pressure that increased the frequency of thrifty variants. If a periodically occurring feast was the norm since early in the history of humans, thrifty variants should be fixed by now. He did, however, assume a functional constraint on variants related to the energy metabolism taking place early in the history of the genus homo, which prevented energy efficient variants to segregate. According to him, this was caused by the need to be able to flee in case of a predator attack, which was less likely to be successful for an obese person. This constraint was lifted when early humans started to use tools and manage fire. Since then, variants related to the energy metabolism were only affected by drift [Speakman, 2008, 2018].

Others arguments state that the differences in susceptibility to MetS and especially obesity were mainly caused by the adaptation to different climate zones [Sellayah et al., 2014]. The clinal distribution observed for some of the aspects linked to MetS are thought to have arisen in consequence to changes in the energy metabolism, linked to thermogenesis. In contrast to adaptations to the extreme cold, were increased fat layers for heat protection are reported [Makinen, 2010], adaptations to moderate cold have been associated with differential distribution of brown adipose tissue (BAT) as well as an increased metabolism [Bakker et al., 2014]. Sellayah et al. [2014] argue that these adaptations could have affected several different genes involved in adipose-tissue distribution and energy metabolism in correlation with climate variables in each distinct habitat.

In another theory, the distribution of MetS and T2D risk alleles has been linked to differential exposure to various diseases in different environments. Since people migrated out of Africa and settled in new climate regions, they were exposed to new diseases. Depending on the exposure, energy efficient variants became beneficial since they allowed the carriers immune system to react even when energy intake was hindered [Wells, 2008]. The spread of obesity was furthermore linked to occurrences of tuberculosis. It is assumed that the pro-inflammatory characteristics associated with excess body-fat helped in combating the infection by increasing the immune response, thus making obesity advan-

tageous in areas with high rates of tuberculosis [Roth, 2009]. Even an association with dairying was proposed [Allen and Cheer, 1996, Cheer and Allen, 1997] according to which the increase in glucose levels in the blood, created by increasing dairy consumption, could have led to the selection on protective variants in dairying populations. Dairying itself is assumed to have caused the increase of a variant that is known to cause lactase persistence in humans [Bersaglieri et al., 2004b]. This is often discussed as an example for culture-gene co-evolution [Gerbault et al., 2011].

The differential susceptibility to high blood pressure is believed to be linked to the need for salt-retention in hot and wet climates. Diets in hunter-gatherer societies are low in sodium. Variants related to an increased excretion of sodium were therefore assumed to be detrimental, which led to a functional constraint. After the migration of humans to cooler climate-zones, this constraint was lifted and allowed for the increase in frequency of "non-retaining" variants. Salt-retaining variants in combination with a high sodium diet, are assumed to be linked to increased risk of hypertension [Young et al., 2005].

Testing the hypotheses

With genetic data becoming more easily available, several studies have investigated the distribution of MetS-related loci in worldwide populations. Data indicates that a selection for a variant in TCF7L2 in Europeans may have occurred in the last 8,000 years [Helgason et al., 2007]. The selected variant was found to be protective against T2D, suggesting that carriers of the SNP were more likely to have an increased BMI, yet less likely to develop T2D. This is an example for the complexity of the underlying genetics of T2D and the MetS, which are only partially understood. Other studies have used known risk alleles to test the "thrifty gene" hypothesis directly, but could not find an excess of risk alleles in favor of protective alleles [Ayub et al., 2014, Wang and Speakman, 2016]. Patterns indicating varying degrees of selection were found for both protective and risk alleles in different populations. Ségurel et al. [2013] investigated the influence of subsistence on the distribution of frequencies in a set of ten T2D related SNPs in two neighboring populations in Central Asia. They found evidence suggesting that selection affected protective variants after the onset of the Neolithic. These patterns were visible in both populations, indicating that subsistence was not necessarily the driving force but rather factors that correlated with eco-region and climate.

An association with the climate was also found for variants in AGT which are known to be associate with hypertension [Balam-Ortiz et al., 2012]. The frequencies of the risk alleles were found to be negatively correlated with latitude, suggesting a pattern facilitated by salt-retention in warm wet climates [Young et al., 2005]. Studies using ancient DNA have furthermore provided evidence for selection affecting markers related to energy metabolism and the processing of foodstuff. Mathieson et al. [2015] found signs for selection affecting lactase persistence and the fatty acid metabolism related to decreased triglyceride levels in the blood. However, so far, ancient DNA has not been used to specifically test for hypotheses related to the prevalence of MetS and T2D. This dataset, for the first time, offers the possibility to test specific hypotheses related to the distribution of MetS risk alleles. The availability of allele frequencies of hunter-gatherers as well as early and later farmers

from Europe allows to investigate the potential influence of the Neolithic as well as the possible effect of climate variables.

4.2 Hypotheses

The data set described here consists of 197 loci associated with MetS, T2D, and the general processing of foodstuff, genotyped in 102 ancient samples. This extensive data set, in combination with genome-wide data from three reference groups was used to test the following hypotheses:

1. The high prevalence of MetS and T2D are caused by alleles that were beneficial during a time when humans lived exclusively as hunter-gatherers and were exposed to periodic cycles of feast and famine (Neel's original thrifty gene hypothesis)

   If true, most, if not all, risk alleles are found in the ancestral state. A high number of risk alleles is found in hunter-gatherer groups compared to farmer groups. After the onset of the Neolithic the frequency of the risk alleles declines with time.

2. The pattern of risk and protective alleles was shaped by a selection associated with eco-region, rather than subsistence.

   If true, high similarities between the distribution of frequencies in the Central European hunter-gatherer group and the Central European reference group can be found. Probable differences in frequencies between hunter-gatherers and early farmers decline over time.

3. The distribution of allele-frequencies related to MetS and T2D were only affected by drift.

   If true, no patterns of selection are detectable by any approach and the differences in F_{st} is similar between MetS and neutral SNPs.

4. The distribution of risk alleles associated with hypertension was shaped by a relaxing of the functional constraint on salt-retention, after the out-of-Africa migration.

   If true, the associated alleles differ in distribution from the other MetS risk alleles. All groups have similar allele frequencies and a difference between the African and non-African groups is visible by means of frequency and F_{st}.

4.3 Results

4.3.1 Allele frequencies

Loci related to the metabolic syndrome

A total of 187 SNPs located in 105 genes were investigated in association with the MetS and T2D. Of all risk alleles, 118 were in the ancestral state, while the remaining 69 were derived. Significant differences were found in the distribution of risk allele frequencies between the Central European reference group (CEU) and the African reference group (YRI, $p \leq 0.01$), as well as to the Chinese reference group (CHS, $p \leq 0.05$), but not between the Chinese and the African group (see figure 25). No significant differences were found when comparing the distributions of the same number of SNPs randomly drawn from the neutralome (see figure 26). For every group investigated here, the distribution of the risk allele frequencies differed significantly from a distribution of frequencies of SNPs randomly drawn from the neutralome ($p \leq 0.05$).

Among the European groups, the highest number of differences in comparison to the Central European reference group was found for the Central European hunter-gatherers (WHG), with 40 SNPs (55% higher, 45% lower), followed by the Central European farmers (CEF) with 26 SNPs (58% higher, 42% lower). For the Eastern European/Russian hunter-gatherers (EHG) 25 SNPs (32% higher, 68% lower) were detected to be different to the CEU reference, 13 SNPs for the farmers from the Aegean/Balkan region (SEF, 79% higher, 21%) and seven for the Bronze Age group (WEZ, 43% higher, 57% lower). Of all the SNPs found to differ from the CEU reference in the three farmer groups, only one was found to differ in all groups, while five were found to differ in both CEF and SEF, and three were different in CEF as well as WEZ. Among the hunter-gatherer groups, 12 SNPs were found to differ in WHG as well as EHG in comparison to the CEU reference.

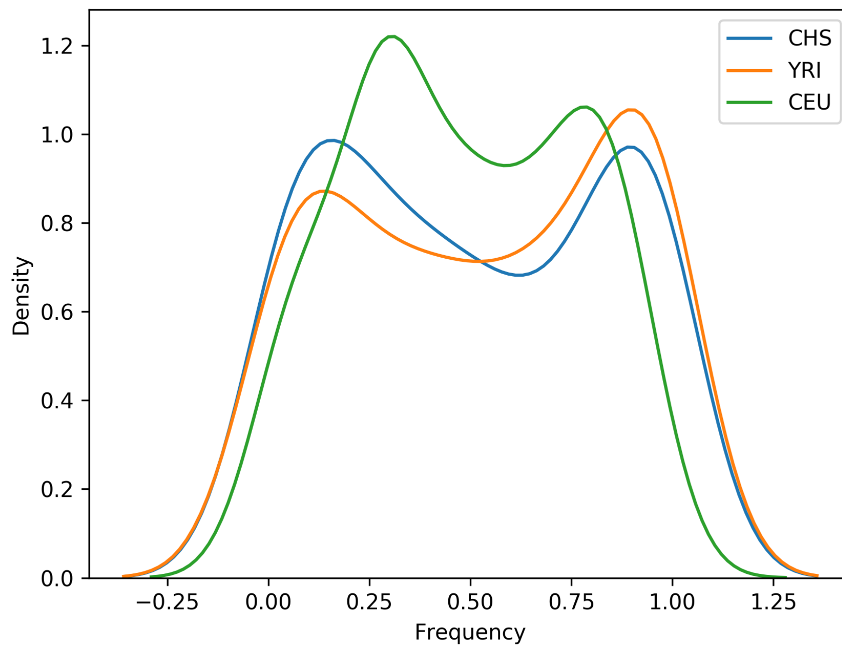

Figure (25) *Kernel density estimation plot of MetS related risk allele frequencies for the three reference groups.*

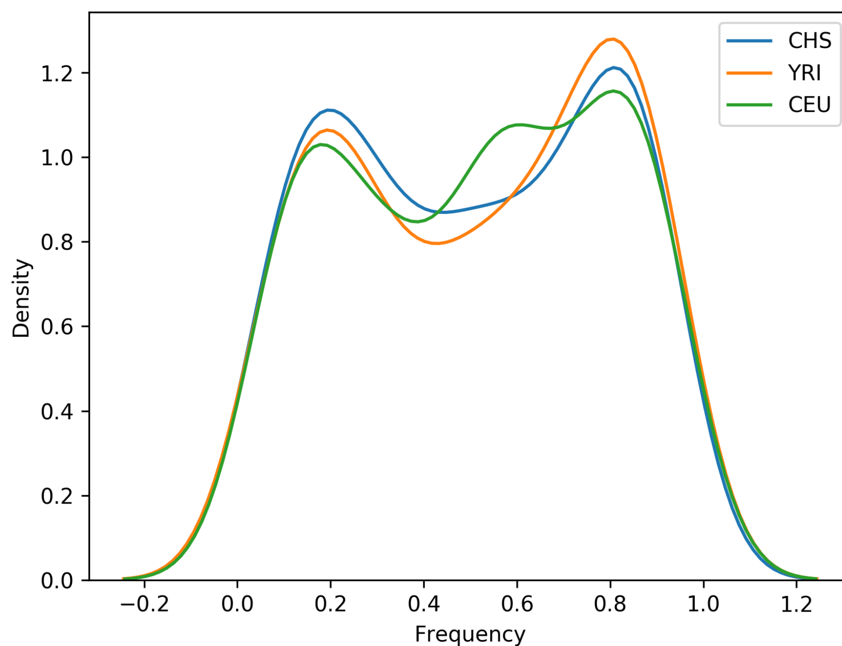

Figure (26) *Kernel density estimation plot of randomly drawn frequencies from the neutralome to match the number of MetS SNPs.*

Calculations of the Hardy-Weinberg equilibrium (HWE) resulted in significant results for every group, except the Middle Neolithic group. In the Eastern European/Russian hunter-gatherers, the SNP rs7756992 in the gene CDKAL1 was found to have a deficit of heterozygous genotypes

(p≤0.05). The only other SNP found with less heterozygotes than expected was rs11196199 in the TCF7L2 gene in the Central European farmers ($p \leq 0.01$). For the Central European farmer group, the SNPs rs6698181 in the gene PKN2, rs2943641 in the gene IRS1, and rs6446482 in gene WFS1 were found with an excess in heterozygotes ($p \leq 0.05$ each). Similarly, in the early farmer group from the Aegean/Balkan region the SNPs rs3758391 (SIRT1, $p \leq 0.05$) and rs7395662 (MADD, $p \leq 0.05$) were found to have an excess of heterozygous genotypes. The only SNP that was found to have a heterozygote excess in more than one group was rs864745 in the gene JAZF1, found in the Central European hunter-gatherers ($p \leq 0.05$) and the Bronze Age group ($p \leq 0.05$). Six SNPs were identified in the Central European reference group that deviated from HWE. For all six SNPs (rs7903146, rs12255372 (TCF7L2), rs11257622 (CDC123), rs1529927 (SLC12A3), rs174570 (FADS2), rs1552224 (CENTD2)), an excess of homozygous genotypes was determined (see table 9).
Derived alleles were absent in all groups dating before the Bronze Age for the SNP rs4988235, which is highly associated with lactase persistence in Europeans. In the WEZ group, derived alleles were found at 0.14 ± 0.18, In the CEU reference group the frequency was at 0.74 ± 0.06, while no derived alleles were found in any other reference group. Of the remaining five MCM6 SNPs, all but rs16855656 were found at significantly higher frequency in at least two ancient groups compared to the CEU reference (see table: 8). SNPs in the ALDH2 and ADH1B genes did not show significant deviation from the Central European reference in any ancient sample group.

Table (8) *Frequencies of SNPs associated with MetS and general metabolism. Significant differences in frequencies between ancient groups and the CEU reference are indicated by colors. Red: higher frequencies, blue: lower frequencies. Colors in Gene column indicate associations to subsets: green: obesity, light blue: hypertension, yellow: T2D. SNPs for which deviations from neutrality were found in CEU are highlighted green int the ID column.*

ID	Gene	Risk	EHG	WHG	SEF	CEF	MN	WEZ	CEU	YRI	CHS
rs2641348	**ADAM30**	**A**	0.42±0.28	0.58±0.19	0.67±0.38	0.88±0.09	1.0±0.6	0.6±0.3	0.92	0.62	0.96
rs4411878	**ADAMTS9**	**C**	1.0±0.31	0.97±0.06	0.7±0.2	0.66±0.12	1.0±0.6	0.88±0.16	0.78	0.79	0.78
rs4607103	**ADAMTS9**	**C**	1.0±0.26	0.97±0.06	0.65±0.21	0.65±0.14	0.83±0.3	0.82±0.16	0.78	0.68	0.64
rs6795735	**ADAMTS9**	**T**	0.33±0.27	0.04±0.07	0.36±0.25	0.56±0.13	0.25±0.42	0.3±0.2	0.48	0.88	0.75
rs11708067	**ADCY5**	**A**	0.83±0.21	0.85±0.16		1.0±0.17	1.0±0.6	0.88±0.23	0.78	0.88	1
rs182052	**ADIPOQ**	**G**	0.1±0.19	0.47±0.17	0.67±0.19	0.7±0.12	0.75±0.42	0.72±0.21	0.59	0.63	0.55
rs7649121	**ADIPOQ**	**A**	0.6±0.3	0.61±0.18	0.78±0.19	0.83±0.1	0.75±0.42	0.71±0.24	0.77	0.87	0.76
rs1042713	**ADRB2**	**G**	0.67±0.27	0.63±0.24	0.5±0.69	0.64±0.2	0.5±0.4	0.5±0.49	0.65	0.46	0.39
rs1042714	**ADRB2**	**G**	0.36±0.25	0.55±0.21		0.34±0.16	0.38±0.34	0.38±0.34	0.46	0.12	0.05
rs2493132	**AGT**	**C**	0.38±0.34	0.3±0.28	0.5±0.69	0.21±0.21	0.5±0.69	0.5±0.49	0.3	0.56	0.67
rs4762	**AGT**	**G**		0.75±0.3		0.5±0.69			0.9	0.97	0.89
rs5046	**AGT**	**A**	0.0±0.31	0.07±0.09	0.11±0.15	0.17±0.1	0.25±0.42	0.11±0.15	0.12	0.15	0.19
rs5049	**AGT**	**C**	1.0±0.31	0.92±0.11	0.83±0.21	0.85±0.1	0.75±0.42	0.9±0.19	0.88	0.66	0.81
rs5050	**AGT**	**T**	0.7±0.28	0.75±0.25		0.8±0.14	1.0±0.37	0.5±0.4	0.85	0.86	0.82
rs5051	**AGT**	**T**	0.3±0.28	0.25±0.25	0.5±0.69	0.35±0.16	0.33±0.35	0.5±0.49	0.39	0.95	0.87
rs699	**AGT**	**G**	0.38±0.34	0.33±0.27	0.5±0.69	0.4±0.3	0.33±0.31		0.41	0.92	0.87
rs7079	**AGT**	**G**	0.6±0.3	0.5±0.17	0.75±0.16	0.72±0.12	0.5±0.38	0.77±0.16	0.68	0.93	0.9
rs943580	**AGT**	**G**	0.38±0.34	0.25±0.21		0.5±0.2	0.5±0.49	0.5±0.28	0.41	0.92	0.87
rs9309473	**ALMS1**	**A**	0.75±0.42	0.5±0.4		0.83±0.21	0.5±0.49	0.5±0.31	0.75	0.38	1
rs2028299	**AP3S2**	**A**	0.92±0.16	0.72±0.16	0.91±0.12	0.8±0.1	0.5±0.4	0.54±0.2	0.74	0.63	0.78
rs662799	**APOA5**	**A**	1.0±0.31	0.96±0.07	0.94±0.12	0.94±0.06	1.0±0.46	1.0±0.26	0.92	0.88	0.73
rs429358	**APOE**	**C**	0.5±0.31	0.3±0.28		0.06±0.12	0.0±0.6		0.18	0.24	0.06

Table 8 continued from previous page

ID	Gene	Risk	EHG	WHG	SEF	CEF	MN	WEZ	CEU	YRI	CHS
rs7412	APOE	C	1.0±0.37	1.0±0.37		0.86±0.18			0.93	0.89	0.92
rs10490072	BCL11A	T	0.58±0.28	0.75±0.16	0.63±0.34	0.81±0.12	1.0±0.84	0.8±0.18	0.71	0.96	1
rs243021	BCL11A	G	0.5±0.28	0.43±0.18	0.58±0.2	0.42±0.14	1.0±0.37	0.68±0.19	0.52	0.56	0.34
rs11634397	BCL2A1	A	0.43±0.26	0.38±0.17	0.42±0.2	0.43±0.13	0.33±0.38	0.27±0.17	0.34	0.57	0.93
rs6265	BDNF	C	1.0±0.31	0.95±0.1	0.88±0.23	0.67±0.13	1.0±0.6	0.7±0.28	0.8	1	0.47
rs4923461	BDNF	A	1.0±0.31	0.97±0.06	0.75±0.16	0.58±0.13	1.0±0.6	0.75±0.19	0.76	0.86	0.53
rs925946	BDNF	G	0.08±0.16	0.18±0.16	0.5±0.49	0.86±0.1	0.25±0.3	0.42±0.28	0.66	0.72	0.94
rs11065987	BRAP	A	1.0±0.31	1.0±0.12	0.88±0.13	0.74±0.12	0.63±0.34	0.73±0.19	0.63	1	1
rs1530440	c19orf107	C	0.83±0.21	0.62±0.16	0.75±0.21	0.74±0.13	0.67±0.38	0.7±0.2	0.81	0.98	0.77
rs7172432	C2CD4A	A	0.5±0.35	0.27±0.17	0.72±0.21	0.59±0.13	1.0±0.84	0.61±0.23	0.59	0.26	0.69
rs6850980	CAMK2D	G	0.25±0.25	0.04±0.07	0.45±0.22	0.26±0.12	0.25±0.42	0.29±0.18	0.31	0.24	0.52
rs11257622	CDC123	T	0.8±0.25	0.86±0.14	0.5±0.49	0.83±0.1	0.67±0.38	0.71±0.24	0.75	0.84	0.65
rs12779790	CDC123	A	0.8±0.25	0.8±0.25	0.5±0.69	0.81±0.19	1.0±0.84	0.83±0.3	0.76	0.87	0.81
rs10440833	CDKAL1	A	0.13±0.23	0.38±0.24	0.5±0.49	0.5±0.31	0.5±0.69	0.13±0.23	0.28	0.25	0.4
rs10946398	CDKAL1	C	0.08±0.16	0.57±0.18	0.36±0.2	0.22±0.11	1.0±0.84	0.25±0.17	0.32	0.67	0.38
rs7754840	CDKAL1	G	0.92±0.16	0.45±0.16	0.62±0.19	0.78±0.11	0.0±0.84	0.85±0.16	0.68	0.33	0.62
rs7756992	CDKAL1	G	0.17±0.21	0.42±0.19	0.33±0.38	0.24±0.13	0.75±0.3	0.13±0.16	0.28	0.62	0.51
rs10811661	CDKN2A/B	C	0.14±0.18	0.21±0.14	0.45±0.21	0.34±0.12	0.33±0.38	0.13±0.13	0.2	0.02	0.42
rs1412829	CDKN2A/B	A	0.3±0.28	0.4±0.18	0.5±0.19	0.57±0.13	1.0±0.84	0.55±0.22	0.56	1	0.9
rs564398	CDKN2A/B	T	0.33±0.27	0.63±0.34		0.5±0.49	0.5±0.35	0.5±nan	0.55	1	0.9
rs10965250	CDKN2A/B	A	0.13±0.23	0.27±0.16	0.33±0.22	0.23±0.15	0.17±0.3	0.17±0.17	0.19	0.02	0.43
rs1333049	CDKN2B	C	0.2±0.25	0.38±0.19	0.5±0.4	0.5±0.14	0.5±0.4	0.5±0.22	0.44	0.21	0.51
rs1552224	CENTD2	A	0.44±0.24	0.5±0.17	0.82±0.16	0.91±0.08	1.0±0.46	0.7±0.2	0.88	1	0.91
rs13292136	CHCHD9	C	0.88±0.23	1.0±0.46		0.83±0.3	1.0±0.6		0.93	0.92	0.93

Table 8 continued from previous page

ID	Gene	Risk	EHG	WHG	SEF	CEF	MN	WEZ	CEU	YRI	CHS
rs1378942	CYP1A2	C	0.67±0.27	0.44±0.16	0.69±0.18	0.63±0.13	0.5±0.35	0.46±0.2	0.32	1	0.8
rs762551	CYP1A2	A	0.5±0.26	0.47±0.16	0.32±0.19	0.45±0.13	0.67±0.38	0.63±0.17	0.73	0.55	0.64
rs1153188	DCD	A	0.5±0.49	0.33±0.19	0.5±0.35	0.42±0.28	0.0±0.84	0.43±0.26	0.74	0.79	0.99
rs2191349	DGKB	T	0.5±0.28	0.5±0.18	0.55±0.21	0.67±0.13	0.5±0.4	0.67±0.19	0.54	0.51	0.64
rs5945326	DUSP9	A	0.5±0.69	0.5±0.35		0.75±0.21	1.0±0.84		0.79	0.81	0.54
rs3759324	EnaCα	T	0.86±0.18	0.62±0.16	0.8±0.18	0.91±0.08	0.83±0.3	0.61±0.23	0.77	0.3	0.46
rs11037909	EXT2	T	0.5±0.31	0.58±0.2	0.9±0.13	0.94±0.06	0.75±0.42	0.59±0.21	0.7	0.82	0.6
rs5896	F2	C	0.63±0.34	1.0±0.46		0.93±0.6	1.0±0.84	1.0±nan	0.9	1	0.36
rs1799883	FABP2	T	0.6±0.3	0.3±0.16	0.75±0.3	0.33±0.14	0.5±0.49	0.1±0.19	0.31	0.23	0.21
rs174570	FADS2/3	C	0.17±0.21	0.35±0.16	0.73±0.17	0.81±0.1	0.5±0.3	0.78±0.19	0.84	0.99	0.42
rs16998073	FGF5	A	0.58±0.28	0.47±0.16	0.75±0.21	0.94±0.06	0.5±0.35	0.88±0.12	0.77	0.89	0.64
rs11642841	FTO	C	0.75±0.25	0.6±0.18	0.56±0.24	0.67±0.13	0.0±0.84	0.56±0.23	0.56	0.94	0.97
rs1421085	FTO	T	0.75±0.3	0.54±0.19	0.46±0.2	0.65±0.13	0.0±0.6	0.38±0.24	0.55	0.94	0.86
rs1477196	FTO	G	0.5±0.4	0.67±0.19	0.5±0.69	0.72±0.21	1.0±0.84	0.56±0.24	0.66	0.97	0.76
rs17817449	FTO	G	0.2±0.25	0.43±0.18	0.5±0.23	0.35±0.13	1.0±0.84	0.44±0.24	0.44	0.4	0.14
rs7202116	FTO	A	0.7±0.28	0.44±0.23	0.75±0.42	0.65±0.16	0.0±0.6	0.4±0.3	0.56	0.49	0.86
rs8050136	FTO	A	0.17±0.21	0.62±0.16	0.42±0.28	0.36±0.13	0.88±0.23	0.4±0.21	0.44	0.47	0.14
rs9921255	FTO	T	1.0±0.46	0.9±0.13	0.58±0.28	0.83±0.3	0.25±0.42	1.0±0.37	0.79	0.75	0.9
rs9930506	FTO	G	0.25±0.3	0.5±0.21	0.33±0.27	0.38±0.24	1.0±0.84	0.5±0.23	0.47	0.21	0.18
rs9939609	FTO	A	0.17±0.21	0.5±0.15	0.5±0.2	0.31±0.13	0.75±0.3	0.5±0.2	0.44	0.51	0.14
rs4607517	GCK	G	0.5±0.35	0.94±0.08	0.77±0.18	0.88±0.09	0.75±0.42	0.75±0.25	0.79	0.93	0.81
rs917793	GCK	A	0.8±0.25	0.92±0.1	0.5±0.31	0.85±0.1	0.67±0.38	0.75±0.21	0.79	0.7	0.81
rs780094	GCKR	C	1.0±0.26	0.94±0.08	0.31±0.23	0.48±0.13	0.33±0.38	0.58±0.28	0.59	0.9	0.51
rs3923113	GRB14	A	0.75±0.3	0.6±0.21	0.6±0.3	0.46±0.19	1.0±0.84	0.45±0.22	0.59	0.22	0.89

Table 8 continued from previous page

ID	Gene	Risk	EHG	WHG	SEF	CEF	MN	WEZ	CEU	YRI	CHS
rs2373115	**GRB2**	**A**	0.1±0.19	0.17±0.21	0.25±0.42	0.28±0.21	0.33±0.38	0.2±0.25	0.13	0.29	0.42
rs1111875	**HHEX**	**T**	0.4±0.3	0.16±0.13	0.15±0.16	0.28±0.12	0.0±0.37	0.56±0.23	0.42	0.12	0.76
rs7923837	**HHEX**	**G**	0.58±0.28	0.96±0.07	0.88±0.13	0.72±0.12	1.0±0.84	0.5±0.2	0.63	1	0.13
rs5015480	**HHEX**	**C**	0.63±0.34	0.81±0.19	0.5±0.69	0.67±0.38	1.0±0.84	0.5±0.28	0.58	0.63	0.14
rs7178572	**HMG20A**	**G**	0.75±0.3	0.93±0.09	0.79±0.16	0.75±0.12	0.5±0.4	0.69±0.18	0.69	0.42	0.39
rs2612067	**HMGA2**	**G**	0.08±0.16	0.03±0.06	0.12±0.12	0.13±0.09	0.33±0.38	0.18±0.14	0.11	0.42	0.12
rs4812829	**HNF4A**	**A**	0.1±0.19	0.14±0.13	0.14±0.14	0.13±0.09	0.0±0.6	0.23±0.16	0.18	0.05	0.44
rs1470579	**IGF2BP2**	**C**	0.13±0.23	0.23±0.16	0.36±0.2	0.25±0.12	0.0±0.6	0.38±0.24	0.31	0.86	0.23
rs4402960	**IGF2BP2**	**G**	0.86±0.18	0.75±0.15	0.63±0.19	0.76±0.11	1.0±0.31	0.67±0.19	0.69	0.46	0.77
rs1800795	**IL6**	**G**	0.67±0.27	0.5±0.19	0.5±0.49	0.75±0.12	0.67±0.38	0.57±0.26	0.48	1	1
rs689	**INS**	**A**	0.5±0.31	0.25±0.25	0.75±0.42	0.48±0.15	0.5±0.4	0.63±0.34	0.27	0.8	0.06
rs2943641	**IRS1**	**T**	0.67±0.27	0.66±0.16	0.63±0.24	0.46±0.14	0.5±0.49	0.31±0.23	0.34	0.23	0.07
rs7578326	**IRS1**	**A**	0.42±0.28	0.5±0.19	0.5±0.69	0.48±0.14	0.6±0.3	0.69±0.23	0.66	0.57	0.83
rs1635852	**JAZF1**	**T**	0.38±0.34	0.5±0.25	0.5±0.49	0.42±0.2	0.75±0.42	0.4±0.3	0.51	0.77	0.79
rs849134	**JAZF1**	**A**	0.38±0.34	0.59±0.17	0.64±0.25	0.39±0.13	0.83±0.3	0.33±0.27	0.51	0.78	0.79
rs864745	**JAZF1**	**T**	0.29±0.24	0.54±0.18	0.45±0.22	0.45±0.15	0.75±0.42	0.28±0.21	0.51	0.78	0.79
rs5215	**KCNJ11**	**T**	0.57±0.26	0.46±0.19	0.75±0.42	0.58±0.13	0.5±0.4	0.56±0.23	0.62	0.99	0.6
rs5219	**KCNJ11**	**C**	0.6±0.3	0.36±0.25		0.39±0.23	0.5±0.4	0.5±0.69	0.62	1	0.6
rs17779747	**KCNJ2**	**G**	0.67±0.27	0.62±0.16	0.5±0.19	0.59±0.13	0.5±0.31	0.69±0.23	0.69	0.96	0.94
rs163184	**KCNQ1**	**G**	0.5±0.35	0.61±0.23	0.5±0.49	0.42±0.14	0.25±0.42	0.5±0.35	0.44	0.13	0.37
rs2074196	**KCNQ1**	**G**	0.9±0.19	0.92±0.16		1.0±0.11	1.0±0.6		0.98	0.84	0.59
rs2283228	**KCNQ1**	**A**	0.9±0.19	0.93±0.09	0.89±0.15	0.98±0.03	1.0±0.46	0.95±0.1	0.91	0.88	0.6
rs231362	**KCNQ1**	**A**	0.4±0.3	0.79±0.21	0.5±0.69	0.59±0.14	0.25±0.42	0.38±0.34	0.49	0.18	0.09
rs368794	**KCTD15**	**T**	0.25±0.25	0.54±0.19	0.32±0.19	0.25±0.19	0.5±0.38	0.57±0.26	0.3	0.33	0.7

Table 8 continued from previous page

ID	Gene	Risk	EHG	WHG	SEF	CEF	MN	WEZ	CEU	YRI	CHS
rs972283	**KLF14**	**G**	0.4±0.3	0.5±0.19	0.5±0.49	0.53±0.16	0.5±0.49	0.38±0.24	0.54	0.91	0.7
rs1137100	**LEPR**	**A**	0.75±0.25	0.31±0.16	0.88±0.13	0.89±0.08	0.67±0.38	0.71±0.18	0.7	0.85	0.16
rs4655518	**LEPR**	**G**	0.25±0.3	0.68±0.17	0.17±0.15	0.13±0.09	0.38±0.34	0.29±0.17	0.31	0.22	0.86
rs7480010	**LOC387761**	**G**	0.25±0.25	0.4±0.18	0.21±0.16	0.27±0.12	0.33±0.38	0.25±0.17	0.28	1	0.22
rs7395662	**MADD**	**G**	0.5±0.31	0.35±0.16	0.55±0.22	0.71±0.12	0.5±0.4	0.56±0.24	0.64	0.4	0.4
rs17782313	**MC4R**	**T**	1.0±0.31	0.77±0.15	0.83±0.15	0.98±0.04	0.5±0.35	0.77±0.16	0.74	0.71	0.84
rs2229616	**MC4R**	**C**	1.0±0.31	1.0±0.14	1.0±0.37	1.0±0.07	1.0±0.6	1.0±0.23	0.99	0.99	0.97
rs7227255	**MC4R**	**A**	0.0±0.31	0.0±0.12	0.0±0.23	0.0±0.07	0.0±0.6	0.0±0.21	0	0.27	0
rs2314349	**MCF2L2**	**G**	1.0±0.31	0.6±0.18	0.95±0.1	0.75±0.25	0.75±0.3	0.83±0.15	0.87	0.23	0.54
rs3812316	**MLXIPL**	**C**	0.75±0.3	0.79±0.21		0.93±0.13	1.0±0.84	0.5±0.69	0.86	0.99	0.93
rs10838738	**MTCH2**	**A**	0.42±0.28	0.5±0.18	0.62±0.19	0.59±0.13	1.0±0.37	0.58±0.19	0.65	0.96	0.69
rs17367504	**MTHFR**	**G**	0.0±0.26	0.14±0.13	0.05±0.09	0.11±0.08	0.33±0.38	0.13±0.16	0.16	0.09	0.1
rs1801133	**MTHFR**	**G**	0.63±0.34	0.83±0.21		0.46±0.18	1.0±0.84	1.0±0.84	0.7	0.91	0.71
rs10830963	**MTNR1B**	**C**	0.5±0.28	0.67±0.22	0.5±0.69	0.56±0.16	1.0±0.6	0.7±0.28	0.74	0.97	0.61
rs1387153	**MTNR1B**	**C**	0.5±0.28	0.56±0.17	0.71±0.18	0.69±0.13	1.0±0.46	0.63±0.19	0.75	0.57	0.55
rs162036	**MTRR**	**A**	1.0±0.26	1.0±0.12	0.78±0.19	0.8±0.1	0.5±0.31	0.9±0.13	0.83	0.57	0.8
rs2568958	**NEGR1**	**G**	0.3±0.28	0.36±0.18	0.42±0.28	0.48±0.13	0.75±0.42	0.39±0.23	0.36	0.48	0.06
rs10923931	**NOTCH2**	**G**	0.43±0.26	0.56±0.16	0.88±0.23	0.9±0.09	1.0±0.46	0.75±0.3	0.92	0.63	0.96
rs1493694	**NOTCH2**	**C**	0.5±0.31	0.53±0.18	0.67±0.38	0.94±0.07	1.0±0.6	0.5±0.49	0.92	0.38	0.96
rs11724320	**NPY1R/5R**	**C**	0.7±0.28	0.68±0.17	0.75±0.19	0.8±0.25	0.67±0.38	0.75±0.21	0.67	0.33	0.26
rs7957197	**OASL**	**T**	1.0±0.46	1.0±0.31		0.71±0.46	1.0±0.46	1.0±nan	0.85	0.88	1
rs2075756	**PAI-1**	**A**	0.0±0.31	0.23±0.18	0.5±0.4	0.31±0.13	0.33±0.38	0.4±0.3	0.25	0.05	0.4
rs6235	**PCSK1**	**C**	0.58±0.28	0.81±0.13	0.81±0.15	0.71±0.12	0.5±0.35	0.65±0.18	0.69	0.9	0.57
rs6698181	**PKN2**	**T**	0.38±0.34	0.44±0.23	0.17±0.3	0.5±0.28	0.0±0.84	0.19±0.19	0.36	0.04	0.33

Table 8 continued from previous page

ID	Gene	Risk	EHG	WHG	SEF	CEF	MN	WEZ	CEU	YRI	CHS
rs12946454	PLCD3	A	1.0±0.84	0.6±0.3		0.0±0.84			0.74	0.95	0.73
rs896854	PLEKHF2	T	0.83±0.21	0.69±0.16	0.6±0.3	0.52±0.14	0.5±0.31	0.4±0.3	0.43	0.77	0.3
rs4751995	PLRP2	G	0.6±0.3	0.54±0.2	0.63±0.24	0.68±0.13	0.38±0.28	0.64±0.2	0.54	0.52	0.28
rs3917498	PON1	G	0.5±0.69	0.75±0.42	0.5±0.4	0.5±0.69	1.0±0.84	0.67±0.38	0.63	0.47	0.35
rs662	PON1	C	0.29±0.24	0.29±0.15	0.44±0.24	0.23±0.11	0.0±0.6	0.36±0.25	0.31	0.78	0.65
rs1525791	POU6F2	C	0.75±0.25	0.93±0.09	0.75±0.17	0.84±0.1	1.0±0.46	0.8±0.18	0.83	0.89	0.84
rs13081389	PPARG	A	0.88±0.23	0.79±0.16	1.0±0.14	1.0±0.1	1.0±0.6	0.86±0.14	0.94	0.98	0.97
rs1801282	PPARG	C	0.67±0.27	0.78±0.19		0.97±0.06	1.0±0.46	0.88±0.23	0.9	1	0.97
rs8042680	PRC1	A	0.33±0.27	0.29±0.15	0.3±0.2	0.41±0.13	0.0±0.6	0.25±0.19	0.28	0.98	1
rs340874	PROX1	T	0.75±0.21	0.57±0.18	0.61±0.23	0.54±0.13	0.83±0.3	0.57±0.18	0.47	0.93	0.55
rs17584499	PTPRD	T	0.0±0.31	0.31±0.23		0.1±0.11	0.0±0.6	0.33±0.38	0.2	0.04	0.1
rs7593730	RBMS1	C	0.92±0.16	0.93±0.1	0.71±0.18	0.74±0.12	0.5±0.49	0.79±0.16	0.82	0.57	0.81
rs10149848	RGS6	A	0.1±0.19	0.0±0.14	0.33±0.38	0.12±0.09	0.25±0.38	0.1±0.19	0.05	0.31	0.01
rs1531343	RPSAP52	G	0.92±0.16	0.97±0.06	0.88±0.13	0.87±0.09	0.63±0.34	0.92±0.11	0.89	0.57	0.88
rs10913469	SEC16B	C	0.67±0.27	0.25±0.17	0.0±0.84	0.03±0.05	0.25±0.42	0.3±0.28	0.24	0.33	0.23
rs12778366	SIRT1	T	0.25±0.3	0.89±0.11	0.81±0.15	0.9±0.08	0.75±0.38	0.96±0.08	0.88	0.93	0.87
rs3758391	SIRT1	T	0.88±0.23	0.25±0.19	0.5±0.28	0.24±0.12	0.5±0.69	0.2±0.25	0.28	0.33	0.85
rs1529927	SLC12A3	G	1.0±0.46	1.0±0.6		1.0±0.37	1.0±0.6	1.0±0.84	0.93	1	1
rs117767867	SLC16A11	C	1.0±0.37	1.0±0.31		1.0±0.42	0.75±0.6	1.0±nan	0.99	1	0.9
rs13342232	SLC16A11	G	0.0±0.37	0.0±0.37		0.0±0.42	0.25±0.46	0.0±nan	0.02	0.38	0.1
rs13342692	SLC16A11	C	0.0±0.31	0.0±0.23	0.0±0.6	0.03±0.06	0.25±0.42	0.0±0.46	0.02	0.38	0.1
rs75493593	SLC16A11	G	1.0±0.37	0.8±0.25		1.0±0.69	0.5±0.84	1.0±nan	0.99	1	0.9
rs312457	SLC16A13	G	0.0±0.31	0.14±0.18	0.0±0.84	0.0±0.11	0.33±0.49	0.0±0.46	0.01	0.15	0.1
rs13266634	SLC30A8	C	0.9±0.19	0.79±0.15	1.0±0.23	0.91±0.08	1.0±0.6	0.63±0.24	0.76	0.93	0.51

Table 8 continued from previous page

ID	Gene	Risk	EHG	WHG	SEF	CEF	MN	WEZ	CEU	YRI	CHS
rs3802177	**SLC30A8**	**G**	0.9±0.19	0.64±0.2	0.96±0.07	0.9±0.08	1.0±0.6	0.68±0.19	0.76	0.93	0.51
rs391300	**SLC30A8**	**C**	0.5±0.31	0.55±0.21	0.5±0.49	0.38±0.16	0.25±0.42	0.42±0.28	0.66	0.42	0.64
rs13373826	**SLC44A5**	**G**	0.0±0.26	0.11±0.15	0.0±0.46	0.0±0.07	0.0±0.6	0.17±0.21	0.15	0.2	0.02
rs1359790	**SPRY2**	**G**	0.88±0.23	0.83±0.13	0.61±0.18	0.67±0.13	0.33±0.38	0.63±0.19	0.73	0.94	0.7
rs16861329	**ST6GAL1**	**C**	0.5±0.28	1.0±0.12	0.77±0.16	0.89±0.08	1.0±0.46	0.68±0.19	0.87	0.95	0.77
rs17036101	**SYNC**	**A**	0.25±0.25	0.25±0.16	0.0±0.13	0.0±0.07	0.0±0.6	0.08±0.11	0.06	0.02	0.03
rs4430796	**TCF2**	**G**	0.25±0.25	0.35±0.16	0.82±0.16	0.72±0.12	1.0±0.46	0.58±0.2	0.47	0.65	0.26
rs757210	**TCF2**	**T**	0.38±0.34	0.07±0.13		0.29±0.4	0.5±0.53	0.5±nan	0.37	0.52	0.26
rs10885406	**TCF7L2**	**G**	0.1±0.19	0.3±0.16	0.42±0.19	0.46±0.14	0.33±0.38	0.54±0.2	0.49	0.86	0.03
rs11196175	**TCF7L2**	**T**	0.9±0.19	0.94±0.11	0.75±0.42	0.69±0.14	1.0±0.6	0.9±0.19	0.69	1	1
rs11196181	**TCF7L2**	**G**	1.0±0.31	1.0±0.15	0.75±0.3	0.92±0.08	1.0±0.6	0.83±0.3	0.92	1	1
rs11196192	**TCF7L2**	**T**	1.0±0.26	0.89±0.11	0.94±0.12	0.89±0.08	1.0±0.46	0.78±0.19	0.91	0.96	0.99
rs11196199	**TCF7L2**	**A**	1.0±0.37	0.92±0.1	0.88±0.16	0.94±0.07	1.0±0.84	0.92±0.16	0.83	0.85	1
rs11196205	**TCF7L2**	**C**	0.17±0.21	0.31±0.16	0.42±0.2	0.44±0.14	0.5±0.35	0.46±0.2	0.47	0.85	0.01
rs11196213	**TCF7L2**	**C**	0.9±0.19	0.8±0.14	0.58±0.2	0.61±0.13	0.5±0.49	0.45±0.22	0.56	0.31	0.99
rs11196228	**TCF7L2**	**T**	0.9±0.19	1.0±0.15	0.56±0.24	0.83±0.1	1.0±0.6	0.92±0.16	0.93	1	1
rs12255372	**TCF7L2**	**G**	1.0±0.37	0.81±0.15	0.6±0.3	0.63±0.14	0.67±0.3	0.68±0.19	0.72	0.7	0.99
rs1555485	**TCF7L2**	**C**	0.8±0.25	0.57±0.18		0.92±0.08	1.0±0.6	0.92±0.16	0.79	1	0.95
rs17685538	**TCF7L2**	**C**	0.38±0.34	0.75±0.15	1.0±0.14	0.96±0.05	1.0±0.6	1.0±0.19	0.85	1	1
rs17747324	**TCF7L2**	**T**	0.9±0.19	0.88±0.16		0.69±0.18	1.0±0.6		0.77	0.98	0.98
rs290494	**TCF7L2**	**T**	0.5±0.35	0.59±0.21	0.75±0.25	0.79±0.21	0.5±0.49	0.75±0.42	0.78	0.97	0.98
rs4506565	**TCF7L2**	**T**	0.07±0.13	0.26±0.15	0.36±0.2	0.37±0.13	0.17±0.28	0.46±0.2	0.33	0.44	0.03
rs7079711	**TCF7L2**	**G**	0.7±0.28	0.65±0.16	0.83±0.15	0.91±0.08	0.75±0.38	0.69±0.23	0.8	0.52	0.98
rs7895340	**TCF7L2**	**G**	0.9±0.19	0.6±0.21	0.5±0.69	0.67±0.17	0.75±0.42	0.25±0.25	0.53	0.15	0.99

Table 8 continued from previous page

ID	Gene	Risk	EHG	WHG	SEF	CEF	MN	WEZ	CEU	YRI	CHS
rs7896811	TCF7L2	C	1.0±0.46	0.67±0.38		0.83±0.21	1.0±0.6	1.0±0.84	0.86	0.85	1
rs7901695	TCF7L2	C	0.08±0.16	0.21±0.15	0.33±0.27	0.31±0.13	0.33±0.38	0.44±0.23	0.33	0.44	0.03
rs7903146	TCF7L2	T	0.08±0.16	0.09±0.12	0.25±0.42	0.33±0.15	0.0±0.46	0.33±0.27	0.31	0.24	0.03
rs7919375	TCF7L2	A	1.0±0.6	0.5±0.4	1.0±0.84	0.88±0.23	0.75±0.42		0.8	0.19	0.2
rs7924080	TCF7L2	T	0.83±0.3	0.75±0.17	0.75±0.25	0.56±0.13	0.5±0.69	0.55±0.21	0.51	0.13	0.99
rs11899863	THADA	C	0.7±0.28	0.75±0.21		1.0±0.15	1.0±0.6	0.88±0.23	0.89	0.89	1
rs7578597	THADA	T	0.7±0.28	0.85±0.12	1.0±0.15	1.0±0.07	1.0±0.6	0.94±0.11	0.88	0.65	0.99
rs6548238	TMEM18	C	0.75±0.25	0.91±0.1	0.67±0.27	0.75±0.11	0.75±0.3	0.77±0.18	0.83	0.94	0.9
rs4760790	TSPAN8	G	0.88±0.23	0.88±0.13	0.38±0.24	0.69±0.13	1.0±0.6	0.67±0.22	0.74	0.87	0.75
rs1353362	TSPAN8	T	0.9±0.19	0.83±0.21	0.5±0.69	0.75±0.21	1.0±0.84	0.88±0.23	0.73	0.96	0.75
rs7961581	TSPAN8	C	0.17±0.21	0.16±0.13	0.55±0.21	0.33±0.13	0.0±0.6	0.22±0.19	0.26	0.19	0.21
rs6780569	UBE2E2	A	0.0±0.26	0.07±0.1	0.12±0.12	0.15±0.09	0.0±0.46	0.04±0.07	0.09	0.66	0.18
rs6536991	UCP1	C	0.1±0.19	0.35±0.21	0.17±0.21	0.42±0.14	0.0±0.84	0.22±0.19	0.25	0.59	0.21
rs9472138	VEGFA	C	0.5±0.28	0.79±0.21		0.79±0.21	0.67±0.38	0.75±0.3	0.77	0.85	0.89
rs1802295	VPS26A	A	0.0±0.3	0.0±0.18	0.0±0.18	0.0±0.11	0.0±0.42	0.0±0.21	0	0	0
rs10010131	WFS1	G	0.9±0.19	0.6±0.3		0.69±0.23	0.5±0.4	0.5±0.4	0.65	0.71	0.87
rs10012946	WFS1	T	0.13±0.23	0.54±0.19	0.5±0.4	0.4±0.13	0.5±0.49	0.58±0.28	0.35	0.29	0.13
rs1801214	WFS1	C		0.5±0.4		0.38±0.34	0.5±0.69		0.35	0.33	0.12
rs6446482	WFS1	C	0.25±0.42	0.5±0.31		0.5±0.26	0.5±0.69	0.5±0.4	0.35	0.29	0.13
rs4457053	ZBED3	A	0.5±0.28	0.77±0.16	0.75±0.42	0.57±0.26	1.0±0.6	0.5±0.21	0.7	0.85	0.96
rs2722425	ZMAT4	C	0.86±0.18	0.85±0.12	0.85±0.14	0.8±0.1	1.0±0.46	0.82±0.16	0.9	0.56	0.73
rs16948048	ZNF652	A	0.33±0.27	0.53±0.17	0.5±0.19	0.59±0.13	0.38±0.34	0.55±0.22	0.57	0.57	0.74
rs7570283	MCM6	C	0.5±0.69	0.25±0.19	0.42±0.28	0.37±0.17	0.0±0.84	0.25±0.21	0.08	0.12	0.39
rs4988235	MCM6	A	0.0±0.26	0.0±0.12	0.0±0.6	0.0±0.07	0.0±0.37	0.14±0.18	0.74	0.00	0.00

Table 8 continued from previous page

ID	Gene	Risk	EHG	WHG	SEF	CEF	MN	WEZ	CEU	YRI	CHS
rs182549	**MCM6**	**T**	0.08±0.16	0.07±0.13		0.0±0.11	0.0±0.46	0.25±0.3	0.74	0.00	0.00
rs584226	**MCM6**	**C**	0.42±0.28	0.38±0.17	0.35±0.18	0.33±0.13	0.33±0.38	0.36±0.2	0.09	0.32	0.40
rs16855656	**MCM6**	**G**		0.2±0.25	0.5±0.69	0.5±0.69	0.0±0.84	0.33±0.38	0.10	0.32	0.40
rs309126	**MCM6**	**C**	0.4±0.3	0.32±0.17	0.5±0.35	0.35±0.13	0.25±0.3	0.5±0.4	0.10	0.35	0.40
rs3099429	**MCM6**	**A**	0.2±0.25	0.3±0.28		0.31±0.18	0.33±0.38	1.0±0.84	0.10	0.23	0.46
rs671	**ALDH2**	**A**	0.0±0.26	0.0±0.12	0.0±0.21	0.0±0.07	0.0±0.37	0.0±0.23	0.00	0.00	0.27
rs1229984	**ADH1Bb**	**T**	0.0±0.26	0.0±0.26		0.0±0.26	0.13±0.23	0.0±0.41	0.02	0.00	0.76
rs3811801	**ADH1Ba**	**A**	0.0±0.26	0.0±0.14	0.0±0.23	0.0±0.08	0.0±0.46	0.0±0.19	0.00	0.00	0.55

Table (9) *Significant results of Hardy-Weinberg equilibrium calculations for MetS related SNPs.*

Group	**ID**	**Gene**	$\mathbf{Ho1_{obs}}$	$\mathbf{He_{obs}}$	$\mathbf{Ho2_{obs}}$	$\mathbf{Ho1_{exp}}$	$\mathbf{He_{exp}}$	$\mathbf{Ho2_{exp}}$	χ^2	**P**	**Het deficit**	**Het excess**
EHG	**rs7756992**	**CDKAL1**	5.00	0.00	1.00	4.17	1.67	0.17	6.00	0.050	1.00	-1.00
SEF	**rs3758391**	**SIRT1**	0.00	6.00	0.00	1.50	3.00	1.50	6.00	0.050	-1.00	1.00
SEF	**rs7395662**	**MADD**	0.00	9.00	1.00	2.03	4.95	3.03	6.69	0.035	-0.82	0.82
CEF	**rs6698181**	**PKN2**	0.00	6.00	0.00	1.50	3.00	1.50	6.00	0.050	-1.00	1.00
CEF	**rs11196199**	**TCF7L2**	22.00	1.00	1.00	21.09	2.81	0.09	9.97	0.007	0.64	-0.64
CEF	**rs2943641**	**IRS1**	2.00	18.00	4.00	5.04	11.92	7.04	6.25	0.044	-0.51	0.51
CEF	**rs6446482**	**WFS1**	0.00	7.00	0.00	1.75	3.50	1.75	7.00	0.030	-1.00	1.00
WHG	**rs864745**	**JAZF1**	1.00	10.00	0.00	3.27	5.45	2.27	7.64	0.022	-0.83	0.83
WEZ	**rs864745**	**JAZF1**	1.00	10.00	0.00	3.27	5.45	2.27	7.64	0.022	-0.83	0.83
CEU	**rs174570**	**FADS2**	73	20	6	69.59	26.83	2.59	6.41	0.04049	0.2545	-0.2545
CEU	**rs1552224**	**CENTD2**	80	15	4	77.34	20.33	1.34	6.80	0.0334	0.2621	-0.2621
CEU	**rs7903146**	**TCF7L2**	53	31	15	47.40	42.21	9.40	6.98	0.03050	0.2655	-0.2655
CEU	**rs12255372**	**TCF7L2**	52	32	15	46.71	42.59	9.71	6.12	0.04695	0.2486	-0.2486
CEU	**rs11257622**	**CDC123**	60	28	11	55.31	37.37	6.31	6.23	0.04443	0.2508	-0.2508
CEU	**rs1529927**	**SLC12A3**	2	9	88	0.43	12.15	86.43	6.64	0.03609	0.2590	-0.259

4.3.2 Neutrality tests

In the Central European reference group, 54 SNPs in 34 genes were identified to deviate from neutral expectations by at least one neutrality test (for details see table 10 and 8). Of those SNPs, 19 differed between one or more ancient sample group and the Central European reference group. The highest number of non-neutral SNPs was identified in the Chinese reference group, while the lowest number was found in the African reference group (see figure 27).

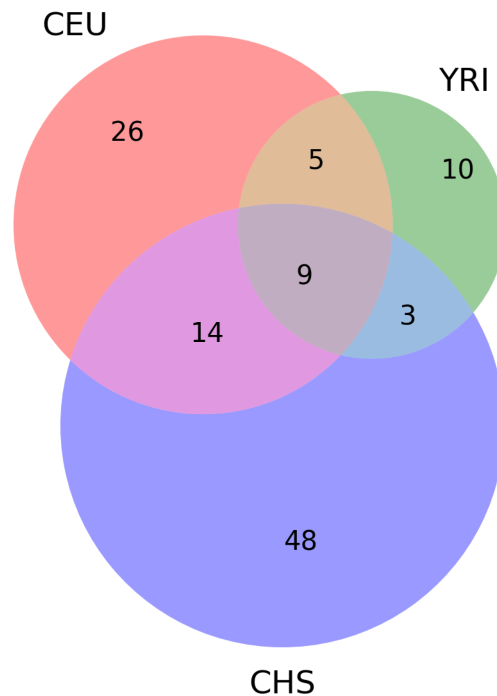

Figure (27) *Venn diagram of the number of SNPs that deviated from neutral expectations in each of the three reference groups. CEU=Central Europeans, YRI=Africans, CHS=Chinese.*

Table (10) *Results of neutrality tests for all three reference groups. Only columns with at least one result in the top or bottom 2.5^{th} percentile are shown.*

ID	**Gene**	CEU_iHS	p	CEU_nSl	p	CEU_D	p	H12	p	H2_H1	p	**Group**
rs1802295	**VPS26A**	-2.3934	1.0635	-1.7616	4.3016	0.7778	76.6875	0.0964	28.7348	0.4286	65.0275	CEU
rs2641348	**ADAM30**	-0.8711	18.7864	-0.9775	16.6077	-1.9042	2.7801	0.5555	99.0706	0.0341	3.6073	
rs10923931	**NOTCH2**	-	-	-	-	-1.9219	2.6247	0.5189	98.4756	0.0144	0.7822	
rs1493694	**NOTCH2**	-	-	-	-	-2.047	1.8006	0.4698	97.2836	0.0282	2.7133	
rs174570	**FADS2**	2.1127	98.2953	2.3381	99.1808	0.5254	67.1837	0.2505	81.6155	0.4305	67.6683	
rs4655518	**LEPR**	-0.2517	39.01	-0.2378	39.5625	2.4373	98.6764	0.1828	59.9796	0.1514	21.9986	
rs17782313	**MC4R**	-2.0514	2.3037	-1.5639	6.0911	2.457	98.7074	0.0745	19.3566	0.5644	86.0045	
rs17036101	**PPARG**	1.2841	90.4186	0.4682	67.3331	0.3522	62.4316	0.0313	2.708	0.6467	94.209	
rs2314349	**MCF2L2**	-2.7231	0.4249	-2.9863	0.1908	-0.4583	31.1957	0.3925	94.0568	0.0273	2.2793	
rs16861329	**ST6GAL1**	0.5712	71.6655	0.627	72.9008	-0.4342	32.0119	0.0219	1.1874	0.6766	95.7771	
rs4411878	**ADAMTS9**	0.2711	60.4521	0.3254	61.9639	1.0763	84.7005	0.0447	5.8463	0.7538	98.5159	
rs6795735	**ADAMTS9**	0.808	79.2215	0.9382	82.4153	1.0763	84.7005	0.0447	5.8463	0.7538	98.5159	
rs1799883	**FABP2**	1.8676	97.1501	1.2921	90.7744	2.3594	97.9336	0.0883	20.7276	0.2391	35.2565	
rs864745	**JAZF1**	-2.0124	2.3878	-1.8675	3.3467	0.2853	57.9177	0.1015	30.0391	0.4681	73.6974	
rs699	**AGT**	1.9184	97.4579	1.9683	98.0818	0.8172	80.0486	0.0663	10.9936	0.6763	96.9349	
rs5051	**AGT**	2.2658	98.8909	2.211	98.9828	0.8172	80.0486	0.0758	14.157	0.362	60.7257	
rs2493132	**AGT**	1.9519	97.6508	1.5127	94.2063	0.8154	79.9981	0.0663	10.9936	0.6763	96.9349	
rs9939609	**FTO**	0.5856	72.1304	0.4723	67.4326	2.6116	99.4569	0.0972	30.6658	0.3271	48.0112	
rs1042713	**ADRB2**	-0.1326	43.9529	0.0422	50.4351	2.1292	97.7143	0.0206	1.0028	0.6632	95.1286	
rs1042714	**ADRB2**	1.6115	95.0118	1.9731	98.0092	2.1292	97.7143	0.0206	1.0028	0.6632	95.1286	
rs12779790	**CDC123**	0.8345	80.4307	0.8643	80.7565	0.5395	68.7751	0.0254	2.2095	0.8033	99.147	
rs10965250	**CDKN2A**	-0.0321	48.0908	0.2331	58.2083	1.6519	94.7322	0.0273	2.6318	0.7337	97.5304	
rs2191349	**DGKB**	0.4617	67.7632	0.3741	63.7124	2.5908	99.0141	0.0314	3.2737	0.5759	87.9229	

Table 10 continued from previous page

ID	Gene	CEU_iHS	p	CEU_nSl	p	CEU_D	p	H12	p	H2_H1	p	Group
rs8050136	**FTO**	-1.1366	12.7956	-1.4964	6.9753	2.4979	99.253	0.0972	30.6658	0.3271	48.0112	
rs7202116	**FTO**	-1.0844	13.8403	-0.8352	19.7999	2.6116	99.4569	0.0972	30.6658	0.3271	48.0112	
rs9930506	**FTO**	1.331	91.297	1.4464	93.2636	2.6116	99.4569	0.0538	11.8376	0.5994	88.6758	
rs17817449	**FTO**	0.5773	71.8231	0.58	71.3475	2.4928	99.2326	0.0972	30.6658	0.3271	48.0112	
rs1421085	**FTO**	0.6736	75.1164	0.5956	71.9037	2.6723	99.5403	0.0336	4.6431	0.6188	90.3144	
rs2373115	**GRB2**	0.8392	80.6733	0.7044	75.6095	-1.5844	4.619	0.4763	97.7134	0.0244	1.7329	
rs2612067	**HMGA2**	2.7834	99.6769	2.7504	99.7448	-1.0902	14.4052	0.4495	96.855	0.166	21.2045	
rs1470579	**IGF2BP2**	-1.473	7.0995	-2.0221	2.3812	-	-	-	-	-	-	
rs7578326	**IRS1**	-1.9363	2.8905	-1.328	9.5174	1.2029	88.5557	0.0218	0.9811	0.775	98.9232	
rs1635852	**JAZF1**	-2.3022	1.1789	-2.1091	1.9471	-0.6477	23.5055	0.2332	78.3535	0.3055	45.7552	
rs7480010	**LOC387761**	-2.7894	0.4512	-2.6656	0.5938	-1.6447	4.0653	0.3416	91.8071	0.0216	1.3534	
rs7227255	**MC4R**	-	-	-	-	-1.7308	3.9373	0.3709	93.5603	0.0162	0.8735	
rs1387153	**MTNR1B**	0.0857	52.7997	0.2575	59.033	-0.3699	31.291	0.012	0.2416	0.7868	99.0847	
rs2568958	**NEGR1**	0.6167	73.244	0.9776	83.7054	2.2484	98.0992	0.2408	76.5629	0.0854	11.3844	
rs662	**PON1**	-2.0003	2.4515	-1.9205	2.9749	1.4362	90.5435	0.0165	0.7866	0.7131	96.7284	
rs1531343	**RPSAP52**	-1.879	3.1774	-1.9319	2.9698	-1.2051	12.1444	0.5111	98.4656	0.2214	30.5205	
rs6536991	**UCP1**	-2.2038	1.5802	-2.0407	2.3954	-0.4061	32.8914	0.2169	70.6221	0.2988	45.428	
rs9472138	**VEGFA**	0.4838	69.3365	0.6955	76.9279	1.1823	85.4212	0.0197	1.4043	0.8311	99.3874	
rs4457053	**ZBED3**	0.9349	83.0265	0.9746	83.8331	0.9458	81.8105	0.0287	2.139	0.6816	96.0373	
rs2722425	**ZMAT4**	0.1475	55.1624	0.4064	64.6846	2.1503	97.6875	0.1193	33.5862	0.224	30.3176	
rs1477196	**FTO**	0.3675	64.0183	0.4573	66.903	2.4611	99.1529	0.0613	14.7643	0.5226	79.5768	
rs3917498	**PON1**	-0.6599	25.1686	-0.7098	23.9072	1.4362	90.5435	0.0116	0.2926	0.9063	99.912	
rs1333049	**CDKN2B**	0.1958	57.2582	0.247	58.7678	1.8557	96.5874	0.0273	2.6318	0.7337	97.5304	
rs11724320	**NPY1R**	1.9459	97.6189	2.2886	99.1827	-0.6538	24.9512	0.263	80.4098	0.3641	56.4875	

Table 10 continued from previous page

ID	**Gene**	CEU_iHS	p	CEU_nSl	p	CEU_D	p	H12	p	H2_H1	p	**Group**
rs7570971	**RAB3GAP1**	1.7306	96.0498	2.3702	99.2456	-2.0824	1.1763	0.6246	99.6321	0.2089	31.3172	
rs7570283	**MCM6**	-1.3399	9.0555	-2.06	2.2879	-1.4636	7.6971	0.6024	99.4613	0.0878	12.179	
rs4988235	**MCM6**	2.8087	99.6986	3.313	99.9484	-1.0086	16.9165	0.5424	98.6588	0.09	12.4805	
rs182549	**MCM6**	-3.3647	0.0878	-3.377	0.0773	-0.8182	21.9142	0.5446	98.6718	0.0957	13.2632	
rs584226	**MCM6**	-1.4474	7.5062	-1.5038	6.9673	-0.6534	26.842	0.5336	98.5151	0.0936	12.9695	
rs16855656	**MCM6**	-1.8714	3.2936	-2.0637	2.2682	-0.9078	19.4768	0.4912	97.5481	0.0992	13.737	
rs309126	**MCM6**	-1.998	2.5359	-2.146	1.8985	-0.9078	19.4768	0.4912	97.5481	0.0992	13.737	
rs3099429	**MCM6**	-1.3847	8.3822	-1.5493	6.3886	-1.1471	13.7207	0.6324	99.6752	0.0651	8.9038	
rs174570	**FADS2**	-	-	-1.0267	14.5109	-2.0419	0.1582	0.3031	99.6416	0.038	1.367	YRI
rs7178572	**HMG20A**	-2.3129	1.3099	-2.6654	0.6168	-1.6792	1.7263	0.1227	82.364	0.4378	60.3298	
rs16861329	**ST6GAL1**	1.0821	86.4446	0.8496	80.5094	-0.0517	77.4748	0.012	2.425	0.699	89.9338	
rs1799883	**FABP2**	2.9931	99.727	2.5812	99.4448	2.5871	99.9232	0.0287	13.3989	0.5294	75.2905	
rs10811661	**CDKN2A/B**	-	-	0.5403	70.6643	-1.0653	19.8607	0.0124	3.4049	0.7762	94.393	
rs9921255	**FTO**	-0.3832	34.1328	-0.3802	33.8843	-0.8481	33.2505	0.0086	1.3623	0.8808	98.1604	
rs7395662	**MADD**	-2.92	0.3557	-2.6695	0.617	0.7985	95.4861	0.1516	86.6275	0.1454	22.3905	
rs1801133	**MTHFR**	-2.0238	2.1578	-2.0032	2.34	-0.3683	63.2867	0.0588	30.9641	0.6058	87.1658	
rs290494	**TCF7L2**	-	-	-	-	-1.0446	20.4003	0.0095	1.4254	0.9277	99.519	
rs1042713	**ADRB2**	-0.4356	32.7151	-0.2065	41.3904	2.0914	99.8186	0.0093	1.0697	0.8419	97.9114	
rs1042714	**ADRB2**	0.1202	55.1026	0.206	58.1055	2.0914	99.8186	0.0093	1.0697	0.8419	97.9114	
rs11642841	**FTO**	-0.6726	23.9764	-0.875	18.1216	-1.0474	22.3332	0.0086	1.3929	0.8767	98.026	
rs1421085	**FTO**	0.157	56.389	0.4391	67.2049	1.1727	98.4764	0.0074	0.7145	0.9407	99.6182	
rs4607517	**GCK**	-0.1159	44.8932	-0.2865	37.8949	-1.4118	7.0097	0.2347	97.7069	0.0222	0.2608	
rs2373115	**GRB2**	2.1008	98.0498	2.161	98.4669	-1.5585	3.4905	0.2157	97.2369	0.0206	0.1044	

Table 10 continued from previous page

ID	**Gene**	CEU_iHS	p	CEU_nSl	p	CEU_D	p	H12	p	H2_H1	p	**Group**
rs7578326	**IRS1**	-0.5402	28.8428	-0.6442	25.425	0.489	93.2032	0.0112	1.8694	0.8683	98.4921	
rs17779747	**KCNJ2**	-	-	-	-	-0.8918	29.5805	0.1909	96.084	0.0402	1.5823	
rs1387153	**MTNR1B**	1.0097	85.0676	1.0175	84.9659	0.0817	82.2533	0.0114	1.9096	0.6799	89.4562	
rs2568958	**NEGR1**	2.2394	98.5444	2.3374	98.9948	1.1582	98.0293	0.02	5.9097	0.4563	73.3692	
rs7593730	**RBMS1/ITGB6**	2.2492	98.5812	2.21	98.6823	-0.0766	78.1825	0.0397	21.4761	0.556	78.3357	
rs6536991	**UCP1**	-0.3875	34.5777	-0.7666	21.6912	0.0847	80.1699	0.0107	2.0018	0.9048	99.0793	
rs9472138	**VEGFA**	-0.1893	42.1219	-0.0571	47.2132	0.0831	79.9443	0.0083	1.048	0.7808	94.4956	
rs3759324	**SCNN1a**	0.7422	77.9367	0.7203	76.8621	-0.1543	72.9554	0.0067	0.3212	0.936	99.6006	
rs1525791	**POU6F2**	-2.5122	0.8267	-2.6241	0.6403	-1.4658	5.6696	0.2266	97.3189	0.0713	7.0722	
rs3917498	**PON1**	-0.103	45.4389	0.0048	49.5681	0.9782	96.81	0.0105	2.2077	0.8524	97.6131	
rs917793	**GCK**	0.0424	51.5519	-0.0281	48.223	-1.631	2.802	0.2347	97.7069	0.0222	0.2608	
rs17685538	**TCF7L2**	-	-	-	-	-2.4039	0.3369	0.6367	99.1207	0.0137	1.091	CHS
rs7895340	**TCF7L2**	-	-	-	-	-2.4039	0.3369	0.6429	99.2109	0.3013	38.9481	
rs7923837	**HHEX**	2.4184	99.3333	2.6834	99.7099	-2.0124	2.1694	0.508	96.2762	0.044	4.8098	
rs2641348	**ADAM30**	-	-	-	-	-1.9161	3.2345	0.59	98.1792	0.0056	0.2369	
rs10923931	**NOTCH2**	-	-	-	-	-2.0575	1.9722	0.6416	99.1405	0.006	0.2781	
rs1493694	**NOTCH2**	-	-	-	-	-2.0373	2.1602	0.634	99.0205	0.006	0.2695	
rs4923461	**BDNF**	2.2064	98.7405	1.8005	96.9462	1.167	82.3843	0.2041	63.4998	0.286	37.3018	
rs11037909	**EXT2**	1.9635	97.7105	1.8242	97.1257	-1.1695	11.5011	0.0723	11.8695	0.6606	94.1468	
rs7178572	**HMG20A**	-2.3452	1.2232	-2.5071	0.8017	-1.1497	12.1561	0.2974	83.7726	0.4125	55.5872	
rs13373826	**SLC44A5**	-	-	-	-	-1.6036	7.0453	0.5512	97.1535	0.1003	12.6284	
rs368794	**KCTD15**	2.2351	98.8344	2.089	98.4615	1.3468	84.9958	0.0725	21.7272	0.5627	82.5342	
rs11899863	**THADA**	-	-	-	-	-2.3834	0.6729	0.6565	98.9804	0.0043	0.197	

Table 10 continued from previous page

ID	**Gene**	CEU_iHS	p	CEU_nSl	p	CEU_D	p	H12	p	H2_H1	p	**Group**
rs7578597	**THADA**	-	-	-	-	-2.3207	0.9751	0.5907	97.7716	0.0528	7.7624	
rs17036101	**PPARG**	-	-	-	-	-0.7919	23.7941	0.0265	1.5437	0.8726	99.8791	
rs2314349	**MCF2L2**	-1.4435	7.6897	-1.5896	5.9176	3.0564	99.3604	0.1318	33.2079	0.5311	80.875	
rs1799883	**FABP2**	1.2378	89.9838	0.9419	82.791	3.8525	99.8351	0.1276	32.1578	0.1951	24.3489	
rs162036	**MTRR**	0.949	83.3565	0.9473	82.8319	0.9249	78.9446	0.0208	0.8238	0.6822	95.7947	
rs7754840	**CDKAL1**	0.1551	54.9516	0.0917	52.1391	3.0172	99.1421	0.0839	16.6479	0.5665	84.9828	
rs896854	**PLEKHF2**	-2.4232	0.9763	-2.6351	0.5702	0.9861	78.3804	0.3736	87.2209	0.2489	30.6815	
rs10946398	**CDKAL1**	0.1849	56.1886	0.068	51.1619	3.0172	99.1421	0.0839	16.6479	0.5665	84.9828	
rs5050	**AGT**	-1.3328	9.3637	-0.604	27.0633	2.7942	98.9925	0.0311	1.9889	0.8406	99.8089	
rs699	**AGT**	1.2137	89.3675	1.0182	84.6118	2.2371	97.2665	0.0311	1.9889	0.8406	99.8089	
rs5051	**AGT**	2.1351	98.5129	1.2108	89.107	2.7942	98.9925	0.0311	1.9889	0.8406	99.8089	
rs2493132	**AGT**	0.1471	55.02	0.1335	54.1018	1.766	94.0367	0.0311	1.9889	0.8406	99.8089	
rs4762	**AGT**	-1.3355	9.3247	-0.9935	16.327	2.2371	97.2665	0.0311	1.9889	0.8406	99.8089	
rs5046	**AGT**	1.3914	92.3929	1.0961	86.5411	2.7942	98.9925	0.0311	1.9889	0.8406	99.8089	
rs5049	**AGT**	1.3914	92.3929	1.0961	86.5411	2.7942	98.9925	0.0311	1.9889	0.8406	99.8089	
rs7079	**AGT**	-0.1003	45.0628	0.0622	51.3164	1.6531	92.9275	0.0452	4.3822	0.7531	98.8659	
rs943580	**AGT**	-0.2344	39.6853	-0.5341	29.2698	1.6531	92.9275	0.0452	4.3822	0.7531	98.8659	
rs7901695	**TCF7L2**	-	-	-	-	-2.455	0.2255	0.6816	99.5823	0.0154	1.3014	
rs7903146	**TCF7L2**	-	-	1.704	96.3088	-2.455	0.2255	0.7368	99.8672	0.0128	0.9782	
rs11196205	**TCF7L2**	-	-	-	-	-2.4039	0.3369	0.658	99.3913	0.313	41.1075	
rs4506565	**TCF7L2**	-	-	1.704	96.3088	-2.455	0.2255	0.7368	99.8672	0.0128	0.9782	
rs7924080	**TCF7L2**	-	-	-	-	-2.5227	0.1065	0.7937	99.9699	0.0018	0.0125	
rs11196199	**TCF7L2**	-	-	-	-	-2.3253	0.5649	0.5453	97.3534	0.0698	7.3838	
rs12255372	**TCF7L2**	-	-	-	-	-2.4039	0.3369	0.658	99.3913	0.313	41.1075	

Table 10 continued from previous page

ID	**Gene**	CEU_iHS	p	CEU_nSl	p	CEU_D	p	H12	p	H2_H1	p	**Group**
rs10885406	**TCF7L2**	-	-	1.8877	97.6704	-2.4664	0.2004	0.7282	99.8322	0.006	0.2593	
rs7079711	**TCF7L2**	-	-	-	-	-2.1088	1.5882	0.3428	87.6204	0.4455	64.906	
rs11196181	**TCF7L2**	-	-	-	-	-2.455	0.2255	0.6574	99.3744	0.0074	0.382	
rs17747324	**TCF7L2**	-	-	0.1113	53.2091	-2.455	0.2255	0.6574	99.3744	0.0074	0.382	
rs7896811	**TCF7L2**	-	-	-	-	-2.2713	0.7816	0.7368	99.8672	0.0128	0.9782	
rs11196192	**TCF7L2**	-	-	-	-	-2.5227	0.1065	0.7937	99.9699	0.0018	0.0125	
rs11196213	**TCF7L2**	-	-	-	-	-2.2777	0.7528	0.6051	98.6322	0.3113	40.7919	
rs1042713	**ADRB2**	0.2749	59.8511	0.5647	70.3883	1.9333	95.139	0.0244	1.1485	0.7531	98.5452	
rs1042714	**ADRB2**	-	-	0.7503	76.7527	1.9333	95.139	0.0244	1.1485	0.7531	98.5452	
rs6265	**BDNF**	2.3885	99.2142	2.0941	98.5415	-0.3783	31.0006	0.0974	20.8973	0.5189	78.0891	
rs11257622	**CDC123**	0.7604	77.8473	1.0601	85.8643	0.6254	66.7447	0.0637	11.4421	0.749	98.0579	
rs10965250	**CDKN2A**	0.946	83.1187	0.9442	82.5471	1.6358	92.6776	0.0223	1.3903	0.6839	95.1991	
rs2191349	**DGKB**	-0.4064	33.3559	-0.4689	31.5353	2.7061	98.7066	0.039	4.459	0.2347	28.9991	
rs11642841	**FTO**	-	-	-	-	0.7593	72.944	0.0123	0.3466	0.6588	92.7749	
rs3923113	**GRB14**	-1.3455	9.2256	-1.4736	7.507	-1.874	4.7696	0.585	97.6484	0.0433	6.4554	
rs2373115	**GRB2**	-1.0788	13.9743	-1.4486	7.7055	-0.3713	31.2502	0.5478	97.6268	0.2206	26.4075	
rs5015480	**HHEX**	1.7558	96.5361	2.0653	98.5528	-0.9027	16.7742	0.0687	13.1812	0.4662	68.5221	
rs1470579	**IGF2BP2**	-2.4067	0.9633	-2.6854	0.4853	-	-	-	-	-	-	
rs7578326	**IRS1**	0.4368	66.5948	0.1304	53.9604	1.306	87.7974	0.0323	2.0473	0.7695	98.9091	
rs5215	**KCNJ11**	0.7577	77.8434	0.8755	80.8562	2.5735	98.277	0.2283	69.9073	0.2893	37.7974	
rs17779747	**KCNJ2**	0.6517	74.2738	0.9257	82.0946	2.8226	99.328	0.3208	83.0091	0.4925	72.1564	
rs2074196	**KCNQ1**	1.0354	85.5758	1.0486	85.5977	1.5028	89.2083	0.0096	0.0649	0.9011	99.9629	
rs7227255	**MC4R**	-	-	-	-	-2.1639	1.1227	0.5547	97.7479	0.0567	5.9269	
rs3812316	**MLXIPL**	-1.2412	10.7965	-1.2747	10.4118	-1.5464	6.5305	0.5288	96.836	0.0168	1.48	

Table 10 continued from previous page

ID	**Gene**	CEU_iHS	p	CEU_nSl	p	CEU_D	p	H12	p	H2_H1	p	**Group**
rs1387153	**MTNR1B**	0.4261	66.2353	0.6684	74.2121	1.687	92.0598	0.0152	0.3714	0.756	98.311	
rs662	**PON1**	-2.0583	2.1975	-1.6008	5.9006	1.4189	87.2486	0.021	1.1073	0.7083	96.4363	
rs9472138	**VEGFA**	-0.0228	47.5032	0.2166	57.0572	1.531	89.481	0.0283	2.2095	0.3393	44.5011	
rs3759324	**SCNN1a**	0.3242	62.4407	0.209	57.1247	0.9446	77.4227	0.0111	0.2611	0.6327	91.0866	
rs3917498	**PON1**	-0.683	24.1123	-0.8765	19.1139	1.4189	87.2486	0.0149	0.4368	0.8367	99.6396	
rs1333049	**CDKN2B**	0.3768	64.278	0.2817	60.0294	1.6358	92.6776	0.0223	1.3903	0.6839	95.1991	
rs4812829	**HNF4A**	1.2684	90.1085	1.4204	92.7066	0.8934	76.5649	0.4244	92.0508	0.3842	52.4235	
rs11724320	**NPY1R**	1.7791	96.6829	1.8696	97.6183	0.9175	74.8543	0.0295	2.0895	0.4895	72.989	
rs1359790	**SPRY2**	-0.4872	30.4705	-0.6012	27.1765	0.9104	75.1553	0.0246	1.482	0.6285	90.8094	
rs17584499	**PTPRD**	-0.323	36.4294	-0.2706	38.6921	0.6996	72.2566	0.0301	2.7392	0.66	93.7119	
rs12778366	**SIRT1**	0.4884	68.3809	0.263	59.094	1.5	89.2794	0.0507	7.5097	0.7433	97.8869	
rs3758391	**SIRT1**	-2.5836	0.7371	-1.8876	3.2177	1.5	89.2794	0.0507	7.5097	0.7433	97.8869	
rs6780569	**UBE2E2**	1.0818	86.5235	1.0019	84.1423	1.725	93.5887	0.0334	2.5472	0.5009	75.5251	
rs671	**ALDH2**	2.0782	98.2262	2.5336	99.5016	0.9187	76.717	0.0836	19.5958	0.4955	73.0446	
rs1229984	**ADH1Bb**	1.97	97.8985	2.1934	98.998	-1.0393	14.8032	0.5063	95.4703	0.061	7.5931	

4.3.3 Average expected heterozygosity D_e

SNPs associated with the TCF7L2 gene

The lowest gene diversity was estimated for the SNPs in the TCF7L2 gene of the African reference group (D_e: 0.2255±0.0062), followed by the hunter-gatherers of Eastern Europe/Russia (D_e: 0.2757±0.0273). The highest diversity was found for the Chinese reference group (D_e: 0.0485±0.0041). For all remaining European groups, similarly elevated values were estimated, with the Central European hunter-gatherers at the lower end (WHG - D_e: 0.3135±0.0194, SEF - D_e: 0.3513±0.0259, CEF - D_e: 0.3275±0.0165, MN - D_e: 0.3576±0.0314, WEZ - D_e: 0.3533±0.0230, CEU - D_e: 0.3558±0.0070). The diversity in the African reference groups was significantly lower compared to all other groups ($p \leq 0.0010$) except for the Eastern European/Russian hunter-gatherers. The diversity in the Chinese reference group was higher compared to every other group ($p \leq 0.0001$). Differences among the European groups were only found between the Central European reference group and the Eastern European/Russian hunter-gatherers ($p \leq 0.0500$, for details see figure 28).

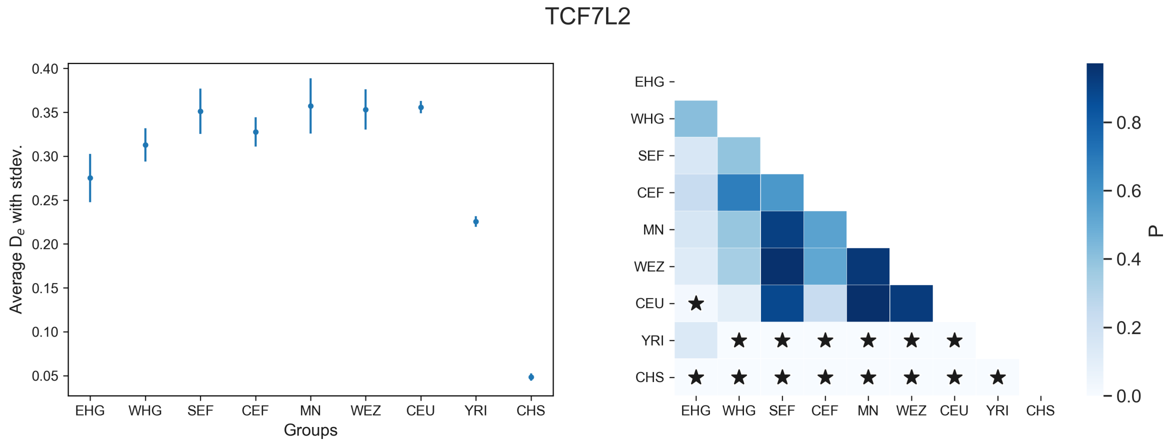

Figure (28) *Average expected heterozygosity D_e for all groups for the SNPs in the TCF7L2 gene. P-values for differences between groups are displayed in the heatmap. Significant values are highlighted by black stars.*

SNPs associated with the FTO gene

For the SNPs in the FTO gene, the highest diversity was found for the European groups (EHG - D_e: 0.3328±0.0467, WHG - D_e: 0.4309±0.0178, SEF - D_e: 0.4258±0.0292, CEF - D_e: 0.4174±0.0258, MN - D_e: 0.3542±0.0456, WEZ - D_e: 0.4318±0.0231, CEU - D_e: 0.4677±0.0060). While the diversity was also elevated in the African reference group (YRI - D_e: 0.3277±0.0086), it was lowest in the Chinese reference group (CHS - D_e: 0.2320±0.0126). The diversity in the African reference group was significantly lower compared to all other groups ($p \leq 0.0500$), except the Middle Neolithic group and the hunter-gatherers of Eastern Europe/Russia. A similar result was found for the Chinese reference group ($p \leq 0.0500$), except that here the diversity was also lower compared to the Middle Neolithic group. Between the European groups, significant differences were found only between the Central European reference group and the Eastern European hunter-gatherers ($p \leq 0.0500$), and the Central European reference group and the Middle neolithic group ($p \leq 0.0500$, for details see figure 29).

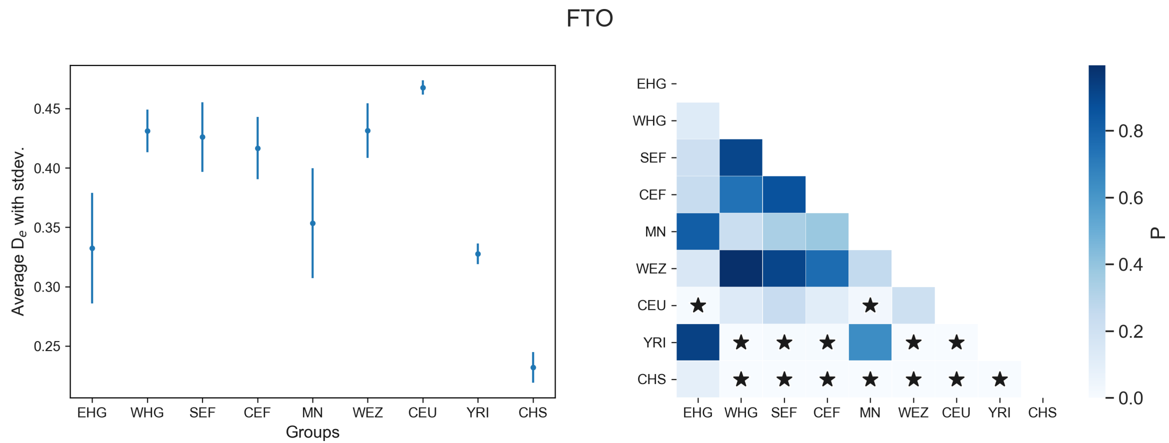

Figure (29) *Average expected heterozygosity D_e for all groups for the SNPs in the FTO gene. P-values for differences between groups are displayed in the heatmap. Significant values are highlighted by black stars.*

SNPs associated with the AGT gene

For the AGT gene, the highest diversity was found in the European groups (EHG - D_e: 0.3541±0.0424, WHG - D_e: 0.3229±0.0399, SEF - D_e: 0.3089±0.0545, CEF - D_e: 0.3545±0.0318, MN - D_e: 0.3564±0.0486, WEZ - D_e: 0.3289±0.0428, CEU - D_e: 0.3484±0.0098). In both non-European reference groups a comparatively lower diversity was found (YRI - D_e: 0.2224±0.0112, CHS - D_e: 0.2662±0.0127). BWhile the European groups did not differ significantly, the diversity in the African reference group was found to be significantly lower in comparison to all European groups ($p \leq 0.0500$), except the farmers from the Aegean/Balkan region. The Chinese group differed in comparison to the Central European reference group ($p \leq 0.0005$), as well as to the farmers from the Aegean/Balkan region ($p \leq 0.0500$, for details see figure 30).

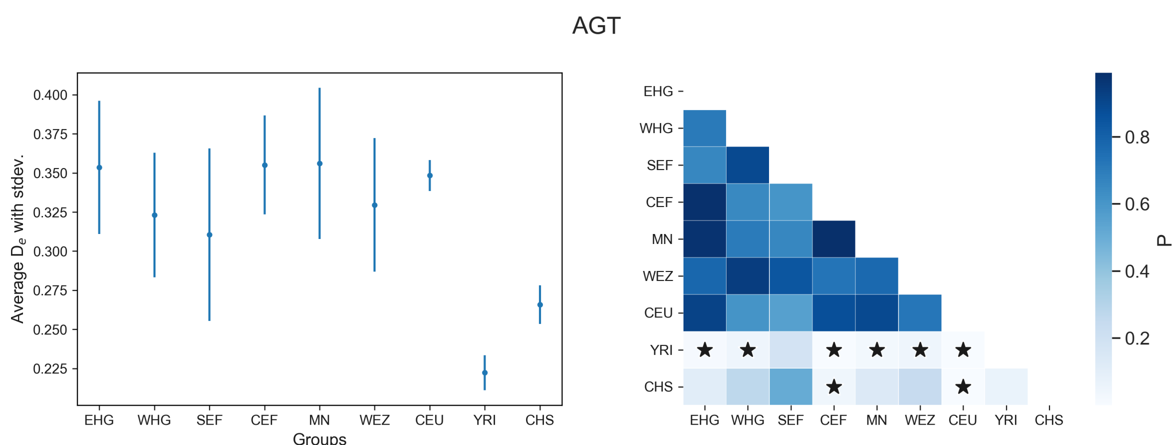

Figure (30) *Average expected heterozygosity D_e for all groups for the SNPs in the AGT gene. P-values for differences between groups are displayed in the heatmap. Significant values are highlighted by black stars.*

SNPs associated with the MCM6 gene

For the MCM6 gene, the diversity was estimated to be similar in all groups, except for the Cen-

tral European and the African reference groups, where a lower diversity was found (EHG - D_e : 0.3200±0.0471, WHG - D_e: 0.3119±0.0383, SEF - D_e: 0.3953±0.0489, CEF - D_e: 0.3189±0.0287, MN - D_e: 0.3468±0.0531, WEZ - D_e: 0.3596±0.0458, CHS - D_e: 0.3480±0.0054, CEU - D_e: 0.2300±0.0138, YRI - D_e: 0.2741±0.0098). Between the ancient sample groups, no significant differences could be found, while the Central European reference group differed from both early Neolithic farmer groups (SEF: $p \leq 0.0010$, CEF: $p \leq 0.0500$), the Bronze Age group ($p \leq 0.0500$) and the Chinese reference group ($p \leq 0.0001$). Furthermore, the African reference differed from the Chinese reference group ($p \leq 0.0001$) and from the early farmers from the Aegean/Balkan region ($p \leq 0.0500$, for details see figure 31).

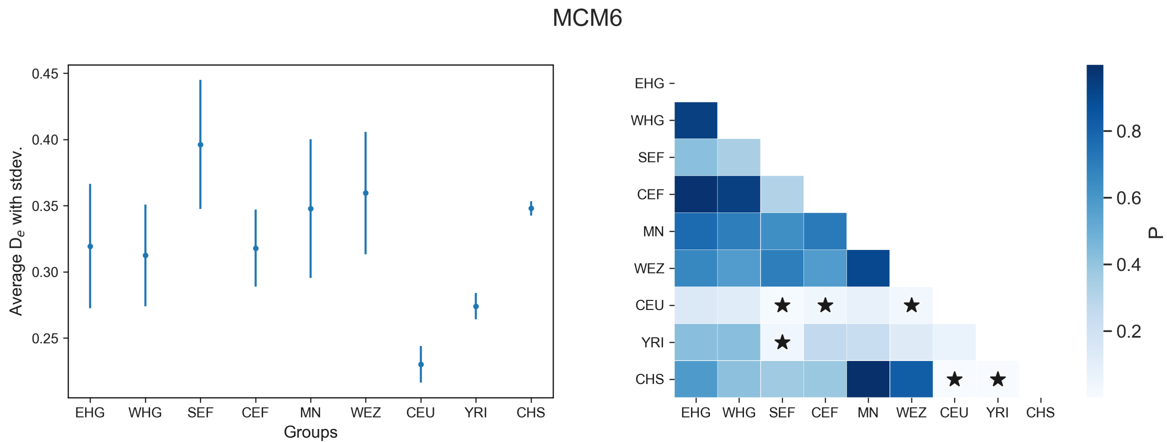

Figure (31) *Average expected heterozygosity D_e for all groups for the SNPs in the MCM6 gene. P-values for differences between groups are displayed in the heatmap. Significant values are highlighted by black stars.*

4.3.4 Pairwise F_{st} comparisons

SNPs associated with the TCF7L2 gene

A significantly higher average F_{st} across the functional SNPs compared to the neutral SNPs was found between the Eastern European/Russian hunter-gatherers and the Central European reference group (diff: 0.0995 $p \leq 0.0001$), the Central European farmer group (0.0636 $p \leq 0.0500$), the farmer group from the Aegean/Balkan region (diff: 0.0636 $p \leq 0.0500$), the Bronze Age group (diff: 0.1444 $p \leq 0.0001$) and the African reference group (diff: 0.2223 $p \leq 0.0001$). The onyl other group with an exclusive European pairing with a significant higher average functional F_{st} consisted of the Central European hunter-gatherers and the farmers from the Aegean/Balkan region (diff: 0.0535 $p \leq 0.0500$). All other pairings with significant differences included the Chinese reference group were the average functional F_{st} was always greater than the average neutral F_{st} (YRI - diff: 0.2268 $p \leq 0.0001$; CEU - diff: 0.1825 $p \leq 0.0001$; WEZ - diff: 0.2259 $p \leq 0.0001$; CEF - diff: 0.1465 $p \leq 0.0001$; SEF - diff: 0.1207 $p \leq 0.0050$).

SNPs associated with the FTO gene

Three pairings were found with significant differences in the pairwise comparisons at the FTO loci.

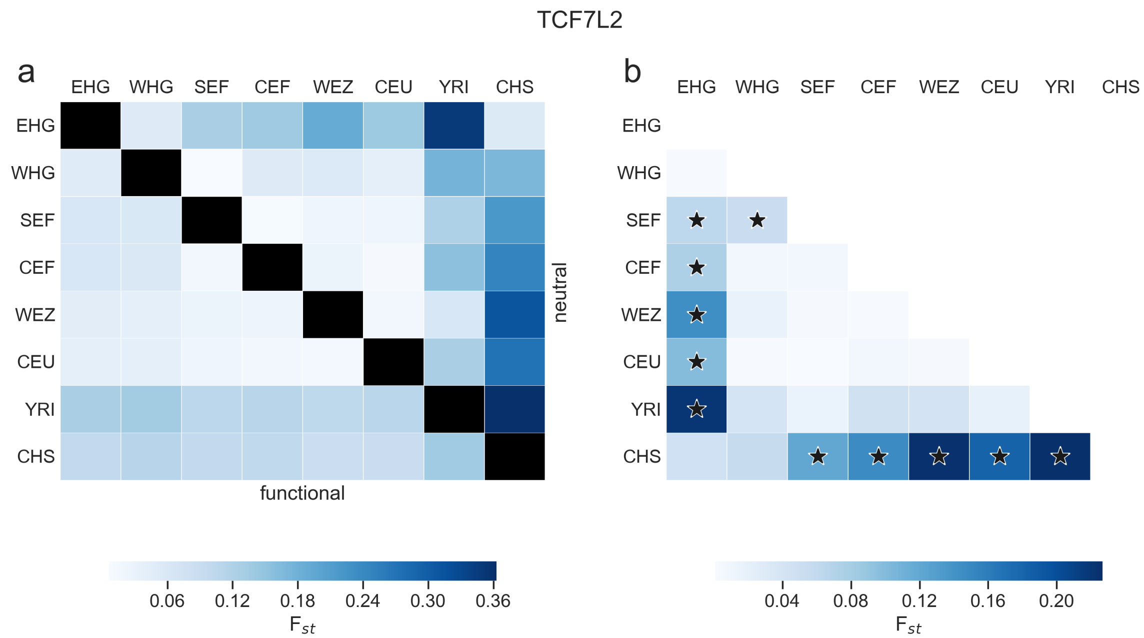

Figure (32) *a) Pairwise F_{st} comparison between SNPs in the TCF7L2 gene and SNPs randomly drawn from the neutralome. b) Absolute differences between F_{st} values. Significant values are highlighted by a black star.*

In all pairings, the F_{st} across the functional SNPs was found to be significantly higher and the pairings always included the Chinese reference group (CEU - diff: 0.0896, $p \leq 0.0100$; WEZ - diff: 0.1383, $p \leq 0.0010$; SEF - diff: 0.1042, $p \leq 0.0500$).

SNPs associated with the AGT gene

All pairings for which significantly higher F_{st} values were found for the functional SNPs in the AGT gene, included either the Chinese (WHG - diff: 0.2057, $p \leq 0.0001$; CEF - diff: 0.1063, $p \leq 0.0100$; CEU - diff: 0.1031, $p \leq 0.0050$; EHG - diff: 0.1693, $p \leq 0.0050$; YRI - diff: -0.1175, $p \leq 0.0010$) or the African reference group (CEF - diff: 0.1391, $p \leq 0.0050$; CEU - diff: 0.1323, $p \leq 0.0050$; EHG - diff: 0.1930, $p \leq 0.0050$; WHG - diff: 0.2284, $p \leq 0.0001$). None of the pairings between the European groups showed any significant differences.

SNPs associated with the MCM6 gene

All pairings for which significant higher F_{st} values were found for the functional SNPs in the AGT gene, included either the Chinese (WHG - diff: 0.2057, $p \leq 0.0001$; CEF - diff: 0.1063, $p \leq 0.0100$; CEU - diff: 0.1031, $p \leq 0.0050$; EHG - diff: 0.1693, $p \leq 0.0050$; YRI - diff: -0.1175, $p \leq 0.0010$) or the African reference group (CEF - diff: 0.1391, $p \leq 0.0050$; CEU - diff: 0.1323, $p \leq 0.0050$; EHG - diff: 0.1930, $p \leq 0.0050$; WHG - diff: 0.2284, $p \leq 0.0001$). None of the pairings between the European groups showed any significant differences.

SNPs related to obesity

The investigation of the obesity related SNPs resulted in 18 pairings where the average functional F_{st} was significantly higher compared to the average neutral F_{st}. The Chinese reference group was

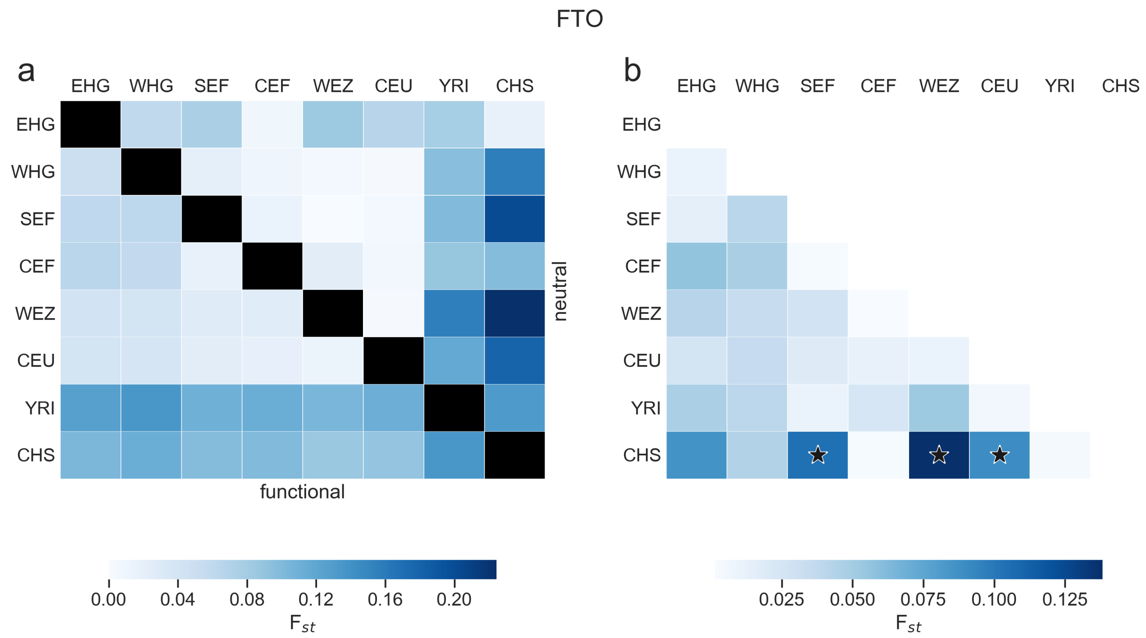

Figure (33) *a) Pairwise F_{st} comparison between SNPs in the FTO gene and SNPs randomly drawn from the neutralome. b) Absolute differences between F_{st} values. Significant values are highlighted by a black star.*

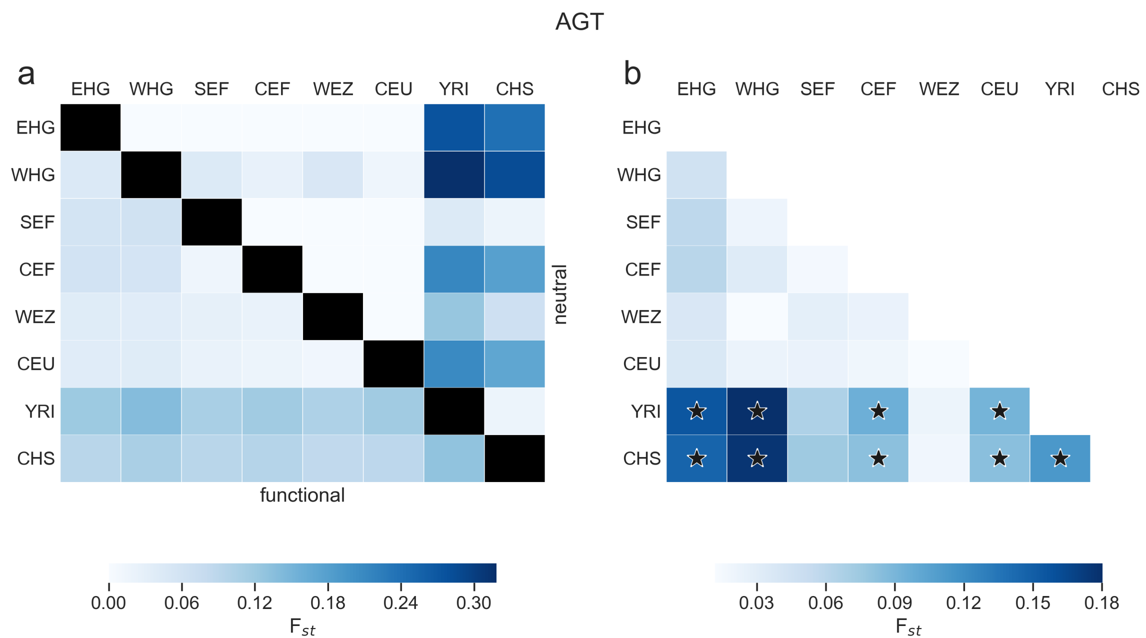

Figure (34) *a) Pairwise F_{st} comparison between SNPs in the AGT gene and SNPs randomly drawn from the neutralome. b) Absolute differences between F_{st} values. Significant values are highlighted by a black star.*

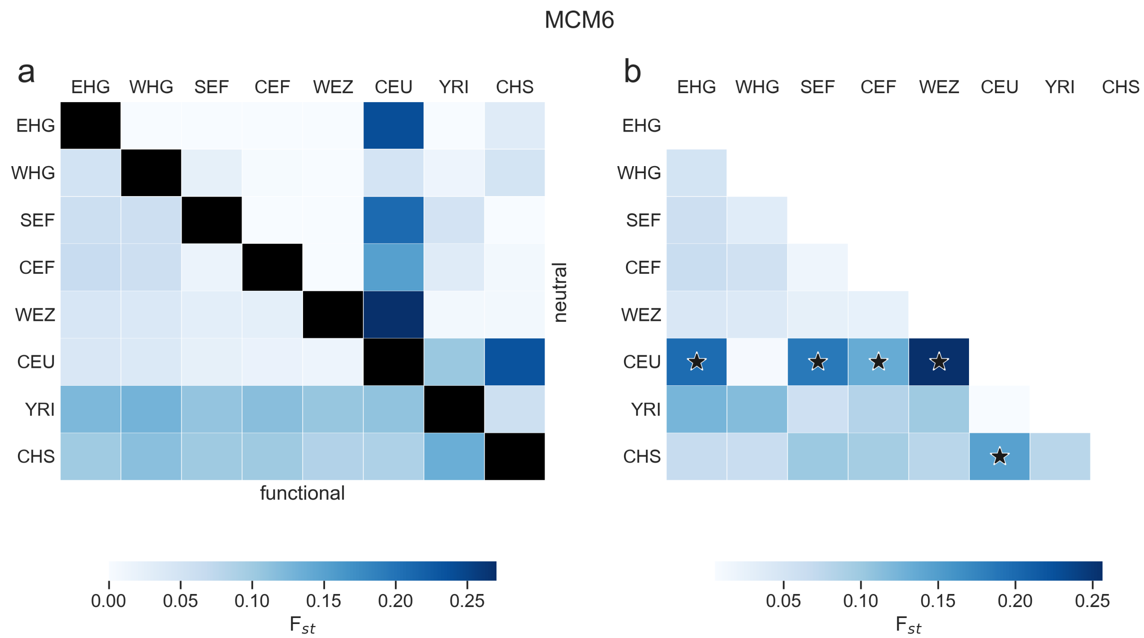

Figure (35) *a) Pairwise F_{st} comparison between SNPs in the MCM6 gene and SNPs randomly drawn from the neutralome. b) Absolute differences between F_{st} values. Significant values are highlighted by a black star.*

found to differ in F_{st} values compared to all groups, except the Central European farmers (EHG - diff: 0.1010, $p \leq 0.0010$; WHG - diff: 0.0882, $p \leq 0.0050$; SEF - diff: 0.1189, $p \leq 0.0001$; WEZ - diff: 0.0684, $p \leq 0.0050$; CEU - diff: 0.0474, $p \leq 0.0500$; YRI diff: 0.0611, $p \leq 0.0500$). The remaining 12 pairings all consisted of groups of European origin. Four pairings included the Central European reference (EHG - diff: -0.0634, $p \leq 0.0000$; CEF - diff: -0.0109, $p \leq 0.0500$; SEF - diff: -0.0180, $p \leq 0.0500$; WHG - diff: -0.0450, $p \leq 0.0001$), while the rest consisted exclusively of ancient sample groups. Of those parings, three included the Central European hunter-gatherers and one of the farmer groups (WEZ - diff: -0.0258, $p \leq 0.0500$; CEF - diff: -0.0943, $p \leq 0.0001$; SEF - diff: -0.0750, $p \leq 0.0050$), while two included the Eastern European/Russian hunter-gatherers and one of the farmer groups (WEZ - diff: -0.0648, $p \leq 0.0010$; CEF - diff: -0.0489, $p \leq 0.0500$). Furthermore, all comparisons between the farmer groups resulted in significant differences (WEZ - SEF diff: -0.0363, $p \leq 0.0050$; SEF - CEF diff: -0.0241, $p \leq 0.0500$; WEZ - CEF diff: -0.0283, $p \leq 0.0050$).

SNPs associated with type 2 diabetes

The assessment of the SNPs exclusively associated with T2D resulted in 15 pairings with significant higher F_{st} values across the functional SNPs compared to the neutral SNPs. Six of those pairings included the Chinese reference group (CEU - diff: 0.0380, $p \leq 0.0500$; YRI - diff: 0.0781, $p \leq 0.0100$; WHG - diff: 0.1246, $p \leq 0.0001$; WEZ - diff: 0.0514, $p \leq 0.0500$; EHG - diff: 0.1384, $p \leq 0.0001$; CEF - diff: 0.0566, $p \leq 0.0500$), three included the Central European reference group (WHG - diff: 0.0625, $p \leq 0.0001$; EHG - diff: 0.0832, $p \leq 0.0001$; CEF - diff: 0.0170, $p \leq 0.0100$), and an additional two pairings included the African reference group (EHG - diff: 0.0653, $p \leq 0.0500$; WEZ - diff: 0.0578, $p \leq 0.0500$). Four pairings consisted exclusively of ancient sample groups; three included the Bronze Age group (EHG - diff: 0.0430, $p \leq 0.0100$; WHG - diff: 0.0477, $p \leq 0.0005$; CEF - diff: 0.0186,

Figure (36) *a) Pairwise F_{st} comparison between SNPs related to obesity and SNPs randomly drawn from the neutralome. b) Absolute differences between F_{st} values. Significant values are highlighted by a black star.*

p≤0.0500), while one was between the early farmers and the hunter-gatherers of Central Europe (diff: 0.0705, p≤0.0001).

SNPs associated with hypertension

For six pairings significantly higher F_{st} across the functional SNPs were found compared to the neutral SNPs. Only one pairing consisted exclusively of ancient sample groups (CEF - WHG diff: -0.1201, p≤0.0500), while the rest always consisted of at least one reference group. One pairing included the African reference (CEF - diff: -0.1953, p≤0.0500), while the rest always included the Central European reference (YRI - diff: -0.2126, p≤0.0100; WHG - diff: -0.1331, p≤0.0001; CHS - diff: -0.1661, p≤0.0500; CEF - diff: -0.0454, p≤0.0100).

4.3.5 Mean difference of risk allele frequencies

When considering a subset of obesity-related SNPs, an higher average frequency in obesity risk alleles was found for the Central European hunter-gatherer group (diff: -0.0850 ± 0.0185, p≤0.0001), while a lower average frequency was found in the Central European farmer group (diff: 0.0494 ± 0.0160, p≤0.0050), compared to the Central European reference group (see figure: 39). The Central European hunter-gatherer group was furthermore found to have an higher average frequency of risk alleles compared to the African (diff: -0.0900 ± 0.0187, p≤0.0001) as well as the Chinese reference (WHG diff: -0.1061 ± 0.0188, p≤0.0001).

By comparing the frequencies of a subset of 27 T2D related SNPs in each of the ancient sample groups to the Central European reference group, only the hunter-gatherer groups were found to have a sig-

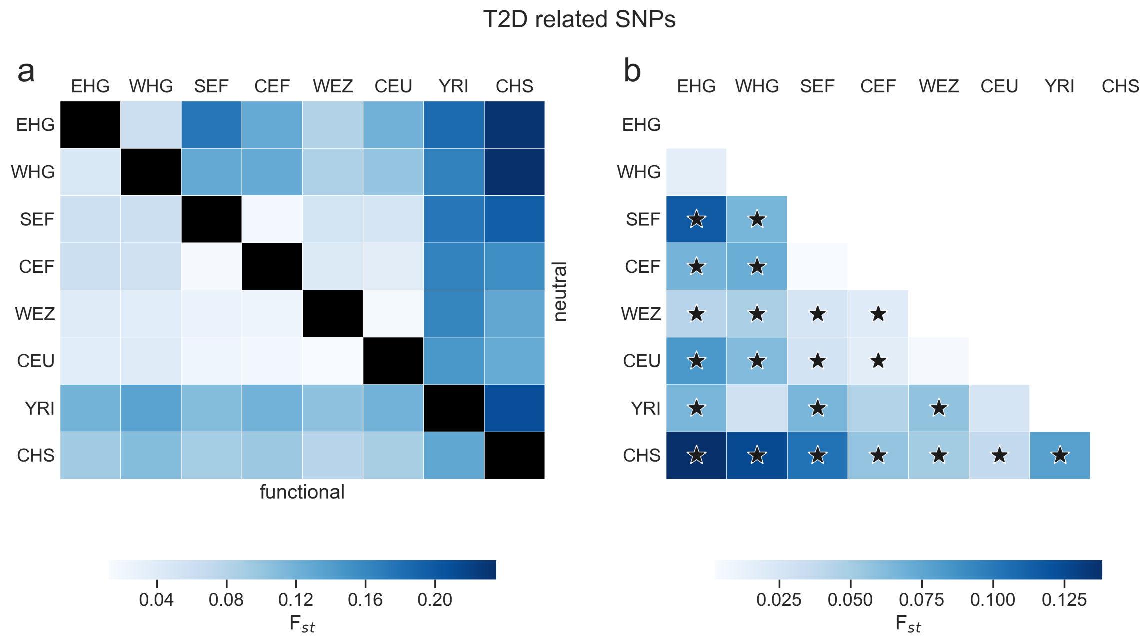

Figure (37) *a) Pairwise F_{st} comparison between SNPs related to T2D and SNPs randomly drawn from the neutralome. b) Absolute differences between F_{st} values. Significant values are highlighted by a black star.*

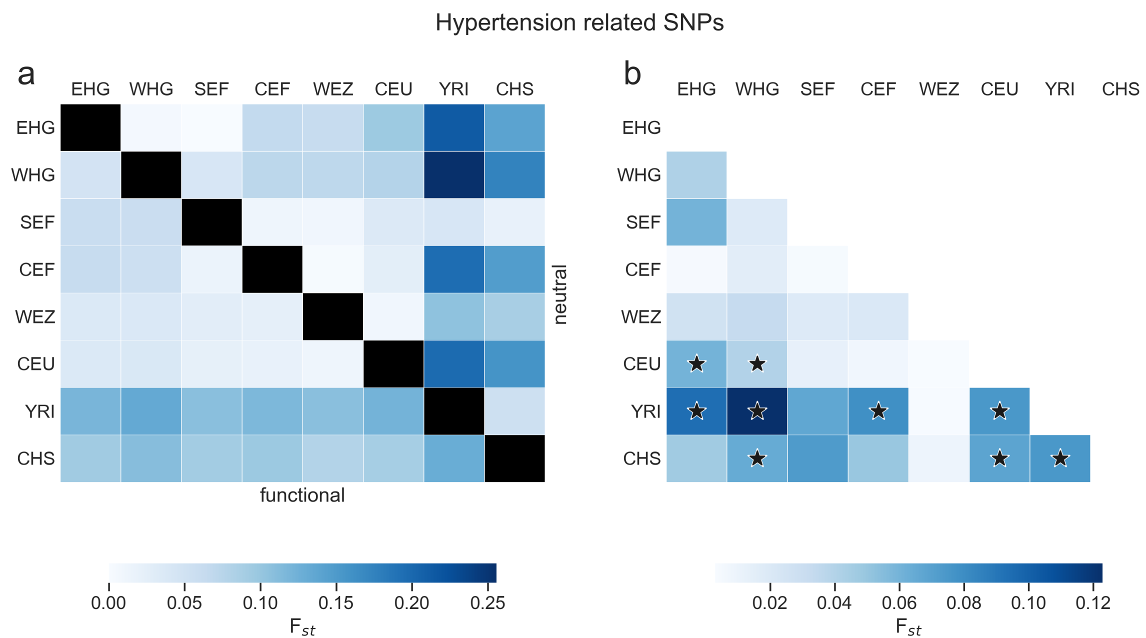

Figure (38) *a) Pairwise F_{st} comparison between SNPs related to hypertension and SNPs randomly drawn from the neutralome. b) Absolute differences between F_{st} values. Significant values are highlighted by a black star.*

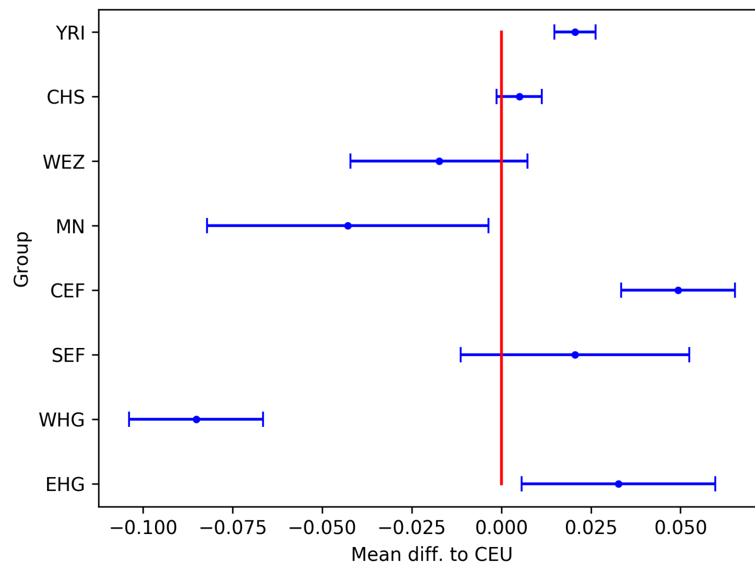

Figure (39) *Results of iterative approach to determine the mean difference of risk allele frequencies of all obesity related SNPs between all ancient sample groups and the reference groups.*

nificantly lower average risk allele frequency (EHG - diff: 0.0627±0.0267, p≤0.0500; WHG - diff: 0.0346±0.0173, p≤0.0500). The same was true for the comparison between the ancient sample groups and the Chinese reference group (EHG - diff: 0.0671±0.0265, p≤0.0500; WHG - diff: 0.0396±0.0176, p≤0.0500). In comparison to the African reference group, all groups but the early farmer group from the Aegean/Balkan region and the Middle Neolithic group, had a significantly lower average risk allele frequency (EHG - diff: 0.1322±0.0263, p≤0.0001; WHG - diff: 0.1037±0.0175, p≤0.0001; CEF - diff: 0.0742±0.0166, p≤0.0001; WEZ - diff: 0.0982±0.0240, p≤0.0001; see figure: 40).

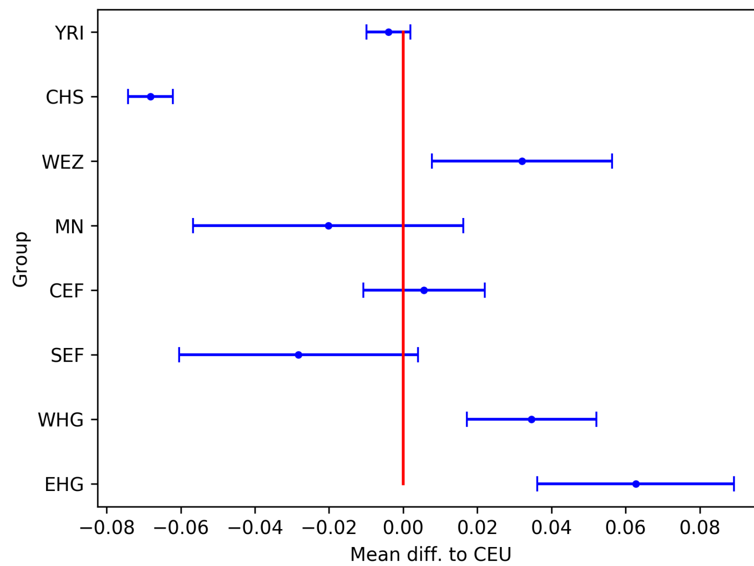

Figure (40) *Results of iterative approach to determine the mean difference of risk allele frequencies of all obesity related SNPs between all ancient sample groups and the reference groups.*

4.3.6 Genetic risk score determination

Three genetic risk scores according to Meigs et al. [2008], Cornelis et al. [2009], Hivert et al. [2011] were determined for each sample in the reference and the ancient data. For the African group, significantly higher average risk scores determined according to Cornelis et al. [2009] and Hivert et al. [2011], were found. Risk scores were non-significant if the African reference group was excluded from the analysis (table: 11). None of the group pairings showed significant dissimilarities for the risk score determined according to Meigs et al. [2008]. When comparing the distribution of risk score results between each ancient sample group and the Central European reference group, significant differences were found for the Middle Neolithic group for the Cornelis et al. [2009] and Meigs et al. [2008] risk scores ($p \leq 0.05$ and $p \leq 0.005$ respectively) and the Central European hunter-gatherer group for theHivert et al. [2011] risk score ($p \leq 0.05$ see table: 12 as well as figure: 41 - 43).

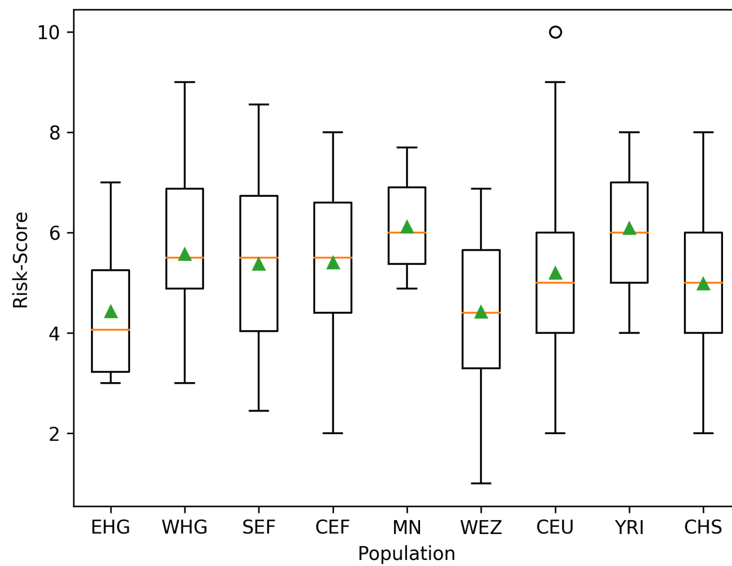

Figure (41) *Box plot of risk score results, determined as described in Cornelis et al. [2009]. Orange line=mean score, green triangle=median score*

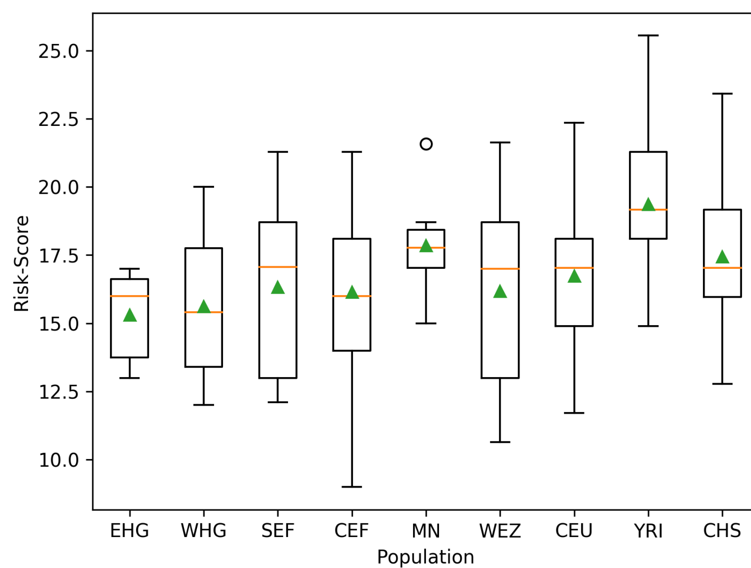

Figure (42) *Box plot of risk score results, determined as described in Hivert et al. [2011]. Orange line=mean score, green triangle=median score*

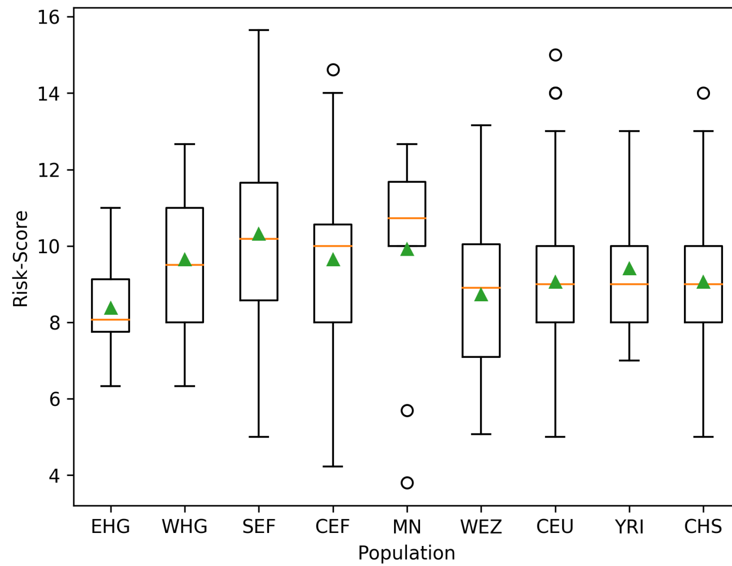

Figure (43) *Box plot of risk score results, determined as described in Meigs et al. [2008]. Orange line=mean score, green triangle=median score*

Table (11) *Results of ANOVA for the three different risk scores*

Group	Stat	P	
All	6.5798	0.0000	Cornelis et al. [2009]
Non-African	1.9556	0.0610	
Europe	1.7036	0.1223	
All	13.9759	0.0000	Hivert et al. [2011]
Non-African	2.9544	0.0053	
Europe	1.3194	0.2504	
All	1.7359	0.0884	Meigs et al. [2008]
Non-African	1.6638	0.1175	
Europe	1.4999	0.1801	

Table (12) *Results of Mann-Whitney U test for the three different risk scores*

Group	Stat	p	Score
EHG	323.500	0.194	Meigs et al. [2008]
WHG	935.500	0.150	
SEF	476.000	0.063	
CEF	1149.000	0.131	
MN	114.000	0.005	
WEZ	738.000	0.331	
EHG	271.000	0.069	Hivert et al. [2011]
WHG	818.000	0.034	
SEF	595.000	0.330	
CEF	1203.000	0.213	
MN	260.000	0.053	
WEZ	819.000	0.431	
EHG	289.500	0.100	Cornelis et al. [2009]
WHG	875.500	0.126	
SEF	509.500	0.364	
CEF	1272.000	0.174	
MN	253.500	0.043	
WEZ	635.500	0.100	

4.3.7 Beta-binomial simulations

For two SNPs drift (S=0) was rejected during beta-binomial simulations as an explanation for the changes in allele frequencies over time. For both SNPs (rs4988235 in MCM6 and rs7570971 in RAB3GAP1) a similar selection strength of 2.00±0.80 was estimated. Furthermore, the SNPs were found to be in linkage disequilibrium in the Central European reference population (R^2 = 0.7849, D' = 0.9439).

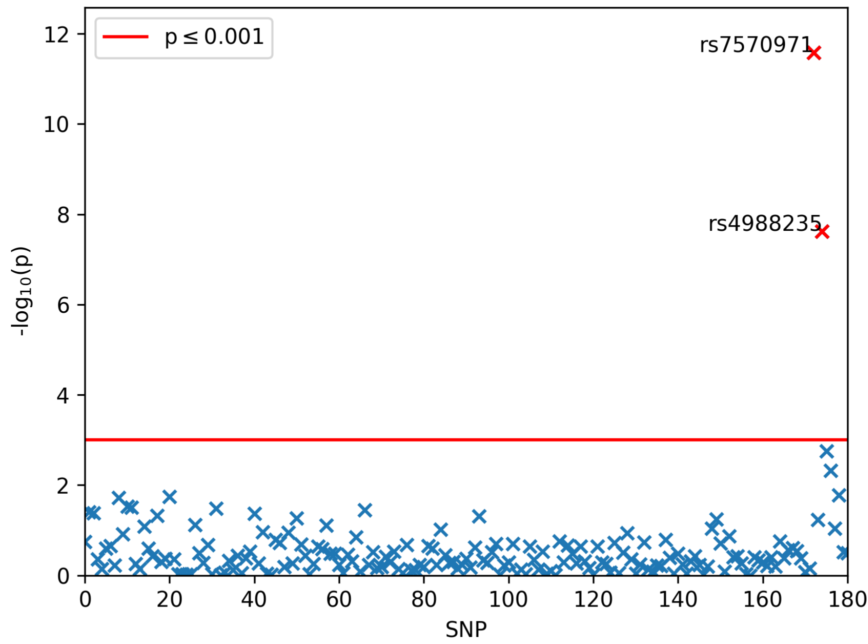

Figure (44) *Negative log_{10} p-values for beta-binomial simulation results for MetS associated SNPs at initial populations size of 1456 individuals for a selection coefficient of S=0. Threshold of p≤0.001 is indicated by red line.*

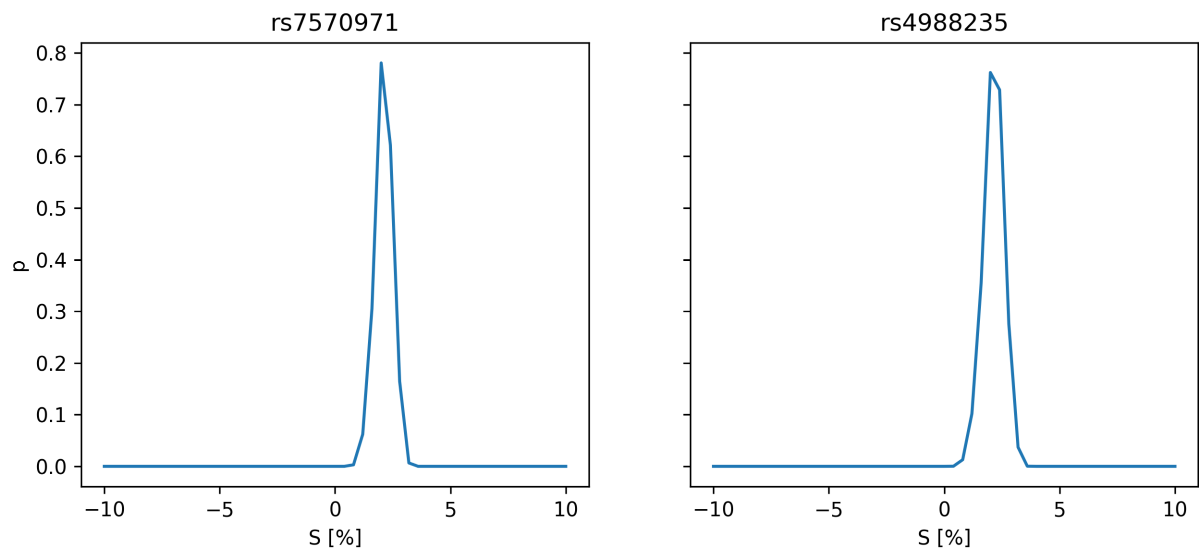

Figure (45) *P-values from beta-binomial simulations for each selection coefficient for an initial population size of 1456 individuals for the two SNPs were drift was rejected.*

4.4 Discussion

The differential distribution of MetS susceptibility in worldwide populations has long been debated. While some theories assume that selection has shaped these patterns, others suggest that differential population histories allowed for drift to create these differences. This chapter investigates the history of selected risk alleles with known associations to civilization diseases. Neutrality tests were applied to genome-wide data from three reference populations to identify loci that underwent potential selection. Ancient DNA data was used to track changes in allele-frequencies over time and thereby correlate signals of selection with a specific time frame. While the majority of alleles likely developed under neutrality, several alleles show signs consistent with selection during both early and later times.

The neutrality tests show that 54 SNPs in 34 genes in the Central European reference group deviate from neutral expectations. These results indicate that the majority of SNPs under investigation, developed under neutral conditions unaffected by selection. Differences at these loci between populations can be attributed to their differential histories. Varying population sizes, growth and decline, expansion and admixture can all have an effect on allele frequencies in a population [Hamilton, 2011]. Depending on the initial frequency of the allele, drift, in combination with these factors, can lead to a high degree of differentiation. Nevertheless, the comparison of the distribution of allele frequencies shows that differences between risk allele frequencies were higher than randomly drawn neutral allele frequencies in the Central European and the African reference group. Similarly, the average risk scores calculated for the African reference groups were higher compared to the non-African groups. These results, in combination with the neutrality tests, indicate that in addition to demographic processes, also selection played a role in shaping the distribution of risk alleles in worldwide populations.

These signs of selection contradict the 'Predatorial release hypothesis', which suggests that differences in risk allele frequencies arose only through drift and demography after a functional constraint was lifted from these genes [Speakman, 2006]. The abundant similarities in allele frequencies for the alleles for which deviations in neutrality were found in the Central European group, between all the European groups of this data set further suggest, that changes in allele frequencies affected by potential selection occurred before the Mesolithic. This can be interpreted as in accordance with Neel's original thrifty gene hypothesis [Neel, 1962]. However, it has to be considered that risk as well as protective variants seemed to be affected alike [Ayub et al., 2014], thus contrasting Neel's proposed increase in thrifty, and therefore risk-increasing alleles.

Early selection on the energy metabolism was probably linked to climate

Several alleles deviated from neutral expectations in both, the Central European and the Chinese reference group. While three SNPs were indicated in the Central European group by high iHS and/or nSL values, all but two SNPs led to high H12 values or additionally high Tajima's D values in the Chinese population. For the three SNPs identified in the Central European group, lower risk allele frequencies were found, compared to the other non-European reference groups, while no differences were detected between any of the ancient sample groups, when compared to the Central European

reference.

Compared to the Central European group, the average risk allele frequency in the AGT gene was higher in both non-European reference groups (YRI - diff: -0.2043±0.0093, p≤0.0001; CHS - diff: -0.2189±0.0097, p≤0.0001), while no significant differences could be found between the ancient sample groups and the Central European reference group. The average F_{st} across the AGT SNPs did not exceed F_{st} values of similar numbers of neutral SNPs between the European groups, while a generally higher diversity was measured at the loci (between: SEF - D_e: 0.3104±0.0551 and MN - D_e: 0.3561±0.0483) compared to the African reference group (D_e: 0.2224±0.0112, p≤0.0500). The diversity in combination with the F_{st} comparisons for the European groups indicate that changes in the AGT gene probably occurred prior to the Mesolithic, as results between groups are highly similar. These results support the salt-retention hypothesis[Nakajima et al., 2004, Thompson et al., 2004]. According to this hypothesis, it was advantageous to retain sodium, which was scarce in hunter-gatherer societies. The AGT gene encodes the protein angiotensinogen, which is a precursor for angiotensin. Angiotensin plays a crucial role in the renin-angiotensin system, which is involved, among other things, in controlling the blood volume via mediating renal functions [Ramalingam et al., 2017]. It was assumed that retaining sodium, which was scarce in hunter-gatherer societies, was advantageous, and associated variants were selected in hot climates. After a migration to more moderate climate-zones several of these variants became disadvantageous and were selected against, or were affected by drift.

In a similar fashion, high iHS (2.7834, $P_{99.6769}$) and nSL scores (2.7504, $P_{99.7448}$) indicate deviations from neutrality for the SNP rs2612067 in the high mobility group gene 2 (HMGA2) in the Central European reference group. The risk allele (G) was found at low frequency in the Central European reference group (0.11±0.04) and the Chinese reference group (0.12±0.04), yet at an intermediate frequency in the African population (0.42±0.07). Similar to the SNPs in the AGT gene, no differences could be found for any ancient sample group when compared to the European reference group. The SNP is associated with an increased risk for developing T2D [Pierce et al., 2011], while the HMGA2 gene acts as transcriptional regulating factor that was also indicated for an involvement in cancer growth [Fedele et al., 2006], and an association with bone mineral density [Kuipers et al., 2009]. This connection could be a potential explanation for the deviations from neutrality and the lower frequencies for the SNP outside of Africa. The reduced UV radiation in habitats further away from the equator is assumed to have led to depigmentation of the skin, to enable to synthesize sufficient amounts of vitamin D [Jablonski, 2012]. The role of vitamin D and its involvement in calcium absorption is crucial in bone formation and general bone health [Wagner et al., 2008]. Additional selection on genes related to bone mineral density or bone health in general in the same context is therefore likely.

Results for a number of other SNPs might also be explained in a similar context. For over 20 SNPs, deviations from neutrality were found in the Central European reference group, while no significant differences in frequencies could be found between any ancient sample group. The majority of those

SNPs had a strong association with the energy and fat metabolism and obesity. For example, for six SNPs in the gene encoding the fat mass and obesity-associated protein (FTO), high positive Tajima's D values (between 2.4611, $P_{99.1529}$ - 2.6723, $P_{99.5403}$) for the Central European reference group indicated an unusual high diversity at the locus. Variants in the gene have a strong association with obesity [González et al., 2012]. For the SNP rs9939609, an additive increase in obesity risk by 23% was estimated per risk allele [Kilpeläinen et al., 2011, Rankinen et al., 2010]. A high diversity was also found in all ancient sample groups, indicated by D_e values (between MN 0.3542±0.0456 and WEZ: 0.4677±0.0060). This high diversity in the ancient sample groups, in combination with the high Tajima's D values in the Central European reference group, indicates diversifying or balancing selection at the locus. Considering the data, these changes probably occurred during an early time period, yet after the split between the populations ancestral to East Asians and Central Europeans between 40ky and 80ky ago. This was indicated by a significant lower diversity found for the Chinese reference group compared to all European groups, except for the Eastern European/Russian hunter-gatherers D_e: 0.2320±0.0126, $p \leq 0.0500$).

Obesity-related SNPs are assumed to be the target of selection in response to cold stress in humans and that not an increased fat mass, but the ratio and distribution of brown adipose tissue (BAT) was affected by selection [Sellayah et al., 2014]. BAT or brown fat, together with white adipose tissue, is the main component in the adipose organs in almost all mammals and is relevant for energy storage and crucial for thermogenesis [Gesta et al., 2007]. It is actively involved in heat production using energy stored in the white adipose tissue. Several of the SNPs for which non-neutrality was detected, were located in genes with a direct association to the BAT metabolism, such as FTO [Ronkainen et al., 2016], MC4R [Berglund et al., 2014], UCP1 [Nishimura et al., 2017] and CDKN2A [Rabhi et al., 2018]. Comparative studies showed that a higher amount of BAT was found in Central Europeans than in East Asians [Bakker et al., 2014]. Considering all non-neutral SNPs without significant differences among the European groups, the average risk allele frequency was slightly higher in the Chinese group than in the Central European reference (diff: 0.0211±0.0058, $p \leq 0.0005$).
This difference was assumed to explain the differential cold tolerance between individuals from both populations. In Central Europeans, a higher amount of BAT was found, which was associated with an increased cold tolerance and an increased metabolic rate [Bakker et al., 2014]. The differential distribution of BAT, which seems to correlate with climate variables, is a potential factor that could explain part of the variance in MetS susceptibility found between people of different ethnic backgrounds [Sellayah et al., 2014]. In addition, the previously mentioned AGT gene also has an association with the fat metabolism and is, as a precursor for angiotensin II, directly involved in the browning of white adipose tissue [Colin Bell et al., 2002, Brown, 2006]. After the migration to more moderate climate zones, when the functional constraint on variants in the AGT gene was lowered, selection could have affected alleles associated with BAT in European populations. This, in turn, could contribute to an explanation of the differences in hypertension-susceptibility between people of different ethnic origin [Colin Bell et al., 2002, Brown, 2006].

In addition to potential selection on genes associated with BAT, an overall selection on SNPs re-

lated to the energy metabolism could be explained by selection in response to a colder climate. A response to cold-stress could have caused an increase in the average metabolic rate. Such an increased metabolic rate is known from humans living in colder climate zones, compared to humans originating from areas closer to the equator [Daanen and Van Marken Lichtenbelt, 2016]. Glucose plays a crucial role as energy supplier in generating heat [Morrison, 2016]. Genes involved in insulin response and blood glucose levels are therefore a potential target of selection in response to temperature dependent stress. In this context, polygenic selection can be assumed, as adaptations to climate and eco-region often are mediated by subtle shifts in allele frequencies [Hancock et al., 2010]. The neutrality test results also indicate an increase in diversity rather than a strong directional selection. Interestingly, of all the potentially non-neutral variants without differences in frequencies between the European groups, only two reached fixation (rs1802295, rs7227255). In both cases, the risk allele was lost from the population.

Selection during the early Neolithic was probably mediated by a shift in diet

Differences among the ancient sample groups could be detected for 23 SNPs that also deviated from neutrality in the Central European reference group. Based on these differences, the SNPs could be grouped in three clusters: (A) difference were found in one or both of the hunter-gatherer groups, (B) differences were found in only one or both of the early farmer groups, (C) differences were found in multiple groups regardless of subsistence.

Set A, which differed only in the early farmer groups, consisted of five SNPs (rs1799883 (FABP2), rs6536991 (UCP1), rs4655518 (LEPR), rs17782313 (MC4R), rs7578326 (IRS1)). Interestingly, all five SNPs are connected to the lipid metabolism or obesity [Meyre et al., 2009, Hancock et al., 2011, Ramos et al., 2012, Rohde et al., 2017]. Furthermore, all SNPs are indicated as non-neutral by either high positive Tajima's D values, low H12 values, or high iHS/nSL score results. These results, in combination with the low to elevated frequencies in the Central European reference (0.25 ± 0.06 – 0.66 ± 0.07) and the Central European hunter-gatherers (0.02 ± 0.04 – 0.48 ± 0.14), are indicative of diversifying selection in response to the Central European climate, similar to the other fat-metabolism related SNPs. These findings are reinforced by two genes previously identified as potential targets of selection in response to climate variables [Hancock et al., 2011]. For the SNP rs1799883 in the gene encoding the fatty acid binding protein 2 (FABP2), for which an association with weight management was previously mentioned, a positive correlation with climate is indicated. Similar results were found in the same study for the SNP rs4655518 in the leptin receptor gene (LEPR). The fact that the majority of the SNPs differed only in the early farmers of Central Europe, but not in the early farmers of the Aegean/Balkan region can probably be explained by the lower number of samples that could be successfully genotyped for the loci in comparison in the Aegean/Balkan group. Although it can not be ruled out that other effects played an additional role. Since the Central European farmers are considered descendants of early farmers from the Aegean region [Hofmanová et al., 2016] that migrated northwards, it is possible that alleles were lost during the migration. Serial founder effects during migration, as well as population expansion, can have severe effects on allele frequencies in a population [Hamilton, 2011].

For set B, where differences were only found in the early farmer groups, an opposite pattern was observable based on seven SNPs. All were found to deviate from neutrality in the Central European reference group by at least one of the tests, while differences in frequencies were only found in one or both hunter-gatherer groups. Two of those SNPs (rs10923931, rs1493694) were located in the gene that encodes the neurogenic locus notch homolog protein 2 (NOTCH2). Previous reports suggest strong linkage for those SNPs outside of Africa and an association with T2D [Zeggini et al., 2008]. This could be confirmed in the data set where both SNPs occurred at the same frequencies in the Chinese and the Central European reference groups (rs10923931 - CEU: 0.08±0.04, CHS: 0.04±0.03; rs1493694 - CEU: 0.08±.04, CHS: 0.04±0.03), but not in the African group (rs10923931: 0.37±0.03, rs1493694: 0.62±0.03). The frequencies in the African reference group further indicated, that a higher diversity was present at this locus than in the non-African groups. No differences in frequencies could be determined between the Central European reference group and any ancient sample group, except for the two hunter gatherer groups. In both groups the frequencies were significantly higher (rs10923931 - EHG: 0.57±0.26, WHG: 0.44±0.16, rs1493694 - EHG: 0.50±0.31, WHG: 0.47±0.18). The high negative Tajima's D value (-2.047, $P_{1.8006}$) for rs1493694 and the high H12 (0.5189, $P_{98.4756}$) and low H2_H1 (0.0144, $P_{0.7822}$) values for rs10923931 further indicate directional selection. There is strong evidence linking the NOTCH2 gene and the LEPR gene to bone mineral density, osteoporosis and general bone health [Kapur et al., 2010, Stathopoulos et al., 2013, Kim et al., 2008, Hannema et al., 2016, Richert et al., 2007, Yang et al., 2017, Koh et al., 2002, Lee et al., 2014]. This potential connection could explain the differential pattern observed for the frequencies at this locus in the data set. Research suggests that deteriorating bone health was a strong selective pressure [Rees and Harding, 2012]. As calcium is crucial for bone mineral density, the increased need to synthesize vitamin D at higher latitudes and its connection to the calcium absorption is thought to be the driver of depigmentation in Europe [Jablonski, 2012, Sturm and Duffy, 2012]. It is therefore realistic to assume that selection affected other variants associated with bone health outside of Africa. This theory is further reinforced by the high H12 (rs10923931: 0.6416, $P_{99.1405}$, rs1493694: 0.6340, $P_{99.0205}$) and low H2_H1 (rs10923931: 0.006, $P_{0.2781}$, rs1493694: 0.006, $P_{0.2695}$) values found for both NOTCH2 variants in the Chinese group. Thus, directional selection may have affected these variants not only in European but also in East Asian populations. In addition, differences in diet may explain the higher frequencies in the hunter-gatherer groups. Similar to variants related to pigmentation, it is possible that the higher vitamin D content in the hunter-gatherer diet lessened potential selective pressures affecting variants associated with bone health [Jablonski and Chaplin, 2012].

The results for the SNP rs174570 in the FADS2 gene encoding the fatty acid desaturase 2 indicate selection related to dietary customs. In the central European reference group, the C allele was found at high frequency (0.84±0.05), together with high iHS (2.1127, $P_{98.2953}$) and nSL (2.3381, $P_{99.1808}$) scores. In addition, an excess of homozygous alleles was found for the SNP. In comparison, no significant differences in allele frequency were found in any of the farmer groups, but a significant lower frequency was found in both hunter-gatherer groups (EHG: 0.17±0.21, WHG: 0.35±0.16). The C

allele, among other SNPs in the FADS1/2/3 gene cluster, is known to have a positive influence on the ratio of long-chained polyunsaturated fatty acids (LCPUFAs) in breast milk [Morales et al., 2011]. If LCPUFAs are not synthesized, they also can be obtained directly from animal-based protein. Adaptations of the FADS2 gene have been found for diets relying mostly on animal protein such as known from the Inuit [Fumagalli et al., 2015], as well as for mostly vegetarian diets of societies originating from South-East Asia [Kothapalli et al., 2016], with opposite effects. An increase in frequency during the early Neolithic seems therefore possible in response to the shift in diet towards a more cereal based nutrition [Richards et al., 2003]. The lowered amount of animal protein could have rendered an increased synthesis of LCPUFAs beneficial, as it would have increased the nutritional value of the infant's diet.

Strong signals of selection on lactase persistence

Strong evidence of selection was found for a SNP strongly associated with lactase persistence in Europeans. The SNP rs4988235 is located in the micro-chromosome-maintenance gene (MCM6), upstream of the gene that encodes the lactase-phlorizin hydrolase (LPH) [Enattah et al., 2002]. As LPH is responsible for hydrolyzing lactose into glucose, a decline in LPH activity after weening causes the inability to digest lactose as an adult. Digesting lactose without a functioning LPH can lead to nausea, internal distress, and diarrhea shortly after ingestion [Heyman, 2006]. As the allele is dominant, one copy of the derived allele is enough for the LPH activity to persist into adulthood. Presumably, it regulates the expression of the gene encoding the LPH, up-regulating its expression [Troelsen, 2005]. The derived allele (A) was absent in both non-European reference populations, but present at high frequency in the Central Europeans (0.74±0.06). In addition, very high iHS (2.8087, $P_{99.6986}$) and nSL (3.313, $P_{99.9484}$) scores for the European reference indicate a strong deviation from neutrality. The derived allele was also absent in all ancient sample groups, with the exception of the Bronze Age group, where it was found at low frequency (0.14±0.18). In addition, during beta-binomial simulations the allele was also indicated as a target of selection. A selection coefficient of 2.00±0.80% was estimated for an initial population size of 1456 people. These findings were consistent with previous studies that report the absence of the derived allele during the early Neolithic [Burger et al., 2007, Plantinga et al., 2012]. This shows a sharp increase in frequency for the SNP over the past 3,000 years in Central Europe. Furthermore, investigations of the locus using modern as well as ancient DNA found strong evidence for selection affecting the locus [Bersaglieri et al., 2004b, Mathieson et al., 2015], consistent with the findings in this study.

The rs182549 SNP is found on the same haplotype as rs4988235, associating it also with lactase persistence in Europeans [Enattah et al., 2002]. Similar neutrality test results (iHS: -3.3647, $P_{0.0878}$, nSL: -3.377, $P_{0.0773}$) and frequencies were found in the Central European reference (0.74±0.06) and the Bronze Age group (0.25±0.30), while the derived allele was absent in all other groups, except for the hunter-gatherers (EHG: 0.08±0.16, WHG: 0.07±0.13). This result is striking, as it indicates a different origin for both variants, since the derived allele was absent in both hunter-gatherer groups. The development of lactase persistence in Europe is often described as gene-culture coevolution [Gerbault et al., 2011]. It is thought to have created a nutritional advantage by allowing adults to consume dairy products. The consumption of cow's milk probably yielded a higher total energy per

cow, compared to consuming its meat. Although not uncontroversial, a beneficial increase in calcium intake, supplementing the reduced vitamin D absorption present at high latitudes, is theorized [Smith et al., 2009, Itan et al., 2010].

The complex nature of the MetS genetics is underlined by the fact that the allele at rs4988235 associated with lactase persistence was found to be protective against MetS [Friedrich et al., 2014], while also being associated with an increase in BMI [Kettunen et al., 2009]. Furthermore, another SNP was found to be in strong linkage with the rs4988235 SNP (R^2 = 0.7849, D' = 0.9439). The SNP rs7570971 is located in the gene encoding the RAB3 GTPase activating protein catalytic subunit 1 (RAB3GAP1) and the distribution of frequencies among the groups in this data set was highly similar compared to rs4988235 (WEZ: 0.14±0.14, CEU: 0.71±0.06, absent in all others). Similar to rs4988235, the allele was also indicated as under selection during the beta-binomial simulations. A similar selection strength was estimated for the allele (2.00±0.80%). A high nSL score (2.3702, $P_{99.2456}$) indicates non-neutrality in the Central European reference group. In combination, the results further indicate that selection affected not only the lactase gene but also RAB3GAP1. As the SNP is associated with positive effects on the fatty acid ratio in the blood [Kalsbeek et al., 2018] and milk contains approximately 400 different fatty acids [Lindmark Månsson, 2008], this could have been advantageous in connection with dairying.

Conclusion

This study shows that, besides factors such as diet and exercise, the differential susceptibility to MetS and T2D between populations is determined by genetics. While these differences may have arisen due to different population histories and drift, other theories suggest that selections in favor of a thrifty energy metabolism played a role. In this chapter, a set of predefined markers with a strong association to MetS risk was used to investigate the development of associated allele frequencies in European populations. Based on neutrality tests it could be shown that the majority of SNPs under investigation evolved under neutral conditions. Differences in frequencies found between populations can therefore be attributed to differential demographic histories that shaped these populations. However, a number of SNPs was found to deviate from neutral expectations. The frequency data obtained from these ancient sample groups could be used to determine the period in time when a potential selection may have occurred.

A large proportion of SNPs appear to be affected by selection in response to climate variables, after humans left Africa. Several loci with an connection to brown adipose tissue (BAT), such as the FTO gene, appeared to be selected during an early time, leading to a uniform distribution of frequencies among the European groups.

Differential distribution of BAT is known to influence cold tolerance, as differences between populations from different climate zones show. Adaptations to cold stress might further offer an explanation for the signs of selection found for other SNPs with an association to the energy metabolism. So are increased metabolic rates often found in people living in cold climate regions [Makinen, 2010, Daanen and Van Marken Lichtenbelt, 2016]. The data further indicates that more regional selec-

tion occurred during the Mesolithic and later times in Central Europe, affecting the genomes of the local hunter-gatherers and, after their arrival, also the early farmers. In addition to cold stress, the lowered UV radiation in regions further away from the equator might have had an effect on variants associated with MetS and T2D. A possible example can be found in the data indicating selection in the locus of the variant of the HMGA2 gene. While the variant is known to increase the risk for developing T2D, the gene is known to influence bone mineral density.

The lowered vitamin D content assumed for the diets of early farmers compared to contemporary hunter-gatherers may have increased the selective pressure on variants associated with the vitamin D and calcium metabolism. The data indicates that selection occurred in the early Neolithic and later periods on variants in NOTCH2 and LEPR, two genes with an association to bone mineral density and osteoporosis. Further evidence indicates that selection affected the FADS1/2 gene family during the same time. The gene influences the ratio of LCPUFAs in breast milk which can either be synthesized or supplemented from animal protein in the diet [Morales et al., 2011]. The plant-based diet of the early farmers is assumed to have rendered the variants beneficial, as they increased the nutritional value of the infants diet during nursing [Mathieson and Mathieson, 2018].

The strongest signals of selection were found for SNPs associated with lactase persistence into adulthood and blood lipid levels. The derived allele at rs4988235 was absent in all groups dating prior to the Bronze Age, where it was found at low frequencies. This shows that the allele sharply increased in frequency during the last 3,000 years in response to directional selection. In addition, a SNP in the RAB3GAP1 gene was found in linkage with rs4988235. The SNP has a positive effect on cholesterol and lipid levels in the blood [Bersaglieri et al., 2004b, Kalsbeek et al., 2018], which could be a connection to the consumption of milk and dairy products.

For a further understanding why selection affected alleles with an association to MetS and T2D risk, additional investigations in the underlying networks of genes involved are needed. With deeper knowledge of the actual function of the identified risk alleles, additional analyses can be designed to make use of the rapidly increasing body of ancient genome data. An increased sample size would also allow a more detailed investigation of polygenic selection, as certainties of frequency estimates would increase. Furthermore, epigenetic studies of ancient genomes could also increase our understanding of differential gene expression between different ancient populations.

5 The influence of subsistence and climate on the immune system

5.1 Introduction

The immune system

The immune system has always been a potential target of adaptations. This chapter investigates the influence of subsistence and eco-region on the history of SNPs in genes associated with the immune system and drug-metabolizing enzymes, in regard to the impact on auto-immune diseases still relevant today.

The body's ability to defend itself against pathogens can be categorized roughly in three major components: The physical barrier in form of the skin and mucous membrane, the innate immune system, and the adaptive immune system. While the innate immune system is present at birth and only discriminate between self and non-self, it is involved in the broad-range responses. In contrast, the adaptive immune system, acquired during a person's lifetime, can differentiate between different pathogens and tailor its response accordingly [Campbell et al., 2003]. Since the body is constantly exposed to a variety of pathogens, a highly functional immune system is crucial for survival.

If the physical barriers are breached and the immune system detects a pathogen, a cascade of responses is set in motion, depending on the pathogen as well as the tissues concerned. A high number of genes and regulatory pathways are involved in orchestrating an appropriate response [Alper et al., 2008, Brodin and Davis, 2017]. Since the human body is constantly exposed to a wide variety of pathogens, the immune system is under continuous pressure. Consequently, pathogens exert a high selective pressure on the immune system [Prugnolle et al., 2005, Wang et al., 2007, Hancock et al., 2010, 2011] Genetic variation, for example, strongly correlates with immune responses and the regional pathogen load in global populations [Fumagalli et al., 2011].

The epidemiological transitions

The co-evolution between humans and pathogens likely went through several key-stages. Omran [2005] proposed a model that describes three major transition in the exposure of humans to pathogens since the initial out-of-Africa migration. According to this model, the first transition correlated with the Neolithic transition, the second happened during the industrial revolution, roughly 200 years ago, and the third transition is associated with the rise of modern medicine and advances in sanitation. It is assumed that when humans lived as hunter-gatherers, the pathogen load was approximately similar for all populations and only correlated with climate-variables, but increased drastically with the onset of the neolithic [Armelagos, 2009].

This increase was linked to four main reasons: increasing population density, animal domestication, food storage and permanent settlements. (1) Increasing population density was a direct consequence of the active management of crops and animals that allowed more people to be fed from a piece of land compared to previous forms of subsistence. In addition, the sessile lifestyle allowed for shorter birth-intervals as small children did not need to be carried during migration periods. Based on

modern-day observations it is assumed that nomadic hunter-gatherers timed their birth-intervals to approximately four years to avoid carrying more than one small child [Diamond, 2002]. These factors allowed for an increase in population growth after the advent of agriculture. This led to the emergence of crowd diseases, which were probably uncommon during the Mesolithic. The small group sizes of hunter-gatherer bands prevented the spread and the re-infection associated with crowd diseases such as measles or smallpox, as their dying out would have simultaneously killed the pathogen [Diamond, 1998].

(2) The domestication of animals led to an increased risk of zoonoses, diseases that spread from one species to another. During the early Neolithic and for a long time afterwards, early peasants lived in close proximity to their livestock. This allowed for parasites and other pathogens, that were previously confined to wild animals, to cross the species barrier, exposing humans to pathogens they had not encountered before [Cook, 2015].

(3) The storage of food for longer periods than before increased the exposure to pathogens and toxins [Harper and Armelagos, 2010]. The availability of grains and the introduction of pottery enabled people to store food for later consumption. Depending on the storage conditions, mold and bacteria could infest the food. In addition, the stored food most likely attracted rodents that are known vectors for diseases [Morand et al., 2015].

(4) The permanent settlements of early farmers and their livestock increased the risk of unsanitary conditions. While hunter-gatherers inhabited each camp site probably only for a limited amount of time, waste and waste-management were only minor problems. A sessile lifestyle increased the risk of contaminating a water- or food source with fecal matter from humans and animals, posing an additional health threat [Diamond, 2002].

While a direct association with infectious diseases is not clear, evidence indicates a population decline in Europe during the later course of the early Neolithic, [Shennan et al., 2013, Kristiansen et al., 2017]. The development of a highly functional immune system was a response to the constant exposure to pathogens, and the role of the Neolithic is thought to have been crucial during this process.
The fact that large-scale animal domestication originated in Eurasia and was absent in the Americas led to a smaller risk of zoonoses in the New World. As a consequence, the immune system of the indigenous people was not prepared for the germs that were brought along by the European migrants, causing epidemics and the subsequent deaths of millions of native Americans [Diamond, 2002]. Genomic studies have shown that Amerindian populations have reduced variation at the HLA-regions in their genomes compared to Eurasian populationss [Cochran and Harpending, 2009]. Variation in these regions is directly linked to variation in immune responses and is thought to be under balancing selection [Prugnolle et al., 2005].

The hygiene hypothesis

Our highly functional immune system has ensured our survival in environments with varying pathogen exposure, but it has likely also led to problems in recent times of high sanitation standards [Zuckerman et al., 2014]. This idea is summarized in the 'hygiene hypothesis', first proposed by Strachan [1989] who noticed a higher prevalence of asthma in urbanized populations. He found a negative correlation between the exposure to infectious agents during childhood and the development of asthma later in life. This hypothesis was later adopted to a broader range of autoimmune diseases such as Type one Diabetes (T1D), hay fever and other allergies, and chronic inflammations such as Crohn's diseases and ulcerative colitis [Okada et al., 2010]. It is assumed that the immune system is adapted to an environment with a high pathogen exposure, which is no longer the case. The lack of exposure during early childhood hinders sufficient "training" of the immune system, which in turn leads to an overreaction caused by the inability to distinguish between self and non-self.

The hypothesis is backed up by findings that the diversity in genes associated with immune responses increases with the diversity of pathogens found in a specific environment [Fumagalli et al., 2011]. In addition, many risk alleles associated with autoimmune diseases are linked to immune responses [Brinkworth and Barreiro, 2014, Karlsson et al., 2014]. For example, alleles in the gene encoding the SH2B adaptor protein 3 (SH2B3) are associated with an increased activation of the NOD2 pathway, which is involved in creating an inflammatory response to the presence of bacteria [Zhernakova et al., 2010]. In addition, the allele is also increasing the risk to develop coeliac disease. Coeliac disease is characterized by a chronic intestinal inflammation, triggered by the consumption of gluten, a protein commonly found in cereal [Dubé et al., 2005]. Moreover, alleles associated with inflammatory bowel syndrome (IBS) are associated with leprosy, although not all IBS risk alleles are protective against leprosy [Karlsson et al., 2014]. IBS describes several conditions, such as Crohn's diseases or ulcerative colitis, were the immune system attacks the gastrointestinal system, causing chronic inflammation [Baumgart and Sandborn, 2012]. In addition, some risk alleles associated with Crohn's disease are known to increase resistance against norovirus infections and show signs of recent positive selection [Raj et al., 2013].

The old friends hypothesis

The 'old friends' hypothesis was proposed as a contrast to the 'hygiene' hypothesis with its strong emphasis on the Neolithic [Rook et al., 2003]. The 'old friends' hypothesis also states that the immune system developed in response to co-evolving pathogens. However, since humans lived the majority of their history as hunter-gatherers, environmental bacteria and parasites, such as helminthes (e.g., hookworms, roundworms etc.), were the main drivers of this arms race, as opposed to infectious crowd-diseases during the Neolithic [Rook, 2012]. A connection between parasites and auto-immune diseases could be shown in Italy, where incidences of multiple sclerosis (MS) increased after the successful eradication of malaria [Bitti et al., 2001, Sotgiu et al., 2008]. Furthermore, a deliberate infestation of IBS patients with helminthes has been shown to alleviate symptoms [Summers et al., 2005, Ramanan et al., 2016], further supporting the evidence for a link between parasites and immunity [Afifi et al., 2015]. Parasitic helminths are known to cause a down-regulation of the immune

response in the host, reducing the inflammation. Infestations with this parasite are furthermore negatively correlated with allergy and asthma prevalence [Yazdanbakhsh et al., 2002]. Helminth infections were common prior to the 1940's in almost all populations and with their decline, allergies and autoimmune diseases rose in frequency [Elliott et al., 2000]. Thus, the association of a rural upbringing and a decreased prevalence of asthma could also be caused by the increased presence of parasites rather than infectious diseases Strachan [1989].
77 Numerous studies have found signs of selection on SNPs associated with autoimmune diseases [Prugnolle et al., 2005, Wang et al., 2007, Hancock et al., 2010, 2011, Fumagalli et al., 2011], but only a few have attempted to time the selective sweeps [Raj et al., 2013, Zhang et al., 2013]. Raj et al. [2013] found evidence for selection on multiple alleles associated with IBS and estimated a rather recent sweep time of 1200 to 2600 years ago. Signs of selection for a risk allele in the SH2B3 gene associated with coeliac disease and T1D can be found in several studies Mathieson et al. [2015], Zhernakova et al. [2010], Pickrell et al. [2009], indicating a recent increase in allele frequency. These results concur with the 'hygiene' hypothesis, since an older sweep time would be expected in the context of the 'old friends' hypothesis.

The effect of food-storage on drug metabolizing enzymes

In addition to auto-immune related genes, genes involved in the drug metabolism, encoding so-called drug metabolizing enzymes (DME), were identified as potential targets of recent selection [Akey et al., 2002, Sabeti et al., 2006, Pimenoff et al., 2012]. It was estimated that variants in genes in the cytochrome P450 family and in the N-acetyl-transferase 2 gene (NAT2), that are associated with a decreased protein activity, increased in frequencies during the last 8,000 years [Patin et al., 2006, Janha et al., 2014]. As this increase in frequency coincides with the Neolithic in Europe, it may have been influenced by the consumption of stored food. The decreased protein activity causes a slower metabolization rate of the enzyme's substrate. Possible benefits of a slower metabolism could include, for example, a slower built-up of toxins in the liver when consuming fungi-infested stored food. The resulting higher tolerance would increase the amount of food a person could ingest compared to someone with a faster metabolism [Janha et al., 2014]. Today, many of these variants are known to affect the drug-response in patients [Wadelius et al., 2007, Shuldiner et al., 2009] as well as increase the personal risk for diseases such as various forms of cancer, Alzheimer's disease, or Parkinson's [Breyer et al., 2009, Zhou et al., 2012].

Although studies using ancient DNA investigated SNPs related to DME, immunity, and auto-immune diseases [Olalde et al., 2014, Mathieson et al., 2015], none of the hypotheses have been formally tested with samples from the relevant periods. Therefore, the data presented here gives an opportunity to use genomic data from early Europeans, dating to different periods of time and stemming from various subsistence backgrounds, to assess the influence of subsistence on the development of the immune system.

5.2 Hypotheses

The dataset presented here consists of SNPs associated with the immune system as well as different autoimmune diseases such as Crohn's disease, coeliac diseases, asthma, and multiple sclerosis. In addition, it contains SNPs associated with drug metabolizing enzymes, which are thought to have been involved in the detoxification of xenobiotics. With samples from different subsistence backgrounds, dating to different periods of time, it is well equipped to test the following hypotheses:

1. The Neolithic lifestyle had a major impact on the development of the immune system and led to an increase in risk alleles associated with autoimmune diseases ('Hygiene' hypothesis).

   If true, allele frequencies in immune system associated SNPs differ between hunter-gatherers and early farmer groups, and risk allele frequencies increase with time.

2. Parasites had a much stronger impact on the development of the immune system compared to the infectious diseases emerging in the Neolithic ('Old friends' hypothesis).

   If true, only minor differences between the hunter-gatherer and farmer groups of Europe exist. No major changes in allele frequencies or diversity coincide with the onset of the Neolithic.

3. Food storage had a major impact on alleles in genes encoding drug metabolizing enzymes.

   If true, allele frequencies in SNPs associated with drug metabolizing enzymes will differ between hunter-gatherers and early farmer groups. Allele frequencies increase further over time as food-storage persists.

5.3 Results

5.3.1 Frequencies

For this study, a total of 82 SNPs in 41 genes were analyzed. When compared to the Central European reference group (CEU), the highest number of differences in frequencies was found for the African reference group (YRI) with 52 SNPs (65% higher frequency, 35% lower frequency), followed by the Chinese reference group (CHS) with 48 SNPs (48% higher, 52% lower). For the ancient sample groups, the highest number of differences compared to CEU was found in the Central European farmer group (CEF) at 21 SNPs (48% higher, 52% lower), followed by the Central European hunter-gatherers (WHG) with 20 SNPs (60% higher, 40% lower). For the farmer group from the Aegean/Balkan region (SEF), 11 SNPs (45% higher, 55% lower) differed in frequency, compared to CEU. The Bronze Age group also differed by 11 SNPs (WEZ, 64% higher, 36% lower) from the CEU, while the hunter-gatherer group from Eastern Europe/Russia differed by 10 SNPs (90% higher, 10% lower). Between the African and the Chinese reference groups, 37 SNPs differed significantly in frequency (see table: 13).

Deviations from HWE were found for two SNPs in the Central European reference group. For the SNP rs16947 an excess of homozygous genotypes was found, while for rs2188962 an excess of heterozygous genotypes was determined. In the early farmer group of Central European an excess of heterozygous genotypes was found for rs1135840, while in the Eastern European/Russian hunter-gatherer group an excess of heterozygous genotypes was found for the SNP rs11096955 (see table: 14).

Table (13) *Frequencies of risk alleles of immune-related SNPs for all groups. Colors in frequency fields indicate significant differences between the group and the CEU reference (red=higher, blue=lower). SNPs for which deviations from neutrality were found in CEU are highlighted green int the ID column.*

ID	Gene	Risk	EHG	WHG	SEF	CEF	MN	WEZ	CEU	YRI	CHS
rs2248359	**CYP24A1**	C	0.40±0.30	0.59±0.17	0.65±0.21	0.71±0.12	0.50±0.35	0.73±0.19	0.61	0.34	0.64
rs72552267	**CYP2c19**	A	0.00±0.31	0.00±0.19	0.00±0.84	0.00±0.08		0.00±0.23			
rs1065852	**CYP2D6**	A	0.25±0.30	0.38±0.34		0.06±0.12	0.00±0.84	0.50±0.69	0.24	0.10	0.61
rs1135840	**CYP2D6**	C	0.40±0.30	0.25±0.21		0.45±0.16	0.25±0.42	0.50±0.49	0.43	0.33	0.26
rs16947	**CYP2D6**	G	0.50±0.35	0.36±0.25		0.56±0.24	0.25±0.42	1.00±0.84	0.68	0.44	0.87
rs3892097	**CYP2D6**	T	0.17±0.30	0.30±0.28		0.06±0.12	0.00±0.84	0.50±0.69	0.23	0.06	0.00
rs2740574	**CYP3A4**	C	0.00±0.60	0.00±0.46		0.00±0.46	0.00±0.84		0.02	0.77	0.00
rs776746	**CYP3A5**	T	0.00±0.26	0.03±0.06	0.04±0.07	0.13±0.09	0.17±0.30	0.11±0.15	0.04	0.84	0.27
rs185819	**HLA class 3**	T	0.33±0.27	0.43±0.26		0.23±0.16	0.00±0.37	0.25±0.42	0.47	0.49	0.44
rs12638253	**LEKR1**	T	0.50±0.69	0.33±0.27	0.50±0.49	0.27±0.17	1.00±0.84	0.40±0.30	0.45	0.11	0.73
rs1041983	**NAT2**	T	0.17±0.21	0.61±0.18	0.50±0.69	0.19±0.10	0.33±0.38	0.25±0.25	0.30	0.50	0.45
rs1208	**NAT2**	G	0.70±0.28	0.14±0.13	0.50±0.20	0.46±0.13	0.50±0.40	0.50±0.20	0.40	0.38	0.04
rs1495741	**NAT2**	G	0.20±0.25	0.20±0.14	0.42±0.19	0.36±0.13	0.33±0.38	0.33±0.19	0.27	0.48	0.50
rs1799929	**NAT2**	T	0.63±0.34	0.13±0.12	0.45±0.21	0.46±0.13	0.50±0.49	0.50±0.21	0.41	0.17	0.04
rs1799930	**NAT2**	A	0.10±0.19	0.55±0.21	0.20±0.25	0.16±0.11	0.25±0.42	0.13±0.16	0.30	0.20	0.25
rs1799931	**NAT2**	A	0.13±0.23	0.00±0.13	0.04±0.08	0.00±0.07	0.00±0.60	0.10±0.13	0.01	0.04	0.19
rs1801279	**NAT2**	A	0.00±0.31	0.00±0.12	0.00±0.15	0.00±0.06	0.00±0.46	0.00±0.19	0.00	0.12	0.00
rs1801280	**NAT2**	C	0.38±0.34	0.50±0.31		0.44±0.17	0.50±0.31	0.50±0.35	0.42	0.24	0.04
rs11096955	**TLR10**	G	0.50±0.28	0.73±0.16	0.60±0.21	0.65±0.13	0.33±0.38	0.68±0.17	0.35	0.53	0.51
rs11096957	**TLR10**	T	0.50±0.31	0.33±0.22	0.42±0.28	0.41±0.15	0.67±0.38	0.36±0.20	0.65	0.47	0.49
rs4129009	**TLR10**	T	1.00±0.26	0.64±0.18	0.80±0.18	0.80±0.11	0.75±0.30	0.61±0.23	0.81	1.00	0.76
rs4833095	**TLR10**	T	0.71±0.24	0.38±0.17	0.60±0.21	0.39±0.13	0.67±0.38	0.62±0.19	0.79	0.09	0.40

Table 13 continued from previous page

ID	Gene	Risk	EHG	WHG	SEF	CEF	MN	WEZ	CEU	YRI	CHS
rs5743618	**TLR10**	C				0.17±0.30	0.00±0.60	0.50±0.49	0.77	0.00	0.00
rs5743810	**TLR6**	G	1.00±0.31	0.96±0.08	0.80±0.18	0.95±0.06	1.00±0.46	0.89±0.15	0.52	1.00	1.00
rs11568729	**CYP2C19**	C	1.00±0.37	1.00±0.15	1.00±0.37	1.00±0.07	1.00±0.60	1.00±0.31	1.00	1.00	1.00
rs12248560	**CYP2C19**	T	0.13±0.23	0.08±0.16	0.00±0.60	0.07±0.07	0.00±0.60	0.21±0.21	0.22	0.25	0.01
rs12778026	**CYP2C19**	T	1.00±0.31	0.96±0.07	0.61±0.23	0.78±0.14	0.75±0.42	0.69±0.18	0.87	0.77	0.65
rs17884712	**CYP2C19**	A	0.00±0.46	0.00±0.31		0.00±0.15	0.00±0.60	0.00±0.60	0.00	0.00	0.00
rs1934967	**CYP2C19**	T	0.50±0.31	0.33±0.17	0.13±0.13	0.06±0.06	0.00±0.60	0.28±0.21	0.24	0.01	0.12
rs28399504	**CYP2C19**	G	0.00±0.31	0.00±0.13	0.00±0.15	0.02±0.04	0.00±0.60	0.00±0.17	0.00	0.00	0.00
rs3758580	**CYP2C19**	C	1.00±0.31	0.94±0.12		0.86±0.18	0.83±0.30	0.70±0.28	0.87	0.83	0.65
rs41291556	**CYP2C19**	C	0.00±0.31	0.00±0.12	0.00±0.23	0.02±0.04	0.00±0.46	0.00±0.21	0.02	0.00	0.00
rs4244285	**CYP2C19**	A	0.00±0.37	0.00±0.37	0.00±0.84	0.10±0.19	0.25±0.42	0.30±0.28	0.13	0.17	0.35
rs4304692	**CYP2C19**	T	0.88±0.23	0.62±0.19	0.88±0.23	0.80±0.11	0.75±0.42	0.89±0.15	0.91	0.47	0.79
rs4417205	**CYP2C19**	G	0.00±0.31	0.00±0.17	0.44±0.24	0.14±0.13	0.25±0.42	0.33±0.19	0.13	0.23	0.35
rs4986893	**CYP2C19**	A	0.00±0.37	0.00±0.14	0.00±0.31	0.00±0.07	0.00±0.60	0.00±0.26	0.00	0.00	0.05
rs4986894	**CYP2C19**	T	1.00±0.31	0.96±0.07	0.67±0.19	0.80±0.11	0.67±0.38	0.63±0.19	0.87	0.83	0.65
rs56337013	**CYP2C19**	T	0.00±0.37	0.00±0.12	0.00±0.31	0.00±0.07		0.00±0.23			
rs6413438	**CYP2C19**	T	0.00±0.37	0.00±0.31	0.00±0.84	0.00±0.37	0.00±0.84	0.00±0.37	0.00	0.00	0.00
rs7067866	**CYP2C19**	G	0.90±0.19	0.60±0.18	0.54±0.19	0.61±0.13	0.75±0.42	0.50±0.19	0.56	0.29	0.55
rs72558184	**CYP2C19**	A	0.00±0.31	0.00±0.19	0.00±0.84	0.00±0.08		0.00±0.23			
rs1128503	**ABCB1a**	A	0.30±0.28	0.44±0.24		0.36±0.25	0.50±0.40	1.00±0.84	0.43	0.14	0.63
rs2032582	**ABCB1b**	A	0.40±0.30	0.59±0.18	0.35±0.21	0.21±0.12	0.50±0.40	0.67±0.22	0.43	0.00	0.34
rs1045642	**ABCB1c**	A	0.67±0.27	0.69±0.18	0.60±0.30	0.29±0.13	0.50±0.40	0.67±0.27	0.57	0.11	0.30
rs17822931	**ABCC11**	A	0.00±0.31	0.17±0.21		0.13±0.16	0.00±0.60	0.00±0.60	0.91	0.96	0.88
rs497116	**CASP12**	G	0.00±0.26	0.00±0.17	0.00±0.84	0.00±0.08	0.00±0.46	0.00±0.46	0.00	0.14	0.00

Table 13 continued from previous page

ID	Gene	Risk	EHG	WHG	SEF	CEF	MN	WEZ	CEU	YRI	CHS
rs1361108	**CEPW**	T	0.33±0.38	0.36±0.20	0.08±0.11	0.35±0.16	0.50±0.40	0.36±0.20	0.42	0.29	1.00
rs1490384	**CEPW**	T	0.70±0.28	0.35±0.21		0.28±0.14	0.50±0.40	0.36±0.25	0.45	0.16	1.00
rs9388489	**CEPW**	G	0.40±0.30	0.33±0.17	0.08±0.11	0.35±0.13	0.50±0.49	0.35±0.21	0.41	0.16	1.00
rs281379	**FUT1/2**	A	0.00±0.31	0.00±0.23		0.37±0.14	0.33±0.38	0.17±0.30	0.58	0.18	0.00
rs601338	**FUT2**	A		0.00±0.60		0.38±0.34	0.50±0.69		0.54	0.53	0.00
rs10786436	**HPSE2**	T	0.60±0.30	0.41±0.21	0.20±0.25	0.31±0.13	0.00±0.37	0.45±0.22	0.32	0.53	0.53
rs2058660	**IL18RAP**	G	0.00±0.31	0.27±0.17	0.50±0.69	0.00±0.19	0.00±0.60	0.29±0.24	0.21	0.06	0.41
rs917997	**IL18RAP**	T	1.00±0.31	0.74±0.15	0.94±0.11	0.96±0.05	1.00±0.84	0.81±0.19	0.21	0.06	0.41
rs6822844	**IL2**	G	0.63±0.34	0.96±0.07	0.88±0.23	0.96±0.05	1.00±0.60	0.83±0.17	0.85	0.99	1.00
rs10210302	**IPP5D**	T	0.00±0.60	0.25±0.42	0.50±0.69	0.70±0.28	0.00±0.46	0.75±0.42	0.56	0.25	0.37
rs9642880	**MYC**	T	0.17±0.21	0.63±0.17	0.32±0.19	0.56±0.13	0.83±0.30	0.38±0.19	0.41	0.81	0.35
rs13003464	**PUS10**	G	0.20±0.25	0.23±0.18	0.13±0.16	0.30±0.13	0.25±0.42	0.29±0.24	0.35	0.94	0.03
rs415890	**RASET2**	C	0.20±0.25	0.23±0.16	0.70±0.20	0.70±0.12	0.33±0.38	0.41±0.21	0.48	0.22	0.38
rs599839	**PSRC1**	A	0.75±0.30	0.62±0.19	0.71±0.18	0.81±0.13	1.00±0.84	0.72±0.21	0.70	0.15	0.93
rs12103	**PUSL1**	T	0.50±0.35	0.25±0.30	0.50±0.69	0.36±0.20	0.25±0.42	0.00±0.84	0.17	0.96	0.99
rs670523	**RIT1**	A	0.30±0.28	0.67±0.19	0.45±0.22	0.46±0.13	0.75±0.42	0.42±0.20	0.35	1.00	0.89
rs675209	**RREB1**	T	0.70±0.28	0.53±0.18	0.04±0.08	0.20±0.10	0.50±0.49	0.41±0.21	0.26	0.45	0.91
rs653178	**SH2B3**	C	0.30±0.28	0.47±0.18	0.96±0.08	0.80±0.10	0.50±0.49	0.59±0.21	0.45	0.00	0.00
rs3184504	**SH2B3**	T	1.00±0.26	1.00±0.14	0.83±0.30	0.75±0.12	0.75±0.30	0.73±0.19	0.45	0.00	0.00
rs4766578	**SH2B3**	T	0.20±0.25	0.00±0.46	0.50±0.69	0.29±0.24	0.25±0.42	0.50±0.40	0.47	0.00	0.00
rs10761659	**ZF365**	G	0.75±0.30	0.40±0.18	0.32±0.19	0.33±0.14	0.50±0.49	0.64±0.20	0.55	0.03	0.77
rs2188962	**SLC22A5**	T	0.00±0.84	0.00±0.60		0.13±0.23	0.00±0.84		0.39	0.00	0.00
rs2269426	**TXB**	A	0.75±0.42	0.38±0.34	0.50±0.69	0.25±0.30	0.50±0.49		0.39	0.28	0.20
rs17810546	**SCHIP1**	G	0.20±0.25	0.00±0.14	0.00±0.23	0.00±0.07	0.00±0.60	0.29±0.24	0.09	0.00	0.00

Table 13 continued from previous page

ID	Gene	Risk	EHG	WHG	SEF	CEF	MN	WEZ	CEU	YRI	CHS
rs2802292	**FOXO3**	G	0.80±0.25	0.35±0.16	0.63±0.24	0.59±0.13	0.67±0.38	0.50±0.26	0.35	0.86	0.30

Table (14) *Significant results of Hardy-Weinberg equilibrium calculations for immune related SNPs.*

Group	**ID**	**Gene**	$\mathbf{Ho1_{obs}}$	$\mathbf{He_{obs}}$	$\mathbf{Ho2_{obs}}$	$\mathbf{Ho1_{exp}}$	$\mathbf{He_{exp}}$	$\mathbf{Ho2_{exp}}$	χ^2	**P**	**Het deficit**	**Het excess**
CEF	**rs1135840**	**CYP2D6**	3	15	1	5.80	9.39	3.80	6.76	0.03399	-0.6000	0.6000
EHG	**rs11096955**	**TLR10**	0	6	0	1.50	3.00	1.50	6.00	0.04979	-1.0000	1.0000
CEU	**rs16947**	**CYP2D6**	16	32	51	10.34	43.31	45.34	6.75	0.03415	0.2612	-0.2612
CEU	**rs2188962**	**SLC22A5**	31	59	9	36.97	47.06	14.97	6.38	0.04119	-0.2538	0.2538

5.3.2 Neutrality tests

In the Central European reference group, 25 SNPs in 14 genes were identified to deviate from neutral expectations by at least one neutrality test (for details see table 15 and fig 46). In the Chinese reference group, eight SNPs were found to deviate from neutrality, of which six deviated also in the Central European reference group. For the African reference group, two SNPs were identified, of which one was also found in the Central European reference.

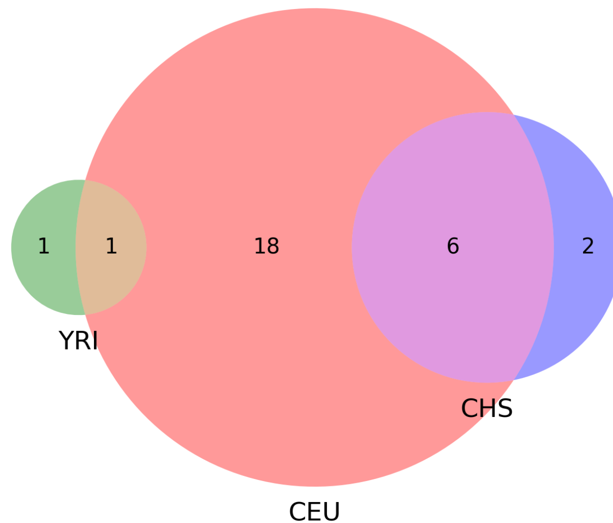

Figure (46) *Venn diagram of the number of SNPs that deviated from neutral expectations in each of the three reference groups. CEU=Central Europeans, YRI=Africans, CHS=Chinese.*

Table (15) *Results from neutrality test, applied to the three reference groups. Only results from the lowest and highest 2.5th percentile are reported.*

ID	**Gene**	**iHS**	[%]	**nSl**	[%]	**Tajima's D**	[%]	**H12**	[%]	**H2_H1**	[%]	**Group**
rs2248359	**CYP24A1**	-1.18	12.07	-1.45	7.67	-0.52	28.42	0.02	1.39	0.7	96.07	CEU
rs1135840	**CYP2D6**	-1.61	5.59	-1.61	5.92	2.56	99.03	0.02	3.63	0.78	97.59	
rs2740574	**CYP3A4**	-	-	-	-	-2.48	0.07	0.65	99.83	0.01	0.09	
rs776746	**CYP3A5**	-	-	-	-	-2.38	0.26	0.53	98.73	0.03	2.38	
rs1041983	**NAT2**	-	-	-	-	2.51	98.95	0.1	24.67	0.52	81.68	
rs1208	**NAT2**	-	-	-	-	2.51	98.95	0.1	24.67	0.52	81.68	
rs1495741	**NAT2**	-	-	-	-	3.12	99.8	0.1	25.9	0.24	33.86	
rs1799929	**NAT2**	-	-	-	-	2.51	98.95	0.1	24.67	0.52	81.68	
rs1799930	**NAT2**	-	-	-	-	2.51	98.95	0.1	24.67	0.52	81.68	
rs1799931	**NAT2**	-	-	-	-	2.51	98.95	0.1	24.67	0.52	81.68	
rs1801279	**NAT2**	-	-	-	-	2.51	98.95	0.1	24.67	0.52	81.68	
rs1801280	**NAT2**	-	-	-	-	2.51	98.95	0.1	24.67	0.52	81.68	
rs11096955	**TLR10**	-2.07	2.14	-2.55	0.8	1.6	92.27	0.52	98.35	0.04	4.72	
rs11096957	**TLR10**	-2.04	2.28	-2.47	0.95	1.6	92.27	0.52	98.35	0.04	4.72	
rs5743618	**TLR10**	-1.7	4.71	-2.29	1.38	-1.03	14.88	0.33	88.63	0.23	33.2	
rs5743810	**TLR6**	1.8	96.67	2.18	98.9	2.24	97.42	0.29	83.88	0.41	63.47	
rs497116	**CASP12**	-	-	-	-	-2.35	0.33	0.51	98.4	0.07	7	
rs281379	**FUT1/FUT2**	1.22	89.43	1.23	89.49	2.66	99.13	0.08	29.48	0.58	84.83	
rs601338	**FUT2**	1.73	96.25	1.59	95.19	2.66	99.13	0.21	74.83	0.54	80.62	
rs2058660	**IL18RAP**	2.1	98.22	1.38	92.18	1.61	94.28	0.05	5.59	0.75	98.31	
rs415890	**RNASET2**	1.24	90.15	1.12	88.37	3.43	99.73	0.08	19.77	0.56	86.33	
rs12103	**CPSf3L**	2.27	98.89	2.83	99.79	-1.94	2.47	0.52	98.5	0.08	10.8	
rs675209	**RREB1**	0.48	69.27	0.74	78.25	-1.4	6.29	0.48	98.24	0.22	30.24	
rs3184504	**SH2B3**	1.78	96.34	2.15	98.61	-1.12	13.75	0.31	89.1	0.57	87.02	

Table 15 continued from previous page

ID	Gene	iHS	[%]	nSl	[%]	Tajima's D	[%]	H12	[%]	H2_H1	[%]	Group
rs3184504	**SH2B3**	1.78	96.34	2.15	98.61	-1.12	13.75	0.31	89.1	0.57	87.02	
rs4417205	**CYP2C19**	-2.1184	2.0537	-2.1429	1.958	-1.9661	0.3207	0.2171	97.3177	0.0612	4.3927	YRI
rs2058660	**IL18RAP**	2.0072	97.6297	1.1573	88.0336	0.1716	86.5412	0.0285	12.9801	0.6143	83.8891	
rs2740574	**CYP3A4**	-	-	-	-	-2.5426	0.0795	0.6721	99.4524	0.0127	0.952	CHS
rs776746	**CYP3A5**	-0.5525	28.2438	-0.4549	32.0065	-1.3561	9.2391	0.6519	99.2812	0.0918	9.8498	
rs4833095	**TLR10**	0.608	72.7468	0.8989	81.5886	2.6258	97.8743	0.0611	8.6908	0.4354	63.8812	
rs497116	**CASP12**	-	-	-	-	-2.239	1.1137	0.3529	87.9334	0.0568	5.8501	
rs415890	**RNASET2**	-0.0642	45.8265	-0.2017	40.5864	3.0797	99.2426	0.1119	27.4143	0.3926	54.309	
rs12103	**CPSf3L**	-	-	-	-	-2.455	0.1895	0.6648	99.4434	0.0204	2.3191	
rs675209	**RREB1**	-2.5741	0.784	-2.6557	0.5928	-2.0472	1.5815	0.5563	98.0179	0.0208	1.8089	
rs3184504	**SH2B3**	-	-	-	-	-1.5968	7.4301	0.6725	99.2987	0.1692	19.5098	
rs3184504	**SH2B3**	-	-	-	-	-1.5968	7.4301	0.6725	99.2987	0.1692	19.5098	
rs3184504	**SH2B3**	-	-	-	-	-1.5968	7.4301	0.6725	99.2987	0.1692	19.5098	
rs3184504	**SH2B3**	-	-	-	-	-1.5968	7.4301	0.6725	99.2987	0.1692	19.5098	

5.3.3 Average expected heterozygosity D_e

SNPs located in the CYP2C19 gene

The diversity estimated for the SNPs in the CYP2C19 gene varied strongly between the groups. While a low diversity was found for the African (0.1973±0.0067) and Central European reference group (0.1744±0.0082), the diversity in the Chinese reference group was slightly elevated (0.2338±0.0055). In the ancient sample groups the diversity also varied, but was generally higher, compared to the non-African reference groups (EHG: 0.2743±0.0326, WHG: 0.2205±0.0218, SEF: 0.3248±0.0300, CEF: 0.2021±0.0210, MN: 0.3602±0.0375, WEZ: 0.3200±0.0284). In comparison, the diversity in the Central European reference group was found to be significantly lower, compared to Eastern European/Russian hunter-gatherers ($p \leq 0.500$), the early farmers from the Aegean/Balkan region ($p \leq 0.0005$), the Middle Neolithic group ($p \leq 0.0001$), the Bronze Age group ($p \leq 0.0001$), as well as the Chinese reference group ($p \leq 0.0001$). Additional pairings also led to significant results (for details see figure: 47)

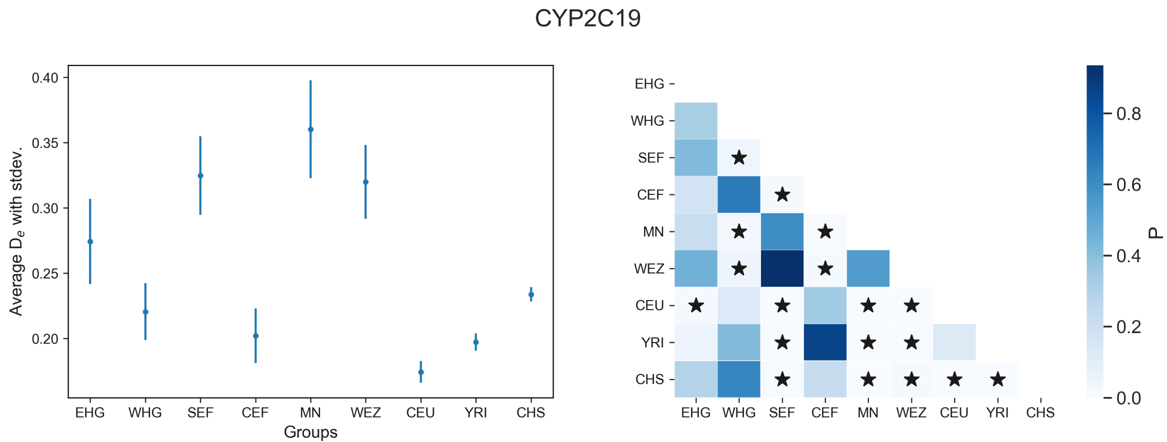

Figure (47) *Average expected heterozygosity D_e for all groups for the SNPs in the CYP2C19 gene. P-values for differences between groups are displayed in the heatmap. Significant values are highlighted by black stars.*

SNPs located in the NAT2 gene

The diversity found for the SNPs in the NAT2 gene was elevated in all groups. The highest diversity could be found in Europe, in the Middle Neolithic group (0.3716±0.0456), while the lowest diversity was found in the Chinese reference group (0.2393±0.0094). The lowest diversity in the ancient sample groups was found for the two hunter-gatherer groups (WHG: 0.2923±0.0273, EHG: 0.3126±0.0463), while a similar high diversity was found for the two early farmer groups (SEF: 0.3277±0.0342, CEF: 0.3225±0.0167). For the Bronze Age group, as well as the Central European and the African reference group an also similar diversity was estimated (WEZ: 0.3383±0.0325; CEU: 0.3385±0.0073, YRI: 0.3386±0.0102). The Chinese reference group was found to have a significant lower diversity, compared to all other groups, except both hunter-gatherers ($p \leq 0.0500$, see figure: 48).

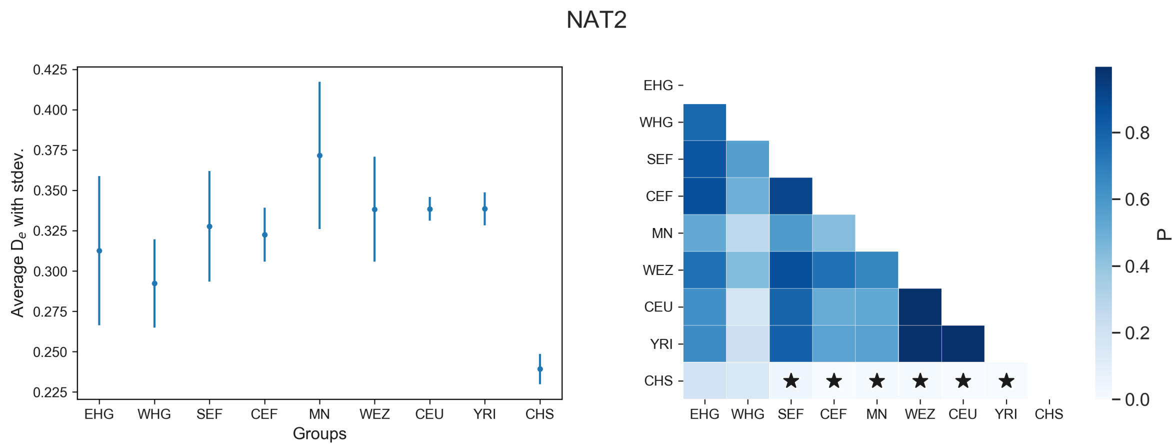

Figure (48) *Average expected heterozygosity D_e for all groups for the SNPs in the NAT2 gene. P-values for differences between groups are displayed in the heatmap. Significant values are highlighted by black stars.*

SNPs located in the TLR10 gene

An elevated diversity was estimated for all groups, when the SNPs at the TLR10 locus were considered. The highest diversity was found in the Bronze Age group (0.4157±0.0388), while the lowest diversity was estimated for the African reference group (0.2365±0.0082). All other groups had a similar high diversity (EHG: 0.3678±0.0454, WHG: 0.4190±0.0369, SEF: 0.4092±0.0409, CEF: 0.4030±0.0356, MN: 0.3653±0.0641, CEU: 0.3780±0.0149, CHS: 0.3683±0.0081). The diversity in the African reference group was found to be significantly lower, compared to all other groups of the data set (p≤0.0500, see figure: 49)

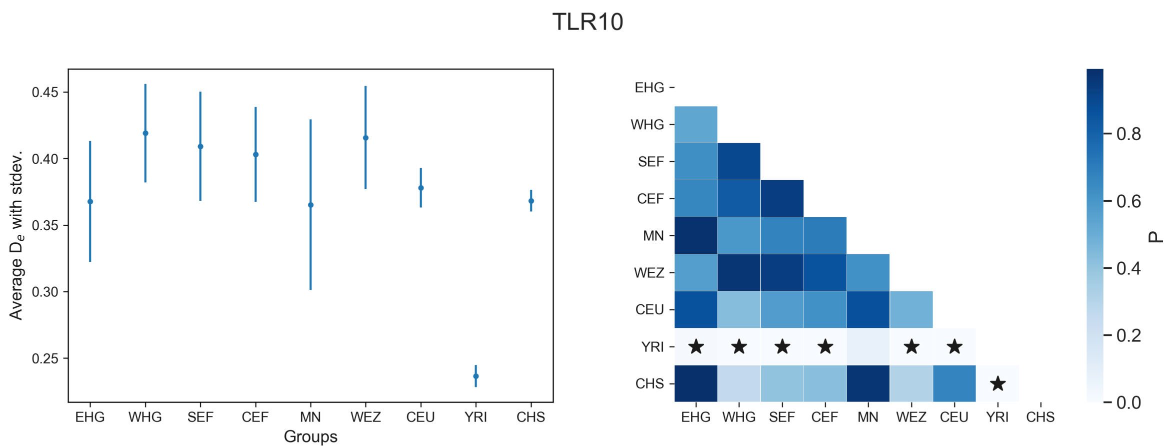

Figure (49) *Average expected heterozygosity D_e for all groups for the SNPs in the TLR10 gene. P-values for differences between groups are displayed in the heatmap. Significant values are highlighted by black stars.*

5.3.4 Pairwise F_{st} comparisons

SNPs located in the CYP2C19 gene

During the investigations of the CYP2C19 related SNPs, four group pairings were found where the average F_{st} differed significantly between the functional and the neutralome SNPs. In comparison

to the Central European reference group, the Bronze Age group (diff: 0.0197, $p \leq 0.0500$) as well as the early farmers from the Aegean/Balkan region (diff: 0.0443, $p \leq 0.0100$) had a significantly higher average F_{st} across the functional SNPs. Furthermore, significant differences were found between the African reference group and the hunter-gatherers from Eastern Europe/Russia (diff: 0.1851, $p \leq 0.0500$). The average F_{st} across the functional SNPs compared between the early farmers from the Aegean/Balkan region and the Chinese reference group was significantly lower than the neutral SNPs (diff: -0.0929, $p \leq 0.0500$, see figure: 50).

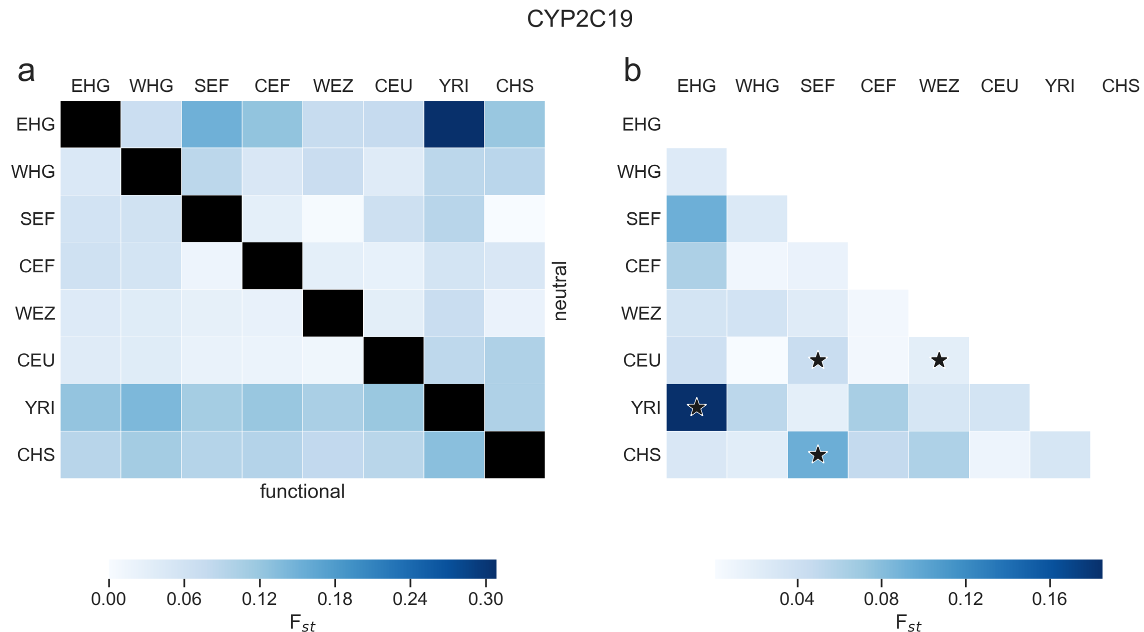

Figure (50) *a) Pairwise F_{st} comparison between SNPs in the CYP2C19 gene and SNPs randomly drawn from the neutralome. b) Absolute differences between F_{st} values. Significant values are highlighted by a black star.*

SNPs located in the NAT2 gene

The pairwise F_{st} comparisons identified 9 group pairings, where the average F_{st} values differed significantly between the functional and the neutral SNPs. Four of the pairings included the Chinese reference group (EHG - diff: 0.2816, $p \leq 0.0001$; SEF - diff: 0.1452, $p \leq 0.0500$; WEZ - diff: 0.2005, $p \leq 0.0001$; CEU - diff: 0.1123, $p \leq 0.0001$; CEF - diff: 0.1902, $p \leq 0.0001$), while the other five consisted only of groups of European origin and always included the Central European hunter-gatherer group (EHG - diff: 0.1375, $p \leq 0.0050$; CEU - diff: 0.0957, $p \leq 0.0500$; WEZ - diff: 0.0999, $p \leq 0.0500$; CEF - diff: 0.1122, $p \leq 0.0500$, see figure: 51).

SNPs located in the TLR10 gene

The F_{st} across the TLR10 SNPs was significantly higher compared to the F_{st} across neutral SNPs between the Central European reference group and the Central European hunter-gatherers (diff: -0.2014, $p \leq 0.0001$), the early farmers of Central Europe (diff: -0.1153, $p \leq 0.0001$), the Bronze Age group (diff: -0.0934, $p \leq 0.0001$), the African reference group (diff: -0.2697, $p \leq 0.0001$) and the Chinese refenrece group (diff: -0.1389, $p \leq 0.0050$, see figure: 52).

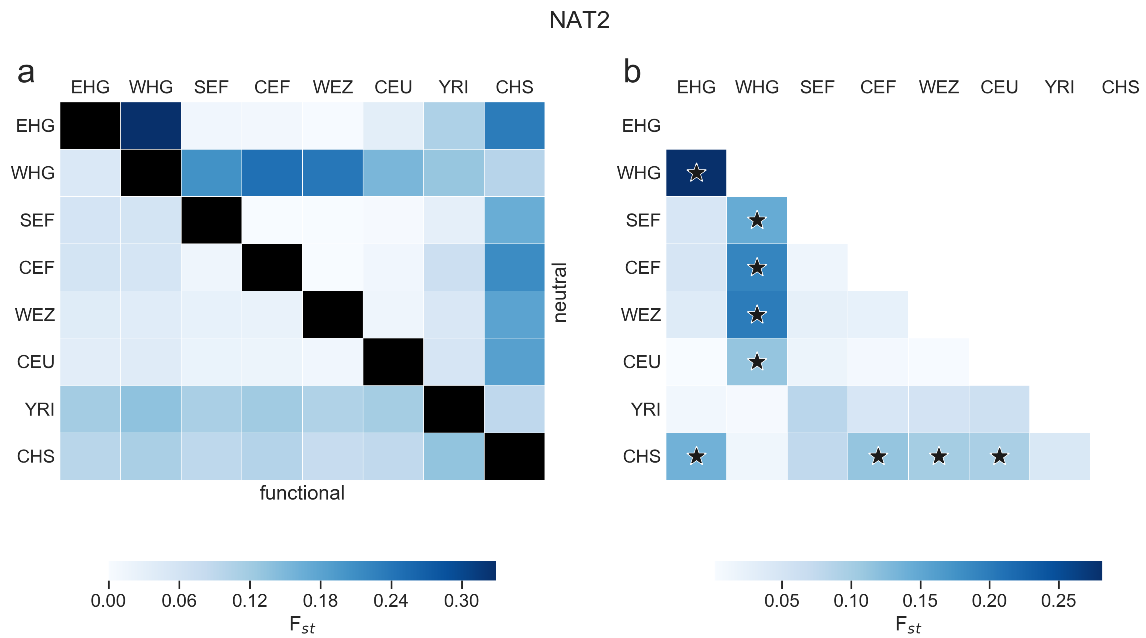

Figure (51) *a) Pairwise F_{st} comparison between SNPs in the NAT2 gene and SNPs randomly drawn from the neutralome. b) Absolute differences between F_{st} values. Significant values are highlighted by a black star.*

5.3.5 NAT2 acetylation-phenotype prediction as described in Sabbagh et al. [2009]

A NAT2 acetylation-phenotype could be predicted for 23 samples. For one sample from the Eastern European/Russian hunter-gatherer group, a rapid-metabolising phenotype was predicted, while for all other samples either a slow or intermediate phenotype was predicted (see table: 16).

5.3.6 Beta-binomial simulations

The results of the beta-binomial simulations for the immune related SNPs indicated that for the SNPs rs5743810 (TLR6) and rs11096955 (TLR10) drift (S=0) was not sufficient to explain the changes in allele frequencies observed between the Bronze Age group and the Central European reference group. For a initial populations size of 1456 individuals selection coefficients of S: 1.00±0.45% (rs11096955) and S: 1.60±0.80% (rs5743810) were estimated (for details see figure: 53 and 54).

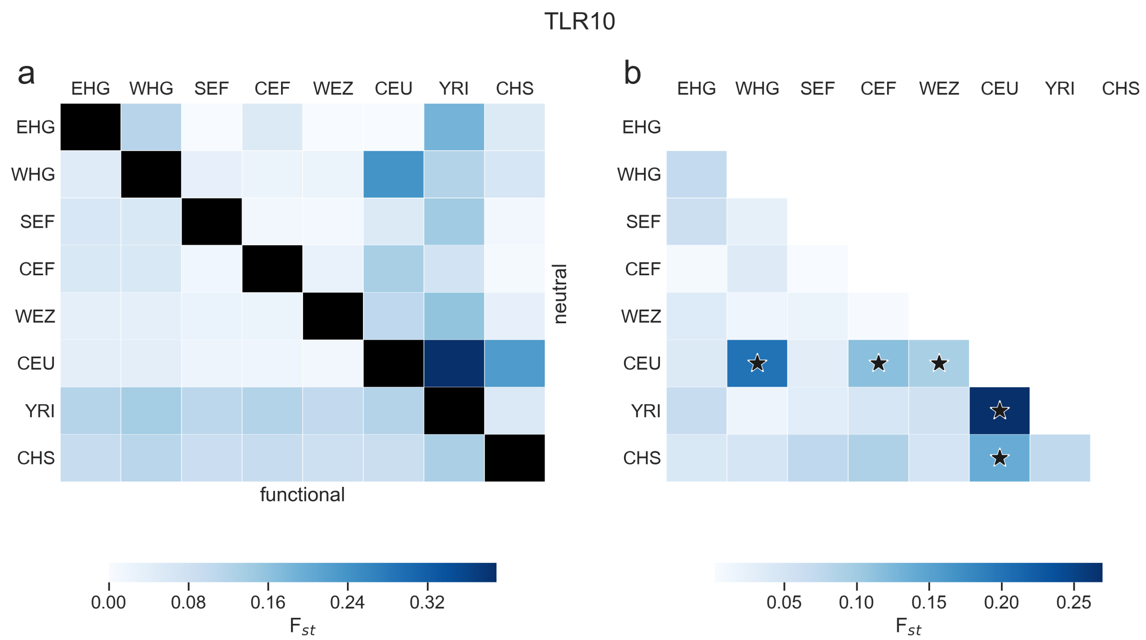

Figure (52) *a) Pairwise F_{st} comparison between SNPs in the TLR10 gene and SNPs randomly drawn from the neutralome. b) Absolute differences between F_{st} values. Significant values are highlighted by a black star.*

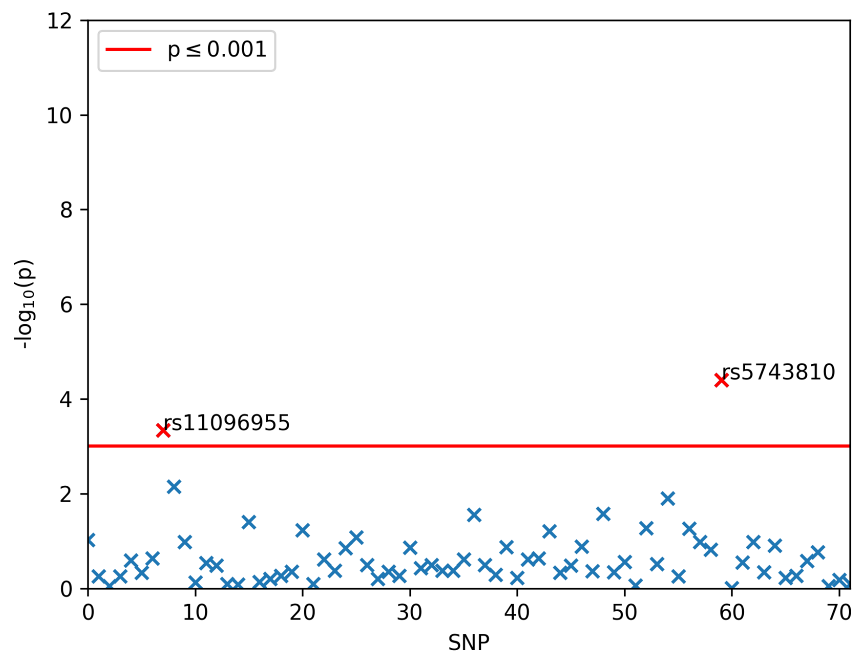

Figure (53) *Negative logtextsubscript10 p-values for beta-binomial simulation results for the SNPs associated with the immune system at initial populations size of 1456 individuals for S=0. Threshold of p≤0.001 is indicated by red line.*

Table (16) *Results from the NAT2 phenotype prediction webtool NAT2PRED. Phenotypes were predicted based on six SNPs. Probabilities for slow (S), intermediate (I) and rapid (R) metabolizing phenotypes are given.*

Sample	**genotype**	**p(R)**	**p(I)**	**p(S)**	**prediction**	**Group**
Min11	1,3,3,1,3,1	0.00164	0.00118	0.99718	S	EHG
Min3	1,1,1,1,2,1	0.97994	0.01221	0.00785	R	
Zv317	2,2,2,1,2,2	0.00001	0.00126	0.99873	S	
Losch	1,2,1,1,2,1	0.00201	0.99646	0.00153	I	WHG
281-19-6	1,2,2,1,2,1	0.00100	0.99696	0.00203	I	CEF
282-104-4	1,2,2,1,2,1	0.00100	0.99696	0.00203	I	
282-126-16	1,2,2,1,2,1	0.00100	0.99696	0.00203	I	
282-126-7	1,2,2,1,2,1	0.00100	0.99696	0.00203	I	
282-23-1	1,2,2,1,2,1	0.00100	0.99696	0.00203	I	
282-94-11	1,2,2,1,2,1	0.00100	0.99696	0.00203	I	
7034	1,2,2,1,2,1	0.00100	0.99696	0.00203	I	
Asp1	2,1,1,1,1,1	0.06598	0.92143	0.01260	I	
Asp2	1,2,2,1,2,1	0.00100	0.99696	0.00203	I	
Asp8	1,2,2,1,2,1	0.00100	0.99696	0.00203	I	
Dil15	2,1,1,2,1,1	0.00095	0.99704	0.00200	I	
Dil16	1,2,2,1,2,1	0.00100	0.99696	0.00203	I	
Klein10	2,2,2,2,2,1	0.00001	0.00126	0.99873	S	
Klein1	2,1,1,2,1,1	0.00095	0.99704	0.00200	I	
Klein8	1,2,2,1,2,1	0.00100	0.99696	0.00203	I	
NE1	1,2,2,1,2,1	0.00100	0.99696	0.00203	I	
Bla32	2,2,2,2,2,1	0.00001	0.00126	0.99873	S	MN
Wez51	1,2,2,1,2,1	0.00100	0.99696	0.00203	I	WEZ
WEZ59	2,2,2,1,2,2	0.00001	0.00126	0.99873	S	

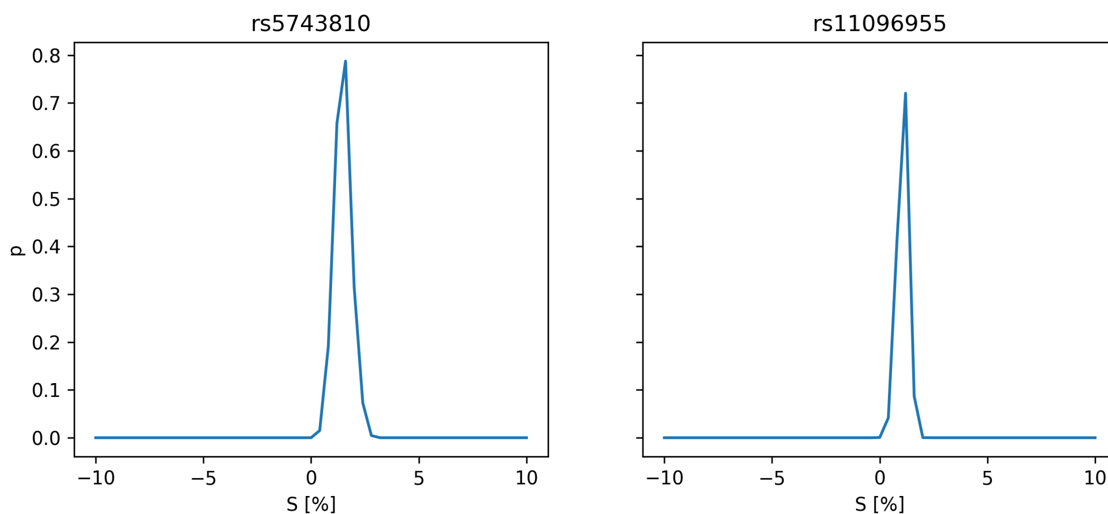

Figure (54) *P-values from beta-binomial simulations for each selection coefficient for an initial population size of 1456 individuals for the two SNPs were drift was rejected.*

5.4 Discussion

Genes involved in immune responses and pathogen recognition have always been thought of as potential targets of adaptations. The human body is constantly exposed to a wide range of environment-specific pathogens and therefore has to rely on a highly functional immune system. Research indicates that the prevalence of autoimmune diseases originates in a highly functional immune system that is, as a consequence of considerably reduced pathogen load, under-challenged in modern westernized societies, [Afifi et al., 2015]. The underlying reasons leading to the differential development of the immune system in populations around the world is still part of debate [Cochran and Harpending, 2009]. While the pathogen load of an environment correlates with climate variables, e.g. humidity and temperature [Prugnolle et al., 2005], it is assumed to have increased after the transition to farming [Armelagos, 2009]. The deteriorating hygienic conditions that are believed to have accompanied the transition from a hunter-gatherer to a farming lifestyle are seen by some researchers as one of the three major epidemiological transitions in human history [Armelagos, 2009]. This is thought to have led to several adaptations in the immune system as well as in genes encoding drug metabolizing enzymes (DMEs).

The data set described here was used to investigate if alleles associated with the immune system and drug metabolizing enzymes were influenced by selection. Neutrality tests were applied to the extensive reference data while the frequency data of the ancient sample groups was used to narrow down a time window where changes in frequencies may have occurred. Thus it was possible to correlate some of the observed changes in frequencies with changes in subsistence as well as with climate regions during different periods in time.

Neutrality tests identified 25 SNPs located in 15 genes to deviate from neutral expectations in the Central European reference group. High test scores, as found in the border areas of the results distribution for each chromosome, are inconsistent with neutrality under several demographic models and can therefore be considered as indications of selection [Akey et al., 2002, Sabeti et al., 2006, Voight et al., 2006, Garud et al., 2015]. This furthermore implies, that the majority of SNPs under investigation did not deviate from neutral expectations. Differences between groups found for those SNPs can be attributed to demographic processes and their general unique histories.

The SNPs discussed in this chapter can be divided into two groups: SNPs related to the immune system and auto-immune disease, and SNPs related to DMEs. While for both sets of SNPs deviations from neutrality could be determined in the Central European reference population, the variation in frequencies was higher for the immune system related SNPs among the ancient sample groups. These results indicate that changes in allele frequencies occurred at loci related to the immune system, most likely in response to selection, in Europe over the course of the last 10,000 years. Remarkably, only two SNPs could be found where one of the alleles reached fixation in the European reference group. In both cases, the alleles were also fixed in all ancient sample groups.

One of those SNPs was found in the caspase 12 gene (CASP12). For rs497116, high negative Tajima's D values in the Central European (-2.3463, $P_{0.3251}$) as well as the Chinese reference group (-2.2390, $P_{1.1137}$) indicate, that directional selection affected the locus outside of Africa. The allele frequencies found for the SNP in all non-African groups were furthermore consistent with this assumption. In all of these groups, the derived A allele had reached fixation. While the ancestral G allele was still present in the African reference group, the derived A allele had also reached elevated frequency (0.74 $\pm$ 0.04). The derived allele at the locus is a mis-sense mutation that leads to a premature stop of the genes transcription, thereby rendering the resulting protein inactive [Saleh et al., 2004]. The CASP12 gene is involved in the mediation of inflammation responses, following infections. The inactivation of the gene leads to an increased inflammatory response, rendering the A allele a beneficial mutation [Man and Kanneganti, 2016], while the G allele is associated with an increased sepsis risk [Saleh et al., 2004]. Age estimations for the derived allele suggest, that it first appeared between 500k and 100k years ago in Africa, with a subsequent selective sweep starting between 100k to 50k years ago [Wang et al., 2006, Xue et al., 2006, Hervella et al., 2012]. These estimates could be substantiated by the results presented here, which furthermore indicate that the A allele was affected by an early sweep, probably reaching fixation before the initial split of the Central European and East Asian populations roughly 40ky ago [Yang and Fu, 2018].

The rest of the immune system related non-neutral SNPs could be divided into three groups. In one group, the frequencies of the SNPs in the Central European hunter-gatherers were already similar to the reference group, yet different from at least one early farmer groups. In another group, the frequencies in the Central European hunter-gatherers differed from the reference, but not from the early farmer groups. In the third and largest group, the SNPs differed in both, the early farmers as well as the hunter-gatherers, and sometimes also in the Bronze Age group in comparison to the modern-day frequencies.

The first group consisted of two SNPs. One SNP was rs2058660 in the Interleukin-18 Receptor Accessory Protein-Like gene (IL18RAP). High iHS values in the Central European (2.0972, $P_{98.2229}$) and the African reference group (2.0072, $P_{97.6297}$) indicated non-neutrality for this SNP. The derived G allele occurred at low frequency in the African group (0.06$\pm$0.03), but was slightly elevated in the Central European reference group (0.21$\pm$0.06) as well as in the Central European hunter-gatherer group (0.27$\pm$0.17). In contrast, no copy of the derived allele could be found in the Central European farmer group. However, for the early farmer group from the Aegean/Balkan region, where only one sample for the SNP could be genotyped, one copy of the derived allele was found. It is possible that the allele has been lost during the migration, as a consequence of a founder effect, especially if it was present at low frequency in the initial population, leading to its absence in the Central European farmer group of this data set. These results show that derived alleles were absent in the new arriving farmers, but again rose in frequency to a level previously found in hunter-gatherers already living in Central Europe for several thousand years. The allele could have been re-introduced into the farmer groups as admixture with local hunter-gatherers increased during the later Neolithic [Lipson et al., 2017]. The G allele at the locus is associated with an increased response to infectious

diseases such as leprosy, while also being associated with increasing the susceptibility to Chron's disease [Franke et al., 2010, Liu et al., 2012]. The high iHS score, in combination with the high H2_H1 values (0.7473, $P_{98.3121}$), indicate a balancing, rather than a strong directional selection at the locus. Similarly, Zhernakova et al. [2010] found signals indicating a selective sweep at the locus during the last 6,500 years. These signals could be explained by selection affecting the locus in early farming populations of Central Europe, in response to a selective pressure that already affected local hunter-gatherers.

The second SNP in the group was rs12103 in the cleavage and polyadenylation-specific factor 3-like gene (CPSf3L). The T allele at the locus is associated with an increased pathogen defense, mediated by an inflammatory response. Via the same mechanism, the allele increases the risk for IBD [Ye and McGovern, 2016]. Signs of non-neutral development were found by high iHS values in the Central European reference group (2.8293,$P_{99.7939}$), as well as high negative Tajima's D values in the Chinese reference (-2.455, $P_{0.1895}$). The risk-increasing allele had almost reached fixation in the African (0.96 $\pm$ 0.03) and the Chinese reference group (0.99 $\pm$ 0.02) while it was at an intermediate to low frequency in the Central European reference group (0.17 $\pm$ 0.05). A significantly higher frequency was found in the Central European farmer group (0.36 $\pm$ 0.20) compared to the Central European reference group, suggesting a decrease in frequency to a level similar to the Central European hunter-gatherers 0.25 $\pm$ 0.30) after the early Neolithic. The high frequencies in combination with the deviations from neutrality in the non-European reference groups could be caused by a functional constraint, associated with an increased pathogen load in warmer climates [Dunn et al., 2010]. During the migration to environments of higher latitudes, this constraint might have been lifted and the T allele became detrimental, potentially leading to a subsequent decrease in frequency over time.

The second group contained two SNPs whose frequencies only differed when compared between the Central European hunter-gatherers and the Central European reference group. One SNP (s415890) was located in the gene which encodes the ribonuclease T2 (rnaset2), while the other (rs3184504) was found in the gene encoding the SH2B adaptor protein 3 (SH2B3). Both SNPs were present at similar intermediate frequencies in the Central European reference group (0.45$\pm$0.07, 0.48$\pm$0.07, respectively), where a high nSL score (2.15, $P_{98.6054}$) was found for rs3184504, while one of the highest Tajima's D values was measured for rs415890 (3.4297, $P_{99.7313}$). No differences were found between the European reference and any of the ancient sample groups, except for both hunter-gatherer groups. For rs3184504, no risk-associated alleles were present in any of the hunter-gatherer groups, while for rs415890 the risk allele was found at high frequencies (EHG: 0.80$\pm$0.25, WHG: 0.77$\pm$0.16). The high positive Tajima's D value is indicative of balancing selection, which could further explain the intermediate frequencies found for both SNPs in the European reference group. The SNP rs3184504 was previously identified to have one of the strongest associations with Coeliac's disease [Hunt et al., 2008], as well as being associated with type 1 diabetes [Ferreira et al., 2010]. In Contrast, for rs415890, an association with increased risk for Chron's disease is known [Franke et al., 2010]. Moreover, both SNPs are known to be involved in mediating immune reactions via up-regulation of inflammation (SH2B3) [Zhernakova et al., 2010] or the activation of T-cells in response

to parasite infections (rnaset2) [Thorn et al., 2012]. Both associations could explain why selection may have affected these alleles. These results were further confirmed by previous studies suggesting that selection affected both loci in association to infections after the onset of the Neolithic [Raj et al., 2013, Zhernakova et al., 2010].

The remaining six non-neutral SNPs were all located in two gene families and the allele frequencies in the Central European reference group were found to differ from all ancient sample groups, including the Bronze Age group.

The SNPs rs601338 and rs281379 are located in the gene encoding the enzyme galactoside 2-alpha-L-fucosyltransferase 2 (FUT2) and are associated with Crohn's and Coeliac's disease [Franke et al., 2010, Parmar et al., 2012]. While the risk allele (A) at rs601338 was present at similar frequency in the African (0.53 ± 0.07) and the European reference group (0.54 ± 0.07), the distribution differed for rs281379. Here, the risk allele (A) frequency was higher in the European group (0.58 ± 0.07) than the African reference group (0.18 ± 0.05). Both SNPs were found on the same holotype spanning 400 SNPs, for which a high Tajima's D value was detected (2.6573, $P_{99.1294}$) in the Central Europeans. In contrast to the other reference groups, no copy of the risk alleles could be found in the Chinese group. The intermediate frequencies, in combination with the high Tajima's D values in the Central European reference group, can be seen as evidence for balancing selection at this locus. No copy of the risk-increasing alleles could be found in the hunter-gatherer groups. However, the genotyping success was low for rs601338 in both groups. While no data was available for the early farmers from the Aegean/Balkan region, the risk allele was already present at elevated frequencies for both SNPs (rs281379: 0.37 ± 0.14, rs601338: 0.38 ± 0.34) in the Central European farmers, although at significant lower frequencies than today. These findings could confirm previous reports suggesting that the FUT2 gene experienced selection during the Neolithic [Raj et al., 2013].

The FUT2 gene is known for its involvement in the release of the ABO antigens in humans [Paré et al., 2008]. The presence of the risk alleles is associated with an inactive FUT2 gene, which prevents the release of the ABO antigens. This so-called non-secreter phenotype is associated with protection against norovirus infections, due to noroviruses using these antigens as receptors [Smyth et al., 2011]. This could explain potential selection, as norovirus infection cause severe stomach flu [Koch-Institut, 2017]. An untreated infection can lead to severe dehydration in children and thus poses a serious health threat. The norovirus is still common in Europe where it affects mostly infants and elderly people. It is transmitted via fluids from person to person or via contaminated food sources [Bernard et al., 2014], but it can also be present in livestock [Mattison et al., 2007]. The signals of non-neutrality, in combination with the absence of the risk alleles in hunter-gatherers and the subsequent increase in frequencies over time, are nevertheless indicative of a selective sweep associated with the Neolithic.

Similar to the SNPs in FUT2, four SNPs in the toll-like receptor gene complex yielded high nSL scores in the Central European reference group. Three SNPs were located in TLR10 (rs11096955,

rs11096957, rs5743618; nSL: -2.5454, $P_{0.8031}$; -2.4662, $P_{0.9519}$; -2.2933, $P_{1.3801}$, respectively), and one SNP was located in TLR6 (rs5743810; nSL: 2.18, $P_{98.8994}$). Furthermore, selection was also indicated by the simulation results for the SNPs rs11096955 and rs5743810. The changes in frequencies over the last 3,000 years since the late Bronze Age could only be explained if selection coefficients of 1.00±0.45% (rs11096955) and 1.60±0.80% (rs5743810) were assumed. The toll-like receptor genes are associated with pathogen pattern recognition and thereby trigger an immune 166response and mediate inflammation [Ozinsky et al., 2000, Delneste et al., 2007, Jiang et al., 2016]. A high diversity was detected for the TLR10 SNPs in the Central European reference group (D_e: 0.3780±0.0149), which, in combination with the high nSL score results, indicates diversifying or balancing selection. For genes involved in immune responses and pattern recognition, gene diversity frequently correlates with pathogen diversity, and signs of selection at TLR genes were found in various populations [Karlsson et al., 2014, Hancock et al., 2011, Novembre and Han, 2012].
While diversity was high in all ancient sample groups too (D_e: between MN: 0.3653±0.0641 and WHG: 0.4190±0.0369), significant differences between frequencies in comparison to the Central European reference indicate changes over time. For the Central European hunter-gatherers (diff: -0.1153, $p \leq 0.0001$) as well as the Central European farmers (diff: -0.2014, $p \leq 0.0001$) and the Bronze Age group (diff: -0.0934, $p \leq 0.0001$), average F_{st} values estimated for the TLR10 SNPs exceeded neutral F_{st} values compared to the Central European reference group. These results show that the changes in frequencies occurred after the onset of the Bronze Age in Europe and can not be explained by drift alone. Previous studies suggest that diversity found in TLR genes in modern populations was partially influenced by introgression from archaic hominins outside of Africa [Jagoda et al., 2017]. In combination with the high diversity found for all European groups, this indicates that selection affected a standing variation in Europe.

The fact that all these SNPs are associated with the susceptibility for infectious diseases such as malaria, tuberculosis and leprosy [Schurz et al., 2015, Wong et al., 2010], could furthermore explain the comparatively late change in frequencies in the data set. The assumed increase in pathogen load during the early Neolithic had probably already led to an increase in selective pressure on variants associated with the immune system. This pressure could have intensified during later periods when population size and mobility started to increase [Linderholm et al., 2014]. The late Neolithic and Bronze Age period are known for demographic changes and vast migrations [Allentoft et al., 2015]. Both factors could have increased the pathogen load while simultaneously providing a vector for pathogen propagation [Diamond, 1998]. It is not clear since when certain diseases were common in Europe, but it is assumed that leprosy was present since the late Neolithic [Stone et al., 2009, Köhler et al., 2017], while tuberculosis is assumed to be much older [Karlsson et al., 2014]. While neither of the diseases are immediately deadly, they can severely impact reproduction and fitness, either in form of a decline in health, or through ostracism and social stigma know from leprosy.

DMEs

Similar to the immune-related genes, genes encoding DMEs show high variability among worldwide populations. Almost all of these DMEs have a naturally occurring substrate, and variation in these

genes is often linked to differences in enzyme activity [Nebert and Dieter, 2000]. Differential enzyme activity manifests in distinct phenotypes, which are characterized by the speed of the metabolic reaction. In some cases, a high enzyme activity can lead to severe reactions in response to a drug dosage that has no side-effects in patients with a slower metabolism [Sistonen et al., 2009]. For several DME encoding genes, signs of selection were found, which in some cases could be dated to the last 10,000 years [Janha et al., 2014]. Based on this, a connection of the Neolithic to the selection on DMEs has been suggested [Magalon et al., 2008, Podgorná et al., 2015]. For the genes NAT2 and CYP2C19 it was assumed that food storage, which increased during the Neolithic, led to the built-up of toxins in the food due to infestation by fungi and bacteria [Janha et al., 2014]. Phenotypes associated with a slower metabolization rate became beneficial as the decreased enzyme activity prevented a rapid uptake of the toxins for the carriers. Others have suggested that the pattern of diversity found for the NAT2 gene can be explained entirely by demographic processes [Sabbagh et al., 2008]. While the highest diversity is found in Africa, it decreases with increasing distance as predicted by migration and serial founder effects [Luca et al., 2008].

The capture array that was used to build this data set contained eigth SNPs located in the NAT2 gene. For all SNPs, high positive Tajima's D values were found in the Central European reference group (between 2.5078, $P_{98.9493}$ - 3.1218, $P_{99.8009}$). The general estimated diversity was high in the Central European group (D_e 0.3385±0.0073), as well as in all ancient sample groups (WHG-D_e: 0.2923±0.0273, MN - D_e: 0.3716±0.0456). For the Chinese group, the estimated diversity was significantly lower compared to all groups ($p \leq 0.0500$) except the hunter-gatherers. These results indicate that the NAT2 locus was under selection in European populations.

Average F_{st} estimates for the locus suggest a selection scenario associated with subsistence. The Central European hunter-gatherers were the only group, for which a higher-than-expected value was found under neutrality (diff: 0.1123, $p \leq 0.0005$) when compared to the Central European reference group. Previous research furthermore indicates, that the NAT2 gene underwent selection in response to different forms of subsistence in populations relying on agriculture, as well as in hunter-gatherers [Sabbagh et al., 2011, Podgorná et al., 2015]. In addition, the gene was suggested to be the target of selection in Europe [Magalon et al., 2008], with an estimated sweep-time starting ca. 6,000 years ago [Luca et al., 2010a]. Other theories state that the gene was not under strong selection until the Bronze Age in Europe and that factors unrelated to diet had a stronger influence [Mathieson and Mathieson, 2018]. Diversity and F_{st} estimates based on the data presented here are more consistent with an earlier sweep time, most likely during the early Neolithic, as no significant differences could be found between any ancient sample groups dating to the Neolithic or later compared to the modern-day Europeans.

Interestingly, no significant differences in diversity of F_{st} could be detected between the Early and Middle Neolithic groups and the Bronze Age group although increasing admixture is known to have happened after the first farmers arrived in Central Europe [Allentoft et al., 2015]. It is furthermore possible, that a re-emergence of hunter-gatherer related ancestry in European farmer populations

influenced the allele frequencies at the NAT2 locus during the Middle Neolithic [Lipson et al., 2017]. While this is indicated by the slightly increased diversity found for the gene in the Middle Neolithic group compared to the early Neolithic groups, the high variance and bad preservation in the Middle Neolithic group prevented a conclusive statement.

A similar scenario was proposed for variation in the CYP2C19 gene, were an increase in slow metabolising phenotypes was associated with the Neolithic lifestyle [Janha et al., 2014]. Here, the results indicate that the majority of diversity found at this locus in European populations could be explained without the assumption of selection. The average diversity was found to be low in the Central European reference group (D_e: 0.1744±0.0082) while a high variation was detected for the ancient sample groups (between WHG - D_e: 0.2205±0.0218 and MN - D_e: 0.3602±0.0375). For all but one SNP, neutrality could be assumed based on the test results, while for rs4304692 elevated H1 (0.3564, $P_{97.8143}$) and low H2_H1 values (0.0203, $P_{1.2776}$) showed that a low haplotype diversity was found surrounding the SNP. In addition, this SNP was suggested to be under selection during the last 10,000 years [Janha et al., 2014]. It was assumed, that the slow metabolizing phenotype the SNP is associated with prevented a rapid activation of potential toxins originating from food storage. The results described here are consistent with such a scenario; lower frequencies were found for the SNP in the Central European hunter-gatherers (0.62±0.19) as well as in the Central European farmers (0.80±0.11) compared to the Central European reference (0.91±0.04).

Furthermore, and in addition to elevated frequencies (0.79±0.06), neutrality tests led to similar results in the Chinese reference group (H1: 0.3564, $P_{97.8143}$, H2_H1: 0.0203, $P_{1.2776}$). This could be the result of an early selection, dating before the split of the populations that became ancestral to Central European and South-east Asians roughly 40ky ago [Yang and Fu, 2018], that was continuous and potentially later intensified in Central Europe. Alternatively, a convergent evolution of the allele in both populations could explain the results. The derived allele was also found at intermediate frequency in the African reference group (0.47±0.07), which suggests that it is rather old. Nevertheless, the low H2Ï values in both non-African populations indicate, that directional rather than diversifying selection affected the CYP2C19 gene, leading to an increase in slow metabolizing phenotypes. A recently found high variation at the CYP2C19 homologue in ancient goat genomes [Daly et al., 2018] was interpreted as coevolution under the assumption that selective pressure has originated from toxins in plant residuals used as fodder.

A selection in response to subsistence was furthermore proposed for the variation in the CYP2D6 gene [Fuselli et al., 2010]. A high haplotype diversity for the SNP rs1135840 was detected in the Central European reference by high Tajima's D values (2.5604, $P_{99.0301}$), accompanied by high H2_H1 values (0.7778, $P_{97.5865}$). The overall lack of success in genotyping for the variants prevented the detection of consistent changes in allele frequencies between the ancient sample groups and the European reference. Only for rs16947 a significant lower frequency could be found in the Central European hunter-gatherers (0.26±0.25) compared to the European reference group (0.68±0.07). Therefore, it could not be determined if a potential selection affected the alleles in response to different subsis-

tence regimes, or if changes in frequencies might have occurred during an earlier time as seen for other DME genes such as CYP3A4 and CYP3A5.

For one SNP in each of those genes deviations neutrality was found in the Central European reference, indicated by high negative Tajime's D values (rs2740574 -2.4820, $P_{0.0657}$; rs776746 -2.3769, $P_{0.2555}$) in addition to low H2_H1 values (rs2740574: 0.0061, $P_{0.0922}$; rs776746: 0.0255, $P_{2.3848}$). Similar results were identified for the Chinese reference based on high H12 (rs2740574: 0.6721, $P_{99.4524}$; rs776746: 0.6519, $P_{99.2812}$), accompanied by low H2_H1 values (rs2740574: 0.0127, $P_{0.952}$). The ancestral alleles for the SNPs rs2740574 in the gene CYP3A4 and rs776746 in the gene CYP3A5 were present at high frequencies in the African reference group (rs2740574: 0.77 ± 0.06 , rs776746: 0.84 ± 0.05), yet at low frequency in the European reference group. In the Chinese reference group, the ancestral allele was at intermediate frequency for rs776746 (0.27 ± 0.06), and absent for rs2740574 (0.00 ± 0.02). Since deviations from neutrality were found for both SNPs in the Chinese and the European reference populations, a shared history of the derived alleles is possible. This scenario was further emphasized by the frequencies in the ancient sample groups. For rs2740574 no derived alleles could be found in any of the ancient samples (EHG: 0.00 ± 0.60, WHG: 0.00 ± 0.46, CEF: 0.00 ± 0.46, MN: 0.00 ± 0.84); however, the overall genotyping success was low. For rs776746 ancestral alleles were present at low frequency in the ancient samples (EHG: 0.00 ± 0.60, WHG: 0.03 ± 0.06, SEF: 0.04 ± 0.07, MN: 0.17 ± 0.30, WEZ: 0.11 ± 0.15), while the frequency in the early farmers of central Europe was significantly higher, compared to the Central European reference. This increased frequency could be caused by lineage sorting during the initial migration process when allele frequencies are often prone to changes. The low H2_H1 values found for both SNPs in the Central European reference, in combination with the low frequencies of the ancestral alleles, indicate directional selection of the derived allele. Based on a similar pattern found in the Chines group, an early selective sweep is highly likely. This is further supported by the frequencies found in the ancient sample population.

While a connection to salt retention was discussed previously for the SNPs in CYP3A4/5 [Thompson et al., 2004], this could not be confirmed by additional work [Ho et al., 2005]. Instead, a connection between CYP3A4/5 and the vitamin D metabolism is suggested [Wang et al., 2013], assuming it to be the target of selection since humans migrated to regions at higher latitudes [Jablonski and Chaplin, 2000, Robins, 2009, Yuen and Jablonski, 2010]. Both CYP3A genes are futhermore involved in metabolizing the active form of vitamin D3 to a yet unknown metabolite [Baranyai, 2014]. Thus, a down-regulation of these genes could lead to a higher level of available vitamin D circulating in the blood [Roizen et al., 2018]. This, in turn, could explain why these alleles may have been beneficial after humans migrated into regions with lower UV radiation levels farther from the equator. The results are also consistent with the age estimation of 76,000 years for the derived allele of the rs776746 SNP [Bains et al., 2013], suggesting that the potential sweep occurred during the first migrations. These findings, in combination with the results discussed in chapter 3, provide additional evidence in favor of the vitamin D hypothesis.

The results for the SNP rs2248359 can be interpreted in a similar context. The SNP is located in the gene CYP24A1, which encodes an enzyme that relies, among others, on 1,25-dihydroxyvitamin D3 as substrate [Jones et al., 2012]. The SNP was found at an elevated frequency in the Central European reference group 0.61±0.07), while an unusually high haplotype diversity was detected through low H12 values (0.0198, $P_{1.3945}$). These results indicate balancing or diversifying selection, in contrast to the SNPs in the CYP3A4/5. Similarly, no differences were found between any of the ancient sample groups and the Central European reference group. This could indicate that the diversification occurred early at this locus, in connection to the viatmin D metabolism, similar to the SNPs in CYP3A4/5. Interestingly, diversity was also high for the region surrounding the SNP in both non-European reference groups, while the frequency found in the Chinese reference group was similar to that in the European reference (0.64±0.07). Derived alleles were also present in the African reference group (0.34±0.05), indicating that a potential selection affected a standing variation rather than a de-novo mutation.

Conclusion

The occurrence of pathogens and toxins varies considerably in correlation with environmental factors and has always posed as a threat to humans. A dependable defense mechanism was therefore crucial for survival, and associated genes have always been seen as target of potential selections [Hancock et al., 2011]. Genes associated with the immune system and the detoxification of xenobiotics are assumed to have been under strong selective pressure. Adaptations in immune-related genes are thought to have led to a hyper-protective immune system that can cause auto-immune diseases in westernized societies [Raj et al., 2013]. In this chapter, the influence of subsistence and climate regions on the evolution of alleles in genes associated with immune-responses and DMEs was tested. Neutrality test were applied to an extensive reference data set, while frequency data obtained from ancient sample groups was used to narrow down a time period in which changes in allele frequencies may have occurred. The results of this study show that the majority of SNPs under investigation developed under neutral expectations. For the immune related SNPs where neutrality was rejected, changes in allele frequencies could be assigned to different periods in time, as well as to different environmental variables.

The results indicate that selection affected gene variants associated with increased inflammatory responses, occurring while, or shortly after humans migrated out of Africa. Additional signals of more regional selection were detected for Central Europe. This indicates an adaptation of the immune system to the regional parasite or pathogen spectrum, which later also affected the genomes of early farmers after their arrival. Signs of non-neutrality in combination with the frequency data furthermore indicate that variants in the genes SH2B3 and RNASET2 were selected during the early Neolithic.

Differences in frequencies between the hunter-gatherers and early farmers of Central Europe indicate a selective pressure originating from the Neolithic lifestyle rather than the European eco-region. Furthermore, the frequencies found in the early farmer groups were highly similar to modern-day frequencies in Central Europe. Up until the early Neolithic, only SNPs associated with increased

inflammation responses appear to be affected by selection while selection on variants associated with crowd diseases such as tuberculosis or norovirus infections emerged at a later time. There is strong evidence that SNPs in the TLR and FUT1/2 gene family were affected by selection during the late Neolithic and Bronze Age. These results confirm earlier theories suggesting that increasing population size and mobility lead to the spread and perseverance of crowd diseases absent before the Neolithic [Armelagos, 2009].

This study furthermore found evidence indicating that variants in DME encoding genes were under selection since humans left Africa. Strong sings of non-neutrality were found for variants in genes such as CYP3A4/5, CYP24A1, CYP2C19, and NAT2. The allele frequency data from the ancient sample groups places changes for NAT2 and CYP2C19 in the early Neolithic. This confirms earlier assumptions that slow metabolizing phenotypes were selected for in both genes in Central Europe. For the remaining variants, an early sweep time was indicated, probably close to the migration out of Africa. All three genes have a strong association to the vitamin D metabolism. Similar to genes such as SLC24A5 and SLC45A2 [Sturm and Duffy, 2012], selection in favor of an optimized vitamin D metabolism is highly likely.

Future research needs to incorporate additional results from different sources. Environmental DNA taken alongside the samples from sampling sides could provide a valuable insight into to the coevolution of humans and pathogens. Dental calculus can also be used to investigate dietary habits as well as pathogen levels. Here a direct estimation of the pathogen load accompanying different forms of subsistence could help understand the rate of exposure early humans have experienced.

6 Synthesis discussion and conclusion

Over the last decade, the advances in sequencing technology have increased the number of available human genomes, ancient and recent alike, at incredible speed. Especially the field of ancient DNA has led to several break-through studies that helped to shed light on the demographic histories of populations. This data showed that the genetic variation in Central European populations living today was strongly impacted by migration and admixture followed by population growth, mainly since the last glacial maximum, approximately 19,000 years ago [Gamba et al., 2014, Fu et al., 2016]. In addition to demography, the selection in, and adaptations to new environments are likely to have played a role in recent human evolution [Hancock et al., 2010]. Several studies using ancient and recent data have found strong signs indicating that present day phenotypic diversity was affected by selection [Hudjashov et al., 2013, Wilde et al., 2014, Mathieson et al., 2015].

Adaptations are assumed to have occurred in response to new environments. For many markers, a strong correlation with geography can be found, indicating selection in response to regional variables [Hancock et al., 2011]. For a long time the influence of the Neolithic on the genetic variation found today has been debated [Harper and Armelagos, 2010]. Several theories suggest that the drastic shift in lifestyles during the Neolithic, going from hunter-gatherers to sedentary farmers, imposed a strong selective pressure on the genomes of early humans [Ségurel et al., 2013]. The new plant-based diet, the increasing population size and the close proximity to domesticated animals is thought of as a new and unique environment that impacted the health of the early farming communities [Harper and Armelagos, 2010, Ash et al., 2016].

This study investigated the impact of climate and subsistence on the genomes of early Europeans. A specially designed capture array was used to compile a diachronic data set based on partial genomes from over 100 individuals. This data was obtained from individuals from the Mesolithic, predating the arrival of agriculture, as well as early and later Neolithic and Bronze Age farmer groups. This enabled the examination of allele frequency changes over time for alleles with known associations to traits assumed to have undergone selection. The phenotypic markers of the capture array can be grouped into three sets: 1) markers associated with pigmentation and other visible traits; 2) Markers associated with the energy metabolism with a known correlation to type 2 diabetes and the metabolic syndrome, as well as markers associated with the processing of foodstuff; and 3) markers associated with the immune system and the detoxification of xenobiotics. By applying neutrality tests such as iHS and Tajima's D to a reference data set of genome-wide data of recent populations, it could be determined if the markers under investigation deviated from neutrality, thus indicating selection. Allele frequency data of the ancient samples was used to narrow down a time frame during which these changes may have occurred.

The majority of SNPs appear to be unaffected by selection despite being specifically chosen based on assumptions of selection as well as high population differentiation. This shows that a large proportion of the genetic and phenotypic variation is the result of differential demographic processes,

drift, and admixture. However, for a subset of markers strong deviations from neutrality were found, and changes in frequencies could be attributed to different periods in time. The results of this study strongly indicate that lowered UV radiation, colder climate and the local pathogen load were driving factors of selection outside of Africa. In addition, the increasing pathogen load and the smaller dietary spectrum associated with the Neolithic transition further increased existent selective pressures.

The vast majority of selection events that dated prior to the Neolithic correlated with a range of climate variables. As humans moved to higher latitudes, they experienced a colder climate with a stronger seasonality. Adaptations to cold-stress are known from several mammals to varying degrees [Makinen, 2010]. Selection in response to a colder climate could explain changes in allele frequencies for a large fraction of SNPs associated with the energy metabolism in this data set. Similarly, increased metabolic rates can be found in humans inhabiting colder regions in comparison to people from warmer climate zones [Makinen, 2010]. Especially brown adipose tissue (BAT) has an important role in cold tolerance and higher levels are known to occur in Central European populations compared to South-East Asians [Bakker et al., 2014]. This indicates that the lipid metabolism could have been under selection for a better distribution of brown adipose tissue rather than increased fat storage. For several SNPs in related genes (e.g., FTO) signs of an early selective sweep, predating the Mesolithic, were found. It appears that while some genes were under selection in all non-African groups of the data set, others have only been affected in European populations. This may be the result of selection in response to the local climate. For other SNPs related to the energy and lipid metabolism frequencies found in the Central European hunter-gatherers were highly similar to those found in Central Europe today. Also other genes may have been under selection because of an association with BAT. A recent study suggested a link of the KITLG gene to the BAT metabolism [Yang et al., 2018], which is otherwise mainly associated with pigmentation phenotypes [Ainger et al., 2017].

In addition to a colder climate, living at higher latitudes is accompanied by lower UV radiation. This is assumed to have created a substantial selective pressure on the vitamin D metabolism. Vitamin D has several biological effects and is highly associated with the absorption of calcium and thereby plays a crucial role in bone formation and general bone health [Holick, 2004]. In this study strong evidence could be found that the reduced vitamin D synthesis at higher latitudes led to selection on the vitamin D metabolism at several levels. Variants in the genes CYP3A4 and CYP3A5 seemed to be selected shortly after humans migrated out of Africa, over 80kya. While the direct effect of the two variants is unclear, both use vitamin D as substrate and are directly involved in the vitamin D metabolism [Baranyai, 2014]. Skin depigmentation also seems to have started during that time. Evidence for selection was found for a variant in the ASIP gene that was at fixation in all non-African populations of this data set. In addition, it could be shown that depigmentation increased as humans migrated further north. In comparison to the African reference group, an increased number of derived alleles were present in the Mesolithic hunter-gatherer groups. In comparison to modern-day populations of Central Europe the frequencies were lower though, indicating a darker complexion for the hunter-gatherers. Depigmentation is believed to be the result of selection towards a higher functioning vitamin D synthesis at lower UV radiation [Jablonski and

Chaplin, 2012]. Furthermore, other variants located in genes such as HMGA2 with an association to bone mineral density appear to have been selected during the same time span. Interestingly, variants associated with a light hair and eye color in genes such as OCA2, HERC2 and TYRP1 were present in the Mesolithic in Central Europe, yet not in Eastern Europe/Russia. Selection was indicated for several of these variants by neutrality test results. It is still unclear why these variants may have been selected since no phenotypic advantage of light hair or eyes is known today [Sturm and Duffy, 2012]. Additional research is needed for a better understanding of the functions of the genes involved.

In addition to selection in response to a colder climate and lower UV radiation, several immune-related variants associated with an increased pathogen resistance seemed to have undergone selection prior to the Mesolithic. A variant in CASP12 showed signs of an early selective sweep, probably shortly after humans left Africa. Further evidence indicated that selection affected variants in the PULS1 gene in European populations prior to the Mesolithic. As the local pathogen level correlates with climate variables such as temperature and humidity [Hancock et al., 2011], it is possible that selection occurred in response to local pathogen levels during migrations.

While the majority of selection events observed in this study that pre-date the Neolithic transition could be correlated with climate variables, selection occurring after the Neolithic transition is strongly influenced by additional factors that arose in connection to the adoption of the new agriculturalist lifestyle. This new lifestyle was characterized by living sedentary in close proximity to livestock, increasing population sizes and a cereal-rich diet. This newly created environment with its altered living conditions is assumed to have further increased already existing selective pressures. In the data of this study, the influence of the Neolithic on the genomes of early humans in Central Europe is clearly visible.

Several immune-related SNPs were indicated as selected after the onset of farming. Interestingly, differences could be found regarding the phenotypic relevance of the SNPs. Sings for selection on variants in SH2B3, RNASET2 and FUT1/2 could be dated to the early Neolithic, while variants in TLR genes seemed to be under selection during the Bronze Age and even later periods. These results indicate that different selective pressures affected the immune system during these different time periods. While the SNPs in SH2B3 and RNASET2 are associated with an increased inflammatory response following infections [Zhernakova et al., 2010, Thorn et al., 2012], the variants in FUT1/2 increase protection against norovirus infections [Smyth et al., 2011, Parmar et al., 2012]. A connection to sanitary conditions of early sedentary farming life is very likely. In contrast, the variants in the TLR genes are associated with pathogen pattern recognition and an increased protection against crowd diseases [Delneste et al., 2007, Schurz et al., 2015]. This may be the result of increasing population sizes and higher mobility during later periods [Armelagos et al., 1991, Linderholm et al., 2014, Sjögren et al., 2016], as population sizes were likely too small for crowd diseases to persist until after the Neolithic transition [Harper and Armelagos, 2010].

The change in diet proposed for the Neolithic [Richards et al., 2003] also seems to have had a sub-

stantial influence on the genomes of the early farmers investigated in this study. This influence was visible on multiple levels. The ingestion of food stored prior to its consumption appears to have created a selective pressure on genes involved in the drug metabolism. Food storage is believed to have led to the accumulation of toxins from fungi and bacteria. It was proposed that this led to adaptations in cytochrome P450 genes that created a higher tolerance towards those toxins [Magalon et al., 2008, Luca et al., 2008, Janha et al., 2014]. Results for several SNPs in the genes NAT2 and CYP2C19 confirm these assumptions and signs of diversifying selection could be observed for both genes in the data set, dating to the early Neolithic.

It appears that the dietary spectrum of the foods consumed furthermore had an influence on genetic variation in early humans in Europe. Compared to contemporary hunter-gatherers, early farmer communities are assumed to have had a reduced dietary spectrum [Richards and Hedges, 1999, Fischer et al., 2007, Oelze et al., 2011], resulting in a reduction in micro-nutrients. The depletion of vitamin D is assumed to have further increased the already present selective pressure on the vitamin D metabolism. This could explain the signs of selection found for variants in the NOTCH2 and LEPR genes, that correlated to the early Neolithic. Both genes have a known association with bone mineral density and osteoporosis [Kapur et al., 2010, Stathopoulos et al., 2013]. In addition, strong evidence was found for selection on pigmentation-related genes between the early Neolithic and the Bronze Age. Signs of non-neutrality for variants in the gene SLC24A5 indicate selection for the gene. The majority of derived alleles located in this gene were already present at high frequencies in the early farmers of this data set. It appears that strong directional selection affected these variants shortly after humans adopted farming. In contrast, a much higher variation was still present in the hunter-gatherers groups, probably due to the higher vitamin D content in their diet. The data further shows that other genes were affected by selection during later periods. Derived alleles increased in frequency in the TYR gene until they reached modern-day levels in the Bronze Age group. Variation in the SLC45A2 gene was also affected during the Neolithic, but selection probably started later compared to TYR. Although derived alleles were present at high frequencies in the Bronze Age group, variation was still higher compared to today. This indicates that selection may have affected the gene beyond the Bronze Age. Nevertheless, a pigmentation phenotype similar to today can be assumed for the individuals from the Bronze Age group based on the presented data.

Further evidence for adaptation to the Neolithic farming-based diet through selection was found in variants in the FADS1/2 genes. The gene is known to influence the levels of long-chained polyunsaturated fatty acids (LCPUFSs) in breast milk [Morales et al., 2011]. While LCPUFAs can be supplemented from animal protein, they can also be synthesized. In this case, an adaptation to a diet depleted in LCPUFAs, similar to other populations [Fumagalli et al., 2015, Kothapalli et al., 2016], is likely. Increased synthesis of LCPUFAs would have the benefit of raising the nutritional value of breast milk for the newborn. Selection on additional, diet-related SNPs could be traced to the Bronze Age. One of the strongest signals of selection was found for SNPs related to lactase persistence into adulthood, confirming previous studies [Bersaglieri et al., 2004b]. This is likely connected to dairying. The selection on lactase persistence in Europe was previously described as an example

for gene-culture coevolution [Gerbault et al., 2011]. In addition, a variant in the RAB3GAP1 gene was selected with similar strength during the same time. The variant has beneficial influences on lipid levels in the blood and reduces cholesterol [Bersaglieri et al., 2004b]. Since the allele was found in linkage with the lactase variant and milk is a rich source of fat, selection in the same context is very likely.

This study investigated the influences that affected human genomes during the population history of Europe. It provides insights into the environmental factors that affected genetic variation and thus increases the understanding of the role of the Neolithic in this process. Challenges during this work were the scarcity of high quality data. While a data set of over 100 individuals is not small, some groups were underrepresented. The sample quality of the Middle Neolithic group was lower compared to the other groups, which led to exclusion from some of the analyses. An extensive analysis of more samples of this period is of particular interest as admixture between early farmers and local hunter-gatherers already occurred at this point in time [Lipson et al., 2017]. Such data would allow the investigation of the influence of admixture on alleles that were presumably under selection. Decreasing sampling costs will hopefully contribute to providing a higher numbers of full-genomes sequences of ancient samples, enabling future research to be conducted on data sets with a higher resolution.

Increasing the sample size of available data would furthermore enable a finer resolution of studies such as this one. In addition, approaches as in Veeramah et al. [2018] or Amorim et al. [2018], where multiple samples from the same site and time are sequenced, provide new opportunities. Higher resolution data will allow for a more detailed analysis of local adaptations in the future. Full genomes also allow the direct application of haplotype based tests on ancient samples. With refined laboratory methods it will be possible to acquire data from individuals of under-sampled areas. Data from Mesolithic samples pre-dating the adoption of agriculture in the core-zone of the Neolithic are still scarce. Such samples would allow exploring the possibility that effects commonly attributed to the Neolithic have a different origin, investigating if, for example, depigmentation was already advanced in the ancestral populations of the first farmers. With dense time-series data of ancient genomes it might be possible to narrow-down the origin and trace the subsequent spread of variants that are assumed to have originated in Europe.

To further investigate the impact of the Neolithic on the immune system, environmental DNA offers a new opportunity. Broad-scale sampling could increase the knowledge about the pathogen environment during that time. By correlating the environmental pathogen load with the genetic variation in humans, the coevolution of the host-defense mechanism could be studied directly. DNA from dental calculus can also provide insight in the same context. In addition, also epi-genetic research based on ancient samples could contribute to the understanding of the recent evolution of the immune system. Methylation patterns from ancient genomes could be used to investigate differential gene expression. It has been shown that differences in methylation patterns between archaic hominins and anatomical modern humans were associated with differential adaptations [Gokhman et al., 2014]. A better

understanding of gene functions and their networks would also allow a more focused investigation of selection. Detailed knowledge about the interaction of genes with different environmental factors could increase the understanding of why and how selection occurred.

An increased focus on archaeological and isotopic evidence should also be considered to improve the understanding of regional lifestyles and how they may have influenced local environments. This could help to investigate local differences and their effects. For example, it is still discussed if lactase persistence occurred in communities that practiced dairying or if dairying increased in response to people being able to digest milk [Gerbault et al., 2011]. Only an interdisciplinary approach to the investigation of the past can provide a detailed picture.

References

Afifi Mohammed A, Jiman-Fatani Asif A, El Saadany Sherif, and Fouad Mahmoud A. Parasites–allergy paradox: Disease mediators or therapeutic modulators. *Journal of microscopy and ultrastructure*, 3(2):53–61, 2015.

Ahn Jiyoung, Yu Kai, Stolzenberg-Solomon Rachael, Simon K Claire, McCullough Marjorie L, Gallicchio Lisa, Jacobs Eric J, Ascherio Alberto, Helzlsouer Kathy, Jacobs Kevin B, and others . Genome-wide association study of circulating vitamin d levels. *Human molecular genetics*, 19(13):2739–2745, 2010.

Ainger Stephen A, Jagirdar Kasturee, Lee Katie J, Soyer H Peter, and Sturm Richard A. Skin pigmentation genetics for the clinic. *Dermatology*, 233(1):1–15, 2017.

Akey Joshua M, Zhang Ge, Zhang Kun, Jin Li, and Shriver Mark D. Interrogating a high-density snp map for signatures of natural selection. *Genome research*, 12(12):1805–1814, 2002.

Alexander David H, Novembre John, and Lange Kenneth. Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, 2009.

Allen John S and Cheer Susan M. The non-thrifty genotype. *Current anthropology*, 37(5):831–842, 1996.

Allentoft Morten E, Sikora Martin, Sjögren Karl-Göran, Rasmussen Simon, Rasmussen Morten, Stenderup Jesper, Damgaard Peter B, Schroeder Hannes, Ahlström Torbjörn, Vinner Lasse, and others . Population genomics of bronze age eurasia. *Nature*, 522(7555):167, 2015.

Alonso Alegre, Izagirre Arribalzaga Neskuts, Smith Zubiaga Isabel, Gardeazabal García Jesús, Díaz Ramón José Luis, Díaz Pérez José Luis, Zelenika Diana, Boyano López María Dolores, Smit Nico, and Rúa Vaca De la Concepción. Complex signatures of selection for the melanogenic loci tyr, tyrp1 and dct in humans. 2008.

Alper Scott, Laws Rebecca, Lackford Brad, Boyd Windy A, Dunlap Paul, Freedman Jonathan H, and Schwartz David A. Identification of innate immunity genes and pathways using a comparative genomics approach. *Proceedings of the national academy of sciences*, 105(19):7016–7021, 2008.

Amorim Carlos Eduardo G, Vai Stefania, Posth Cosimo, Modi Alessandra, Koncz István, Hakenbeck Susanne, La Rocca Maria Cristina, Mende Balazs, Bobo Dean, Pohl Walter, and others . Understanding 6th-century barbarian social organization and migration through paleogenomics. *bioRxiv*, page 268250, 2018.

Aranow Cynthia. Vitamin d and the immune system. *Journal of investigative medicine*, 59(6):881–886, 2011.

Armelagos George J. The paleolithic disease-scape, the hygiene hypothesis, and the second epidemiological transition. In *The hygiene hypothesis and Darwinian medicine*, pages 29–43. Springer, 2009.

Armelagos George J, Goodman Alan H, and Jacobs Kenneth H. The origins of agriculture: Population growth during a period of declining health. *Population and environment*, 13(1):9–22, 1991.

Aronesty Erik. Comparison of sequencing utility programs. *The Open Bioinformatics Journal*, 7(1), 2013.

Ash Abigail, Francken Michael, Pap Ildikó, Tvrdỳ Zdeněk, Wahl Joachim, and Pinhasi Ron. Regional differences in health, diet and weaning patterns amongst the first neolithic farmers of central europe. *Scientific Reports*, 6:29458, 2016.

Ayub Qasim, Moutsianas Loukas, Chen Yuan, Panoutsopoulou Kalliope, Colonna Vincenza, Pagani Luca, Prokopenko Inga, Ritchie Graham RS, Tyler-Smith Chris, McCarthy Mark I, and others . Revisiting the thrifty gene hypothesis via 65 loci associated with susceptibility to type 2 diabetes. *The American Journal of Human Genetics*, 94(2): 176–185, 2014.

Bains Ripudaman K, Kovacevic Mirna, Plaster Christopher A, Tarekegn Ayele, Bekele Endashaw, Bradman Neil N, and Thomas Mark G. Molecular diversity and population structure at the cytochrome p450 3a5 gene in africa. *BMC genetics*, 14(1):34, 2013.

Bakker Leontine EH, Boon Mariëtte R, Linden van der Rianne AD, Arias-Bouda Lenka Pereira, Klinken van Jan B, Smit Frits, Verberne Hein J, Jukema J Wouter, Tamsma Jouke T, Havekes Louis M, and others . Brown adipose tissue volume in healthy lean south asian adults compared with white caucasians: a prospective, case-controlled observational study. *The lancet Diabetes & endocrinology*, 2(3):210–217, 2014.

Balam-Ortiz Eros, Esquivel-Villarreal Adolfo, Huerta-Hernandez David, Fernandez-Lopez Juan Carlos, Alfaro-Ruiz Luis, Muñoz-Monroy Omar, Gutierrez Ruth, Figueroa-Genis Enrique, Carrillo Karol, Elizalde Adela, and others . Hypercontrols in genotype-phenotype analysis reveal ancestral haplotypes associated with essential hypertension. *Hypertension*, pages Hypertension AHA – 111, 2012.

Baranyai Dorothea. *Die Expression und Regulation der CYP3A in menschlicher Haut: deren mögliche Beteiligung am Vitamin D-Stoffwechsel als Ursache für die positive Selektion des CYP3A-Lokus in Europäern*. PhD thesis, Universitätsbibliothek Mainz, 2014.

Barreiro Luis B, Laval Guillaume, Quach Hélene, Patin Etienne, and Quintana-Murci Lluís. Natural selection has driven population differentiation in modern humans. *Nature genetics*, 40(3):340, 2008.

Barsh Greg, Gunn Teresa, He Lin, Schlossman Stuart, and Duke-Cohan Jonathan. Biochemical and genetic studies of pigment-type switching. *Pigment cell research*, 13:48–53, 2000.

Baumgart Daniel C and Sandborn William J. Crohn's disease. *The Lancet*, 380(9853):1590–1605, 2012.

Behr Aaron A, Liu Katherine Z, Liu-Fang Gracie, Nakka Priyanka, and Ramachandran Sohini. pong: Fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics*, 32(18):2817–2823, 2016.

Beleza Sandra, Santos António M, McEvoy Brian, Alves Isabel, Martinho Cláudia, Cameron Emily, Shriver Mark D, Parra Esteban J, and Rocha Jorge. The timing of pigmentation lightening in europeans. *Molecular biology and evolution*, 30(1):24–35, 2012.

Bellwood Peter and Oxenham Marc. The expansions of farming societies and the role of the neolithic demographic transition. In *The Neolithic demographic transition and its consequences*, pages 13–34. Springer, 2008.

Bentley R Alexander, Price T Douglas, Lning Jens, Gronenborn Detlef, Wahl Joachim, and Fullagar PaulD. Prehistoric migration in europe: strontium isotope analysis of early neolithic skeletons. *Current Anthropology*, 43(5):799–804, 2002.

Bentley R Alexander, Bickle Penny, Fibiger Linda, Nowell Geoff M, Dale Christopher W, Hedges Robert EM, Hamilton Julie, Wahl Joachim, Francken Michael, Grupe Gisela, and others . Community differentiation and kinship among europe's first farmers. *Proceedings of the National Academy of Sciences*, 109(24):9326–9330, 2012.

Benyshek Daniel C and Watson James T. Exploring the thrifty genotype's food-shortage assumptions: a cross-cultural comparison of ethnographic accounts of food security among foraging and agricultural societies. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, 131(1): 120–126, 2006.

Berglund Eric D, Liu Tiemin, Kong Xingxing, Sohn Jong-Woo, Vong Linh, Deng Zhuo, Lee Charlotte E, Lee Syann, Williams Kevin W, Olson David P, and others . Melanocortin 4 receptors in autonomic neurons regulate thermogenesis and glycemia. *Nature neuroscience*, 17(7):911, 2014.

Bernard Helen, Faber Mirko, Wilking Hendrik, Haller S, Höhle Michael, Schielke Anika, Ducomble Tanja, Siffczyk C, Merbecks SS, Fricke G, and others . Large multistate outbreak of norovirus gastroenteritis associated with frozen strawberries, east germany, 2012. *Eurosurveillance*, 19(8), 2014.

Bersaglieri Todd, Sabeti Pardis C, Patterson Nick, Vanderploeg Trisha, Schaffner Steve F, Drake Jared A, Rhodes Matthew, Reich David E, and Hirschhorn Joel N. Genetic signatures of strong recent positive selection at the lactase gene. *The American Journal of Human Genetics*, 74(6):1111–1120, 2004a.

Bersaglieri Todd, Sabeti Pardis C, Patterson Nick, Vanderploeg Trisha, Schaffner Steve F, Drake Jared A, Rhodes Matthew, Reich David E, and Hirschhorn Joel N. Genetic signatures of strong recent positive selection at the lactase gene. *The American Journal of Human Genetics*, 74(6):1111–1120, 2004b.

Bhatia Gaurav, Patterson Nick J, Sankararaman Sriram, and Price Alkes L. Estimating and interpreting fst: the impact of rare variants. *Genome research*, pages gr–154831, 2013.

Bickle Penny and Whittle Alasdair. *The first farmers of central Europe: diversity in LBK lifeways*. Oxbow Books, 2013.

Bikle Daniel D. Vitamin d and the immune system: role in protection against bacterial infection. *Current opinion in nephrology and hypertension*, 17(4):348–352, 2008.

Bitti PP, Murgia BS, Ticca A, Ferrai R, Musu L, Piras ML, Puledda E, Campo S, Durando S, Montomoli C, and others . Association between the ancestral haplotype hla a30b18dr3 and multiple sclerosis in central sardinia. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society*, 20(2):271–283, 2001.

Blaustein Mordecai P, Zhang Jin, Chen Ling, and Hamilton Bruce P. How does salt retention raise blood pressure? *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 290(3):R514–R523, 2006.

Bochud Murielle, Bovet Pascal, Burnier Michel, and Eap Chin B. Cyp3a5 and abcb1 genes and hypertension. 2009.

Bocquet-Appel Jean-Pierre. The agricultural demographic transition during and after the agriculture inventions. *Current Anthropology*, 52(S4):S497–S510, 2011.

Bogaard Amy, Fraser Rebecca, Heaton Tim HE, Wallace Michael, Vaiglova Petra, Charles Michael, Jones Glynis, Evershed Richard P, Styring Amy K, Andersen Niels H, and others . Crop manuring and intensive land management by europe's first farmers. *Proceedings of the National Academy of Sciences*, 110(31):12589–12594, 2013.

Bollongino Ruth, Nehlich Olaf, Richards Michael P, Orschiedt Jörg, Thomas Mark G, Sell Christian, Fajkošová Zuzana, Powell Adam, and Burger Joachim. 2000 years of parallel societies in stone age central europe. *Science*, page 1245049, 2013.

Bonsall C, Cook GT, Hedges REM, Higham TFG, Pickard C, and Radovanović I. Radiocarbon and stable isotope evidence of dietary change from the mesolithic to the middle ages in the iron gates: new results from lepenski vir. *Radiocarbon*, 46(1):293–300, 2004.

Boroneanţ A, Bonsall Clive, Kogarlniceanu R, Curcar R, Gligor M, and Stratton S. Burial practices in the iron gates mesolithic. In *Homines, Funera, Astra. Proceedings of the International Symposium on Funerary Anthropology. Oxford: Archaeopress, BAR International Series*, volume 2410, pages 45–56, 2012.

Boroneanţ Adina and Dinu Alexandru. The romanian mesolithic and the transition to farming. a case study: The iron gates. *Studii de Preistorie*, 3:41–76, 2006.

Botigué Laura R, Song Shiya, Scheu Amelie, Gopalan Shyamalika, Pendleton Amanda L, Oetjens Matthew, Taravella Angela M, Seregély Timo, Zeeb-Lanz Andrea, Arbogast Rose-Marie, and others . Ancient european dog genomes reveal continuity since the early neolithic. *Nature communications*, 8:16082, 2017.

Brace Selina, Diekmann Yoan, Booth Thomas J., Faltyskova Zuzana, Rohland Nadin, Mallick Swapan, Ferry Matthew, Michel Megan, Oppenheimer Jonas, Broomandkhoshbacht Nasreen, Stewardson Kristin, Walsh Susan, Kayser Manfred, Schulting Rick, Craig Oliver E, Sheridan Alison, Pearson Mike Parker, Stringer Chris, Reich David, Thomas Mark G, and Barnes Ian. Population replacement in early neolithic britain. *bioRxiv*, page 267443, 2018.

Bramanti Barbara, Thomas Mark G, Haak Wolfgang, Unterländer Martina, Jores Pia, Tambets Kristiina, Antanaitis-Jacobs Indre, Haidle Miriam N, Jankauskas Rimantas, Kind C-J, and others . Genetic discontinuity between local hunter-gatherers and central europe's first farmers. *Science*, 326(5949):137–140, 2009.

Branicki Wojciech, Brudnik Urszula, and Wojas-Pelc Anna. Interactions between herc2, oca2 and mc1r may influence human pigmentation phenotype. *Annals of human genetics*, 73(2):160–170, 2009.

Breyer Joan P, McReynolds Kate M, Yaspan Brian L, Bradley Kevin M, Dupont William D, and Smith Jeffrey R. Genetic variants and prostate cancer risk: candidate replication and exploration of viral restriction genes. *Cancer Epidemiology and Prevention Biomarkers*, 18(7):2137–2144, 2009.

Brinker Ute, Flohr Stefan, Piek Jürgen, and Orschiedt Jörg. Human remains from a bronze age site in the tollense valley. *The Routledge Handbook of the Bioarchaeology of Human Conflict*, 2013.

Brinkworth Jessica F and Barreiro Luis B. The contribution of natural selection to present-day susceptibility to chronic inflammatory and autoimmune disease. *Current opinion in immunology*, 31:66–78, 2014.

Brodin Petter and Davis Mark M. Human immune system variation. *Nature reviews immunology*, 17(1):21, 2017.

Broushaki Farnaz, Thomas Mark G., Link Vivian, López Saioa, Dorp van Lucy, Kirsanow Karola, Hofmanová Zuzana, Diekmann Yoan, Cassidy Lara M., Molino Díez-del David, Kousathanas Athanasios, Sell Christian, Robson Harry K., Martiniano Rui, Blöcher Jens, Scheu Amelie, Kreutzer Susanne, Bollongino Ruth, Bobo Dean, Davoudi Hossein, Munoz Olivia, Currat Mathias, Abdi Kamyar, Biglari Fereidoun, Craig Oliver E., Bradley Daniel G., Shennan Stephen, Veeramah Krishna R., Mashkour Marjan, Wegmann Daniel, Hellenthal Garrett, and Burger Joachim. Early neolithic genomes from the eastern fertile crescent. *Science*, page aaf7943, 2016.

Brown Audrey E and Walker Mark. Genetics of insulin resistance and the metabolic syndrome. *Current cardiology reports*, 18(8):75, 2016.

Brown Morris J. Hypertension and ethnic group. *Bmj*, 332(7545):833–836, 2006.

Budd Chelsea, Lillie Malcolm, Alpaslan-Roodenberg Songül, Karul Necmi, and Pinhasi Ron. Stable isotope analysis of neolithic and chalcolithic populations from aktopraklık, northern anatolia. *Journal of archaeological science*, 40 (2):860–867, 2013.

Burger Joachim, Kirchner Martina, Bramanti Barbara, Haak Wolfgang, and Thomas Mark G. Absence of the lactase-persistence-associated allele in early neolithic europeans. *Proceedings of the National Academy of Sciences*, 104(10): 3736–3741, 2007.

Campbell Neil A, Reece Jane B, and Markl Jürgen. Biologie. 6. *Aufl. Spektrum Akademischer, Heidelberg*, 2003.

Candille Sophie I, Absher Devin M, Beleza Sandra, Bauchet Marc, McEvoy Brian, Nanibaa A Garrison, Li Jun Z, Myers Richard M, Barsh Gregory S, Tang Hua, and others . Genome-wide association studies of quantitatively measured skin, hair, and eye pigmentation in four european populations. *PLoS One*, 7(10):e48294, 2012.

Canfield Victor A, Berg Arthur, Peckins Steven, Wentzel Steven M, Ang Khai Chung, Oppenheimer Stephen, and Cheng Keith C. Molecular phylogeography of a human autosomal skin color locus under natural selection. *G3: Genes, Genomes, Genetics*, pages g3–113, 2013.

Carrera-Bastos Pedro, Fontes-Villalba Maelan, OKeefe James H, Lindeberg Staffan, and Cordain Loren. The western diet and lifestyle and diseases of civilization. *Res Rep Clin Cardiol*, 2(2):2–15, 2011.

Chaplin George. Geographic distribution of environmental factors influencing human skin coloration. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, 125(3): 292–302, 2004.

Cheer Susan M and Allen John S. Lactose digestion capacity and perceived symptomatic response after dairy product consumption in tokelau island migrants. *American Journal of Human Biology: The Official Journal of the Human Biology Association*, 9(2):233–246, 1997.

Chen Rong, Corona Erik, Sikora Martin, Dudley Joel T, Morgan Alex A, Moreno-Estrada Andres, Nilsen Geoffrey B, Ruau David, Lincoln Stephen E, Bustamante Carlos D, and others . Type 2 diabetes risk alleles demonstrate extreme directional differentiation among human populations, compared to other diseases. *PLoS genetics*, 8(4): e1002621, 2012.

Chen Tai C, Chimeh Farhad, Lu Zhiren, Mathieu Jeffrey, Person Kelly S, Zhang Anqi, Kohn Nathan, Martinello Stephen, Berkowitz Roger, and Holick Michael F. Factors that influence the cutaneous synthesis and dietary sources of vitamin d. *Archives of biochemistry and biophysics*, 460(2):213–217, 2007.

Christensen Jonas. Warfare in the european neolithic. *Acta Archaeologica*, 75(2):129–156, 2004.

Clark Peter U, Dyke Arthur S, Shakun Jeremy D, Carlson Anders E, Clark Jorie, Wohlfarth Barbara, Mitrovica Jerry X, Hostetler Steven W, and McCabe A Marshall. The last glacial maximum. *science*, 325(5941):710–714, 2009.

Cleave TL. The saccharine disease. *New Canaan, CT: Keats Publishing*, 36, 1974.

Cochran Gregory and Harpending Henry. *The 10,000 year explosion: How civilization accelerated human evolution.* Basic Books, 2009.

Colagiuri Stephen and Miller J Brand. The 'carnivore connection'–evolutionary aspects of insulin resistance. *European Journal of Clinical Nutrition*, 56(S1):S30, 2002.

Colin Bell A, Adair Linda S, and Popkin Barry M. Ethnic differences in the association between body mass index and hypertension. *American Journal of Epidemiology*, 155(4):346–353, 2002.

Congdon Peter. Estimating diabetes prevalence by small area in england. *Journal of Public Health*, 28(1):71–81, 2006.

Consortium 1000 Genomes Project and others . A global reference for human genetic variation. *Nature*, 526(7571):68, 2015.

Cook C Justin. The natural selection of infectious disease resistance and its effect on contemporary health. *Review of Economics and Statistics*, 97(4):742–757, 2015.

Coop Graham, Pickrell Joseph K, Novembre John, Kudaravalli Sridhar, Li Jun, Absher Devin, Myers Richard M, Cavalli-Sforza Luigi Luca, Feldman Marcus W, and Pritchard Jonathan K. The role of geography in human adaptation. *PLoS genetics*, 5(6):e1000500, 2009.

Cordain Loren, Eaton S Boyd, Sebastian Anthony, Mann Neil, Lindeberg Staffan, Watkins Bruce A, O'Keefe James H, and Brand-Miller Janette. Origins and evolution of the western diet: health implications for the 21st century. *The American journal of clinical nutrition*, 81(2):341–354, 2005.

Cornelis Marilyn C, Qi Lu, Zhang Cuilin, Kraft Peter, Manson JoAnn, Cai Tianxi, Hunter David J, and Hu Frank B. Joint effects of common genetic variants on the risk for type 2 diabetes in us men and women of european ancestry. *Annals of internal medicine*, 150(8):541–550, 2009.

Crawford Nicholas G, Kelly Derek E, Hansen Matthew EB, Beltrame Marcia H, Fan Shaohua, Bowman Shanna L, Jewett Ethan, Ranciaro Alessia, Thompson Simon, Lo Yancy, and others . Loci associated with skin pigmentation identified in african populations. *Science*, 358(6365):eaan8433, 2017.

Crombé Philippe and Robinson Erick. European mesolithic: geography and culture. In *Encyclopedia of Global Archaeology*, pages 2623–2645. Springer, 2014.

Daanen Hein AM and Van Marken Lichtenbelt Wouter D. Human whole body cold adaptation. *Temperature*, 3(1): 104–118, 2016.

Daly Kevin G., Maisano Delser Pierpaolo, Mullin Victoria E., Scheu Amelie, Mattiangeli Valeria, Teasdale Matthew D., Hare Andrew J., Burger Joachim, Verdugo Marta Pereira, Collins Matthew J., Kehati Ron, Erek Cevdet Merih, Bar-Oz Guy, Pompanon François, Cumer Tristan, Çakırlar Canan, Mohaseb Azadeh Fatemeh, Decruyenaere Delphine, Davoudi Hossein, Çevik Özlem, Rollefson Gary, Vigne Jean-Denis, Khazaeli Roya, Fathi Homa, Doost Sanaz Beizaee, Rahimi Sorkhani Roghayeh, Vahdati Ali Akbar, Sauer Eberhard W., Azizi Kharanaghi Hossein, Maziar Sepideh, Gasparian Boris, Pinhasi Ron, Martin Louise, Orton David, Arbuckle Benjamin S., Benecke Norbert, Manica Andrea, Horwitz Liora Kolska, Mashkour Marjan, and Bradley Daniel G. Ancient goat genomes reveal mosaic domestication in the fertile crescent. *Science*, 361(6397):85–88, 2018.

Delahanty Ryan J, Zhang Yanfeng, Bichell Terry Jo, Shen Wangzhen, Verdier Kelienne, Macdonald Robert L, Xu Lili, Boyd Kelli, Williams Janice, and Kang Jing-Qiong. Beyond epilepsy and autism: disruption of gabrb3 causes ocular hypopigmentation. *Cell reports*, 17(12):3115–3124, 2016.

Delneste Yves, Beauvillain Celine, and Jeannin Pascale. Innate immunity: structure and function of tlrs. *Medecine sciences: M/S*, 23(1):67–73, 2007.

Delong Robert. Gaba (a) receptor alpha5 subunit as a candidate gene for autism and bipolar disorder: a proposed endophenotype with parent-of-origin and gain-of-function features, with or without oculocutaneous albinism. *Autism*, 11(2):135–147, 2007.

Deng Lian and Xu Shuhua. Adaptation of human skin color in various populations. *Hereditas*, 155(1):1, 2018.

DePristo Mark A, Banks Eric, Poplin Ryan, Garimella Kiran V, Maguire Jared R, Hartl Christopher, Philippakis Anthony A, Del Angel Guillermo, Rivas Manuel A, Hanna Matt, and others . A framework for variation discovery and genotyping using next-generation dna sequencing data. *Nature genetics*, 43(5):491, 2011.

Diamond Jared. Evolution, consequences and future of plant and animal domestication. *Nature*, 418(6898):700, 2002.

Diamond Jared. The double puzzle of diabetes. *Nature*, 423(6940):599, 2003.

Diamond Jared M. *Guns, germs and steel: a short history of everybody for the last 13,000 years*. Random House, 1998.

Dobrovolskaya Maria V. Upper palaeolithic and late stone age human diet. *Journal of physiological anthropology and applied human science*, 24(4):433–438, 2005.

Donnelly Michael P, Paschou Peristera, Grigorenko Elena, Gurwitz David, Barta Csaba, Lu Ru-Band, Zhukova Olga V, Kim Jong-Jin, Siniscalco Marcello, New Maria, and others . A global view of the oca2-herc2 region and pigmentation. *Human genetics*, 131(5):683–696, 2012.

D'orazio John A, Nobuhisa Tetsuji, Cui Rutao, Arya Michelle, Spry Malinda, Wakamatsu Kazumasa, Igras Vivien, Kunisada Takahiro, Granter Scott R, Nishimura Emi K, and others . Topical drug rescue strategy and skin protection based on the role of mc1r in uv-induced tanning. *Nature*, 443(7109):340, 2006.

Driscoll Carlos A and Macdonald David W. Top dogs: wolf domestication and wealth. *Journal of biology*, 9(2):10, 2010.

Dubé Catherine, Rostom Alaa, Sy Richmond, Cranney Ann, Saloojee Navaaz, Garritty Chantelle, Sampson Margaret, Zhang Li, Yazdi Fatemeh, Mamaladze Vasil, and others . The prevalence of celiac disease in average-risk and at-risk western european populations: a systematic review. *Gastroenterology*, 128(4):S57–S67, 2005.

Duffy David L, Montgomery Grant W, Chen Wei, Zhao Zhen Zhen, Le Lien, James Michael R, Hayward Nicholas K, Martin Nicholas G, and Sturm Richard A. A three–single-nucleotide polymorphism haplotype in intron 1 of oca2 explains most human eye-color variation. *The American Journal of Human Genetics*, 80(2):241–252, 2007.

Dunn Robert R, Davies T Jonathan, Harris Nyeema C, and Gavin Michael C. Global drivers of human pathogen richness and prevalence. *Proceedings of the Royal Society B: Biological Sciences*, 277(1694):2587–2595, 2010.

Eiberg Hans, Troelsen Jesper, Nielsen Mette, Mikkelsen Annemette, Mengel-From Jonas, Kjaer Klaus W, and Hansen Lars. Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the herc2 gene inhibiting oca2 expression. *Human genetics*, 123(2):177–187, 2008.

Elliott David E, Urban Jr Joe F, Argo Curtis K, and Weinstock Joel V. Does the failure to acquire helminthic parasites predispose to crohn's disease? *The FASEB Journal*, 14(12):1848–1855, 2000.

Enattah Nabil Sabri, Sahi Timo, Savilahti Erkki, Terwilliger Joseph D, Peltonen Leena, and Järvelä Irma. Identification of a variant associated with adult-type hypolactasia. *Nature genetics*, 30(2):233, 2002.

Eriksson Gunilla, Lõugas Lembi, and Zagorska Ilga. Stone age hunter–fisher–gatherers at zvejnieki, northern latvia: radiocarbon, stable isotope and archaeozoology data. *Before Farming*, 2003(1):1–25, 2003a.

Eriksson Gunilla, Lõugas Lembi, and Zagorska Ilga. Stone age hunter–fisher–gatherers at zvejnieki, northern latvia: radiocarbon, stable isotope and archaeozoology data. *Before Farming*, 2003(1):1–25, 2003b.

Eshed Vered, Gopher Avi, Pinhasi Ron, and Hershkovitz Israel. Paleopathology and the origin of agriculture in the levant. *American Journal of Physical Anthropology*, 143(1):121–133, 2010.

Facchinello Nicola, Tarifeño-Saldivia Estefania, Grisan Enrico, Schiavone Marco, Peron Margherita, Mongera Alessandro, Ek Olivier, Schmitner Nicole, Meyer Dirk, Peers Bernard, and others . Tcf7l2 plays pleiotropic roles in the control of glucose homeostasis, pancreas morphology, vascularization and regeneration. *Scientific reports*, 7(1):9605, 2017.

Fedele Monica, Pierantoni Giovanna Maria, Visone Rosa, and Fusco Alfredo. Critical role of the hmga2 gene in pituitary adenomas. *Cell cycle*, 5(18):2045–2048, 2006.

Felsenstein Joseph. Theoretical evolutionary genetics joseph felsenstein. *University of Washington, Seattle*, 2005.

Ferreira Manuel AR, Mangino Massimo, Brumme Chanson J, Zhao Zhen Zhen, Medland Sarah E, Wright Margaret J, Nyholt Dale R, Gordon Scott, Campbell Megan, McEvoy Brian P, and others . Quantitative trait loci for cd4: Cd8 lymphocyte ratio are associated with risk of type 1 diabetes and hiv-1 immune control. *The American Journal of Human Genetics*, 86(1):88–92, 2010.

Ferrer-Admetlla Anna, Liang Mason, Korneliussen Thorfinn, and Nielsen Rasmus. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Molecular biology and evolution*, 31(5):1275–1291, 2014.

Field Yair, Boyle Evan A, Telis Natalie, Gao Ziyue, Gaulton Kyle J, Golan David, Yengo Loic, Rocheleau Ghislain, Froguel Philippe, McCarthy Mark I, and others . Detection of human adaptation during the past 2000 years. *Science*, 354(6313):760–764, 2016.

Fischer Anders, Richards Mike, Olsen Jesper, Robinson David Earle, Bennike Pia, Kubiak-Martens Lucyna, and Heinemeier Jan. The composition of mesolithic food: Evidence from the submerged settlement on the argus bank, denmark. *Acta Archaeologica*, 78(2):163–178, 2007.

Flannery Kent V and others . Origins and ecological effects of early domestication in iran and the near east. *Origins and ecological effects of early domestication in Iran and the Near East.*, 1969.

Florez Jose C, Hirschhorn Joel, and Altshuler David. The inherited basis of diabetes mellitus: implications for the genetic analysis of complex traits. *Annual review of genomics and human genetics*, 4(1):257–291, 2003.

Franke Andre, McGovern Dermot PB, Barrett Jeffrey C, Wang Kai, Radford-Smith Graham L, Ahmad Tariq, Lees Charlie W, Balschun Tobias, Lee James, Roberts Rebecca, and others . Genome-wide meta-analysis increases to 71 the number of confirmed crohn's disease susceptibility loci. *Nature genetics*, 42(12):1118, 2010.

Friedrich Deise C, Andrade de Fabiana M, Fiegenbaum Marilu, Almeida Silvana de, Mattevi Vanessa S, Callegari-Jacques Sidia M, and Hutz Mara H. The lactase persistence genotype is a protective factor for the metabolic syndrome. *Genetics and molecular biology*, 37(4):611–615, 2014.

Frost Peter. European hair and eye color: a case of frequency-dependent sexual selection? *Evolution and Human Behavior*, 27(2):85–103, 2006.

Fu Qiaomei, Hajdinjak Mateja, Moldovan Oana Teodora, Constantin Silviu, Mallick Swapan, Skoglund Pontus, Patterson Nick, Rohland Nadin, Lazaridis Iosif, Nickel Birgit, and others . An early modern human from romania with a recent neanderthal ancestor. *Nature*, 524(7564):216, 2015.

Fu Qiaomei, Posth Cosimo, Hajdinjak Mateja, Petr Martin, Mallick Swapan, Fernandes Daniel, Furtwängler Anja, Haak Wolfgang, Meyer Matthias, Mittnik Alissa, and others . The genetic history of ice age europe. *Nature*, 534 (7606):200, 2016.

Fu Wenqing, D OĆonnor Timothy, and Akey Joshua M. Genetic architecture of quantitative traits and complex diseases. *Current opinion in genetics & development*, 23(6):678–683, 2013.

Fumagalli Matteo, Sironi Manuela, Pozzoli Uberto, Ferrer-Admettla Anna, Pattini Linda, and Nielsen Rasmus. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS genetics*, 7(11):e1002355, 2011.

Fumagalli Matteo, Moltke Ida, Grarup Niels, Racimo Fernando, Bjerregaard Peter, Jørgensen Marit E, Korneliussen Thorfinn S, Gerbault Pascale, Skotte Line, Linneberg Allan, and others . Greenlandic inuit show genetic signatures of diet and climate adaptation. *Science*, 349(6254):1343–1347, 2015.

Furumura Minao, Sakai Chie, Abdel-Malek Zalfa, Barsh Gregory S, and Hearing Vincent J. The interaction of agouti signal protein and melanocyte stimulating hormone to regulate melanin formation in mammals. *Pigment cell research*, 9(4):191–203, 1996.

Fuselli Silvia, Filippo de Cesare, Mona Stefano, Sistonen Johanna, Fariselli Piero, Destro-Bisol Giovanni, Barbujani Guido, Bertorelle Giorgio, and Sajantila Antti. Evolution of detoxifying systems: the role of environment and population history in shaping genetic diversity at human cyp2d6 locus. *Pharmacogenetics and genomics*, 20(8): 485–499, 2010.

Gamba Cristina, Jones Eppie R, Teasdale Matthew D, McLaughlin Russell L, Gonzalez-Fortes Gloria, Mattiangeli Valeria, Domboróczki László, Kővári Ivett, Pap Ildikó, Anders Alexandra, and others . Genome flux and stasis in a five millennium transect of european prehistory. *Nature communications*, 5:5257, 2014.

Garn Stanley M and Leonard William R. What did our ancestors eat? *Nutrition Reviews*, 47(11):337–345, 1989.

Garud Nandita R, Messer Philipp W, Buzbas Erkan O, and Petrov Dmitri A. Recent selective sweeps in north american drosophila melanogaster show signatures of soft sweeps. *PLoS genetics*, 11(2):e1005004, 2015.

Gerbault Pascale, Liebert Anke, Itan Yuval, Powell Adam, Currat Mathias, Burger Joachim, Swallow Dallas M, and Thomas Mark G. Evolution of lactase persistence: an example of human niche construction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1566):863–877, 2011.

Gerritsen Fokke A, Özbal Rana, and Thissen Laurens C. The earliest neolithic levels at barcin höyük, northwestern turkey. *Anatolica*, 39:53–92, 2013.

Gesta Stephane, Tseng Yu-Hua, and Kahn C Ronald. Developmental origin of fat: tracking obesity to its source. *Cell*, 131(2):242–256, 2007.

Gheorghiu Dragos. Introduction. early pottery: a concise overview. In *Early farmers, late foragers, and ceramic traditions: on the beginning of pottery in the Near East and Europe*, pages 1–21. Cambridge Scholars Publishing, 2009.

Giardina Emiliano, Pietrangeli Ilenia, Martínez-Labarga Cristina, Martone Claudia, Angelis Flavio d, Spinella Aldo, Stefano Gianfranco De, Rickards Olga, and Novelli Giuseppe. Haplotypes in slc24a5 gene as ancestry informative markers in different populations. *Current genomics*, 9(2):110–114, 2008.

Gillespie John H. *Population genetics: a concise guide*. JHU Press, 2004.

Gnirke Andreas, Melnikov Alexandre, Maguire Jared, Rogov Peter, LeProust Emily M, Brockman William, Fennell Timothy, Giannoukos Georgia, Fisher Sheila, Russ Carsten, and others . Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature biotechnology*, 27(2):182, 2009.

Gokhman David, Lavi Eitan, Prüfer Kay, Fraga Mario F, Riancho José A, Kelso Janet, Pääbo Svante, Meshorer Eran, and Carmel Liran. Reconstructing the dna methylation maps of the neandertal and the denisovan. *Science*, 344 (6183):523–527, 2014.

Goldberg Amy, Günther Torsten, Rosenberg Noah A, and Jakobsson Mattias. Ancient x chromosomes reveal contrasting sex bias in neolithic and bronze age eurasian migrations. *Proceedings of the National Academy of Sciences*, 114 (10):2657–2662, 2017.

González Juan R, González-Carpio Marta, Hernández-Sáez Rosario, Vargas Victoria Serrano, Hidalgo Guadalupe Torres, Rubio-Rodrigo Marta, García-Nogales Ana, Estévez Manuela Núñez, Pérez Luis M Luengo, and Rodríguez-López Raquel. Fto risk haplotype among early onset and severe obesity cases in a population of western spain. *Obesity*, 20(4):909–915, 2012.

González-Fortes Gloria, Jones Eppie R, Lightfoot Emma, Bonsall Clive, Lazar Catalin, Grandal-d'Anglade Aurora, Garralda María Dolores, Drak Labib, Siska Veronika, Simalcsik Angela, and others . Paleogenomic evidence for multi-generational mixing between neolithic farmers and mesolithic hunter-gatherers in the lower danube basin. *Current Biology*, 27(12):1801–1810, 2017.

Green Richard E, Krause Johannes, Briggs Adrian W, Maricic Tomislav, Stenzel Udo, Kircher Martin, Patterson Nick, Li Heng, Zhai Weiwei, Fritz Markus Hsi-Yang, and others . A draft sequence of the neandertal genome. *Science*, 328(5979):710–722, 2010.

Greenwood Darren C, Threapleton Diane E, Evans Charlotte EL, Cleghorn Christine L, Nykjaer Camilla, Woodhead Charlotte, and Burley Victoria J. Glycemic index, glycemic load, carbohydrates, and type 2 diabetes: systematic review and dose–response meta-analysis of prospective studies. *Diabetes care*, 36(12):4166–4171, 2013.

Gross Lee S, Li Li, Ford Earl S, and Liu Simin. Increased consumption of refined carbohydrates and the epidemic of type 2 diabetes in the united states: an ecologic assessment. *The American journal of clinical nutrition*, 79(5): 774–779, 2004.

Günther Torsten, Valdiosera Cristina, Malmström Helena, Ureña Irene, Rodriguez-Varela Ricardo, Sverrisdóttir Óddny Osk, Daskalaki Evangelia A, Skoglund Pontus, Naidoo Thijessen, Svensson Emma M, and others . Ancient genomes link early farmers from atapuerca in spain to modern-day basques. *Proceedings of the National Academy of Sciences*, 112(38):11917–11922, 2015.

Günther Torsten, Malmström Helena, Svensson Emma M, Omrak Ayça, Sánchez-Quinto Federico, Kılınç Gülşah M, Krzewińska Maja, Eriksson Gunilla, Fraser Magdalena, Edlund Hanna, and others . Population genomics of mesolithic scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLoS biology*, 16(1):e2003703, 2018.

Gutenkunst Ryan N, Hernandez Ryan D, Williamson Scott H, and Bustamante Carlos D. Inferring the joint demographic history of multiple populations from multidimensional snp frequency data. *PLoS genetics*, 5(10):e1000695, 2009.

Haack Fabian. Die frühneolithische grabenanlage von herxheim bei landau: Architektur, verfüllungsprozesse und nutzungsdauer. 2016.

Haak Wolfgang, Lazaridis Iosif, Patterson Nick, Rohland Nadin, Mallick Swapan, Llamas Bastien, Brandt Guido, Nordenfelt Susanne, Harney Eadaoin, Stewardson Kristin, and others . Massive migration from the steppe was a source for indo-european languages in europe. *Nature*, 522(7555):207, 2015.

Haber Marc, Mezzavilla Massimo, Xue Yali, and Tyler-Smith Chris. Ancient dna and the rewriting of human history: be sparing with occam's razor. *Genome biology*, 17(1):1, 2016.

Hamilton Matthew. *Population genetics.* John Wiley & Sons, 2011.

Han Jiali, Kraft Peter, Nan Hongmei, Guo Qun, Chen Constance, Qureshi Abrar, Hankinson Susan E, Hu Frank B, Duffy David L, Zhao Zhen Zhen, and others . A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS genetics*, 4(5):e1000074, 2008.

Hancock Angela M, Witonsky David B, Ehler Edvard, Alkorta-Aranburu Gorka, Beall Cynthia, Gebremedhin Amha, Sukernik Rem, Utermann Gerd, Pritchard Jonathan, Coop Graham, and others . Human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. *Proceedings of the National Academy of Sciences*, page 200914625, 2010.

Hancock Angela M, Witonsky David B, Alkorta-Aranburu Gorka, Beall Cynthia M, Gebremedhin Amha, Sukernik Rem, Utermann Gerd, Pritchard Jonathan K, Coop Graham, and Di Rienzo Anna. Adaptations to climate-mediated selective pressures in humans. *PLoS genetics*, 7(4):e1001375, 2011.

Hannema Sabine E, Wit Jan M, Houdijk Mieke ECAM, Haeringen van Arie, Bik Elsa C, Verkerk Annemieke JMH, Uitterlinden André G, Kant Sarina G, Oostdijk Wilma, Bakker Egbert, and others . Novel leptin receptor mutations identified in two girls with severe obesity are associated with increased bone mineral density. *Hormone research in paediatrics*, 85(6):412–420, 2016.

Harper Kristin and Armelagos George. The changing disease-scape in the third epidemiological transition. *International Journal of Environmental Research and Public Health*, 7(2):675–697, 2010.

Hart Katie L, Kimura Shey L, Mushailov Vladimir, Budimlija Zoran M, Prinz Mechthild, and Wurmbach Elisa. Improved eye-and skin-color prediction based on 8 snps. *Croatian medical journal*, 54(3):248–256, 2013.

Helgason Agnar, Pálsson Snæbjörn, Thorleifsson Gudmar, Grant Struan FA, Emilsson Valur, Gunnarsdottir Steinunn, Adeyemo Adebowale, Chen Yuanxiu, Chen Guanjie, Reynisdottir Inga, and others . Refining the impact of tcf7l2 gene variants on type 2 diabetes and adaptive evolution. *Nature genetics*, 39(2):218, 2007.

Hellenthal Garrett, Busby George BJ, Band Gavin, Wilson James F, Capelli Cristian, Falush Daniel, and Myers Simon. A genetic atlas of human admixture history. *Science*, 343(6172):747–751, 2014.

Hervella Montserrat, Plantinga Theo S, Alonso Santos, Ferwerda Bart, Izagirre Neskuts, Fontecha Lara, Fregel Rosa, Meer van der Jos WM, Rúa de-la Concepcion, and Netea Mihai G. The loss of functional caspase-12 in europe is a pre-neolithic event. *PloS one*, 7(5):e37022, 2012.

Hewison Martin. Vitamin d and the immune system: new perspectives on an old theme. *Rheumatic Disease Clinics*, 38(1):125–139, 2012.

Heyd Volker. Yamnaya groups and tumuli west of the black sea. *Publications de la Maison de l'Orient et de la Méditerranée*, 58(1):535–555, 2012.

Heyd Volker. Kossinna's smile. *antiquity*, 91(356):348–359, 2017.

Heyman Melvin B. Lactose intolerance in infants, children, and adolescents. *Pediatrics*, 118(3):1279–1286, 2006.

Hivert Marie-France, Jablonski Kathleen A, Perreault Leigh, Saxena Richa, McAteer Jarred B, Franks Paul W, Hamman Richard F, Kahn Steven E, Haffner Steven, Meigs James B, and others . An updated genetic score based on 34 confirmed type 2 diabetes loci is associated with diabetes incidence and regression to normoglycemia in the diabetes prevention program. *Diabetes*, page DB_101119, 2011.

Hlusko Leslea J, Carlson Joshua P, Chaplin George, Elias Scott A, Hoffecker John F, Huffman Michaela, Jablonski Nina G, Monson Tesla A, O'Rourke Dennis H, Pilloud Marin A, and others . Environmental selection during the last ice age on the mother-to-infant transmission of vitamin d and fatty acids through breast milk. *Proceedings of the National Academy of Sciences*, 115(19):E4426–E4432, 2018.

Ho Herbert, Pinto Amar, Hall Stephen D., Flockhart David A., Li Lang, Skaar Todd C., Cadman Peter, O'Connor Daniel T., Wagner Urban, Fineberg Naomi S., and Weinberger Myron H. Association between the cyp3a5 genotype and blood pressure. *Hypertension*, 45(2):294–298, 2005.

Hofmanová Zuzana. *Palaeogenomic and Biostatistical Analysis of Ancient DNA Data from Mesolithic and Neolithic Skeletal Remains.* PhD thesis, Johannes Gutenberg-Universität Mainz, 2016.

Hofmanová Zuzana, Kreutzer Susanne, Hellenthal Garrett, Sell Christian, Diekmann Yoan, Molino Díez-del David, Dorp van Lucy, López Saioa, Kousathanas Athanasios, Link Vivian, Kirsanow Karola, Cassidy Lara M., Martiniano Rui, Strobel Melanie, Scheu Amelie, Kotsakis Kostas, Halstead Paul, Triantaphyllou Sevi, Kyparissi-Apostolika Nina, Urem-Kotsou Dushka, Ziota Christina, Adaktylou Fotini, Gopalan Shyamalika, Bobo Dean M., Winkelbach Laura, Blöcher Jens, Unterländer Martina, Leuenberger Christoph, Çilingiroğlu Çiler, Horejs Barbara, Gerritsen Fokke, Shennan Stephen J., Bradley Daniel G., Currat Mathias, Veeramah Krishna R., Wegmann Daniel, Thomas Mark G., Papageorgopoulou Christina, and Burger Joachim. Early farmers from across europe directly descended from neolithic aegeans. *Proceedings of the National Academy of Sciences*, 113(25):6886–6891, 2016.

Holick Michael F. Sunlight and vitamin d for bone health and prevention of autoimmune diseases, cancers, and cardiovascular disease. *The American journal of clinical nutrition*, 80(6):1678S–1688S, 2004.

Holick Michael F. Vitamin d deficiency. *New England Journal of Medicine*, 357(3):266–281, 2007.

Holick Michael F, Smith Ellen, and Pincus Stephanie. Skin as the site of vitamin d synthesis and target tissue for 1, 25-dihydroxyvitamin d3: use of calcitriol (1, 25-dihydroxyvitamin d3) for treatment of psoriasis. *Archives of dermatology*, 123(12):1677–1683a, 1987.

Holst Daniela. Hazelnut economy of early holocene hunter–gatherers: a case study from mesolithic duvensee, northern germany. *Journal of Archaeological Science*, 37(11):2871–2880, 2010.

Hublin Jean-Jacques. The origin of neandertals. *Proceedings of the National Academy of Sciences*, 106(38):16022–16027, 2009.

Hudjashov Georgi, Villems Richard, and Kivisild Toomas. Global patterns of diversity and selection in human tyrosinase gene. *PLoS One*, 8(9):e74307, 2013.

Hunt Karen A, Zhernakova Alexandra, Turner Graham, Heap Graham AR, Franke Lude, Bruinenberg Marcel, Romanos Jihane, Dinesen Lotte C, Ryan Anthony W, Panesar Davinder, and others . Newly identified genetic risk variants for celiac disease related to the immune response. *Nature genetics*, 40(4):395, 2008.

Ibarrola-Villava Maider, Fernandez Lara P, Pita Guillermo, Bravo Jerónimo, Floristan Uxua, Sendagorta Elena, Feito Marta, Avilés José A, Martin-Gonzalez Manuel, Lázaro Pablo, and others . Genetic analysis of three important genes in pigmentation and melanoma susceptibility: Cdkn2a, mc1r and herc2/oca2. *Experimental dermatology*, 19 (9):836–844, 2010.

Walde im Th Schulte, Freundlich JC, Schwabedissen Hermann, and Taute Wolfgang. Köln radiocarbon dates iii. *Radiocarbon*, 28(1):134–140, 1986.

Itan Yuval, Jones Bryony L, Ingram Catherine JE, Swallow Dallas M, and Thomas Mark G. A worldwide correlation of lactase persistence phenotype and genotypes. *BMC evolutionary biology*, 10(1):36, 2010.

Jablonski Nina G. Human skin pigmentation as an example of adaptive evolution. *Proceedings of the American Philosophical Society*, 156(1):45–57, 2012.

Jablonski Nina G and Chaplin George. The evolution of human skin coloration. *Journal of Human Evolution*, 39(1): 57–106, 2000.

Jablonski Nina G and Chaplin George. Human skin pigmentation, migration and disease susceptibility. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1590):785–792, 2012.

Jablonski Nina G and Chaplin George. The colours of humanity: the evolution of pigmentation in the human lineage. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1724):20160349, 2017.

Jagoda Evelyn, Lawson Daniel J, Wall Jeffrey D, Lambert David, Muller Craig, Westaway Michael, Leavesley Matthew, Capellini Terence D, Mirazón Lahr Marta, Gerbault Pascale, and others . Disentangling immediate adaptive introgression from selection on standing introgressed variation in humans. *Molecular biology and evolution*, 35(3): 623–630, 2017.

Janha Ramatoulie E, Worwui Archibald, Linton Kenneth J, Shaheen Seif O, Sisay-Joof Fatoumatta, and Walton Robert T. Inactive alleles of cytochrome p450 2c19 may be positively selected in human evolution. *BMC evolutionary biology*, 14(1):71, 2014.

Jantzen Detlef, Brinker Ute, Orschiedt Jörg, Heinemeier Jan, Piek Jürgen, Hauenstein Karlheinz, Krüger Joachim, Lidke Gundula, Lübke Harald, Lampe Reinhard, and others . A bronze age battlefield? weapons and trauma in the tollense valley, north-eastern germany. *Antiquity*, 85(328):417–433, 2011.

Jiang Song, Li Xinyan, Hess Nicholas J, Guan Yue, and Tapping Richard I. Tlr10 is a negative regulator of both myd88-dependent and-independent tlr signaling. *The Journal of Immunology*, 196(9):3834–3841, 2016.

Jones Emily Lena. Archaeofaunal evidence of human adaptation to climate change in upper paleolithic iberia. *Journal of Archaeological Science: Reports*, 2:257–263, 2015.

Jones Eppie R, Gonzalez-Fortes Gloria, Connell Sarah, Siska Veronika, Eriksson Anders, Martiniano Rui, McLaughlin Russell L, Llorente Marcos Gallego, Cassidy Lara M, Gamba Cristina, and others . Upper palaeolithic genomes reveal deep roots of modern eurasians. *Nature communications*, 6:8912, 2015.

Jones Glenville, Prosser David E, and Kaufmann Martin. 25-hydroxyvitamin d-24-hydroxylase (cyp24a1): its important role in the degradation of vitamin d. *Archives of biochemistry and biophysics*, 523(1):9–18, 2012.

Jowitt LM. Ethnicity and type 2 diabetes in pacific island adults in new zealand. *International Journal of Diabetes and Clinincal Reseaerch*, 1:1–5, 2014.

Kalsbeek Anya, Veenstra Jenna, Westra Jason, Disselkoen Craig, Koch Kristin, McKenzie Katelyn A, O'Bott Jacob, Vander Woude Jason, Fischer Karen, Shearer Greg C, and others . A genome-wide association study of red-blood cell fatty acids and ratios incorporating dietary covariates: Framingham heart study offspring cohort. *PloS one*, 13 (4):e0194882, 2018.

Kanetsky Peter A, Swoyer Jennifer, Panossian Saarene, Holmes Robin, Guerry DuPont, and Rebbeck Timothy R. A polymorphism in the agouti signaling protein gene is associated with human pigmentation. *The American Journal of Human Genetics*, 70(3):770–775, 2002.

Kapur Sonia, Amoui Mehran, Kesavan Chandrasekhar, Wang Xiaoguang, Mohan Subburaman, Baylink David J, and Lau K-H William. Leptin receptor (lepr) is a negative modulator of bone mechanosensitivity and genetic variations in lepr may contribute to the differential osteogenic response to mechanical stimulation in the c57bl/6j and c3h/hej pair of mouse strains. *Journal of Biological Chemistry*, 285(48):37607–37618, 2010.

Karlsson Elinor K, Kwiatkowski Dominic P, and Sabeti Pardis C. Natural selection and infectious disease in human populations. *Nature Reviews Genetics*, 15(6):379, 2014.

Kettunen Johannes, Silander Kaisa, Saarela Olli, Amin Najaf, Müller Martina, Timpson Nicholas, Surakka Ida, Ripatti Samuli, Laitinen Jaana, Hartikainen Anna-Liisa, and others . European lactase persistence genotype shows evidence of association with increase in body mass index. *Human molecular genetics*, 19(6):1129–1136, 2009.

Kilpeläinen Tuomas O, Qi Lu, Brage Soren, Sharp Stephen J, Sonestedt Emily, Demerath Ellen, Ahmad Tariq, Mora Samia, Kaakinen Marika, Sandholt Camilla Helene, and others . Physical activity attenuates the influence of fto variants on obesity risk: a meta-analysis of 218,166 adults and 19,268 children. *PLoS medicine*, 8(11):e1001116, 2011.

Kim Sun Mie, Kim Seok Hyun, Lee Jung Ryeol, Jee Byung Chul, Ku Seung-Yup, Suh Chang Suk, Choi Young Min, Kim Jung Gu, and Moon Shin Yong. Association of leptin receptor polymorphisms lys109arg and gln223arg with serum leptin profile and bone mineral density in korean women. *American journal of obstetrics and gynecology*, 198 (4):421–e1, 2008.

Kimura Motoo and Ohta Tomoko. The average number of generations until fixation of a mutant gene in a finite population. *Genetics*, 61(3):763, 1969.

Kircher Martin. Analysis of high-throughput ancient dna sequencing data. In *Ancient DNA*, pages 197–228. Springer, 2012.

Kircher Martin, Sawyer Susanna, and Meyer Matthias. Double indexing overcomes inaccuracies in multiplex sequencing on the illumina platform. *Nucleic acids research*, 40(1):e3–e3, 2011.

Koch-Institut Robert. Infektionsepidemiologisches jahrbuch meldepflichtiger krankheiten für 2017. *Robert Koch-Institut, Berlin*, 2017.

Koh Jung-Min, Kim Duk J, Hong Jeong S, Park Joong Y, Lee Ki-Up, Kim Shin-Yoon, and Kim Ghi S. Estrogen receptor alpha gene polymorphisms pvu ii and xba i influence association between leptin receptor gene polymorphism (gln223arg) and bone mineral density in young men. *European journal of endocrinology*, 147(6):777–783, 2002.

Köhler Kitti, Marcsik Antónia, Zádori Péter, Biro Gergely, Szeniczey Tamás, Fábián Szilvia, Serlegi Gábor, Marton Tibor, Donoghue Helen D, and Hajdu Tamás. Possible cases of leprosy from the late copper age (3780-3650 cal bc) in hungary. *PloS one*, 12(10):e0185966, 2017.

Kothapalli Kumar SD, Ye Kaixiong, Gadgil Maithili S, Carlson Susan E, O'brien Kimberly O, Zhang Ji Yao, Park Hui Gyu, Ojukwu Kinsley, Zou James, Hyon Stephanie S, and others . Positive selection on a regulatory insertion–deletion polymorphism in fads2 influences apparent endogenous synthesis of arachidonic acid. *Molecular biology and evolution*, 33(7):1726–1739, 2016.

Kousathanas Athanasios, Leuenberger Christoph, Link Vivian, Sell Christian, Burger Joachim, and Wegmann Daniel. Inferring heterozygosity from ancient and low coverage genomes. *Genetics*, pages genetics–116, 2016.

Kristiansen Kristian, Allentoft Morten E, Frei Karin M, Iversen Rune, Johannsen Niels N, Kroonen Guus, Pospieszny Lukasz, Price T Douglas, Rasmussen Simon, Sjögren Karl-Göran, Sikora Martin, and Willerslev Eske. Re-theorising mobility and the formation of culture and language among the corded ware culture in europe. *Antiquity*, 91(356): 334–347, 2017.

Kuipers Allison, Zhang Yingze, Cauley Jane A, Nestlerode Cara S, Chu Yanxia, Bunker Clareann H, Patrick Alan L, Wheeler Victor W, Hoffman Andrew R, Orwoll Eric S, and others . Association of a high mobility group gene (hmga2) variant with bone mineral density. *Bone*, 45(2):295–300, 2009.

Kuznetsov Igor B, McDuffie Michael, and Moslehi Roxana. A web server for inferring the human n-acetyltransferase-2 (nat2) enzymatic phenotype from nat2 genotype. *Bioinformatics*, 25(9):1185–1186, 2009.

Lamason Rebecca L, Mohideen Manzoor-Ali PK, Mest Jason R, Wong Andrew C, Norton Heather L, Aros Michele C, Jurynec Michael J, Mao Xianyun, Humphreville Vanessa R, Humbert Jasper E, and others . Slc24a5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science*, 310(5755):1782–1786, 2005.

Land Edward J and Riley Patrick A. Spontaneous redox reactions of dopaquinone and the balance between the eumelanic and phaeomelanic pathways. *Pigment Cell Research*, 13(4):273–277, 2000.

Lao O, De Gruijter JM, Duijn van Kate, Navarro Arcadi, and Kayser Manfred. Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. *Annals of human genetics*, 71(3):354–369, 2007.

Larsen Clark Spencer. The agricultural revolution as environmental catastrophe: Implications for health and lifestyle in the holocene. *Quaternary International*, 150(1):12–20, 2006.

Latham Katherine J. Human health and the neolithic revolution: an overview of impacts of the agricultural transition on oral health, epidemiology, and the human body. 2013.

Lazaridis Iosif, Patterson Nick, Mittnik Alissa, Renaud Gabriel, Mallick Swapan, Kirsanow Karola, Sudmant Peter H, Schraiber Joshua G, Castellano Sergi, Lipson Mark, and others . Ancient human genomes suggest three ancestral populations for present-day europeans. *Nature*, 513(7518):409, 2014.

Lazaridis Iosif, Nadel Dani, Rollefson Gary, Merrett Deborah C, Rohland Nadin, Mallick Swapan, Fernandes Daniel, Novak Mario, Gamarra Beatriz, Sirak Kendra, and others . Genomic insights into the origin of farming in the ancient near east. *Nature*, 536(7617):419, 2016.

Lee Hee Jun, Kim Hoon, Ku Seung-Yup, Choi Young Min, Kim Jong Hak, and Kim Jung Gu. Association between polymorphisms in leptin, leptin receptor, and β-adrenergic receptor genes and bone mineral density in postmenopausal korean women. *Menopause*, 21(1):67–73, 2014.

Lenneis Christine and Neugebauer-Maresch Eva. *Das linearbandkeramische Gräberfeld von Kleinhadersdorf*, volume 82. Verlag der Österreichischen Akademie der Wissenschaften, 2015.

Li Heng and Durbin Richard. Fast and accurate short read alignment with burrows–wheeler transform. *bioinformatics*, 25(14):1754–1760, 2009.

Li Heng, Handsaker Bob, Wysoker Alec, Fennell Tim, Ruan Jue, Homer Nils, Marth Gabor, Abecasis Goncalo, and Durbin Richard. The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–2079, 2009.

Li Jun Z, Absher Devin M, Tang Hua, Southwick Audrey M, Casto Amanda M, Ramachandran Sohini, Cann Howard M, Barsh Gregory S, Feldman Marcus, Cavalli-Sforza Luigi L, and others . Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, 319(5866):1100–1104, 2008.

Lidke Gundula. Untersuchungen zur bedeutung von gewalt und aggression im neolithikum deutschlands unter besonderer berücksichtigung norddeutschlands. 2005.

Liem Edwin B, Joiner Teresa V, Tsueda Kentaro, and Sessler Daniel I. Increased sensitivity to thermal pain and reduced subcutaneous lidocaine efficacy in redheads. *Anesthesiology: The Journal of the American Society of Anesthesiologists*, 102(3):509–514, 2005.

Lillie Malcolm and Jacobs Kenneth. 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, 2006.

Lin Jennifer Y and Fisher David E. Melanocyte biology and skin pigmentation. *Nature*, 445(7130):843, 2007.

Linderholm Anna, Fornander Elin, Eriksson Gunilla, Mörth Carl-Magnus, and Lidén Kerstin. Increasing mobility at the neolithic/bronze age transition-sulphur isotope evidence from öland, sweden. *Internet Archaeology*, 37, 2014.

Lindmark Månsson Helena. Fatty acids in bovine milk fat. *Food & nutrition research*, 52(1):1821, 2008.

Link Vivian, Kousathanas Athanasios, Veeramah Krishna, Sell Christian, Scheu Amelie, and Wegmann Daniel. Atlas: analysis tools for low-depth and ancient samples. *bioRxiv*, page 105346, 2017.

Lipson Mark, Szécsényi-Nagy Anna, Mallick Swapan, Pósa Annamária, Stégmár Balázs, Keerl Victoria, Rohland Nadin, Stewardson Kristin, Ferry Matthew, Michel Megan, and others . Parallel palaeogenomic transects reveal complex genetic history of early european farmers. *Nature*, 551(7680):368, 2017.

Littlewood Roland. Commentary: Globalization, culture, body image, and eating disorders. *Culture, medicine and psychiatry*, 28(4):597–602, 2004.

Liu Hong, Irwanto Astrid, Tian Hongqing, Fu Xi'an, Yu Yongxiang, Yu Gongqi, Low Huiqi, Chu Tongsheng, Li Yi, Shi Benqing, and others . Identification of il18rap/il18r1 and il12b as leprosy risk genes demonstrates shared pathogenesis between inflammation and infectious diseases. *The American Journal of Human Genetics*, 91(5): 935–941, 2012.

Liu Lijun, Li Jingjie, Yan Mengdan, Li Jing, Chen Junyu, Zhang Yi, Zhu Xikai, Wang Li, Kang Longli, Yuan Dongya, and others . Tcf7l2 polymorphisms and the risk of schizophrenia in the chinese han population. *Oncotarget*, 8(17): 28614, 2017.

Liu Simin, Willett Walter C, Stampfer Meir J, Hu Frank B, Franz Mary, Sampson Laura, Hennekens Charles H, and Manson JoAnn E. A prospective study of dietary glycemic load, carbohydrate intake, and risk of coronary heart disease in us women. *The American journal of clinical nutrition*, 71(6):1455–1461, 2000.

Loog Liisa, Thomas Mark G, Barnett Ross, Allen Richard, Sykes Naomi, Paxinos Ptolemaios D, Lebrasseur Ophélie, Dobney Keith, Peters Joris, Manica Andrea, and others . Inferring allele frequency trajectories from ancient dna indicates that selection on a chicken gene coincided with changes in medieval husbandry practices. *Molecular biology and evolution*, 34(8):1981–1990, 2017.

Lu Z, Chen TC, Zhang A, Persons KS, Kohn N, Berkowitz R, Martinello S, and Holick MF. An evaluation of the vitamin d3 content in fish: Is the vitamin d content adequate to satisfy the dietary requirement for vitamin d? *The Journal of steroid biochemistry and molecular biology*, 103(3-5):642–644, 2007.

Luca F, Perry GH, and Di Rienzo A. Evolutionary adaptations to dietary changes. *Annual review of nutrition*, 30: 291–314, 2010a.

Luca Francesca, Bubba Giuseppina, Basile Massimo, Brdicka Radim, Michalodimitrakis Emmanuel, Rickards Olga, Vershubsky Galina, Quintana-Murci Lluis, Kozlov Andrey I, and Novelletto Andrea. Multiple advantageous amino acid variants in the nat2 gene in human populations. *PloS one*, 3(9):e3136, 2008.

Luca Sabin Adrian, Suciu Cosmin Ioan, and Dumitrescu-Chioar Florian. Starčevo-criş culture in western part of romania–transylvania, banat, crişana, maramureş, oltenia and western muntenia: repository, distribution map, state of research and chronology, în neolithization of the carpathian basin: northernmost distribution of the starčevo/körös culture. *Neolithization of the Carparthian Basin: Northernmost Distribution of theE StarČevo/Körös Culture*, page 103, 2010b.

Magalon Hélene, Patin Etienne, Austerlitz Frédéric, Hegay Tatyana, Aldashev Almaz, Quintana-Murci Lluís, and Heyer Evelyne. Population genetic diversity of the nat2 gene supports a role of acetylation in human adaptation to farming in central asia. *European Journal of Human Genetics*, 16(2):243, 2008.

Makinen Tiina Maria. Different types of cold adaptation in humans. *Front Biosci (Schol Ed)*, 2(660):1047–1067, 2010.

Man Si Ming and Kanneganti Thirumala-Devi. Converging roles of caspases in inflammasome activation, cell death and innate immunity. *Nature Reviews Immunology*, 16(1):7, 2016.

Marciniak Stephanie and Perry George H. Harnessing ancient genomes to study the history of human adaptation. *Nature Reviews Genetics*, 18(11):659, 2017.

Mathieson Iain, Lazaridis Iosif, Rohland Nadin, Mallick Swapan, Patterson Nick, Roodenberg Songül Alpaslan, Harney Eadaoin, Stewardson Kristin, Fernandes Daniel, Novak Mario, and others . Genome-wide patterns of selection in 230 ancient eurasians. *Nature*, 528(7583):499, 2015.

Mathieson Iain, Alpaslan-Roodenberg Songül, Posth Cosimo, Szécsényi-Nagy Anna, Rohland Nadin, Mallick Swapan, Olalde Iñigo, Broomandkhoshbacht Nasreen, Candilio Francesca, Cheronet Olivia, and others . The genomic history of southeastern europe. *Nature*, 555(7695):197, 2018.

Mathieson Sara and Mathieson Iain. Fads1 and the timing of human adaptation to agriculture. *Molecular biology and evolution*, 35(12):2957–2970, 2018.

Mattison Kirsten, Shukla Anu, Cook Angela, Pollari Frank, Friendship Robert, Kelton David, Bidawid Sabah, and Farber Jeffrey M. Human noroviruses in swine and cattle. *Emerging infectious diseases*, 13(8):1184, 2007.

Maty Siobhan C, James Sherman A, and Kaplan George A. Life-course socioeconomic position and incidence of diabetes mellitus among blacks and whites: the alameda county study, 1965–1999. *American Journal of Public Health*, 100(1):137–145, 2010.

May Hila and Ruff Christopher. Physical burden and lower limb bone structure at the origin of agriculture in the levant. *American journal of physical anthropology*, 161(1):26–36, 2016.

McEvoy Brian, Beleza Sandra, and Shriver Mark D. The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model. *Human molecular genetics*, 15(suppl_2):R176–R181, 2006.

McKenna Aaron, Hanna Matthew, Banks Eric, Sivachenko Andrey, Cibulskis Kristian, Kernytsky Andrew, Garimella Kiran, Altshuler David, Gabriel Stacey, Daly Mark, and others . The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data. *Genome research*, 2010.

Meigs James B, Cupples L Adrienne, and Wilson Peter WF. Parental transmission of type 2 diabetes: the framingham offspring study. *Diabetes*, 49(12):2201–2207, 2000.

Meigs James B, Shrader Peter, Sullivan Lisa M, McAteer Jarred B, Fox Caroline S, Dupuis Josée, Manning Alisa K, Florez Jose C, Wilson Peter WF, D'Agostino Sr Ralph B, and others . Genotype score in addition to common risk factors for prediction of type 2 diabetes. *New England Journal of Medicine*, 359(21):2208–2219, 2008.

Mensink Ronald P and Katan Martijn B. Effect of dietary trans fatty acids on high-density and low-density lipoprotein cholesterol levels in healthy subjects. *New England Journal of Medicine*, 323(7):439–445, 1990.

Meyre David, Delplanque Jérôme, Chèvre Jean-Claude, Lecoeur Cécile, Lobbens Stéphane, Gallina Sophie, Durand Emmanuelle, Vatin Vincent, Degraeve Franck, Proença Christine, and others . Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in european populations. *Nature genetics*, 41(2): 157, 2009.

Milici Nicoleta. A short history of the metabolic syndrome definitions. In *Proc Rom Acad*, volume 1, pages 13–20, 2010.

Mittnik Alissa, Wang Chuan-Chao, Pfrengle Saskia, Daubaras Mantas, Zariņa Gunita, Hallgren Fredrik, Allmäe Raili, Khartanovich Valery, Moiseyev Vyacheslav, Tõrv Mari, and others . The genetic prehistory of the baltic sea region. *Nature communications*, 9(1):442, 2018.

Morales Eva, Bustamante Mariona, Gonzalez Juan Ramon, Guxens Monica, Torrent Maties, Mendez Michelle, Garcia-Esteban Raquel, Julvez Jordi, Forns Joan, Vrijheid Martine, and others . Genetic variants of the fads gene cluster and elovl gene family, colostrums lc-pufa levels, breastfeeding, and child cognition. *PloS one*, 6(2):e17181, 2011.

Morand Serge, Jittapalapong Sathaporn, and Kosoy Michael. Rodents as hosts of infectious diseases: Biological and ecological characteristics, 2015.

Morrison Shaun F. Central control of body temperature. *F1000Research*, 5, 2016.

Mostafa Wedad Z and Hegazy Rehab A. Vitamin d and the skin: Focus on a complex relationship: A review. *Journal of advanced research*, 6(6):793–804, 2015.

Mozaffarian Dariush, Micha Renata, and Wallace Sarah. Effects on coronary heart disease of increasing polyunsaturated fat in place of saturated fat: a systematic review and meta-analysis of randomized controlled trials. *PLoS medicine*, 7(3):e1000252, 2010.

Nakajima Toshiaki, Wooding Stephen, Sakagami Takuro, Emi Mitsuru, Tokunaga Katsushi, Tamiya Gen, Ishigami Tomoaki, Umemura Satoshi, Munkhbat Batmunkh, Jin Feng, and others . Natural selection and population history in the human angiotensinogen gene (agt): 736 complete agt sequences in chromosomes from around the world. *The American Journal of Human Genetics*, 74(5):898–916, 2004.

Nakayama Kazuhiro, Fukamachi Shoji, Kimura Hiroshi, Koda Yoshiro, Soemantri Augustinus, and Ishida Takafumi. Distinctive distribution of aim1 polymorphism among major human populations with different skin color. *Journal of human genetics*, 47(2):92, 2002.

Nebert Daniel W and Dieter Matthew Z. The evolution of drug metabolism. *Pharmacology*, 61(3):124–135, 2000.

Neel James V. Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? *American journal of human genetics*, 14(4):353, 1962.

Nei Masatoshi. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, 70(12):3321–3323, 1973.

Nessvi Sofia, Johansson Lisa, Jopson Jan, Stewart Alistair, Reeder Anthony, McKenzie Richard, and Scragg Robert K. Association of 25-hydroxyvitamin d3 levels in adult new zealanders with ethnicity, skin color and self-reported skin sensitivity to sun exposure. *Photochemistry and photobiology*, 87(5):1173–1178, 2011.

Nishimura Takayuki, Katsumura Takafumi, Motoi Midori, Oota Hiroki, and Watanuki Shigeki. Experimental evidence reveals the ucp1 genotype changes the oxygen consumption attributed to non-shivering thermogenesis in humans. *Scientific reports*, 7(1):5570, 2017.

Norton Heather L, Kittles Rick A, Parra Esteban, McKeigue Paul, Mao Xianyun, Cheng Keith, Canfield Victor A, Bradley Daniel G, McEvoy Brian, and Shriver Mark D. Genetic evidence for the convergent evolution of light skin in europeans and east asians. *Molecular biology and evolution*, 24(3):710–722, 2006.

Novembre John and Han Eunjung. Human population structure and the adaptive response to pathogen-induced selection pressures. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1590):878–886, 2012.

Novembre John, Johnson Toby, Bryc Katarzyna, Kutalik Zoltán, Boyko Adam R, Auton Adam, Indap Amit, King Karen S, Bergmann Sven, Nelson Matthew R, and others . Genes mirror geography within europe. *Nature*, 456 (7218):98, 2008.

O'Connor Michelle Y, Thoreson Caroline K, Ramsey Natalie LM, Ricks Madia, and Sumner Anne E. The uncertain significance of low vitamin d levels in african descent populations: a review of the bone and cardiometabolic literature. *Progress in cardiovascular diseases*, 56(3):261–269, 2013.

Oelze Vicky M, Siebert Angelina, Nicklisch Nicole, Meller Harald, Dresely Veit, and Alt Kurt W. Early neolithic diet and animal husbandry: stable isotope evidence from three linearbandkeramik (lbk) sites in central germany. *Journal of Archaeological Science*, 38(2):270–279, 2011.

Okada Hea, Kuhn C, Feillet H, and Bach J-F. The 'hygiene hypothesis' for autoimmune and allergic diseases: an update. *Clinical & Experimental Immunology*, 160(1):1–9, 2010.

Olalde Inigo, Allentoft Morten E, Sánchez-Quinto Federico, Santpere Gabriel, Chiang Charleston WK, DeGiorgio Michael, Prado-Martinez Javier, Rodríguez Juan Antonio, Rasmussen Simon, Quilez Javier, and others . Derived immune and ancestral pigmentation alleles in a 7,000-year-old mesolithic european. *Nature*, 507(7491):225, 2014.

Olalde Iñigo, Schroeder Hannes, Sandoval-Velasco Marcela, Vinner Lasse, Lobón Irene, Ramirez Oscar, Civit Sergi, García Borja Pablo, Salazar-García Domingo C, Talamo Sahra, and others . A common genetic origin for early farmers from mediterranean cardial and central european lbk cultures. *Molecular biology and evolution*, 32(12): 3132–3142, 2015.

Olalde Iñigo, Brace Selina, Allentoft Morten E, Armit Ian, Kristiansen Kristian, Booth Thomas, Rohland Nadin, Mallick Swapan, Szécsényi-Nagy Anna, Mittnik Alissa, and others . The beaker phenomenon and the genomic transformation of northwest europe. *Nature*, 555(7695):190, 2018.

Omrak Ayça, Günther Torsten, Valdiosera Cristina, Svensson Emma M, Malmström Helena, Kiesewetter Henrike, Aylward William, Storå Jan, Jakobsson Mattias, and Götherström Anders. Genomic evidence establishes anatolia as the source of the european neolithic gene pool. *Current Biology*, 26(2):270–275, 2016.

Omran Abdel R. The epidemiologic transition: a theory of the epidemiology of population change. *The Milbank Quarterly*, 83(4):731–757, 2005.

Orschiedt Joerg and Haidle Miriam Noel. The lbk enclosure at herxheim: theatre of war or ritual centre? references from osteoarchaeological investigations. *Journal of Conflict Archaeology*, 2(1):153–167, 2006.

Orschiedt Jörg. Ergebnisse einer neuen untersuchung der spätmesolithischen kopfbestattungen aus süddeutschland. *Aktuelle Forschungen zum Mesolithikum/Current Mesolithic Research. Urgeschichtliche Materialhefte*, 12:147–160, 1998.

Orschiedt Jörg, Gehlen Birgit, SchöN Werner, and Gröning Flora. The neolithic and mesolithic cave site "blätterhöhle" in westphalia (d).

Orton David. Herding, settlement, and chronology in the balkan neolithic. *European Journal of Archaeology*, 15(1): 5–40, 2012.

Ottoni Claudio, Girdland Flink Linus, Evin Allowen, Geörg Christina, De Cupere Bea, Van Neer Wim, Bartosiewicz László, Linderholm Anna, Barnett Ross, Peters Joris, and others . Pig domestication and human-mediated dispersal in western eurasia revealed through ancient dna and geometric morphometrics. *Molecular biology and evolution*, 30 (4):824–832, 2012.

Özdoğan Mehmet. Archaeological evidence on the westward expansion of farming communities from eastern anatolia to the aegean and the balkans. *Current Anthropology*, 52(S4):S415–S430, 2011a.

Özdoğan Mehmet. Archaeological evidence on the westward expansion of farming communities from eastern anatolia to the aegean and the balkans. *Current Anthropology*, 52(S4):S415–S430, 2011b.

Ozinsky Adrian, Underhill David M, Fontenot Jason D, Hajjar Adeline M, Smith Kelly D, Wilson Christopher B, Schroeder Lea, and Aderem Alan. The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between toll-like receptors. *Proceedings of the National Academy of Sciences*, 97(25): 13766–13771, 2000.

Paré Guillaume, Chasman Daniel I, Kellogg Mark, Zee Robert YL, Rifai Nader, Badola Sunita, Miletich Joseph P, and Ridker Paul M. Novel association of abo histo-blood group antigen with soluble icam-1: results of a genome-wide association study of 6,578 women. *PLoS genetics*, 4(7):e1000118, 2008.

Parmar AS, Alakulppi N, Paavola-Sakki P, Kurppa K, Halme L, Färkkilä M, Turunen U, Lappalainen M, Kontula K, Kaukinen K, and others . Association study of fut2 (rs601338) with celiac disease and inflammatory bowel disease in the finnish population. *Tissue antigens*, 80(6):488–493, 2012.

Patin Etienne, Barreiro Luis B, Sabeti Pardis C, Austerlitz Frédéric, Luca Francesca, Sajantila Antti, Behar Doron M, Semino Ornella, Sakuntabhai Anavaj, Guiso Nicole, and others . Deciphering the ancient and complex evolutionary history of human arylamine n-acetyltransferase genes. *The American Journal of Human Genetics*, 78(3):423–436, 2006.

Pedrosa Susana, Uzun Metehan, Arranz Juan-José, Gutiérrez-Gil Beatriz, San Primitivo Fermín, and Bayón Yolanda. Evidence of three maternal lineages in near eastern sheep supporting multiple domestication events. *Proceedings of the Royal Society of London B: Biological Sciences*, 272(1577):2211–2217, 2005.

Pickrell Joseph K, Coop Graham, Novembre John, Kudaravalli Sridhar, Li Jun Z, Absher Devin, Srinivasan Balaji S, Barsh Gregory S, Myers Richard M, Feldman Marcus W, and others . Signals of recent positive selection in a worldwide sample of human populations. *Genome research*, 2009.

Pierce Brandon L, Austin Melissa A, and Ahsan Habibul. Association study of type 2 diabetes genetic susceptibility variants and risk of pancreatic cancer: an analysis of panscan-i data. *Cancer Causes & Control*, 22(6):877–883, 2011.

Pimenoff Ville N, Laval Guillaume, Comas David, Palo Jukka U, Gut Ivo, Cann Howard, Excoffier Laurent, and Sajantila Antti. Similarity in recombination rate and linkage disequilibrium at cyp2c and cyp2d cytochrome p450 gene regions among europeans indicates signs of selection and no advantage of using tagsnps in population isolates. *Pharmacogenetics and genomics*, 22(12):846–857, 2012.

Plantinga Theo S, Alonso Santos, Izagirre Neskuts, Hervella Montserrat, Fregel Rosa, Van Der Meer Jos WM, Netea Mihai G, and De La Rúa Concepcion. Low prevalence of lactase persistence in neolithic south-west europe. *European journal of human genetics*, 20(7):778, 2012.

Podgorná Eliška, Diallo Issa, Vangenot Christelle, Sanchez-Mazas Alicia, Sabbagh Audrey, Černý Viktor, and Poloni Estella S. Variation in nat2 acetylation phenotypes is associated with differences in food-producing subsistence modes and ecoregions in africa. *BMC evolutionary biology*, 15(1):263, 2015.

Prentice Andrew M. The emerging epidemic of obesity in developing countries. *International journal of epidemiology*, 35(1):93–99, 2005.

Price T Douglas, Bentley R Alexander, Lüning Jens, Gronenborn Detlef, and Wahl Joachim. Prehistoric human migration in the linearbandkeramik of central europe. *Antiquity*, 75(289):593–603, 2001.

Pritchard Jonathan K, Pickrell Joseph K, and Coop Graham. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current biology*, 20(4):R208–R215, 2010.

Prugnolle Franck, Manica Andrea, Charpentier Marie, Guégan Jean François, Guernier Vanina, and Balloux François. Pathogen-driven selection and worldwide hla class i diversity. *Current biology*, 15(11):1022–1027, 2005.

Purcell Shaun, Neale Benjamin, Todd-Brown Kathe, Thomas Lori, Ferreira Manuel AR, Bender David, Maller Julian, Sklar Pamela, De Bakker Paul IW, Daly Mark J, and others . Plink: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3):559–575, 2007.

Rabhi Nabil, Hannou Sarah Anissa, Gromada Xavier, Salas Elisabet, Yao Xi, Oger Frédérik, Carney Charlène, Lopez-Mejia Isabel C, Durand Emmanuelle, Rabearivelo Iandry, and others . Cdkn2a deficiency promotes adipose tissue browning. *Molecular metabolism*, 8:65–76, 2018.

Raj Towfique, Kuchroo Manik, Replogle Joseph M, Raychaudhuri Soumya, Stranger Barbara E, and De Jager Philip L. Common risk alleles for inflammatory diseases are targets of recent positive selection. *The American Journal of Human Genetics*, 92(4):517–529, 2013.

Ramalingam Latha, Menikdiwela Kalhara, LeMieux Monique, Dufour Jannette M, Kaur Gurvinder, Kalupahana Nishan, and Moustaid-Moussa Naima. The renin angiotensin system, oxidative stress and mitochondrial function in obesity and insulin resistance. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1863(5):1106–1114, 2017.

Ramanan Deepshika, Bowcutt Rowann, Lee Soo Ching, San Tang Mei, Kurtz Zachary D, Ding Yi, Honda Kenya, Gause William C, Blaser Martin J, Bonneau Richard A, and others . Helminth infection promotes colonization resistance via type 2 immunity. *Science*, 352(6285):608–612, 2016.

Ramos Adauto V, Bastos-Rodrigues Luciana, Resende Bruna A, Friedman Eitan, Campanha-Versiani Luciana, Miranda Debora M, Sarquis Marta, and De Marco Luiz. The contribution of fto and ucp-1 snps to extreme obesity, diabetes and cardiovascular risk in brazilian individuals. *BMC medical genetics*, 13(1):101, 2012.

Rankinen Tuomo, Rice Treva, Teran-Garcia Margarita, Rao Dabeeru C, and Bouchard Claude. Fto genotype is associated with exercise training–induced changes in body composition. *Obesity*, 18(2):322–326, 2010.

Redding Richard W. A general explanation of subsistence change: From hunting and gathering to food production. *Journal of Anthropological Archaeology*, 7(1):56–97, 1988.

Rees Jonathan L and Harding Rosalind M. Understanding the evolution of human pigmentation: recent contributions from population genetics. *Journal of Investigative Dermatology*, 132(3):846–853, 2012.

Reich David, Green Richard E, Kircher Martin, Krause Johannes, Patterson Nick, Durand Eric Y, Viola Bence, Briggs Adrian W, Stenzel Udo, Johnson Philip LF, and others . Genetic history of an archaic hominin group from denisova cave in siberia. *Nature*, 468(7327):1053, 2010.

Revedin Anna, Aranguren Biancamaria, Becattini Roberto, Longo Laura, Marconi Emanuele, Lippi Marta Mariotti, Skakun Natalia, Sinitsyn Andrey, Spiridonova Elena, and Svoboda Jiří. Thirty thousand-year-old evidence of plant food processing. *Proceedings of the National Academy of Sciences*, 107(44):18815–18819, 2010.

Richards Michael P and Hedges Robert EM. Stable isotope evidence for similarities in the types of marine foods used by late mesolithic humans at sites along the atlantic coast of europe. *Journal of Archaeological Science*, 26(6): 717–722, 1999.

Richards Michael P, Schulting Rick J, and Hedges Robert EM. Archaeology: sharp shift in diet at onset of neolithic. *Nature*, 425(6956):366, 2003.

Richards Michael P, Jacobi RM, Cook J, Pettitt Paul B, and Stringer Chris B. Isotope evidence for the intensive use of marine foods by late upper palaeolithic humans. *Journal of Human Evolution*, 49(3):390–394, 2005.

Richert Laura, Chevalley Thierry, Manen Danielle, Bonjour Jean-Philippe, Rizzoli René, and Ferrari Serge. Bone mass in prepubertal boys is associated with a gln223arg amino acid substitution in the leptin receptor. *The Journal of Clinical Endocrinology & Metabolism*, 92(11):4380–4386, 2007.

Ridush B. Bear caves in ukraine. *Ukraine. Slovenskỳ Kras. Acta Carstologica Slovaca*, 47(Suppl 1):67–84, 2009.

Rigaud Solange, d'Errico Francesco, and Vanhaeren Marian. Ornaments reveal resistance of north european cultures to the spread of farming. *PloS one*, 10(4):e0121166, 2015.

Robins Ashley H. The evolution of light skin color: role of vitamin d disputed. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, 139(4):447–450, 2009.

Rohde Kerstin, Klös Matthias, Hopp Lydia, Liu Xuanshi, Keller Maria, Stumvoll Michael, Dietrich Arne, Schön Michael R, Gärtner Daniel, Lohmann Tobias, and others . Irs1 dna promoter methylation and expression in human adipose tissue are related to fat distribution and metabolic traits. *Scientific reports*, 7(1):12369, 2017.

Roizen Jeffrey D, Li Dong, O'lear Lauren, Javaid Muhammad K, Shaw Nicholas J, Ebeling Peter R, Nguyen Hanh H, Rodda Christine P, Thummel Kenneth E, Thacher Tom D, and others . Cyp3a4 mutation causes vitamin d–dependent rickets type 3. *The Journal of clinical investigation*, 128(5):1913, 2018.

Ronkainen Justiina, Mondini Eleonora, Cinti Francesca, Cinti Saverio, Sebért Sylvain, Savolainen Markku, and Salonurmi Tuire. Fto-deficiency affects the gene and microrna expression involved in brown adipogenesis and browning of white adipose tissue in mice. *International journal of molecular sciences*, 17(11):1851, 2016.

Rook Graham AW. Hygiene hypothesis and autoimmune diseases. *Clinical reviews in allergy & immunology*, 42(1): 5–15, 2012.

Rook Graham AW, Martinelli Roberta, and Brunet Laura Rosa. Innate immune responses to mycobacteria and the downregulation of atopic responses. *Current opinion in allergy and clinical immunology*, 3(5):337–342, 2003.

Roth Jesse. Evolutionary speculation about tuberculosis and the metabolic and inflammatory processes of obesity. *Jama*, 301(24):2586–2588, 2009.

Sabbagh Audrey, Langaney André, Darlu Pierre, Gérard Nathalie, Krishnamoorthy Rajagopal, and Poloni Estella S. Worldwide distribution of nat2 diversity: implications for nat2 evolutionary history. *BMC genetics*, 9(1):21, 2008.

Sabbagh Audrey, Darlu Pierre, and Vidaud Michel. Evaluating nat2pred for inferring the individual acetylation status from unphased genotype data. *BMC medical genetics*, 10(1):148, 2009.

Sabbagh Audrey, Darlu Pierre, Crouau-Roy Brigitte, and Poloni Estella S. Arylamine n-acetyltransferase 2 (nat2) genetic diversity and traditional subsistence: a worldwide population survey. *PloS one*, 6(4):e18507, 2011.

Sabeti Pardis C, Reich David E, Higgins John M, Levine Haninah ZP, Richter Daniel J, Schaffner Stephen F, Gabriel Stacey B, Platko Jill V, Patterson Nick J, McDonald Gavin J, and others . Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419(6909):832, 2002.

Sabeti Pardis C, Schaffner Stephen F, Fry Ben, Lohmueller Jason, Varilly Patrick, Shamovsky Oleg, Palma Alejandro, Mikkelsen TS, Altshuler D, and Lander ES. Positive natural selection in the human lineage. *science*, 312(5780): 1614–1620, 2006.

Saleh Maya, Vaillancourt John P, Graham Rona K, Huyck Matthew, Srinivasula Srinivasa M, Alnemri Emad S, Steinberg Martin H, Nolan Vikki, Baldwin Clinton T, Hotchkiss Richard S, and others . Differential modulation of endotoxin responsiveness by human caspase-12 polymorphisms. *Nature*, 429(6987):75, 2004.

Scherer Dominique and Kumar Rajiv. Genetics of pigmentation in skin cancer–a review. *Mutation Research/Reviews in Mutation Research*, 705(2):141–153, 2010.

Scheu Amelie, Powell Adam, Bollongino Ruth, Vigne Jean-Denis, Tresset Anne, Çakırlar Canan, Benecke Norbert, and Burger Joachim. The genetic prehistory of domesticated cattle from their origin to the spread across europe. *BMC genetics*, 16(1):54, 2015.

Schlebusch Carina M, Malmström Helena, Günther Torsten, Sjödin Per, Coutinho Alexandra, Edlund Hanna, Munters Arielle R, Vicente Mário, Steyn Maryna, Soodyall Himla, and others . Southern african ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science*, 358(6363):652–655, 2017.

Schrider Daniel R and Kern Andrew D. Soft sweeps are the dominant mode of adaptation in the human genome. *Molecular biology and evolution*, 34(8):1863–1877, 2017.

Schurz Haiko, Daya Michelle, Möller Marlo, Hoal Eileen G, and Salie Muneeb. Tlr1, 2, 4, 6 and 9 variants associated with tuberculosis susceptibility: a systematic review and meta-analysis. *PloS one*, 10(10):e0139711, 2015.

Scott James C. *Against the grain: a deep history of the earliest states*. Yale University Press, 2017.

Seguin-Orlando Andaine, Korneliussen Thorfinn S, Sikora Martin, Malaspinas Anna-Sapfo, Manica Andrea, Moltke Ida, Albrechtsen Anders, Ko Amy, Margaryan Ashot, Moiseyev Vyacheslav, and others . Genomic structure in europeans dating back at least 36,200 years. *Science*, 346(6213):1113–1118, 2014.

Ségurel Laure, Austerlitz Frederic, Toupance Bruno, Gautier Mathieu, Kelley Joanna L, Pasquet Patrick, Lonjou Christine, Georges Myriam, Voisin Sarah, Cruaud Corinne, and others . Positive selection of protective variants for type 2 diabetes from the neolithic onward: a case study in central asia. *European Journal of Human Genetics*, 21 (10):1146, 2013.

Sell Christian. Addressing challenges of ancient dna sequence data obtained with next generation methods. *Dissertation*, 1(1):0–150, 2017.

Sellayah Dyan, Cagampang Felino R, and Cox Roger D. On the evolutionary origins of obesity: a new hypothesis. *Endocrinology*, 155(5):1573–1588, 2014.

Shennan Stephen. Evolutionary demography and the population history of the european early neolithic. *Human Biology*, 81(3):339–355, 2009.

Shennan Stephen, Downey Sean S, Timpson Adrian, Edinborough Kevan, Colledge Sue, Kerig Tim, Manning Katie, and Thomas Mark G. Regional population collapse followed initial agriculture booms in mid-holocene europe. *Nature Communications*, 4:2486, 2013.

Shuldiner Alan R, O'Connell Jeffrey R, Bliden Kevin P, Gandhi Amish, Ryan Kathleen, Horenstein Richard B, Damcott Coleen M, Pakyz Ruth, Tantry Udaya S, Gibson Quince, and others . Association of cytochrome p450 2c19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. *Jama*, 302(8):849–857, 2009.

Sistonen Johanna, Fuselli Silvia, Palo Jukka U, Chauhan Neelam, Padh Harish, and Sajantila Antti. Pharmacogenetic variation at cyp2c9, cyp2c19, and cyp2d6 at global and microgeographic scales. *Pharmacogenetics and genomics*, 19(2):170–179, 2009.

Sjögren Karl-Göran, Price T Douglas, and Kristiansen Kristian. Diet and mobility in the corded ware of central europe. *PloS one*, 11(5):e0155083, 2016.

Skoglund Pontus, Malmström Helena, Raghavan Maanasa, Storå Jan, Hall Per, Willerslev Eske, Gilbert M Thomas P, Götherström Anders, and Jakobsson Mattias. Origins and genetic legacy of neolithic farmers and hunter-gatherers in europe. *Science*, 336(6080):466–469, 2012.

Skoglund Pontus, Storå Jan, Götherström Anders, and Jakobsson Mattias. Accurate sex identification of ancient human remains using dna shotgun sequencing. *Journal of Archaeological Science*, 40(12):4477–4482, 2013.

Slatkin Montgomery and Racimo Fernando. Ancient dna and human history. *Proceedings of the National Academy of Sciences*, page 201524306, 2016.

Smith George Davey, Lawlor Debbie A, Timpson Nic J, Baban Jamil, Kiessling Matt, Day Ian NM, and Ebrahim Shah. Lactase persistence-related genetic variant: population substructure and health outcomes. *European journal of human genetics*, 17(3):357, 2009.

Smyth Deborah J, Cooper Jason D, Howson Joanna MM, Clarke Pamela, Downes Kate, Mistry Trupti, Stevens Helen, Walker Neil M, and Todd John A. Fut2 nonsecretor status links type 1 diabetes susceptibility and resistance to infection. *Diabetes*, 60(11):3081–3084, 2011.

Soejima Mikiko and Koda Yoshiro. Population differences of two coding snps in pigmentation-related genes slc24a5 and slc45a2. *International Journal of Legal Medicine*, 121(1):36–39, 2007.

Sotgiu Stefano, Angius Andrea, Embry Ashton, Rosati Giulio, and Musumeci Salvatore. Hygiene hypothesis: innate immunity, malaria and multiple sclerosis. *Medical hypotheses*, 70(4):819–825, 2008.

Speakman John R. Thrifty genes for obesity and the metabolic syndrome–time to call off the search? *Diabetes and Vascular Disease Research*, 3(1):7–11, 2006.

Speakman John R. Thrifty genes for obesity, an attractive but flawed idea, and an alternative perspective: the 'drifty gene' hypothesis. *International journal of obesity*, 32(11):1611, 2008.

Speakman John R. The evolution of body fatness: trading off disease and predation risk. *Journal of Experimental Biology*, 221(Suppl 1):jeb167254, 2018.

Spichenok Olga, Budimlija Zoran M, Mitchell Adele A, Jenny Andreas, Kovacevic Lejla, Marjanovic Damir, Caragine Theresa, Prinz Mechthild, and Wurmbach Elisa. Prediction of eye and skin color in diverse populations using seven snps. *Forensic Science International: Genetics*, 5(5):472–478, 2011.

Stathopoulos Ioannis P, Trovas George, Lampropoulou-Adamidou Kalliopi, Koromila Theodora, Kollia Panagoula, Papaioannou Nikolaos A, and Lyritis George. Severe osteoporosis and mutation in notch2 gene in a woman with hajdu–cheney syndrome. *Bone*, 52(1):366–371, 2013.

Stone Anne C, Wilbur Alicia K, Buikstra Jane E, and Roberts Charlotte A. Tuberculosis and leprosy in perspective. *American journal of physical anthropology*, 140(S49):66–94, 2009.

Strachan David P. Hay fever, hygiene, and household size. *BMJ: British Medical Journal*, 299(6710):1259, 1989.

Street Martin, Baales Michael, Cziesla Erwin, Hartz Sönke, Heinen Martin, Jöris Olaf, Koch Ingrid, Pasda Clemens, Terberger Thomas, and Vollbrecht Jürgen. Final paleolithic and mesolithic research in reunified germany. *Journal of World Prehistory*, 15(4):365–453, 2001.

Sturm Richard A. A golden age of human pigmentation genetics. *Trends in genetics*, 22(9):464–468, 2006.

Sturm Richard A and Duffy David L. Human pigmentation genes under environmental selection. *Genome biology*, 13 (9):248, 2012.

Sturm Richard A and Larsson Mats. Genetics of human iris colour and patterns. *Pigment cell & melanoma research*, 22(5):544–562, 2009.

Sulem Patrick, Gudbjartsson Daniel F, Stacey Simon N, Helgason Agnar, Rafnar Thorunn, Magnusson Kristinn P, Manolescu Andrei, Karason Ari, Palsson Arnar, Thorleifsson Gudmar, and others . Genetic determinants of hair, eye and skin pigmentation in europeans. *Nature genetics*, 39(12):1443, 2007.

Summers Robert W, Elliott DE, Urban JF, Thompson Robin, and Weinstock JV. Trichuris suis therapy in crohn's disease. *Gut*, 54(1):87–90, 2005.

Svizzero S. Persistent controversies about the neolithic revolution. *Journal of Historical Archaeology & Anthropological Sciences*, 1(2):00013, 2017.

Tajima Fumio. Statistical method for testing the neutral mutation hypothesis by dna polymorphism. *Genetics*, 123 (3):585–595, 1989.

Tallavaara Miikka, Luoto Miska, Korhonen Natalia, Järvinen Heikki, and Seppä Heikki. Human population dynamics in europe over the last glacial maximum. *Proceedings of the National Academy of Sciences*, 112(27):8232–8237, 2015.

Terberger Thomas, Kotula Andreas, Lorenz Sebastian, Schult Manuela, Burger Joachim, and Jungklaus Bettina. Standing upright to all eternity–the mesolithic burial site at groß fredenwalde, brandenburg (ne germany). *Quartär*, 62:133, 2015.

Teschler-Nicola M, Gerold F, Bujatti-Narbeshuber M, Prohaska T, Latkoczy Ch, Stingederº G, and Watkins M. Evidence of genocide 7000 bp–neolithic paradigm and geo-climatic. *Coll. Antropol*, 23(2):437–450, 1999.

Thompson Emma E, Kuttab-Boulos Hala, Witonsky David, Yang Limei, Roe Bruce A, and Di Rienzo Anna. Cyp3a variation and the evolution of salt-sensitivity variants. *The American Journal of Human Genetics*, 75(6):1059–1069, 2004.

Thorn Andrea, Steinfeld Robert, Ziegenbein Marc, Grapp Marcel, Hsiao He-Hsuan, Urlaub Henning, Sheldrick George M, Gärtner Jutta, and Krätzner Ralph. Structure and activity of the only human rnase t2. *Nucleic acids research*, 40(17):8733–8742, 2012.

Torres Sofía, García-Palmero Irene, Marín-Vicente Consuelo, Bartolome Ruben A, Calviño Eva, Fernández-Aceñero María Jesús, and Casal J Ignacio. Proteomic characterization of transcription and splicing factors associated with a metastatic phenotype in colorectal cancer. *Journal of proteome research*, 17(1):252–264, 2017.

Troelsen Jesper T. Adult-type hypolactasia and regulation of lactase expression. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1723(1-3):19–32, 2005.

Turck R, Kober B, Kontny J, Haack F, and Zeeb-Lanz A. "widely travelled people" at herxheim? sr isotopes as indicators of mobility. *Populations dynamics in prehistory and early history: new approaches using stable isotopes and genetics*, pages 149–163, 2012.

Vallee Alexandre, Vallée Jean-Noël, Guillevin Remy, and Lecarpentier Yves. Interactions between the canonical wnt/beta-catenin pathway and ppar gamma on neuroinflammation, demyelination, and remyelination in multiple sclerosis. *Cellular and molecular neurobiology*, 38(4):783–795, 2018.

Vancampfort Davy, Sweers K, Probst Michel, Maurissen K, Knapen Jan, Minguet P, and De Hert Marc. Association of the metabolic syndrome with physical activity performance in patients with schizophrenia. *Diabetes & metabolism*, 37(4):318–323, 2011.

Vanmontfort Bart. Forager–farmer connections in an 'unoccupied'land: First contact on the western edge of lbk territory. *Journal of Anthropological Archaeology*, 27(2):149–160, 2008.

Vassy JL, Durant NH, Kabagambe EK, Carnethon Mercedes R, Rasmussen-Torvik Laura Jarmila, Fornage M, Lewis CE, Siscovick DS, and Meigs JB. A genotype risk score predicts type 2 diabetes from young adulthood: the cardia study. *Diabetologia*, 55(10):2604–2612, 2012.

Veeramah Krishna R., Rott Andreas, Groß Melanie, Dorp van Lucy, López Saioa, Kirsanow Karola, Sell Christian, Blöcher Jens, Wegmann Daniel, Link Vivian, Hofmanová Zuzana, Peters Joris, Trautmann Bernd, Gairhos Anja, Haberstroh Jochen, Päffgen Bernd, Hellenthal Garrett, Haas-Gebhard Brigitte, Harbeck Michaela, and Burger Joachim. Population genomic analysis of elongated skulls reveals extensive female-biased immigration in early medieval bavaria. *Proceedings of the National Academy of Sciences*, 115(13):3494–3499, 2018. doi: 10.1073/pnas.1719880115.

Villalba-Mouco Vanessa, Loosdrecht van de Marieke S, Posth Cosimo, Mora Rafael, Martínez-Moreno Jorge, Rojo-Guerra Manuel, Salazar-García Domingo C, Royo-Guillén José I, Kunst Michael, Rougier Hélene, and others . Survival of late pleistocene hunter-gatherer ancestry in the iberian peninsula. *Current Biology*, 2019.

Voight Benjamin F, Kudaravalli Sridhar, Wen Xiaoquan, and Pritchard Jonathan K. A map of recent positive selection in the human genome. *PLoS biology*, 4(3):e72, 2006.

Voight Benjamin F, Scott Laura J, Steinthorsdottir Valgerdur, Morris Andrew P, Dina Christian, Welch Ryan P, Zeggini Eleftheria, Huth Cornelia, Aulchenko Yurii S, Thorleifsson Gudmar, and others . Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature genetics*, 42(7):579, 2010.

Wadelius Mia, Chen Leslie Y, Eriksson Niclas, Bumpstead Suzannah, Ghori Jilur, Wadelius Claes, Bentley David, McGinnis Ralph, and Deloukas Panos. Association of warfarin dose with genes involved in its action and metabolism. *Human genetics*, 121(1):23–34, 2007.

Wagner Carol L, Greer Frank R, and others . Prevention of rickets and vitamin d deficiency in infants, children, and adolescents. *Pediatrics*, 122(5):1142–1152, 2008.

Walsh Susan, Chaitanya Lakshmi, Clarisse Lindy, Wirken Laura, Draus-Barini Jolanta, Kovatsi Leda, Maeda Hitoshi, Ishikawa Takaki, Sijen Titia, Knijff de Peter, and others . Developmental validation of the hirisplex system: Dna-based eye and hair colour prediction for forensic and anthropological usage. *Forensic Science International: Genetics*, 9:150–161, 2014.

Walsh Susan, Chaitanya Lakshmi, Breslin Krystal, Muralidharan Charanya, Bronikowska Agnieszka, Pospiech Ewelina, Koller Julia, Kovatsi Leda, Wollstein Andreas, Branicki Wojciech, and others . Global skin colour prediction from dna. *Human genetics*, 136(7):847–863, 2017.

Wang Chaolong, Zhan Xiaowei, Bragg-Gresham Jennifer, Kang Hyun Min, Stambolian Dwight, Chew Emily Y, Branham Kari E, Heckenlively John, Study The FUSION, Fulton Robert, and others . Ancestry estimation and control of population stratification for sequence-based association studies. *Nature genetics*, 46(4):409, 2014.

Wang Guanlin and Speakman John R. Analysis of positive selection at single nucleotide polymorphisms associated with body mass index does not support the "thrifty gene" hypothesis. *Cell metabolism*, 24(4):531–541, 2016.

Wang Haidong, Dwyer-Lindgren Laura, Lofgren Katherine T, Rajaratnam Julie Knoll, Marcus Jacob R, Levin-Rector Alison, Levitz Carly E, Lopez Alan D, and Murray Christopher JL. Age-specific and sex-specific mortality in 187 countries, 1970–2010: a systematic analysis for the global burden of disease study 2010. *The Lancet*, 380(9859): 2071–2094, 2012.

Wang Haijian, Ding Keyue, Zhang Yang, Jin Li, Kullo Iftikhar J, and He Fuchu. Comparative and evolutionary pharmacogenetics of abcb1: complex signatures of positive selection on coding and regulatory regions. *Pharmacogenetics and genomics*, 17(8):667–678, 2007.

Wang Thomas J, Zhang Feng, Richards J Brent, Kestenbaum Bryan, Van Meurs Joyce B, Berry Diane, Kiel Douglas P, Streeten Elizabeth A, Ohlsson Claes, Koller Daniel L, and others . Common genetic determinants of vitamin d insufficiency: a genome-wide association study. *The Lancet*, 376(9736):180–188, 2010.

Wang Xiaoxia, Grus Wendy E, and Zhang Jianzhi. Gene losses during human origins. *PLoS biology*, 4(3):e52, 2006.

Wang Zhican, Schuetz Erin G, Xu Yang, and Thummel Kenneth E. Interplay between vitamin d and the drug metabolizing enzyme cyp3a4. *The Journal of steroid biochemistry and molecular biology*, 136:54–58, 2013.

Weiss Ehud, Wetterstrom Wilma, Nadel Dani, and Bar-Yosef Ofer. The broad spectrum revisited: evidence from plant remains. *Proceedings of the National Academy of Sciences*, 101(26):9551–9555, 2004.

Weiss Ehud, Kislev Mordechai E, and Hartmann Anat. Autonomous cultivation before domestication. *Science*, 312 (5780):1608–1610, 2006.

Weissensteiner Hansi, Pacher Dominic, Kloss-Brandstätter Anita, Forer Lukas, Specht Günther, Bandelt Hans-Jürgen, Kronenberg Florian, Salas Antonio, and Schönherr Sebastian. Haplogrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic acids research*, 44(W1):W58–W63, 2016.

Wellington Jean Susorney. Ancient europe 8000 bc-ad 1000: Encyclopedia of the barbarian world. *Reference & User Services Quarterly*, 44(1):81–83, 2004.

Wells Jonathan CK. Ethnic variability in adiposity and cardiovascular risk: the variable disease selection hypothesis. *International journal of epidemiology*, 38(1):63–71, 2008.

Weninger Bernhard, Clare Lee, Gerritsen Fokke, Horejs Barbara, Krauß Raiko, Linstädter Jörg, Özbal Rana, Rohling Eelco J, and others . Neolithisation of the aegean and southeast europe during the 6600–6000 calbc period of rapid climate change. 2014.

Wilde Sandra, Timpson Adrian, Kirsanow Karola, Kaiser Elke, Kayser Manfred, Unterländer Martina, Hollfelder Nina, Potekhina Inna D, Schier Wolfram, Thomas Mark G, and others . Direct evidence for positive selection of skin, hair, and eye pigmentation in europeans during the last 5,000 years. *Proceedings of the National Academy of Sciences*, page 201316513, 2014.

Wong Sunny H, Gochhait Sailesh, Malhotra Dheeraj, Pettersson Fredrik H, Teo Yik Y, Khor Chiea C, Rautanen Anna, Chapman Stephen J, Mills Tara C, Srivastava Amit, and others . Leprosy and the adaptation of human toll-like receptor 1. *PLoS pathogens*, 6(7):e1000979, 2010.

Wood RE, Higham Thomas FG, Buzilhova A, Suvorov A, Heinemeier J, and Olsen Jesper. Freshwater radiocarbon reservoir effects at the burial ground of minino, northwest russia. *Radiocarbon*, 55(1):163–177, 2013.

Xue Yali, Daly Allan, Yngvadottir Bryndis, Liu Mengning, Coop Graham, Kim Yuseob, Sabeti Pardis, Chen Yuan, Stalker Jim, Huckle Elizabeth, and others . Spread of an inactive form of caspase-12 in humans is due to recent positive selection. *The American Journal of Human Genetics*, 78(4):659–670, 2006.

Yang Melinda A and Fu Qiaomei. Insights into modern human prehistory using ancient genomes. *Trends in Genetics*, 2018.

Yang Mengyu, Arai Atsushi, Udagawa Nobuyuki, Hiraga Toru, Lijuan Zhao, Ito Susumu, Komori Toshihisa, Moriishi Takeshi, Matsuo Koichi, Shimoda Kouji, and others . Osteogenic factor runx2 marks a subset of leptin receptor-positive cells that sit atop the bone marrow stromal cell hierarchy. *Scientific reports*, 7(1):4928, 2017.

Yang Min, Ren Yan, Lin Zhimin, Tang Chenchen, Jia Yanjun, Lai Yerui, Zhou Tingting, Wu Shaobo, Liu Hua, Yang Gangyi, and others . Krüppel-like factor 14 increases insulin sensitivity through activation of pi3k/akt signal pathway. *Cellular signalling*, 27(11):2201–2208, 2015.

Yang Zhaohui, Shi Hong, Ma Pengcheng, Zhao Shilei, Kong Qinghong, Bian Tianhao, Gong Chao, Zhao Qi, Liu Yuan, Qi Xuebin, and others . Darwinian positive selection on the pleiotropic effects of kitlg explain skin pigmentation and winter temperature adaptation in eurasians. *Molecular biology and evolution*, 35(9):2272–2283, 2018.

Yazdanbakhsh Maria, Kremsner Peter G, and Van Ree Ronald. Allergy, parasites, and the hygiene hypothesis. *Science*, 296(5567):490–494, 2002.

Ye Byong Duk and McGovern Dermot PB. Genetic variation in ibd: progress, clues to pathogenesis and possible clinical utility. *Expert review of clinical immunology*, 12(10):1091–1107, 2016.

Young J Hunter, Chang Yen-Pei C, Kim James Dae-Ok, Chretien Jean-Paul, Klag Michael J, Levine Michael A, Ruff Christopher B, Wang Nae-Yuh, and Chakravarti Aravinda. Differential susceptibility to hypertension is due to selection during the out-of-africa expansion. *PLoS genetics*, 1(6):e82, 2005.

Yuen AWC and Jablonski NG. Vitamin d: in the evolution of human skin colour. *Medical hypotheses*, 74(1):39–44, 2010.

Zeder Melinda A. The origins of agriculture in the near east. *Current Anthropology*, 52(S4):S221–S235, 2011.

Zeeb-Lanz Andrea, Arbogast Rose-Marie, Haack Fabian, Haidle Miriam N, Jeunesse Christian, fiedt Jörg, Schimmelpfennig Dirk, and Van Willingen S. The lbk settlement with pit enclosure at herxheim near landau (palatinate)–first results. *Creating communities. New advances in central european Neolithic research*, pages 202–220, 2008.

Zeggini Eleftheria, Scott Laura J, Saxena Richa, Voight Benjamin F, Marchini Jonathan L, Hu Tianle, Bakker de Paul IW, Abecasis Gonçalo R, Almgren Peter, Andersen Gitte, and others . Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature genetics*, 40(5):638, 2008.

Zhang Ge, Muglia Louis J, Chakraborty Ranajit, Akey Joshua M, and Williams Scott M. Signatures of natural selection on genetic variants affecting complex human traits. *Applied & translational genomics*, 2:78–94, 2013.

Zhernakova Alexandra, Elbers Clara C, Ferwerda Bart, Romanos Jihane, Trynka Gosia, Dubois Patrick C, De Kovel Carolien GF, Franke Lude, Oosting Marije, Barisani Donatella, and others . Evolutionary and functional analysis of celiac risk loci reveals sh2b3 as a protective factor against bacterial infection. *The American Journal of Human Genetics*, 86(6):970–977, 2010.

Zhong Ming, Zhang Yiwei, Lange Kenneth, and Fan Ruzong. A cross-population extended haplotype-based homozygosity score test to detect positive selection in genome-wide scans. *Statistics and its Interface*, 4(1):51, 2011.

Zhou Li-Ping, Luan Hong, Dong Xi-Hua, Jin Guo-Jiang, Man Dong-Liang, and Shang Hong. Genetic variants of cyp2d6 gene and cancer risk: a huge systematic review and meta-analysis. *Asian Pacific Journal of Cancer Prevention*, 13 (7):3165–3172, 2012.

Zuckerman Molly, Harper Kristin, Barrett Ronald, and Armelagos George. The evolution of disease: anthropological perspectives on epidemiologic transitions. *Global health action*, 7(1):23303, 2014.

Raw data and scripts are available upon request from **jbloech@uni-mainz.de**