**Population genomic analysis of elongated skulls reveals extensive female-biased immigration in Early Medieval Bavaria**

# SI Appendix

Krishna R. Veeramah[a], Andreas Rott[b,1], Melanie Groß[c,1], Lucy van Dorp[d], Saioa López[e], Karola Kirsanow[c], Christian Sell[c], Jens Blöcher[c], Daniel Wegmann[f,g], Vivian Link[f,g], Zuzana Hofmanová[f,g], Joris Peters[b,h], Bernd Trautmann[b], Anja Gairhos[i], Jochen Haberstroh[j], Bernd Päffgen[k], Garrett Hellenthal[d], Brigitte Haas-Gebhard[i], Michaela Harbeck[b,2, 3], and Joachim Burger[c,2,3]

[1,2] These authors contributed equally to this work

[3] Corresponding authors (Michaela.Harbeck@extern.lrz-muenchen.de; jburger@uni-mainz.de)

[a] Department of Ecology and Evolution, Stony Brook University, Stony Brook, New York, 11794-5245, USA.
[b] SNSB, State Collection for Anthropology and Palaeoanatomy, 80333 Munich, Germany
[c] Palaeogenetics Group, Institute of Organismic and Molecular Evolution (iOME), Johannes Gutenberg University Mainz, 55099 Mainz, Germany.
[d] UCL Genetics Institute, Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK.
[e] Cancer Institute, University College London, London WC1E 6DD, UK.
[f] Department of Biology, University of Fribourg, 1700 Fribourg, Switzerland.
[g] Swiss Institute of Bioinformatics, 1700 Fribourg, Switzerland.
[h] ArchaeoBioCenter and Institute for Palaeoanatomy, Domestication Research and the History of Veterinary Medicine, Ludwig-Maximilian University, 80539 Munich, Germany
[i] Bavarian State Archaeological Collection, 80538 Munich, Germany
[j] Bavarian State Department of Monuments and Sites, 80539 Munich, Germany
[k] Institute of Prehistoric and Protohistoric Archaeology, Ludwig-Maximilian University, 80799 Munich, Germany

**This PDF file includes:**

# SI Appendix

## SI Appendix

## SI1. Archaeological and Morphological information

### Early Medieval graves from Bavaria

*Brigitte Haas-Gebhard, Bernd Trautmann, Michaela Harbeck, Andreas Rott*

For the epoch of the late 5th/early 6th century, grave goods are the almost exclusive source material available for archaeological research due to the fact that hardly any contemporary settlements have so far been identified. During this time period the dead were placed on their backs with their arms alongside the body and with their heads generally positioned to the West. They were interred in burial groups, which are understood to represent the cemeteries of small social units. From the mid-6th century onwards these cemeteries then merged into large necropoles which can comprise more than 1000 inhumation burials. A large percentage of the dead were buried with a sex-specific assemblage; in the case of males this consisted of weapons, in the case of females it was jewelry. It is thought that burying the dead was accompanied by an elaborate laying-out ceremony and other burial rites, which demonstrated the social position of the dead and therefore also cemented that of their family (1). It was quite common to re-open the graves after inhumation and to remove the precious grave goods.

By using methods of combination statistics, the large quantity of archaeological finds available for this time period allows for a reliable reconstruction of the sequence of many groups of grave goods. Combined with absolute dateable material such as coins, it enables absolute dating to a time-frame of 30-40 years (2). Dates estimated this way are shown in Table S1. Dendrochronological data is not yet available for this time period in most parts of central Europe. Due to the shape of the calibration curve, the data collected from radiocarbon analyses is also only of limited suitability for exact dating (3; see Table S1).

In the scholarly tradition of the 19th century, the alleged ethnic origin of the wearer of a certain type of jewelry was determined on the basis of specific embellishments on the so-called bow-brooch, a highly ornamented garment fastener usually made of gilded silver and in function similar to a modern-day safety pin (e.g. 4). This approach is now highly criticized due to the fact that these brooches apparently were not personal property of the wearer and the distribution types of Early Medieval jewelry have still not been fully identified (5). At best, the archaeological-antiquarian analysis of a grave goods assemblage can tell us in which cultural context the organizers of the funeral - in most cases probably the closest relatives – placed the deceased at the time of their death. Nevertheless, “cultural labels” were assigned to the bow-brooches, that were part of the study in order to be able to recognize any evolving patterns (see Table S1).

All individuals from Bavarian sites (36 individuals in total) were re-examined in an anthropological survey including the determination of sex and age-at-death as explained in Grupe et al. (6), as well as paleopathological features with a special focus on re-evaluation of artificial skull deformation (Table S1; 6).

Before the project described here, a total of 30 skulls from Bavaria were reported as artificially modified, of which 21 were potentially available for DNA analysis. However, before the analysis of the genetic data started, all of the 36 examined skulls underwent an analysis for modification. For the detailed procedure see Trautmann et al. (7). To differentiate artificially deformed from non-deformed skulls, the criteria described in the studies of Clark et al. (8), Ginzburg & Žirov (9) in Molnár et al. (10) and O´Brien & Stanley (11) were used. These studies allow determining the level as well as the kind of deformation by means of skull measurements. In addition, a morphological examination was carried out to assess and classify the exterior shape of the skulls according to the descriptions in Cocilovo et al. (12). Skulls were classified as artificially deformed when at least two of the metrical studies or one of the metrical and the morphological examination revealed positive results (7). Skulls exhibiting a suspicious exterior shape and with no positive result from any of the metric analyses were classified as "intermediate". Undeformed skulls neither yielded positive results from metric analyses nor do they exhibit a suspicious exterior shape. From the 21 available crania previously classified as deformed, only nine could be confirmed as artificially modified (Table S1) while five skulls were categorized as "intermediate", and seven have now to be considered as not modified in any way (marked as "not deformed" in Table S1). In addition to individuals with potentially deformed skulls, contemporary individuals without skull deformation were also included in the morphological re-examination.

Radiocarbon dating was carried out for all individuals with deformed skulls and many of the individuals from the site of Altheim (see below). The results of these analyses (Table S1) support the archaeological dating of these women to around 500 AD. An example of the grave goods of women with artificially deformed skulls is shown in Fig. S1.

The 36 individuals from Bavaria were buried in six different graveyards:

Comprising more than 1,400 graves containing 1,520 individuals (13), Altenerding-Klettham (AED) is one of the biggest Early Medieval burial places in Bavaria. Its use begins in the middle of the 5$^{th}$ century AD and lasts until the end of the 7$^{th}$ century AD (14) and it is located near today's Munich in the middle of the former Roman province *Raetia II* (see Fig. 1B). Three individuals exhibiting an artificially deformed skull (graves 125, 513 and 1108), and seven individuals without an artificially deformed skull were included in genetic analyses.

Another such necropole was excavated at Straubing-Bajuwarenstraße (STR) which is located at the river Danube directly at the border of the former Roman province *Raetia II*. This burial place comprised 819 graves dating from the mid-5$^{th}$ to the mid-7$^{th}$ century AD (15). 17 individuals, three of which (graves 228, 328, and 535) exhibited an artificially deformed skull, of this burial place were included in the presented analyses. Altheim-Andreasweg (Alh; see Fig. 1B) includes 404 inhumations dating from the first half of the 5$^{th}$ century AD to the second half of the 7$^{th}$ century AD. Radiocarbon dating suggests that this burial place could already have been in use since Roman times (16). The site was included in the present study because the occurrence of individuals with artificially deformed crania was also reported here (16). Yet, none of the individuals in question proved to have an artificially deformed skull (7). 4 individuals were included in the present study.

In addition to these large burial places, several smaller burial groups, in which the presence of individuals with artificially deformed skulls were reported, were also included in the current study. All of these are located in the vicinity of Regensburg, which is situated at the river Danube directly at the border of the former Roman province *Raetia II* (see Fig. 1B).

Barbing-Irlmauth (BIM) is a small burial group of 29 inhumation graves dating to the first half of the 6$^{th}$ century AD (17). Two individuals, one with an artificially deformed skull (grave 33) and one without artificial skull deformation (grave 37) were included in the analyses. Due to the fact that this burial site was excavated in the 1930s, some inconsistencies regarding the archeological finds appeared: while grave inventory 33, without any doubt the equipment of a male, does not seem to match with the female with an artificially deformed skull, the artificially deformed skull might well be attributed to the female inventory of grave 31.

Burgweinting Nordwest II (NW) is one of three contemporary 5$^{th}$ century AD burial places that were oriented alongside a former Roman road connecting Regensburg to the south of the Roman province *Raetia II*. The burial site Nordwest II comprised 15 graves, some of which were equipped with rich grave goods. Among the 15 burials of Nordwest II one burial of a woman with an artificially deformed skull was recovered (grave 10254, NW 54; 18, 19). Additionally, the woman of grave 10255 was also included in this study.

Another individual with an artificially deformed skull stems from Alteglofsheim (AEH). This skeleton represents one of the very rare singular graves known from this time-period. Yet, it remains possible that further graves in its vicinity are still to be found (20).

In summary, all other individuals with artificially deformed skulls from Bavaria were found on regular, Early Medieval cemeteries.

None of the graves showed peculiarities in burial practice. There are no signs of any special treatments during the burial ceremony or regarding the duration of the laying-out. A secondary re-opening of the graves seems to have happened to two graves from Straubing-Bajuwarenstraße (15). It seems plausible that during this re-opening grave goods were removed at least from grave 228, while removal of grave goods cannot be excluded for grave 328 (7).

**Sarmatian samples**

*Melanie Groß*

Two Early Sarmatian individuals from Russian Pokrovka, Southern Ural (Fig. 1B), were chosen to complement the internal reference data set. They have previously been described in Unterländer et al. (21).

Burials were excavated from kurgans located within the territory of Russia's Orenburg Region (22). Both samples (PR_4, PR_10) were assigned to Pokrovka cemetry 02 that is divided in two groups, situated above the Khobda River near Pokrovka Village (23) and graves are dated to 4$^{th}$-2$^{nd}$ century BC. Most of the burials were individual graves with clay vessels (22).

Sample PR_4 was morphologically classified as a male adult, 30-40 years old, buried with some sheep bones, bronze arrowheads, one leather quiver, one iron dagger, one iron sword and one iron knife. The burial of the male individual PR_10 was found next to a 1.5-2-year-

old child, buried with sheep bones, glass beads under the child's chin, one iron knife, one iron fragment and sherds hand.

**Late Roman grave from Bavaria**

*Jochen Haberstroh, Bernd Trautmann, Andreas Rott*

As there is no final analysis published or available yet, the preliminary findings regarding the site of Freiham-Nord (FN) and especially the individual from grave 1335 shall be briefly presented here:

In the course of archeological investigations previous to construction works in the newly developed Munich district "Freiham-Nord", a previously unknown Late Roman cemetery containing inhumation graves was discovered in autumn 2014. In total, 20 graves could be recovered. They were orientated in different directions with the majority (13 individuals) lying on an east-west axis with the head to the east. The majority of the graves contained sex-specific grave goods typical of Southern Bavaria. Necropoles of this type in Southern Bavaria can be found as of the mid-3$^{rd}$ century AD and were typically in use until the first half of the 5$^{th}$ century AD. Not unusual for such cemeteries in the area surrounding Munich, the find material also contains elements of non-Roman origin. Grave goods originating in the so-called *Barbaricum* (brooches, ceramics) were already present at the onset of this type of Late Roman cemeteries.

Individuals were already sampled for subsequent population genetic analyses during excavation work. One individual was selected for analysis in the present study. With regard to the historical developments of the settlement in the 4$^{th}$ and 5$^{th}$ century AD, the selected burial had to be classified as "local", i.e. "Roman" in the present setting.

Grave 1335 in the eastern half of the cemetery was chosen. It contained the burial of a senile (aged 60+) man, who was buried in east-west direction on his back with his arms crossed over his pelvis. The grave cavity was very well differentiated from the surrounding gravel. It was 0.9 m wide and with 2.40 m exceptionally long. The man was probably buried on a wooden board ("Totenbrett") and there were no signs of burial offerings in the drawn in head section of the grave cavity. Meat offerings of chicken and pork were deposited at four positions next to the shoulders and the feet of the man.

Close to his left thigh a pierced bronze sheet and a thin iron ring could be found, as well as a big iron knife with a bronze plate pommel on the right of his right knee. A crossbow brooch of type 1 according to Keller/Pröttel could be recovered at the man's right ankle (Fig. S2). Finally, a soapstone bowl with antique traces of repair and a small bowl of the form Alzey 2/Chenet 320 a (Fig. S2), which probably resembles an imitation, could be recovered from the foot section of the grave cavity that was also slightly drawn in. Bowls comparable to the small one described here were found in Late Roman graves from Potzham, county of Munich, and Gilching, county of Starnberg near Munich. The crack in the bottom of the soapstone bowl was repaired using a riveted bronze clamp while the bowl's rim was fixed with an iron metal sheet. This allowed for further use of the bowl, with some limitations. The bronze crossbow brooch (Pröttel type 1) that was probably deposited in the foot section of the grave cavity together with an associated coat, is crucial for the dating of the burial. Similarly, cast pieces could be recovered from sites at Passau, Eining or Burghöfe where they were probably

manufactured. It is among the oldest types of crossbow brooches, whose wearers were very often identified to be executives in the Late Roman military. Dating around 300 AD, grave 1335 is among the oldest inhumation burials of the small cemetery at Freiham.

**Viminacium-Više Grobalja**
*Andreas Rott*

The burial place of Viminacium-Više Grobalja (VIM) is located in today's Serbia and is referred to as a Medieval Gepidian cemetery dating to around the middle of the 6th century AD. Of 103 skeletons excavated there, 46 could be identified as adult men, 27 as adult women, and 16 as children (24). In contrast to the Bavarian sites mentioned above, about one third of the individuals exhibited artificially deformed skulls. Here, both men and women of all ages could be diagnosed with this phenomenon (24). From this burial place, the individual from grave 2022, a subadult individual with artificially deformed skull (24), was included as a reference sample in the genetic analysis.

**Kerch**
*Bernd Päffgen, Andreas Rott*

A find which is especially interesting and important to the context of our samples, is that of an artificially deformed skull from antique Pantikapaion, today's Kerch in Crimea. As very few details about the find's history of discovery have been published so far, we will briefly describe it here:
The skull was discovered around 1900 in a burial chamber at Pantikapaion's Mithridates hill and reached the Collection of the Romano-Germanic Museum in Cologne on ways that are best described as adventurous.
Pantikapaion (Greek: Παντικάπαιον; Latin: *Panticapaeum*) is recognized as a Greek-Milesian colony that was established in the 6th century BC at the western shores of the Cimmerian Bosporus. Later, Pantikapaion gained importance as capital of the Bosporan Kingdom. In its golden period the city reached from the port at the road coming from Kerch, which connects the Black Sea to the Sea of Azov, up to the Acropolis on the Mithridates hill (25).
In Late Antique times the city's importance ceased and the settlement structure changed. Under King Rheskuporis VI. (reign: 314/15-341/42) Pantikapaion's mintage came to an end; yet, in his regency the city became newly fortified. Around the middle of the 4th century AD Gothic Tetraxits established their rule (26) before it was overtaken by the Huns in the seventies of the same century (27, 28). After they had destroyed the city to an unclear extent, the Huns and their allies were probably also aware of its strategically important location. In 534 Emperor Justinian I. conquered Pantikapaion, which was still important, and integrated it into the Byzantine area of authority. The city was also renamed "Bosporus" (Procop, De Aedif. III, 7; Bell. Pers. I, 12; bell. Goth. IV, 5). After roughly one generation the Byzantine rule ended in 576 when Angai, a prince of the Utigurs, conquered the city.
It is assumed that more than 2,000 graves were uncovered at and around the Mithridates hill and in the district of Glinisce. Yet, no exact inventories of the finds were passed on. Only exceptional finds were recognized, such as the "Kerch vases" which today are very well-

known in the community of antiquarians (29, 30), and the terracotta which were discovered there (31–33). Burial chambers that were described as catacomb graves were mainly laid out at the northern side of the Mithridates hill. Some of these graves were situated within the city boundaries of the Classic Epoch. The burial chambers were laid out during the Bosporan Kingdom and were used until Late Antique times. Some burial chambers that were painted or otherwise decorated were published by Rostovtsev (34).

Around 1900 (the exact time is unknown) a burial chamber containing several inhumation burials was discovered and attracted attention due to the find of a diadem with multi-colored stones. It was manufactured from three golden plates and its superelevated front plate resembles the form of two eagle heads. The diadem that is decorated with more than 250 individually set almandines and green glass (resembling the eyes of the eagle heads) can be dated to around 400 AD or into the first half of the $5^{th}$ century AD (35–39). Such golden diadems with precious stone inlays are typically found in female burials of the Hunnic elite (36, 39–44). According to Anke (45) the custom of wearing such diadems in the $4^{th}/5^{th}$ century AD may be seen as a stimulus from the Sarmatian cultural sphere.

The diadem and an artificially deformed skull from the same burial chamber which was not the skull of the diadem's wearer reached the Museum's collection via detours. Both are still held there today. It remains without doubt that both the diadem and the skull were discovered within the same burial chamber around 1900. From today's perspective it is very unfortunate that only the exceptional finds, diadem and artificially deformed skull, were paid attention to while the remaining grave inventory including several skeletons was not even properly documented.

In 2009 Dr. Anja Staskiewicz re-examined the skull from Kerch (Fig. S3) at the State Collection for Anthropology and Palaeoanatomy. The results of this re-examination are briefly summarized here:

As obvious in Fig. S3, the calvarium of the skull (labeled 43,95) is mounted to a wire rack that is attached to the occiput. A mandible is attached to the skull with a wire on both sides and is labeled "[belonging] to 43,95". Yet, doubts remain whether the mandible originally belonged to the skull. Due to missing mandibular condyles it was impossible to test how well these fit into the respective jaw sockets. Nevertheless the mandible seems less robust and masculine than the skull. It is furthermore conspicuous that 11 teeth are present in the mandible and most probably no tooth was intravitally lost, while not a single tooth was present in the maxilla and at least seven of these were intravitally lost.

Almost all features relevant for sex estimation of the calvarium point to a male individual due to their extreme robusticity (relevant criteria for sex estimation were taken from Herrmann et al. (46)). Additionally, age at death was roughly estimated by the degree of ossification of cranial sutures (relevant criteria were also taken from Herrmann et al. (46)). Under normal conditions the degree of ossification indicates a higher, possibly (late) mature age of ca. 50 to 60 years. Such an age-at-death would also be in accordance with the many intravital maxillary tooth losses.

For the genetic analyses presented here, a petrous bone from the calvarium was sampled. Radiocarbon dating from a piece of the same petrous bone yielded an age of 256 – 401 cal AD (1709 ± 23 years).

## SI2. Sample preparation and sequencing

*Melanie Groß, Andreas Rott, Joachim Burger*

Ancient DNA work on the skeletal samples from southern Bavaria, Germany, was carried out in dedicated facilities of the Palaeogenetics Group at Johannes Gutenberg-University Mainz. Decontamination and first steps of sample preparation (drilling and milling) were performed as described in Scheu et al. (47). For contamination monitoring and decontamination efficiency of used devices blank controls during milling (47) and all further steps were processed. A PCR on a mitochondrial fragment was performed on milling and extraction controls. In order to find the most suitable samples for nuclear target enrichment and whole genome sequencing, all sampled petrous bones were tested for high complexity and endogenous DNA content as described below (Table S2).

### DNA extraction

Extraction was performed according to Scheu et al. (47) and Hofmanová et al. (48) with slight modifications: 6.7 ml of EDTA (0.5M, pH8; Ambion/Applied Biosystems, Life technologies, Darmstadt, Germany), N-laurylsarcosine (250µl, 0.5%; Merck Millipore, Darmstadt, Germany) and proteinase K (30µl, 18U/µl; Roche, Mannheim, Germany) were added to bone powder in a lysis step. This solution was incubated on rocking shakers at 37°C for 72-120 hours. The DNA was isolated via phenol/chloroform/isoamyl alcohol (25:24:1, Roth, Karlsruhe, Germany), then desalted by stepwise washes with HPLC-water (10-12 ml) and concentrated to approximately 300µl using Amicon Ultra-15 Centrifugal Filter Units (Merck Millipore, Darmstadt, Germany).
For STR_288, STR_355 and AED_1108 two extractions were performed from one tooth and one petrous bone respectively.
PCRs on mitochondrial DNA fragment were performed for all extracts to monitor extraction success and cleanness of blank controls.

### Library Preparation

All libraries were prepared based on Kircher et al. (49) and as described in Hofmanová et al. (48). Two libraries of AED_432 and STR_300 were treated with USER-Enzyme and pooled with 4 non-treated libraries for target enrichment respectively (Table S3).
Index-Primer P7 and IS7 were used for library PCR with AmpliTaq Gold® DNA Polymerase (Applied Biosystems) in at least three PCR parallels and 10-12 cycles for screening libraries and 10-16 cycles for libraries prepared for target enrichment. A second round of PCR was performed with Herculase II Fusion DNA Polymerase (Agilent) and enabled us:
I.) to add individual P5 -index on short end of library molecules for MiSeq sequencing
II.) to increase molecule numbers for further target enrichment.
The cycle number was adjusted to the concentration of each library as measured by Qubit® Fluorometric quantitation (dsDNA HS assay, Invitrogen) and the Agilent 2100 Bioanalyzer System (HS, Agilent Technologies) following the manufacturer's protocols.

**Sample selection**

For 63 samples in total, quantity and quality of molecules of libraries prepared from each DNA extract were determined using a quantitative real-time PCR and shallow shotgun sequencing. Only DNA extracts with at least more than $10^8$ molecules per µl of library product and endogenous content > 30% were selected for subsequent nuclear target enrichment and whole genome shotgun sequencing (Table S2; Fig. S4).
All individuals with skull deformation and two Sarmatian samples were included in this study regardless of their screening results because these were of special interest.

A Quantitative real-time PCR

Quantitative real-time PCR was performed on each library fill-in product with KAPA Sybr Fast Universal Mastermix (PeqLab, VWR International) on a Step One PlusTM Real-Time PCR system (Applied Biosystems, Thermo Fisher Scientific) with primer pair IS7/IS8 as described in Hofmanová et al. (48).
As an additional monitoring step, molecule numbers for most of the libraries prepared for capture enrichment were gauged with quantitative real-time PCR.

B MiSeq screening

Endogenous content was estimated by calculating the percentage of aligned reads against all reads after quality filtering.
Among the selected samples, eleven show an endogenous content of above 30% and 23 samples above 50% endogenous molecules. Four samples (STR_241, STR_266, STR_360, STR_491) were selected despite an endogenous DNA content below 30% (28.58%, 27.66%, 20.93%, 15.63%) due to their high molecule number per µl of library fill-in product determined by qPCR. Two Sarmatian samples PR_4 and PR_10 were included in this study as internal references although they do not meet previously set quality criteria. The biological sex of the samples was determined using the method described in Skoglund et al. (50). Even in cases with low read numbers, results are consistent with morphologically classified sex. Post-mortem damage patterns were determined for all samples with the software package MapDamage 2.0 (51). For all ancient samples a typical increase of C/T deamination at the 5'-end of reads could be observed. Due to relatively good biomolecular preservation of Bavarian samples, deamination rates range from 0.13 to 0.35.

The bioinformatic pipeline for analyzing the data is described in SI4.

C blank controls

Extraction and library blank controls were treated as samples separately and transformed into libraries. All blank controls were measured on a high sensitivity chip with the Agilent Bioanalyzer. If a low concentration could be measured, either a qPCR and / or a MiSeq sequencing was performed. For shallow sequencing on MiSeq, complete blank controls were vaporized and resupended in 2 µl UV-HPLC-$H_2O$ to examine the whole library product.

Molecules in fill-in product of library blank control ranging from $1.32x10^5$ to $6.86x10^5$, represent low estimates of contamination between 0.004 to 1.25% per library (0.0003613 ng - 0.0025056 ng).
While evaluating the MiSeq screening results, between 2 and 10 unique molecules were counted that could be aligned against hg19 reference genome. For one extraction control (B_Ex_4) 3,687 aligned reads were detected.

**Mitochondrial Capture Enrichment**

All samples that were screened showed a DNA preservation that was good enough for enrichment of whole mitochondrial DNA. Sample DNA from at least two independent libraries (Table S2, S3) was pooled for every sample to achieve 20.13 ng to 277.40 ng of starting amount per sample. Whole mitochondrial genome capture was carried out using the custom designed SureSelectXT in solution target enrichment kit (Agilent) based on Gnirke et al. (52) according to the modified protocol described in Hofmanová et al. (48) without lowering the washing temperature. Samples were double-captured and PCR-amplified in three parallels. An individual P5-index was added to every sample after the second enrichment. Purification of the capture products between different steps was carried out using MSB® Spin PCRapace (Invitek, Stratec Molecular, Berlin, Germany).
Mitochondrial DNA of samples ALH_1, ALH_10, and FN_2 was retrieved from shotgun sequencing runs of these samples (Table S7).

**Nuclear capture enrichment**

Samples were pooled together from at least 4 libraries to achieve a DNA starting quantity of approximately 1 µg per sample per reaction and in order to maintain complexity (Table S4). Each library pool was concentrated using a vacuum centrifuge (SpeedVac) and subsequently diluted in 5.4 µl UV-HPLC-H2O. Double captures were carried out following the MYbaits protocol v2. Incubation for the first round of enrichment was set to 36 h, followed by a clean-up of the captured products and 7 - 10 cycles of PCR. Adapter sequences were blocked on the reverse strand only with a custom blocking oligo-mix around the varying index sequence including 1 µg/µl Human Cot-1 DNA and 1 µg/µl Salmon Sperm DNA. Incubation in the second capture cycle was set to 24 h, followed by purification of the products and 12 - 14 cycles of PCR in which also an individual P5-index was added to every sample. Purification of capture products was performed using MinElute™ PCR Purification Kit (Qiagen, Hilden, Germany) and PCR products were purified using MSB® Spin PCRapace (Invitek, Stratec Molecular, Berlin, Germany), both following the manufacturer's instructions. Enriched libraries were quantified with Qubit and visualized on Agilent Bioanalyzer HS chip to exclude dimer peaks and overamplification.

**Sequencing**

A Screening
MiSeq screening of library products was carried out on Illumina's MiSeq with 50 bp read length and single-end runs. All double-indexed libraries were diluted and pooled in equimolar

proportions on several lanes respectively, resulting in 300k to 1.3 million raw reads per sample.

B Enrichment

Sequencing of enriched mitochondrial DNA was carried out on an Illumina MiSeq machine (150 bp, single end) at StarSeq, Johannes Gutenberg-University Mainz, Germany. Samples were pooled together in equal quantity.
Enriched nuclear DNA of samples AED_1119, ALH_1, ALH_3, ALH_10, FN_2, PR_4, STR_316, and VIM_2 was carried out on an Illumina HiSeq 2500 platform at the Institute for Molecular Genetics, Johannes Gutenberg-University Mainz, Germany, using the rapid mode (100 bp, paired end; see Table S4). The other samples were also sequenced on an Illumina HiSeq 2500 platform at the same institute (100 bp, paired end). Four samples each were pooled together in equal quantity and subsequently sequenced on one lane per pool. Possible dimer peaks in prepared sequencing pools were eliminated while cleaning with AMPure SPRI beads.

C Shotgun

Next generation shotgun sequencing was carried out on an Illumina HiSeq 2500 platform at the Institute of Genetics, Johannes Gutenberg-University Mainz, Germany. To achieve this, two libraries each were pooled together for samples ALH_1, ALH_10, and six libraries for FN_2 and directly sequenced (100 bp, paired end) on one, one, and two lanes, respectively.
Additionally, samples NW_54, KER_1, STR_220, STR_300, STR_310, STR_355, STR_486, and VIM_2 were also shotgun sequenced on Illumina HiSeq 2500 (100 bp, single end). Four to seven libraries were pooled for every sample, the samples were pooled together and finally sequenced on two lanes. See Tables S3 and S4 for detailed information on pooling and sequencing strategy.

Average coverage calculation of hg19 for all samples is calculated including mitochondrial genome and refers to 3.75-35.81% coverage of whole genome for capture and 17.95-85.47% coverage for shotgun data.
See Table S6 for detailed sequencing results.

## SI3. Description and implementation of the 5 Mb capture assay

*Christian Sell, Melanie Groß, Karola Kirsanow, Joachim Burger, Krishna Veeramah*

In order to apply population genetic methods that make inferences from the allele frequency spectrum (53, 54) it is desirable to analyze specific regions of the genome that are considered putatively neutral (i.e. where patterns of genetic variation are not skewed due to the impact of mutations under natural selection either within or linked to the region of interest) across multiple individuals. Due to generally disparate conservation status of ancient DNA resulting in low coverage fragments of the genome, the intersection of such regions with high coverage could be comparatively small when performing shotgun sequencing. Therefore we designed an ~5 Mb capture approach to enrich samples of interest for multiple genomic regions to

create a reliable and comparable data resource for further usage of population genetic methods. The bait library were designed by Palaeogenetic Group at the Institute of Organismic and Molecular Evolution of Johannes-Gutenberg University in Mainz and by the Veeramah Lab at Stony Brook University in cooperation with MYcroarray (USA). Implementation of 5 Mb capture assay was carried out at post-PCR Laboratories of Palaeogenetic Group in Mainz and described in Groß (PhD dissertation; work in progress) and Sell (55).

**Identifying the regions**

Gronau et al. (56) previously identified 37,574 1 kb "neutral" regions on chromosomes 1 to 22. These were chosen such that they encompass regions that are syntenic with chimp (syntenic net track) and were subject to the following filters:

a) Recent segmental duplications (UCSC segmental dups track)

b) Recent transposable elements (RepeatMasker with elements with < 20% divergence from consensus)

c) Exons of protein coding genes, UTRs and 1000 bp flanking region (RefSeq, ENSEMBLE, UCSC)

d) Noncoding RNA and 1000 bp flanking region (RNA Genes track)

e) Conserved noncoding elements (phastCons elements and 100 bp flanking)

These loci were lifted over from hg18 to hg19, with four regions being lost due to the changes in the genome assembly. The liftover regions were then used as an additional inclusion mask within the program NRE (57). We then identified all regions identified by NRE that also overlapped 100% with the hg19 Gronau 1 kb regions. There is some redundancy between the filters of NRE and the original Gronau et al. filter (Known Genes, SegDups, EST, CNVs etc). Thus utilizing NRE provides an extra layer of stringency in identifying putatively neutral regions while accounting for the change in reference builds. We then applied the following extra filters to remove individual 1 kb loci:

a) Background selection (58) value < 0.85 (any more stringent than this and we start losing a lot more loci).

b) Recombination rate < 0.01 and > 10 cm/Mb

c) > 0.01 cM from genes

d) Extreme fastCons and simple repeats scores overlap (these were already low from the filtering by Gronau et al., but a few slipped through from hg18-hg19)

Finally we ensured that all loci were at least 50 kb from each other. After the filtering there were 4,687 1 kb loci. These were supplemented with 429 500 bp loci using similar but slightly more stringent criteria. With a background selection value of 0.95 and a distance of at least 0.5 cM between each of the regions and towards genes. In addition we have identified a

small set (n=26) of X-linked loci, which may be useful to look at sex differences in demographic history. These were found using NRE and are all separated by at least 500 kb and are not in the pseudoautosomal region.

**Identifying possible probe specificity/mapping issues**

We chose loci based on criteria that excludes regions containing repeats similar to the known consensus sequence (20% divergent). However, there may be still some issues with sequence similarity across the chromosome because of the recent replication of divergent repeats that could affect how well aDNA fragments hybridize to the designed probes or map to the reference genomes, especially if they contain substantial polymorphism or sequence errors. To explore this further we performed the following multi-alignment experiment using BLAST for each 1 kb loci.

a) Extract 100 bp sequences every 20 bp sliding along the entire length of the 1 kb region.

b) BLAST each 100 bp sequence against hg19.

c) For each 100 bp sequence determine the alignment score for its true genomic position and, if one is found, the alignment score for the next best match.

d) For each 100 bp sequence with at least one alternative alignment, extract 100 bp sequences 1-19 basepairs up and downstream (so we narrow down the precise problem regions after the coarser sliding window search in set a).

e) BLAST each 100 bp sequence against hg19

f) For each 100 bp sequence determine the alignment score for its true genomic position and, if found, the alignment score for the next best match.

We also repeated this analysis using 50 bp sequences with 10 bp sliding windows to reflect shorter aDNA fragments. Thus we should obtain almost all 100 bp and 50 bp windows where there are potential multiple alignments with hg19 such that they may cause problems in probe specificity or read mapping. Table S4 shows how many loci contain at least one 100 bp or 50 bp sequence that can align to an alternative region along the hg19 genome, as well as the total number of base pairs across all regions where these matches were identified. For 100 bp sequences there was no match that was as good as the true position, while for the 50 bp windows, there were 101 1 kb loci and a total 15,292 bp (0.3% of the total target sequence) where there was a match with another region of hg19. Further breakdowns are given in Table S5. For 100 bp sequences the probability of finding a match outside the target region with at least 90% of the alignment score of the true position is very low (only 35 1 kb loci with at least one alternative match, and 0.13% of the total target sequence). Unsurprisingly there is more alternative matching for 50 bp windows, though it still only encompasses ~1.2% of the total target sequence. To give some perspective on the values in the table below, with 90%, 80%, and 70% alignment scores the alternative region contains approximately 5, 10, 15 differences from the original 100 bp sequence and 3, 6 and 10 differences from the original 50 bp sequence, though the exact relationship changes due to the introduction of gaps and

appropriate gap penalties. Based on the figures below, probe specificity and mismapping should not be a major issue, though, depending on the aDNA fragment size distribution, there may be some small (~1-5%) amount of sequence across the total of ~5 Mb where we may expect poorer coverage after bioinformatics processing. From our BLAST results we have identified these regions such that we can account for them in any downstream analysis

In addition, to explore how actual fastq reads might be affected during aligning by BWA, we performed the following experiment.
For each region, 100 fastq reads were generated by randomly picking a range between 40-180 bp starting from anywhere in the region. To ensure the fastq reads aligned with BWA, a random base quality between 40-60 Phred-scaled quality score was assigned to each base. The reads were aligned using BWA default parameters and filtering for a mapping quality of 25 (bwa places multiple hits randomly and assigns a mapping quality of 0). Only 352 reads of 468,700 could not be aligned.

The divergence of the regions was checked further by adding up mismatches, deletions and insertion divergence from repeatmasker and only 50 regions were found that overlapped a repeat with a divergence less than 20% with more than 500 bp. From the 352 not aligning sequences 216 could be assigned to those regions.

Finally, mrna databases were also checked against all regions. 317 regions overlap with the mrna and EST database.

**Phenotypic Markers**

486 phenotypic informative markers were added to the capture assay. These SNPs were selected on the basis of association with aspects of the visible (e.g. pigmentation, morphology) or metabolic (*e.g.* lactase and alcohol digestion, diabetes risk) phenotype, or their association with certain medical conditions (e.g. non-infectious and inflammatory diseases), in modern humans. Many of the SNPs are also hypothesized to have been under relatively recent selection. Each SNP Position was extended for 60 bp on each side to achieve 120bp long sequence regions to enable further enrichment.

Bait design
In total 4,915,000 bases (0.16% of the whole genome) are covered by 80 bp long RNA-baits.
Overlapping regions were combined into continous sequences. For neutral regions and prolonged SNP positions a 5x tiling was applied. Under the assumption of recommended hybridization temperature and conditions from MYbait user manual in silico designed bait sequences were blasted against the human genome and melting temperatures were calculated. This allows multiple matching sequences to be marked according to six bins of Tm: 40-60°C, 60-62.5°C, 62.5-65°C, 65-67.5°C, 67.5-70°C and > 70°C. Various stringency filters could be chosen for bait design. In the present work, only baits were selected according to the strictest guidelines.

Following stringency filters were applied for most of the RNA-Baits:

A. No blast hit with a Tm above 60°C
B. ≤ 2 hits at 62.5 – 65°C or 10 hits in the same Bait interval and at least one neighbor candidate being rejected
C. ≤ 2 hits at 65 - 67.5°C and 10 hits at 62.5 – 65°C and two neighbor candidates on at least one side being rejected
D. ≤ 1 hit at or above 70°C and no more than 1 hit at 65 -67.5°C and 2 hits at 62.5 – 65°C and two neighbor candidates on at least on side being rejected

Some baits for functional markers were chosen with less stringency to increase the amount of baits covering the loci. In order to achieve better hybridization conditions of the bait sequences, the region of critical marker positions was expanded to 300 bp and less stringent filter criteria was applied (≤ 10 hits at 62.5 – 65°C or 10 hits and not surrounded by two candidates being rejected). If it could not be guaranteed under these relaxed filter criteria that all 5 baits covering SNP positions those bait sequences closest to original positions were selected to enable a possible by-catch.

## SI4. Read processing

*Christian Sell, Melanie Groß, Andreas Rott, Joachim Burger*

All sequence reads were processed as described previously in Hofmanová et al. (48) and Broushaki et al. (59), with the addition that reads that could not be merged prior to the alignment using ea-utils (60) were aligned using bwa aln and bwa sampe. All paired-end reads gained from bwa sampe were filtered for proper pairs.

For further population genetic analysis, bam files of whole genome sequencing and nuclear capture enrichment were merged together and processed as one.

See Table S6 for read counts and coverage information of the nuclear captures and shotguns and Table S7 for the mitochondrial captures.

## SI5. SNP and genotype call

*Krishna Veeramah*

Genotype calling

Genotype calling was performed for each of the 41 ancient individuals using the likelihood model described in Hofmanová et al. (48) but using weibull distribution to model post-mortem damage as described in Botigue et al. (61).

We took advantage of the high endogenous content of our ancient samples by determining allele-presence (i.e. haploid) calls at specific SNP positions identified in the POPRES and HellBus reference datasets (see SI6). These calls take advantage of "off-target" reads that overlap known SNPs. Note that these allele-presence calls will usually only be determined by a single overlapping read, and there will also be very little overlap amongst ancient samples (i.e. each ancient individual will have its own random distribution of off target reads); thus

between ancient sample comparisons cannot be made using these calls, only comparisons between an ancient samples and modern reference samples can be made.

In addition, because of the high-coverage obtained, we determined full diploid genotypes for the entirety of our 5 Mb capture region across all samples. Mean coverage at these regions ranged from 13x to 148x, with a mean 73x across the 41 samples (Fig S5). Excluding the two samples with lowest coverage (STR_328 = 16x, PR_4 = 13x), we were able to obtain reasonably consistent titv ratio at heterozygous and homozygous non-reference sites for all samples (mean 1.99) when we filtered out (in this case by making them homozygous reference) sites with a variant quality score (62) of <30 (Fig S6). Not including this filter led to some samples with excessive post-mortem damage (such as STR_502) and/or lower coverage also demonstrating excess titv ratios (suggesting post-mortem damage was excessive) and behaving abnormally in subsequent analyses.

## SI6. Modern Reference Datasets

We assembled a variety of reference datasets for comparison to our 41 ancient samples.

### Imputed POPRES and SGDP

The Michigan Imputation Server (63) was used to impute additional SNPs from the original 197,108 SNPs in 1,385 European individuals analyzed in Novembre et al. (64) that form part of the POPRES project (65). After accounting for potential strand flips there were 165,443 usable SNPs that could be orientated with human reference genome build 37. Phasing was performed using Eagle2 (66) using the worldwide sample set from 1000 Genomes Phase 3 (67) as the reference panel. Following imputation, biallelic SNPs were filtered to be >=5% minor allele frequency with a $r^2$ imputation of value of >= 0.9, resulting in a final imputed SNP set of 2,113,389 SNPs. These SNPs were then extracted from 279 publically available whole genomes from the Simon Genome Diversity Panel (68) and merged using PLINK (69). This modern reference dataset was merged with the allele presence calls for the 41 ancient genomes generated here as well as seven previously published 5-7th century low to medium coverage genomes from Britain (70). As these genomes were UDG treated, allele presence calls did not model post-mortem damage. Instead the first and last five bp of each read were removed before applying the standard GATK likelihood model (62).

### Imputed HellBus, 1000 Genomes and SGDP

The same imputation procedure was carried out on 512,368 SNPs in 1,582 European and Asian individuals from Hellenthal et al. (71) (hence forth known as the HellBus dataset). Following imputation, biallelic SNPs were filtered to be >=5% minor allele frequency with a $r^2$ imputation of value of >=0.9, resulting in a final imputed SNP set of 4,883,514 SNPs. These SNPs were then extracted from 279 publically available whole genomes from the Simon Genome Diversity Panel (68) as well as 504 East Asian genomes from the 1000 Genomes project (67) and merged with the HellBus dataset using PLINK (69). This modern reference dataset was merged with the allele presence calls for the 41 ancient genomes as well as the seven 5-7th century British genomes (70).

**1000 Genomes, SGDP, Turkish, and GoNL**

142,834 biallelic variants were found within the 5 Mb neutral region that were segregating amongst 1000 Genomes samples from the 15 Eurasian populations (European, South Asian, East Asian). Phased haplotypes at these variants were extracted and used as a reference panel for phasing these sites and specific alleles (i.e. we condition our other data on 1000 Genomes variants) for the 41 ancient samples, 264 Eurasian SGDP, 16 Turkish and 100 Dutch GoNL genomes.

The high coverage (32-48x) Turkish genomes were originally mapped to gr38 (72). Therefore, Liftover was used to identify the 5 Mb regions in this new build, fastq files overlapping these regions with 1,000 bp padding either side were extracted using bedtools bamtofastq, reads were remapped to gr37 using bwa (73), realigned using GATK Indelrealigner and multisample calling performed using GATK Unified Genotyper (62).

BAM files containing reads overlapping our 5 Mb region for 250 Dutch trios was kindly provided the GoNL consortium (74). The 100 Dutch parental samples with the highest coverage (17-26x) were used for multisample calling performed using GATK Unified Genotyper (62).

Phasing was performed at both a whole genome level using Eagle2 (66) and for individual 1 kb loci using PHASE (75) under default parameters, with 1000 Genomes phasing used to supervise the phasing of other genomes.

**GoNL AFS**

In order to construct a robust allele frequency spectrum AFS) from our medium to high coverage Dutch genomes (for which we do not care about haplotype phase), single-sample called vcf files for all 750 GoNL individuals (250 trios) for the 5 Mb regions were grouped into trios, phased by transmission using GATK (62) and then for each site the most likely allele from each parent (the highest homozygous genotype likelihood) in a trio combined to create an AFS across the 5 Mb region for 250 pseudo-diploid individuals.

## SI7. Functional Variants

*Andreas Rott, Jens Blöcher, Melanie Groß, Michaela Harbeck, Karola Kirsanow, Joachim Burger*

419 single nucleotide polymorphisms (SNPs) included in the capture described in SI3 were called for all individuals as described in SI5 and genotypes were assessed (67 capture positions located on the X and Y chromosomes were excluded from these analyses).
The Bavarian individuals were each assigned to one of the following two groups:

1. The “non-deformed” group, which also includes three individuals of morphological intermediate status but not STR_300, STR_502, and STR_310, which showed substantial or complete non-central European ancestry. STR_491 was also excluded from this group as it is a sibling of STR_355 (see also SI15).

2. The “deformed” group with all individuals displaying unquestionable skull deformation.

**Markers associated with pigmentation phenotype**

All markers except for rs12821296 (*KITLG*) of the HIrisplex complex (76) were investigated. Only genotypes with QUAL ≥ 30 were submitted to the HIrisplex online tool (https://hirisplex.erasmusmc.nl/) in order to assess individual iris and hair color phenotypes. Furthermore, several markers affecting skin pigmentation were assessed and skin color was reconstructed according to the 7-SNP approach described by Spichenok *et al.* (77): Individuals homozygous for the derived allele of at least two loci out of rs12913832, rs1545397, rs16891982, rs1426654, rs885479, or rs12203592 were classified as “non-dark”, while individuals homozygous for the ancestral allele at rs6119471 were classified as “non-white”.

Using the HIrisplex eye and hair color assignment scheme (Table S8), the “deformed” group appears different from the “non-deformed” group (Table S9; Fig. S7A, B). These differences are less pronounced when male individuals are compared to female individuals. Regarding eye color, observed differences are significant for the groups “deformed” vs. “non-deformed” (Fisher Exact Test; $p = 0.0026$). Although there are far more blonde individuals in the “non-deformed” group, the difference is not statistically significant. If only *HERC2* (rs12913832) is considered, there is also a significant difference (Fisher Exact Test; $p = 0.0052$) between the frequencies of the derived allele (depigmentation) in the deformed group (55.6%; 95% CI: 0.326-0.785) and the non-deformed group (89.1%; 95% CI: 0.801-0.981). Again there is no significance when females are compared to males.

Most individuals’ skin tone was classified as “non-dark” (Table S10). The falling along spectrum from depigmented to partially depigmented skin color predicted for the Bavarian individuals therefore falls within the typical modern range for central Europeans. Individuals AED_1108, AED_1135, KER_1, STR_328, STR_486, and VIM_2, are all heterozygous at most of the loci affecting skin color and might therefore have had a relatively darker skin tone than the Bavarian individuals homozygous for the derived allele at all positions. All of the Bavarian individuals carry at least one copy of the C11-haplotype of *SLC24A5* carrying the derived allele at rs1426654 (*A111T*), which is almost fixed in all European populations (78; Table S11). Previous analyses indicate that this allele is the largest single contributor to depigmentation differences observed in modern Europeans relative to modern Africans (79, 80). Only individuals AED_1108, PR_10, and VIM_2 were heterozygous at some positions (see Table S11), which is consistent with these individuals each carrying one copy of C11 and C9, C3, and C7, respectively. The C9, C3, and C7 haplotypes are not observed in modern European HapMap populations and have distribution peaks in Central/East Asia today (78). We additionally assessed a 3-SNP *SLC24A5* haplotype as described and identified in Giardina *et al.* (81) (Table S11). Consistent with the 16-SNP method, AED_1108, PR_10, and VIM_2 show patterns that today are absent in Europe but appear in Asia with frequencies of 13.9% (AED_1108) and 23.5% (PR_10, VIM_2).

The derived allele of an additional SNP (rs1800414) in the *OCA2-HERC2* region, which causes light skin color via a missense mutation and is restricted to East Asian populations

(82), did not appear in any of the samples investigated.
Two individuals (AED_92, ALH_3) are heterozygous for the derived allele of *EDAR* rs3827760 (Table S10). This allele has a frequency of 87.3% in East Asia and 1.1% in Europe (EUR; Ensembl GRCh37 release 88, last retrieved on 23 May 2017), and is associated with incisor shape and hair morphology (83, 84).
*OCA2-HERC2* haplotypes were determined following the 13-SNP approach by Eiberg *et al.* (85) and via determination of the 3 "blue-eye haplotypes" (BEH) defined by Donnelly *et al.* (82). Results (Table S12) for the *OCA2-HERC2* haplotypes were consistent with eye color prediction of HIrisplex; all individuals predicted as having the highest probability for blue eyes in HIrisplex are most probably homozygous for a blue-eye haplotype according to both methods (according to the Donnelly-method, they are either homozygous for only BEH2 or for all BEH). Individuals that were assigned brown-eyed by HIrisplex at least carried one brown-eye haplotype. Interestingly, five individuals (AED_1135, AEH_I, STR_300, STR_328, VIM_2) can be excluded from the blue-eyed haplotypes identified by Eiberg *et al.* (85), but could not be conclusively assigned to one of the 6 commonly observed brown-eyed haplotypes.
The derived allele of an additional SNP (rs1800414) in the *OCA2-HERC2* region which causes light skin color via a missense mutation and is restricted to East Asian populations (82), did not appear in any of the samples investigated.

**Markers associated with metabolic phenotype**

Xenobiotic metabolism
All individuals were genotyped at 8 SNPs in the *NAT2* region, which is associated with xenobiotic metabolism through acetylation speed (slow/ intermediate/ rapid). Acetylation phenotypes are associated with the risk of developing certain cancers and the efficacy of drug metabolism (86). It has been hypothesized that the *NAT2* region might have been under recent selection as a consequence of altered human exposure to toxins and carcinogens in the course of transition to agriculture (87).
The Bavarian individuals' most probable acetylation phenotypes were determined using the online tool nat2pred (http://nat2pred.rit.albany.edu; 88), which infers the phenotype from 6 SNPs within the *NAT2* region (Table S13). The acetylation phenotypes inferred using the online tool were consistent with phenotype determinations made using a single SNP (rs1495741) that correlates with a 7-SNP *NAT2* haplotype (89). In accordance with previous observations, the ancient Bavarian population and its different subgroups show a higher proportion of the slow acetylation phenotypes (i.e. slow and intermediate phenotypes), which is similar to the pattern observed in modern central Europeans and typical for agriculturalists in general (Fig. S8; Table S13; Table S14; 87, 89, 90). Differences between the deformed and the non-deformed group are not significant (Fisher Exact test; $p > 0.05$). However, there is a significant difference between the male and female group (Fisher Exact test; $p = 0.03$). The observation of relatively high frequencies of intermediate acetylation phenotypes could be the result of balancing selection acting on heterozygotes: as both homozygous slow and homozygous fast acetylators are at higher risk of developing certain cancers, being heterozygous might be advantageous over any homozygous state at this locus (86, 91).
Proteins of the cytochrome P450 family are involved in drug metabolism and bioactivation

(92–94). Modern populations display a great deal of phenotypic variability related to cytochrome enzyme activity, which may be related to past selection (e.g. 93, 95, 96). We determined the haplotypes of several cytochrome genes in order to characterize the Bavarian individuals.

Results of the genotyping of cytochrome P450 enzyme 2C19 (CYP2C19) are shown in Table S15. Phenotypes were assigned according to the nomenclature of 'The Human Cytochrome P450 (*CYP*) Allele Nomenclature Database' (http://www.cypalleles.ki.se/) as far as possible. This locus is of modern medical relevance because the ultra-rapid metabolizer allele (*CYP2C19*17*) is associated with certain serious drug complications, and the various deficiency alleles have been associated with poor drug response (see 97). As with *NAT2*, different *CYP2C19* phenotypes are associated with different cancer risks. The majority of the Bavarian individuals (75%, 95% CI: 0.609-0.891) appear to be homozygous for the wild-type allele (*CYP2C19*1*, extensive metabolizers) with minor frequencies of heterozygous *CYP2C19*17* carriers (intermediate and indeterminate metabolizers), which is the pattern observed in a pan-ethnic sample of modern populations (see 97). The Asian-specific deficiency allele *CYP2C19*3* (93, 97) could not be detected within the sample. Differences in phenotype frequencies (Table S16) among the different groups of individuals (Fig. S9) were not significant (Fisher Exact test, $p > 0.05$). However, high-confidence phenotypes could not be determined for all of the ancient individuals, and there is a huge complexity of possible allele combinations with allele **17* (see 97), which could not always be properly assessed for the ancient samples. Therefore, statistical testing between modern and ancient samples was not conducted.

Like *CYP2C19, CYP2D6* is a gene in the cytochrome family having important implications for drug efficacy in modern populations. DNA preservation did not allow determining *CYP2D6* alleles for every individual of the sample (Table S17). In 14 Bavarian individuals that could be genotyped, eight displayed a haplotype combination associated with normal enzymatic function (i.e. a combination of haplotypes *1, and *2), five with intermediate function (i.e. a combination of *1 or *2 with haplotype *4), and one with slow function (i.e. one haplotype with decreased function [*10] together with one haplotype with no function [*4]) (94).

Ancient Bavarian haplotype frequencies were 44.8% (95% CI: 0.267-0.629) for *CYP2D6*1* (95% CI: 0.267-0.629), 31.0% for *CYP2D6*2* (95% CI: 0.142-0.478), 20.7% for *CYP2D6*4* (95% CI: 0.059-0.354), and 3.4% for *CYP2D6*10* (95% CI: 0-0.101). This pattern is similar to the allele frequency spectrum of modern Germans/ Europeans (94). It was not possible to assess differences in haplotype and/or phenotype frequencies between ancient Bavarian sub-groups because too few individuals could be genotyped.

Frequencies of the deficiency alleles of *CYP3A4* and *CYP3A5* vary with geographical latitude (98). Although the products of both genes are of modern clinical significance because of their importance in drug metabolism, it has been suggested that their modern patterns of variation were shaped by their roles in sodium reabsorption (*CYP3A5*) and vitamin D metabolism (*CYP3A4*). It is hypothesized that the advantage in salt and water retention conferred by the ancestral allele *CYP3A5*1* decreases with distance from the equator, while the risk of pre-eclampsia associated with the same allele increases, creating a selection pressure favoring the derived nonexpressor allele *CYP3A5*3* at high latitudes (98). In the case of *CYP3A4*,

selection probably acted against the high-expressor allele *CYP3A4*1B* as it potentially increases the risk for rickets through its involvement in vitamin D homeostasis (95). The derived alleles *CYP3A5*3* and *CYP3A4*1* appear linked in the four most common haplotypes in modern Europeans, which together account for >80% of all haplotypes (95). Consistent with these modern observations, *CYP3A5*3* accounts for 86.1% (95% CI: 0.781-0.941) of alleles present in the ancient Bavarian sample (Table S18), while the ancestral allele *CYP3A5*1* only appears at 13.9% (95% CI: 0.059-0.219) with only one individual (ALH_2) homozygous for this variant. Of the 14 Bavarian individuals which could be genotyped for *CYP3A4*, 13 appear homozygous for allele *CYP3A4*1*, while only one individual (AED_204) is heterozygous at this locus (Table S18).

The *ABCB1* or multidrug resistance (*MDR1*) gene plays an important role in xenobiotic elimination. In Asian and European populations, haplotypes carrying derived alleles at three positions (rs1128503, rs2032582, and rs1045642) are hypothesized to be selected and make up 37% and 41% of all haplotypes found in these populations, respectively (99, 100). 23 of the Bavarian individuals exhibit genotypes consistent with each of them carrying at least one copy of this 3-SNP haplotype (Table S18). 18 individuals have at least one copy of ancestral allele at all three positions, which may represent haplotype *ABCB1*1*, the second most abundant allele among Asians and Europeans (haplotype frequency of 15% in both populations; 99). Precise haplotype determinations and haplotype frequency estimations for this locus are not possible with the SNP coverage in the current unphased dataset.

Similar to *CYP3A5*, the angiotensinogen gene (*AGT*) also contributes to hypertension risk in humans (101, 102). One particular mutation in the gene's promoter region (A-6G) is assumed to play an important role in reduction of hypertension risk in most populations outside Africa (101). We examined all individuals at 9 positions in the *AGT* region (Table S19) in order to determine the presence/ absence of four risk alleles (A-6, C4072, C6309, G12775) and possibly extrapolate haplotypes (102). 27 individuals carry at least one copy of the possibly protective haplotypes H2 or H4 (102), 16 of these are either homozygous for one of the protective haplotypes or potentially carry both.

Alcohol metabolism

We genotyped all samples at *ADH1Bb* and *ADH1Ba* (rs1229984 and rs3811801) and *ALDH2* (rs671). These genes encode enzymes that are crucial in the metabolism of alcohol into acetaldehyde and acetaldehyde into acetate. Generally, derived alleles at these loci are suspected to be under recent positive selection in East Asians but virtually absent in other regions of the world (103–105). At *ALDH2* all individuals appear homozygous for the ancestral allele (Table S20). In the case of *ADH1Ba* and *ADH1Bb* most of the individuals were also homozygous for the ancestral allele. The only exceptions are seen in NW_54 who is heterozygous at *ADH1Bb*, and VIM_2 who appears heterozygous at *ADH1Ba*. These genotypes indicate that all of the ancient Baiuvarians exhibited the ancestral phenotype common in modern Europeans (e.g. 103, 105).

Lactase persistence

We typed individuals at rs4988235 (*LCTa/MCM6* -13910C>T) and rs182549 (*LCTb* -22018A>G) in order to assess their ability to digest lactose during adulthood (lactase

persistence) (106). The genotypes are presented in Table S20, the allele frequencies are shown in Table S21 and Fig. S7C. Lactase persistence status could not be assessed for AED_106 and AED_432 as both *LCT* SNPs could not be determined for these individuals. All other individuals belonging to the non-deformed group except for AED_1135, ALH_2, ALH_3, NW_255, STR_360, and STR_393 (15/21 total non-deformed individuals where *LCT* SNPs could be determined) are lactase persistent. Among the individuals with artificially deformed skulls, AED_513, AED_1108, and STR_328 appear non-persistent, while all other females with artificially deformed skulls are lactase persistent. Among the reference samples, PR_10 and KER_1 are non-persistent, while PR_4, FN_2, and VIM_2 are lactase persistent. As for modern samples (e.g. 107), *LCT* SNPs are in high LD in the ancient sample as well. The only exception is ALH_1 (Table S20).

The -13910C>T allele frequency difference between the deformed and non-deformed groups is close to significance ($p = 0.0501$, Fisher Exact Test, Table S22). Compared to modern European populations from the 1000 Genomes Project, the non-deformed group is significantly different from the Central European (CEU) and the Tuscany (TSI) population while the deformed group is significantly different from all European populations except Iberians (IBS) and Europeans in total (EUR) (Table S22). This indicates that the non-deformed group falls into the allelic spectrum of north/ central European populations while the deformed group falls into the spectrum of Southern European populations. We also observed a significant difference in allele frequency between males and females ($p = 0.0125$; Fisher Exact test; Table S22). Compared to the modern 1000 Genomes populations, males only differ significantly from IBS and TSI, while females differ significantly from all populations except IBS (Table S22).

Metabolic Syndrome – associated markers

It has been hypothesized that susceptibility to a number of diseases, including the various facets of metabolic syndrome (MetS) (*e.g.* obesity, hypertension, type 2 diabetes, dyslipidaemia), have varied in recent human evolutionary history, and that this variability stems from gene-environment interaction (e.g. 108). Here we used modern type 2 diabetes genetic risk scores to compare ancient Bavarian variability at MetS-associated loci to that observed in modern Europeans (the average modern genotype is modeled here by the score representing "normal risk") .We determined individual risk scores using scoring metrics developed by Meigs *et al.* (109), Cornelis *et al.* (110), and Hivert *et al.* (111), and by assessing the *TCF7L2* haplotype of each individual following the definition of Helgason *et al.* (112). Of the 34 SNPs taken into account by Hivert *et al.* (111), only 33 SNPs could maximally be assessed per individual as rs4747969 (*CDC123*) was not part of the Nuclear Capture assay. All of these risk scores are not independent of each other as there is a certain degree of overlap between markers used for each of them.

Genetic risk scores (GRS) were calculated by summing up the numbers of risk alleles per individual. Missing SNPs were accounted for by calculating a standardized GRS following Cornelis *et al.* (110): count GRS = [total number of risk alleles/number of alleles present] × number of maximum alleles per model. This standardization was only applied to individuals with at least half of the loci present per score model. Standardized ancient Bavarian genetic risk scores according to the Meigs *et al.* model (109) ranged from 14.78 to 27.31 (Table S23),

and mean values for the different groups we identified in our sample ranged between 20.43 and 20.89. As Fig. S10 demonstrates, these values are well distributed around 17.7 which is the mean risk score of modern Americans developing diabetes in the study by Meigs *et al.* (109). The same pattern can be seen when T2D risk scores are determined following the Hivert *et al.* (111) model: the standardized GRS of the Bavarian individuals reach values between 26.97 and 46.90 and are distributed around the modern weighted median score of 37 (111; Fig. S11, Table S24). Mean values of all different groups range from 35.25 to 36.43 and are therefore also in good accordance with the published value for modern individuals (see above). Applying the Cornelis *et al.* (110) model, ancient Bavarian standardized GRS values vary between 7.14 and 16.67 (Table S25). Mean values for all groups range from 10.26 to 11.76 and are comparable to the median value which was 11 for both cohorts of modern individuals examined by Cornelis *et al.* (110).

If the GRS results reported here occurred on a modern European genetic background, they would be consistent with average T2D risk. Nevertheless, development of T2D is not only dependent on the genetic risk but also on several environmental factors, such as physical activity and nutrition (110). The risk score alone is therefore not sufficient to infer an individual's T2D status. Although standardized GRS tend to be higher in the deformed than in the non-deformed group (Table S26), these differences are not significant (t-test; $p > 0.05$ for all groupings tested).

*TCF7L2* haplotypes were determined as defined by Helgason *et al.* (112). Haplotype A (HapA) which is assumed to confer some protection against T2D, is defined by the derived allele (C) at rs7903146 and the derived allele (A) at rs10885406. HapA furthermore correlates with the ancestral allele (T) at rs7924080. Haplotype B (HapB) comprises haplotypes carrying the ancestral allele (T) at rs7903146. From all haplotypes determined for Bavarian individuals (Table S27), HapA accounts for 44.4% (95% CI: 0.330-0.560) and HapB for 55.6% (95% CI: 0.441-0.670) of the sample. This is in the same range as CEU samples, where HapA frequency is at 58%, while it accounts for 95% of all alleles in East Asians (112). Generally, the number of individuals carrying at least one copy of the protective haplotype seems to be a little bit lower in the deformed group but these differences are not significant (Fisher Exact test; $p > 0.05$).

We furthermore assessed a haplotype in *SLC16A11* which is associated with T2D risk in Latin American populations and also appears at low frequencies in Europe and Asia (113). An additional SNP (rs312457) located in *SLC16A13*, which is associated with T2D risk in East Asian populations (114), was also genotyped. None of the individuals genotyped for *SLC16A13* except for AED_1119 displayed the T2D risk associated SNP (Table S28). Regarding *SLC16A11* haplotypes, most of the individuals showed the reference haplotype or SNP combinations that are consistent with carrying the reference haplotype (in cases where only few SNPs of the haplotype could be genotyped; see Table S28). Individual AEH_I was heterozygous at rs13342232 and rs13342692 which indicates that this individual carried one copy of the ancestral haplotype and one copy of the "2 SNP" haplotype defined by SIGMA (113), which appears at frequencies of < 1% in Europe and is absent in Asia. Individual STR_266 was also heterozygous at rs13342692 and homozygous for the derived allele at rs117767867 which would be consistent with this individual carrying at least one copy of the "5 SNP" haplotype defined by SIGMA (113) which is associated with T2D risk and appears

at frequencies < 2% in European populations and about 12% in Asian populations (113). Yet, as other positions of the haplotype could not be genotyped for this individual, definite determination of the haplotype is not possible in that case.

Markers associated with non-infectious and inflammatory diseases

We furthermore assessed the status at several loci (Table S29) associated with the development of different diseases, such as Crohn's disease (CD), type 1 diabetes (T1D), multiple sclerosis (MS), and celiac disease (CeD). Inflammatory-disease associated alleles are believed to have risen to higher frequencies in tandem with changes in pathogen load, as they may confer resistance to various disease organisms (115). In all markers allele frequencies were similar between the Medieval sample and in modern European (Table S30). The only exception from this is rs2058660 with an estimated sweep date ~ 7,500 ya (115). Here, the non-deformed and male individuals show significantly lower frequencies than modern European samples ($p < 0.05$, Fisher Exact Test), while the deformed and female group do not. Differences between groups were not significant.

Rs2188962, rs6822844, and rs17810546 have been shown to have undergone recent positive selection along with four other markers forming a core network. Raj *et al.* (115) suggested that the selective pressure behind this sweep may have been a *Yersinia pestis* pandemic, the so-called Black Death. As our Medieval sample clearly predates the era of the Black Death, we might expect to find allele differences between our sample and modern populations. However, no such differences in the frequency of the derived alleles were observed (Fig. S7D, Tables S30, S31).

The Medieval Bavarian sample also predates the earlier Justinianic Plague, an Early Medieval *Y. pestis* pandemic that reached southern Germany during the 6th century (116–118). It is possible that earlier *Y. pestis* epidemics were already driving the supposedly protective alleles to higher frequencies. Several earlier plagues - potentially *Y. pestis* pandemics - were recorded for the Roman Empire (e.g. 119, 120), and a recent study demonstrated the presence of *Y. pestis* in human skeletons as early as the Bronze Age (121). However, such early dates are inconsistent with the estimated date of the selective sweep.

Although many of the Bavarian individuals carry risk alleles for the above mentioned diseases, the "hygiene hypothesis" poses that these alleles only became deleterious very recently due to the reduction of microbial load by modern hygiene standards. Therefore, it cannot be inferred whether any of the individuals in the sample suffered from Crohn's disease or not.

## SI8. Analysis of uniparental markers

*Andreas Rott, Melanie Groß*

### Mitochondrial haplogroups

Mitochondrial DNA was investigated for all individuals that were part of this study (see Table S7). Determination of mitochondrial haplogroups (mtDNA hg) was achieved using Haplofind (122). Mitochondrial genomes yielded average coverages of 123.49 to 557.68x. The ancient Bavarian individuals' haplogroups belong to clades U4, U5a, U5b, W, X, J, I, H, HV, T1, T2,

K, and C. These haplogroups, except for the one from the C-clade, are not uncommon among modern central Europeans (123, 124).

Between-group comparisons were conducted for the groups “non-deformed” and “deformed” (see SI 7 & 15). Haplotype and nucleotide diversity, as determined with Arlequin ver 3.5.2.2 (125) according to Nei (126) and Tajima (127) (Table S32), indicate a rather heterogeneous population, which is in accordance with the assumptions made regarding composition of the Baiuvarii Genetic distances (according to Slatkin (128)) between groups and between Bavarian samples and other contemporary samples from eastern and southern Europe are negligible (Table S33).

Among individuals with artificially deformed skulls we find hgs belonging to clades C4a, H5, H7, HV9, K1a, T2b, U4a, U5a, and U5b. All of these haplogroups, with the exception of hg C4a, appear throughout Europe at least at low frequencies. Nevertheless, U4 just like hg C also appears in northwestern Siberia at a rather high frequency (129). Furthermore, also hgs U5a, U5b, T1, K, HV, and H5 appear at higher frequencies in eastern Europe and the Caucasus (130, 131) and might therefore hint at an influx of people from eastern Europe/ Western Asia into today’s Bavaria/ central Europe. Yet, among the individuals without artificially deformed skulls we also find hgs V, U5b, H2, H1, which might hint at an eastern origin as today these hgs are more frequent in eastern Europe and Western Asia (123, 130–134).

Seven Bavarian individuals with deformed skulls (out of nine; 77.8%; 95% CI: 50.6 – 100%) exhibit hgs that are among those with a possible eastern affinity. In the non-deformed group only eight (out of 23; 34.8%; 95% CI: 15.3 – 54.3%) exhibit hgs that appear at higher frequencies in eastern Europe/ Western Asia than in western/ central Europe (130, 131).

In summary, the distribution of mitochondrial haplogroups would be in accordance with the initial hypothesis that there is genetic influx from eastern regions to central Europe during the Migration Period and that women with deformed skulls might be morphologically visible traces of such an influx.

**Mitochondrial and X-chromosome contamination estimates**

Mitochondrial and X-chromosomal reads were processed as described in SI4. The mitochondrial contamination rate was assessed for all mitochondrial genomes using the likelihood approach described in Fu et al. (135). The genomes exhibit potential contamination ranging from 0.006 – 7.094% (Table S34). X-chromosomal contamination was assessed for all male individuals for which shotgun sequencing data were available (ALH_1, FN_2; KER_1, STR_486, and VIM_2). This was achieved using ANGSD (136). As described in the software’s documentation, a binary count file of positions X:5000000-154900000 was generated for every sample from which contamination was estimated using a Fisher’s Exact Test and jackknife (“contamination.R”). The analysis was restricted to unique sites on the X-chromosome (“RES/ChrX.unique.gz") and known HapMap polymorphic sites ("RES/HapMapChrX.gz"). Contamination as assessed from X-chromosomal DNA ranged

from 0.1779 – 1.5904% (Table S34).

## SI9. Principal component analysis

*Krishna Veeramah*

All principal components analysis (PCA) was conducted for the ancient Bavarian samples alongside each of the three reference dataset using smartpca (137).When analyzing the off-target calls, individual pseudo-haploid PCAs were conducted for each ancient sample separately using a maximum of 200,000 overlapping SNPs with the reference dataset (if more than 200,000 SNPs were overlapping a random set was sampled), and then individual analyses were combined using a Procrustes transformation in R using the vegan package as described previously in (138). In addition haploid calls in the seven 5-7th century genomes from Britain (70) were included in these analysis. We constructed three such PCA plots, one using the POPRES reference dataset (Fig. S12), one using the HellBus, SGDP and 1000G East Asians reference dataset (Fig. S13), and one using the HellBus, SGDP and 1000G East Asians reference dataset minus any modern individuals of South Asian origin (Fig. S14).

When analyzing data from the 5 Mb neutral regions, as suggested in Patterson et al. for microsatellite data (137) each distinct 1 kb haplotype from each of the ~5,000 regions was coded as a distinct biallelic SNP allele in plink format, i.e. either an individual had two copies of the haplotype (coded 2), one copy (coded 1) or no copies (coded 0). There were ~90,000 such haplotypes amongst the combined modern and ancient sample dataset. Unlike for the off-target SNPs, a single diploid-based PCA was constructed using all modern and ancient samples without the need for a procrustes transformation. However, PR_4 and STR_328 were clear outliers in this analyses of PCs 1 and 2 (Fig. S15). As there are the two samples with the lowest coverage and more extreme titv ratio, we concluded this was a result of poor phasing for these samples, and they were removed from further analysis.

We then conducted three PCA analyses: a) all ancient samples as well as 1000 Genomes and SGDP Eurasian samples (Fig. S16), b) all ancient samples as well as 1000 Genomes samples from Europe and SGDP Western Eurasian samples (Fig. S17) and c) all ancient samples as well as 1000 Genomes and SGDP samples from Europe (Fig. S18). Despite being constructed using very different data types, this plots were highly concordant with the Procrustes transformed PCAs using off-target SNPs.

## SI10. Model-based clustering analysis

*Krishna Veeramah*

Model-based clustering analysis was applied two different ways to the haplotypes from 5 Mb region. In the first a supervised analysis was performed on the ancient samples using ADMIXTURE (139), treating the data as in the PCA analysis with each haplotype coded as 0,1 or 2. Given our PCA results, the following reference populations were utilized: 1000 Genomes CEU, GBR, IBS, TSI, FIN, along with SAS and EAS (all South Asian and East

Asians pooled into single populations) and GoNL. These populations were chosen as they possessed large sample sizes (>100 individuals, and thus should provide high resolution in the difference of haplotype frequencies, even when they are subtle). 100 independent runs were performed for the 38 ancient samples (we excluded PR_4 and STR_328 because of their low coverage and STR_491 because of their relatedness to STR_355). To examine consistency across runs, we plotted the estimated admixture coefficient for each reference population across the 100 runs. Estimates for GBR, CEU and GoNL were highly variable across runs (Fig. S19). However, summing coefficients for these populations into a single northern/central European provided more consistent results (Fig. S20), demonstrating the ADMIXTURE found it difficult to assign ancestry to one of these three specific populations because of their genetic similarity. However, in general GoNL had higher coefficients, followed by CEU and GBR.

Estimates of EAS, SAS and to a lesser extent IBS and FIN were consistent across runs, suggesting these results are fairly reliable. Our final plot (Fig. 2C) is based on the run with the highest likelihood and with CEU, GBR and GoNL merged into a single population.

The same analyses was conducted on old world SGDP individuals from Europe, the Caucasus, Central Asia, the Middle East, South Asia and East Asia in order to better understand the expected admixture coefficients of modern samples given our approach of eight reference populations. Only European individuals from south and southeast Europe demonstrated TSI ancestry (Fig. S21). No modern European possessed any EAS or SAS ancestry (making the result for AED_1108 highly unusual), except one individual from Hungary that showed a very small East Asian component. As neither a Caucasus nor Middle Eastern reference panel was present, all individuals from these two regions (Fig. S22 and Fig. S23) demonstrated a very high TSI component, though the latter also had South Asian ancestry. Central Asian individuals demonstrate a highly mixed ancestry (Fig. S24), while East Asians (Fig. S26) and South Asians (Fig. S25) tended to derive most of their ancestry from their respective reference populations.

In addition we performed an unsupervised analysis using the variable haplotypes only for the 38 ancient samples. Our first attempt using ADMIXTURE failed to identify consistent clusters. However, the use of STRUCTURE (140) allowing for correlated allele frequencies successfully identified two consistent clusters of individuals across 100 runs. When testing for the most likely value of K across K=2-6 using both the mean probability of the data across runs and the run with the highest probability, K=2 had a probability of 1.0 (all other K probabilities were negligible). There was a clear correspondence between these two cluster in our unsupervised analysis and the northern/central European and TSI clusters in the supervised ADMIXTURE analysis (Pearson's correlation coefficient 0.86 and 0.74, p-value $7.17 \times 10^{-13}$, $1.05 \times 10^{-7}$) (Fig. S27) that is also correlated with normal versus elongated skulls and makes up the bulk of the ancestry in that analysis. Given how diverse the rest of the ancestry inferred from the supervised analysis was, it is perhaps unsurprising that further clear clustering could not be identified.

## SI11. Population Assignment Analysis (PAA)

*Krishna Veeramah*

In order to obtain a more precise estimate of the modern population most closely resembling our ancient samples we applied the following likelihood framework to our off-target ancient sample data matching the HellBus reference dataset.

For every HellBus reference population, *k*, with at least 10 individuals and for which a geographic latitude and longitude could be assigned we estimated for every SNP, *i*, the allele frequency of an arbitrary allele ($q_{ik}$) for a randomly drawn set of 20 chromosomes (so sample sizes were equal across populations). Then, for each ancient sample we determined the most likely population of origin by estimating the log-likelihood of observing the pseudo-haploid call, $D_i$, given a particular reference population for each SNP position, which is simply the log of $q_{ik}$ for the observed allele, and summing across loci. In order to account for a reference population being fixed for the allele not observed in a particular ancient sample (which may happen because of either the low sample size of the reference population or a sequencing error for the ancient sample), we allowed an 0.1% error rate, *e*, such that the log likelihood for each SNP used was:

$$lnLL( k \mid D_i) = ln[q_{ik} \times (1-e)) + ((1-q_{ik}) \times e)] \quad ,$$

where $q_{ik}$ is the frequency of the observed allele, $D_i$. However, the results were robust to different choices of *e* (both smaller and larger). The most likely reference population is shown in Table S35.

In order to obtain an estimate of uncertainty in our most likely reference population and take into account correlation amongst neighboring SNPs, we performed 100 bootstrap iterations, where for each iteration we resampled with replacement 5 Mb non-overlapping windows of SNPs from across the genome. For each bootstrap iteration we noted the reference population with the highest likelihood using the above expression and scored the total number of times each population obtained the highest log likelihood across the 100 iterations. We visualized this uncertainty on a geographic map of Eurasia using a 1,000 x 1,000 rasterized grid (longitude limits -20-140, latitude limits 30-62) using interpolation over the surface of the grid via a thin plate spline regression with lambda set to 0.0001 based on the score for each reference population considered using the R package *fields*. We performed both individual sample interpolation, and grouped based interpolation within 5 phenotypic class (by summing scores across individuals within a group): males with normal skulls (Fig. 4A), females with normal skulls Fig. 4B, females with intermediate skulls (Fig. S28), females with elongated skulls (Fig. 4C) and non-Bavarians with elongated skulls (Fig. 4D).

All males and females with normal skulls had estimated origins with north and central Europe, apart from STR_300 and STR_502, with their most likely geographic origins being Greece and Turkey respectively (Fig. S29).

As they demonstrated substantial heterogeneity within southeast Europe and West Asia, individual maps for fully elongated skulls, as well as STR_310 (the only non-northern or central European with an intermediate skull) are shown in Fig. S30.

Finally, individual maps for our five non-Bavarian and the Anglo-Saxon samples are shown in Figs. S31 and S32, respectively.

## SI12. Outlier analysis

*Krishna Veeramah*

Given how both diverse and low the non-European ancestry inferred from the supervised ADMIXTURE analysis and population identity analysis was (for example only AED_1108 and VIM_2 appear to possess substantial East Asian ancestry, it is perhaps unsurprising that further clear clustering could not be identified in the unsupervised STRUCTURE analysis). To determine whether we could identify outliers from the ancient samples without the use of a reference panel we conducted the following analysis on the 1 kb haplotype data for all 38 ancient samples. We first treated these 38 samples as a single population and determined for each locus,k, the haplotype frequencies ($p_{1k}$,$p_{2k}$...$p_{ik}$). We then determined the likelihood that each of these 38 samples in turn belonged to this single population by estimating the probability of observing a particular pair of haplotypes in the target individual given the population haplotype frequencies, which is simply $p_{ik}^2$ for a homozygous locus and $2p_{ik}p_{jk}$ for a heterozygote locus. Assuming each locus is independent, we then summed the log probability across loci to determine the likelihood for the individual.

As it is not clear how to form a null hypothesis to determine whether a particular likelihood is indicative of being an outlier or not, we also used the population haplotype frequencies to generate 1,000 random individuals from the population and estimated their likelihood to provide us with a null distribution to estimate an empirical probability that one our target individual came from our overall population or not (Table S36). Based on this analysis, VIM_2, PR_10, KER_1, AED_1108 and STR_502 are all highly significantly different from our null distribution, and consistent with our supervised ADMIXTURE analysis that these particular samples have a very different ancestry than the rest of the ancient individuals, who primarily resemble modern Europeans (either NC_Europe, TSI or in the case of FN_2, IBS).

## SI13. Private East Asian Haplotypes

*Krishna Veeramah*

In order to examine how robust the inferred East Asian ancestry observed in AED_1108 was based on the ADMIXTURE analysis, we scored the number of 1 kb haplotypes found in the 5 Mb neutral loci found in our ancient and SGDP samples that were private to EAS individuals from the 1000 Genomes project when compared to CEU, TSI, IBS, FIN and SAS individuals. 12,197 haplotypes were unique to EAS. VIM_2, AED_1108, PR_4, STR_328, PR_10 and KER_1 had 14, 12, 10, 7, 7, and 6 of these haplotypes respectively, within the range found Central Asians, and substantial more than any other Bavarian samples (Fig. S33). ALH_3 had 4 haplotypes, consistent with the very small amount of inferred EAS ancestry, STR_248 and STR_502 had 3 haplotypes, while the remaining had 2 EAS haplotypes or less. Other than 2

Russian samples (with 5 and 3 EAS haplotypes), no modern European samples had more than 2 EAS haplotypes either.

## SI14. $F_{IS}$ and $F_{ST}$ analysis

*Krishna Veeramah*

We took advantage of the presence of the same regions being called across all ancient samples to calculate classic population genetic parameters, Wright's F-statistics within population $F_{IS}$ and population pairwise $F_{ST}$ (141). As well as the 1000 Genomes, GoNL and Turkish populations we grouped our ancient samples into male normal skulls, female normal skulls, female intermediate skulls and females elongated skulls (Fig. S34). While it is arguable that these four groupings do not reflect natural randomly sampled populations for which Wright's parameters were originally intended for, the estimation of these parameters are still informative with regard to the structure of these groupings. Phased 1 kb haplotypes were used for this analysis as above. We used the $F_{ST}$ estimator introduced by Bhatia et al. (142) as this has been shown to be the most robust to differences in sample size. The ratio of averages of the numerator (variance of allele frequencies between populations) and denominator (total variance of allele frequencies in the ancestral population) was used to combine results across multiple loci. Standard errors and 95% CI's were constructed by performing 100 bootstraps for individual loci. As would be expected $F_{ST}$ was smallest between populations within continents. In this regards the four ancient populations were mostly in line with $F_{ST}$ expected for modern European samples compared to non-European samples (Table S37).

However, there was also clear differences when comparing the four ancient populations to modern European populations and Turkey. Ignoring the intermediate population because of their low sample size, both non-deformed males and females were genetically most similar to CEU and GoNL, while the deformed female population was closest to TSI, consistent with the substantial southern ancestry we have observed in previous analysis. They were also noticeably closer modern Turkish populations. In addition non-deformed females were closer to IBS, TSI and Turkey and more distant to FIN, consistent with the increased southern European ancestry observed in STR_502 and STR_300.

$F_{IS}$ was calculated for each individual population (Fig. S35) using the estimator from Nei 1978 (143) that corrects for small sample size. Standard errors and 95% CI's were constructed by performing 1000 bootstraps for individual loci. The male non-deformed population had levels consistent with other modern populations and with CIs that overlapped 0. The female deformed population was amongst a set of elevated $F_{IS}$ estimates. While this could be the result of inbreeding (an excess of homozygotes), the more likely explanation is that these values reflect the increased population structure of this group with individuals having various origins based on our other analyses. Interestingly, the female non-deformed population also had increased $F_{IS}$ values. While STR_300 and STR_502 likely contribute to this result, the values are still elevated even when these two samples are removed from the analysis (f_non-def_rm), suggesting that the female non-deformed population is more structured than the male population, though the 95% CIs now overlap 0.

## SI15. Inference based on the variation in allele frequency spectra

*Krishna Veeramah*

### Constructing a joint allele frequency spectrum

In order to construct a robust allele frequency spectrum (AFS) from our medium to high coverage Dutch genomes (for which we do not care about haplotype phase), single-sample called vcf files for all 738 GoNL individuals (246 trios, 4 trios were corrupted during data transfer and thus discarded) for the 5 Mb regions were grouped into trios, phased by transmission using GATK (62) and then for each site the most likely allele from each parent (the highest homozygous genotype likelihood) in a trio combined to create an AFS across the 5 Mb region for 246 pseudo-diploid individuals. In this way we should dramatically reduce false negative and positives, as we do not require a heterozygote genotype quality threshold, we only need to correctly identify one callable allele at each site in a diploid allele.

We constructed an AFS for the Medieval samples using all normal and intermediate skulled males and females with >20x (Fig. 1B) excluding the three samples with significant non-central/northern European ancestry (STR_310, STR_300, STR_502), giving us a total of 23 diploid individuals. For any given site we counted the reference and alternate alleles for all genotypes with at least 20x coverage depth of coverage. We discarded sites for which at least 20 individuals did not meet this criteria.

Site with evidence of three or more different nucleotides (i.e. trinucleotide SNPs) across both GoNL and Medieval datasets were excluded from any downstream analysis. The ancestral state for each site were taken from the 1000 Genomes ancestral fasta sequence for GR37 (144). ∂a∂i's (53) 'Spectrum.from_data_dict' function with a correction for ancestral state mis-identification (145) and projecting down to 40 chromosomes (from a maximum of 46) for the Medieval data was utilized to construct a unfolded joint 40 by 492 chromosome 2D-AFS for further downstream analysis based on 4,100,927 sites. Projecting sample size down this way provided the best balance of number of callable sites versus chromosome number.

### Inference Procedure for ∂a∂i

We used the diffusion approximation approach of ∂a∂i (53) to fit various models of demographic history to the 2D-AFS, assuming that the 492 GoNL samples would represent modern central and northern Europeans, and that the 23 Medieval samples would represent the same population 1,500 years (60 generations assuming 25 years per generation) prior. As in recent studies that have been able to successfully fit modern European sequence data using large sample sizes (146, 147) we applied an approach that involved only modelling very recent European population history rather than also modelling African and Asian populations simultaneously, with these inter-continental features accounted for via a fixed set of parameters in our underlying model.

In particular we attempted to fit the model of Gazave et al. (146) that included two population bottlenecks 4,720 and 720 generations ago, followed by recent exponential growth (Fig. S36). The free parameters in our model were as in Gao et al. (147), the effective populations size prior to recent growth $N_{e\_growth}$, the start time of recent growth, $t_{growth}$ and the current effective population size $N_{e\_present}$. These three parameters inherently describe a per generation exponential growth rate.

The model from Gazave was based on parameters estimated in (148), which itself assumed a human mutation rate, $\mu$, of 2.59 x $10^{-8}$ per site per generation (presented in the paper as $\mu$ =1.99 x $10^{-8}$ when excluding CpG sites, assuming a 1.3x correction). Therefore we rescaled all $N_e$ and time in generation, $t_g$, parameters to $\theta = 4N_e\mu$ and time in coalescent, $t_c = t_g/2N_e$, units respectively prior to analysis. This allowed us to infer parameters using more recent estimates of $\mu$ (1.2 x $10^{-8}$), as well as ultimately estimate $\mu$ calibrating using our ancient time point (see below). All analyses were performed with and without transitions, the class most prone to post-mortem damage, with $\theta$ and $\mu$ scaled assuming a *titv* ratio of 2.1.

As the time of the start of growth in Europe has previously been estimated to occur well before the age of our Medieval sample (~200 generations), we limit the timing of growth to be between the interval of the final bottleneck 720 generation ago (or equivalent coalescent scaling of 0.031) to 60 generations ago, which we rescale as 0-1 (0 being immediately following the bottleneck). We explore both $\theta_{growth}/\theta_{B2}$ and $\theta_{present}/\theta_{growth}$ on $\log_{10}$ scale.

Free parameters were fitted using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) optimizer. Because of the large size of the AFS we used a two-tier process to fit free parameters for any given model. In an initial run we set the grid sizes for extrapolating the approximate solution to the partial differential equation from Gutenkunst et al. (53) to 50, 60 and 70. This optimization was performed five times from random start points in parameter space. This allowed us to identify an appropriate start point when using a larger grid (which is much more computationally intensive). Parameters from the run with the highest likelihood were then used as a starting point for a final run with grid sizes of 500, 750, 1000 (i.e. a fine-tune step). A fixed value of $\theta$ was utilized in all analyses when fitting the 2D-AFS, and likelihoods were calculated using the Poisson count approach. Parameter bounds are shown in Table S38. Confidence intervals for all parameters were estimated by applying the entire inference framework on 100 individual bootstrapped datasets based on sampling data for individual loci with replacement.

Usually fitting the AFS of modern populations does not depend on a specific $\mu$, as units are estimated in units of $\theta$ and $t_c$. $\mu$ is only required to convert these units into time in years/generations and diploid population size. However, in order to include the AFS in the inference from a past population that is separated by a specific number of years from the present day population, the use of a specific $\mu$ becomes necessary in order to convert this time into coalescent units. Therefore, we performed our ∂a∂i analysis assuming both the original rate from Keinan et al. (148) which essentially captures the phylogenetic-based estimate of $\mu$ by Nachman and Crowell (149) of 2.5 x $10^{-8}$, as well as the more recent pedigree-based estimates derived from whole genome sequencing of trios and quartets (150–152) of ~1 x $10^{-8}$ (we specifically utilize the 1.2 x $10^{-8}$ estimate of Kong et al. (152)). However, the presence of population samples from two time points and the very large samples size of the modern

population also presents the possibility to estimate $\mu$ as an additional free-parameter along with the other demographic parameters, as the two sampled time points provide a specific per generation interval within which a certain number of mutations must arise given the underlying demographic model. As the addition of this extra parameter increases the complexity of the likelihood space that must be traversed, our BFGS optimization includes two additional pre-runs using the smaller 50, 60 and 70 grid combination in which each run is initiated to the best parameter estimates when assuming $\mu$ is fixed at $1.2 \times 10^{-8}$ or $2.59 \times 10^{-8}$, as well as the original five iterations based on random draws from the space. The parameter is explored on a $\log_{10}$ scale.

**Results**

Estimates for the growth rate (2.8%) and time of growth (~3,000 years) in the European population assuming the original phylogenetic estimate of $\mu$ =~$2.5 \times 10^{-8}$ (149) were consistent with previous estimates that utilized AFS data (146, 147, 153, 154) (Table S39). While these it did not greatly affect the resulting demographic parameter estimates (growth rate = 1.8% and time of growth =~5,000) years, higher log likelihoods were obtained using the pedigree rate of $\mu = 1.2 \times 10^{-8}$ (152) both with and without transitions, though the time of growth. When $\mu$ was allowed to freely vary we obtained an estimate of $\mu$ of $1.14 \times 10^{-8}$ (95% CI: $6.51 \times 10^{-8}$ - $1.53 \times 10^{-8}$) much more in line with the pedigree-based estimate. Even when the starting $\mu$ in the optimization procedure began at the higher phylogenetic rate, it tended towards the lower rate pedigree rate. Finally, the lower pedigree rate was also obtained when excluding transitions (i.e. potential damaged bases) $1.1 \times 10^{-8}$ (95% CI: $4.62 \times 10^{-9}$ - $3.46 \times 10^{-8}$), albeit with wider confidence intervals that included the phylogenetic rate. Utilizing a generation time of 30 years only moderately changes our results, with an estimate $\mu = 1.29 \times 10^{-8}$ and $1.34 \times 10^{-8}$ with and without transitions respectively. Our estimates of $\mu$ calibrated using our ancient AFS are not only in line with recent pedigree-based rates, but also an estimate ($1.3 \times 10^{-8}$) based on large individuals sequenced at ~200 drug metabolizing enzymes (154), while being closer to an estimate ($1.6 \times 10^{-8}$) calibrated using recombination rates (155) than the phylogenetic rate.

Our best-fit parameters under this $\mu$ are consistent with a recent exponential growth rate of 1.81% (95% CI: 1.00-2.36%), with a current $N_e$ of 1.4 million (750K-2.8 million) with growth beginning 5.924 years ago (95% CI: 4,402-10,402) from an initial $N_e$ of 19,694 (95% CI: 14,445-34,458). The $N_e$ at the time of sampling of the Medieval samples (1,500 years ago) would have been ~500,000. Both the individual and 2D-AFS for our data were well fit by the model (Figs. S37-S41), in particular when compared to a model with constant effective population size (a difference of ~3000 log likelihood units) and even a model with no fixed bottlenecks (a difference of ~80 log likelihood units), the latter of which had trouble fitting the large singleton and doubleton classes of the AFS. Estimates were also highly consistent when excluding transitions suggesting our results are fairly robust.

## SI16. Comparing patterns of allele and haplotype sharing between ancient and modern samples

*Lucy van Dorp, Saioa López, Garrett Hellenthal*

### Introduction

The aim of this section is to apply additional techniques to explore the ancestry of sampled ancient DNA individuals in relation to modern populations. In particular, we compare haplotype and allele sharing patterns to identify which sampled groups are most related genetically, reflecting shared common ancestry relative to other sampled groups.

### Methods

We merged our aDNA samples with modern reference panels, using three different datasets which consider either only the high coverage neutralome or high coverage whole genome data.

As previously described in Supplementary Section SI6, we merged our 41 ancient genomes together with 5-7$^{th}$ century British genomes (70) in a modern dataset merge we name "HellBus". In particular we merged our aDNA samples with imputed data (as described in Section SI6) from 1,582 European and Asian individuals from Hellenthal *et al*. (71), 279 publicly available whole genomes from the Simons Genome Diversity Panel (68) together with 504 East Asian genomes from the 1000 Genomes project (67). This dataset contained 8,692 neutralome SNPs. We also utilized the 1000 Genomes, SGDP, Turkish, and GoNL dataset described in Section SI6 which also considers only the neutralome but without imputation of modern genomes. Individual 1 kb loci in this dataset were phased using the program PHASE (20) with 1000 Genomes data used to supervise the phasing. This dataset comprised 264 Eurasian SGDP, 16 Turkish and 100 Dutch GoNL genomes and in this section we refer to this merge as "Phase_loci".

As discussed in Section SI5, PR_4 and STR_328 were excluded from all analyses based on diploid calls due to their low coverage. STR_355 and STR_491 were identified as siblings but included in the analyses presented in this section given individuals were analyzed independently.

In addition we performed a further merge that utilized only our novel samples which were shot-gun sampled to high coverage (>10X) as described in Section SI2. Our new aDNA samples ALH_10, ALH_1 and FN_2 were combined with genotype data from all individuals in Busby et al. (156) and Hellenthal et al. (71) genotyped on the Illumina Infinium series chip. Throughout the labels used in these original publications are used to refer to modern groups. We term this merge "HellBus_SG", which comprised 402,586 non-missing autosomal SNPs, across 2,457 individuals. The HellBus_SG dataset was phased using SHAPEIT v2 (157) with build 37 genetic maps.

### Inferring Allele Sharing Profiles

We applied the CHROMOPAINTER "unlinked" model ("*–u"* switch) to each of the datasets (HellBus, Phase_loci and HellBus_SG), as described in Lawson *et al.* (158). Briefly, this program compares the DNA patterns in a "recipient" chromosome to that in a collection of "donor" chromosomes. Under the "unlinked" model, CHROMOPAINTER calculates, separately for each SNP, the probability that a "recipient" chromosome is most closely related to a particular "donor" individual given their genotype data at that SNP. Throughout we use modern populations as donors, while analyzing each ancient and modern individual independently as a recipient. We note that when running CHROMOPAINTER with a modern individual as a recipient, that modern individual was removed from the set of donors.

For each recipient *r*, we define $y_d^r$ to be the total *amount* (measured in SNP count) of DNA for which individual *r* is inferred to be most closely related to a donor chromosome from group *d* (i.e. the ".chunkcounts.out" output from the unlinked CHROMOPAINTER model). We define $f_d^r$ to be the total *proportion* of DNA for which individual *r* is inferred to be most closely related to a donor chromosome from group *d*. For the HellBus and Phase_loci datasets, $y_d^r$ and $f_d^r$ are formed from genetic information exclusively across the neutralome. Here *r* can refer to a single individual (e.g. for an ancient sample) or can represent an average across individuals with the same group label (e.g. for modern groups). We let each distinct population label represent a different donor group *d,* leading to *D*=231 total donor groups for the HellBus dataset, *D*= 98 total donor groups for the Phase_loci dataset and *D*=161 total donor groups for the HellBus_SG dataset.

We refer to the $f_d^r$ as "allele sharing profiles" and the $y_d^r$ as the "unscaled allele sharing profiles", which can be used to assess the relative proportion or amount, respectively, of DNA that recipient *r* shares with modern groups from different continental regions. Uncertainty in these profiles was assessed by performing jack-knifing dropping one chromosome at a time (159).

**Inferring Haplotype Sharing Profiles**

Additionally, for the Phase_loci and HellBus_SG datasets we applied a "linked" CHROMOPAINTER analysis that can utilize the rich information provided by considering haplotype information (158). Under this "linked" model, CHROMOPAINTER calculates the genome-wide amount of DNA (in cM) for which a "recipient" chromosome is most closely related to a particular "donor" individual given their joint genotype data across all SNPs, thus explicitly modeling linkage disequilibrium information to inform the inference. As before, we use modern individuals as donors and modern and ancient individuals as recipients, again excluding a modern individual as a donor when analyzing that modern individual as a recipient. Also, as before, though now defined using "linked" CHROMOPAINTER, for each recipient *r*, we define $y_d^r$ and $f_d^r$ to be the total amount (in cM) and proportion, respectively, of DNA for which individual *r* is inferred to be most closely related to a donor chromosome from group *d* (i.e. the "chunklengths.out" output from CHROMOPAINTER linked model). We refer to these $f_d^r$ and $y_d^r$ under the linked model as "haplotype sharing profiles" and "unscaled haplotype sharing profiles" respectively.

Following Lawson *et al.* (158), we estimated the CHROMOPAINTER switch rate (N, "-n" switch) and mutation rate (M, "-M" switch) on every $10^{th}$ individual on chromosomes 1, 4, 15 and 22 using E-M with 10 iterations (i.e. "-i 10 -in -iM") in the HellBus_SG dataset. We weight-averaged these inferred values across these four chromosomes, weighting by the number of SNPs per chromosome. This resulted in estimates of N =231.0991 and M= 0.0006024927. For the Phase_loci dataset we used default values of switch rate and mutation rate.

## Principal Components Analysis (PCA) of inferred allele sharing and haplotype sharing profiles

We perform principal components analysis (PCA) separately on the matrices of unscaled allele sharing and haplotype sharing profiles inferred using CHROMOPAINTER generated as described above. PCA was implemented using the *prcomp* function in R.

## Measuring differences in inferred allele sharing and haplotype sharing profiles (TVD)

In order to quantify differences in the inferred haplotype or allele profiles between each of the ancient samples, including those with and without a deformed skull phenotype, we apply the total variation distance (*TVD*) measure as originally applied in Leslie *et al*. (160). As before, let $f_d^r$ be the genome-wide proportion of DNA that recipient individual (or group) *r* copies from each of donor groups $d \in [1, \ldots, D]$. We then calculated the *TVD* between any 2 recipients *X* and *Y* as:

$$TVD_{XY} = 0.5\sum_{d=1}^{D}\left|\left(f_d^X - f_d^Y\right)\right|.$$

## Discussion

The PCA of our allele sharing profile (Fig. S42) demonstrates that these profiles clearly resolve large continental groupings worldwide despite the low SNP count of the neutralome, suggesting this dataset provides some power to resolve regional population structure. Within our Baiuvarii samples, a subset: AED_125, STR_328 (low coverage), STR_502, AED_204 and AED_92 tend to cluster with present-day populations from the Middle East and West Asia, whilst AED_1108 also clusters with Middle Eastern and West Asian populations but is positioned more closely to modern populations from Central and East Asia across component 1. The majority of our other Baiuvarii samples tend to cluster more closely with modern populations from South and North West Europe, with little resolution to distinguish these modern groupings using this low-SNP dataset. The Crimea individual (KER_1), the Sarmatian individuals (PR_4 and PR_10) and the Gepid individual (VIM_2) all appear genetically distinct from the core group, though only the latter two exhibit deformed skull phenotypes.

Consistent with Fig. S42, the largest differences we infer (highest TVD) between any 2 ancient Bavarian individuals is found when comparing AED_1108 to other individuals from

the archeological sites of Altenerding-Klettham (AED) and Straubing-Bajuwarenstraße (STR) (Fig. S43). However, pairwise TVD values among individuals with a deformed skull phenotype were not notably higher or lower than the pairwise TVD values among individuals without a deformed skull phenotype.

To explore more closely the most marked genetic differences between AED_1108 compared to other sampled individuals, we evaluated the proportions of DNA matched to different regional groupings of modern populations in the allele-sharing profiles inferred in the HellBus dataset and haploptype sharing profiles inferred in the Phase_loci dataset (Figs. S44-S46). We note that the differences we infer across ancient individuals are subtle and will be due in part to differences in sample size of the regional groupings. We also note that there are some inconsistencies between results for the HellBus and Phase_loci dataset likely relating to both differences in the type of data analyzed as well as the differences in modern donor groups present in these datasets. Therefore the inferred differences should not be interpreted as a robust measure of the proportion of recent shared ancestry with a particular modern group. However, given each individual is analyzed in the same way, relative patterns of donor contributions can be informative on general trends in relatedness to modern populations.

Notably, Figs. S44-S46 highlight that the Baiuvarii individual with the largest amount of ancestry matching to modern East Asian groups is AED_1108 across datasets. This is consistent with this individual sharing relatively more East Asian-like ancestry compared to our other sampled ancient individuals, and is supported by the observation that this individual carries the derived Asian *SLC24A5* haplotype (Section SI7). Interesting this is also the individual with the strongest 'deformed' phenotype, Fig. 1b main text. Most of the differences in inferred ancestry between AED_1108 and AED_1119, which is from the same archeological site of Altenerding without a deformed phenotype, appear to relate to differences in East Asian-like ancestry (Fig. S47). The Sarmatian individual PR_10 and Serbian Gepid VIM_2 also carry the derived *SLC24A5* haplotype and are amongst our individuals with the highest amount of DNA matching to East Asia. Indeed, AED_1108, PR_10 and VIM_2 are the individuals with not only the highest East Asian ancestry but also the lowest Middle Eastern and North West European contributions. Interestingly, NW_54 from Burgweinting is amongst the individuals with the highest Central Asian ancestry, consistent with mitochondrial data showing this individual carries a haplogroup from the C-clade found commonly in modern-day Central Asia.

We also note that, despite not exhibiting a deformed phenotype, ALH_3 and AED_92 are inferred to carry the derived East Asian EDAR gene, whilst all other individuals analyzed were homozygous for the non-derived gene (Section SI7). Consistently, these two individuals are also amongst the individuals with the highest levels of allele sharing with East Asian populations (Figs. S44-S46).

As noted, KER_1 is our most easterly sampled individual whilst also pre-dating AED_1108 and VIM_2. However, we do not detect a strong signal of East Asian ancestry in this individual (Figs. S44-S46).

Amongst other individuals with deformed skulls, there are a mixture of ancestries with some (e.g. AED_513, STR_228 and AEH_1) appearing more central to eastern European, others slightly more Central South Asian (AED_125) and others more East Asian (AED_1108). This is consistent with a diversity of ancestries amongst individuals adopting the same cultural practice of skull deformation. Overall with our sample sizes, we found no evidence for a higher amount of matching to modern East Asian groups in the 10 deformed skull individuals relative to 29 individuals without deformed skulls when analyzing the neutralome (Wilcoxon rank-sum test two-sided p-value = 0.84).

We also more closely considered the ancestry of those samples for which we had high coverage shotgun data, in particular ALH_10 (12.17X), ALH_1 (13.27X) and FN_2 (11.08X), using our HellBus_SG dataset. The high SNP overlap between these individuals and our world-wide reference populations offers the opportunity to utilize genome-wide haplotype information in these samples, potentially allowing more refined ancestry inference.

As before, but now using the haplotype sharing profiles we evaluated the proportion of DNA matching to relevant regional groupings (Figs. S48-S49). In particular we considered how the haplotype sharing patterns inferred for our two samples from Altheim (ALH_1, ALH_10) differed from that of the Roman soldier sampled from Munich (FN_2). We find some significant differences in regional ancestry comparing the Altheim individuals to FN2. For example, FN2 exhibits substantially more West Asian-like ancestry than ALH_1 and ALH_10 (Fig. S48), also consistent with Fig. S44-S46, and substantially more Middle Eastern-like and Southern European-like ancestry. Conversely ALH_1 and ALH_10 match significantly more to modern populations from northwest and east Europe. We do not, however, always replicate these results when analyzing the neutralome (Figs. S44-S46), especially when considering allele-matching, perhaps illustrating the lack of precision with these techniques when analyzing a relatively small proportion of the genome and/or ignoring haplotype information.

## SI Appendix references


1. Brather S (2009) Memoria und Repräsentation. Frühmittelalterliche Bestattungen zwischen Erinnerung und Erwartung. *Historia Archaeologica. Festschrift Für Heiko Steuer Zum 70. Geburtstag (Reallexikon Der Germanischen Altertumskunde – Ergänzungsbände, Band 70)*, ed Brather S (De Gruyter, Berlin), pp 247–284.

2. Koch U (2001) *Das alamannisch-fränkische Gräberfeld bei Pleidelsheim.(Forschungen u. Berichte z. Vor-u. Frühgeschichte Baden-Württemberg 60)* (Konrad Theiss, Stuttgart).

3. Hines J, Bayliss A (2013) *Anglo-Saxon Graves and Grave Goods of the 6th and 7th Centuries AD: A chronological Framework* (The Society of Medieval Archaeology, London).

4. Koch A (1998) *Bügelfibeln der Merowingerzeit im westlichen Frankenreich Teil 1* (Römisch-Germanisches Zentralmuseum, Mainz).

5. Haas-Gebhard B, von Looz G (2009) Neue Beobachtungen an der Bügelfibel aus Altenerding Grab 512. *Archäologisches Korrespondenzblatt* 39(4):579–588.

6. Grupe G, McGlynn G, Harbeck M (2015) *Prähistorische Anthropologie* (Springer, Berlin).

7. Trautmann B, et al. (2017) Eine Reevaluation artifiziell deformierter Schädel des Frühen Mittelalters aus Bayern. *Archäologisches Korrespondenzblatt, Band 47, Nr 2*.

8. Clark JL, et al. (2007) Identifying artificially deformed crania. *Int J Osteoarchaeol* 17(6):596–607.

9. Ginzburg VV, Žirov EV (1949) Antropologicseszkije materialii iz Kenkolszkogo katakombnogo moglinika v doline r. Talasz Kirgizszkoj SzSzR. *Szbornik Muzeja antropologii I etnografii* 10:213–265.

10. Molnár M, János I, Szűcs L, Szathmáry L (2014) Artificially deformed crania from the Hun-Germanic Period (5th-6th century ad) in northeastern Hungary: historical and morphological analysis. *Neurosurg Focus* 36(4):E1.

11. O'Brien TG, Stanley AM (2013) Boards and Cords: Discriminating Types of Artificial Cranial Deformation in Prehispanic South Central Andean Populations. *Int J Osteoarchaeol* 23(4):459–470.

12. Cocilovo JA, Varela HH, O'Brien TG (2011) Effects of artificial deformation on cranial morphogenesis in the south central Andes. *Int J Osteoarchaeol* 21(3):300–312.

13. Sage W (1984) *Das Reihengräberfeld von Altenerding in Oberbayern I* (Mann Verlag, Berlin).

14. Pleterski A (2003) *Altenerding in Oberbayern II* (scrîpvaz, Berlin, Bamberg, Ljubljana).

15. Geisler H (1998) *Das frühbairische Gräberfeld Straubing – Bajuwarenstrasse I* (Verlag Marie Leidorf, Rahden/Westf).

16. Sebrich J (2017) *Das spätantik-frühmittelalterliche Gräberfeld Altheim-Andreasweg, Lkr. Landshut/Niederbayern* (Dissertation Ludwig-Maximilians-Universität München).

17. Koch U (1968) *Die Grabfunde der Merowingerzeit aus dem Donautal um Regensburg* (de Gruyter).

18. Codreanu-Windauer S, Schleuder R (2013) Gräber des 5. Jahrhunderts von Regensburg-Burgweinting aus archäologischer und anthropologischer Sicht. *Bericht der Bayerischen Bodendenkmalpflege* 54:351–362.

19. Codreanu-Windauer S, Harbeck M (2016) Neue Untersuchungen zu Gräbern des 5. Jahrhunderts: Der Fall Burgweinting. *Wandel Durch Migration? 26. Internationales Symposium „Grundprobleme Der Frühgeschichtlichen Entwicklung Im Mittleren Donauraum“*, ed Geisler H (Verlag Dr. Faustus, Büchenbach), pp 243–260.

20. Bierbrauer V, Osterhaus U, Gerhardt K (1973) Ein Frauengrab des frühen 6. Jahrhunderts aus Alteglofsheim, Ldkr. Regensburg. *Bayerische Vorgeschichtsblätter* 38:94–100.

21. Unterländer M, et al. (2017) Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe. *Nat Commun* 8:14615.

22. Malashev VY, Yablonsky LT (2004) Early Nomads in the Southern Foothills of the Urals Based on Materials from the Pokrovka Burial-Ground. *Ancient Civilizations from Scythia to Siberia* 10(3):259–292.

23. Davis-Kimball J, Yablonky LT, Demkin VA (1995) *Kurgans on the Left Bank of the Ilek: Excavations at Pokrovka, 1990-1992* (Zinat Press, Berkeley).

24. Mikić Z (1999) Die Gepiden von Viminacium in der Völkerwanderungszeit. Anthropologischer Beitrag. *Anthropol Anz* 57(3):257–268.

25. Tolstikov VP (2002) Pantikapaion. Ein archäologisches Porträt der Hauptstadt des Kimmerischen Bosporus. *Das Bosporanische Reich. Der Nordosten Des Schwarzen Meeres in Der Antike*, eds Fornasier J, Böttger B (Verlag Phillip von Zabern, Mainz), pp 39–58.

26. Rosen K (2016) *Attila. Der Schrecken der Welt* (C.H. Beck, München).

27. Altheim F, Haussig HW (1958) *Die Hunnen in Osteuropa* (Grimm, Baden-Baden).

28. Shchukin MB, Kazanski M, Sharov O (2006) *Des les goths aux huns: le nord de la mer Noire au Bas-Empire et a l'époque des grandes migrations* (British Archaeological Reports, Oxford).

29. Fless F (2002) *Rotfigurige Keramik als Handelsware: Erwerb und Gebrauch attischer Vasen im mediterranen und pontischen Raum während des 4. Jhs. v. Chr* (Verlag Marie Leidorf, Rahden/Westf).

30. Fless F, Lorenz A (2005) Die Nekropolen Pantikapaions im 4. Jh. v. Chr. *Bilder Und Objekte Als Träger Kultureller Identität Und Interkultureller Kommunikation Im Schwarzmeergebiet, Kolloquium in Zschortau/ Sachsen Vom 13.-15. Februar 2003*, eds Fless F, Treister M (Verlag Marie Leidorf, Rahden/Westf), pp 17–26.

31. Derewitzky A, Pavlowsky A, von Stern E (1897/1898) *Das Museum der Kaiserlich Odessaer Gesellschaft für Geschichte u. Altertumskunde. Lieferung 1 und 2 Terracotten* (Frankfurt a. Main, Odessa).

32. Kobylina MM (1961) *Terrakotowie statuėtki Pantikapeja i Phanagorii* (Moskau).

33. Kobylina MM (1974) *Terrakotovye statuėtki Teil Č. 3 Pantikapej* (Moskau).

34. Rostovtsev M (1914) *Antichnaya dekorativnaya Zhivopis na yuge Rossii* (St. Petersburg).

35. Schramm PE (1954) *Herrschaftszeichen und Staatssymbolik. Beiträge zu ihrer Geschichte vom dritten bis zum sechzehnten Jahrhundert* (Hiersemann, Stuttgart).

36. Werner J (1956) *Beiträge zur Archäologie des Attila-Reiches* (Philosophical-Historical Class, München).

37. Roth H (1979) *Kunst der Völkerwanderungszeit* (Propyläen Verlag, Berlin).

38. Damm IG (1988) *Goldschmiedearbeiten der Völkerwanderungszeit aus dem nördlichen Schwarzmeergebiet* (Gebr. Mann, Berlin).

39. Schmauder M (2009) *Die Hunnen. Ein Reitervolk in Europa* (Wissenschaftliche Buchgesellschaft, Darmstadt).

40. Kovrig I (1985) Das Diadem von Csorna. *Folia Archaeologica* 36:107–145.

41. Bóna I (1991) *Das Hunnenreich* (Corvina, Budapest).

42. Anke B (1998) *Studien zur reiternomadischen Kultur des 4. und 5.Jahrhunderts* (Beier & Beran, Weissbach).

43. Hardt M (2003) The Nomad's Greed for Gold: From the Fall of the Burgundians to the Avar Treasure. *Construction of Communities in the Early Middle Ages: Texts, Resources and Artefacts*, eds Corradini R, Diesenberger M, Reimitz H (BRILL, Amsterdam), pp 95–107.

44. Hardt M (2004) *Gold und Herrschaft. Die Schätze europäischer Könige und Fürsten im ersten Jahrtausend. Europa im Mittelalter* (Akademie Verlag, Berlin).

45. Anke B (2000) Hunnen §6. Archäologisches. *Reallexikon der Germanischen Altertumskunde* 15:256–261.

46. Herrmann B, Grupe G, Hummel S, Piepenbrink H, Schutkowski H (1990) *Prähistorische Anthropologie. Leitfaden der Feld- und Labormethoden* (Springer, Berlin).

47. Scheu A, et al. (2015) The genetic prehistory of domesticated cattle from their origin to the spread across Europe. *BMC Genet* 16:54.

48. Hofmanová Z, et al. (2016) Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc Natl Acad Sci U S A* 113(25):6886–6891.

49. Kircher M, Sawyer S, Meyer M (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* 40(1):e3.

50. Skoglund P, Storå J, Götherström A, Jakobsson M (2013) Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J Archaeol Sci* 40(12):4477–4482.

51. Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L (2013) mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29(13):1682–1684.

52. Gnirke A, et al. (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 27(2):182–189.

53. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* 5(10):e1000695.

54. Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genet* 9(10):e1003905.

55. Sell C (2017) Addressing challenges of ancient DNA sequence data obtained with next generation methods. PhD (JGU Mainz). Available at: http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:hebis:77-diss-1000012793.

56. Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A (2011) Bayesian inference of ancient human demography from individual genome sequences. *Nat Genet* 43(10):1031–1034.

57. Arbiza L, Zhong E, Keinan A (2012) NRE: a tool for exploring neutral loci in the human genome. *BMC Bioinformatics* 13:301.

58. McVicker G, Gordon D, Davis C, Green P (2009) Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet* 5(5):e1000471.

59. Broushaki F, et al. (2016) Early Neolithic genomes from the eastern Fertile Crescent. *Science* 353(6298):499–503.

60. Aronesty E (2011) ea-utils: Command-line tools for processing biological sequencing data. *Expression Analysis, Durham, NC*.

61. Botigué LR, et al. (2017) Ancient European dog genomes reveal continuity since the Early Neolithic. *Nat Commun* 8:ncomms16082.

62. DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491–498.

63. Das S, et al. (2016) Next-generation genotype imputation service and methods. *Nat Genet* 48(10):1284–1287.

64. Novembre J, et al. (2008) Genes mirror geography within Europe. *Nature* 456(7218):98–101.

65. Nelson MR, et al. (2008) The Population Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics research. *Am J Hum Genet* 83(3):347–358.

66. Loh P-R, et al. (2016) Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* 48(11):1443–1448.

67. 1000 Genomes Project Consortium, et al. (2015) A global reference for human genetic variation. *Nature* 526(7571):68–74.

68. Mallick S, et al. (2016) The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* 538(7624):201–206.

69. Purcell S, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559–575.

70. Schiffels S, et al. (2016) Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *Nat Commun* 7:10408.

71. Hellenthal G, et al. (2014) A genetic atlas of human admixture history. *Science* 343(6172):747–751.

72. Alkan C, et al. (2014) Whole genome sequencing of Turkish genomes reveals functional private alleles and impact of genetic interactions with Europe, Asia and Africa. *BMC Genomics* 15:963.

73. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589–595.

74. Genome of the Netherlands Consortium (2014) Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet* 46(8):818–825.

75. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68(4):978–989.

76. Walsh S, et al. (2014) Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. *Forensic Sci Int Genet* 9:150–161.

77. Spichenok O, et al. (2011) Prediction of eye and skin color in diverse populations using seven SNPs. *Forensic Sci Int Genet* 5(5):472–478.

78. Canfield VA, et al. (2013) Molecular phylogeography of a human autosomal skin color locus under natural selection. *G3* 3(11):2059–2067.

79. Lamason RL, et al. (2005) SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* 310(5755):1782–1786.

80. Beleza S, et al. (2013) Genetic architecture of skin and eye color in an African-European admixed population. *PLoS Genet* 9(3):e1003372.

81. Giardina E, et al. (2008) Haplotypes in SLC24A5 Gene as Ancestry Informative Markers in Different Populations. *Curr Genomics* 9(2):110–114.

82. Donnelly MP, et al. (2012) A global view of the OCA2-HERC2 region and pigmentation. *Hum Genet* 131(5):683–696.

83. Fujimoto A, et al. (2008) A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. *Hum Mol Genet* 17(6):835–843.

84. Kimura R, et al. (2009) A common variation in EDAR is a genetic determinant of shovel-shaped incisors. *Am J Hum Genet* 85(4):528–535.

85. Eiberg H, et al. (2008) 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. *Hum Genet* 123(2):177–187.

86. Magalon H, et al. (2008) Population genetic diversity of the NAT2 gene supports a role of acetylation in human adaptation to farming in Central Asia. *Eur J Hum Genet* 16(2):243–251.

87. Patin E, et al. (2006) Deciphering the ancient and complex evolutionary history of human arylamine N-acetyltransferase genes. *Am J Hum Genet* 78(3):423–436.

88. Kuznetsov IB, McDuffie M, Moslehi R (2009) A web server for inferring the human N-acetyltransferase-2 (NAT2) enzymatic phenotype from NAT2 genotype. *Bioinformatics* 25(9):1185–1186.

89. García-Closas M, et al. (2011) A single nucleotide polymorphism tags variation in the arylamine N-acetyltransferase 2 phenotype in populations of European background. *Pharmacogenet Genomics* 21(4):231–236.

90. Sabbagh A, Darlu P, Crouau-Roy B, Poloni ES (2011) Arylamine N-acetyltransferase 2 (NAT2) genetic diversity and traditional subsistence: a worldwide population survey. *PLoS One* 6(4):e18507.

91. Patillon B, et al. (2014) A Homogenizing Process of Selection Has Maintained an “Ultra-Slow” Acetylation NAT2 Variant in Humans. *Hum Biol* 86(3):185–214.

92. Martis S, et al. (2013) Multi-ethnic distribution of clinically relevant CYP2C genotypes and haplotypes. *Pharmacogenomics J* 13(4):369–377.

93. Janha RE, et al. (2014) Inactive alleles of cytochrome P450 2C19 may be positively selected in human evolution. *BMC Evol Biol* 14:71.

94. Gaedigk A, Sangkuhl K, Whirl-Carrillo M, Klein T, Leeder JS (2017) Prediction of CYP2D6 phenotype from genotype across world populations. *Genet Med* 19(1):69–76.

95. Schirmer M, et al. (2006) Genetic signature consistent with selection against the CYP3A4*1B allele in non-African populations. *Pharmacogenet Genomics* 16(1):59–71.

96. Bains RK, et al. (2013) Molecular diversity and population structure at the Cytochrome P450 3A5 gene in Africa. *BMC Genet* 14:34.

97. Strom CM, et al. (2012) Testing for variants in CYP2C19: population frequencies and testing experience in a clinical laboratory. *Genet Med* 14(1):95–100.

98. Thompson EE, et al. (2004) CYP3A variation and the evolution of salt-sensitivity variants. *Am J Hum Genet* 75(6):1059–1069.

99. Kroetz DL, et al. (2003) Sequence diversity and haplotype structure in the human ABCB1 (MDR1, multidrug resistance transporter) gene. *Pharmacogenetics* 13(8):481–494.

100. Wang H, et al. (2007) Comparative and evolutionary pharmacogenetics of ABCB1: complex signatures of positive selection on coding and regulatory regions. *Pharmacogenet Genomics* 17(8):667–678.

101. Nakajima T, et al. (2004) Natural selection and population history in the human angiotensinogen gene (AGT): 736 complete AGT sequences in chromosomes from around the world. *Am J Hum Genet* 74(5):898–916.

102. Balam-Ortiz E, et al. (2012) Hypercontrols in genotype-phenotype analysis reveal ancestral haplotypes associated with essential hypertension. *Hypertension* 59(4):847–853.

103. Oota H, et al. (2004) The evolution and population genetics of the ALDH2 locus: random genetic drift, selection, and low levels of recombination. *Ann Hum Genet* 68(Pt 2):93–109.

104. Peng Y, et al. (2010) The ADH1B Arg47His polymorphism in east Asian populations and expansion of rice domestication in history. *BMC Evol Biol* 10:15.

105. Li H, et al. (2011) Diversification of the ADH1B gene during expansion of modern humans. *Ann Hum Genet* 75(4):497–507.

106. Bersaglieri T, et al. (2004) Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet* 74(6):1111–1120.

107. Poulter M, et al. (2003) The causal element for the lactase persistence/non-persistence polymorphism is located in a 1 Mb region of linkage disequilibrium in Europeans. *Ann Hum Genet* 67(Pt 4):298–311.

108. Zhang G, Muglia LJ, Chakraborty R, Akey JM, Williams SM (2013) Signatures of natural selection on genetic variants affecting complex human traits. *Appl Transl Genom* 2:78–94.

109. Meigs JB, et al. (2008) Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med* 359(21):2208–2219.

110. Cornelis MC, et al. (2009) Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry. *Ann Intern Med* 150(8):541–550.

111. Hivert M-F, et al. (2011) 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* 60(4):1340–1348.

112. Helgason A, et al. (2007) Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. *Nat Genet* 39(2):218–225.

113. SIGMA Type 2 Diabetes Consortium, et al. (2014) Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature* 506(7486):97–101.

114. Hara K, et al. (2014) Genome-wide association study identifies three novel loci for type 2 diabetes. *Hum Mol Genet* 23(1):239–246.

115. Raj T, et al. (2013) Common risk alleles for inflammatory diseases are targets of recent positive selection. *Am J Hum Genet* 92(4):517–529.

116. Wiechmann I, Grupe G (2005) Detection of Yersinia pestis DNA in two early medieval skeletal finds from Aschheim (Upper Bavaria, 6th century A.D.). *Am J Phys Anthropol* 126(1):48–55.

117. Harbeck M, et al. (2013) Yersinia pestis DNA from Skeletal Remains from the 6th Century AD Reveals Insights into Justinianic Plague. *PLoS Pathog* 9(5):e1003349.

118. Feldman M, et al. (2016) A High-Coverage Yersinia pestis Genome from a Sixth-Century Justinianic Plague Victim. *Mol Biol Evol* 33(11):2911–2923.

119. Harper K (2015) Pandemics and passages to late antiquity: rethinking the plague of c.249–270 described by Cyprian. *Journal of Roman Archaeology* 28:223–260.

120. McCormick M (2015) Tracking mass death during the fall of Rome's empire (I). *Journal of Roman Archaeology* 28:325–357.

121. Rasmussen S, et al. (2015) Early divergent strains of Yersinia pestis in Eurasia 5,000 years ago. *Cell* 163(3):571–582.

122. Vianello D, et al. (2013) HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment. *Hum Mutat* 34(9):1189–1194.

123. Maca-Meyer N, González AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* 2:13.

124. Torroni A, Achilli A, Macaulay V, Richards M, Bandelt H-J (2006) Harvesting the fruit of the human mtDNA tree. *Trends Genet* 22(6):339–345.

125. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10(3):564–567.

126. Nei M (1987) *Molecular evolutionary genetics* (Columbia University Press, New York).

127. Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105(2):437–460.

128. Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139(1):457–462.

129. Pimenoff VN, et al. (2008) Northwest Siberian Khanty and Mansi in the junction of West and East Eurasian gene pools as revealed by uniparental markers. *Eur J Hum Genet* 16(10):1254–1264.

130. Roostalu U, et al. (2007) Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. *Mol Biol Evol* 24(2):436–448.

131. Derenko M, et al. (2007) Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *Am J Hum Genet* 81(5):1025–1041.

132. Torroni A, et al. (2001) A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* 69(4):844–852.

133. Loogväli E-L, et al. (2004) Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol Biol Evol* 21(11):2012–2021.

134. Tambets K, et al. (2004) The western and eastern roots of the Saami--the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. *Am J Hum Genet* 74(4):661–682.

135. Fu Q, et al. (2013) A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr Biol* 23(7):553–559.

136. Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15:356.

137. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2(12):e190.

138. Veeramah KR, Hammer MF (2014) The impact of whole-genome sequencing on the reconstruction of human population history. *Nat Rev Genet* 15(3):149–162.

139. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655–1664.

140. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–959.

141. Wright S (1951) THE GENETICAL STRUCTURE OF POPULATIONS. *Ann Eugen* 15(1):323–354.

142. Bhatia G, Patterson N, Sankararaman S, Price AL (2013) Estimating and interpreting FST: the impact of rare variants. *Genome Res* 23(9):1514–1521.

143. Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89(3):583–590.

144. 1000 Genomes Project Consortium, et al. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491(7422):56–65.

145. Hernandez RD, Williamson SH, Bustamante CD (2007) Context dependence, ancestral misidentification, and spurious signatures of natural selection. *Mol Biol Evol* 24(8):1792–1800.

146. Gazave E, et al. (2014) Neutral genomic regions refine models of recent rapid human population growth. *Proc Natl Acad Sci U S A* 111(2):757–762.

147. Gao F, Keinan A (2016) Inference of Super-exponential Human Population Growth via Efficient Computation of the Site Frequency Spectrum for Generalized Models. *Genetics* 202(1):235–245.

148. Keinan A, Mullikin JC, Patterson N, Reich D (2007) Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nat Genet* 39(10):1251–1255.

149. Nachman MW, Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics* 156(1):297–304.

150. Conrad DF, et al. (2011) Variation in genome-wide mutation rates within and between human families. *Nat Genet* 43(7):712–714.

151. Roach JC, et al. (2010) Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 328(5978):636–639.

152. Kong A, et al. (2012) Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 488(7412):471–475.

153. Tennessen JA, et al. (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337(6090):64–69.

154. Nelson MR, et al. (2012) An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* 337(6090):100–104.

155. Lipson M, et al. (2015) Calibrating the Human Mutation Rate via Ancestral Recombination Density in Diploid Genomes. *PLoS Genet* 11(11):e1005550.

156. Busby GBJ, et al. (2015) The Role of Recent Admixture in Forming the Contemporary West Eurasian Genomic Landscape. *Curr Biol* 25(21):2878.

157. Delaneau O, Zagury J-F, Marchini J (2013) Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* 10(1):5–6.

158. Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of population structure using dense haplotype data. *PLoS Genet* 8(1):e1002453.

159. Kunsch HR (1989) The Jackknife and the Bootstrap for General Stationary Observations. *Ann Stat* 17(3):1217–1241.

160. Leslie S, et al. (2015) The fine-scale genetic structure of the British population. *Nature* 519(7543):309–314.

161. Alt K (2006) Die artifizielle Schädeldeformation bei den Westgermanen. *Artificial Deformation of Human Head in Eurasian Past. OPUS: Interdisciplinary Investigation in Archaeology*, ed Mednikova M (Institute of Archaeology RAS, Moscow), pp 115–126.

162. Batieva EF (2006) Skulls with artificial deformations in the lower Don cemeteries of the first centuries AD. *Artificial Deformation of Human Head in Eurasian Past. OPUS: Interdisciplinary Investigation in Archaeology*, ed Mednikova M (Institute of Archaeology RAS, Moscow), pp 53–72.

163. Bernštam A (1940) *Kenkol'skii mogil'nik arkheologicheskoi ekspeditsii Ermitazha* (Leningrad).

164. Dubova NA (2006) Artificial head deformation among the Bronze Age farmers. *Artificial Deformation of Human Head in Eurasian Past. OPUS: Interdisciplinary Investigation in Archaeology*, ed Mednikova M (Institute of Archaeology RAS, Moscow), pp 22–36.

165. Hotz G (2011) Attilas kulturelles Erbe aus den Steppen Eurasiens. Künstliche Schädeldeformierungen in der Spätantike Europas. *Schädelkult: Kopf Und Schädel in Der Kulturgeschichte Des Menschen*, eds Wieczorek A, Rosendahl W (Schnell & Steiner, Regensburg), pp 109–113.

166. Kiszely I (1978) *The origins of artificial cranial formation in Eurasia from the sixth millennium BC to the seventh century AD* (British Archaeological Reports).

167. Sellenk J (2017) Das Phänomen der artifiziellen Schädeldeformation im archäologischen Kontext – Ein Überblick zur Verbreitung im 4. bis 7. Jahrhundert. Master (University of Leipzig).

168. Wagner G (2013) Die künstlich deformierten Schädel von Österreich in der Frühgeschichte. Diploma (University of Vienna).

169. Rott A (2017) Phänomene der frühmittelalterlichen Reihengräberzeit aus molekularbiologischer Sicht. PhD (LMU Munich).

170. Rott A, et al. (2017) Early medieval stone-lined graves in Southern Germany: analysis of an emerging noble class. *Am J Phys Anthropol* 162(4):794–809.

171. Sofeso C, Vohberger M, Wisnowsky A, Päffgen B, Harbeck M (2012) Verifying archaeological hypotheses: Investigations on origin and genealogical lineages of a privileged society in Upper Bavaria from Imperial Roman times (Erding, Kletthamer Feld). *Population Dynamics in Prehistory and Early History*, eds Kaiser E, Burger J, Schier W (DE GRUYTER, Berlin, Boston).

172. Csősz A, Szécsényi-Nagy A, Csákyová V, Langó P, Bódis V, Köhler K, Tömöry G, Nagy M, Mende BG (2016) Maternal Genetic Ancestry and Legacy of 10th Century AD Hungarians. *Scientific reports* 6:33446.

173. Vai S, et al. (2015) Genealogical relationships between early medieval and modern inhabitants of Piedmont. *PLoS One* 10(1):e0116801.

174. Alt KW, et al. (2014) Lombards on the move--an integrative study of the migration period cemetery at Szólád, Hungary. *PLoS One* 9(11):e110793.

## SI Appendix figures

**Figure S1** Grave inventory of individual Altenerding (AED) 1108. As many of the other individuals with an artificially deformed skull, AED 1108 was buried with a ("Thuringian") bow brooch that shows their close connection to an East Germanic complex. © Bavarian State Archaeological Collection.


**Figure S2** Two of the finds associated with the Late Roman burial from Freiham, grave 1335. The left panel shows the crossbow brooch (Pröttel type 1) which is crucial for the individual's dating to the 3rd century AD. The right panel shows the small bowl of type Alzey/Chenet 320 a that was recovered from the foot section of the grave cavity. © Bavarian State Department of Monuments and Sites.

**Figure S3** Artificially deformed skull from Kerch, Crimea. As described in the text, the skull is mounted to a wire rack and a mandible was attached to it. According to the morphological analysis by Dr. A. Staskiewicz, doubts remain about the correct assignment of the mandible to the male skull. A petrous bone could be sampled for both radiocarbon and ancient DNA analysis. Photo: A. Staskiewicz.

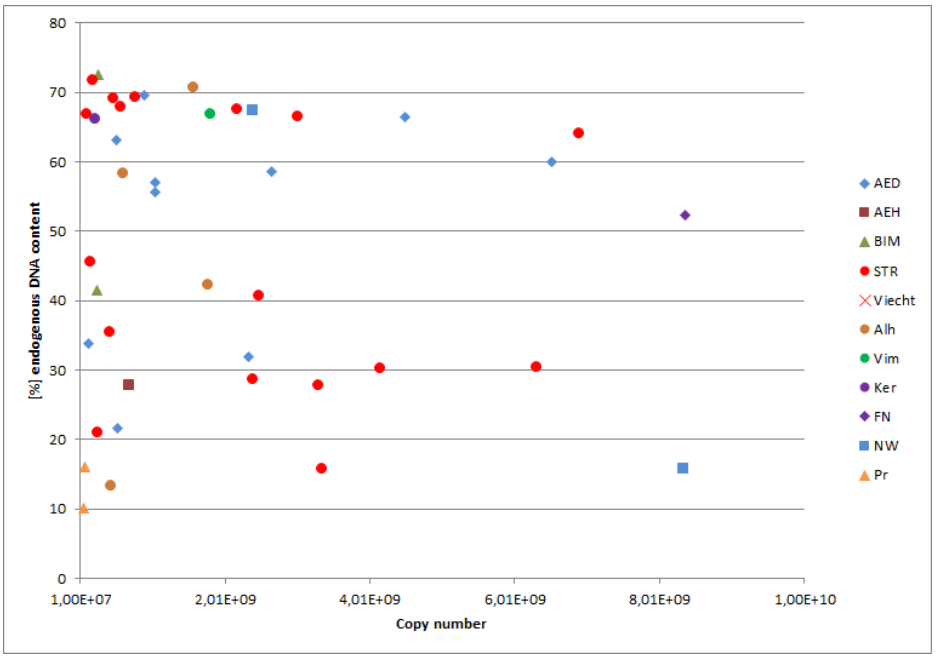


**Figure S4** Copy number of library molecules after Fill-In reaction are plotted against their corresponding endogenous DNA content of screening libraries analyzed in this study. Samples are grouped according to their respective sites.

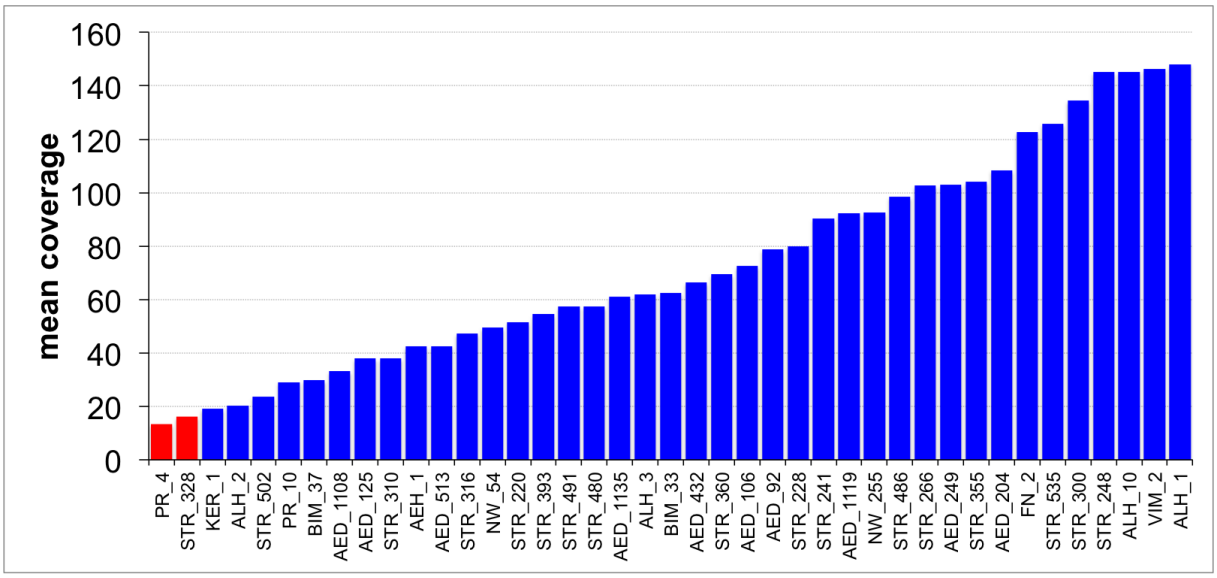


**Figure S5** Mean coverage for each of the 41 ancient samples at the 5 Mb capture region. Two samples in red were excluded from any analyses utilizing diploid calls due to their lower coverage (<20X).

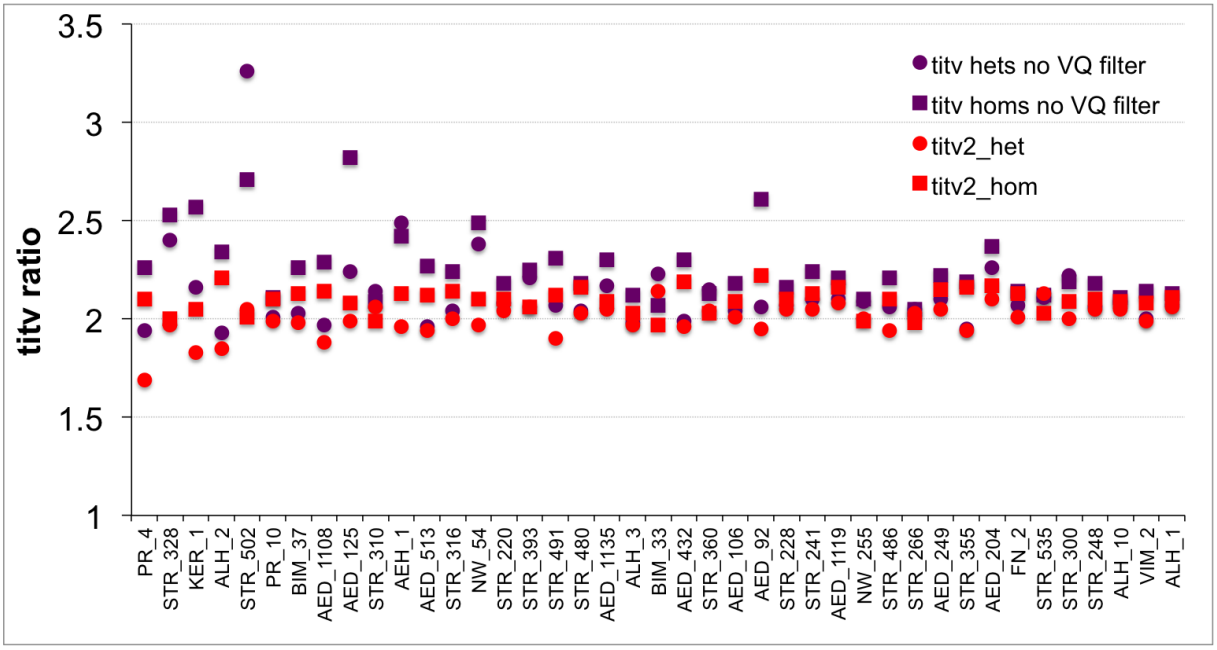


**Figure S6** TITV ratios at the 5 Mb capture region for heterozygous and non-reference homozygous sites with and without a variant quality filter of 30.

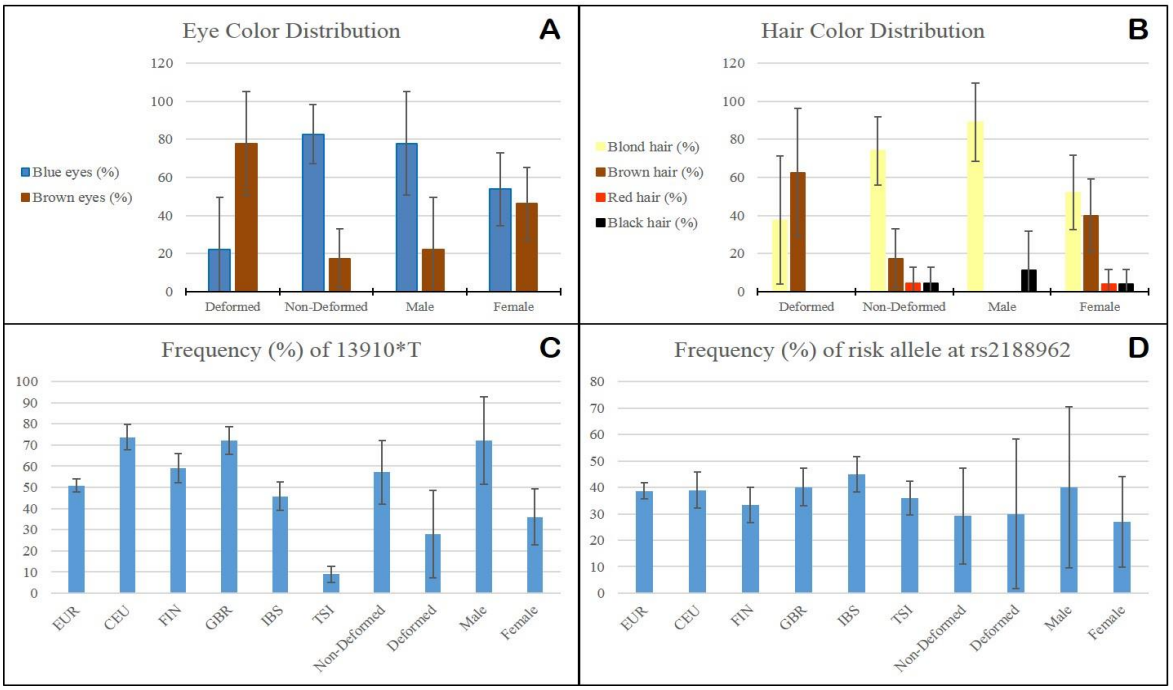


**Figure S7** A & B: Distribution of eye and hair color among the different groups of individuals. Error bars indicate 95% CIs. To calculate frequencies and 95% CIs, each individual was assigned the most likely iris and hair color of the HIrisplex output and the numbers of appearance of each color were determined on the group level. C: Derived allele frequencies of rs4988235 per group of individuals. Error bars indicate 95% CIs. D: Derived allele frequencies of rs2188962 per group of individuals. Error bars indicate 95% CIs.

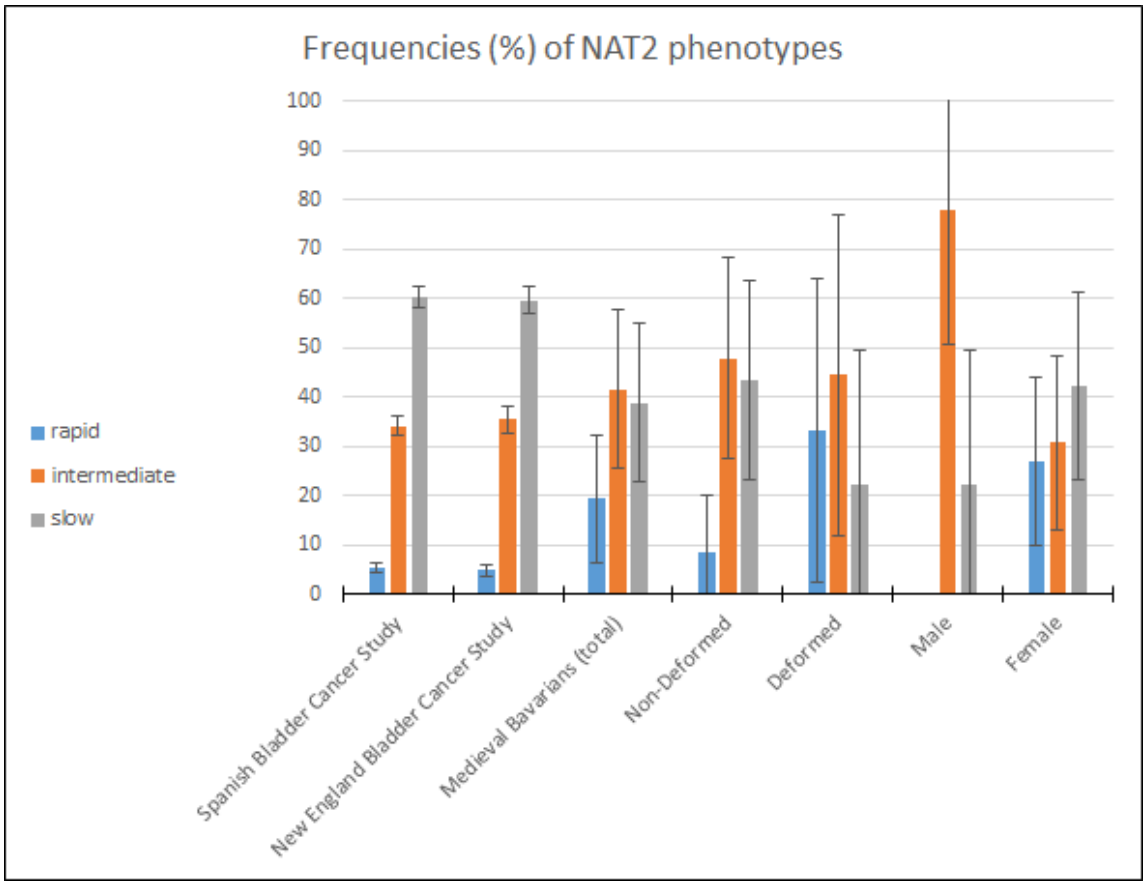


**Figure S8** Frequencies (%) of rapid, intermediate, and slow acetylation phenotypes among the different groups of individuals. Error bars indicate 95% confidence intervals. Furthermore, modern phenotype frequencies from both the Spanish Bladder Cancer Study and for individuals of European descent from the New England Bladder Cancer Study as published in García-Closas et al. (89) are shown.

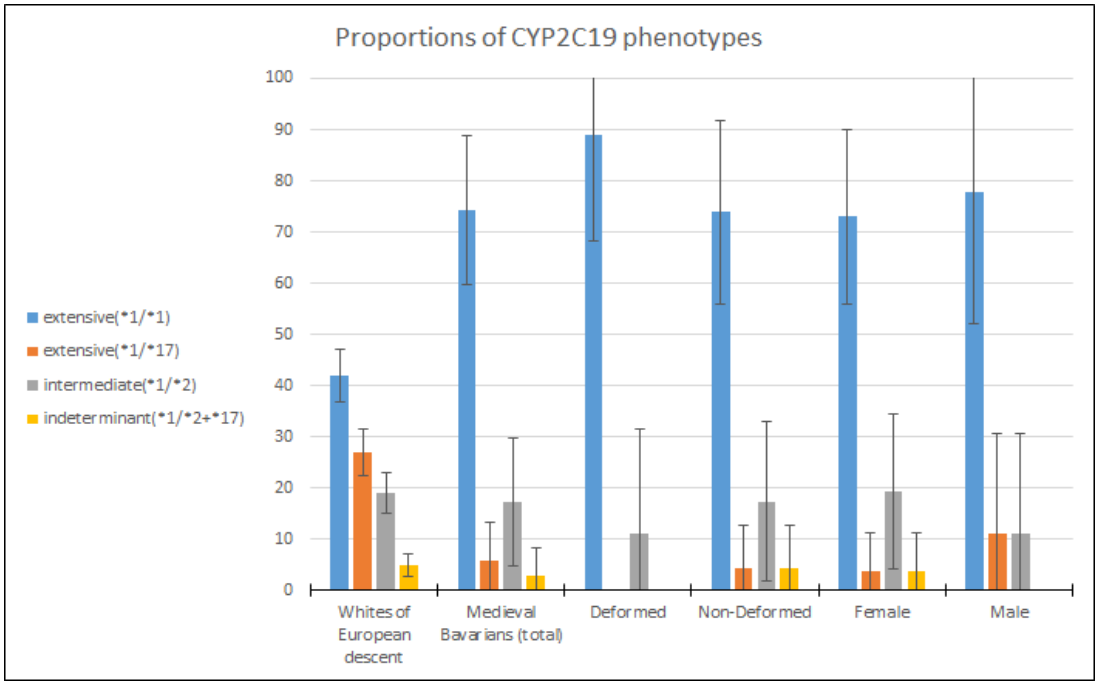


**Figure S9** Frequencies (%) of CYP2C19 phenotypes in the different groups of individuals. Error bars indicate 95% confidence intervals. Frequencies of whites of European descent were taken from Strom et al. (97). Numbers of ultrarapid and poor metabolizers are not shown as they did not occur in the ancient samples.

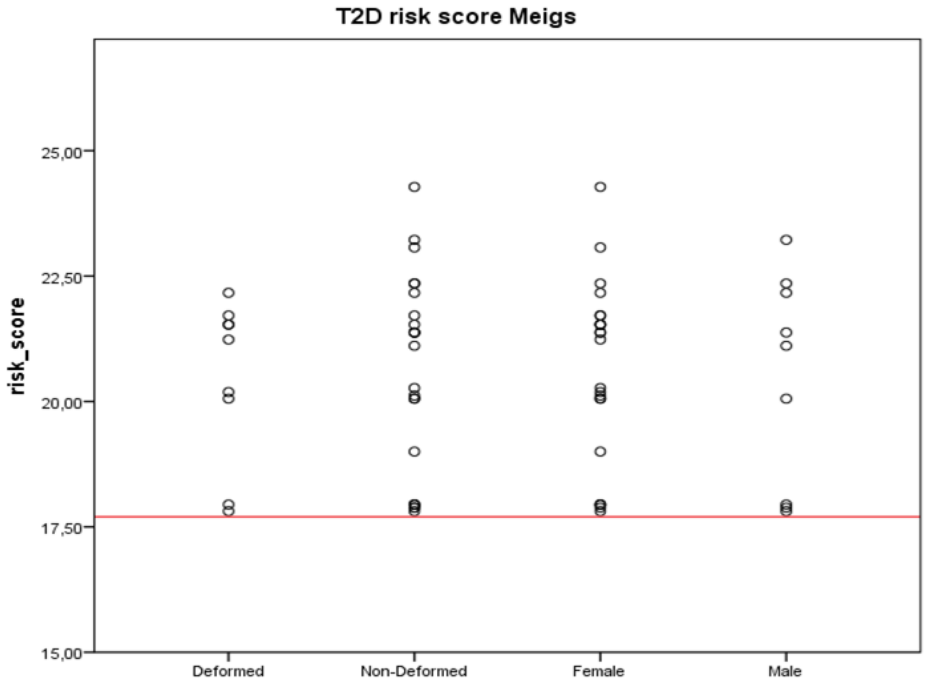


**Figure S10** GRS for development of T2D according to Meigs et al. (109). Individual data points are shown for every group, as well as the modern threshold for developing T2D of 17.7 (red line).

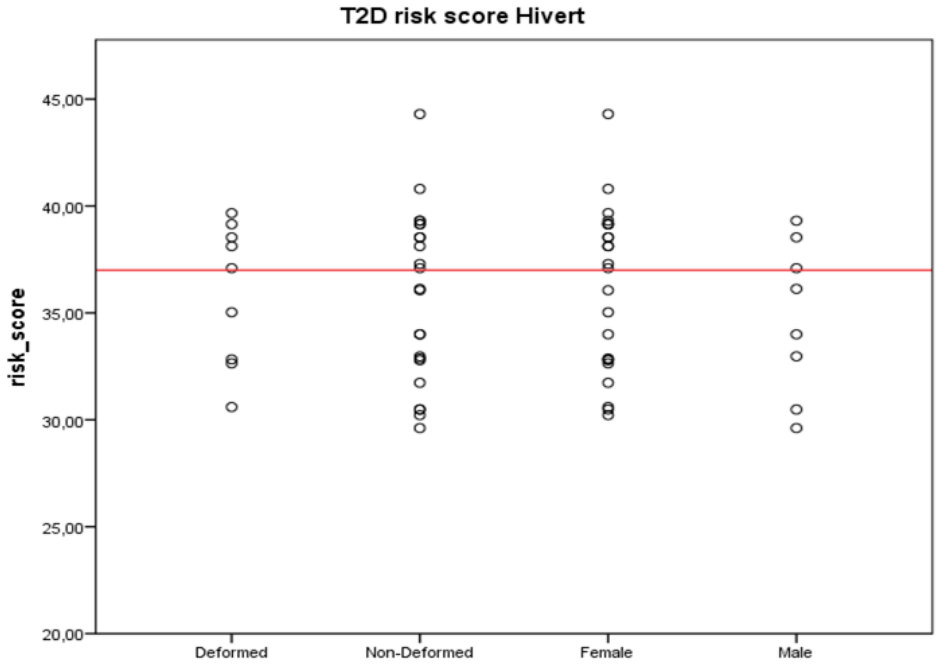


**Figure S11** GRS for development of T2D according to Hivert et al. (111). Individual values are shown for every group, as well as the median value of modern individuals (red line).

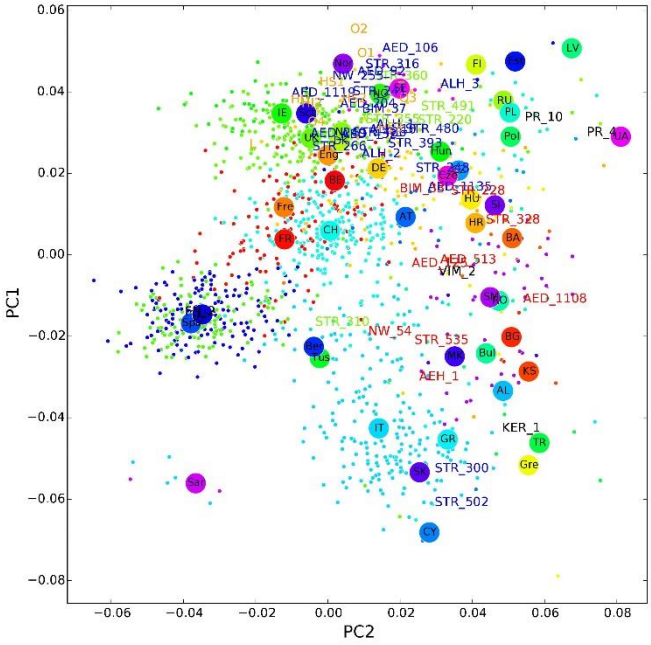


**Figure S12** Procrustes transformed PCA of ancient samples against POPRES imputed SNP dataset. Percentage of variation explained PCs 1 and 2 for modern populations only is 0.25% and 0.15% respectively. Two letter and three codes for POPRES samples: AL=Albania, AT=Austria, BA=Bosnia-Herzegovina, BE=Belgium, BG=Bulgaria, CH=Switzerland, CY=Cyprus, CZ=Czech Republic, DE=Germany, DK=Denmark, ES=Spain, FI=Finland, FR=France, GB=United Kingdom, GR, Greece, HR=Croatia, HU=Hungary, IE=Ireland, IT=Italy, KS=Kosovo, LV=Latvia, MK=Macedonia, NO=Norway, NL=Netherlands, PL=Poland, PT=Portugal, RO=Romania, SM=Serbia and Montenegro, RU=Russia, Sct=Scotland, SE=Sweden, SI=Slovenia, SK=Slovakia, TR=Turkey, UA=Ukraine. Three letter codes for SGDP samples: Ber=Bergamo, Bul=Bulgaria, Cze=Czech Republic, Eng=England, Est=Estonia, Fre=France, Gre=Greece, Hun=Hungary, Nor=Norway, Pol=Poland, Sar=Sardinian, Spa=Spain, Tus=Tuscan.

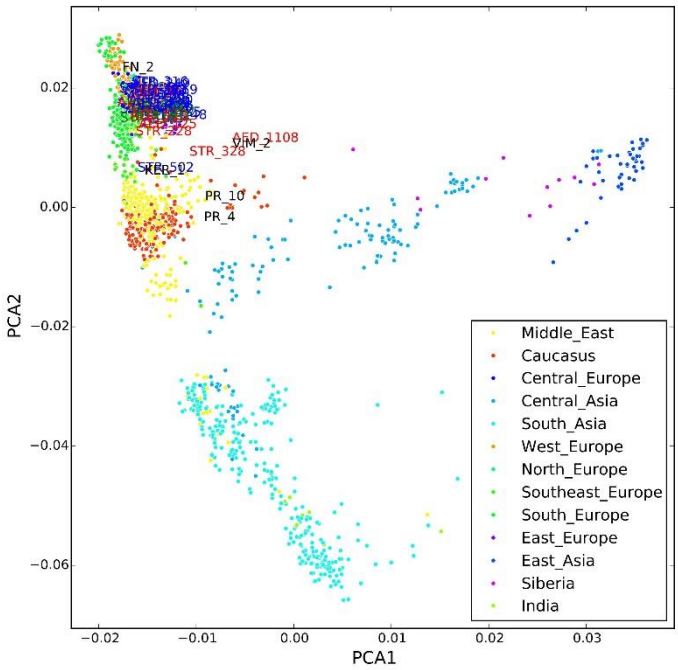


**Figure S13** Procrustes transformed PCA of ancient samples against HellBus imputed SNP dataset plus 1000 Genomes East Asian populations. Percentage of variation explained PCs 1 and 2 for modern populations only is 5.17% and 0.77% respectively.

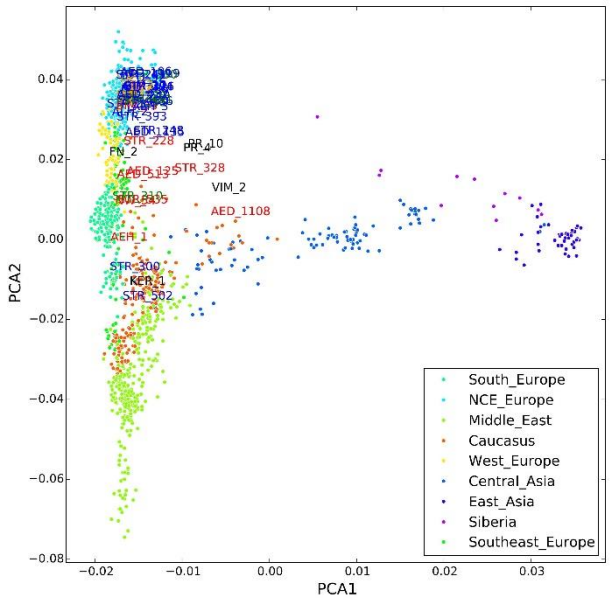


**Figure S14** Procrustes transformed PCA of ancient samples against HellBus imputed SNP dataset excluding South Asian populations plus 1000 Genomes East Asian populations. Percentage of variation explained PCs 1 and 2 for modern populations only is 5.12% and 0.50%.

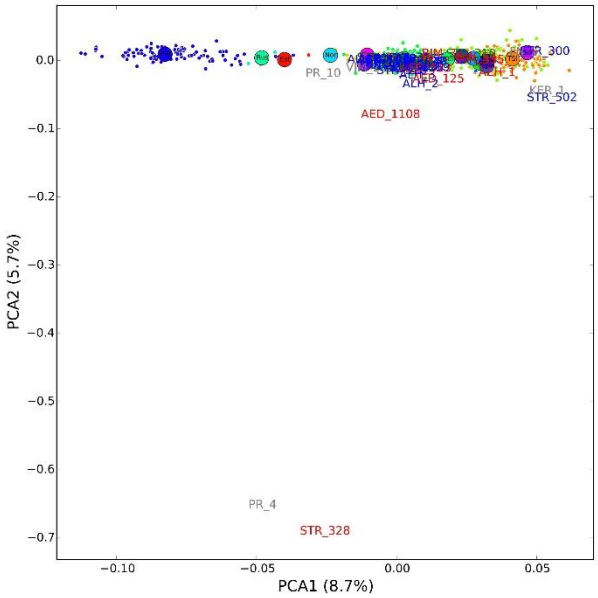


**Figure S15** PCA of ancient samples, 1000 Genomes European populations, GoNL, SGDP West Eurasian and Turkish genomes based on 5 Mb haplotypes.

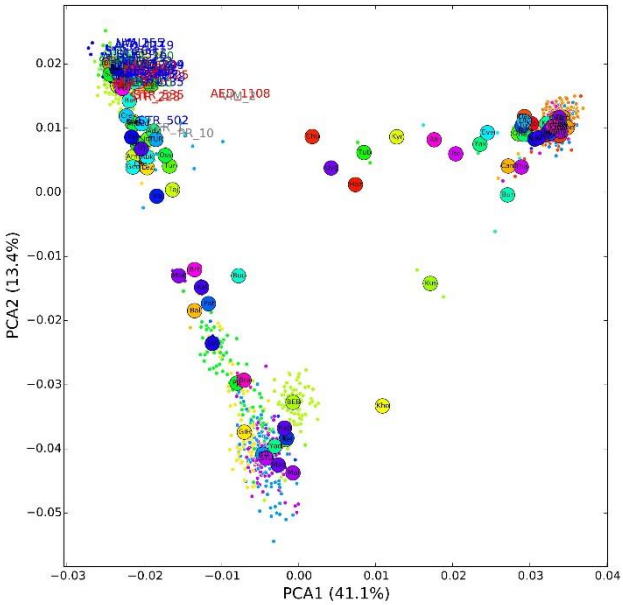


**Figure S16** PCA of ancient samples, 1000 Genomes Eurasian populations, GoNL, SGDP Eurasian and Turkish genomes based on 5 Mb haplotypes.

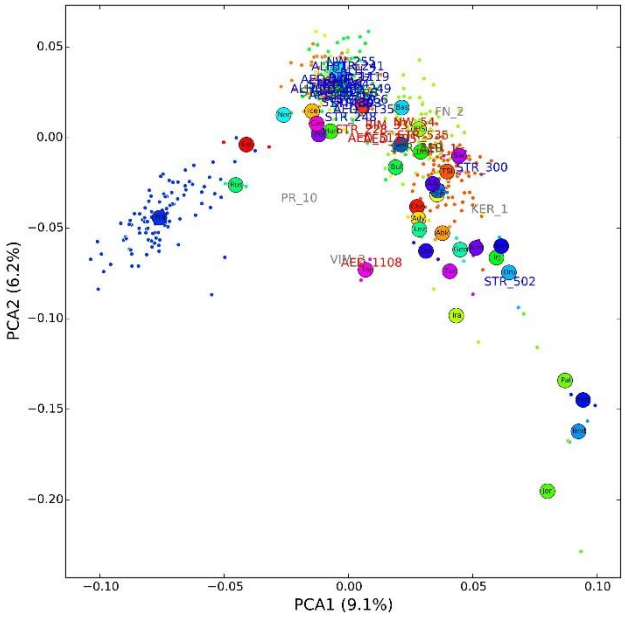


**Figure S17** PCA of ancient samples, 1000 Genomes European populations, GoNL, SGDP West Eurasian and Turkish genomes based on 5 Mb haplotypes.

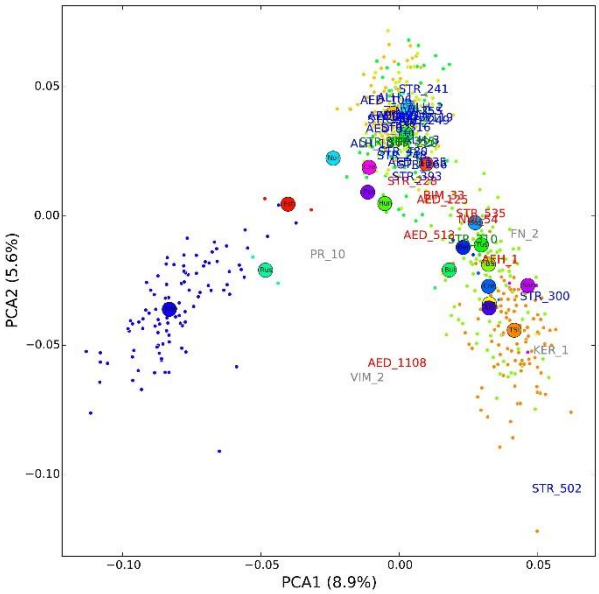


**Figure S18** PCA of ancient samples, 1000 Genomes European populations, GoNL and SGDP European genomes based on 5 Mb haplotypes.

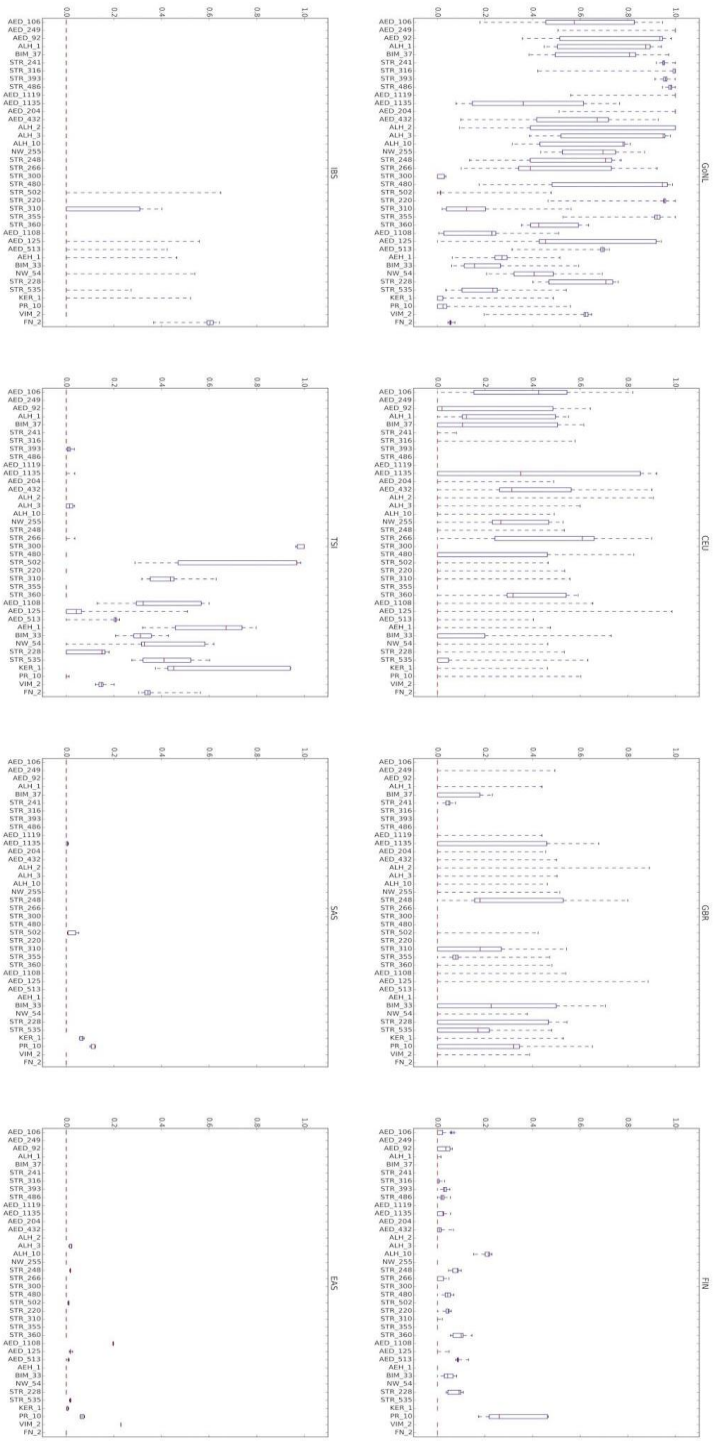


**Figure S19** Box plot distribution of ADMIXTURE coefficient estimates for ancient samples against each individual supervised population (GoNL, CEU, GBR, FIN, IBS, TSI, SAS and SAS) across 100 independent runs based on 5 Mb haplotypes.

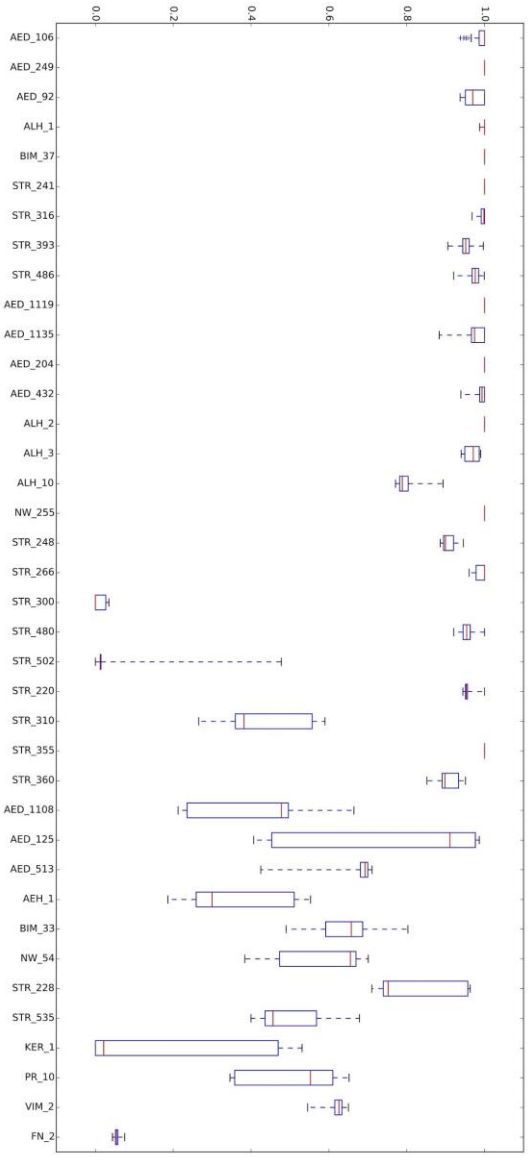


**Figure S20** Box plot distribution of ADMIXTURE coefficient estimates for ancient samples summed for GoNL, CEU and GBR across 100 independent runs based on 5 Mb haplotypes.

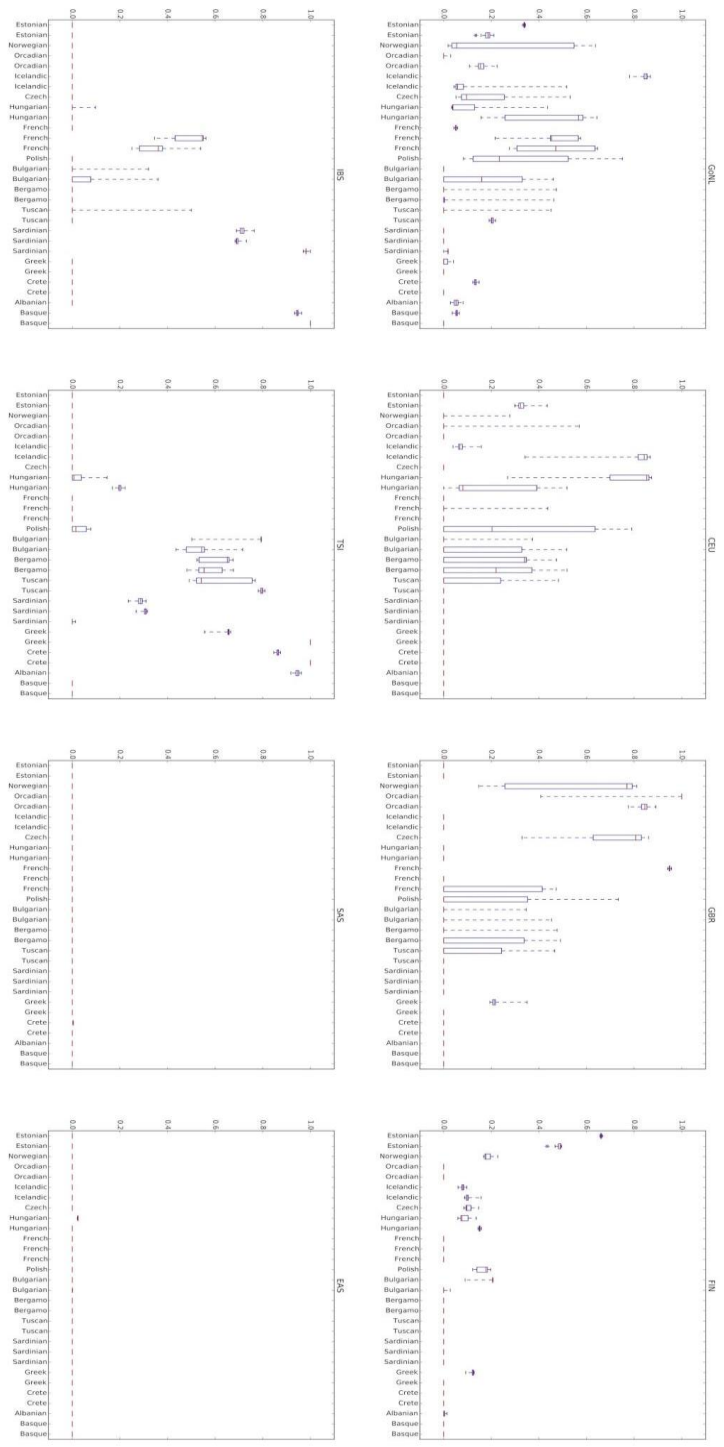


**Figure S21** Box plot distribution of ADMIXTURE coefficient estimates for modern SGDP European samples against each individual supervised population (GoNL, CEU, GBR, FIN, IBS, TSI, SAS and SAS) across 100 independent runs based on 5 Mb haplotypes.

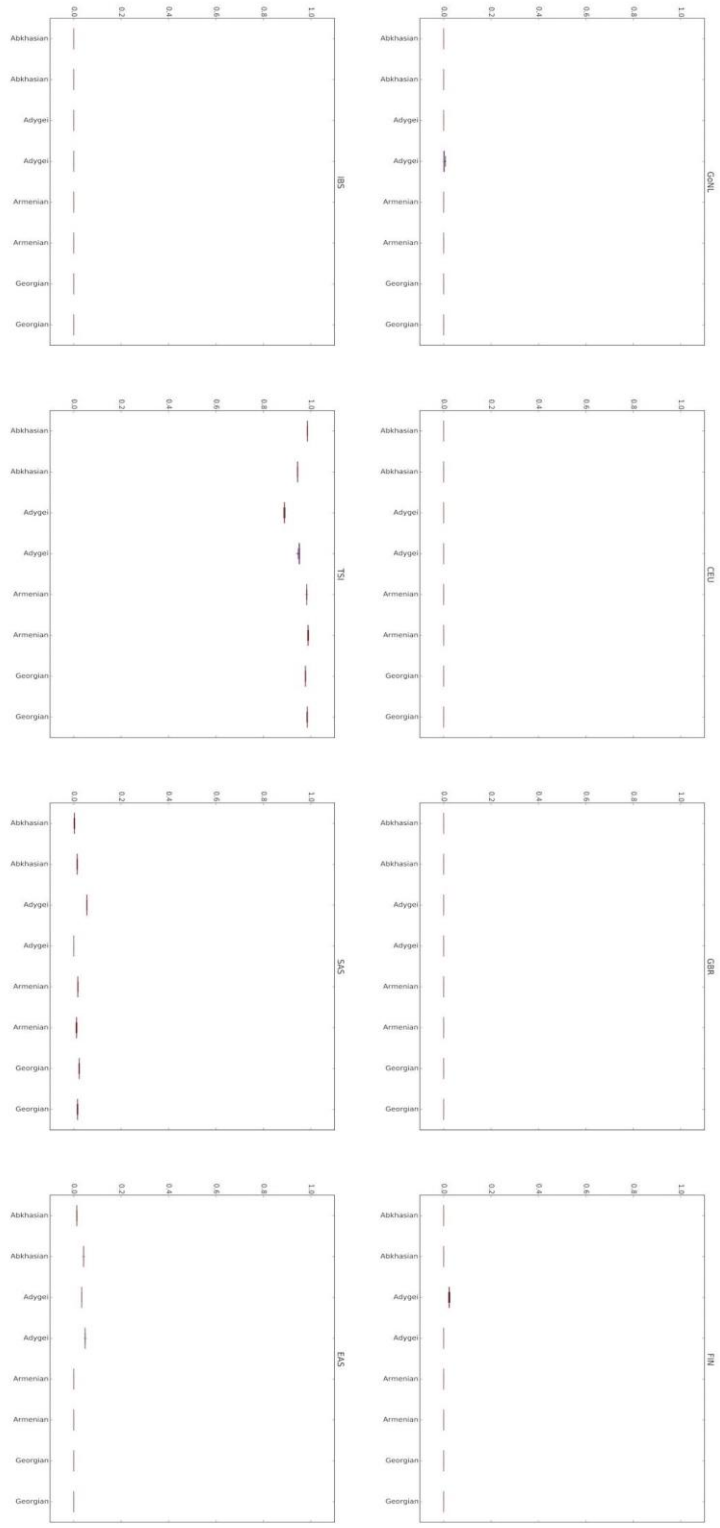


**Figure S22** Box plot distribution of ADMIXTURE coefficient estimates for modern SGDP Caucasus samples against each individual supervised population (GoNL, CEU, GBR, FIN, IBS, TSI, SAS and SAS) across 100 independent runs based on 5 Mb haplotypes.

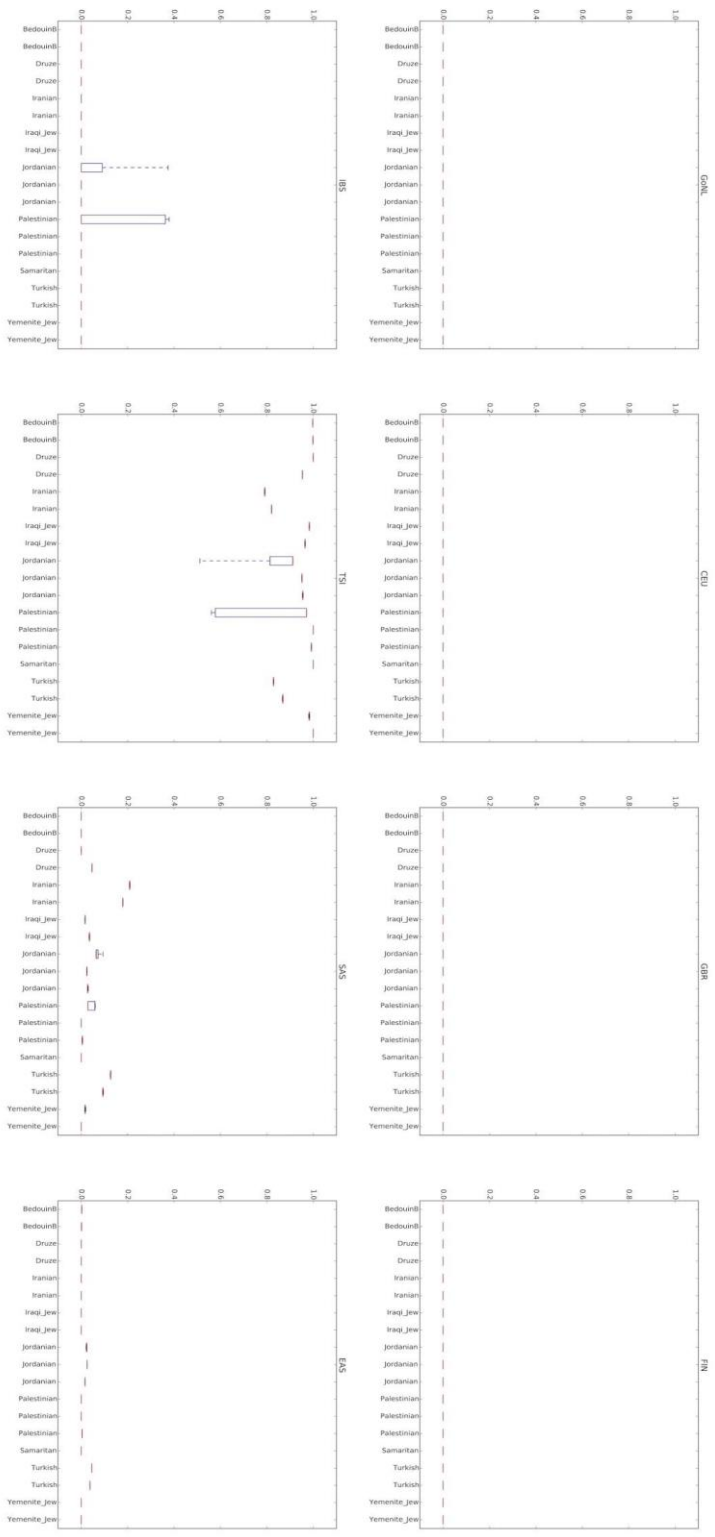


**Figure S23** Box plot distribution of ADMIXTURE coefficient estimates for modern Middle Eastern SGDP samples against each individual supervised population (GoNL, CEU, GBR, FIN, IBS, TSI, SAS and SAS) across 100 independent runs based on 5 Mb haplotypes.

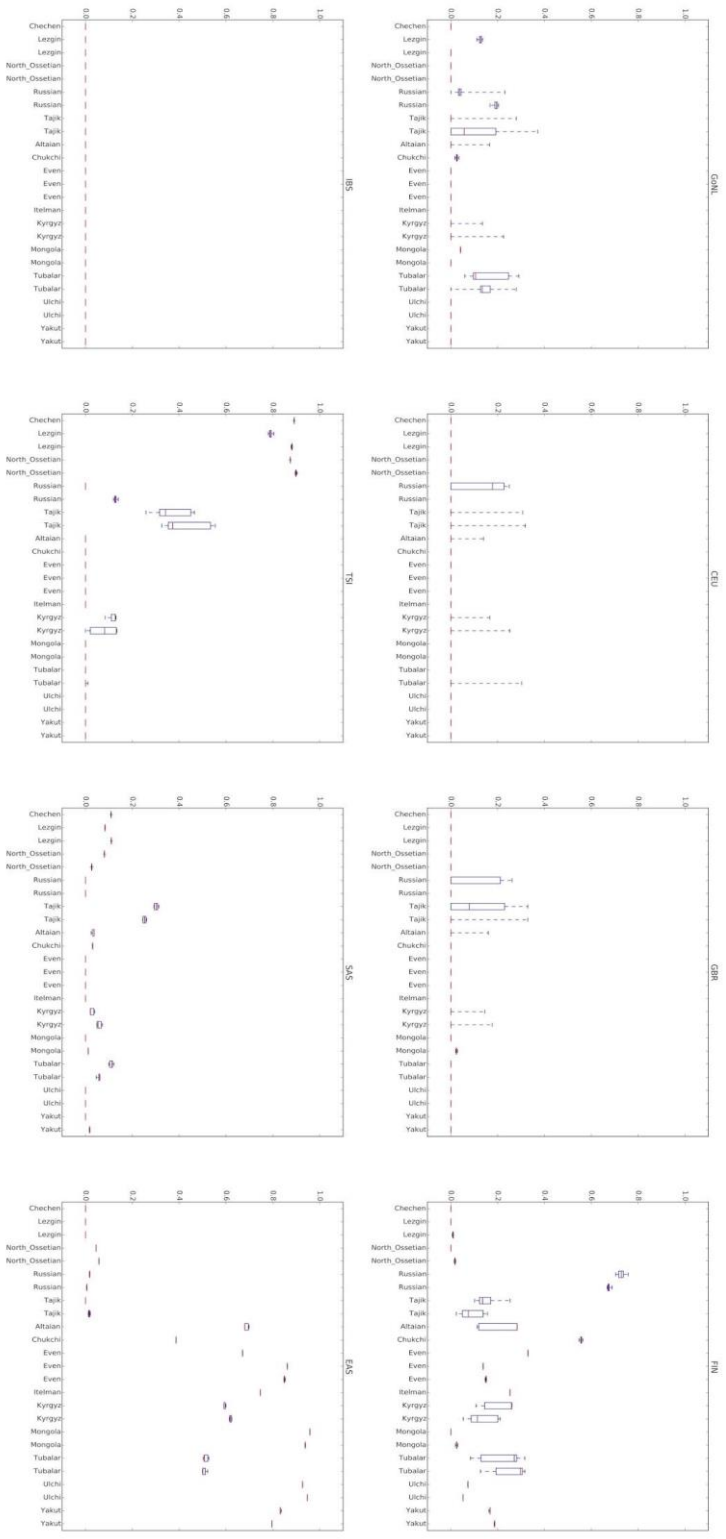


**Figure S24** Box plot distribution of ADMIXTURE coefficient estimates for modern SGDP Central Asian samples against each individual supervised population (GoNL, CEU, GBR, FIN, IBS, TSI, SAS and SAS) across 100 independent runs based on 5 Mb haplotypes.

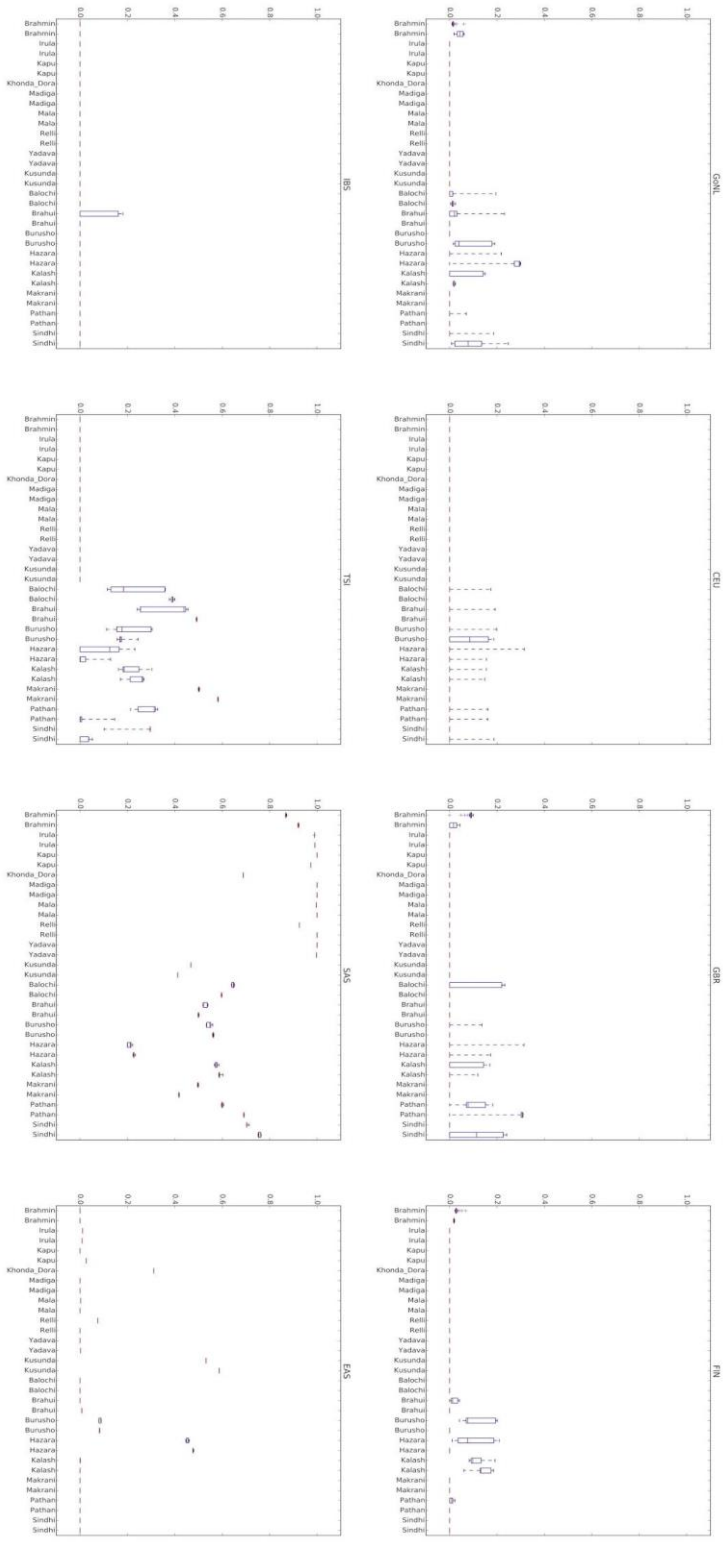


**Figure S25** Box plot distribution of ADMIXTURE coefficient estimates for modern SGDP South Asian samples against each individual supervised population (GoNL, CEU, GBR, FIN, IBS, TSI, SAS and SAS) across 100 independent runs based on 5 Mb haplotypes.

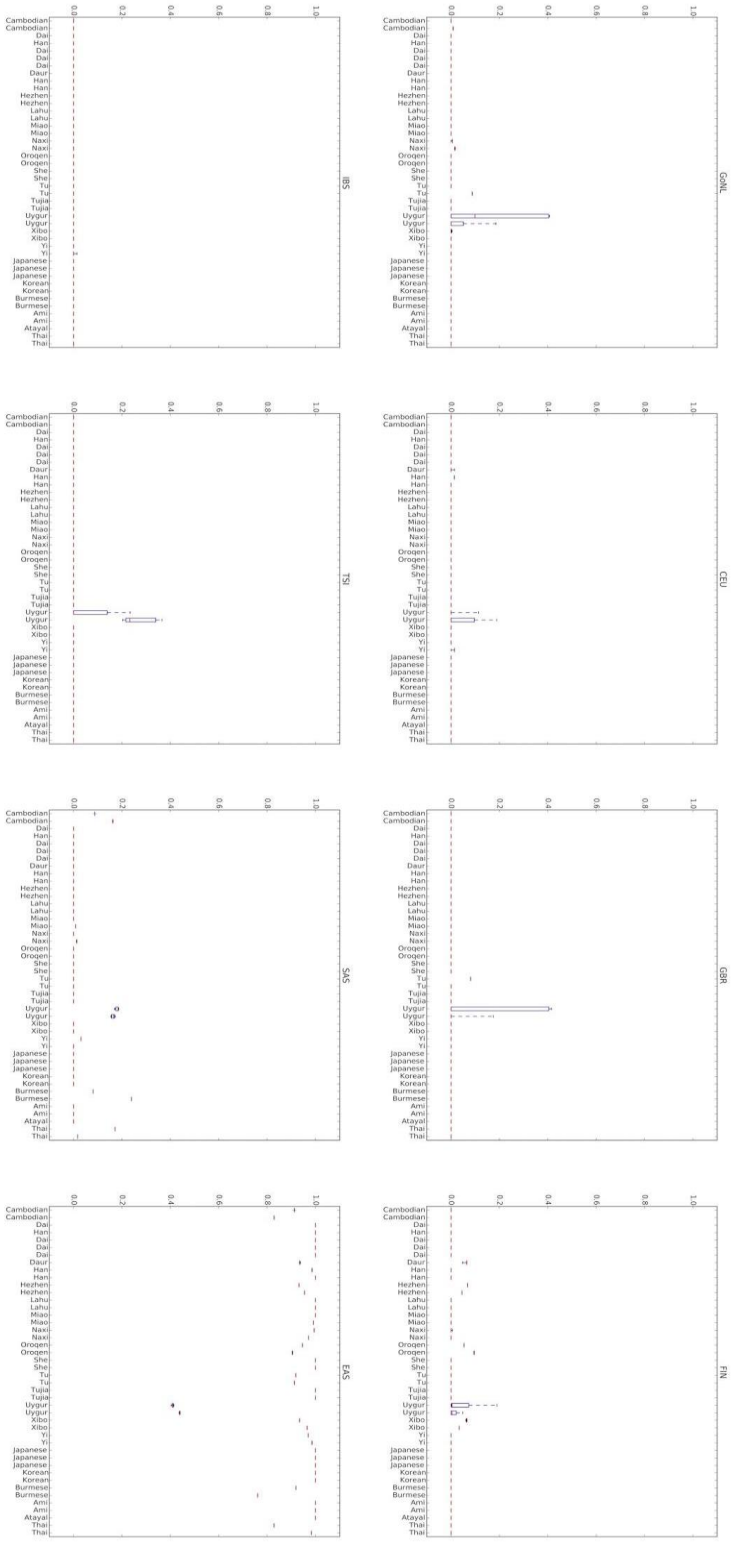


**Figure S26** Box plot distribution of ADMIXTURE coefficient estimates for modern SGDP East Asian samples against each individual supervised population (GoNL, CEU, GBR, FIN, IBS, TSI, SAS and SAS) across 100 independent runs based on 5 Mb haplotypes.

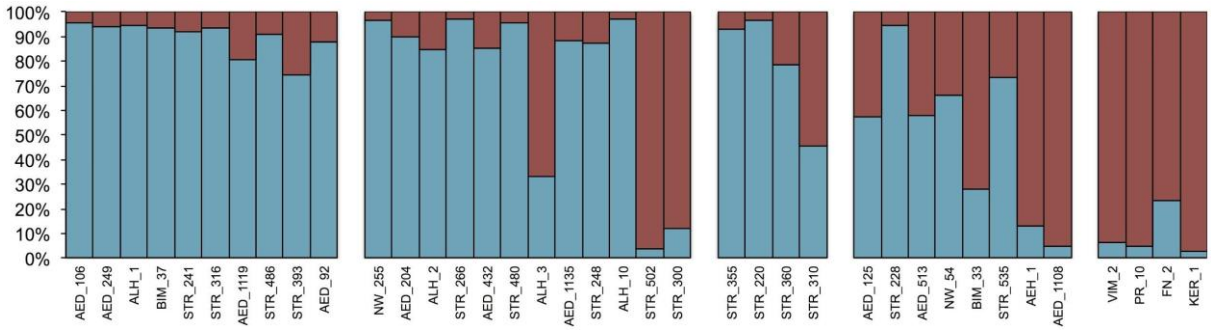


**Figure S27** Unsupervised STRUCTURE analysis for ancient samples for K=2

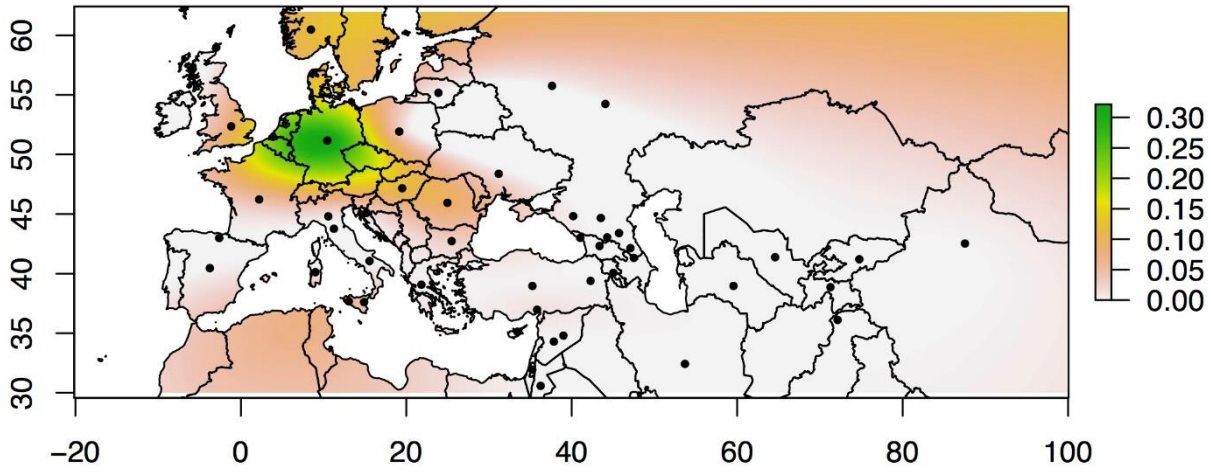


**Figure S28** Geographic distribution of population similarity analysis for grouped Bavarian samples with intermediate heads.

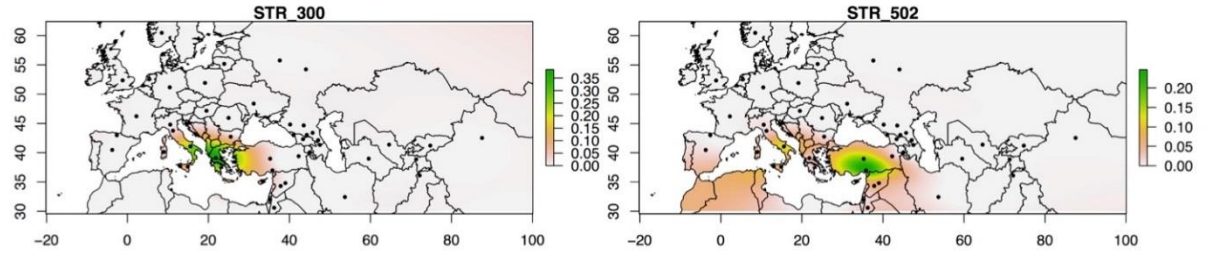


**Figure S29** Geographic distribution of population similarity analysis for individual Bavarian samples with normal head and non-central European median location.

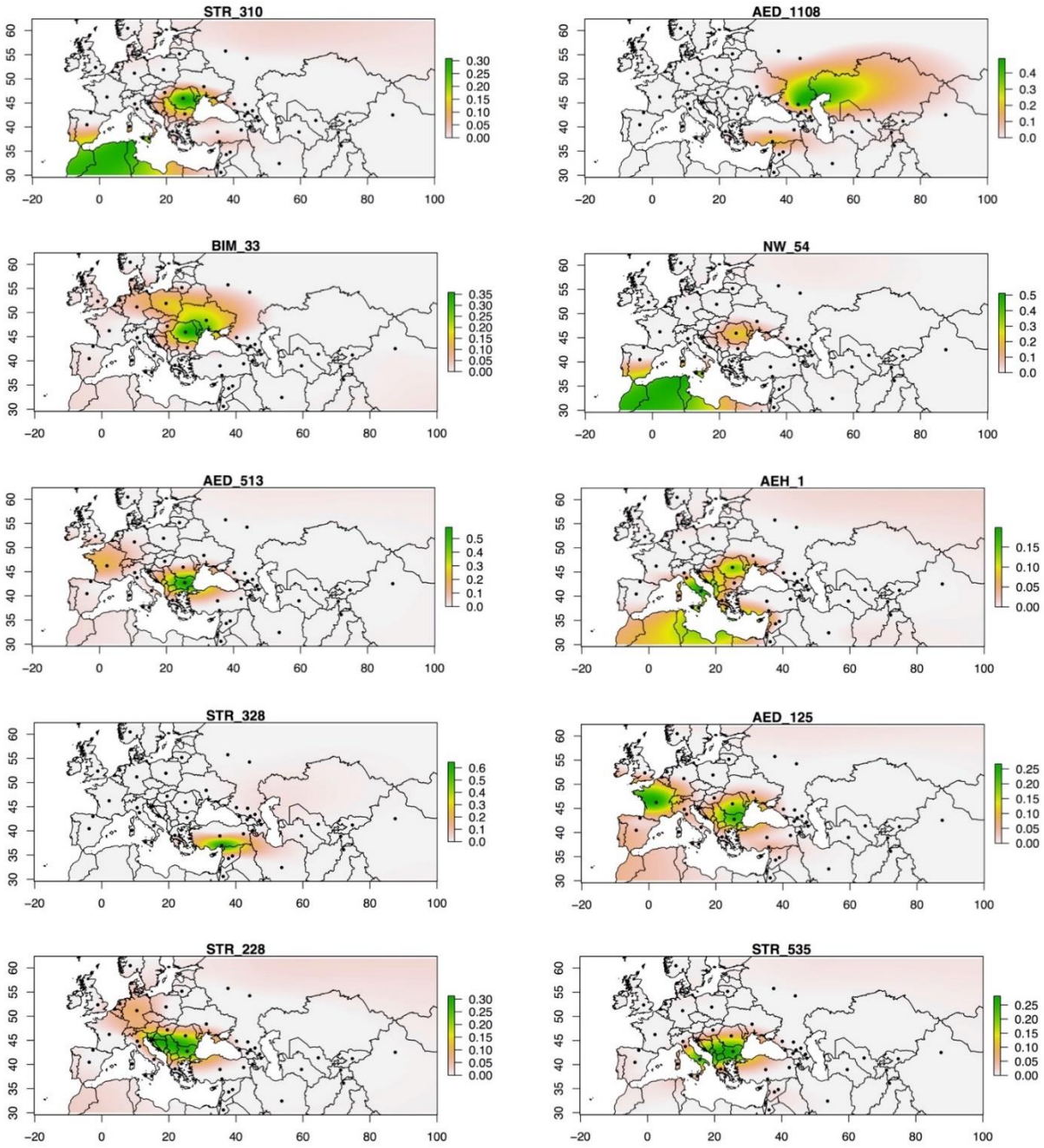


**Figure S30** Geographic distribution of population similarity analysis for individual Bavarian samples with elongated heads as well as STR_310.

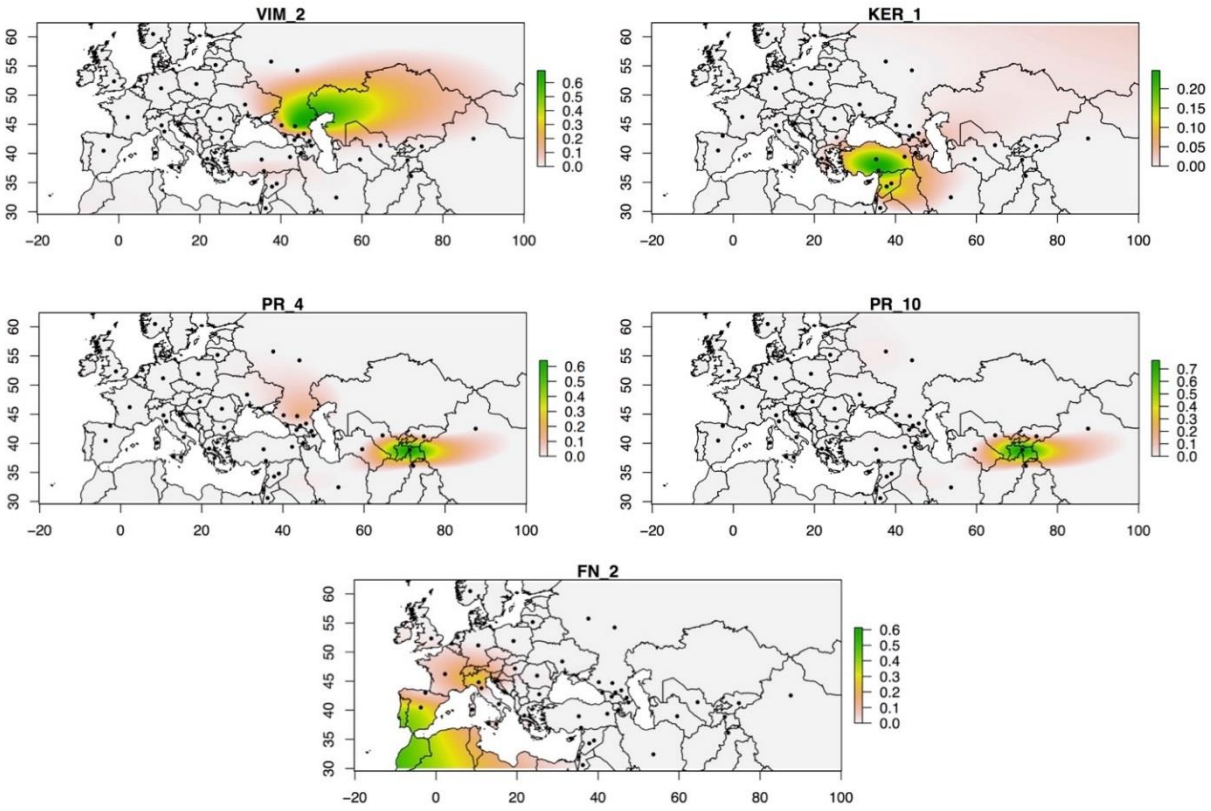


**Figure S31** Geographic distribution of population similarity analysis for individual non-Bavarian ancient samples.

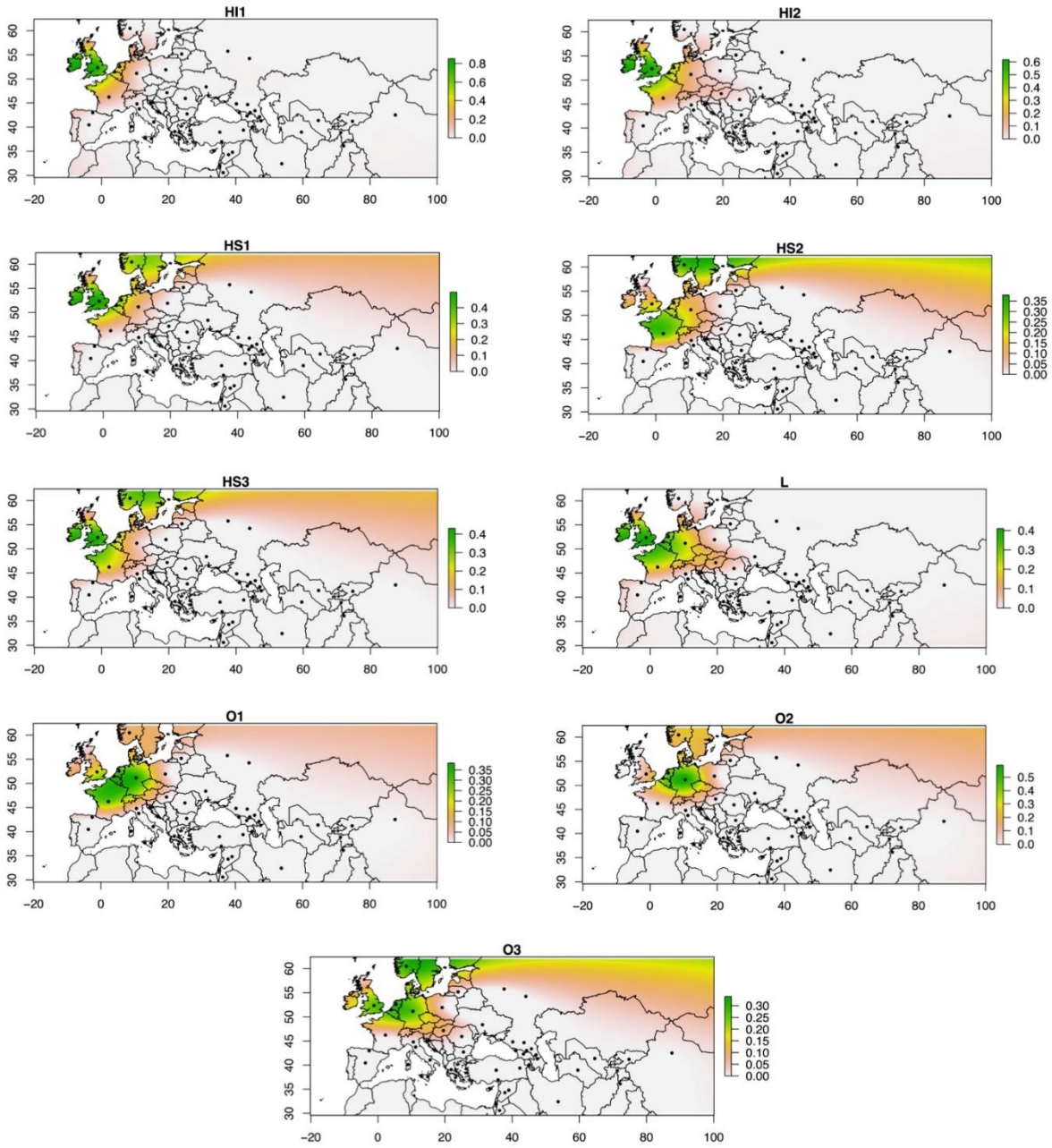


**Figure S32** Geographic distribution of population similarity analysis for individual Anglo-Saxon ancient samples.

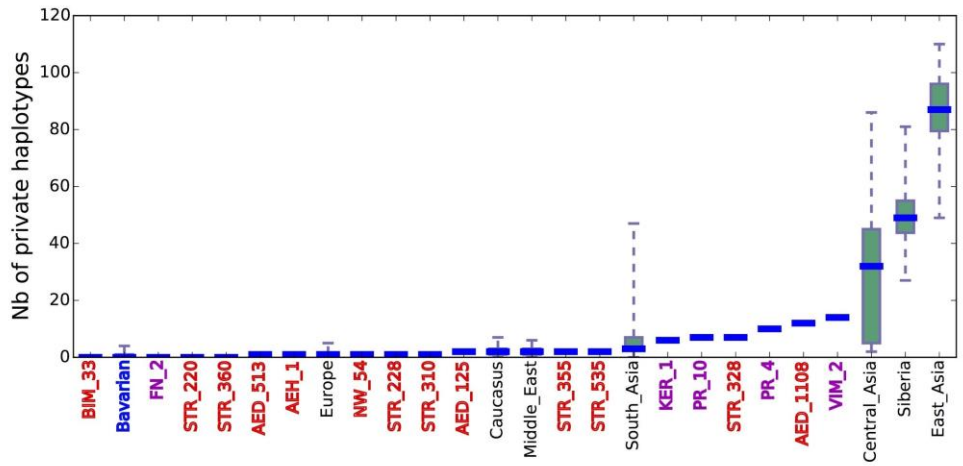


**Figure S33** Boxplot showing number of private haplotypes for individual 1k loci from the 5 Mb neutralome in the 1000 Genomes East Asian populations for ancient samples and SGDP samples grouped by modern region. Ancient Bavarian samples are either grouped together (all males and females with normal skulls) or shown individually (all females with elongated skulls, red labelling, and non-Bavarian ancient samples, purple labelling). Order is sorted by point estimate of median for groups of individuals.

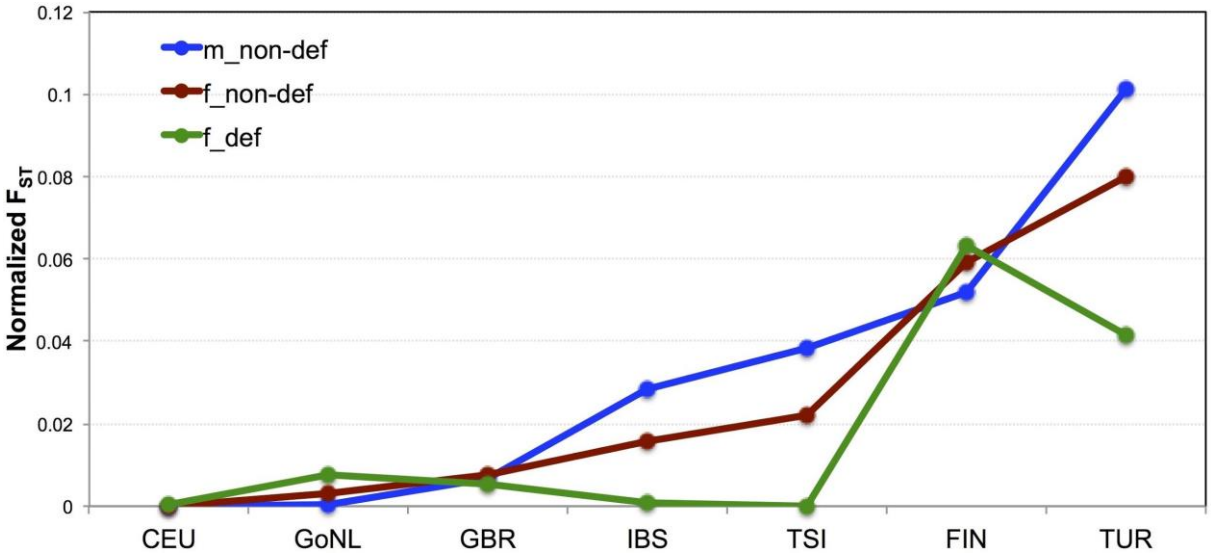


**Figure S34** Pairwise $F_{ST}$ estimates for haplotype data at the 5 Mb region for Medieval populations against each of the 1000 Genome European populations as well at the Dutch and Turkish samples.

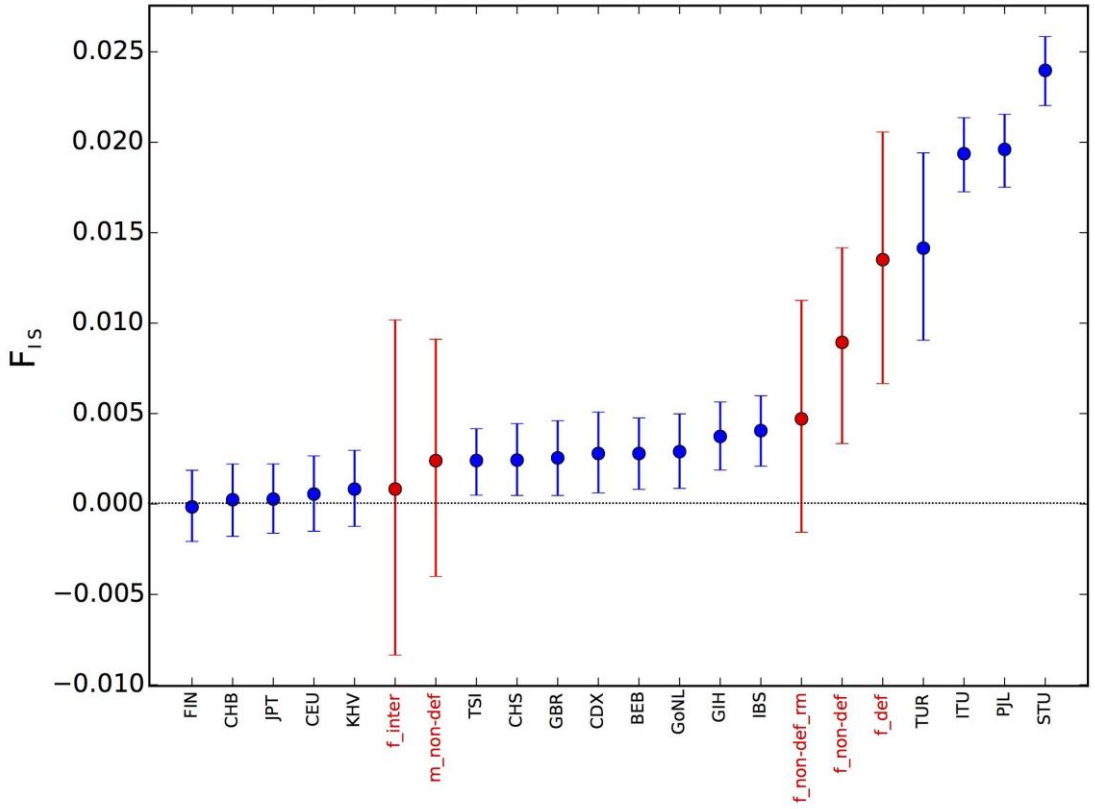


**Figure S35** $F_{IS}$ estimates for haplotype data at the 5 Mb region. Note that f_non_def_rm presents the $F_{IS}$ estimate when removing the two outlier individuals STR_300 and STR_502.

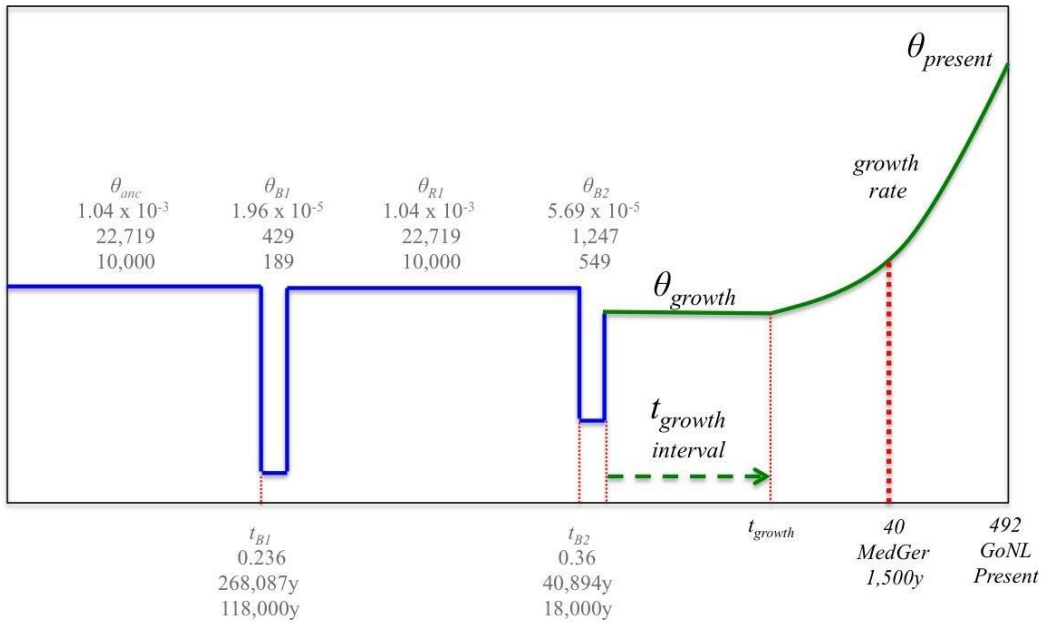


**Figure S36** Demographic model of European populations under recent exponential growth. Blue line represent portion of model fixed based on previous estimates of Gazave et al.. Fixed parameters are represented by three numbers, top: units of θ and time in coalescent, $t_c$, middle: assuming μ =2.59 x $10^{-8}$, bottom: assuming μ =1.2 x $10^{-8}$. Green line represents portion of model fit in this study using ∂a∂i using a 2D-AFS. Years were based on assuming a generation time of 25 years. Figure not to scale.

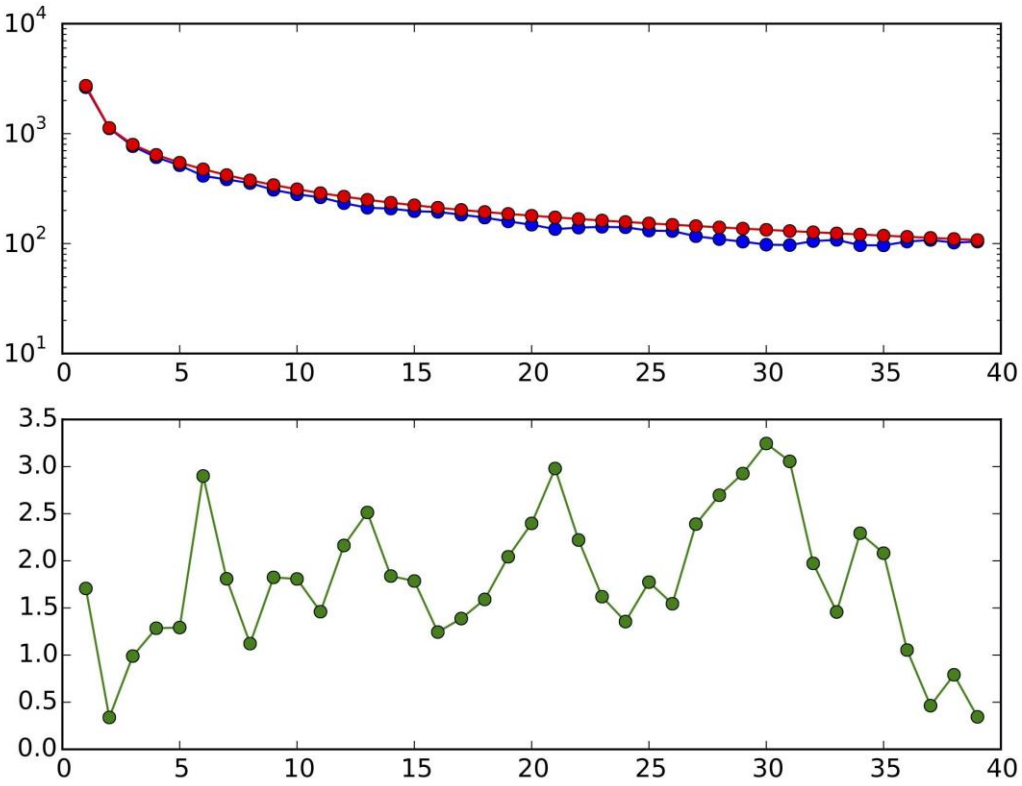


**Figure S37** Top - AFS of Medieval Bavarian genomes (blue) and best fit model (red), Bottom -residual difference between model and data.

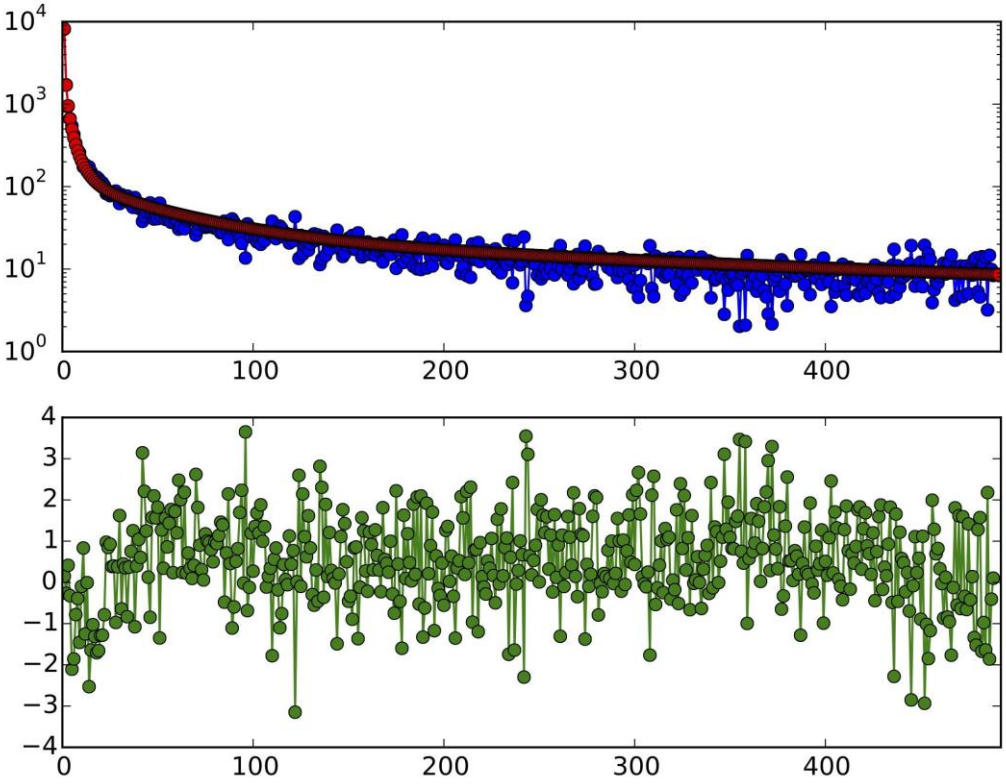


**Figure S38** Top - AFS of modern Dutch (GoNL) genomes (blue) and best fit model (red), Bottom -residual difference between model and data.

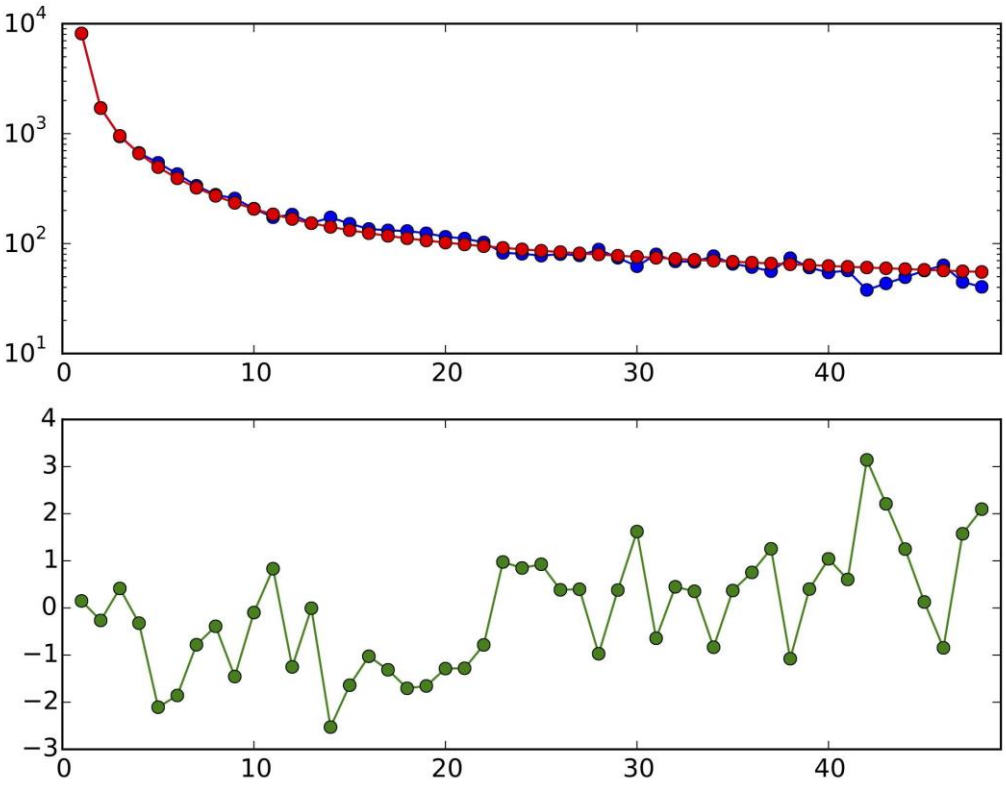


**Figure S39** Top - AFS of modern Dutch (GoNL) genomes (blue) for the first 50 frequency classes and best fit model (red), Bottom -residual difference between model and data.

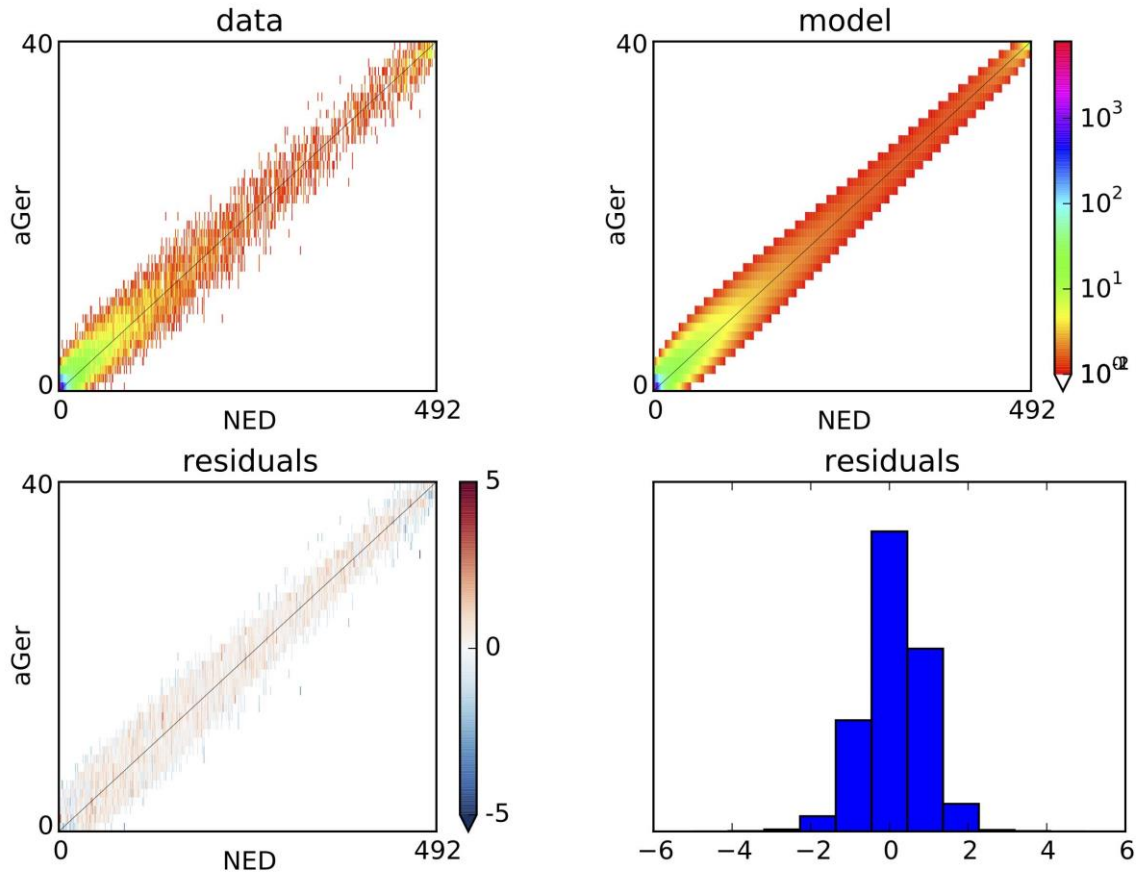


**Figure S40** 2D-AFS of modern Dutch and Medieval Bavarian genomes (top left) and best fit model (top right), Bottom row is residual difference between model and data.

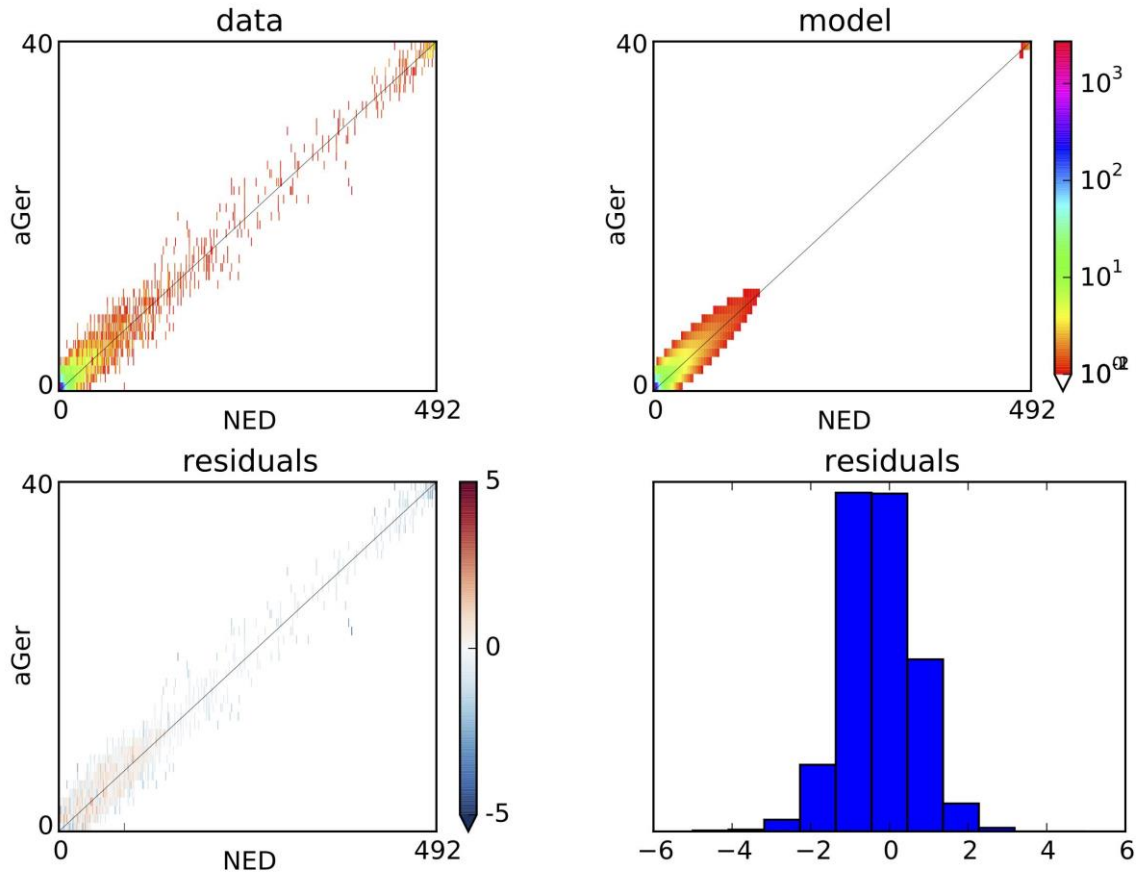


**Figure S41** 2D-AFS of modern Dutch and Medieval Bavarian genomes excluding transitions (top left) and best fit model (top right), Bottom row is residual difference between model and data.

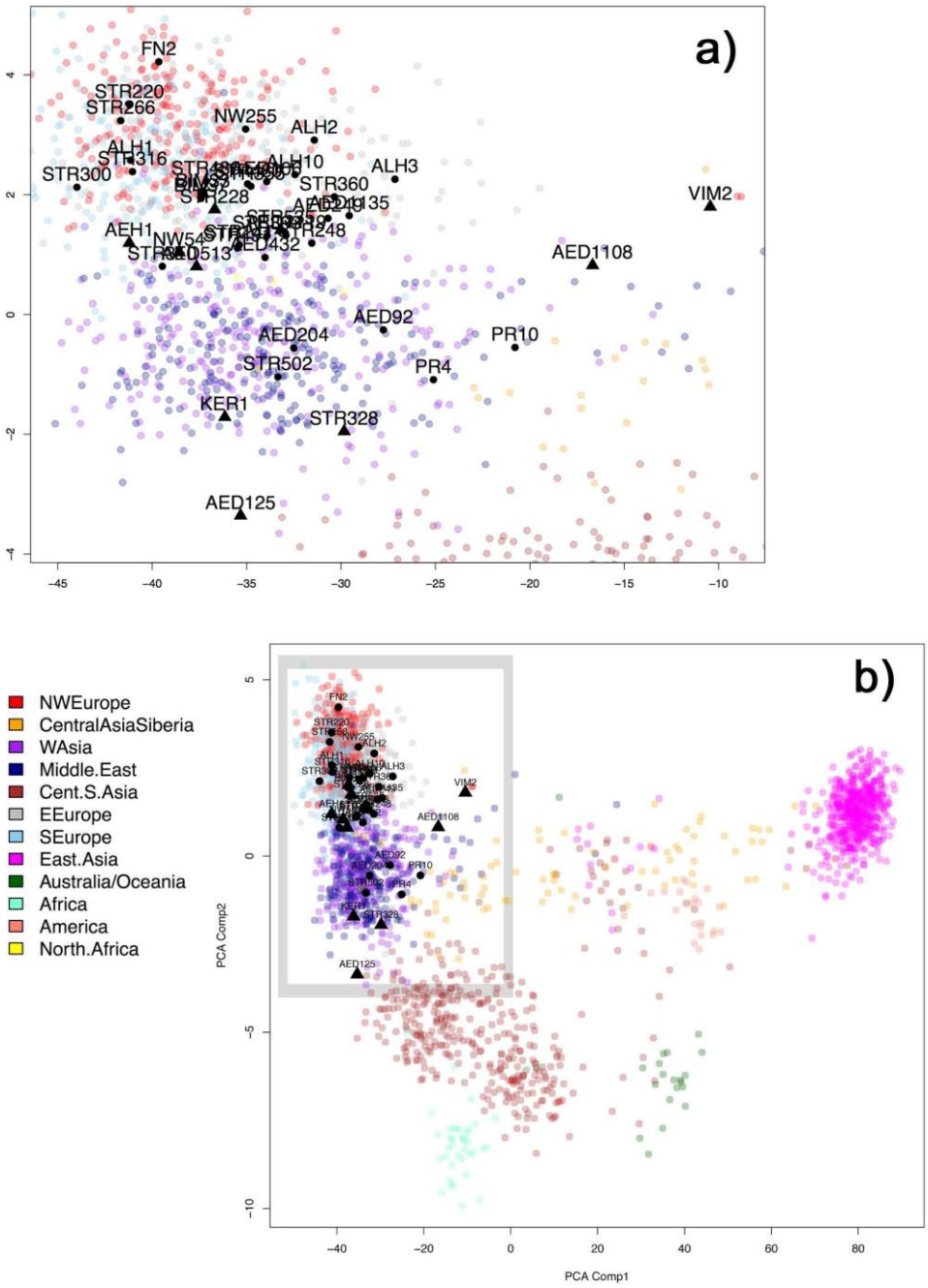


**Figure S42** b) PCA (components 1 and 2) applied to the CHROMOPAINTER inferred allele-sharing profiles based on the HellBus capture dataset across 8692 SNPs. Modern individuals are colored roughly according to the regional clusters identified in Hellenthal et al.(71) as per the legend. Ancient samples are given as black points with those individuals exhibiting a deformed phenotype depicted with a triangle. (a) provides a zoomed insert of (b) as highlighted in the grey box.

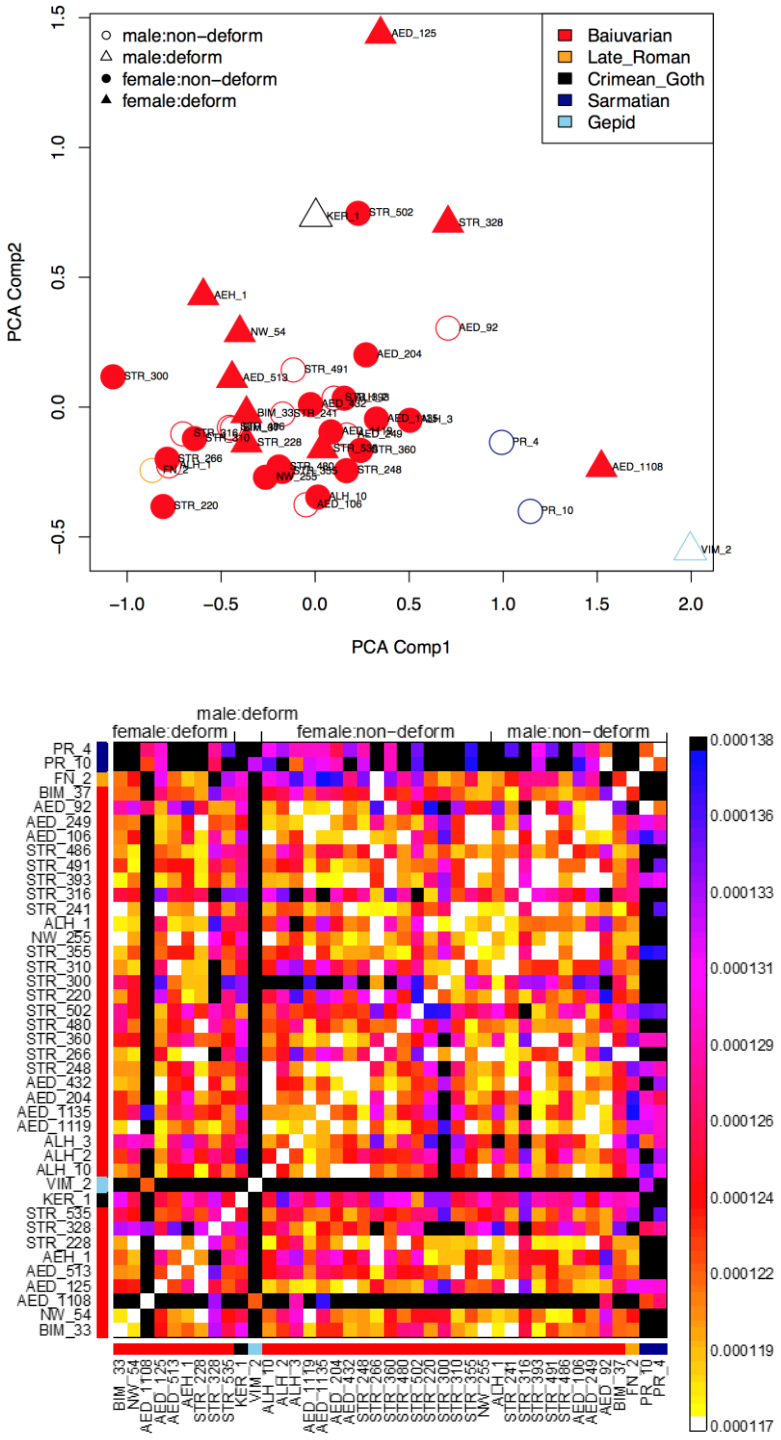


**Figure S43** (top) PCA of the inferred allele sharing profiles for ancient samples based on the HellBus dataset. Individuals are colored according to culture with those with a deformed skull phenotype plotted with triangular symbols (bottom) Heatmap providing the pairwise inferred TVD between each of the ancient samples, excluding those from Schiffels et al. (70), based on the allele sharing profiles inferred using the HellBus dataset. Individuals are ordered according to their gender and skull phenotype.

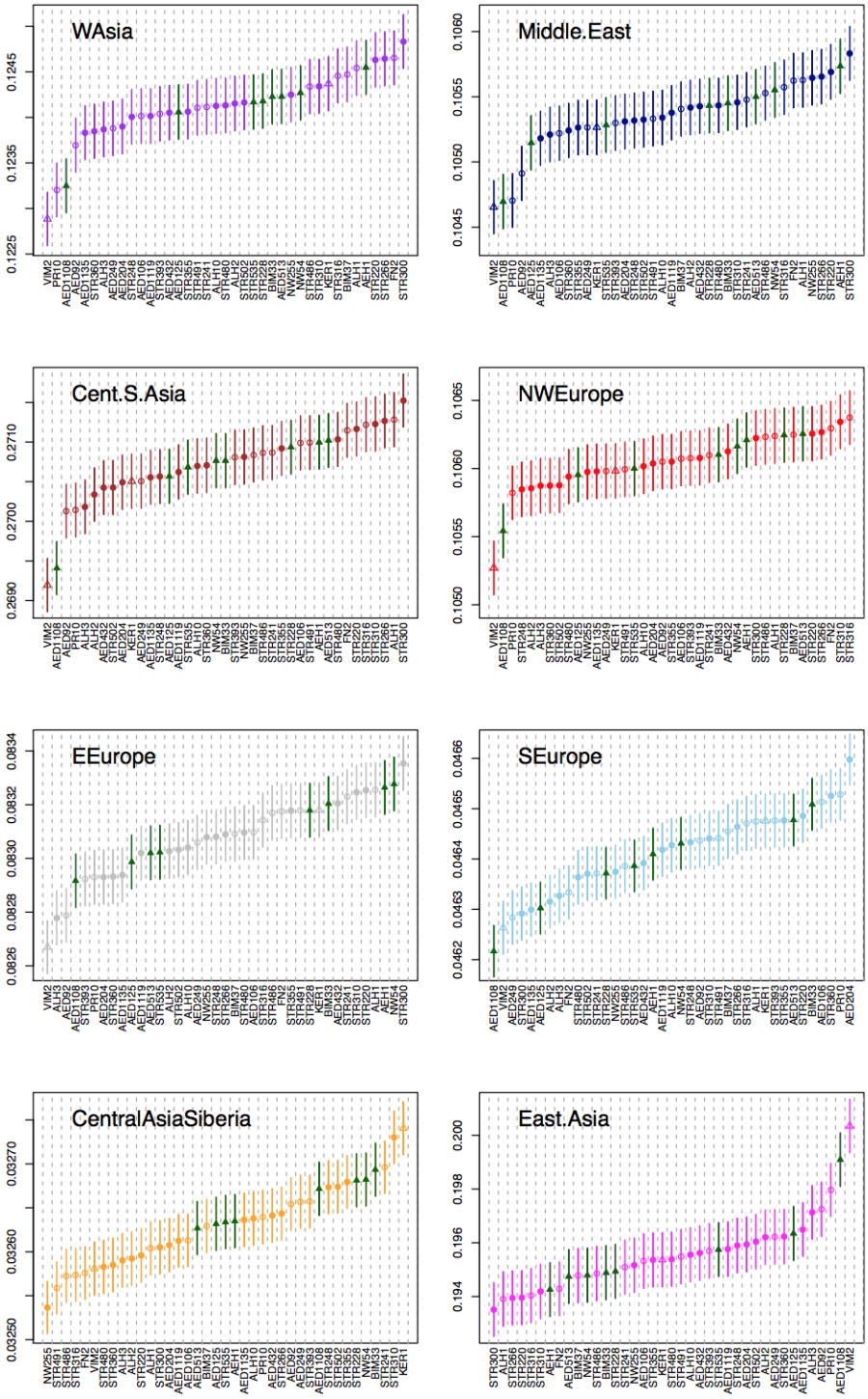


**Figure S44** Proportion of regional ancestry (y-axis) assigned to each ancient sample (x-axis) in the allele-sharing profiles using the HellBus dataset. Regions are grouped according to the clusters inferred in Hellenthal et al. (71) and provided at top-left with the colored points as in Fig. S42 and symbols as in Fig. S43. Ancient samples are ordered according to the inferred proportion for each set of donor contributions with individuals with deformed skulls are highlighted with a triangle. Female individuals with a deformed skull are depicted in dark green. Error bars give the standard error after jack-knife resampling dropping one chromosome at a time.

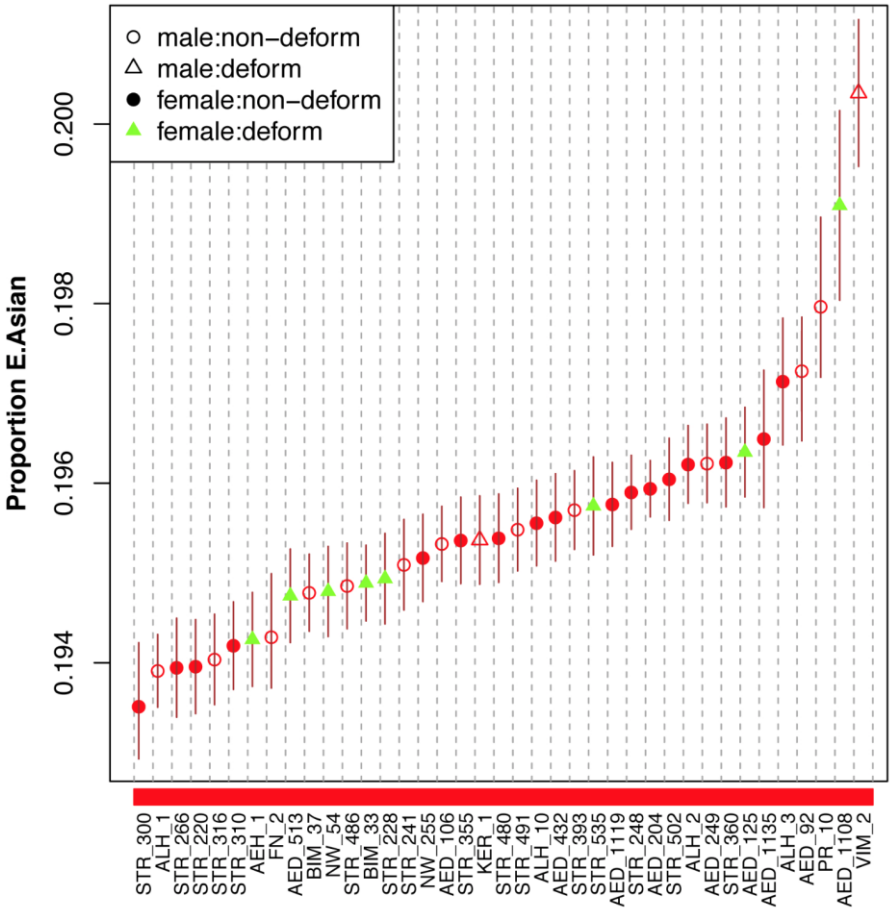


**Figure S45** Proportion of East Asian ancestry based on allele-sharing profiles inferred for the HellBus dataset. The axis at left provides the proportion of East Asian ancestry where East Asian includes contributions from: Ami, Atayal, Beijing_Han_Chinese, Cambodian, Chinese_Dai, Dai, Daur, Han, Hezhen, Japan, Japanese, Kinh, Korean, Kyrgyz, Lahu, Miao, Mongola, Naxi, Oroqen, She, Southern_Han_Chinese, Tu, Tujia, Uygur, Vietnam, Xibo, and Yi (as also provided in Fig. S44). Ancient samples are ordered according to the inferred proportion for each dataset with individuals with deformed skulls highlighted with a triangle. Female individuals with a deformed skull are colored green. As in Fig. S44, error bars give the standard error after jack-knife resampling dropping one chromosome at a time.

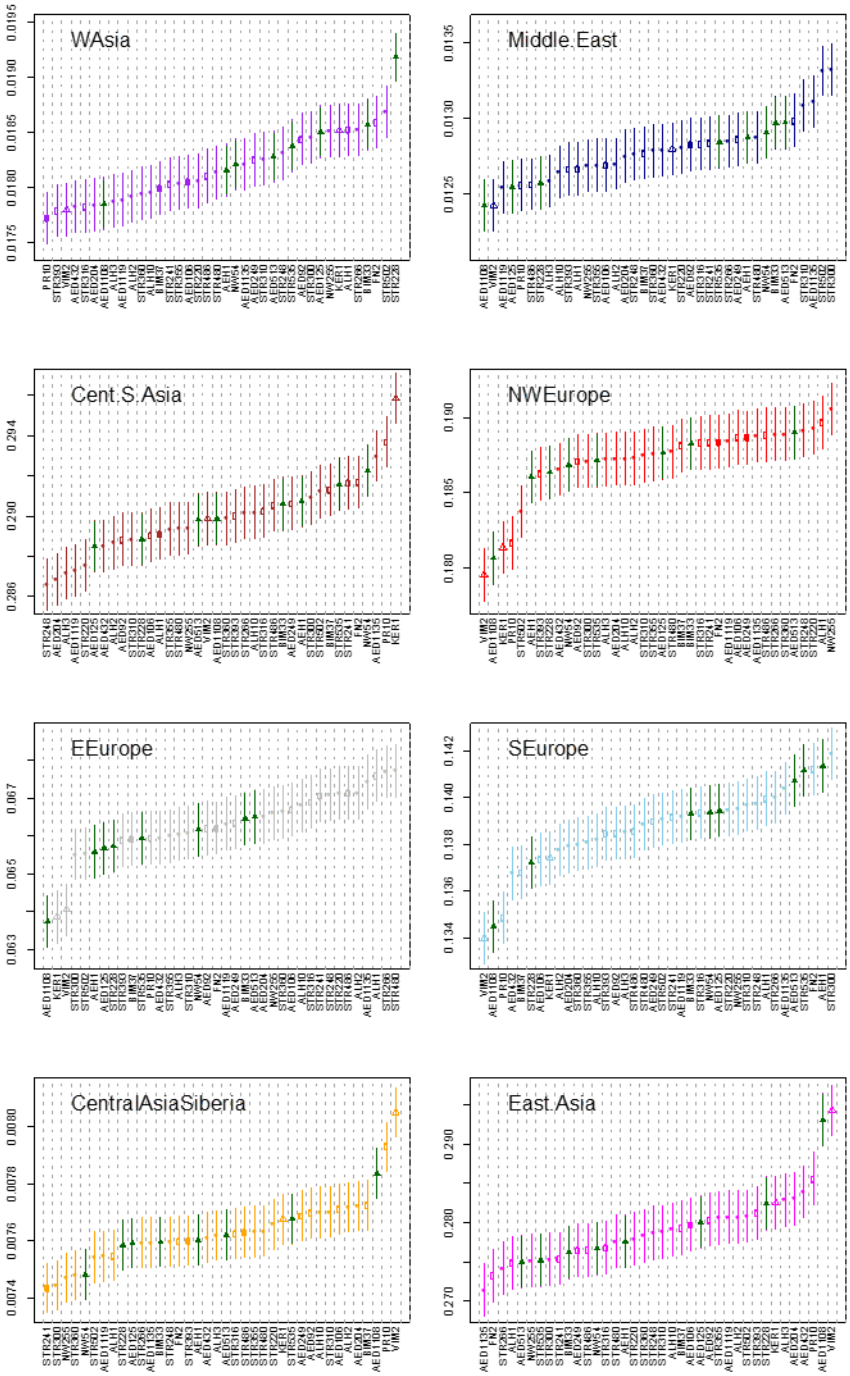


**Figure S46** Proportion of regional ancestry (y-axis) assigned to each ancient sample (x-axis) in the haplotype sharing profiles using the Phase_loci dataset. Regions are grouped according to the clusters inferred in Hellenthal et al. (71) and provided at top-left with the colored points as in Fig. S42 and symbols as in Fig. S43. Ancient samples are ordered according to the inferred proportion for each set of donor contributions with individuals with deformed skulls are highlighted with a triangle. Female individuals with a deformed skull are depicted in dark green. Error bars give the standard error after jack-knife resampling dropping one chromosome at a time.

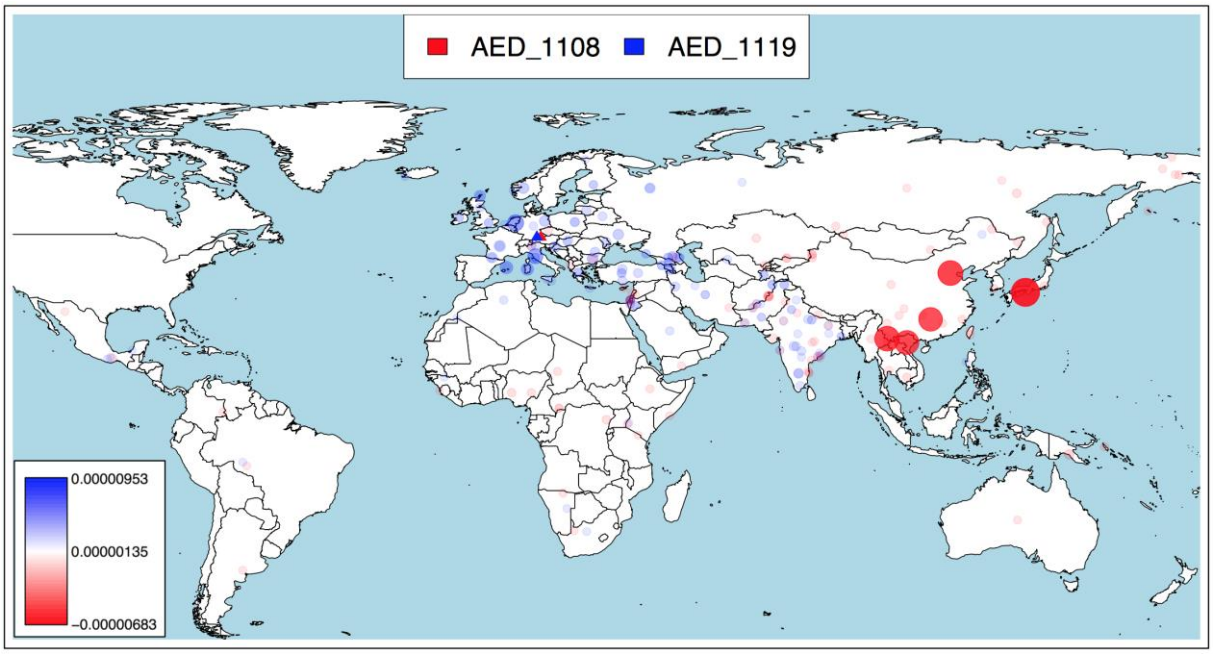


**Figure S47** Difference in CHROMOPAINTER inferred allele-sharing profiles for pairwise comparison of AED_1108 (red triangle - strong deformed phenotype) and AED_1119 (blue triangle - non-deformed phenotype) for the HellBus dataset. The size and color of the depicted circles corresponds to the degree of difference between the two samples being compared (so for example a higher relative proportion of matching in AED_1108 is indicated with darker red as given in the scale at bottom-left).

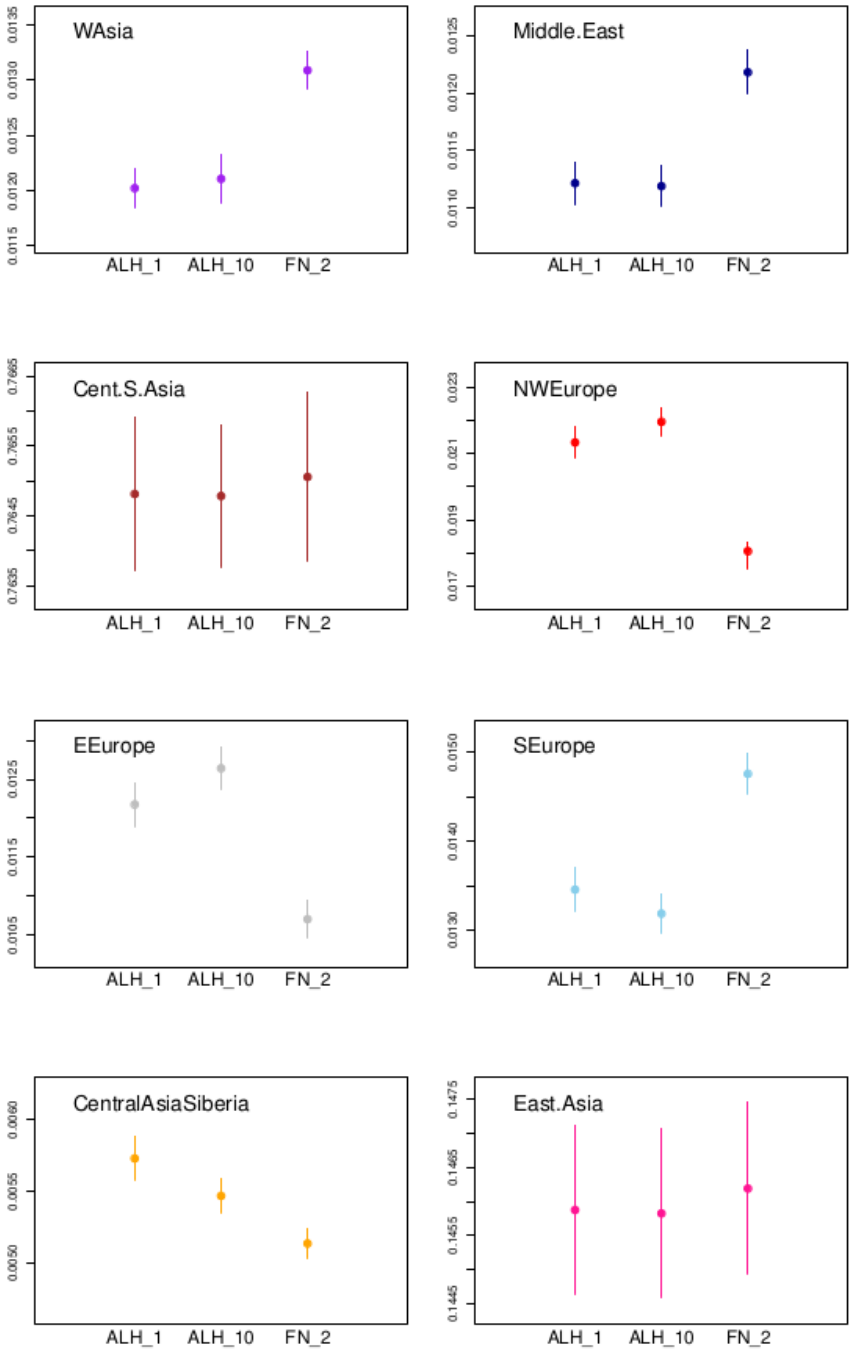


**Figure S48** Proportion of regional ancestry (y-axis) assigned to each ancient sample (x-axis) based on the haplotype-sharing profiles using the HellBus_SG dataset. Regions are grouped according to the clusters inferred in Hellenthal et al. (71) and provided at top-left with the colored points as in Fig. S42. Error bars give the standard error after jack-knife resampling dropping one chromosome at a time.

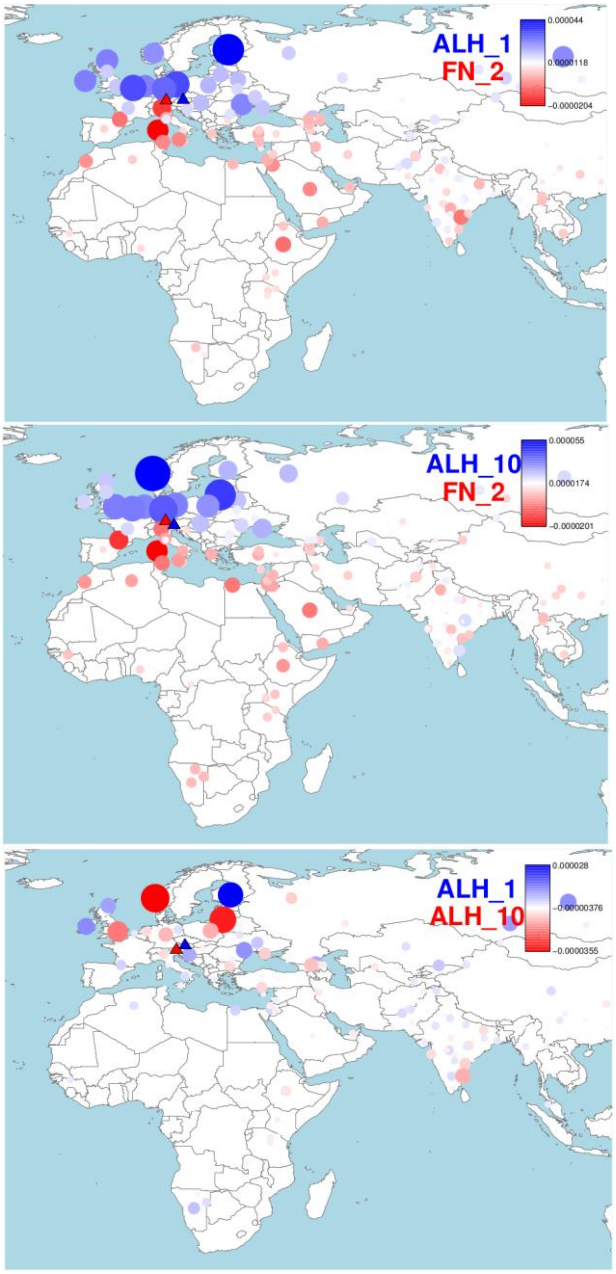


**Figure S49** Difference in CHROMOPAINTER inferred haplotype-sharing profiles for pairwise comparison of Altheim individuals ALH_1 and ALH_10 with the Roman individual FN_2 based on the HellBus_SG dataset painting profiles. The size and color of the depicted circles corresponds to the degree of difference between the two samples being compared as given in the scale at top-right.

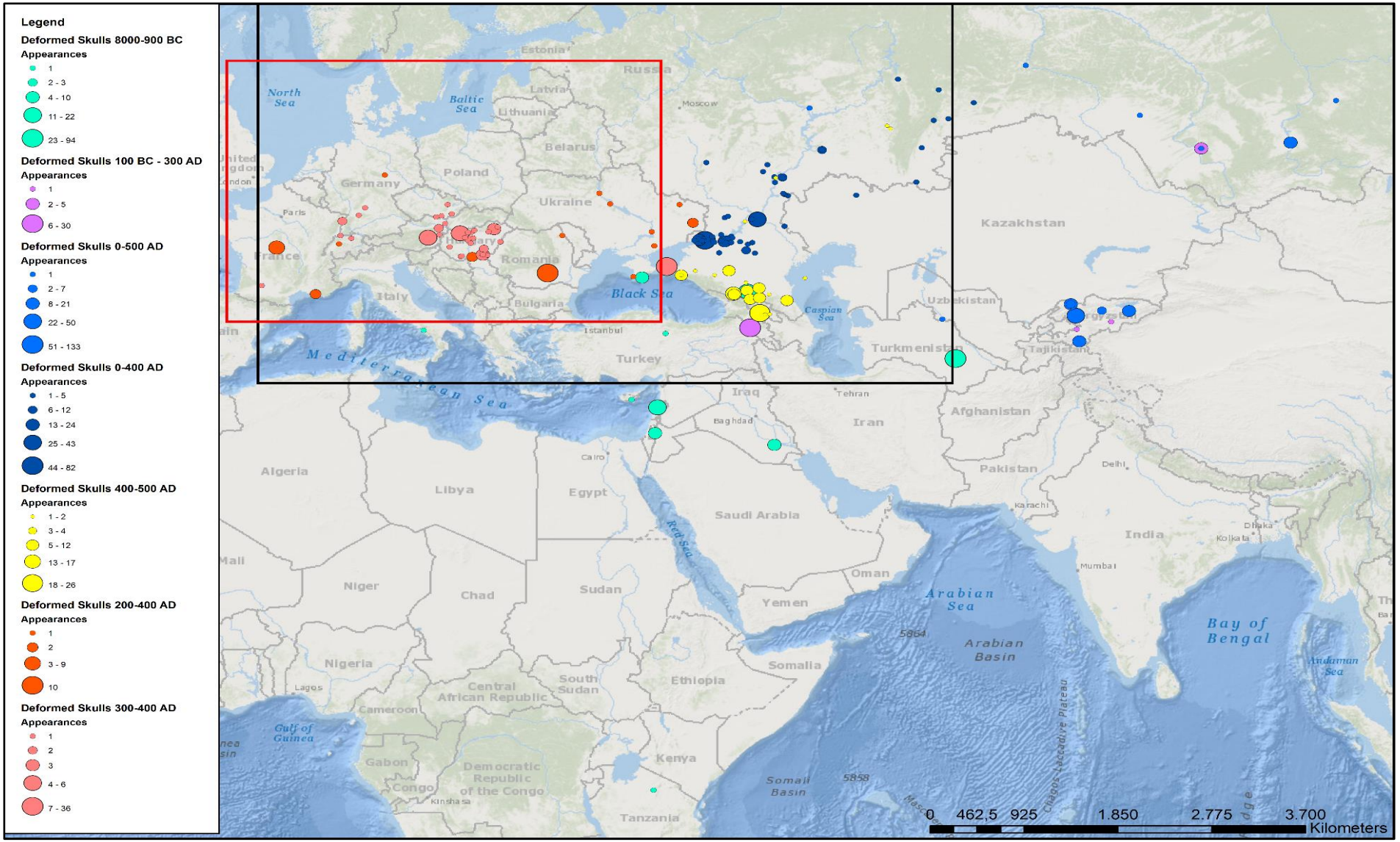
Legend
Deformed Skulls 8000-900 BC
Appearances
1
2 - 3
4 - 10
11 - 22
23 - 94
Deformed Skulls 100 BC - 300 AD
Appearances
1
2 - 5
6 - 30
Deformed Skulls 0-500 AD
Appearances
1
2 - 7
8 - 21
22 - 50
51 - 133
Deformed Skulls 0-400 AD
Appearances
1 - 5
6 - 12
13 - 24
25 - 43
44 - 82
Deformed Skulls 400-500 AD
Appearances
1 - 2
3 - 4
5 - 12
13 - 17
18 - 26
Deformed Skulls 200-400 AD
Appearances
1
2
3 - 9
10
Deformed Skulls 300-400 AD
Appearances
1
2
3
4 - 6
7 - 36
North Sea
Baltic Sea
Estonia
Latvia
Lithuania
Russia
Moscow
Belarus
Poland
Germany
Ukraine
Paris
Romania
Black Sea
Italy
Bulgaria
Istanbul
Turkey
Caspian Sea
Kazakhstan
Uzbekistan
Turkmenistan
Tajikistan
Mediterranean Sea
Iraq
Tehran
Iran
Afghanistan
Baghdad
Cairo
Algeria
Libya
Egypt
Pakistan
Delhi
Karachi
Saudi Arabia
Red Sea
India
Dhaka
Kolkata
Mumbai
Mali
Niger
Chad
Sudan
Oman
Yemen
Arabian Sea
Arabian Basin
5864
Bay of Bengal
Andaman Sea
Nigeria
Lagos
Somalia
Ethiopia
South Sudan
Central African Republic
Cameroon
Gulf of Guinea
Gabon
Congo
Kinshasa
Democratic Republic of the Congo
Kenya
Somali Basin
5858
Tanzania
Chagos-Laccadive Plateau
0 462,5 925 1.850 2.775 3.700
Kilometers

**Figure S50** Artificial skull deformation in Africa and Eurasia until 500 AD. The first appearances (turquoise) span an area from Tanzania up to Turkmenistan and the Crimean Peninsula. Around the beginning of the first millennium AD (purple) cranial deformation was practiced both in Central Asia and in the Caucasus. After that, some finds are unsharply dated to the first half of the first millennium AD (light blue) while others are dated between 0 and 400 AD (dark blue). Beginning in 400 AD artificial skull deformation shifts south into the Northern Caucasus (yellow). In central Europe the earliest crania with artificial deformation appear since 200 AD (orange), while most of the sites harboring artificially deformed skulls set in since 300 AD (light red). The distribution in the time between 0 and 500 AD (light blue and dark blue) looks like some sort of bifurcation, however, this could mainly be an artifact as basically all of Kazakhstan is spared which might be related to the current state of research there. Burial sites spanning two "time periods" were always included in the era when they set in. As for few burial sites the numbers of artificially deformed were unsharply defined as "several" and "numerous", approximate numbers were given to these sites after comparison to terms used for burial sites where clear numbers were given ("several" ≈ 10; "numerous" ≈ 50), locations without any definition of the quantity of artificially deformed skulls were included with count 1. Furthermore, the extents of Fig. S49 (black frame) and Fig. S50 (red frame) are indicated. Dating and determination of skull deformation is based on information given in the literature, which were not re-evalutated. Sources for artificially deformed skulls: Alt (161), Batieva (162), Bernštam (163), Dubova (164), Ginzburg & Žirov (9), Hotz (165), Kiszely (166), Sellenk (167), Wagner (168), Werner (36). Map created with ArcMap 10.5.

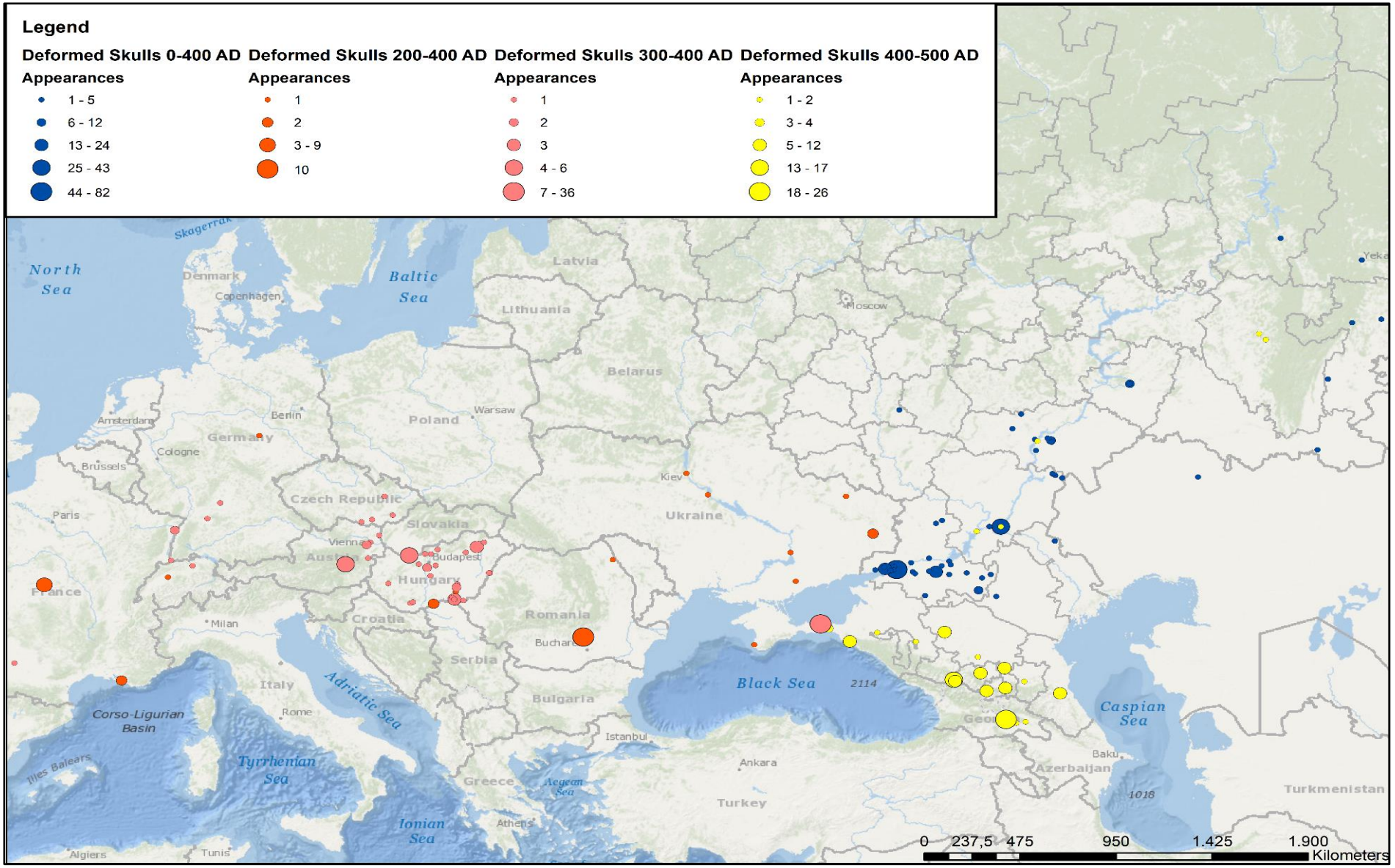
Legend
Deformed Skulls 0-400 AD
Appearances
1 - 5
6 - 12
13 - 24
25 - 43
44 - 82
Deformed Skulls 200-400 AD
Appearances
1
2
3 - 9
10
Deformed Skulls 300-400 AD
Appearances
1
2
3
4 - 6
7 - 36
Deformed Skulls 400-500 AD
Appearances
1 - 2
3 - 4
5 - 12
13 - 17
18 - 26
North Sea
Baltic Sea
Denmark
Copenhagen
Latvia
Lithuania
Moscow
Belarus
Poland
Warsaw
Berlin
Amsterdam
Germany
Cologne
Brussels
Paris
France
Czech Republic
Slovakia
Vienna
Budapest
Hungary
Kiev
Ukraine
Milan
Croatia
Romania
Serbia
Bulgaria
Black Sea
2114
Italy
Adriatic Sea
Corso-Ligurian Basin
Rome
Tyrrhenian Sea
Illes Baleares
Istanbul
Ankara
Greece
Aegean Sea
Turkey
Ionian Sea
Athens
Algiers
Tunis
Caspian Sea
Baku
Azerbaijan
1018
Turkmenistan
0 237,5 475 950 1.425 1.900
Kilometers

**Figure S51** Artificial skull deformation in eastern and central Europe. Earliest appearances of artificially deformed skulls are mainly located at the river Don (dark blue, see also Fig. S50). After the inroad of the Huns to that area, the appearance of artificially deformed skulls shifts towards the Caucasus Mountains (yellow). In Western Europe, only very few deformed skulls appear before the 4$^{th}$ century AD (orange), while the custom rises in appearance beginning in the 4$^{th}$ century AD (light red). Burial sites spanning two "time periods" were always included in the era when they set in. As for few burial sites the numbers of artificially deformed were unsharply defined as "several" and "numerous", approximate numbers were given to these sites after comparison to terms used for burial sites where clear numbers were given ("several" ≈ 10; "numerous" ≈ 50), locations without any definition of the quantity of artificially deformed skulls were included with count 1. Dating and determination of skull deformation is based on information given in the literature, which were not re-evalutated. Sources for artificially deformed skulls: Alt (161), Batieva (162), Bernštam (163), Dubova (164), Ginzburg & Žirov (9), Hotz (165), Kiszely (166), Sellenk (167), Wagner (168), Werner (36). Map created with ArcMap 10.5.

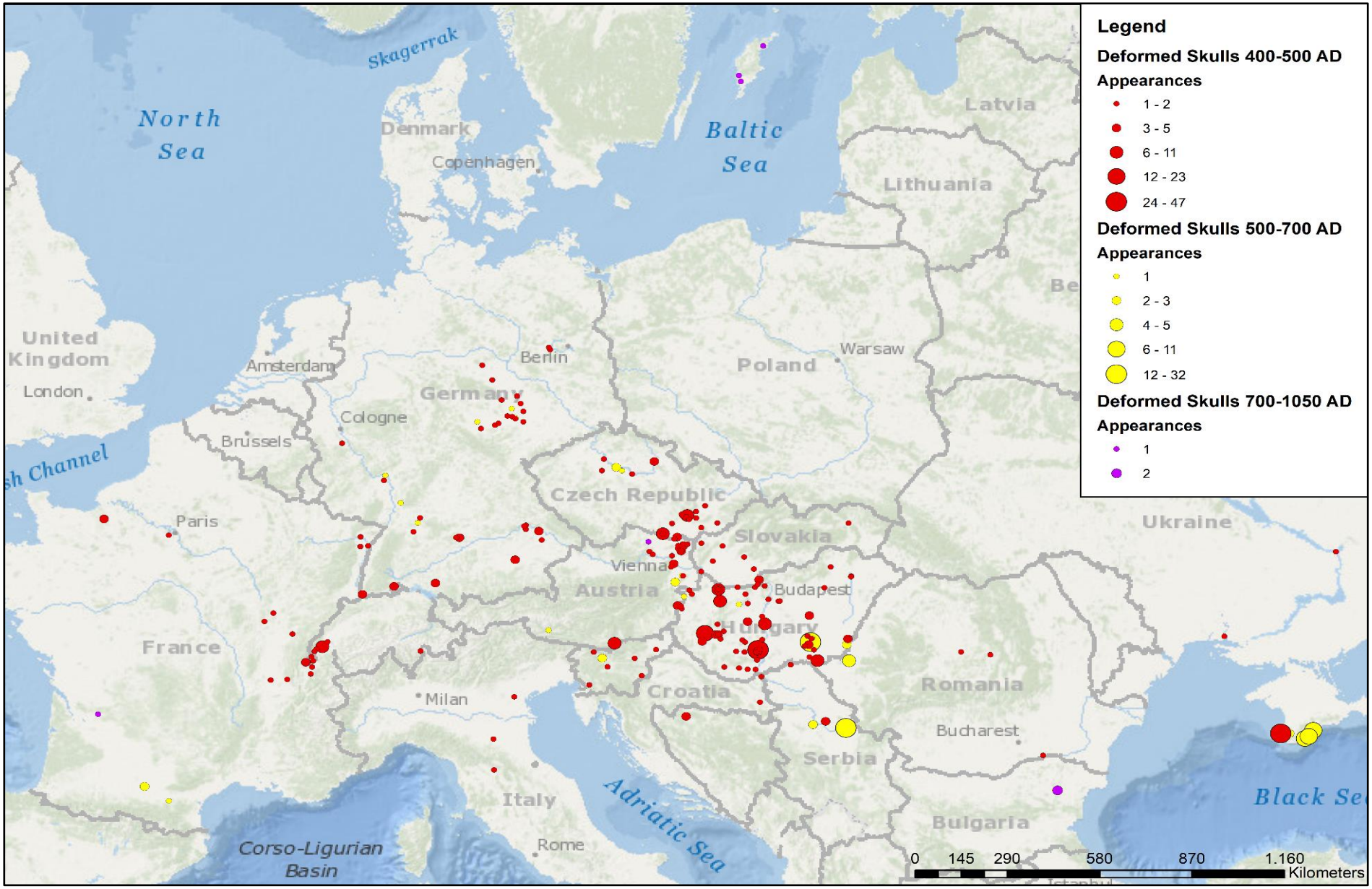
Legend
Deformed Skulls 400-500 AD
Appearances
1 - 2
3 - 5
6 - 11
12 - 23
24 - 47
Deformed Skulls 500-700 AD
Appearances
1
2 - 3
4 - 5
6 - 11
12 - 32
Deformed Skulls 700-1050 AD
Appearances
1
2
North Sea
Skagerrak
Denmark
Copenhagen
Baltic Sea
Latvia
Lithuania
United Kingdom
London
Amsterdam
Berlin
Germany
Cologne
Brussels
Poland
Warsaw
Czech Republic
Slovakia
Ukraine
Paris
Vienna
Austria
Budapest
Hungary
France
Milan
Croatia
Romania
Bucharest
Serbia
Italy
Adriatic Sea
Black Sea
Bulgaria
Corso-Ligurian Basin
Rome
0 145 290 580 870 1.160
Kilometers

**Figure S52** Appearances of artificially deformed skulls in central and western Europe. The time of the greatest distribution of this custom in Europe starts in the 5$^{th}$ century AD (red). While some of these sites also contain skulls dating to later centuries, the number of new sites with artificially deformed skulls decreases beginning in the 6$^{th}$ century AD (yellow and purple). Burial sites spanning two "time periods" were always included in the era when they set in. As for few burial sites the numbers of artificially deformed were unsharply defined as "several" and "numerous", approximate numbers were given to these sites after comparison to terms used for burial sites where clear numbers were given ("several" ≈ 10; "numerous" ≈ 50), locations without any definition of the quantity of artificially deformed skulls were included with count 1. Dating and determination of skull deformation is based on information given in the literature, which were not re-evalutated. Sources for artificially deformed skulls: Alt (161), Batieva (162), Bernštam (163), Dubova (164), Ginzburg & Žirov (9), Hotz (165), Kiszely (166), Sellenk (167), Wagner (168), Werner (36). Map created with ArcMap 10.5.

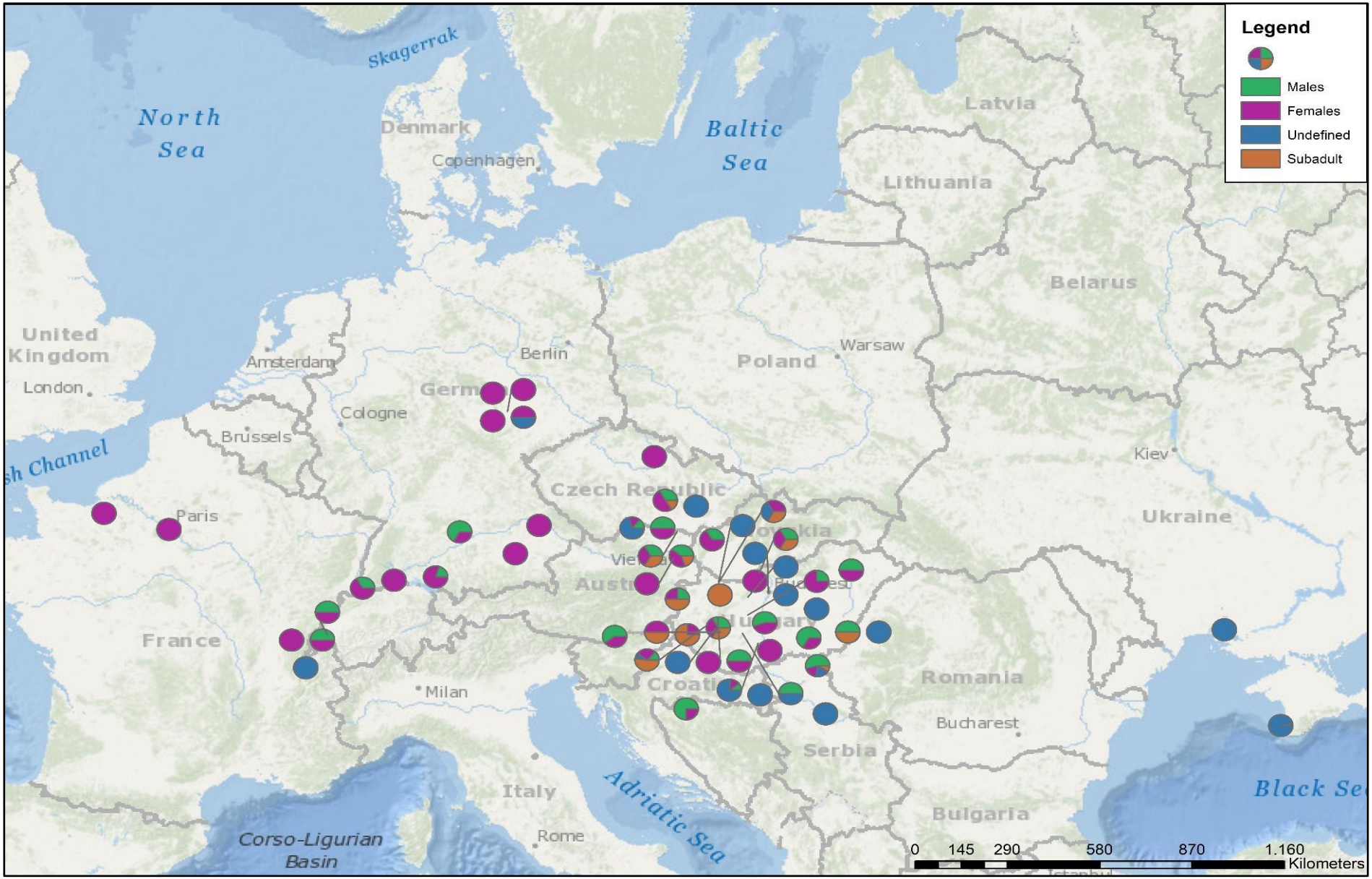
Legend
Males
Females
Undefined
Subadult
North Sea
Skagerrak
Denmark
Copenhagen
Baltic Sea
Latvia
Lithuania
Belarus
United Kingdom
London
Amsterdam
Berlin
Poland
Warsaw
Germ
Cologne
Brussels
Channel
Paris
Czech Republic
Kiev
Ukraine
Vie
Austr
France
Milan
Croati
Romania
Bucharest
Serbia
Italy
Adriatic Sea
Black Se
Bulgaria
Rome
Corso-Ligurian Basin
0 145 290 580 870 1.160
Kilometers

**Figure S53** Distribution of female, male and subadult individuals with artificially deformed crania in central Europe. The map shows all sites in the time frame 400 – 500 AD (see Fig. S52) from which at least 2 individuals with artificially deformed skulls could be recovered. While, except for a few sites at France's eastern border, in Western Europe mainly adult female individuals exhibit artificially deformed skulls, the proportion of male and subadult individuals showing this modification increases in the more eastern sites in and around Hungary. Dating and determination of skull deformation is based on information given in the literature, which were not re-evalutated. Sources for artificially deformed skulls: Alt (161), Batieva (162), Bernštam (163), Dubova (164), Ginzburg & Žirov (9), Hotz (165), Kiszely (166), Sellenk (167), Wagner (168), Werner (36). Map created with ArcMap 10.5.

## SI Appendix tables

**Table S1** Overview of the samples of the present study. The table lists the site, grave and sample number as well as the basic morphological data (sex, age at death, status of skull deformation) for each individual. Furthermore an overview of the grave goods is given and, where possible a "cultural label" is assigned to each individual. Datings of the samples are either based on archaeological dating (AD) or on $^{14}$C dating (cal AD). *information only given for skulls that were formerly discussed to be deformed, individuals without information show no skull deformation (7); °$^{14}$C dates taken from Trautmann et al. (7); M=male, F=female.

| site | grave no. | sample no. | sex | age-at-death | artificially deformed * | Grave good | Cultural label of bow-brooches | $^{14}$C-date° | Dating according to grave finds and $^{14}$C dates° |
|---|---|---|---|---|---|---|---|---|---|
| **Straubing-Bajuwarenstraße** | 220 | STR_220 | F | 30-50 | intermediate | 2 bow brooches, 2 eagle brooches, beads, chatelaine | ostrogothic | | 480-530 AD |
| **Straubing-Bajuwarenstraße** | 228 | STR_228 | F | 30-50 | deformed | disturbed grave | | 1606 ± 23 | 382 – 546 cal AD |
| **Straubing-Bajuwarenstraße** | 241 | STR_241 | M | 50-60 | | belt, bag? | | | 480 – 510 AD |
| **Straubing-Bajuwarenstraße** | 248 | STR_248 | F | 40-60 | | 2 S-shaped brooches, chatelaine, beads | | | 530 – 560 AD |
| **Straubing-Bajuwarenstraße** | 266 | STR_266 | F | 20-30 | | 2 brooches, 2 bird brooches, 2 earrings, needle, beads, chatelaine | ostrogothic | | 510 – 530 AD |
| **Straubing-Bajuwarenstraße** | 300 | STR_300 | F | 40-60 | | 2 bow brooches, chatelaine | alamannic-frankish | | 510 – 530 AD |
| **Straubing-Bajuwarenstraße** | 310 | STR_310 | F | 60+ | intermediate | 2 bow brooches | nordic type | | 510 – 530 AD |
| **Straubing-Bajuwarenstraße** | 316 | STR_316 | M | 40-60 | | belt, bag, vessel | | | 480 – 510 AD |
| **Straubing-Bajuwarenstraße** | 328 | STR_328 | F | 40-60 | deformed | 2 bow brooches | ostrogothic | 1635 ± 19 | 350 – 367 cal. AD<br>380 – 432 cal. AD<br>491 – 530 cal. AD<br>480 – 510 AD |
| **Straubing-Bajuwarenstraße** | 355 | STR_355 | F | 30-50 | intermediate | 2 bow brooches, 2 bird brooches, needle, chatelaine, beads | Alamannic-Frankish | | 510 – 530 AD |
| **Straubing-** | 360 | STR_360 | F | 50-60 | intermediate | 2 bow brooches, 1 bird | Alamannic - | | 510 – 530 AD |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Bajuwarenstraße** | | | | | | brooch, beads, chatelaine | Frankish | | |
| **Straubing-Bajuwarenstraße** | 393 | STR_393 | M | 40-55 | | spatha | | | 460 – 530 AD |
| **Straubing-Bajuwarenstraße** | 480 | STR_480 | F | 60+ | | 2 S-shaped brooches, chatelaine, beads | | | 510 – 530 AD |
| **Straubing-Bajuwarenstraße** | 486 | STR_486 | M | 60+ | | spatha, seax, belt, bag | | | 460 – 530 AD |
| **Straubing-Bajuwarenstraße** | 491 | STR_491 | M | 60+ | intermediate | spatha, lance, belt, bag | | | 510 – 530 AD |
| **Straubing-Bajuwarenstraße** | 502 | STR_502 | F | 60+ | not deformed | 2 earrings, beads, spindle whorl | | 1492 ± 24 | 430 – 492 cal AD<br>530 – 647 cal AD<br>580 – 620 AD |
| **Straubing-Bajuwarenstraße** | 535 | STR_535 | F | 50-60 | deformed | 2 bow brooches , 2 bird brooches, chatelaine, beads | nordic type | 1625 ± 19 | 348 – 368 cal AD<br>380 – 538 cal AD<br>510 – 530 AD |
| **Altenerding-Klettham** | 92 | AED_92 | M | 20-30 | | spatha, seax, lance, shield, belt, bag | | | 480 – 510 AD |
| **Altenerding-Klettham** | 106 | AED_106 | M | 60+ | | spatha, belt, bag, vessel, glass | | | 480 – 510 AD |
| **Altenerding-Klettham** | 125 | AED_125 | F | 40-60 | deformed | no grave goods | | 1636 ± 20 | 338 – 475 cal AD<br>485 – 535 cal AD |
| **Altenerding-Klettham** | 204 | AED_204 | F | 60+ | | 2 bird brooches, beads, chatelaine | | | 480 – 510 AD |
| **Altenerding-Klettham** | 249 | AED_249 | M | 40-55 | | belt, bag, arrowheads | | | 460 – 510 AD |
| **Altenerding-Klettham** | 432 | AED_432 | F | 60+ | not deformed | disturbed grave;<br>1 S-shaped brooch | | | 480 – 530 AD |
| **Altenerding-Klettham** | 513 | AED_513 | F | 30-50 | deformed | spindle whorl, belt | | 1649 ± 23 | 261 – 279 cal AD<br>326 – 475 cal AD<br>485 – 535 cal AD<br>480 – 510 AD |
| **Altenerding-Klettham** | 1108 | AED_1108 | F | 60+ | deformed | 1 bow brooch, chatelaine, comb | Thuringian | 1601 ± 19 | 355 – 365 cal AD<br>381 – 433 cal AD<br>489 – 532 cal AD<br>460 – 510 AD |
| **Altenerding-Klettham** | 1119 | AED_1119 | F | 30-45 | | disturbed grave;<br>1 bow brooch, chatelaine, beads | Alamannic-Frankish | | 510 – 530 AD |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Altenerding-Klettham** | 1135 | AED_1135 | F | 30-50 | not deformed | 2 bird brooches, chatelaine, beads | | | 480 – 530 AD |
| **Altheim** | 55 | ALH_10 | F | 30-50 | | no grave goods | | 1619 ± 22 | 387 – 475 cal AD<br>485 – 535 cal AD |
| **Altheim** | 73 | ALH_2 | F | 40-55 | not deformed | disturbed grave | | 1651 ± 40 | 260 – 535 cal AD |
| **Altheim** | 74 | ALH_3 | F | 30-50 | | disturbed grave | | 1631 ± 19 | 356 – 535 cal. AD<br>480 – 510 AD |
| **Altheim** | 84 | ALH_1 | M | 50-60 | | disturbed grave | | | Non-datebable |
| **Barbing-Irlmauth** | „33“ | BIM_33 | F | 30-50 | deformed | no grave goods attributable | | 1616 ± 23 | 349 – 368 cal AD<br>380 – 542 cal AD |
| **Barbing-Irlmauth** | 37 | BIM_37 | M | 60+ | | spatha, belt, bag | | | 580 – 620 AD |
| **Alteglofsheim** | 145 | AEH_1 | F | 30-50 | deformed | 1 bow brooch, bracelet | Ostrogothic | 1639 ± 20 | 346 – 370 cal AD<br>377 – 430 cal AD<br>492 – 529 cal AD<br>480 – 510 AD |
| **Burgweinting – Nord-West II** | 10254 | NW_54 | F | 38-45 | deformed | 1 Pelta brooch | | 1624 ± 20 | 388 – 532 cal AD |
| **Burgweinting – Nord-West II** | 10255 | NW_255 | F | 45-60 | | 2 silver-gilt bow brooches, 2 bird brooches | Alamannic-Frankish | 1594 ± 23 | 417 – 536 cal AD |

**Table S2** Results of quality assessment with shallow MiSeq sequencing and quantitative real-time PCR of library fill-in products (mD5 = deamination rate of first position at 5' molecule end, C/T).

| sample | sex | skull deformation | unique library molecules | endogenous DNA content [%] | aligned reads Mkdup | mD5 |
|---|---|---|---|---|---|---|
| AED_106 | m | non-def | $1.06x10^{9}$ | 66.51 | 255372 | 0.28 |
| AED_1108 | f | def | $1.73x10^{8}$ | 62.52 | 143822 | 0.22 |
| AED_1119 | f | non-def | $1.62x10^{9}$ | 58.7 | 576008 | 0.26 |
| AED_1135 | f | non-def | $5.68x10^{8}$ | 55.72 | 438996 | 0.29 |
| AED_125 | f | def | $2.30x10^{9}$ | 21.65 | 152871 | 0.30 |
| AED_204 | f | non-def | $1.84x10^{9}$ | 31.84 | 259558 | 0.23 |
| AED_249 | m | non-def | $4.57x10^{8}$ | 63.14 | 301736 | 0.26 |
| AED_432 | f | non-def | $2.89x10^{9}$ | 60.04 | 242187 | 0.30 |
| AED_513 | f | def | $4.17x10^{8}$ | 69.57 | 269402 | 0.19 |
| AED_92 | m | non-def | $1.49x10^{9}$ | 57.06 | 460128 | 0.28 |
| AEH_I | f | def | $1.02x10^{9}$ | 27.78 | 85444 | 0.27 |
| Alh_1 | f | non-def | $1.79x10^{9}$ | 42.29 | 497240 | 0.28 |
| Alh_10 | m | non-def | $1.59x10^{9}$ | 70.63 | 311719 | 0.24 |
| Alh_2 | f | non-def | $4.38x10^{8}$ | 13.19 | 80909 | 0.13 |
| Alh_3 | f | non-def | $6.11x10^{8}$ | 58.28 | 308223 | 0.14 |
| BIM_33 | f | def | $3.49x10^{8}$ | 72.64 | 369168 | 0.21 |
| BIM_37 | m | non-def | $2.38x10^{8}$ | 41.45 | 389009 | 0.21 |
| NW_255 | f | non-def | $3.03x10^{9}$ | 67.34 | 431987 | 0.30 |
| NW_54 | f | def | $5.35x10^{9}$ | 15.65 | 67865 | 0.35 |
| STR_220 | f | interm. | $1.26x10^{8}$ | 66.75 | 223146 | 0.21 |
| STR_228 | f | def | $4.71x10^{8}$ | 67.86 | 445253 | 0.29 |
| STR_241 | m | non-def | $2.39x10^{9}$ | 28.58 | 119490 | 0.23 |
| STR_248 | f | non-def | $2.78x10^{9}$ | 30.4 | 215236 | 0.23 |
| STR_266 | f | non-def | $4.56x10^{9}$ | 27.66 | 164990 | 0.24 |
| STR_300 | f | non-def | $2.69x10^{9}$ | 40.64 | 233026 | 0.30 |
| STR_310 | f | interm. | $4.34x10^{8}$ | 67.49 | 270427 | 0.25 |
| STR_316 | m | non-def | $3.69x10^{8}$ | 69.29 | 253100 | 0.22 |
| STR_328 | f | def | $6.61x10^{8}$ | 66.5 | 345320 | 0.23 |
| STR_355 | f | interm. | $2.61x10^{9}$ | 63.97 | 227284 | 0.09 |
| STR_360 | f | interm. | $1.31x10^{9}$ | 20.93 | 79885 | 0.22 |
| STR_393 | m | non-def | $5.88x10^{8}$ | 35.46 | 277044 | 0.30 |
| STR_480 | f | non-def | $3.21x10^{8}$ | 69.12 | 247233 | 0.19 |
| STR_486 | m | non-def | $2.01x10^{9}$ | 30.22 | 227868 | 0.24 |
| STR_491 | m | interm. | $5.33x10^{9}$ | 15.63 | 64224 | 0.25 |
| STR_502 | f | non-def | $3.06x10^{8}$ | 45.51 | 170147 | 0.33 |
| STR_535 | f | def | $2.57x10^{8}$ | 71.67 | 230403 | 0.22 |
| FN_2 | m | non-def | $4.72x10^{9}$ | 52.37 | 399463 | 0.29 |
| KER_I | m | def | $2.30x10^{8}$ | 66.15 | 258420 | 0.22 |
| Vim_2 | m | def | $1.82x10^{9}$ | 66.79 | 420021 | 0.22 |
| PR_10 | m | non-def | $7.54x10^{7}$ | 16.1 | 113769 | 0.08 |
| PR_4 | m | non-def | $1.03x10^{8}$ | 10.41 | 68161 | 0.12 |

**Table S3** Overview of libraries that were produced per sample. Library code: S = screened library; mt = library also used for mitochondrial capture, USER = library treated with USER enzyme. PCR cycles: Taq PCR/ Herculase PCR/ Reamplification with Herculase. For nuclear capture all libraries were used.

| Sample | Extraction Code | Libraries produced | PCR cycles |
|---|---|---|---|
| AED_92 | B_EX_4_2 | LM6_2$^{S}$ | 10/14/0 |
| | | LBC8_1$^{mt}$, LBC8_2$^{mt}$, LBC8_3 | 10/12/0 |
| | | LBC18_1, LBC18_2$^{mt}$, LBC18_3$^{mt}$, LBC18_4 | 16/0/0 |
| AED_106 | B_EX_6_4 | LM7_3$^{S,mt}$ , LBC10_10$^{mt}$ , LBC10_11$^{mt}$ , LBC10_12, LBC15_10, LBC15_11$^{mt}$ | 10/12/0 |
| AED_125 | B_EX_4_10 | LM6_9$^{S}$ | 10/14/0 |
| | | LBC8_4, LBC8_5$^{mt}$, LBC8_6$^{mt}$ | 10/12/0 |
| | | LBC18_5, LBC18_6$^{mt}$, LBC18_7$^{mt}$, LBC18_8 | 16/0/0 |
| AED_204 | B_EX_5_10 | LM5_10$^{S,mt}$, LBC9_7$^{mt}$, LBC9_9$^{mt}$ | 10/12/0 |
| | | LBC9_8$^{mt}$ | 10/12/6 |
| AED_249 | B_EX_6_5 | LM7_4$^{S,mt}$, LBC15_5$^{mt}$, LBC15_6$^{mt}$ | 10/12/0 |
| | | LBC10_13, LBC10_14$^{mt}$, LBC10_15 | 10/12/10 |
| AED_432 | B_EX_6_2 | LM7_11$^{S,mt}$ | 10/14/0 |
| | | LBC11_6, LBC11_7$^{mt}$, LBC11_8$^{mt}$, LBC16_12$^{USER,mt}$, LBC16_13$^{USER,mt}$ | 10/12/0 |
| AED_513 | B_EX_6_3 | LM7_2$^{S,mt}$ | 10/14/0 |
| | | LBC10_4$^{mt}$ | 10/12/0 |
| | | LBC10_5$^{mt}$ , LBC10_6$^{mt}$ | 10/12/6 |
| AED_1108 | B_EX_5_9 | LM5_9$^{S}$ | 10/14/0 |
| AED_1108 | B_EX_7_1 | LM8_1$^{S,mt}$ | 10/14/0 |
| | | LBC12_7, LBC12_8$^{mt}$, LBC12_9$^{mt}$ | 10/12/5 |
| | | LBC17_12$^{mt}$ | 10/12/0 |
| AED_1119 | B_EX_4_11 | LM6_10$^{S,mt}$ | 10/12/0 |
| | | LBC14_1$^{mt}$ | 10/12/4 |
| AED_1119 | B_EX_8_6 | LBC14_2$^{mt}$ , LBC14_3$^{mt}$ | 10/12/4 |
| | | LBC20_1, LBC20_2 | 16/0/0 |
| AED_1135 | B_EX_5_5 | LM5_5$^{S,mt}$ | 10/14/0 |
| | | LBC9_4$^{mt}$, LBC9_5$^{m}$, LBC9_6$^{mt\ t}$ | 10/12/4 |
| AEH_I | B_EX_5_2 | LM5_2$^{S,mt}$ | 10/12/0 |
| | | LBC9_1$^{mt}$, LBC9_2$^{mt}$, LBC9_3$^{mt}$ | 10/12/5 |
| ALH_1 | EX_4 | LM4_2$^{S}$, LBC4_1, LBC4_2, LBC4_3, LC3_1, LC3_2, LC3_3 | 10/9/0 |
| ALH_2 | EX_2 | LM2_7$^{S,mt}$, LBC5_1$^{mt}$, LBC6_1$^{mt}$, LBC6_2$^{mt}$, LBC7_1$^{mt}$, LBC7_2$^{mt}$ | 10/9/0 |
| ALH_3 | EX_1 | LM1_4$^{S,mt}$, LBC5_3, LBC5_4$^{mt}$, LBC6_6, LBC7_3$^{mt}$, LBC7_4$^{mt}$ | 10/9/0 |

| | | | |
|---|---|---|---|
| ALH_10 | EX_4 | LM4_5$^{S}$ | 10/9/0 |
| | | LBC4_7, LBC4_8 | 12/8/0 |
| | | LC3_7, LC3_8, LC3_9, LC3_10 | 10/9/0 |
| BIM_33 | B_EX_6_8 | LM7_7$^{S,mt}$, LBC15_4 | 10/12/0 |
| | | LBC10_7$^{mt}$ , LBC10_8, LBC10_9$^{mt}$, LBC15_3$^{mt}$ | 10/12/8 |
| BIM_37 | B_EX_4_4 | LM6_4$^{S,mt}$ | 10/14/0 |
| | | LBC14_13$^{mt}$ | 10/12/5 |
| | B_EX_8_3 | LBC14_14$^{mt}$ | 10/12/5 |
| | | LBC14_15$^{mt}$ | 10/12/0 |
| FN_2 | B_EX_5_12 | LM5_12$^{S}$, LBC8_7 | 10/9/0 |
| | | LBC8_8, LBC8_9, LBC8_10, LBC8_11, LBC8_12 | 10/12/0 |
| KER_1 | B_EX_6_6 | LM7_5$^{S,mt}$ | 10/14/0 |
| | | LBC11_1$^{mt}$, LBC11_2 | 10/12/6 |
| | B_EX_8_9 | LBC11_12 | 10/12/6 |
| | | LBC16_10$^{mt}$, LBC16_11$^{t}$ | 10/12/0 |
| | | LBC17_8$^{mt}$ | 16/0/0 |
| NW_54 | B_EX_6_7 | LM7_6$^{S,mt}$, LBC10_1$^{mt}$, LBC10_2$^{mt}$, LBC10_3$^{mt}$ | 10/12/0 |
| NW_255 | B_EX_8_2 | LBC11_15$^{S,mt}$ | 10/12/0 |
| | | LBC17_4$^{mt}$, LBC17_5$^{mt}$, LBC17_6$^{mt}$, LBC18_11, LBC18_12 | 16/0/0 |
| PR_4 | Ex18 | LBC15_15 | 10/12/5 |
| | | LBC16_1, LBC16_2, LBC16_3, LBC16_4 | 10/12/6 |
| | Ex11 | LBC17_15 | 16/0/6 |
| PR_10 | Ex18 | LBC16_5, LBC16_6 | 10/12/6 |
| | Ex12 | LBC16_7, LBC16_8 | 10/12/6 |
| STR_220 | B_EX_6_9 | LM7_8$^{S,mt}$ | 10/14/0 |
| | | LBC13_4$^{mt}$, LBC13_5, LBC13_6$^{mt}$ | 10/12/8 |
| | | LBC17_13$^{mt}$ | 16/0/0 |
| STR_228 | B_EX_4_14 | LM6_12$^{S,mt}$, LBC15_7$^{mt}$, LBC15_8 | 10/12/0 |
| | B_EX_8_8 | LBC12_4$^{mt}$, LBC12_5$^{mt}$, LBC12_6 | 10/12/0 |
| STR_241 | B_EX_7_9 | LM8_9$^{S,mt}$, LBC14_4$^{mt}$, LBC14_5$^{mt}$, LBC14_6$^{mt}$ | 10/12/0 |
| STR_248 | B_EX_5_1 | LM5_1$^{S,mt}$, LBC9_10, LBC9_11$^{mt}$, LBC9_12$^{mt}$, LBC16_9$^{mt}$ | 10/12/0 |
| STR_266 | B_EX_4_13 | LM6_11$^{S,mt}$, LBC12_10$^{mt}$, LBC15_9$^{mt}$ | 10/12/0 |
| | B_EX_8_7 | LBC12_11, LBC12_12$^{mt}$ | 10/12/0 |
| STR_300 | B_EX_4_7 | LM6_6$^{S,mt}$, LBC14_10, LBC14_11$^{mt}$, LBC14_12$^{mt}$, LBC16_14$^{USER}$, LBC16_15$^{USER,mt}$ | 10/12/0 |
| STR_310 | B_EX_6_10 | LM7_9$^{S,mt}$ | 10/14/0 |

| | | | |
|---|---|---|---|
| | | LBC12_13$^{mt}$ | 10/12/8 |
| | | LBC12_14, LBC12_15$^{mt}$, LBC15_12, LBC15_14$^{mt}$ | 10/12/0 |
| | | LBC15_13 | 10/12/5 |
| STR_316 | B_EX_7_10 | LM8_10$^{S}$ | 10/12/0 |
| | | LBC14_7$^{mt}$, LBC14_8$^{mt}$, LBC14_9$^{mt}$ | 10/12/8 |
| | | LBC17_14$^{mt}$ | 16/0/0 |
| STR_328 | B_EX_7_3 | LM8_3$^{S,mt}$ | 10/14/0 |
| | | LBC13_7$^{mt}$, LBC13_8$^{mt}$, LBC13_9$^{mt}$ | 10/12/8 |
| | | LBC17_16 | 16/0/0 |
| STR_355 | B_EX_5_8 | LM5_8$^{S}$ | 10/9/0 |
| STR_355 | B_EX_7_11 | LM8_11$^{S,mt}$, LBC13_1$^{mt}$, LBC13_2$^{mt}$, LBC13_3$^{mt}$ | 10/12/0 |
| STR_360 | B_EX_7_4 | LM8_4$^{S,mt}$, LBC13_13$^{mt}$, LBC13_14$^{mt}$, LBC13_15$^{mt}$ | 10/12/0 |
| STR_393 | B_EX_5_4 | LM5_4$^{S,mt}$ | 10/14/0 |
| | | LM8_12, LM8_13$^{mt}$ | 10/9/5 |
| | | LBC15_1, LBC15_2$^{mt}$ | 10/12/5 |
| | | LBC17_7$^{mt}$ | 16/0/0 |
| STR_480 | B_EX_7_8 | LM8_8$^{S,mt}$ | 10/14/0 |
| | | LBC11_9, LBC11_10$^{mt}$, LBC11_11$^{mt}$ | 10/12/0 |
| | | LBC17_11$^{mt}$ | 16/0/0 |
| STR_486 | B_EX_5_3 | LM5_3$^{S,mt}$, LBC9_13$^{mt}$, LBC9_14$^{mt}$, LBC9_15$^{mt}$ | 10/12/0 |
| STR_491 | B_EX_7_6 | LM8_6$^{S,mt}$, LBC12_1$^{mt}$, LBC12_2$^{mt}$, LBC12_3$^{mt}$ | 10/12/0 |
| STR_502 | B_EX_7_2 | LM8_2$^{S,mt}$ | 10/14/0 |
| | | LBC13_10$^{mt}$, LBC13_11$^{mt}$, LBC13_12$^{mt}$ | 10/12/0 |
| STR_535 | B_EX_7_7 | LM8_7$^{S,mt}$ | 10/14/0 |
| | | LBC11_3$^{mt}$, LBC11_4$^{mt}$, LBC11_5 | 10/12/0 |
| | | LBC17_9, LBC17_10$^{mt}$ | 16/0/0 |
| VIM_2 | EX_4 | LM4_7$^{S,mt}$, LBC4_4, LBC4_5, LBC4_6$^{mt}$, LC3_4$^{mt}$, LC3_5$^{mt}$, LC3_6 | 10/9/0 |

**Table S4** Overview of pooling and sequencing strategy for nuclear capture and shotgun sequencing. NE = nuclear enrichment; PE = paired end; SE = single end; all nuclear enrichment products are sequenced with 100 bp and PE mode.

| Sample | n libraries pooled for NE | starting quantity NE (ng) | seq. mode NE | n libraries pooled for shotgun seq. | seq. mode shotgun |
|---|---|---|---|---|---|
| AED_106 | 6 | 978.51 | HS | - | - |
| AED_1108 | 5 | 924.30 | HS | - | - |
| AED_1119 | 6 | 853.70 | RM | - | - |
| AED_1135 | 4 | 1014.20 | HS | - | - |
| AED_125 | 5 | 915.20 | HS | - | - |
| AED_204 | 4 | 900.00 | HS | - | - |
| AED_249 | 6 | 1708.00 | HS | - | - |
| AED_432 | 6 | 923.10 | HS | - | - |
| AED_513 | 4 | 871.80 | HS | - | - |
| AED_92 | 5 | 797.54 | HS | - | - |
| AEH_I | 4 | 1058.30 | HS | - | - |
| ALH_10 | 7 | 2173.00 | RM | 2 | HS 100 bp PE |
| ALH_1 | 6 | 1494.00 | RM | 2 | HS 100 bp PE |
| ALH_2 | 6 | 767.18 | HS | - | - |
| ALH_3 | 6 | 1474.72 | RM | - | - |
| BIM_33 | 6 | 1422.42 | HS | - | - |
| BIM_37 | 4 | 964.60 | HS | - | - |
| FN_2 | 6 | 1187.25 | RM | 6 | HS 100 bp PE |
| KER_1 | 7 | 852.03 | HS | 7 | HS 100 bp SE |
| NW_255 | 6 | 1077.95 | HS | - | - |
| NW_54 | 4 | 1590.00 | HS | 4 | HS 100 bp SE |
| PR_10 | 4 | 709.00 | HS | - | - |
| PR_4 | 6 | 1288.75 | RM | - | - |
| STR_220 | 5 | 994.83 | HS | 5 | HS 100 bp SE |
| STR_228 | 6 | 1604.20 | HS | - | - |
| STR_241 | 4 | 950.00 | HS | - | - |
| STR_248 | 5 | 1008.00 | HS | - | - |
| STR_266 | 5 | 1023.15 | HS | - | - |
| STR_300 | 6 | 1950.00 | HS | 6 | HS 100 bp SE |
| STR_310 | 7 | 886.10 | HS | 7 | HS 100 bp SE |
| STR_316 | 5 | 862.65 | RM | - | - |
| STR_328 | 5 | 520.33 | HS | - | - |
| STR_355 | 4 | 810.80 | HS | 4 | - |
| STR_360 | 4 | 1471.50 | HS | - | - |
| STR_393 | 6 | 1413.80 | HS | - | - |
| STR_480 | 5 | 670.75 | HS | - | - |
| STR_486 | 4 | 898.50 | HS | 4 | HS 100 bp SE |
| STR_491 | 4 | 943.75 | HS | - | - |
| STR_502 | 4 | 953.80 | HS | - | - |
| STR_535 | 6 | 800.50 | HS | - | - |
| VIM_2 | 6 | 1672.30 | RM | 6 | HS 100 bp SE |

**Table S5** Summary of number and percentage of 1 kb loci and total bp count and total target percentage where a 100 bp or 50 bp target sequence aligns to an alternative region of the genome with an alignment score reaching at least a certain % of the alignment score for the true position.

| **100 bp region** | **>70%** | **>80%** | **>90%** |
|---|---|---|---|
| Total loci | 194 | 100 | 35 |
| % of loci | 4.14% | 2.13% | 0.75% |
| Total bp count | 83,809 | 40,854 | 6,522 |
| % bp count | 1.79% | 0.87% | 0.14% |

| **50 bp region** | **>70%** | **>80%** | **>90%** |
|---|---|---|---|
| Total loci | 895 | 581 | 251 |
| % of loci | 19.10% | 12.40% | 5.36% |
| Total bp count | 186,036 | 122,883 | 57,138 |
| % bp count | 3.97% | 2.62% | 1.22% |

**Table S6** Overview of the sequencing results. Given are the number of raw reads, aligned reads after duplicate removal and filtering for proper-pairs (MkdupFULLpp), and the sequencing method (Cap = capture, Shot = shotgun, PE = paired end, SE = single end). Furthermore average sequencing depth (Av.depth) with standard deviation (stdv.) is calculated including mtGenome, the number of bases covered (Bases covered) and the percentage covered (Coverage [%]) are shown both for the Capture regions and the whole genome (Hg19). For Samples with available WGS and capture data these statistics are calculated for merged BAM-files.

| | | | | Capture regions | | | | Hg19 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Sample** | **Reads raw** | **MkdupFULLpp** | **Cap/ Shot** | **Av.depth** | **stdv** | **Bases covered** | **Cov. [%]** | **Av.depth** | **stdv** | **Bases covered** | **Cov. [%]** |
| AED_106 | 104212408 | 10458190 | Cap PE | 72.63 | 31.52 | 4907350 | 99.84 | 2.32 | 7.39 | 459732085 | 14.85 |
| AED_1108 | 113534672 | 8982907 | Cap PE | 33.23 | 14.51 | 4894117 | 99.58 | 1.96 | 3.65 | 462614282 | 14.94 |
| AED_1119 | 88512400 | 8388113 | Cap PE | 92.44 | 47.04 | 4895932 | 99.61 | 3.13 | 12.65 | 254696370 | 8.23 |
| AED_1135 | 98361148 | 10824787 | Cap PE | 61.6 | 31.8 | 4897467 | 99.64 | 2.49 | 6.9 | 437081179 | 14.12 |
| AED_125 | 110811420 | 4762635 | Cap PE | 38.4 | 26.95 | 4854267 | 98.76 | 2.73 | 7.91 | 133234560 | 4.3 |
| AED_204 | 79717468 | 10604895 | Cap PE | 109.24 | 71.05 | 4890234 | 99.5 | 2.86 | 12.98 | 386147806 | 12.47 |
| AED_249 | 115229492 | 12192831 | Cap PE | 103.16 | 51.59 | 4907430 | 99.85 | 2.7 | 10.73 | 457239608 | 14.77 |
| AED_432 | 99599846 | 8563133 | Cap PE | 66.52 | 35.67 | 4890905 | 99.51 | 2.45 | 8.68 | 291710240 | 9.42 |
| AED_513 | 86634502 | 7993500 | Cap PE | 42.71 | 18.28 | 4893723 | 99.57 | 1.94 | 4.71 | 404173044 | 13.06 |
| AED_92 | 106989782 | 9737385 | Cap PE | 79.47 | 53.81 | 4891001 | 99.51 | 2.87 | 11.43 | 284752124 | 9.2 |
| AEH_I | 104045520 | 8226930 | Cap PE | 42.95 | 23.92 | 4886291 | 99.42 | 2.1 | 5.52 | 386147806 | 12.47 |
| | 63540778 | 165882072 | Cap PE | | | | | | | | |
| Alh_10 | 869005670 | 143966599 | Shot PE | 145.6 | 83.48 | 4905167 | 99.8 | 12.17 | 10.79 | 2653997341 | 85.73 |
| | 115255770 | 173032798 | Cap PE | | | | | | | | |
| Alh_1 | 502612874 | 145655720 | Shot PE | 148.12 | 66.49 | 4914469 | 99.99 | 13.27 | 10.81 | 2645754801 | 85.47 |
| Alh_2 | 88999638 | 3782115 | Cap PE | 20.45 | 8.35 | 4895521 | 99.6 | 2.28 | 3.32 | 215823622 | 6.97 |
| Alh_3 | 151874886 | 20412034 | Cap PE | 61.72 | 32.2 | 4889728 | 99.49 | 3.79 | 5.3 | 1108608771 | 35.81 |
| BIM_33 | 114245484 | 9979595 | Cap PE | 62.68 | 27.22 | 4897088 | 99.64 | 2.16 | 6.57 | 437576306 | 14.13 |
| BIM_37 | 116847674 | 6898036 | Cap PE | 29.83 | 11.27 | 4905709 | 99.81 | 1.88 | 3.59 | 347767347 | 11.23 |
| | 100423988 | 166625673 | Cap PE | | | | | | | | |
| FN_2 | 661497130 | 153139685 | Shot PE | 122.95 | 64.19 | 4914426 | 99.99 | 11.08 | 9.24 | 2621363406 | 84.68 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 88745526 | 23290846 | Cap PE | | | | | | | | |
| KER_I | 26290049 | 17686401 | Shot SE | 19.22 | 9.82 | 4890773 | 99.51 | 2.36 | 2.08 | 1139501652 | 36.81 |
| NW_255 | 112035972 | 13417629 | Cap PE | 92.87 | 49.46 | 4898283 | 99.66 | 2.8 | 9.76 | 479526058 | 15.49 |
| | 108407546 | 11896192 | Cap PE | | | | | | | | |
| NW_54 | 36991270 | 5869938 | Shot SE | 50.01 | 27.79 | 4890107 | 99.49 | 2.26 | 5.26 | 555771265 | 17.95 |
| PR_10 | 114870126 | 5974673 | Cap PE | 29.08 | 9.7 | 4907878 | 99.86 | 2.53 | 3.79 | 361354505 | 11.67 |
| PR_4 | 88230506 | 2728956 | Cap PE | 13.38 | 5.79 | 4897770 | 99.65 | 2.56 | 2.79 | 176526733 | 5.7 |
| | 117926744 | 22728574 | Cap PE | | | | | | | | |
| STR_220 | 15654924 | 10662652 | Shot SE | 51.69 | 19.26 | 4900994 | 99.72 | 2.96 | 4.09 | 1224068654 | 39.54 |
| STR_228 | 125186938 | 14550582 | Cap PE | 80.36 | 37.17 | 4899080 | 99.68 | 2.44 | 7.19 | 649084478 | 20.97 |
| STR_241 | 87373228 | 8208222 | Cap PE | 90.9 | 43.67 | 4910541 | 99.91 | 3.35 | 12.6 | 250288449 | 8.08 |
| STR_248 | 101333860 | 12245470 | Cap PE | 146.08 | 74.92 | 4902501 | 99.75 | 3.6 | 17.05 | 354547567 | 11.45 |
| STR_266 | 65456960 | 7023738 | Cap PE | 102.95 | 55.73 | 4897418 | 99.64 | 4.27 | 17.35 | 168056318 | 5.43 |
| | 95746172 | 25565967 | Cap PE | | | | | | | | |
| STR_300 | 31517778 | 13577549 | Shot SE | 135.43 | 85.17 | 4900894 | 99.71 | 3.55 | 11.01 | 960184340 | 31.02 |
| | 97301284 | 32277898 | Cap PE | | | | | | | | |
| STR_310 | 32048462 | 21626267 | Shot SE | 38.06 | 15.65 | 4898384 | 99.66 | 3.7 | 3.48 | 1581991988 | 51.1 |
| STR_316 | 103800972 | 10378302 | Cap PE | 47.38 | 20.08 | 4906211 | 99.82 | 2.2 | 4.76 | 523470757 | 16.9 |
| STR_328 | 101455199 | 12242555 | Cap PE | 16.23 | 7.69 | 4871783 | 99.12 | 1.96 | 1.96 | 699788757 | 22.6 |
| | 110261624 | 30311257 | Cap PE | | | | | | | | |
| STR_355 | 28479585 | 19251247 | Shot SE | 104.34 | 48.73 | 4901126 | 99.72 | 3.33 | 6.99 | 1385950767 | 44.77 |
| STR_360 | 96803297 | 6232900 | Cap PE | 69.53 | 31.94 | 4895387 | 99.6 | 3.09 | 10.9 | 182921705 | 5.91 |
| STR_393 | 106633500 | 7593951 | Cap PE | 55.07 | 27.69 | 4907706 | 99.85 | 2.53 | 7.38 | 285908772 | 9.24 |
| STR_480 | 96617542 | 10368056 | Cap PE | 57.58 | 23.42 | 4899089 | 99.68 | 2.15 | 5.62 | 526352689 | 17 |
| | 108822884 | 18700690 | Cap PE | | | | | | | | |
| STR_486 | 20670362 | 6964981 | Shot SE | 99.29 | 51.81 | 4911259 | 99.92 | 2.79 | 8.05 | 864902028 | 27.94 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| STR_491 | 75860290 | 4633068 | Cap PE | 57.41 | 33.35 | 4897344 | 99.64 | 3.52 | 11.71 | 116059302 | 3.75 |
| STR_502 | 103782545 | 5043188 | Cap PE | 23.85 | 11.6 | 4875952 | 99.21 | 2.05 | 3.84 | 206428437 | 6.67 |
| STR_535 | 92309750 | 12613019 | Cap PE | 126.81 | 65.47 | 4902940 | 99.75 | 3.01 | 13.39 | 439297888 | 14.19 |
| Vim_2 | 85102788 | 24873215 | Cap PE | 146.55 | 77.05 | 4909035 | 99.88 | 3.69 | 10.46 | 1122639173 | 36.26 |
| | 16692542 | 11182639 | Shot SE | | | | | | | | |

**Table S7** Overview of the results of the mitochondrial captures. The table shows the number of raw reads, and aligned reads after duplicate removal (Mkdup). The numbers of positions of rCRS that could not be covered are given as well as average coverage depth (Av.depth) with standard deviation (stdv.), average read length (Av.length), and deamination rate (mD5) at the 5' end of the molecules (C>T). Furthermore the mitochondrial haplogroup (Hg) as determined with Haplofind (122) is presented for every sample. *mD5 value very low due to USER treatment of libraries.

| Sample | Raw reads | Mkdup | Not covered | Av.depth | stdv. | Av.length | mD5 | Hg |
|---|---|---|---|---|---|---|---|---|
| AED_92 | 783869 | 54621 | 0 | 380.15 | 73.58 | 115.32 | 0.26 | U4a |
| AED_106 | 427042 | 28648 | 0 | 232.41 | 29.47 | 134.54 | 0.21 | V3 |
| AED_125 | 713424 | 52848 | 0 | 341.24 | 63.85 | 107 | 0.26 | K1a |
| AED_204 | 453877 | 29951 | 0 | 247.91 | 34.32 | 137.29 | 0.20 | X2b |
| AED_249 | 807922 | 31854 | 0 | 263.83 | 23.14 | 137.24 | 0.21 | T2c1d1a |
| AED_432 | 843872 | 55187 | 0 | 384.73 | 61.63 | 115.53 | 0.15 | H1h1 |
| AED_513 | 515293 | 28706 | 0 | 234.28 | 28.82 | 135.36 | 0.20 | H7a |
| AED_1108 | 590764 | 27291 | 0 | 214.59 | 29.44 | 130.41 | 0.24 | U4a2a |
| AED_1119 | 654199 | 30625 | 0 | 255.74 | 30.25 | 138.49 | 0.20 | W1e1a |
| AED_1135 | 431310 | 28675 | 0 | 228.33 | 33.68 | 132.06 | 0.26 | H60a |
| AEH_I | 925952 | 46783 | 0 | 322.38 | 52.77 | 114.19 | 0.26 | H5 |
| ALH_1 | extracted from shotgun | 39011 | 105 | 196.55 | 92.16 | 83.48 | 0.32 | J2a2b |
| ALH_2 | 793978 | 31641 | 0 | 252.11 | 19.70 | 132.03 | 0.22 | H2a2a1 |
| ALH_3 | 626400 | 31782 | 0 | 279.12 | 22.35 | 145.52 | 0.16 | H2a2b |
| ALH_10 | extracted from shotgun | 30461 | 148 | 160.80 | 76.85 | 87.46 | 0.30 | I1 |
| BIM_33 | 288980 | 24076 | 0 | 189.41 | 38.42 | 130.49 | 0.24 | U5b1b |
| BIM_37 | 310033 | 23935 | 0 | 183.41 | 34.02 | 127.08 | 0.22 | T2b16 |
| FN_2 | extracted from shotgun | 37086 | 63 | 169.61 | 84.09 | 75.78 | 0.35 | H3 |
| KER_1 | 280050 | 18809 | 0 | 123.39 | 29.83 | 108.79 | 0.24 | HV9a |
| NW_54 | 433135 | 26196 | 0 | 198.90 | 36.41 | 125.93 | 0.32 | C4a1a |
| NW_255 | 600149 | 30100 | 0 | 238.12 | 30.55 | 131.19 | 0.28 | H18b |
| STR_220 | 388191 | 27877 | 0 | 229.69 | 30.19 | 136.65 | 0.23 | U5b2b |
| STR_228 | 529897 | 29346 | 0 | 242.38 | 28.80 | 136.97 | 0.25 | T2b21 |
| STR_241 | 288920 | 26989 | 0 | 220.45 | 36.94 | 135.47 | 0.21 | H1 |
| STR_248 | 501335 | 29822 | 0 | 251.39 | 30.92 | 139.81 | 0.20 | H11a2 |
| STR_266 | 646073 | 76859 | 0 | 557.68 | 86.43 | 120.23 | 0.26 | J1c5 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| STR_300 | 646822 | 30950 | 0 | 235.89 | 27.60 | 126.4 | 0.09* | J1c10 |
| STR_310 | 642024 | 29719 | 0 | 232.36 | 19.08 | 129.55 | 0.24 | H7 |
| STR_316 | 767155 | 31331 | 0 | 244.16 | 20.09 | 129.13 | 0.22 | H3 |
| STR_328 | 508046 | 27351 | 0 | 192.33 | 22.97 | 116.52 | 0.26 | U5a1h |
| STR_355 | 722970 | 32215 | 0 | 257.87 | 19.71 | 132.64 | 0.22 | T2b |
| STR_360 | 612490 | 29716 | 0 | 226.47 | 24.93 | 126.29 | 0.23 | H5 |
| STR_393 | 836998 | 53281 | 0 | 395.41 | 52.85 | 122.97 | 0.26 | H5a1 |
| STR_480 | 714467 | 30790 | 0 | 255.82 | 23.01 | 137.67 | 0.21 | J1c2o |
| STR_486 | 747838 | 50819 | 0 | 399.05 | 60.18 | 130.11 | 0.24 | T2b |
| STR_491 | 561070 | 30011 | 0 | 234.94 | 26.87 | 129.72 | 0.23 | T2b |
| STR_502 | 605930 | 31131 | 0 | 212.69 | 21.27 | 113.21 | 0.31 | T1 |
| STR_535 | 467471 | 27313 | 0 | 224.42 | 31.26 | 136.28 | 0.23 | HV9 |
| VIM_2 | 849940 | 32620 | 0 | 283.19 | 18.88 | 143.86 | 0.20 | H7 |

**Table S8** Results of HIrisplex analysis for all individuals subjected to nuclear target enrichment.

| Sample | blue eye | | intermediate eye | | brown eye | | blonde hair | | brown hair | | red hair | | black hair | | light hair | | dark hair | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P | AUC loss | P | AUC loss | P | AUC loss | P | AUC loss | P | AUC loss | P | AUC loss | P | AUC loss | P | AUC loss | P | AUC loss |
| AED_92 | 0.917 | 0.000 | 0.052 | 0.000 | 0.031 | 0.000 | 0.671 | 0.003 | 0.163 | 0.003 | 0.163 | 0.000 | 0.003 | 0.004 | 0.997 | 0.004 | 0.003 | 0.004 |
| AED_106 | 0.863 | 0.002 | 0.079 | 0.009 | 0.058 | 0.005 | 0.690 | 0.007 | 0.269 | 0.006 | 0.012 | 0.008 | 0.030 | 0.008 | 0.948 | 0.006 | 0.052 | 0.006 |
| AED_125 | 0.052 | 0.000 | 0.116 | 0.000 | 0.832 | 0.000 | 0.265 | 0.003 | 0.623 | 0.003 | 0.004 | 0.000 | 0.107 | 0.004 | 0.707 | 0.004 | 0.293 | 0.004 |
| AED_204 | 0.936 | 0.000 | 0.043 | 0.000 | 0.021 | 0.000 | 0.666 | 0.006 | 0.281 | 0.003 | 0.032 | 0.008 | 0.021 | 0.005 | 0.955 | 0.004 | 0.045 | 0.004 |
| AED_249 | 0.936 | 0.000 | 0.043 | 0.000 | 0.021 | 0.000 | 0.686 | 0.004 | 0.279 | 0.005 | 0.003 | 0.001 | 0.032 | 0.006 | 0.946 | 0.005 | 0.054 | 0.005 |
| AED_432 | 0.963 | 0.002 | 0.026 | 0.009 | 0.010 | 0.005 | 0.663 | 0.007 | 0.307 | 0.004 | 0.010 | 0.007 | 0.021 | 0.006 | 0.977 | 0.005 | 0.023 | 0.005 |
| AED_513 | 0.144 | 0.003 | 0.171 | 0.016 | 0.686 | 0.007 | 0.206 | 0.056 | 0.574 | 0.053 | 0.003 | 0.010 | 0.217 | 0.019 | 0.500 | 0.036 | 0.500 | 0.036 |
| AED_1108 | 0.001 | 0.000 | 0.018 | 0.000 | 0.981 | 0.000 | 0.102 | 0.006 | 0.636 | 0.003 | 0.001 | 0.008 | 0.262 | 0.005 | 0.213 | 0.004 | 0.787 | 0.004 |
| AED_1119 | 0.937 | 0.002 | 0.042 | 0.009 | 0.020 | 0.005 | 0.753 | 0.005 | 0.205 | 0.004 | 0.003 | 7.728E-05 | 0.038 | 0.009 | 0.956 | 0.006 | 0.044 | 0.006 |
| AED_1135 | 0.014 | 0.000 | 0.051 | 0.000 | 0.935 | 0.000 | 0.151 | 0.003 | 0.491 | 0.003 | 0.002 | 0.000 | 0.356 | 0.004 | 0.295 | 0.004 | 0.705 | 0.004 |
| AEH_I | 0.143 | 0.000 | 0.155 | 0.000 | 0.702 | 0.000 | 0.449 | 0.006 | 0.410 | 0.003 | 0.001 | 0.008 | 0.140 | 0.005 | 0.752 | 0.004 | 0.248 | 0.004 |
| ALH_1 | 0.143 | 0.000 | 0.155 | 0.000 | 0.702 | 0.000 | 0.466 | 0.003 | 0.406 | 0.003 | 0.059 | 0.000 | 0.069 | 0.004 | 0.858 | 0.004 | 0.142 | 0.004 |
| ALH_2 | 0.858 | 0.000 | 0.081 | 0.000 | 0.061 | 0.000 | 0.161 | 0.026 | 0.175 | 0.017 | 0.658 | 0.085 | 0.006 | 0.019 | 0.954 | 0.011 | 0.046 | 0.011 |
| ALH_3 | 0.924 | 0.010 | 0.057 | 0.029 | 0.019 | 0.005 | 0.392 | 0.004 | 0.455 | 0.004 | 0.005 | 0.000 | 0.147 | 0.005 | 0.765 | 0.005 | 0.235 | 0.005 |
| ALH_10 | 0.075 | 0.000 | 0.136 | 0.000 | 0.788 | 0.000 | 0.319 | 0.003 | 0.543 | 0.003 | 0.001 | 0.000 | 0.136 | 0.004 | 0.638 | 0.004 | 0.362 | 0.004 |
| BIM_33 | 0.893 | 0.002 | 0.067 | 0.009 | 0.040 | 0.005 | 0.687 | 0.012 | 0.275 | 0.010 | 0.005 | 0.010 | 0.034 | 0.011 | 0.944 | 0.006 | 0.056 | 0.006 |
| BIM_37 | 0.958 | 0.002 | 0.030 | 0.003 | 0.013 | 0.001 | 0.754 | 0.003 | 0.214 | 0.003 | 0.006 | 0.000 | 0.027 | 0.004 | 0.961 | 0.004 | 0.039 | 0.004 |
| FN_2 | 0.052 | 0.000 | 0.116 | 0.000 | 0.832 | 0.000 | 0.340 | 0.005 | 0.538 | 0.005 | 0.006 | 5.983E-05 | 0.117 | 0.008 | 0.656 | 0.005 | 0.344 | 0.005 |
| KER_1 | 0.014 | 0.000 | 0.051 | 0.000 | 0.935 | 0.000 | 0.144 | 0.003 | 0.527 | 0.003 | 0.001 | 0.000 | 0.328 | 0.004 | 0.300 | 0.004 | 0.700 | 0.004 |
| NW_54 | 0.052 | 0.000 | 0.116 | 0.000 | 0.832 | 0.000 | 0.389 | 0.003 | 0.493 | 0.003 | 0.003 | 0.000 | 0.115 | 0.004 | 0.806 | 0.004 | 0.194 | 0.004 |
| NW_255a | 0.963 | 0.002 | 0.026 | 0.009 | 0.010 | 0.005 | 0.799 | 0.005 | 0.155 | 0.004 | 0.010 | -0.001 | 0.036 | 0.006 | 0.961 | 0.005 | 0.039 | 0.005 |
| PR_4 | 0.297 | 0.002 | 0.175 | 0.009 | 0.528 | 0.005 | 0.455 | 0.027 | 0.394 | 0.023 | 0.018 | 0.078 | 0.133 | 0.016 | 0.771 | 0.013 | 0.229 | 0.013 |
| PR_10 | 0.919 | 0.002 | 0.051 | 0.009 | 0.030 | 0.005 | 0.631 | 0.011 | 0.306 | 0.006 | 0.008 | 0.009 | 0.055 | 0.010 | 0.905 | 0.006 | 0.095 | 0.006 |
| STR_220 | 0.919 | 0.002 | 0.051 | 0.009 | 0.030 | 0.005 | 0.655 | 0.009 | 0.316 | 0.005 | 0.011 | 0.008 | 0.018 | 0.009 | 0.958 | 0.007 | 0.042 | 0.007 |
| STR_228 | 0.937 | 0.002 | 0.042 | 0.009 | 0.020 | 0.005 | 0.792 | 0.007 | 0.175 | 0.005 | 0.004 | 0.007 | 0.029 | 0.008 | 0.968 | 0.006 | 0.032 | 0.006 |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| STR_241 | 0.891 | 0.000 | 0.068 | 0.000 | 0.041 | 0.000 | 0.706 | 0.003 | 0.245 | 0.003 | 0.013 | 0.000 | 0.036 | 0.004 | 0.961 | 0.004 | 0.039 | 0.004 |
| STR_248 | 0.951 | 0.000 | 0.035 | 0.000 | 0.014 | 0.000 | 0.586 | 0.003 | 0.322 | 0.003 | 0.003 | 0.000 | 0.088 | 0.004 | 0.855 | 0.004 | 0.145 | 0.004 |
| STR_266 | 0.851 | 0.005 | 0.081 | 0.018 | 0.068 | 0.008 | 0.532 | 0.008 | 0.201 | 0.005 | 0.261 | 0.008 | 0.005 | 0.007 | 0.991 | 0.005 | 0.009 | 0.005 |
| STR_300 | 0.003 | 0.000 | 0.054 | 0.000 | 0.943 | 0.000 | 0.036 | 0.003 | 0.610 | 0.003 | 0.001 | 0.000 | 0.353 | 0.004 | 0.106 | 0.004 | 0.894 | 0.004 |
| STR_310 | 0.069 | 0.002 | 0.136 | 0.009 | 0.795 | 0.005 | 0.492 | 0.009 | 0.381 | 0.005 | 0.113 | 0.010 | 0.015 | 0.007 | 0.972 | 0.007 | 0.028 | 0.007 |
| STR_316 | 0.953 | 0.002 | 0.032 | 0.009 | 0.015 | 0.005 | 0.783 | 0.011 | 0.186 | 0.006 | 0.004 | 0.009 | 0.027 | 0.007 | 0.969 | 0.005 | 0.031 | 0.005 |
| STR_328 | 0.165 | 0.019 | 0.157 | 0.048 | 0.678 | 0.017 | NA | 0.052 | NA | 0.045 | NA | 0.255 | NA | 0.050 | 0.538 | 0.030 | 0.462 | 0.030 |
| STR_355 | 0.917 | 0.000 | 0.052 | 0.000 | 0.031 | 0.000 | 0.687 | 0.003 | 0.253 | 0.003 | 0.016 | 0.000 | 0.044 | 0.004 | 0.931 | 0.004 | 0.069 | 0.004 |
| STR_360 | 0.910 | 0.002 | 0.068 | 0.009 | 0.022 | 0.005 | 0.354 | 0.026 | 0.496 | 0.020 | 0.074 | 0.079 | 0.075 | 0.012 | 0.799 | 0.010 | 0.201 | 0.010 |
| STR_393 | 0.917 | 0.000 | 0.052 | 0.000 | 0.031 | 0.000 | 0.655 | 0.024 | 0.246 | 0.022 | 0.006 | 0.084 | 0.092 | 0.012 | 0.901 | 0.009 | 0.099 | 0.009 |
| STR_480 | 0.858 | 0.000 | 0.081 | 0.000 | 0.061 | 0.000 | 0.743 | 0.007 | 0.223 | 0.004 | 0.003 | 0.009 | 0.030 | 0.006 | 0.961 | 0.005 | 0.039 | 0.005 |
| STR_486 | 0.001 | 0.000 | 0.019 | 0.000 | 0.981 | 0.000 | 0.018 | 0.003 | 0.229 | 0.003 | 6.577E-05 | 0.000 | 0.753 | 0.004 | 0.029 | 0.004 | 0.971 | 0.004 |
| STR_491 | 0.851 | 0.005 | 0.081 | 0.018 | 0.068 | 0.008 | 0.629 | 0.037 | 0.309 | 0.036 | 0.046 | 0.079 | 0.015 | 0.024 | 0.969 | 0.020 | 0.031 | 0.020 |
| STR_502 | 0.003 | 0.005 | 0.095 | 0.018 | 0.902 | 0.008 | 0.004 | 0.024 | 0.431 | 0.027 | 0.000 | 0.017 | 0.565 | 0.025 | 0.007 | 0.018 | 0.993 | 0.018 |
| STR_535 | 0.087 | 0.000 | 0.192 | 0.000 | 0.721 | 0.000 | 0.162 | 0.003 | 0.635 | 0.003 | 0.088 | 0.000 | 0.116 | 0.004 | 0.559 | 0.004 | 0.441 | 0.004 |
| VIM_2 | 0.001 | 0.002 | 0.017 | 0.011 | 0.983 | 0.003 | 0.049 | 0.013 | 0.269 | 0.013 | 0.000 | 0.004 | 0.682 | 0.016 | 0.086 | 0.016 | 0.914 | 0.016 |

**Table S9** Frequencies (%) of eye and hair colors in the respective groupings of the ancient Bavarian individuals. 95% CIs are given in brackets.

| Group | blue eyes (%) | brown eyes (%) | blonde hair (%) | brown hair (%) | black hair (%) | red hair (%) |
|---|---|---|---|---|---|---|
| **Deformed** | 22.2 (0-49.3) | 77.8 (50.6-100) | 37.5 (4.0-71.1) | 62.5 (29.0-96.1) | 0 | 0 |
| **Non-Deformed** | 82.6 (67.1-98.1) | 17.4 (1.9-32.9) | 73.9 (56.0-91.9) | 17.4 (1.9-32.9) | 4.35 (0-12.7) | 4.35 (0-12.7) |
| **Male** | 77.8 (50.6-100) | 22.2 (0-49.4) | 88.9 (68.4-100) | 0 | 0 | 11.1 (0-31.6) |
| **Female** | 53.8 (34.7-73.0) | 46.2 (27.0-65.3) | 52.0 (32.4-71.6) | 40.0 (20.8-59.2) | 4.0 (0-11.7) | 4.0 (0-11.7) |

**Table S10** Markers applied for prediction of skin color (77) and genotypes for EDAR (rs3827760) for all individuals subjected to nuclear target enrichment.

| **SNP** | **rs12913832** | **rs1545397** | **rs16891982** | **rs885479** | **rs1426654** | **rs12896399** | **rs6119471** | **rs12203592** | | | **rs3827760** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Gene** | **HERC2** | **OCA2** | **SLC45A2** | **MC1R** | **SLC24A5** | **SLC24A4** | **ASIP** | **IRF4** | | | **EDAR** |
| **Ancestral/ Derived** | **A/G** | **A/T** | **C/G** | **G/A** | **G/A** | **G/T** | **G/C** | **C/T** | | | **A/G** |
| AED_92 | G/G | A/A | G/G | G/G | A/A | G/T | C/C | C/C | non-dark | light-medium | A/G |
| AED_106 | G/G | A/A | G/G | - | A/A | G/G | C/C | C/C | non-dark | light-medium | A/A |
| AED_125 | A/G | A/A | G/G | G/G | A/A | G/G | C/C | C/C | non-dark | light-medium | A/A |
| AED_204 | G/G | A/A | G/G | G/G | A/A | G/T | C/C | C/C | non-dark | light-medium | A/A |
| AED_249 | G/G | A/A | G/G | G/G | A/A | G/T | C/C | C/C | non-dark | light-medium | A/A |
| AED_432 | G/G | A/A | G/G | G/G | A/A | T/T | C/C | C/C | non-dark | light-medium | - |
| AED_513 | A/G | A/A | G/G | - | A/A | G/T | C/C | - | non-dark | light-medium | A/A |
| AED_1108 | A/A | A/T | G/G | G/G | A/G | G/T | C/C | C/C | inconclusive | | - |
| AED_1119 | G/G | A/A | G/G | G/G | A/A | G/T | C/C | C/C | non-dark | light-medium | A/A |
| AED_1135 | A/G | A/A | C/G | G/G | A/A | G/G | C/C | C/C | inconclusive | | - |
| AED_204 | G/G | A/A | G/G | G/G | A/A | G/T | C/C | C/C | non-dark | light-medium | A/A |
| AEH_I | A/G | A/A | G/G | G/G | A/A | G/T | C/C | C/C | non-dark | light-medium | A/A |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ALH_1 | A/G | A/A | G/G | G/G | A/A | G/T | C/C | C/C | non-dark | light-medium | A/A |
| ALH_2 | G/G | A/A | G/G | G/G | A/A | G/G | C/C | C/C | non-dark | light-medium | A/A |
| ALH_3 | G/G | A/A | G/G | - | A/A | - | C/C | C/T | non-dark | light-medium | A/G |
| ALH_10 | A/G | A/A | G/G | G/G | A/A | G/G | C/C | C/C | non-dark | light-medium | A/A |
| BIM_33 | G/G | A/T | G/G | G/G | A/A | G/G | C/C | C/C | non-dark | light-medium | A/A |
| BIM_37 | G/G | A/A | G/G | G/G | A/A | T/T | C/C | C/C | non-dark | light-medium | A/A |
| FN_2 | A/G | A/A | G/G | G/G | A/A | G/G | C/C | C/C | non-dark | light-medium | A/A |
| KER_1 | A/G | A/A | C/G | G/G | A/A | G/G | C/C | C/C | inconclusive | | - |
| NW_54 | A/G | A/A | G/G | G/G | A/A | G/G | C/C | C/C | non-dark | light-medium | A/A |
| NW_255 | G/G | A/A | G/G | G/G | A/A | T/T | C/C | C/C | non-dark | light-medium | A/A |
| PR_4 | A/G | A/A | G/G | - | A/A | T/T | - | C/C | non-dark | light-medium | - |
| PR_10 | G/G | A/A | G/G | G/G | A/G | G/T | C/C | C/C | non-dark | light-medium | A/A |
| STR_220 | G/G | A/A | G/G | G/G | A/A | G/T | C/C | C/C | non-dark | light-medium | A/A |
| STR_228 | G/G | A/T | G/G | G/G | A/A | G/T | C/G | C/C | non-dark | light-medium | A/A |
| STR_241 | G/G | A/A | G/G | G/G | A/A | G/G | C/C | C/C | non-dark | light-medium | A/A |
| STR_248 | G/G | A/A | G/G | A/G | A/A | G/T | C/C | C/C | non-dark | light-medium | A/A |
| STR_266 | G/G | A/A | - | G/G | A/A | G/G | C/C | C/C | non-dark | light-medium | A/A |
| STR_300 | A/A | A/A | G/G | G/G | A/A | T/T | C/C | C/T | non-dark | light-medium | A/A |
| STR_310 | A/G | A/A | G/G | G/G | A/A | G/G | C/C | C/C | non-dark | light-medium | A/A |
| STR_316 | G/G | A/A | G/G | G/G | A/A | T/T | - | C/C | non-dark | light-medium | A/A |
| STR_328 | A/G | A/T | - | - | A/A | - | C/C | C/C | inconclusive | | - |
| STR_355 | G/G | A/A | G/G | G/G | A/A | G/T | C/C | C/C | non-dark | light-medium | A/A |
| STR_360 | G/G | A/A | G/G | G/G | A/A | G/G | C/C | C/T | non-dark | light-medium | A/A |
| STR_393 | G/G | A/A | G/G | A/A | A/A | G/T | C/C | C/C | non-dark | light-medium | A/A |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| STR_480 | G/G | A/T | G/G | G/G | A/A | G/G | C/C | C/C | non-dark | light-medium | A/A |
| STR_486 | A/A | A/A | C/G | A/G | A/A | G/T | C/C | C/C | inconclusive | | - |
| STR_491 | G/G | A/A | - | - | A/A | G/G | - | C/C | non-dark | light-medium | A/A |
| STR_502 | A/A | A/A | - | - | A/A | G/G | C/C | T/T | non-dark | light-medium | A/A |
| STR_535 | A/G | A/A | G/G | G/G | A/A | G/G | C/C | C/T | non-dark | light-medium | A/A |
| VIM_2 | A/A | A/T | - | - | A/G | G/T | C/C | C/C | inconclusive | | - |

**Table S11** Results of determination of SLC24A5 haplotypes following the approaches by Canfield et al. (78) and Giardina et al. (81). Haplotypes marked with "?" are not completely secure.

| SNP | rs1834640 | rs2675345 | rs2469592 | rs2470101 | rs938505 | rs2433354 | rs2459391 | rs2433356 | rs2675347 | rs2675348 | rs1426654 | rs2470102 | rs16960631 | rs2675349 | rs3817315 | rs7163587 | | rs16960620 | rs2555364 | rs1426654 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ancestral | G | G | G | C | C | T | G | A | G | G | G | G | A | G | T | T | | A | C | G | |
| Derived | A | A | A | T | T | C | A | G | A | A | A | A | G | A | C | C | **Haplotype** | G | G | A | **Haplotype** |
| AED_92 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| AED_106 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| AED_125 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| AED_204 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| AED_249 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| AED_432 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| AED_513 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| AED_1108 | A/G | A/G | A/A | T/T | C/C | C/C | A/A | G/G | - | A/A | A/G | - | A/A | - | C/C | C/C | **C11/C9** | A/A | G/G | A/G | **Ho1/Ho2/He** |
| AED_1119 | A/G | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| AED_1135 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| AED_204 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| AEH_I | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| ALH_1 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | A/A | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| ALH_2 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | A/A | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| ALH_3 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | (A/A) | G/G | A/A | **Ho1?/Ho2/Ho1** |
| ALH_10 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | A/A | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| BIM_33 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| BIM_37 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| FN_2 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | A/A | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| KER_1 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| NW_54 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NW_255 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| PR_4 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| PR_10 | A/A | A/G | A/G | C/C* | C/T | C/T | A/G | - | - | A/G | A/G | A/G | A/A | - | C/C* | C/T | **C11/C3?** | A/A | C/G | A/G | **Ho1/He/He** |
| STR_220 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| STR_228 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| STR_241 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | A/A | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| STR_248 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| STR_266 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| STR_300 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| STR_310 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| STR_316 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | - | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| STR_328 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | - | - | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| STR_355 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| STR_360 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| STR_393 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| STR_480 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | - | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| STR_486 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| STR_491 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | - | A/A | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| STR_502 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | - | A/A | A/A | A/A | A/A | - | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| STR_535 | A/A | A/A | A/A | T/T | C/C | C/C | A/A | G/G | A/A | A/A | A/A | A/A | A/A | A/A | C/C | C/C | **C11** | A/A | G/G | A/A | **Ho1/Ho2/Ho1** |
| VIM_2 | A/G | A/G | A/A | T/T | C/C | C/T | A/A | G/G | A/A* | A/G | A/G | G/G* | A/G | - | C/T | C/C | **C11/C7?** | A/A | C/G | A/G | **Ho1/He/He** |

**Table S12** Determination of OCA-HERC haplotypes according to the approach by Eiberg et al. (85) and the blue eye haplotypes (BEH) following Donnelly et al. (82). Haplotypes marked with "?" are not completely secure.

| | | | | | | | | | | | | | | | BEH1 | | | BEH2 | | BEH3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SNP | rs4778241 | rs1129038 | rs12593929 | rs12913832 | rs7183877 | rs3935591 | rs7170852 | rs2238289 | rs3940272 | rs8028689 | rs2240203 | rs11631797 | rs916977 | **13-SNP haplotype** | rs4778138 | rs4778241 | rs7495174 | rs1129038 | rs12913832 | rs916977 | rs1667394 | **BEH** | rs1800414 |
| Ancestral | A | C | G | A | C | T | T | A | T | C | C | A | T | | G | A | G | C | A | T | C | | T |
| Derived | C | T | A | G | A | C | A | G | G | T | T | G | C | | A | C | A | T | G | C | T | | C |
| AED_92 | C/C | T/T | A/A | G/G | C/C | C/C | A/A | A/A | G/G | T/T | T/T | G/G | C/C | **h-1** | A/A | C/C | A/A | T/T | G/G | C/C | T/T | **BEH1/2/3** | T/T |
| AED_106 | C/C | - | - | G/G | C/C | C/C | A/A | - | - | T/T | T/T | - | C/C | **h-1** | A/A | C/C | A/A | - | G/G | C/C | T/T | **BEH1/2?/3** | T/T |
| AED_125 | C/C | - | A/A | A/G | C/C | C/C | A/A | A/A | G/G | T/T | T/T | G/G | C/C | **h-1/h-5** | A/G | C/C | A/A | - | A/G | C/C | T/T | **BEH3** | T/T |
| AED_204 | C/C | - | A/A | G/G | C/C | C/C | A/A | A/A | G/G | T/T | T/T | G/G | C/C | **h-1** | A/A | C/C | A/A | - | G/G | C/C | T/T | **BEH1/2?/3** | T/T |
| AED_249 | C/C | - | - | G/G | C/C | - | A/A | - | - | T/T | T/T | - | C/C | **h-1?, h-3?** | A/A | C/C | A/A | - | G/G | C/C | T/T | **BEH1/2/3** | T/T |
| AED_432 | C/C | - | - | G/G | C/C | - | A/T | - | - | T/T | T/T | - | C/C | **h-1/h-3** | A/A | C/C | A/A | - | G/G | C/C | T/T | **BEH1/2/3** | T/T |
| AED_513 | - | - | - | A/G | C/C | - | A/A | - | - | T/T | T/T | - | C/C | **h-1?/h-3?,h-5?/h-6?** | A/G | - | - | - | A/G | C/C | T/T | **BEH3** | T/T |
| AED_1108 | C/C | - | - | A/A | C/C | - | A/A | - | - | T/T | T/T | - | C/C | **h-5** | A/A | C/C | - | - | A/A | C/C | T/T | **BEH1?/2?/3** | T/T |
| AED_1119 | C/C | - | - | G/G | C/C | - | A/A | - | - | T/T | T/T | - | C/C | **h-1** | A/A | C/C | A/A | - | G/G | C/C | T/T | **BEH1/2?/3** | T/T |
| AED_1135 | A/C | C/T | A/G | A/G | C/C* | C/T | A/T | A/G | G/T | C/T | C/T | A/G | C/T | **h-1/h-10?, h-3/h-9?** | A/G | A/C | A/G | C/T | A/G | C/T | C/T | **het for all BEH** | T/T |
| AED_204 | C/C | - | A/A | G/G | C/C | C/C | A/A | A/A | G/G | T/T | T/T | G/G | C/C | **h-1** | A/A | C/C | A/A | - | G/G | C/C | T/T | **BEH1/2?/3** | T/T |
| AEH_I | A/A | - | A/G | A/G | C/C* | - | A/T | A/G | - | C/T | C/T | A/G | C/T | **h-1/h-10** | G/G | A/A | A/A | - | A/G | C/T | C/T | **het BEH2/3?** | T/T |
| ALH_1 | C/C | C/T | A/A | A/G | C/C | C/C | A/A | A/A | G/G | T/T | T/T | G/G | C/T | **h-2/h-5?** | A/G | C/C | A/A | C/T | A/G | C/T | C/T | **het for all BEH** | T/T |
| ALH_2 | C/C | T/T | - | G/G | C/C | C/C | A/A | A/A | G/G | T/T | T/T | G/G | C/C | **h-1** | A/A | C/C | A/A | T/T | G/G | C/C | T/T | **BEH1/2/3** | T/T |
| ALH_3 | C/C | - | - | G/G | C/C | - | A/A | A/A | - | T/T | T/T | - | C/C | **h-1** | A/A | C/C | A/A | - | G/G | C/C | T/T | **BEH1/2?/3** | T/T |
| ALH_10 | C/C | - | A/A | A/G | C/C | C/C | A/A | A/A | G/G | T/T | T/T | G/G | C/T | **h-2/h-5** | A/G | C/C | A/A | - | A/G | C/T | C/T | **het for all BEH** | T/T |
| BIM_33 | C/C | - | - | G/G | C/C | - | A/A | - | - | T/T | T/T | - | C/C | **h-1** | A/A | C/C | - | - | G/G | C/C | T/T | **BEH1/2?/3** | T/T |
| BIM_37 | C/C | - | - | G/G | C/C | C/C | A/A | - | - | T/T | T/T | - | C/C | **h-1** | A/A | C/C | - | - | G/G | C/C | T/T | **BEH1/2?/3** | T/T |
| FN_2 | A/C | T/T° | A/A | A/G | C/C | C/T | A/T | A/A | - | T/T | T/T | - | C/T | **h-3/h-7?** | A/G | A/C | A/G | T/T | A/G | C/T | C/T | **het for all BEH?** | T/T |
| KER_1 | C/C | - | A/A | A/G | C/C | C/C | A/A | A/A | - | T/T | T/T | G/G | C/T | **h-2/h-5** | A/G | C/C | A/A | - | A/G | C/T | C/T | **het for all BEH** | T/T |

| | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NW_54 | C/C | C/T | A/A | A/G | C/C | C/T | A/A | A/A | G/G | T/T | T/T | G/G | C/C | **h-1/h-5?** | A/A | C/C | A/A | C/T | A/G | C/C | T/T | **BEH1/3; het BEH2?** | T/T |
| NW_255 | C/C | - | A/A | G/G | C/C | - | A/A | - | - | T/T | T/T | - | C/C | **h-1** | A/A | C/C | A/A | - | G/G | C/C | T/T | **BEH1/2?/3** | T/T |
| PR_4 | C/C | - | - | A/G | C/C | - | A/A | - | - | T/T | T/T | - | C/C | **h-1/h-5** | A/A | C/C | - | - | A/G | C/C | T/T | **BEH1/3; het BEH2?** | T/T |
| PR_10 | C/C | - | - | G/G | C/C | - | A/A | - | - | T/T | T/T | - | C/C | **h-1** | A/A | C/C | - | - | G/G | C/C | T/T | **BEH1/2?/3** | - |
| STR_220 | C/C | - | - | G/G | C/C | - | A/A | - | - | T/T | T/T | - | C/C | **h-1** | A/A | C/C | A/A | - | G/G | C/C | T/T | **BEH1/2?/3** | T/T |
| STR_228 | A/C | - | - | G/G | C/C | C/C | A/A | - | - | T/T | T/T | - | C/C | **h-1/h-1** | A/G | A/C | A/A | - | G/G | C/C | T/T | **BEH2?/3; het BEH1** | T/T |
| STR_241 | C/C | T/T | A/A | G/G | C/C | C/C | A/A | A/A | G/G | T/T | T/T | G/G | C/C | **h-1** | A/A | C/C | A/A | T/T | G/G | C/C | T/T | **BEH1/2/3** | T/T |
| STR_248 | C/C | T/T | A/A | G/G | C/C | C/C | A/A | A/A | G/G | T/T | T/T | G/G | C/C | **h-1** | A/A | C/C | A/A | T/T | G/G | C/C | T/T | **BEH1/2/3** | T/T |
| STR_266 | C/C | - | - | G/G | C/C | - | A/A | - | - | T/T | T/T | - | C/C | **h-1** | A/A | C/C | A/A | - | G/G | C/C | T/T | **BEH1/2?/3** | T/T |
| STR_300 | A/A | C/C | A/G | A/A | C/C* | C/T | A/T | A/G | G/T | C/T | C/T | A/G | C/T | **h-5/h-10?** | G/G | A/A | - | C/C | A/A | C/T | C/T | **het BEH3** | T/T |
| STR_310 | A/C | - | - | A/G | C/C | - | A/T | - | - | T/T | T/T | - | C/T | **h-3/h-7** | A/G | A/C | A/A | - | A/G | C/T | C/T | **het BEH1/2/3?** | T/T |
| STR_316 | - | - | - | G/G | - | - | A/A | - | - | T/T | T/T | - | C/C | **h-1** | A/A | - | A/A | - | G/G | C/C | T/T | **BEH1?/2?/3** | T/T |
| STR_328 | - | - | - | A/G | C/C* | - | A/T | - | - | C/T | C/C | - | C/T | **h-4/h-10?** | A/G | - | - | - | A/G | C/T | - | **all BEHs only possible het** | T/T |
| STR_355 | C/C | - | A/A | G/G | C/C | C/C | A/A | - | - | T/T | T/T | - | C/C | **h-1** | A/A | C/C | - | - | G/G | C/C | T/T | **BEH1?/2?/3** | T/T |
| STR_360 | C/C | - | - | G/G | C/C | - | A/A | - | - | T/T | T/T | - | C/C | **h-1** | A/A | C/C | - | - | G/G | C/C | T/T | **BEH1/2?/3** | T/T |
| STR_393 | C/C | T/T | A/A | G/G | C/C | C/C | A/A | A/A | G/G | T/T | T/T | G/G | C/C | **h-1** | A/G | C/C | A/A | T/T | G/G | C/C | T/T | **BEH2/3; BEH1 het** | T/T |
| STR_480 | C/C | - | - | G/G | C/C | - | A/A | - | - | T/T | T/T | - | C/C | **h-1** | A/A | C/C | A/A | - | G/G | C/C | T/T | **BEH1/2?/3** | T/T |
| STR_486 | A/A | C/C | A/A | A/A | A/A | T/T | T/T | G/G | T/T | T/T | T/T | A/A | T/T | **h-8** | A/A | A/A | A/A | C/C | A/A | T/T | C/C | **no BEH** | T/T |
| STR_491 | C/C | - | - | G/G | C/C | - | A/A | - | - | T/T | T/T | - | C/C | **h-1** | A/A | C/C | - | - | G/G | C/C | T/T | **BEH1?/2?/3** | T/T |
| STR_502 | C/C | - | - | A/A | C/C | - | A/A | - | - | T/T | T/T | - | C/C | **h-5** | A/A | C/C | - | - | A/A | C/C | T/T | **BEH1?/3** | T/T |
| STR_535 | A/C | C/T | A/A | A/G | C/C | C/C | A/A | A/A | G/G | T/T | T/T | G/G | C/C | **h-1/h-5** | A/G | A/C | G/G | C/T | A/G | C/C | T/T | **BEH3; BEH2 het** | T/T |
| VIM_2 | A/A | C/C | - | A/A | A/C* | - | T/T | - | - | C/T | C/T | - | T/T | **h-8/h-10?** | G/G | A/A | G/G | C/C | A/A | T/T | C/C | **no BEH** | T/T |

**Table S13** Determination of NAT2 genotypes and acetylation phenotypes for all individuals subjected to nuclear target enrichment. Genotypes marked with * were included in order to allow for analysis with the online tool nat2pred (http://nat2pred.rit.albany.edu; 88) and exhibited QUAL > 12 and sequence depth (DP) > 3.

| **Sample** | rs1041983 (282C>T) | rs1801280 (341T>C) | rs1799929 (481C>T) | rs1799930 (590G>A) | rs1208 (803A<G) | rs1799931 (857G>A) | rs1495741 (tag) | rs1801279 (191G>A) | **acetylation phenotype** | p as determined with nat2pred-onlinetool |
|---|---|---|---|---|---|---|---|---|---|---|
| **AED_92** | C/T | T/T | C/C | G/G | A/A | G/A | G/A | G/G | **intermediate** | 0.99704 |
| **AED_106** | C/C | C/T | C/T | G/G | G/A | G/G | G/A | G/G | **intermediate** | 0.99696 |
| **AED_125** | C/C | C/T | C/T | G/G | G/A | G/G | G/A | G/G | **intermediate** | 0.99696 |
| **AED_204** | C/T | C/T | C/T | G/A | G/A | G/G | A/A | G/G | **slow** | 0.99873 |
| **AED_249** | C/T | C/C | C/T | G/A | A/A | G/G | A/A | G/G | **slow** | 1 |
| **AED_432** | C/C | C/C | C/T | G/G | G/A | G/G | G/A | G/G | **slow** | 0.99718 |
| **AED_513** | C/C | T/T | C/C | G/G | A/A | G/G | G/G | G/G | **rapid** | 0.97995 |
| **AED_1108** | C/T* | C/C | C/T | G/A* | G/A | G/G | A/A | G/G | **slow** | 1 |
| **AED_1119** | C/C | T/T* | C/T | G/G | G/A | G/G | G/A | G/G | **rapid** | 0.61491 |
| **AED_1135** | C/C | C/C | T/T | G/G | G/G | G/G | A/A | G/G | **slow** | 0.99718 |
| **AEH_I** | C/T | C/T | C/T | G/A | G/A | G/G | A/A | G/G | **slow** | 0.99873 |
| **ALH_1** | C/C | C/T | C/T | G/G | G/A | G/G | G/A | G/G | **intermediate** | 0.99696 |
| **ALH_2** | C/C | C/C | T/T | G/G | G/G | G/G | A/A | G/G | **slow** | 0.99718 |
| **ALH_3** | C/T | T/T | C/C | G/A | A/A | G/G | G/A | G/G | **intermediate** | 0.99704 |
| **ALH_10** | C/C | C/C | T/T | G/G | G/G | G/G | A/A | G/G | **slow** | 0.99718 |
| **BIM_33** | C/C | T/T | C/C | G/G | A/A | G/G | G/A | G/G | **rapid** | 0.97995 |
| **BIM_37** | C/C | C/T | C/T | G/G | G/A | G/G | G/A | G/G | **intermediate** | 0.99696 |
| **FN_2** | T/T | T/T | C/C | A/A | A/A | G/G | A/A | G/G | **slow** | 0.99758 |
| **KER_1** | T/T | C/T | C/C | A/A | A/A | G/G | A/A | G/G | **slow** | 0.99994 |
| **NW_54** | C/C | C/T | C/C | G/G | G/A | G/G | G/A | G/G | **intermediate** | 0.99646 |
| **NW_255** | C/C | C/T | C/T | G/G | G/A | G/G | G/A | G/G | **intermediate** | 0.99696 |
| **PR_4** | C/T | C/C | C/T | G/A* | G/A | G/G | A/A | G/G | **slow** | 1 |
| **PR_10** | C/C | C/T | C/T | G/G | G/A | G/G | G/A | G/G | **intermediate** | 0.99696 |
| **STR_220** | C/T | T/T | C/C | G/A | A/A | G/G | G/A | G/G | **intermediate** | 0.99704 |
| **STR_228** | C/C | T/T | C/C | G/G | A/A | G/G | G/G | G/G | **rapid** | 0.97995 |
| **STR_241** | C/T | T/T | C/C | G/A | A/A | G/G | G/A | G/G | **intermediate** | 0.99704 |
| **STR_248** | T/T | T/T | C/C | A/A | A/A | G/G | A/A | G/G | **slow** | 0.99758 |
| **STR_266** | C/C | T/T | C/C | G/G | A/A | G/G | G/G | G/G | **rapid** | 0.97995 |
| **STR_300** | C/T | C/T | C/T | G/A | G/A | G/G | A/A | G/G | **slow** | 0.99873 |
| **STR_310** | C/C | T/T | C/C | G/G | A/A | G/G | G/G | G/G | **rapid** | 0.97995 |
| **STR_316** | C/C | C/C | T/T | G/G | G/G | G/G | A/A | G/G | **slow** | 0.99718 |
| **STR_328** | C/T* | T/T | C/C | G/G* | A/A | G/G | G/G | G/G | **intermediate** | 0.92143 |
| **STR_355** | C/T | C/C | C/T | G/A | G/A | G/G | A/A | G/G | **slow** | 1 |
| **STR_360** | C/C | C/C | T/T | G/G | G/G | G/G | A/A | G/G | **slow** | 0.99718 |
| **STR_393** | C/T | T/T | C/C | G/A | A/A | G/G | A/A | G/G | **intermediate** | 0.99704 |
| **STR_480** | C/C | C/T | C/T | G/G | G/A | G/G | A/A | G/G | **intermediate** | 0.99696 |
| **STR_486** | C/C | C/T | C/T | G/G | G/A | G/G | G/A | G/G | **intermediate** | 0.99696 |
| **STR_491** | C/C | C/C | T/T | G/G | G/G | G/G | A/A | G/G | **slow** | 0.99718 |
| **STR_502** | C/C | T/T | C/C | G/G | A/A | G/G | G/G | G/G | **rapid** | 0.97995 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **STR_535** | C/T | T/T | C/C | G/A | A/A | G/G | G/A | G/G | **intermediate** | 0.99704 |
| **VIM_2** | C/T | C/T | C/T | G/A | G/A | G/G | A/A | G/G | **slow** | 0.99873 |

**Table S14** NAT2 phenotype frequencies (%) of the modern samples as published by García-Closas et al. (89) and subgroups of ancient Bavarian samples. 95% CIs are given in parentheses.

| Group | rapid (%) | intermediate (%) | slow (%) |
|---|---|---|---|
| New England Bladder Cancer Study | 5.3<br>(4.4-6.2) | 34.2<br>(32.2-36.2) | 60.3<br>(58.2-62.4) |
| Spanish Bladder Cancer Study | 4.9<br>(3.7-6.1) | 35.4<br>(32.8-38) | 59.7<br>(57.0-62.4) |
| Medieval Bavarians (total) | 19.4<br>(6.5-32.3) | 41.7<br>(25.6-57.8) | 38.9<br>(23.0-54.8) |
| Deformed | 33.3<br>(2.5-64.1) | 44.4<br>(11.9-76.9) | 22.2<br>(0-49.4) |
| Non-Deformed | 8.7<br>(0-20.2) | 47.8<br>(27.4-68.2) | 43.5<br>(23.2-63.8) |
| Male | 0 | 77.8<br>(50.6-100) | 22.2<br>(0-49.4) |
| Female | 26.9<br>(9.9-43.9) | 30.8<br>(13.1-48.5) | 42.3<br>(23.3-61.3) |

**Table S15** Determination of CYP2C19 phenotypes. Haplotypes marked with "?" are not completely secure.

| SNP | rs12248560 | rs28399504 | rs41291556 | rs72558184 | rs4986893 | rs4244285 | rs72558186 | rs56337013 | rs17884712 | rs6413438 | rs4417205 | rs3758580 | rs12778026 | rs4304692 | rs7067866 | rs1934967 | rs11568729 | rs4986894 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ancestral/ Derived | C/T,A | A/G | T/C | G/A | G/A | G/A | T/A | C/T | G/A | T/C | G/C | T/C | A/T | C/T | T/G | C/T | T/C | C/T | Haplotype | Phenotype |
| AED_92 | C/C | - | T/T | G/G | G/G | G/A | - | C/C | G/G | C/C | C/G | C/T | T/A | T/T | G/T | C/T | C/C | C/T | *1/*2? | intermediate |
| AED_106 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | - | T/T | T/T | G/G | C/C | C/C | T/T | *1/*1? | extensive |
| AED_125 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | C/C | T/T | T/T | G/T | C/C | C/C | T/T | *1/*1? | extensive |
| AED_204 | C/C | - | T/T | G/G | G/G | G/A | - | C/C | G/G | C/C | C/G | C/T | A/T | T/T | G/T | C/C | C/C | C/T | *1/*2? | intermediate |
| AED_249 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | - | T/T | C/T | G/G | C/C | C/C | C/C | *1/*1? | extensive |
| AED_432 | - | - | T/T | G/G | G/G | G/G | - | C/C | - | C/C | C/C | - | T/T | T/T | G/T | C/T | C/C | T/T | *1/*1? | extensive |
| AED_513 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | - | T/T | C/C | G/T | C/C | C/C | T/T | *1/*1? | extensive |
| AED_1108 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | - | T/T | T/T | G/G | C/T | C/C | T/T | *1/*1? | extensive |
| AED_1119 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | - | C/C | - | T/T | T/T | G/G | C/C | C/C | T/T | *1/*1? | extensive |
| AED_1135 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | C/C | T/T | C/C | G/G | C/T | C/C | T/T | *1/*1? | extensive |
| AEH_I | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | C/C | T/T | T/T | G/T | C/C | C/C | T/T | *1/*1? | extensive |
| ALH_1 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | C/C | T/T | T/T | G/T | C/T | C/C | T/T | *1/*1? | extensive |
| ALH_2 | C/T | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | C/C | T/T | T/T | G/T | C/C | C/C | T/T | *1/*1+*17? | extensive |
| ALH_3 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | - | - | T/T | T/T | G/G | C/C | C/C | T/T | *1/*1? | extensive |
| ALH_10 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | C/C | T/T | T/T | T/T | C/C | C/C | T/T | *1/*1? | extensive |
| BIM_33 | C/C | - | T/T | G/G | G/G | - | - | C/C | G/G | C/C | C/G | - | A/T | T/T | G/T | C/T | C/C | C/T | *1/*2? | intermediate |
| BIM_37 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | - | T/T | T/T | G/G | C/T | C/C | T/T | *1/*1? | extensive |
| FN_2 | C/T | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | C/C | T/T | T/T | G/T | C/C | C/C | T/T | *1/*1+*17? | extensive |
| KER_1 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | C/C | T/T | T/T | G/T | C/C | C/C | T/T | *1/*1? | extensive |
| NW_54 | C/T | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | C/C | T/T | T/T | G/T | C/T | C/C | T/T | *1/*1? | extensive |
| NW_255 | C/C | - | T/T | G/G | G/G | - | - | C/C | G/G | C/C | C/G | - | A/T | T/T | G/T | C/C | C/C | C/T | *1/*2? | intermediate |
| PR_4 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | - | T/T | T/T | G/G | C/T | C/C | T/T | *1/*1? | extensive |
| PR_10 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | - | T/T | T/T | G/G | C/T | C/C | T/T | *1/*1? | extensive |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| STR_220 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | - | T/T | T/T | G/G | C/C | C/C | T/T | *1/*1? | extensive |
| STR_228 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | - | T/T | T/T | G/G | C/T | C/C | T/T | *1/*1? | extensive |
| STR_241 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | C/C | T/T | T/T | G/T | C/C | C/C | T/T | *1/*1? | extensive |
| STR_248 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | C/C | T/T | T/T | G/G | C/T | C/C | T/T | *1/*1? | extensive |
| STR_266 | C/C | - | T/T | G/G | G/G | - | - | C/C | G/G | C/C | C/C | - | T/T | T/T | G/T | C/C | C/C | T/T | *1/*1? | extensive |
| STR_300 | C/T | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | C/C | T/T | T/T | T/T | C/C | C/C | T/T | *1/*1+*17? | extensive |
| STR_310 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | - | C/C | C/C | - | T/T | T/T | G/T | C/C | C/C | T/T | *1/*1? | extensive |
| STR_316 | - | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | - | T/T | T/T | G/T | C/T | C/C | T/T | *1/*1? | extensive |
| STR_328 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | - | C/C | C/C | - | T/T | T/T | G/G | C/T | C/C | T/T | *1/*1? | extensive |
| STR_355 | - | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | - | T/T | T/T | G/T | C/T | C/C | T/T | *1/*1? | extensive |
| STR_360 | C/T | - | T/T | G/G | G/G | G/A | - | C/C | G/G | C/C | C/G | - | A/T | T/T | T/T | C/C | C/C | C/T | *1/*2+*17? | undetermined |
| STR_393 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | C/C | T/T | T/T | G/G | C/T | C/C | T/T | *1/*1? | extensive |
| STR_480 | - | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | - | T/T | T/T | T/T | C/C | C/C | T/T | *1/*1? | extensive |
| STR_486 | C/C | - | T/T | G/G | G/G | G/A | - | C/C | G/G | C/C | C/G | C/T | A/T | T/T | G/T | C/T | C/C | C/T | *1/*2? | intermediate |
| STR_491 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | - | T/T | T/T | G/T | C/T | C/C | T/T | *1/*1? | extensive |
| STR_502 | C/C | - | T/T | G/G | G/G | - | - | C/C | G/G | C/C | - | - | A/T | - | G/T | C/C | C/C | C/T | *1/*2? | intermediate |
| STR_535 | C/C | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | C/C | T/T | T/T | G/T | C/T | C/C | T/T | *1/*1? | extensive |
| VIM_2 | - | - | T/T | G/G | G/G | G/G | - | C/C | G/G | C/C | C/C | - | T/T | T/T | G/T | C/C | C/C | T/T | *1/*1? | extensive |

**Table S16** Frequencies (%) of the major CYP2C19 phenotypes in modern as well as ancient samples. Differences in the frequencies of the two extensive phenotypes between ancient and modern samples due to the fact that high-confidence phenotypes could not be determined for all ancient samples and that there is a huge complexity of different possible combinations with allele *17 (see SI text above). 95% CIs are shown in brackets.

| Group | extensive (*1/*1) | extensive (*1/*17) | intermediate (*1/*2) | indeterminate (*1/*2+*17) |
|---|---|---|---|---|
| Whites of European descent | 42.0<br>(36.9-47.1) | 27.0<br>(22.4-31.6) | 19.0<br>(14.9-23.1) | 4.8<br>(2.6-7.0) |
| Medieval Bavarians (total) | 74.3<br>(59.8-88.8) | 5.7<br>(0-13.4) | 17.1<br>(4.6-29.6) | 2.9<br>(0-8.4) |
| Deformed | 88.9<br>(68.4-100) | 0 | 11.1<br>(0-31.6) | 0 |
| Non-Deformed | 73.9<br>(56.0-91.8) | 4.3<br>(0-12.6) | 17.4<br>(1.9-32.9) | 4.3<br>(0-8.3) |
| Male | 77.8<br>(52.0-100) | 11.1<br>(0-30.6) | 11.1<br>(0-30.6) | 0 |
| Female | 73.1<br>(56.1-90.1) | 3.8<br>(0-11.2) | 19.2<br>(4.1-34.3) | 3.8<br>(0-11.2) |

**Table S17** Genotyping results for SNPs at CYP2D6. Only 15 individuals yielded enough markers to allow for determination of CYP2D6 phenotypes. Haplotypes marked with "?" are not completely secure.

| **SNP** | **rs3892097** | **rs1065852** | **rs16947** | **rs1135840** | | |
|---|---|---|---|---|---|---|
| **Ancestral/Derived** | **C/T** | **G/A** | **G/A** | **G/C** | | |
| **Mutation** | **1846G>A** | **100C>T** | **2850C>T** | **4180G>C** | **Haplotype** | **Phenotype** |
| AED_92 | C/T | G/A | G/A | G/G | *2/*4? | intermediate |
| AED_106 | - | - | - | G/G | ? | |
| AED_125 | C/T | A/A | G/G | G/G | *4/*10 | slow |
| AED_204 | C/T | A/G | G/G | C/G | *1/*4 | intermediate |
| AED_249 | - | - | - | G/G | ? | |
| AED_432 | - | - | - | G/G | ? | |
| AED_513 | - | - | - | - | | |
| AED_1108 | - | - | - | C/G | ? | |
| AED_1119 | - | - | - | C/C | ? | |
| AED_1135 | C/C | G/G | A/G | C/G | *1/*2 | normal |
| AEH_I | C/C | G/G | A/G | C/G | *1/*2 | normal |
| ALH_1 | - | G/G | G/G | C/C | *1/*1 | normal |
| ALH_2 | - | - | A/G | G/G | *2/? | |
| ALH_3 | - | - | - | C/G | ? | |

| | | | | | | |
|---|---|---|---|---|---|---|
| ALH_10 | - | G/G | G/G | C/C | *1/*1? | normal |
| BIM_33 | - | - | - | G/G | ? | |
| BIM_37 | - | - | - | - | | |
| FN_2 | - | - | - | C/C | ? | |
| KER_1 | C/C | G/G | A/G | C/G | *1/*2 | normal |
| NW_54 | C/C | G/G | A/G | C/G | *1/*2 | normal |
| NW_255 | - | - | - | G/G | ? | |
| PR_4 | - | - | - | - | | |
| PR_10 | - | - | - | G/G | ? | |
| STR_220 | - | - | - | C/C | ? | |
| STR_228 | - | - | - | - | | |
| STR_241 | C/T | A/G | G/G | C/G | *1/*4 | intermediate |
| STR_248 | C/C | G/G | A/A | G/G | *2/*2? | normal |
| STR_266 | - | - | - | - | | |
| STR_300 | C/C | G/G | G/G | C/C | *1/*1 | normal |
| STR_310 | - | - | - | - | | |
| STR_316 | - | - | - | - | | |
| STR_328 | - | - | - | - | | |
| STR_355 | - | - | - | C/C | ? | |
| STR_360 | - | - | - | G/G | ? | |
| STR_393 | C/T | A/G | G/G | C/G | *1/*4 | intermediate |
| STR_480 | - | - | - | - | | |
| STR_486 | C/C | G/G | A/A | G/G | *2/*2? | normal |
| STR_491 | - | - | - | C/C | ? | |
| STR_502 | - | - | - | - | | |
| STR_535 | C/T | A/G | G/G | C/G | *1/*4 | intermediate |
| VIM_2 | - | - | - | C/C | ? | |

**Table S18** Genotyping results for SNPs associated with CYP3A5, CYP3A4, and ABCB1, as well as determination of CYP3A alleles.

| **SNP** | **rs776746** | **rs2740574** | **rs1128503** | **rs2032582** | **rs1045642** | | |
|---|---|---|---|---|---|---|---|
| **Gene** | **CYP3A5** | **CYP3A4** | **ABCB1a** | **ABCB1b** | **ABCB1c** | | |
| **Ancestral/Derived** | **T/C** | **C/T** | **G/A** | **C/A,T** | **G/A** | **Haplotypes** | |
| AED_92 | C/C | - | G/G | C/C | A/G | CYP3A5*3 | |
| AED_106 | C/C | T/T | G/G | A/C | A/G | CYP3A5*3 | CY3A4*1 |
| AED_125 | C/C | T/T | G/G | C/C | G/G | CYP3A5*3 | CYP3A4*1 |
| AED_204 | C/T | C/T | A/G | A/C | A/G | CYP3A5*1/*3 | CYP3A4*1/*1B |
| AED_249 | C/T | - | - | A/C | A/G | CYP3A5*1/*3 | |
| AED_432 | C/C | - | - | A/C | A/A | CYP3A5*3 | |
| AED_513 | C/C | - | - | A/A | A/A | CYP3A5*3 | |
| AED_1108 | C/C | - | - | C/T | A/A | CYP3A5*3 | |
| AED_1119 | C/T | - | A/A | A/C | A/G | CYP3A5*1/*3 | |
| AED_1135 | C/C | T/T | A/G | A/C | A/G | CYP3A5*3 | CYP3A4*1 |
| AEH_I | C/C | T/T | A/G | A/C | A/G | CYP3A5*3 | CYP3A4*1 |
| ALH_1 | C/T | T/T | A/G | A/C | A/A | CYP3A5*1/*3 | CYP3A4*1 |
| ALH_2 | T/T | - | A/G | A/C | A/A | CYP3A5*1 | |
| ALH_3 | C/C | - | - | A/A | A/A | CYP3A5*3 | |
| ALH_10 | C/C | T/T | A/A | A/A | A/A | CYP3A5*3 | CYP3A4*1 |
| BIM_33 | C/C | - | - | C/C | G/G | CYP3A5*3 | |
| BIM_37 | C/C | - | - | C/C | G/G | CYP3A5*3 | |
| FN_2 | C/C | T/T | A/G | A/C | A/A | CYP3A5*3 | CYP3A4*1 |
| KER_1 | C/C | - | A/G | A/C | A/G | CYP3A5*3 | |
| NW_54 | C/C | T/T | A/G | A/C | A/G | CYP3A5*3 | CYP3A4*1 |
| NW_255 | C/T | - | G/G | A/C | A/G | CYP3A5*1/*3 | |
| PR_4 | C/C | - | - | A/A | A/A | CYP3A5*3 | |
| PR_10 | C/C | T/T | A/A | A/A | A/G | CYP3A5*3 | CYP3A4*1 |
| STR_220 | C/C | - | - | C/C | A/G | CYP3A5*3 | |
| STR_228 | C/C | - | - | A/A | A/A | CYP3A5*3 | |
| STR_241 | C/C | T/T | A/A | A/A | A/A | CYP3A5*3 | CYP3A4*1 |
| STR_248 | C/C | T/T | A/G | A/C | A/G | CYP3A5*3 | CYP3A4*1 |
| STR_266 | C/C | - | - | A/C | A/G | CYP3A5*3 | |
| STR_300 | C/C | T/T | G/G | C/C | G/G | CYP3A5*3 | CYP3A4*1 |
| STR_310 | C/C | - | - | C/C | A/G | CYP3A5*3 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| STR_316 | C/T | - | - | A/A | A/A | CYP3A5*1/*3 | |
| STR_328 | C/C | - | - | A/C | A/G | CYP3A5*3 | |
| STR_355 | C/C | T/T | - | C/C | G/G | CYP3A5*3 | CYP3A4*1 |
| STR_360 | C/C | - | - | A/C | A/G | CYP3A5*3 | |
| STR_393 | C/T | - | A/G | A/C | A/A | CYP3A5*1/*3 | |
| STR_480 | C/C | - | A/A | A/A | A/A | CYP3A5*3 | |
| STR_486 | C/T | T/T | G/G | C/T | A/G | CYP3A5*1/*3 | CYP3A4*1 |
| STR_491 | C/C | - | - | C/C | G/G | CYP3A5*3 | |
| STR_502 | C/C | - | - | A/C | A/A | CYP3A5*3 | |
| STR_535 | C/C | T/T | A/G | A/C | A/G | CYP3A5*3 | CYP3A4*1 |
| VIM_2 | C/T | - | - | A/A | A/A | CYP3A5*1/*3 | |

**Table S19** Determination of AGT haplotypes via genotyping of 9 relevant markers. Crucial for distinguishing between protective and risk haplotypes are SNPs coding for substitutions A-6G, C4072T, C6309T, and G12775A. Haplotypes marked with "?" are not completely secure.

| **SNP** | **rs5046** | **rs5049** | **rs5050** | **rs5051** | **rs4762** | **rs699** | **rs2493132** | **rs7079** | **rs9431580** | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Ancestral / Derived** | **A/G** | **C/T** | **T/G** | **T/C** | **G/A** | **G/A** | **C/T** | **G/T** | **G/A** | |
| **Mutation** | **C-532T** | **G-217A** | **A-20C** | **A-6G** | **C3389T** | **C4072T** | **C6309T** | **C11535A** | **G12775A** | |
| AED_92 | G/G | C/C | G/T | C/T | - | A/G | C/T | G/G | A/G | H4/H3?,H5? |
| AED_106 | G/G | C/C | T/T | C/C | - | A/A | T/T | G/T | - | H2/H4 |
| AED_125 | G/G | C/C | T/T | C/C | - | A/A | T/T | G/G | A/G | H4/H8? |
| AED_204 | A/G | C/C | T/T | C/C | - | A/G | T/T | G/T | A/G | H2/H6 |
| AED_249 | G/G | C/C | G/T | C/T | - | - | - | G/G | G/G | H3?,H5?/H7 |
| AED_432 | G/G | C/C | T/T | C/C | - | - | - | G/T | A/A | H2/H4 |
| AED_513 | A/G | C/T | T/T | T/T | - | - | - | G/T | - | H2/H6 |
| AED_1108 | G/G | C/C | T/T | C/C | - | - | - | T/T | A/A | H2/H2 |
| AED_1119 | G/G | C/C | G/T | - | - | - | - | G/T | G/G | H2?/H3?,H5? |
| AED_1135 | G/G | C/C | G/T | T/T | - | G/G | C/C | G/G | G/G | H1/H3,5? |
| AEH_I | A/G | C/T | T/T | - | - | A/G | T/T | G/G | A/G | H4/H6 |
| ALH_1 | G/G | C/C | G/T | C/T | - | A/G | C/T | G/T | A/G | H2/H3?,H5? |
| ALH_2 | G/G | C/C | T/T | T/T | - | G/G | C/C | G/G | G/G | H1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ALH_3 | G/G | C/C | T/T | T/T | G/G | - | - | G/G | G/G | H1?, H8? |
| ALH_10 | G/G | C/C | G/G | T/T | - | G/G | C/C | G/G | G/G | H3?, H5? |
| BIM_33 | G/G | C/C | T/T | C/C | - | - | - | T/T | - | H2 |
| BIM_37 | G/G | C/C | T/T | C/C | G/G | - | - | G/T | A/A | H2/H4 |
| FN_2 | A/G | C/T | T/T | C/T | G/G | - | T/T | G/G | A/G | H4/H6 |
| KER_1 | A/G | C/T | T/T | C/T | - | A/G | T/T | G/T | A/G | H2/H6 |
| NW_54 | A/A | T/T | T/T | T/T | G/G | G/G | T/T | G/G | G/G | H6 |
| NW_255 | G/G | C/C | T/T | - | - | A/A | - | T/T | - | H2 |
| PR_4 | A/A | T/T | T/T | T/T | - | - | - | G/G | G/G | H6 |
| PR_10 | G/G | C/C | T/T | T/T | - | - | - | G/G | G/G | H1, H8? |
| STR_220 | G/G | C/C | G/G | - | - | - | - | G/G | G/G | H3, H5? |
| STR_228 | G/G | C/C | T/T | C/C | - | - | T/T | G/T | A/A | H2/H4 |
| STR_241 | G/G | C/C | T/T | C/C | G/G | A/A | T/T | G/T | A/A | H2/H4 |
| STR_248 | A/G | C/T | G/T | T/T | - | G/G | C/T | G/G | G/G | H3?,H5?/H6 |
| STR_266 | G/G | C/C | T/T | C/C | - | - | T/T | G/T | A/A | H2/H4 |
| STR_300 | G/G | C/C | G/T | C/T | G/G | A/G | T/T | G/T | A/A | H2/H4? |
| STR_310 | G/G | C/C | T/T | - | - | - | T/T | G/T | - | H2/H4?,H8? |
| STR_316 | G/G | C/C | T/T | C/C | G/G | - | T/T | G/T | A/A | H2/H4 |
| STR_328 | - | - | - | - | - | - | - | G/G | G/G | |
| STR_355 | G/G | C/C | T/T | C/T | - | - | - | G/G | A/G | H3/H4 |
| STR_360 | G/G | C/C | T/T | C/C | - | - | - | G/G | A/A | H4 |
| STR_393 | G/G | C/C | G/T | C/T | G/G | A/G | C/T | G/T | A/G | H2/H5 |
| STR_480 | G/G | C/C | T/T | - | - | - | - | G/T | - | |
| STR_486 | G/G | C/C | T/T | C/C | - | A/A | T/T | G/T | A/A | H2/H4 |
| STR_491 | G/G | C/C | T/T | - | - | - | - | G/G | - | |
| STR_502 | G/G | C/C | T/T | C/C | - | - | - | T/T | A/A | H2 |
| STR_535 | G/G | C/C | G/T | T/T | - | G/G | C/T | G/T | A/G | H2/H3?,H5? |
| VIM_2 | G/G | C/C | T/T | C/C | - | - | T/T | T/T | - | H2 |

**Table S20** Genotyping of markers relevant for alcohol metabolism (ADH1Bb, ADH1Ba, ALDH2) and lactase persistence (LCTb, MCM6/LCTa). + = lactase persistence, - = lactase non-persistence, / = persistence status indeterminable. *Allele nomenclature for the LCT SNPs follows traditional nomenclature; bases in reference to GRCh37 are given in brackets.

| **SNP** | **rs1229984** | **rs3811801** | **rs671** | **rs182549*** | **rs4988235*** | |
|---|---|---|---|---|---|---|
| **Gene** | ***ADH1Bb*** | ***ADH1Ba*** | ***ALDH2*** | ***LCTb -22018A>G*** | ***MCM6/LCTa -13910C>T*** | **Lactase persistence** |
| **Ancestral/Derived** | **C/T** | **G/A** | **G/A** | **A/G (C/T)** | **C/T (G/A)** | |
| AED_92 | C/C | G/G | G/G | A/A (T/T) | T/T (A/A) | + |
| AED_106 | C/C | G/G | G/G | - | - | / |
| AED_125 | C/C | G/G | G/G | G/A (C/T) | C/T (G/A) | + |
| AED_204 | C/C | G/G | G/G | A/A (T/T) | T/T (A/A) | + |
| AED_249 | C/C | G/G | G/G | A/A (T/T) | T/T (A/A) | + |
| AED_432 | C/C | G/G | G/G | - | - | / |
| AED_513 | C/C | G/G | G/G | G/G (C/C) | C/C (G/G) | - |
| AED_1108 | C/C | G/G | G/G | G/G (C/C) | C/C (G/G) | - |
| AED_1119 | C/C | G/G | G/G | - | T/T (A/A) | + |
| AED_1135 | C/C | G/G | G/G | G/G (C/C) | C/C (G/G) | - |
| AEH_I | C/C | G/G | G/G | G/G (C/C) | C/C (G/G) | - |
| ALH_1 | C/C | G/G | G/G | A/A (T/T) | C/T (G/A) | + |
| ALH_2 | C/C | G/G | G/G | G/G (C/C) | C/C (G/G) | - |
| ALH_3 | C/C | - | G/G | G/G (C/C) | C/C (G/G) | - |
| ALH_10 | C/C | G/G | G/G | A/A (T/T) | T/T (A/A) | + |
| BIM_33 | C/C | G/G | G/G | - | C/T (G/A) | + |
| BIM_37 | C/C | G/G | G/G | - | C/T (G/A) | + |
| FN_2 | C/C | G/G | G/G | G/A (C/T) | C/T (G/A) | + |
| KER_1 | C/C | G/G | G/G | G/G (C/C) | C/C (G/G) | - |
| NW_54 | C/T | G/G | G/G | G/A (C/T) | C/T (G/A) | + |
| NW_255 | C/C | G/G | G/G | G/G (C/C) | C/C (G/G) | - |
| PR_4 | - | G/G | G/G | - | T/T (A/A) | + |
| PR_10 | C/C | G/G | G/G | G/G (C/C) | C/C (G/G) | - |
| STR_220 | C/C | G/G | G/G | G/G (C/C) | C/T (G/A) | + |
| STR_228 | C/C | G/G | G/G | G/A (C/T) | C/T (G/A) | + |
| STR_241 | C/C | G/G | G/G | A/A (T/T) | T/T (A/A) | + |
| STR_248 | C/C | G/G | G/G | G/A (C/T) | C/T (G/A) | + |
| STR_266 | - | G/G | G/G | - | C/T (G/A) | + |

| | | | | | | |
|---|---|---|---|---|---|---|
| STR_300 | C/C | G/G | G/G | G/G (C/C) | C/C (G/G) | - |
| STR_310 | - | G/G | G/G | - | C/C (G/G) | - |
| STR_316 | C/C | G/G | G/G | - | T/T (A/A) | + |
| STR_328 | C/C | G/G | G/G | - | C/C (G/G) | - |
| STR_355 | C/C | G/G | G/G | A/A (T/T) | T/T (A/A) | + |
| STR_360 | C/C | G/G | G/G | G/G (C/C) | C/C (G/G) | - |
| STR_393 | C/C | G/G | G/G | G/G (C/C) | C/C (G/G) | - |
| STR_480 | C/C | G/G | G/G | A/A (T/T) | T/T (A/A) | + |
| STR_486 | C/C | G/G | G/G | G/A (C/T) | C/T (G/A) | + |
| STR_491 | C/C | G/G | G/G | - | T/T (A/A) | + |
| STR_502 | - | G/G | G/G | G/G (C/C) | C/C (G/G) | - |
| STR_535 | C/C | G/G | G/G | G/A (C/T) | C/T (G/A) | + |
| VIM_2 | - | G/A | G/G | - | C/T (G/A) | + |

**Table S21** Derived allele frequency of -13910C>T at rs4988235 within modern populations and the different subgroups of ancient Bavarian samples.

| **Group** | **Frequency (%) derived allele rs4988235** |
|---|---|
| EUR | 50.8 (47.7 – 53.9) |
| CEU | 73.7 (67.6 – 79.9) |
| FIN | 59.1 (52.2 – 65.9) |
| GBR | 72.0 (65.5 – 78.5) |
| IBS | 45.8 (39.1 – 52.5) |
| TSI | 8.9 (5.1 – 12.7) |
| Non-Deformed | 57.1 (42.2 – 72.1) |
| Deformed | 27.8 (7.1 – 48.5) |
| Male | 72.2 (51.5 – 92.9) |
| Female | 36.0 (22.7 – 49.3) |

**Table S22** Fisher Exact results of the comparison of the Baiuvarii allele frequency spectrum of rs4988235 (-13910C>T) with allele frequencies of modern European populations. Allele frequencies of modern populations as well as naming of the populations were taken from the Ensembl GRCh37 release 88 (http://grch37.ensembl.org/Homo_sapiens/Variation/Population?db=core;r=2:136608146-136609146;v=rs4988235;vdb=variation;vf=3096507; last retrieved on May 24, 2017).

| | **EUR** | **CEU** | **FIN** | **GBR** | **IBS** | **TSI** | **Deformed** | **Female** |
|---|---|---|---|---|---|---|---|---|
| **Non-Deformed** | 0.4358 | 0.0398 | 0.8638 | 0.0662 | 0.2367 | 1.84E-11 | 0.0501 | |
| **Deformed** | 0.0595 | 0.0001 | 0.0128 | 0.0003 | 0.2160 | 0.0263 | | |
| **Male** | 0.0950 | 1 | 0.3236 | 1 | 0.0471 | 3.03E-09 | | 0.0125 |
| **Female** | 0.0433 | 1.36E-06 | 0.0041 | 4.83E-06 | 0.2679 | 7.19E-06 | | |

**Table S23** Determination of type 2 diabetes genetic risk score (GRS) according to Meigs et al. (109).

| SNP | rs7903146 | rs1470579 | rs10811661 | rs864745 | rs5219 | rs12779790 | rs7578597 | rs7754840 | rs7961581 | rs4607103 | rs1111875 | rs10923931 | rs13266634 | rs1153188 | rs1801282 | rs9472138 | rs10490072 | rs689 | rs5215* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Locus | TCF7L2 | IGF2BP2 | CDKN2A/B | JAFZ1 | KCNJ11 | CDC123, CAMK1D | THADA | CDKAL1 | ISPAN8; LGR5 | ADAMTS9 | HHEX | NOTCH2 | SLC30A8 | DCD | PPARg | VEGFA | BCL11A | INS | KCNJ11 | | |
| risk allele/non-risk allele | T/C | C/A | T/C | T/C | T/G | G/A | T/C | C/G | C/T | C/T | C/T | T/G | C/T | A/T | C/G | T/C | T/C | T/A | C/T | Count GRS | Standardized GRS |
| AED_92 | C/T | A/A | T/T | T/T | T/G | A/A | T/C | G/G | C/T | C/C | C/C | G/G | T/T | - | C/G | - | T/C | T/T | C/T | 16 | 17.88 |
| AED_106 | C/T | A/C | T/T | C/C | - | A/A | T/T | C/C | C/T | C/T | T/T | T/T | T/T | A/A | C/C | - | T/T | T/T | T/T | 20 | 22.35 |
| AED_125 | T/T | A/A | T/C | C/C | G/G | A/A | T/T | G/C | C/C | C/T | C/T | G/T | C/T | A/A | C/C | - | T/T | A/T | - | 19 | 21.24 |
| AED_204 | C/C | A/C | T/T | T/C | G/G | A/A | T/T | C/C | T/T | C/C | C/T | G/G | C/C | T/T | C/G | C/T | T/C | A/T | - | 17 | 17.94 |
| AED_249 | C/C | A/A | T/T | T/C | - | - | T/T | G/G | T/T | C/C | C/T | G/G | C/C | A/A | C/C | - | - | A/T | C/C | 16 | 17.94 |
| AED_432 | - | C/C | T/C | T/C | - | G/G | T/C | G/G | T/T | C/C | C/C | G/G | - | A/A | C/G | C/C | T/C | A/T | T/T | 15 | 20.27 |
| AED_513 | C/T | A/C | T/T | T/C | - | - | T/C | G/G | T/T | T/T | C/T | G/G | - | A/A | C/C | - | T/C | T/T | C/T | 17 | 17.81 |
| AED_1108 | C/C | A/A | T/C | T/C | - | - | T/T | C/C | T/T | C/C | C/C | G/G | C/C | T/T | C/C | - | T/T | A/T | T/T | 18 | 21.53 |
| AED_1119 | C/T | A/A | T/T | T/C | - | - | T/T | C/C | T/T | C/T | C/C | G/G | C/C | A/A | G/G | - | T/C | A/T | C/T | 18 | 21.38 |
| AED_1135 | C/C | A/C | T/T | T/T | T/T | A/A | T/T | G/G | T/T | C/T | C/T | G/G | C/T | T/A | C/C | C/C | T/T | A/T | - | 17 | 21.38 |
| AEH_I | C/T | A/A | T/T | T/C | - | A/A | T/T | G/C | T/T | C/T | C/C | G/G | C/C | A/A | C/C | C/T | T/T | A/T | C/T | 21 | 22.17 |
| ALH_1 | C/T | A/C | T/T | T/C | T/G | A/A | T/C | G/C | C/T | C/T | C/T | G/T | C/C | T/A | C/G | C/T | T/T | A/T | - | 20 | 21.11 |
| ALH_2 | C/T | A/C | T/C | C/C | T/T | A/A | T/T | G/G | T/T | C/C | C/C | G/G | C/T | T/A | C/C | - | T/T | A/T | - | 18 | 20.12 |
| ALH_3 | - | A/A | T/T | T/T | G/G | G/G | T/T | G/G | C/T | C/C | C/C | G/G | C/C | - | C/G | - | - | A/T | - | 17 | 23.07 |
| ALH_10 | C/C | C/C | T/T | T/C | - | A/A | T/C | G/C | C/C | C/T | C/C | G/T | C/C | A/A | C/C | C/C | T/T | A/T | C/T | 23 | 24.28 |
| BIM_33 | T/T | A/C | T/T | T/T | - | A/A | T/T | G/G | C/T | C/C | C/T | G/T | T/T | - | C/C | - | - | A/T | T/T | 17 | 21.53 |
| BIM_37 | C/C | A/A | T/C | T/C | T/T | - | T/T | G/G | C/T | C/C | C/T | G/G | C/T | T/A | - | C/C | T/T | A/T | - | 15 | 17.81 |
| FN_2 | C/C | A/A | T/C | C/C | G/G | A/A | T/T | G/C | T/T | C/T | C/T | G/G | C/T | A/A | C/C | - | T/C | T/T | T/T | 14 | 14.78 |
| KER_1 | C/C | A/A | T/T | T/T | T/T | A/G | T/T | G/G | T/T | C/T | C/C | G/G | C/T | A/A | C/C | C/C | T/C | T/T | - | 20 | 21.11 |

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NW_54 | C/T | A/A | T/C | T/C | T/T | A/A | T/T | G/G | T/T | C/T | C/T | G/G | C/T | A/A | C/C | C/T | T/T | T/T | - | 19 | 20.06 |
| NW_255 | C/C | A/C | T/T | T/T | - | A/A | T/C | G/C | T/T | C/T | C/T | G/G | C/T | T/A | C/C | - | T/C | A/T | C/T | 16 | 17.88 |
| PR_4 | C/C | A/A | T/T | T/C | - | - | T/T | G/C | C/T | - | - | G/T | C/C | A/A | C/C | - | T/C | T/T | T/T | 17 | 23.07 |
| PR_10 | C/C | C/C | T/T | T/C | - | - | T/T | C/C | T/T | C/T | C/T | G/G | C/C | A/A | C/G | C/C | T/C | T/T | - | 19 | 22.56 |
| STR_220 | - | A/A | T/T | T/C | G/G | - | T/T | G/C | C/T | C/T | T/T | G/G | - | A/A | C/C | - | T/T | T/T | - | 16 | 21.71 |
| STR_228 | C/C | A/C | T/T | T/T | G/G | - | T/T | G/G | C/T | C/T | T/T | G/G | C/T | A/A | C/C | - | T/T | A/T | - | 17 | 20.19 |
| STR_241 | C/C | A/A | T/T | T/T | T/T | A/A | T/T | G/C | T/T | C/C | C/T | G/G | C/C | A/A | C/G | C/T | T/T | T/T | - | 22 | 23.22 |
| STR_248 | C/T | A/C | T/T | T/C | T/G | A/A | T/T | G/C | T/T | C/T | T/T | G/G | C/C | T/A | C/C | C/C | T/C | T/C | - | 17 | 17.94 |
| STR_266 | C/C | A/C | T/C | T/C | T/T | A/A | T/C | G/C | C/T | C/T | C/T | G/G | C/C | A/A | C/C | C/C | T/C | T/T | - | 19 | 20.06 |
| STR_300 | C/T | A/C | T/C | T/C | T/T | A/G | T/T | G/G | C/T | T/T | C/T | G/G | C/C | A/A | C/C | C/C | T/T | A/T | - | 20 | 21.11 |
| STR_310 | T/T | C/C | T/T | T/T | - | - | T/T | G/G | C/T | C/C | T/T | G/G | C/C | T/T | C/C | C/C | T/C | A/T | C/T | 20 | 22.35 |
| STR_316 | - | A/A | T/C | T/C | - | A/A | T/C | C/C | T/T | C/T | C/C | G/T | C/C | T/A | C/C | - | T/T | T/T | T/T | 18 | 21.38 |
| STR_328 | C/T | A/C | T/C | T/C | - | - | T/T | G/C | - | - | T/T | G/G | C/C | A/A | C/C | - | T/T | A/A | C/T | 16 | 21.71 |
| STR_355 | C/T | C/C | T/T | T/C | G/G | A/A | T/T | G/G | T/T | C/C | C/T | - | C/C | A/A | C/G | C/C | T/T | T/T | - | 20 | 22.35 |
| STR_360 | - | A/A | T/T | T/C | T/T | - | T/T | G/G | T/T | C/C | T/T | G/G | - | T/A | C/G | - | T/T | A/T | - | 14 | 19.00 |
| STR_393 | C/C | A/A | T/T | T/C | G/G | A/G | T/C | G/G | T/T | C/C | C/C | G/G | C/C | A/A | C/C | C/C | T/T | T/T | - | 19 | 20.06 |
| STR_480 | C/C | A/A | T/T | T/C | - | G/G | T/T | G/C | C/T | C/C | C/T | G/G | C/C | A/A | - | - | C/C | - | C/T | 17 | 21.53 |
| STR_486 | C/T | A/A | T/T | T/T | T/G | A/A | T/T | G/G | T/T | C/C | C/T | G/G | C/C | A/A | C/C | C/C | T/T | T/T | - | 21 | 22.17 |
| STR_491 | C/C | A/A | T/T | T/C | - | - | T/T | G/G | T/T | C/C | C/T | G/G | - | A/A | C/C | - | T/C | - | T/T | 13 | 17.64 |
| STR_502 | T/T | A/C | T/T | T/T | - | - | T/T | G/G | T/T | C/C | C/C | G/G | C/C | T/A | C/C | - | T/T | - | C/T | 21 | 26.60 |
| STR_535 | C/T | A/C | T/C | T/T | G/G | A/A | T/T | G/G | T/T | C/C | T/T | G/G | C/C | T/A | C/C | C/T | T/C | A/T | - | 17 | 17.94 |
| VIM_2 | C/C | C/C | T/T | T/C | - | - | T/T | C/C | T/T | C/T | C/C | G/G | C/C | A/A | C/C | - | T/T | T/T | C/T | 23 | 27.31 |

**Table S24** Genotypes for calculation of T2D GRS according to Hivert et al. (111). Presented are the genotypes, the GRS, and the standardized GRS accounting for missing genotypes. *rs5215 was used as a proxy for rs5219.

| SNP | rs7903146 | rs10811661 | rs7754840 | rs1801282 | rs1111875 | rs1470579 | rs5219 | rs13266634 | rs7578597 | rs1552224 | rs10923931 | rs11708067 | rs10010131 | rs7578326 | rs13292136 | rs864745 | rs757210 | rs1531343 | rs4607103 | rs7961581 | rs10830963 | rs243021 | rs231362 | rs4457053 | rs340874 | rs917793 | rs972283 | rs8042680 | rs7957197 | rs2191349 | rs780094 | rs896854 | rs11634397 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gene | TCF7L2 | CDKN2A/2B | CDKAL1 | PPARG | HHEX | IGF2BP2 | KCNJ11 | SLC30A8 | THADA | CENTD2 | NOTCH2 | ADCY5 | WFS1 | IRS1 | CHCHD9 | JAZF1 | HNF1B | HMGA2 | ADAMTS9 | TSPAN8 | MTNR1B | BCL11A | SLC22A18AS | ZBED3 | PROX1 | GCK | TSGA13 | VPS33B | HNF1A | DGKB | GCKR | PLEKHF2 | BCL2A1 | | |
| Risk/Non-Risk | T/C | T/C | C/G | C/G | C/T | C/A | T/C | C/T | T/C | A/C | T/G | A/G | G/A | A/G | C/T | T/C | T/C | C/G | C/T | C/T | G/C | A/G | G/A | G/A | C/T | T/A | G/A | A/C | T/A | T/G | C/T | T/C | G/A | | |
| OR | 1,37 | 1,26 | 1,25 | 1,18 | 1,17 | 1,17 | 1,16 | 1,15 | 1,15 | 1,14 | 1,13 | 1,12 | 1,11 | 1,11 | 1,11 | 1,1 | 1,1 | 1,1 | 1,09 | 1,09 | 1,09 | 1,08 | 1,08 | 1,08 | 1,07 | 1,07 | 1,07 | 1,07 | 1,07 | 1,06 | 1,06 | 1,06 | 1,06 | GRS | Standardized GRS |
| AED_92 | C/T | T/T | G/G | C/G | C/C | A/A | C/T | T/T | C/T | A/A | G/G | A/A | A/G | A/A | C/C | T/T | C/T | G/G | C/C | C/T | C/C | G/G | G/G | A/G | C/T | A/T | G/G | C/C | T/T | G/T | C/T | C/T | A/G | 36 | 37.09 |
| AED_106 | C/T | T/T | C/C | C/C | T/T | A/C | C/C* | T/T | T/T | A/C | G/G | A/A | - | A/A | - | C/C | - | G/G | C/T | C/T | C/C | A/G | A/G | A/A | T/T | A/A | A/G | A/A | | G/G | C/T | C/T | G/G | 26 | 30.48 |
| AED_125 | T/T | C/T | C/G | C/C | C/T | A/A | C/C | C/T | T/T | A/A | G/T | A/G | A/G | A/G | C/C | C/C | T/T | C/G | C/T | C/C | C/G | A/A | A/G | G/G | T/T | A/T | A/G | A/C | T/T | G/G | T/T | T/T | A/G | 38 | 39.15 |
| AED_204 | C/C | T/T | C/C | C/G | C/T | A/C | C/C | C/C | T/T | A/A | G/G | A/A | G/G | G/G | C/C | C/T | C/T | G/G | C/C | T/T | G/G | A/G | G/G | A/A | C/T | A/T | G/G | C/C | T/T | G/T | C/T | C/T | A/G | 38 | 39.15 |
| AED_249 | C/C | T/T | G/G | C/C | C/T | A/A | T/T* | C/C | T/T | A/C | G/G | A/A | - | A/G | C/C | C/T | - | G/G | C/C | T/T | C/G | G/G | A/G | A/A | T/T | A/A | A/G | A/C | A/A | G/G | C/C | C/T | A/A | 27 | 29.61 |
| AED_432 | - | C/T | G/G | C/G | C/C | C/C | C/C* | - | C/T | A/A | G/G | A/A | - | A/G | C/C | C/T | - | G/G | C/C | T/T | C/G | A/G | A/A | A/A | C/T | A/A | A/A | C/C | T/T | G/G | C/C | C/C | G/G | 26 | 30.48 |
| AED_513 | C/T | T/T | G/G | C/C | C/T | A/C | C/T* | - | C/T | A/A | G/G | A/A | - | A/A | - | C/T | T/T | - | T/T | T/T | C/C | G/G | G/G | A/A | C/T | A/A | A/G | A/C | T/T | G/T | C/T | C/C | A/G | 28 | 32.83 |
| AED_1108 | C/C | C/T | C/C | C/C | C/C | A/A | C/C* | C/C | T/T | A/A | G/G | A/A | - | A/G | C/C | C/T | - | G/G | C/C | T/T | - | G/G | A/A | A/A | C/C | A/A | G/G | A/A | T/T | T/T | C/T | T/T | A/G | 35 | 39.67 |
| AED_1119 | C/T | T/T | C/C | G/G | C/C | A/A | C/T* | C/C | T/T | C/C | G/G | A/G | - | G/G | C/C | C/T | - | G/G | C/T | T/T | C/C | A/A | G/G | A/G | C/C | A/A | A/G | A/C | T/T | G/T | C/C | T/T | A/G | 34 | 37.29 |
| AED_1135 | C/C | T/T | G/G | C/C | C/T | A/C | T/T | C/T | T/T | A/A | G/G | A/G | A/A | A/A | C/C | T/T | C/C | C/C | C/T | T/T | C/C | A/G | A/G | A/G | T/T | A/A | G/G | A/C | T/T | G/T | C/C | C/C | A/G | 35 | 36.06 |
| AEH_I | C/T | T/T | C/G | C/C | C/C | A/A | C/T* | C/C | T/T | A/A | G/G | A/A | A/A | A/A | C/C | C/T | C/T | G/G | C/T | T/T | C/C | A/A | G/G | A/G | C/T | A/A | A/G | A/C | A/T | G/T | T/T | C/T | G/G | 37 | 38.12 |
| ALH_1 | C/T | T/T | C/G | C/G | C/T | A/C | C/T | C/C | C/T | A/A | G/T | A/A | G/G | A/G | - | C/T | C/C | G/G | C/T | C/T | C/G | A/G | A/G | A/G | C/C | A/T | A/G | A/C | T/T | G/T | C/C | C/T | A/A | 37 | 39.31 |
| ALH_2 | C/T | C/T | G/G | C/C | C/C | A/C | T/T | C/T | T/T | A/A | G/G | A/A | A/G | A/A | C/C | C/C | - | C/C | C/C | T/T | C/C | G/G | A/A | A/G | C/T | A/A | A/G | A/C | T/T | T/T | C/T | C/T | G/G | 37 | 39.31 |
| ALH_3 | - | T/T | G/G | C/G | C/C | A/A | C/C | C/C | T/T | A/C | G/G | A/A | - | A/G | C/C | T/T | C/C | G/G | C/C | C/T | C/G | A/G | A/G | - | T/T | A/T | G/G | C/C | T/T | T/T | C/C | C/T | A/G | 34 | 38.53 |
| ALH_10 | C/C | T/T | C/G | C/C | C/C | C/C | C/T* | C/C | C/T | A/A | G/T | A/A | G/G | A/A | C/C | C/T | C/C | C/C | C/T | C/C | C/G | A/A | G/G | A/A | C/C | A/T | A/A | C/C | T/T | G/T | C/C | C/C | A/A | 43 | 44.30 |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| BIM_33 | T/T | T/T | G/G | C/C | C/T | A/C | C/C* | T/T | T/T | A/A | G/T | A/A | - | A/A | C/C | T/T | - | G/G | C/C | C/T | - | G/G | A/G | A/A | C/C | A/A | G/G | A/C | T/T | G/G | C/T | C/C | A/G | 34 | 38.53 |
| BIM_37 | C/C | C/T | G/G | - | C/T | A/A | T/T | C/T | T/T | A/A | G/G | A/A | - | A/A | C/C | C/T | - | G/G | C/C | C/T | C/C | A/G | G/G | A/G | C/T | A/A | G/G | C/C | T/T | T/T | C/T | C/T | G/G | 34 | 38.53 |
| FN_2 | C/C | C/T | C/G | C/C | C/T | A/A | C/C | C/T | T/T | A/A | G/G | A/A | G/G | A/G | C/C | C/C | C/C | G/G | C/T | T/T | C/C | A/G | A/G | A/A | C/T | A/A | A/G | A/C | T/T | T/T | C/T | C/C | G/G | 30 | 30.91 |
| KER_1 | C/C | T/T | G/G | C/C | C/C | A/A | T/T | C/T | T/T | A/A | G/G | A/G | A/A | A/G | C/C | T/T | C/C | C/G | C/T | T/T | C/G | A/G | - | A/A | C/T | A/A | A/A | A/C | T/T | G/T | T/T | C/C | A/G | 29 | 30.81 |
| NW_54 | C/T | C/T | G/G | C/C | C/T | A/A | T/T | C/T | T/T | A/A | G/G | A/G | A/G | A/G | C/C | C/T | C/C | C/G | C/T | T/T | G/G | A/A | A/G | A/A | C/T | A/T | A/G | C/C | A/T | G/G | C/C | C/T | G/G | 34 | 35.03 |
| NW_255 | C/C | T/T | C/G | C/C | C/T | A/C | C/T* | C/T | C/T | A/C | G/G | G/G | - | A/A | C/C | T/T | - | G/G | C/T | T/T | - | A/G | A/G | A/A | T/T | A/A | A/G | A/C | T/T | G/G | C/T | T/T | G/G | 29 | 32.87 |
| PR_4 | C/C | T/T | C/G | C/C | - | A/A | C/C* | C/C | T/T | A/A | G/T | A/A | - | A/A | - | C/T | - | - | - | C/T | C/C | G/G | G/G | G/G | C/T | A/T | G/G | A/C | T/T | G/T | C/C | C/C | A/G | 33 | 41.56 |
| PR_10 | C/C | T/T | C/C | C/G | C/T | C/C | - | C/C | T/T | A/A | G/G | A/A | - | A/A | - | C/T | - | C/C | C/T | T/T | - | A/G | G/G | A/G | C/T | A/A | A/G | C/C | T/T | G/G | C/T | C/T | G/G | 34 | 41.29 |
| STR_220 | - | T/T | C/G | C/C | T/T | A/A | C/C | - | T/T | A/A | G/G | G/G | A/A | A/G | C/C | C/T | - | G/G | C/T | C/T | C/G | A/A | A/A | A/A | C/T | A/A | A/G | C/C | T/T | G/T | C/C | C/T | G/G | 28 | 31.73 |
| STR_228 | C/C | T/T | G/G | C/C | T/T | A/C | C/C | C/T | T/T | A/C | G/G | A/A | - | - | - | T/T | C/C | G/G | C/T | C/T | C/C | A/G | G/G | A/G | C/C | A/A | A/G | A/C | A/T | G/T | T/T | C/T | A/G | 27 | 30.60 |
| STR_241 | C/C | T/T | C/G | C/G | C/T | A/A | T/T | C/C | T/T | A/A | G/G | G/G | - | A/A | C/C | T/T | T/T | G/G | C/C | T/T | C/G | G/G | A/G | G/G | C/T | A/A | A/A | C/C | T/T | G/G | C/C | C/C | G/G | 34 | 36.13 |
| STR_248 | C/T | T/T | C/G | C/C | T/T | A/C | C/T | C/C | T/T | A/C | G/G | A/A | A/G | G/G | C/C | C/T | C/T | C/G | C/T | T/T | C/G | A/G | G/G | A/G | C/C | A/A | A/A | A/C | T/T | T/T | C/C | C/C | A/G | 37 | 38.12 |
| STR_266 | C/C | C/T | C/G | C/C | C/T | A/C | T/T | C/C | C/T | A/C | G/G | A/A | A/A | A/G | - | C/T | - | G/G | C/T | C/T | G/G | G/G | A/G | A/A | C/T | A/A | G/G | C/C | T/T | G/G | C/C | C/T | G/G | 31 | 34.00 |
| STR_300 | C/T | C/T | G/G | C/C | C/T | A/C | T/T | C/C | T/T | A/A | G/G | A/A | A/G | G/G | C/C | C/T | C/C | G/G | T/T | C/T | C/C | A/G | G/G | A/A | T/T | A/A | G/G | C/C | A/T | G/G | T/T | C/T | A/G | 29 | 29.88 |
| STR_310 | T/T | T/T | G/G | C/C | T/T | C/C | C/T* | C/C | T/T | A/A | G/G | - | - | A/G | - | T/T | - | G/G | C/C | C/T | C/C | A/G | G/G | A/G | C/T | A/T | A/A | A/C | T/T | G/T | C/C | C/T | A/G | 35 | 41.03 |
| STR_316 | - | C/T | C/C | C/C | C/C | A/A | C/C* | C/C | C/T | A/A | G/T | G/G | - | A/A | C/T | - | T/T | G/G | C/T | T/T | C/C | A/A | A/A | A/A | C/T | A/A | A/G | A/C | T/T | G/T | C/T | C/T | A/G | 30 | 34.00 |
| STR_328 | C/T | C/T | C/G | C/C | T/T | A/C | C/T* | C/C | T/T | A/A | G/G | A/A | - | G/G | - | C/T | - | G/G | - | - | - | G/G | - | A/A | C/C | - | A/A | C/C | T/T | G/T | C/C | C/C | A/G | 24 | 32.64 |
| STR_355 | C/T | T/T | G/G | C/G | C/T | C/C | C/C | C/C | T/T | A/A | - | A/A | - | A/A | C/C | C/T | - | G/G | C/C | T/T | C/G | G/G | G/G | A/A | C/C | A/A | G/G | A/C | T/T | G/T | C/T | C/T | A/G | 36 | 40.80 |
| STR_360 | - | T/T | G/G | C/G | T/T | A/A | T/T | - | T/T | A/C | G/G | A/A | - | A/G | - | C/T | - | G/G | C/C | T/T | C/C | A/G | - | A/G | T/T | A/T | A/G | C/C | T/T | G/T | C/C | C/T | A/A | 24 | 30.22 |
| STR_393 | C/C | T/T | G/G | C/C | C/C | A/A | C/C | C/C | C/T | A/A | G/G | A/A | A/A | A/A | C/C | C/T | T/T | G/G | C/C | T/T | G/G | A/G | A/G | A/A | T/T | A/A | A/G | A/C | T/T | G/G | C/T | C/T | A/A | 32 | 32.97 |
| STR_480 | C/C | T/T | C/G | C/G | - | A/A | C/T* | C/C | T/T | A/C | G/G | A/G | - | G/G | - | C/T | - | G/G | C/C | C/T | C/G | G/G | G/G | A/G | C/T | A/T | - | C/C | T/T | G/T | T/T | C/T | G/G | 27 | 32.79 |
| STR_486 | C/T | T/T | G/G | C/C | C/T | A/A | C/T | C/C | T/T | A/A | G/G | A/A | A/G | A/A | C/C | T/T | C/T | G/G | C/C | T/T | C/C | A/G | A/G | A/A | T/T | A/A | G/G | A/A | T/T | T/T | C/T | C/C | G/G | 38 | 39.15 |
| STR_491 | C/C | T/T | G/G | C/C | C/T | A/A | C/C* | - | T/T | A/A | G/G | G/G | - | A/A | - | C/T | - | G/G | C/C | T/T | C/C | G/G | A/A | A/A | C/T | A/A | A/G | A/C | T/T | G/T | C/T | C/T | A/G | 23 | 26.97 |
| STR_502 | T/T | T/T | G/G | C/C | C/C | A/C | C/T* | C/C | T/T | A/A | G/G | A/A | - | A/A | C/C | T/T | - | G/G | C/C | T/T | - | A/G | - | A/A | C/T | A/T | A/G | A/A | T/T | G/G | C/C | T/T | G/G | 40 | 46.90 |
| STR_535 | C/T | C/T | G/G | C/C | T/T | A/C | C/C | C/C | T/T | C/C | G/G | A/G | A/A | A/G | C/C | T/T | T/T | G/G | C/C | T/T | C/C | A/G | G/G | A/G | C/C | A/T | G/G | A/C | T/T | T/T | T/T | C/T | G/G | 36 | 37.09 |
| VIM_2 | C/C | T/T | C/C | C/C | C/C | C/C | C/T* | C/C | T/T | A/A | G/G | A/A | - | A/A | C/C | C/T | - | G/G | C/T | T/T | G/G | A/A | A/G | A/A | T/T | A/T | G/G | A/C | T/T | T/T | C/T | C/T | A/G | 41 | 44.97 |

**Table S25** T2D GRS according to Cornelis et al. (110). Presented are genotypes, count GRS, and standardized GRS accounting for missing genotypes for each individual.

| SNP | rs564398 | rs10010131 | rs7754840 | rs4402960 | rs1801282 | rs5215 | rs1111875 | rs13266634 | rs10811661 | rs12255372 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gene | CDKN2A/B | WFS1 | CDKAL1 | IGF2BP2 | PPARG | KCNJ11 | HHEX | SLC30A8 | CDKN2A/B | TCF7L2 | | |
| Risk/ Non-Risk allele | T/C | G/A | C/G | T/G | C/G | C/T | C/T | C/T | T/C | T/G | GRS | Standardized GRS |
| AED_92 | C/T | A/G | G/G | G/G | C/G | C/T | C/C | T/T | T/T | G/T | 9 | 9.00 |
| AED_106 | - | - | C/C | G/T | C/C | T/T | T/T | T/T | T/T | G/T | 8 | 10.00 |
| AED_125 | C/T | A/G | C/G | G/G | C/C | T/T | C/T | C/T | C/T | T/T | 10 | 10.00 |
| AED_204 | T/T | G/G | C/C | G/T | C/G | T/T | C/T | C/C | T/T | G/G | 13 | 13.00 |
| AED_249 | C/C | - | G/G | G/G | C/C | C/C | C/T | C/C | T/T | G/G | 9 | 10.00 |
| AED_432 | - | - | G/G | T/T | C/G | T/T | C/C | - | C/T | T/T | 8 | 11.43 |
| AED_513 | - | - | G/G | G/T | C/C | C/T | C/T | - | T/T | G/T | 8 | 11.43 |
| AED_1108 | T/T | - | C/C | G/G | C/C | T/T | C/C | C/C | C/T | G/G | 11 | 12.22 |
| AED_1119 | T/T | - | C/C | G/G | G/G | C/T | C/C | C/C | T/T | G/T | 12 | 13.33 |
| AED_1135 | T/T | A/A | G/G | G/T | C/C | C/C | C/T | C/T | T/T | G/G | 11 | 11.00 |
| AEH_I | C/T | A/A | C/G | G/G | C/C | C/T | C/C | C/C | T/T | G/T | 12 | 12.00 |
| ALH_1 | C/T | G/G | C/G | G/T | C/G | C/T | C/T | C/C | T/T | G/T | 13 | 13.00 |
| ALH_2 | C/T | A/G | G/G | G/T | C/C | C/T | C/C | C/T | C/T | T/T | 12 | 12.00 |
| ALH_3 | T/T | - | G/G | G/G | C/G | C/T | C/C | C/C | T/T | G/T | 11 | 12.22 |
| ALH_10 | - | G/G | C/G | T/T | C/C | C/T | C/C | C/C | T/T | G/G | 14 | 15.56 |
| BIM_33 | C/C | - | G/G | G/T | C/C | T/T | C/T | T/T | T/T | T/T | 8 | 8.89 |
| BIM_37 | - | - | G/G | G/G | - | C/C | C/T | C/T | C/T | G/G | 5 | 7.14 |
| FN_2 | T/T | G/G | C/G | G/G | C/C | T/T | C/T | C/T | C/T | G/G | 10 | 10.00 |
| KER_1 | T/T | A/A | G/G | G/G | C/C | C/C | C/C | C/T | T/T | G/G | 11 | 11.00 |
| NW_54 | C/C | A/G | G/G | G/G | C/C | C/C | C/T | C/T | C/T | G/T | 9 | 9.00 |
| NW_255 | C/T | - | C/G | G/T | C/C | C/T | C/T | C/T | T/T | G/G | 10 | 11.11 |
| PR_4 | T/T | - | C/G | G/G | C/C | T/T | - | C/C | T/T | G/G | 9 | 11.25 |
| PR_10 | - | - | C/C | T/T | C/G | - | C/T | C/C | T/T | G/G | 10 | 14.29 |
| STR_220 | C/T | A/A | C/G | G/G | C/C | T/T | T/T | - | T/T | G/T | 7 | 7.78 |
| STR_228 | C/T | - | G/G | G/T | C/C | T/T | T/T | C/T | T/T | G/T | 8 | 8.89 |
| STR_241 | C/T | - | C/G | G/G | C/G | C/C | C/T | C/C | T/T | G/G | 10 | 11.11 |
| STR_248 | C/C | A/G | C/G | G/T | C/C | C/T | T/T | C/C | T/T | G/T | 11 | 11.00 |
| STR_266 | C/T | A/A | C/G | G/T | C/C | C/C | C/T | C/C | C/T | G/G | 11 | 11.00 |
| STR_300 | C/T | A/G | G/G | G/T | C/C | C/C | C/T | C/C | C/T | G/T | 12 | 12.00 |
| STR_310 | C/T | - | G/G | G/T | C/C | C/T | T/T | C/C | T/T | T/T | 11 | 12.22 |
| STR_316 | T/T | - | C/C | G/G | C/C | T/T | T/T | C/C | C/T | G/T | 10 | 11.11 |
| STR_328 | - | - | C/G | G/T | C/C | C/T | T/T | C/C | C/T | G/T | 9 | 11.25 |
| STR_355 | T/T | - | G/G | T/T | C/G | T/T | C/T | C/C | T/T | G/T | 11 | 12.22 |
| STR_360 | - | - | G/G | G/G | C/G | C/C | T/T | - | T/T | G/T | 6 | 8.57 |
| STR_393 | C/T | A/A | G/G | G/G | C/C | T/T | C/C | C/C | T/T | G/G | 9 | 9.00 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| STR_480 | C/T | - | C/G | G/G | - | C/T | C/T | C/C | T/T | G/G | 8 | 10.00 |
| STR_486 | T/T | A/G | G/G | G/G | C/C | C/T | C/T | C/C | T/T | G/T | 12 | 12.00 |
| STR_491 | - | - | G/G | G/G | C/C | T/T | C/T | - | T/T | G/T | 6 | 8.57 |
| STR_502 | T/T | - | G/G | G/T | C/C | C/T | C/C | C/C | T/T | T/T | 14 | 15.56 |
| STR_535 | T/T | A/A | G/G | G/T | C/C | T/T | T/T | C/C | C/T | G/T | 9 | 9.00 |
| VIM_2 | T/T | - | C/C | T/T | C/C | C/T | C/C | C/C | T/T | G/G | 15 | 16.67 |

**Table S26** Mean values of standardized GRS and standard deviation (sd) for groups of individuals according to Meigs et al. (109), Cornelis et al. (110), and Hivert et al. (111).

| | **Meigs et al. (2008)** | | **Cornelis et al. (2009)** | | **Hivert et al. (2011)** | |
|---|---|---|---|---|---|---|
| **Group** | **mean** | **sd** | **mean** | **sd** | **mean** | **sd** |
| all | 20.69 | 2.08 | 10.93 | 1.91 | 35.87 | 4.39 |
| deformed | 20.46 | 1.53 | 10.30 | 1.34 | 35.96 | 3.11 |
| north | 20.56 | 1.94 | 10.98 | 1.88 | 35.78 | 3.86 |
| female | 20.89 | 2.05 | 11.26 | 1.90 | 36.43 | 4.36 |
| female without deformed/south | 20.64 | 1.90 | 11.44 | 1.87 | 36.12 | 4.07 |
| male | 20.44 | 1.99 | 10.26 | 1.67 | 35.25 | 3.64 |

**Table S27** Determination of TCF7L2 haplotypes according to the method of Helgason et al. (112). Haplotypes marked with "?" are not completely secure.

| **SNP** | **rs7903146** | **rs10885406** | **rs12255372** | **rs7924080** | **rs11196199** | |
|---|---|---|---|---|---|---|
| **Ancestral/ Derived** | **T/C** | **G/A** | **G/T** | **T/C** | **A/G** | **Haplotype** |
| AED_92 | C/T | G/G | G/T | C/C | A/A | B/B? |
| AED_106 | C/T | A/G | G/T | C/T | A/A | A/B |
| AED_125 | T/T | G/G | T/T | C/C | A/A | B/B |
| AED_204 | C/C | A/G | G/G | C/T | A/G | A/B? |
| AED_249 | C/C | A/G | G/G | C/T | A/A | A/B |
| AED_432 | T/T | - | T/T | C/C | A/A | B/B |
| AED_513 | C/T | A/G | G/T | C/T | A/A | A/B |
| AED_1108 | C/C | A/A | G/G | T/T | A/A | A/A |
| AED_1119 | C/T | A/G | G/T | C/T | A/A | A/B |
| AED_1135 | C/C | A/A | G/G | T/T | A/A | A/A |
| AEH_I | C/T | G/G | G/T | C/C | A/G | B/B? |
| ALH_1 | C/T | A/G | G/T | C/T | A/A | A/B |

| | | | | | | |
|---|---|---|---|---|---|---|
| ALH_2 | C/T | A/G | T/T | C/T | A/A | A/B |
| ALH_3 | - | G/G | G/T | G/G | - | B?/B? |
| ALH_10 | C/C | A/A | G/G | T/T | A/A | A/A |
| BIM_33 | T/T | G/G | T/T | C/C | A/A | B/B |
| BIM_37 | C/C | A/G | G/G | C/T | A/A | A?/B |
| FN_2 | C/C | A/A | G/G | T/T | A/A | A/A |
| KER_1 | C/C | A/A | G/G | T/T | A/A | A/A |
| NW_54 | C/T | A/G | G/T | C/T | A/A | A/B |
| NW_255 | C/C | A/A | G/G | T/T | A/A | A/A |
| PR_4 | C/C | A/A | G/G | - | A/A | A/A |
| PR_10 | C/C | A/A | G/G | T/T | A/A | A/A |
| STR_220 | C/T | - | G/T | C/T | A/A | A/B |
| STR_228 | C/C | A/G | G/T | C/T | A/A | A/B? |
| STR_241 | C/C | A/A | G/G | T/T | A/A | A/A |
| STR_248 | C/T | A/G | G/T | C/T | A/A | A/B |
| STR_266 | C/C | A/A | G/G | T/T | A/A | A/A |
| STR_300 | C/T | G/G | G/T | C/C | A/G | B?/B |
| STR_310 | T/T | G/G | T/T | C/C | A/A | B/B |
| STR_316 | - | G/G | G/T | C/C | A/G | B/B |
| STR_328 | C/T | A/G | G/T | T/T | A/A | A/B |
| STR_355 | C/T | A/G | G/T | C/T | A/A | A/B |
| STR_360 | - | A/G | G/T | C/T | A/G | A/B |
| STR_393 | C/C | A/A | G/G | T/T | A/A | A/A |
| STR_480 | C/C | A/A | G/G | T/T | A/A | A/A |
| STR_486 | C/T | G/G | G/T | C/C | A/A | B/B |
| STR_491 | C/C | A/G | G/T | C/T | A/A | A/B? |
| STR_502 | T/T | G/G | T/T | C/C | A/A | B/B |
| STR_535 | C/T | G/G | G/T | C/C | A/A | B/B? |
| VIM_2 | C/C | A/A | G/G | T/T | A/A | A/A |

**Table S28** Genotyping of markers relevant for determination of a T2D risk haplotype in SLC16A11 (113), and an additional risk conferring SNP in SLC16A13 (114). Haplotypes marked with "?" are not completely secure.

| **SNP** | **rs75493593** | **rs75418188** | **rs13342232** | **rs13342692** | **rs117767867** | **rs312457** | | |
|---|---|---|---|---|---|---|---|---|
| **Ancestral / Derived** | **G/T** | **C/T** | **G/A** | **C/T** | **C/T** | **G/A** | | |
| **Gene** | ***SLC16A11*** | | | | | ***SLC16A13*** | | |
| **Mutation** | **P443T** | **G340S** | **L187L** | **D127G** | **V113I** | | ***SLC16A11* haplotype** | ***SLC16A13*** |
| AED_92 | G/G | - | A/A | T/T | C/C | A/A | non-risk | non-risk |
| AED_106 | - | - | - | T/T | C/C | A/A | non-risk? | non-risk |
| AED_125 | G/G | - | A/A | T/T | C/C | A/A | non-risk | non-risk |
| AED_204 | G/G | - | A/A | T/T | C/C | A/A | non-risk | non-risk |
| AED_249 | G/G | - | - | - | C/C | A/A | non-risk? | non-risk |
| AED_432 | G/G | - | - | T/T | C/C | A/A | non-risk? | non-risk |
| AED_513 | - | - | - | - | - | - | | |
| AED_1108 | - | - | - | T/T | C/C | A/A | non-risk? | non-risk |
| AED_1119 | - | - | A/A | T/T | C/C | G/G | non-risk | non-risk |
| AED_1135 | G/G | - | A/A | T/T | C/C | A/A | non-risk | non-risk |
| AEH_I | G/G | - | A/G | C/T | C/C | A/A | non-risk/2 SNP | non-risk |
| ALH_1 | G/G | - | A/A | T/T | C/C | A/A | non-risk | non-risk |
| ALH_2 | G/G | - | A/A | T/T | C/C | A/A | non-risk | non-risk |
| ALH_3 | G/G | - | A/A | T/T | C/C | A/A | non-risk | non-risk |
| ALH_10 | G/G | - | A/A | - | C/C | A/A | non-risk? | non-risk |
| BIM_33 | - | - | - | T/T | C/C | A/A | non-risk? | non-risk |
| BIM_37 | G/G | - | - | T/T | C/C | A/A | non-risk? | non-risk |
| FN_2 | G/G | - | A/A | T/T | C/C | A/A | non-risk | non-risk |
| KER_1 | G/G | - | A/A | T/T | C/C | A/A | non-risk | non-risk |
| NW_54 | G/G | - | A/A | T/T | C/C | A/A | non-risk | non-risk |
| NW_255 | G/G | - | A/A | T/T | C/C | A/A | non-risk | non-risk |
| PR_4 | G/G | - | - | - | - | A/A | ? | non-risk |
| PR_10 | G/G | - | - | T/T | C/C | A/A | non-risk? | non-risk |
| STR_220 | G/G | - | - | T/T | - | A/A | non-risk? | non-risk |
| STR_228 | G/G | C/C | - | T/T | C/C | A/A | non-risk | non-risk |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| STR_241 | G/G | - | A/A | T/T | C/C | A/A | non-risk | non-risk |
| STR_248 | G/G | - | A/A | T/T | C/C | A/A | non-risk | non-risk |
| STR_266 | - | - | - | C/T | T/T | - | risk? | |
| STR_300 | G/G | - | A/A | T/T | C/C | A/A | non-risk | non-risk |
| STR_310 | G/G | - | - | T/T | C/C | A/A | non-risk? | non-risk |
| STR_316 | G/G | - | - | T/T | C/C | A/A | non-risk? | non-risk |
| STR_328 | - | - | A/A | T/T | C/C | A/A | non-risk? | non-risk |
| STR_355 | G/G | - | A/A | T/T | C/C | A/A | non-risk | non-risk |
| STR_360 | - | - | A/A | T/T | C/C | A/A | non-risk? | non-risk |
| STR_393 | G/G | - | A/A | T/T | C/C | A/A | non-risk | non-risk |
| STR_480 | - | - | - | T/T | C/C | A/A | non-risk? | non-risk |
| STR_486 | G/G | - | A/A | T/T | C/C | A/A | non-risk | non-risk |
| STR_491 | - | - | A/A | T/T | - | A/A | non-risk? | non-risk |
| STR_502 | - | - | - | T/T | - | A/A | ? | non-risk |
| STR_535 | G/G | - | A/A | T/T | C/C | A/A | non-risk | non-risk |
| VIM_2 | G/G | C/C | A/A | T/T | - | A/A | non-risk | non-risk |

**Table S29** Genotyping results for SNPs associated with common inflammatory diseases according to Raj et al. (115). CD = Crohn's disease, T1D = Type 1 diabetes, MS = Multiple sclerosis, CeD = Celiac disease.

| SNP | rs281379 | rs2188962 | rs10210302 | rs415890 | rs2058660 | rs10761659 | rs13003464 | rs10786436 | rs9388489 | rs12638253 | rs2248359 | rs6822844 | rs17810546 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gene | FUT1/ FUT2 | SLC22A5/C5orf56/ IRF1 | INPP5D/ Atg16L1 | RNASET2 | IL18RAP | ZNF365 | PUS10 | HPSE2 | CENPW | LEKR1 | CYP24A1 | IL2/ IL21 | SCHIP1/ IL12A |
| Ancestral/ Derived | A/G | C/T | T/C | C/G | A/G | A/G | A/G | C/T | A/G | C/T | C/T | G/T | A/G |
| Disease association | CD | CD | CD | CD | CD | CD | CD | T1D | T1D | MS | MS | CeD | CeD |
| AED_92 | A/G | C/C | C/C | C/C | A/A | G/G | A/A | C/T | G/G | | T/T | G/G | A/A |
| AED_106 | A/G | - | - | C/G | A/A | A/G | A/G | C/C | A/A | C/T | C/C | G/G | A/A |
| AED_125 | A/G | C/C | T/T | C/C | A/A | G/G | A/G | C/T | A/G | T/T | C/T | G/G | A/A |
| AED_204 | G/G | C/C | C/T | C/G | A/A | A/G | A/G | C/T | A/A | T/T | C/C | G/G | A/A |
| AED_249 | A/A | - | - | C/G | A/A | A/G | A/A | C/C | G/G | C/T | T/T | G/G | A/A |
| AED_432 | G/G | - | - | C/G | A/A | A/A | A/G | - | A/G | T/T | C/T | G/G | A/A |
| AED_513 | A/A | - | - | C/G | A/A | A/A | A/A | C/C | A/A | T/T | C/C | G/G | A/A |
| AED_1108 | G/G | C/C | - | C/G | - | G/G | A/A | - | G/G | - | C/C | G/G | A/A |
| AED_1119 | G/G | - | - | C/C | A/A | A/G | A/A | C/T | A/A | - | C/C | G/G | A/G |
| AED_1135 | G/G | - | C/T | C/G | A/G | A/G | A/A | T/T | A/A | C/T | T/T | G/G | A/A |
| AEH_I | G/G | C/T | - | C/G | A/A | A/G | A/G | C/C | A/G | C/T | C/C | G/T | A/A |
| ALH_1 | - | C/T | T/T | C/C | A/A | A/G | A/G | - | A/G | C/T | C/T | G/T | A/A |
| ALH_2 | A/G | - | C/C | C/C | A/G | G/G | A/G | C/C | A/G | T/T | C/T | G/T | A/A |
| ALH_3 | G/G | - | - | C/G | - | A/G | A/G | C/C | A/G | - | C/T | G/G | A/A |
| ALH_10 | A/G | T/T | C/T | C/C | A/A | A/A | A/G | C/T | A/A | C/T | C/T | G/G | A/A |
| BIM_33 | A/G | - | - | C/G | A/G | A/A | A/G | T/T | A/G | T/T | T/T | G/T | A/A |
| BIM_37 | G/G | - | - | G/G | A/A | G/G | A/G | C/C | A/G | T/T | C/C | G/G | A/A |
| FN_2 | A/G | C/C | - | C/G | A/G | A/G | G/G | C/C | A/G | C/T | C/T | G/G | A/A |
| KER_1 | A/G | C/C | T/T | C/G | A/A | A/G | G/G | C/T | G/G | C/T | C/C | G/G | A/G |
| NW_54 | G/G | T/T | C/T | C/G | A/G | A/G | A/A | C/C | A/G | C/T | C/T | G/T | A/A |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NW_255 | A/G | C/C | - | C/G | A/A | A/G | G/G | C/C | A/G | - | C/T | G/G | A/G |
| PR_4 | G/G | - | - | C/G | G/G | A/G | A/A | - | A/G | - | C/T | G/G | A/A |
| PR_10 | A/A | - | - | G/G | A/A | A/A | A/A | - | A/G | T/T | C/C | G/G | A/A |
| STR_220 | A/G | C/C | - | C/G | A/A | A/G | A/A | C/T | A/G | - | - | G/G | A/G |
| STR_228 | A/G | - | - | C/G | A/A | G/G | A/A | C/T | A/G | - | C/C | G/G | A/A |
| STR_241 | A/G | T/T | C/T | C/G | A/A | G/G | A/G | C/T | A/A | C/T | T/T | G/G | A/G |
| STR_248 | A/A | C/T | T/T | C/G | A/G | G/G | A/A | C/T | A/G | C/T | C/T | G/G | A/A |
| STR_266 | A/A | - | - | C/G | A/A | A/A | A/G | C/T | A/G | C/C | C/T | G/G | A/A |
| STR_300 | A/G | C/T | T/T | C/G | A/A | A/G | A/A | C/T | A/G | C/T | C/T | G/G | A/G |
| STR_310 | A/G | - | - | G/G | A/A | G/G | A/A | C/C | G/G | C/C | C/T | G/G | A/G |
| STR_316 | - | - | - | C/G | A/A | G/G | A/A | C/T | A/A | C/C | C/C | G/G | A/G |
| STR_328 | - | C/C | - | G/G | A/A | A/G | - | C/C | A/G | C/C | C/T | G/G | A/G |
| STR_355 | G/G | C/C | - | G/G | A/A | A/G | A/G | C/T | A/G | C/T | C/C | T/T | A/A |
| STR_360 | A/A | C/C | - | C/G | A/A | G/G | A/G | C/T | A/G | C/T | C/C | G/T | A/A |
| STR_393 | A/G | C/C | C/T | C/C | A/A | A/G | G/G | C/T | A/G | C/C | T/T | G/G | A/A |
| STR_480 | G/G | - | - | G/G | - | A/G | A/A | C/T | A/G | C/C | C/T | G/G | A/A |
| STR_486 | G/G | C/T | C/T | C/C | A/A | A/G | A/A | C/C | A/A | C/T | T/T | G/T | A/A |
| STR_491 | G/G | - | - | C/G | A/A | A/A | A/G | - | A/G | - | - | G/T | A/A |
| STR_502 | G/G | - | - | G/G | A/A | A/G | - | C/T | A/G | T/T | - | G/G | A/A |
| STR_535 | A/G | - | T/T | C/C | A/A | A/G | A/G | T/T | A/A | C/T | T/T | G/T | A/A |
| VIM_2 | A/A | - | - | C/C | A/A | G/G | G/G | C/C | A/G | - | C/C | G/T | A/G |

**Table S30** Derived allele frequencies for markers associated with common inflammatory diseases according to Raj et al. (115). CD = Crohn's disease, T1D = Type 1 diabetes, CeD = Celiac disease. Frequencies for EUR, CEU, FIN, GBR, IBS, and TSI were retrieved from the Ensembl GRCh37 release 87 (http://grch37.ensembl.org/Homo_sapiens/Info/Index). 95% confidence intervals for ancient Bavarian allele frequencies are given in brackets.

| SNP | rs10210302 | rs2058660 | rs2188962 | rs281379 | rs415890 | rs10786436 | rs12638253 | rs2248359 | rs6822844 | rs17810546 |
|---|---|---|---|---|---|---|---|---|---|---|
| Gene | *INPP5D/Atg16L1* | *IL18RAP* | *SLC22A5/C5orf56/IRF1* | *FUT1/FUT2* | *RNASET2* | *HPSE2* | *LEKR1* | *CYP24A1* | *IL2/IL21* | *SCHIP1/IL12A* |
| Associated disease | CD | CD | CD | CD | CD | T1D | MS | MS | CeD | CeD |
| Selective sweep (ya) | n.a. | ~7500 | ~1380 | n.a. | n.a. | ~6560 | ~5100 | ~8500 | ~2150 | ~2310 |
| Ancestral/Derived | T/C | A/G | C/T | A/G | C/G | C/T | C/T | T/C | G/T | A/G |
| Medieval Bavarians (total) | 0.393 (0.212-0.574) | 0.076 (0.012-0.140) | 0.306 (0.155-0.456) | 0.636 (0.520-0.752) | 0.458 (0.343-0.573) | 0.359 (0.242-0.477) | 0.536 (0.405-0.666) | 0.545 (0.425-0.666) | 0.153 (0.070-0.236) | 0.111 (0.039-0.184) |
| deformed | 0.167 (0-0.465) | 0.125 (0-0.287) | 0.300 (0.016-0.584) | 0.625 (0.388-0.862) | 0.444 (0.215-0.674) | 0.375 (0.138-0.612) | 0.643 (0.392-0.894) | 0.611 (0.386-0.836) | 0.222 (0.030-0.414) | 0.056 (0-0.161) |
| north | 0.500 (0.281-0.719) | 0.071 (0-0.149) | 0.292 (0.110-0.474) | 0.619 (0.472-0.766) | 0.413 (0.271-0.555) | 0.357 (0.212-0.502) | 0.500 (0.337-0.663) | 0.532 (0.375-0.670) | 0.130 (0.033-0.228) | 0.109 (0.019-0.199) |
| male | 0.500 (0.190-0.810) | 0.000 (0-0) | 0.400 (0.096-0.704) | 0.571 (0.312-0.830) | 0.333 (0.115-0.551) | 0.250 (0.038-0.462) | 0.438 (0.194-0.681) | 0.389 (0.164-0.614) | 0.111 (0-0.256) | 0.111 (0-0.256) |
| female | 0.667 (0.449-0.884) | 0.109 (0.019-0.199) | 0.269 (0.099-0.440) | 0.640 (0.507-0.773) | 0.500 (0.364-0.636) | 0.396 (0.257-0.534) | 0.575 (0.422-0.728) | 0.604 (0.466-0.743) | 0.154 (0.056-0.252) | 0.115 (0.029-0.202) |
| EUR | 0.462 | 0.221 | 0.387 | 0.540 | 0.494 | 0.330 | 0.525 | 0.593 | 0.153 | 0.094 |
| CEU | 0.439 | 0.207 | 0.389 | 0.424 | 0.520 | 0.323 | 0.449 | 0.611 | 0.152 | 0.091 |
| FIN | 0.545 | 0.202 | 0.333 | 0.702 | 0.551 | 0.298 | 0.500 | 0.641 | 0.146 | 0.106 |
| GBR | 0.473 | 0.253 | 0.401 | 0.500 | 0.505 | 0.401 | 0.566 | 0.621 | 0.220 | 0.115 |
| IBS | 0.411 | 0.229 | 0.449 | 0.556 | 0.439 | 0.327 | 0.561 | 0.621 | 0.117 | 0.084 |
| TSI | 0.449 | 0.215 | 0.360 | 0.514 | 0.463 | 0.308 | 0.547 | 0.481 | 0.140 | 0.079 |

**Table S31** Results of Fisher exact tests to test for allele frequency differences between groups of ancient Bavarian samples and modern European groups at rs2188962. To conduct the tests, numbers of homozygous and heterozygous individuals for modern European populations were retrieved from Ensembl GRCh37 release 88 (http://grch37.ensembl.org/Homo_sapiens/Variation/Population?db=core;r=5:131770305-131771305;v=rs2188962;vdb=variation;vf=1575398).

| | **EUR** | **CEU** | **FIN** | **GBR** | **IBS** | **TSI** | **Deformed** | **Female** |
|---|---|---|---|---|---|---|---|---|
| **Non-Deformed** | 0.4014 | 0.384 | 0.8195 | 0.3758 | 0.1923 | 0.6536 | 1 | |
| **Deformed** | 0.749 | 0.7442 | 1 | 0.7425 | 0.5183 | 1 | | |
| **Male** | 1 | 1 | 0.7355 | 1 | 1 | 0.7505 | | 0.4539 |
| **Female** | 0.3072 | 0.2853 | 0.6574 | 0.2811 | 0.09515 | 0.3944 | | |

**Table S32** Diversity indices for the groups of samples that were used for analysis of mitochondrial DNA. Mitochondrial HVR I data for comparison were taken from [1]Rott (169), [2]Rott et al. (170), [3]Sofeso et al. (171), [4]Csősz et al. (172), [5]Unterländer et al. (21), [6]Vai et al. (173), [7]Alt et al. (174).

| **Population** | **n** | **k** | **Haplotype Diversity** | **Nucleotide diversity** | **FS** | **FS p-value** |
|---|---|---|---|---|---|---|
| Deformed | 9 | 9 | 1.0000 ± 0.0524 | 0.016788 ± 0.010133 | -4.62107 | 0.009 |
| Non-Deformed | 23 | 19 | 0.9802 ± 0.0198 | 0.013865 ± 0.007910 | -13.32554 | 0 |
| Baiuvars 6th/7th century[1] | 17 | 14 | 0.9706 ± 0.0323 | 0.011407 ± 0.006795 | -8.97715 | 0 |
| Baiuvars 7th/8th century[1,2] | 11 | 11 | 1.0000 ± 0.0388 | 0.013707 ± 0.008266 | -7.6522 | 0 |
| Baiuvars 4th/5th century[3] | 8 | 6 | 0.9286 ± 0.0844 | 0.015576 ± 0.009644 | -0.4842 | 0.37 |
| Avars 6th/7th century[4] | 8 | 8 | 1.0000 ± 0.0625 | 0.011237 ± 0.007252 | -4.9581 | 0 |
| Avars 8th/9th century[4] | 18 | 16 | 0.9869 ± 0.0229 | 0.017083 ± 0.009641 | -9.45658 | 0.001 |
| Avars 9th/10th century[4] | 5 | 5 | 1.0000 ± 0.1265 | 0.016199 ± 0.011021 | -1.28257 | 0.129 |
| Initial Scythians[5] | 3 | 3 | 1.0000 ± 0.2722 | 0.010384 ± 0.009026 | -0.07696 | 0.257 |
| Scythians[5] | 19 | 17 | 0.9883 ± 0.0210 | 0.019092 ± 0.010618 | -9.77518 | 0 |
| Sarmatians[5] | 11 | 11 | 1.0000 ± 0.0388 | 0.022883 ± 0.013089 | -5.32228 | 0.008 |
| Italian Longobards[6] | 28 | 18 | 0.9418 ± 0.0298 | 0.010631 ± 0.006240 | -10.96882 | 0 |
| Hungarian Longobards[7] | 26 | 20 | 0.9723 ± 0.0209 | 0.012279 ± 0.007079 | -14.53894 | 0 |

**Table S33** Results of determination of $F_{st}$ values between "Deformed" and "Non-Deformed" according to Slatkin (128). $F_{st}$ values (lower left triangle) and according p-values (upper right triangle) were determined using Arlequin ver 3.5.2.2. For sources of mitochondrial data for comparison see Table S32.

| Population | Non-Deformed | Deformed | Bavaria 6-7th century | Bavaria 7-8th century | Bavaria 4-5th century | Avars 6-7th century | Avars 8-9th century | Avars 9-10th century | Initial Scythians | Scythians | Sarmatians | Italian Longobards | Hungarian Longobards |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Non-Deformed** | 0 | 0.676 ± 0.046 | 0.270 ± 0.033 | 0.541 ± 0.033 | 0.441 ± 0.039 | 0.207 ± 0.047 | 0.586 ± 0.053 | 0.045 ± 0.020 | 0.703 ± 0.048 | 0.171 ± 0.025 | 0.063 ± 0.014 | 0.198 ± 0.030 | 0.198 ± 0.038 |
| **Deformed** | 0.001 | 0 | 0.973± 0.013 | 0.901± 0.027 | 0.685± 0.058 | 0.838± 0.027 | 0.568± 0.043 | 0.027± 0.014 | 0.964± 0.014 | 0.523± 0.054 | 0.685± 0.028 | 0.396± 0.055 | 0.541± 0.038 |
| **Bavaria 6-7th century** | 0.009 | 0.001 | 0 | 0.496± 0.053 | 0.189± 0.032 | 0.811± 0.036 | 0.117± 0.031 | 0.009± 0.009 | 0.631± 0.053 | 0.216± 0.045 | 0.144± 0.034 | 0.054± 0.020 | 0.009± 0.009 |
| **Bavaria 7-8th century** | 0.001 | 0.001 | 0.001 | 0 | 0.838± 0.039 | 0.333± 0.033 | 0.514± 0.041 | 0.063± 0.024 | 0.613± 0.049 | 0.459± 0.044 | 0.793± 0.033 | 0.514± 0.047 | 0.243± 0.043 |
| **Bavaria 4-5th century** | 0.001 | 0.001 | 0.020 | 0.001 | 0 | 0.324± 0.049 | 0.694± 0.039 | 0.081± 0.029 | 0.595± 0.041 | 0.396± 0.034 | 0.910± 0.021 | 0.162± 0.042 | 0.045± 0.020 |
| **Avars 6-7th century** | 0.022 | 0.001 | 0.001 | 0.010 | 0.046 | 0 | 0.081± 0.025 | 0.000± 0.000 | 0.883± 0.033 | 0.811± 0.030 | 0.198± 0.030 | 0.090± 0.030 | 0.459± 0.052 |
| **Avars 8-9th century** | 0.001 | 0.001 | 0.018 | 0.001 | 0.001 | 0.040 | 0 | 0.027± 0.019 | 0.919± 0.023 | 0.342± 0.038 | 0.721± 0.051 | 0.000± 0.000 | 0.000± 0.000 |
| **Avars 9-10th century** | 0.086 | 0.114 | 0.187 | 0.093 | 0.120 | 0.212 | 0.097 | 0 | 0.270± 0.036 | 0.027± 0.014 | 0.063± 0.019 | 0.000± 0.000 | 0.000± 0.000 |
| **Initial Scythians** | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.056 | 0 | 0.991± 0.003 | 0.928± 0.024 | 0.306± 0.045 | 0.559± 0.056 |
| **Scythians** | 0.016 | 0.001 | 0.015 | 0.001 | 0.001 | 0.001 | 0.006 | 0.087 | 0.001 | 0 | 0.721± 0.038 | 0.000± 0.000 | 0.036± 0.015 |
| **Sarmatians** | 0.022 | 0.001 | 0.020 | 0.001 | 0.001 | 0.020 | 0.001 | 0.071 | 0.001 | 0.001 | 0 | 0.018± 0.012 | 0.018± 0.012 |
| **Italian Longobards** | 0.012 | 0.001 | 0.028 | 0.001 | 0.022 | 0.037 | 0.044 | 0.142 | 0.034 | 0.039 | 0.054 | 0 | 0.135± 0.031 |
| **Hungarian Longobards** | 0.009 | 0.001 | 0.030 | 0.013 | 0.057 | 0.001 | 0.050 | 0.132 | 0.001 | 0.027 | 0.050 | 0.016 | 0 |

**Table S34** Results from contamination estimation based on mitochondrial DNA for all individuals and based on X-chromosomal DNA for all male individuals of which shotgun data was available.

| Sample | Mitochondrial DNA | | | X-chromosomal DNA | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | estimated error rate | authentic (%) | contamination estimate (%) | Method 1 (%) | SE | p-Value | Method 2 (%) | SE | p-Value |
| AED_92 | 0.0016 | 99.03 | 0.638 – 1.439 | | | | | | |
| AED_106 | 0.004 | 98.80 | 0.659 – 1.844 | | | | | | |
| AED_125 | 0.0012 | 99.23 | 0.442 – 1.095 | | | | | | |
| AED_204 | 0.0019 | 97.86 | 1.501 – 3.129 | | | | | | |
| AED_249 | 0.0028 | 99.14 | 0.522 – 1.288 | | | | | | |
| AED_432 | 0.0011 | 99.61 | 0.187 - 0,684 | | | | | | |
| AED_513 | 0.0048 | 94.75 | 4.260 – 6.332 | | | | | | |
| AED_1108 | 0.0042 | 99.42 | 0.270 – 1.006 | | | | | | |
| AED_1119 | 0.0035 | 98.84 | 0.797 – 1.713 | | | | | | |
| AED_1135 | 0.0028 | 99.01 | 0.566 – 1.790 | | | | | | |
| AEH_I | 0.0029 | 98.84 | 0.783 – 1.799 | | | | | | |
| ALH_1 | 0.0017 | 99.57 | 0.159 – 1.014 | 1.447 | 0.0993 | 2.20E-16 | 1.282 | 0.1779 | 2.20E-16 |
| ALH_2 | 0.003 | 99.24 | 0.353 – 1.512 | | | | | | |
| ALH_3 | 0.004 | 99.10 | 0.490 – 1.384 | | | | | | |
| ALH_10 | 0.0018 | 99.97 | 0.006 – 0.477 | | | | | | |
| BIM_33 | 0.0038 | 98.26 | 1.192 – 2.519 | | | | | | |
| BIM_37 | 0.0045 | 99.41 | 0.237 – 1.095 | | | | | | |
| FN_2 | 0.0012 | 99.73 | 0.033 – 0.954 | 1.567 | 0.1132 | 2.20E-16 | 1.316 | 0.1867 | 1.80E-13 |
| KER_1 | 0.0043 | 94.84 | 3.887 – 7.094 | 1.590 | 0.4503 | 0.0004 | 1.934 | 0.7483 | 0.0019 |
| NW_54 | 0.0034 | 98.42 | 1.133 – 2.190 | | | | | | |
| NW_255 | 0.0032 | 98.83 | 0.541 – 2.006 | | | | | | |
| STR_220 | 0.0046 | 98.71 | 0.830 – 1.880 | | | | | | |
| STR_228 | 0.0043 | 99.52 | 0.200 – 0.875 | | | | | | |
| STR_241 | 0.0041 | 99.73 | 0.044 – 1.026 | | | | | | |
| STR_248 | 0.0036 | 99.22 | 0.388 – 1.397 | | | | | | |
| STR_266 | 0.0012 | 99.34 | 0.418 – 1.081 | | | | | | |
| STR_300 | 0.0009 | 99.33 | 0.374 – 1.058 | | | | | | |
| STR_310 | 0.0026 | 99.03 | 0.548 – 1.549 | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| STR_316 | 0.0022 | 99.27 | 0.216 – 1.626 | | | | | | |
| STR_328 | 0.002 | 98.45 | 0.955 – 2.467 | | | | | | |
| STR_355 | 0.0037 | 99.54 | 0.208 – 0.849 | | | | | | |
| STR_360 | 0.0026 | 99.02 | 0.557 – 1.710 | | | | | | |
| STR_393 | 0.002 | 99.43 | 0.285 – 1.093 | | | | | | |
| STR_480 | 0.0028 | 98.35 | 1.092 – 2.328 | | | | | | |
| STR_486 | 0.0022 | 99.65 | 0.172 – 0.671 | 0.736 | 0.5104 | 0.2192 | 1.522 | 0.9224 | 0.0713 |
| STR_491 | 0.0027 | 99.93 | 0.009 – 0.449 | | | | | | |
| STR_502 | 0.0015 | 99.59 | 0.144 – 0.948 | | | | | | |
| STR_535 | 0.0039 | 97.35 | 1.859 – 3.588 | | | | | | |
| VIM_2 | 0.0023 | 98.92 | 0.548 – 1.762 | 1.274 | 0.4469 | 0.0056 | 0.560 | 0.5812 | 0.4851 |

**Table S35** Most likely population origin from population similarity analysis for all ancient samples.

| Sample | Phenotype | Best refpop | Probability | Log Likelihood | Nb of SNPs used |
|---|---|---|---|---|---|
| BIM_37 | m_non-def | UK | 0.53 | -325208 | 601.326 |
| AED_249 | m_non-def | FRE | 0.56 | -421935 | 784.135 |
| STR_486 | m_non-def | FRE | 0.56 | -817831 | 1.504.828 |
| STR_393 | m_non-def | FRE | 0.31 | -265561 | 490.457 |
| AED_106 | m_non-def | GER | 0.86 | -427149 | 790.203 |
| ALH_1 | m_non-def | GER | 0.64 | -2460495 | 4.621.155 |
| STR_241 | m_non-def | GER | 0.3 | -234857 | 435.988 |
| AED_1119 | f_non-def | GER | 0.66 | -232654 | 435.102 |
| AED_92 | m_non-def | GER | 0.48 | -266321 | 492.745 |
| STR_316 | m_non-def | NOW | 0.29 | -493841 | 912.025 |
| | | | | | |
| ALH_2 | f_non-def | FRE | 0.36 | -199988 | 369.189 |
| AED_432 | f_non-def | FRE | 0.39 | -264622 | 491.462 |
| ALH_3 | f_non-def | FRE | 0.51 | -1028164 | 1.903.507 |
| NW_255 | f_non-def | GER | 0.62 | -438121 | 808.395 |
| STR_266 | f_non-def | GER | 0.58 | -154454 | 287.642 |
| STR_480 | f_non-def | GER | 0.52 | -482392 | 891.251 |
| AED_1135 | f_non-def | GER | 0.39 | -396899 | 732.071 |
| ALH_10 | f_non-def | GER | 0.67 | -2463425 | 4.634.247 |
| STR_248 | f_non-def | HUN | 0.41 | -329217 | 604.708 |
| AED_204 | f_non-def | NOW | 0.28 | -352715 | 653.560 |
| STR_502 | f_non-def | TUS | 0.31 | -187039 | 341.464 |
| STR_300 | f_non-def | GRE | 0.44 | -891249 | 1.641.970 |
| | | | | | |
| STR_491 | m_intermediate | GER | 0.47 | -108342 | 200.526 |

| | | | | | |
|---|---|---|---|---|---|
| STR_355 | f_intermediate | GER | 0.34 | -1286343 | 2.373.280 |
| STR_220 | f_intermediate | GER | 0.42 | -1131870 | 2.095.121 |
| STR_360 | f_intermediate | GER | 0.68 | -167764 | 309.943 |
| STR_310 | f_intermediate | ROM | 0.38 | -1483461 | 2.705.440 |
| | | | | | |
| AED_513 | f_def | BUL | 0.69 | -371517 | 684.245 |
| STR_535 | f_def | BUL | 0.3 | -409691 | 748.453 |
| STR_228 | f_def | CRO | 0.39 | -590457 | 1.090.644 |
| AED_125 | f_def | FRE | 0.3 | -123648 | 227.299 |
| BIM_33 | f_def | ROM | 0.43 | -399480 | 738.213 |
| AEH_1 | f_def | ROM | 0.18 | -297956 | 549.882 |
| NW_54 | f_def | SIC | 0.71 | -510023 | 940.779 |
| STR_328 | f_def | TUS | 0.91 | -649438 | 1.178.656 |
| AED_1108 | f_def | NOG | 0.67 | -425068 | 777.077 |
| | | | | | |
| VIM_2 | Other | NOG | 0.94 | -1075867 | 1.956.521 |
| PR_10 | Other | TAJ | 0.96 | -345301 | 631.544 |
| PR_4 | Other | TAJ | 0.8 | -169088 | 307.669 |
| FN_2 | Other | ITN | 0.38 | -2450237 | 4.582.464 |
| KER_1 | Other | TUS | 0.29 | -1084016 | 1.977.874 |
| HI1 | Other | UK | 0.83 | -1504087 | 2.722.385 |
| HS1 | Other | UK | 0.52 | -2180386 | 3.888.836 |
| HS2 | Other | NOW | 0.34 | -2235851 | 3.967.747 |
| HI2 | Other | UK | 0.61 | -2532684 | 4.592.425 |
| HS3 | Other | UK | 0.43 | -1144959 | 2.085.379 |
| O1 | Other | GER | 0.4 | -2085931 | 3.758.121 |
| O2 | Other | GER | 0.67 | -1820179 | 3.265.587 |
| O3 | Other | GER | 0.29 | -2528987 | 4.512.789 |
| O4 | Other | UK | 0.5 | -2496737 | 4.489.572 |
| L | Other | UK | 0.41 | -1173405 | 2.133.620 |

**Table S36** Individual log likelihoods and associated p-values for each individual ancient sample for outlier analysis.

| Sample | Log Likelihood | Empirical P-value |
|---|---|---|
| AED_1108 | -5917 | 0 |
| KER_1 | -5883 | 0 |
| PR_10 | -5850 | 0 |
| VIM_2 | -5844 | 0 |
| STR_502 | -5787 | 0.006 |
| AED_125 | -5723 | 0.091 |
| STR_535 | -5702 | 0.161 |
| STR_310 | -5677 | 0.268 |
| AED_1135 | -5673 | 0.289 |
| AEH_1 | -5670 | 0.306 |
| ALH_1 | -5667 | 0.325 |
| STR_300 | -5661 | 0.369 |
| ALH_3 | -5656 | 0.402 |
| STR_355 | -5652 | 0.427 |
| AED_432 | -5641 | 0.506 |
| STR_486 | -5641 | 0.506 |
| ALH_2 | -5635 | 0.533 |
| AED_1119 | -5634 | 0.538 |
| BIM_37 | -5615 | 0.656 |
| STR_248 | -5611 | 0.677 |
| STR_393 | -5610 | 0.683 |
| NW_54 | -5607 | 0.697 |
| AED_204 | -5604 | 0.713 |
| FN_2 | -5603 | 0.721 |
| BIM_33 | -5602 | 0.723 |
| STR_228 | -5597 | 0.74 |
| AED_513 | -5595 | 0.752 |
| STR_316 | -5581 | 0.828 |
| AED_106 | -5580 | 0.834 |

| | | |
|---|---|---|
| STR_360 | -5580 | 0.834 |
| AED_92 | -5576 | 0.846 |
| STR_220 | -5568 | 0.882 |
| STR_241 | -5558 | 0.912 |
| STR_480 | -5522 | 0.972 |
| AED_249 | -5513 | 0.981 |
| ALH_10 | -5447 | 0.997 |
| NW_255 | -5447 | 0.997 |
| STR_266 | -5453 | 0.997 |

**Table S37** Pairwise $F_{ST}$ estimates for haplotype data at the 5 Mb region for all populations. Colors overlaying numbers indicate smallest (orange) to largest (green) $F_{ST.}$

| FST x 100 | CHB | JPT | CHS | CDX | KHV | GIH | PJL | BEB | STU | ITU | TUR | FIN | TSI | IBS | CEU | GoNL | GBR | m_non-def | f_non-def | f_inter | f_def |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CHB | | | | | | | | | | | | | | | | | | | | | |
| JPT | 0.658 | | | | | | | | | | | | | | | | | | | | |
| CHS | 0.096 | 0.800 | | | | | | | | | | | | | | | | | | | |
| CDX | 0.810 | 1.591 | 0.461 | | | | | | | | | | | | | | | | | | |
| KHV | 0.588 | 1.300 | 0.303 | 0.167 | | | | | | | | | | | | | | | | | |
| GIH | 6.774 | 6.863 | 6.992 | 7.005 | 6.608 | | | | | | | | | | | | | | | | |
| PJL | 6.620 | 6.661 | 6.821 | 6.874 | 6.463 | 0.365 | | | | | | | | | | | | | | | |
| BEB | 5.177 | 5.286 | 5.351 | 5.352 | 4.986 | 0.425 | 0.362 | | | | | | | | | | | | | | |
| STU | 6.315 | 6.386 | 6.498 | 6.496 | 6.105 | 0.456 | 0.363 | 0.229 | | | | | | | | | | | | | |
| ITU | 6.403 | 6.461 | 6.591 | 6.599 | 6.205 | 0.399 | 0.332 | 0.204 | 0.115 | | | | | | | | | | | | |
| TUR | 8.583 | 8.730 | 8.861 | 8.924 | 8.498 | 2.266 | 1.810 | 2.652 | 2.766 | 2.663 | | | | | | | | | | | |
| FIN | 9.460 | 9.600 | 9.719 | 9.854 | 9.405 | 3.255 | 2.754 | 3.602 | 3.842 | 3.754 | 1.589 | | | | | | | | | | |
| TSI | 10.113 | 10.262 | 10.399 | 10.415 | 9.969 | 3.083 | 2.572 | 3.623 | 3.697 | 3.594 | 0.540 | 1.155 | | | | | | | | | |
| IBS | 10.022 | 10.145 | 10.297 | 10.331 | 9.900 | 3.173 | 2.681 | 3.658 | 3.782 | 3.682 | 0.814 | 0.980 | 0.142 | | | | | | | | |
| CEU | 10.138 | 10.263 | 10.409 | 10.447 | 10.015 | 3.174 | 2.653 | 3.662 | 3.795 | 3.704 | 0.974 | 0.636 | 0.317 | 0.222 | | | | | | | |
| GoNL | 10.295 | 10.399 | 10.573 | 10.605 | 10.175 | 3.300 | 2.742 | 3.773 | 3.913 | 3.803 | 1.013 | 0.710 | 0.481 | 0.363 | 0.110 | | | | | | |
| GBR | 10.113 | 10.242 | 10.404 | 10.432 | 10.004 | 3.124 | 2.604 | 3.634 | 3.775 | 3.673 | 0.967 | 0.674 | 0.369 | 0.221 | 0.038 | 0.139 | | | | | |
| m_non-def | 10.347 | 10.484 | 10.639 | 10.721 | 10.229 | 3.451 | 2.872 | 3.925 | 4.091 | 3.939 | 1.325 | 0.810 | 0.670 | 0.564 | 0.269 | 0.275 | 0.337 | | | | |
| f_non-def | 10.021 | 10.124 | 10.264 | 10.323 | 9.877 | 3.219 | 2.606 | 3.599 | 3.775 | 3.652 | 0.929 | 0.719 | 0.339 | 0.274 | 0.112 | 0.144 | 0.191 | 0.217 | | | |
| f_inter | 10.222 | 10.470 | 10.551 | 10.541 | 10.122 | 3.350 | 2.745 | 3.801 | 3.868 | 3.798 | 0.829 | 0.647 | 0.289 | 0.259 | 0.021 | 0.064 | 0.103 | 0.252 | 0.004 | | |
| f_def | 9.547 | 9.640 | 9.824 | 9.874 | 9.457 | 2.922 | 2.346 | 3.322 | 3.408 | 3.304 | 0.570 | 0.779 | 0.165 | 0.173 | 0.170 | 0.240 | 0.217 | 0.307 | -0.061 | 0.150 | |

**Table S38** Free parameter estimates for our demographic model against the observed GoNL-Medieval Barbarian 2D-AFS under two fixed mutation rates and a variable mutation rate.

| **Parameter** | **Bounds** | **$\mu$ = 1.2 $x$ $10^{-8}$** | | **$\mu$ = 2.59 $x$ $10^{-8}$** | | **Variable $\mu$** | |
|---|---|---|---|---|---|---|---|
| | **Lower**<br>**Upper** | **Point** | **5% CI**<br>**95% CI** | **Point** | **5% CI**<br>**95% CI** | **Point** | **5% CI**<br>**95% CI** |
| **Log likelihood** | | **-7591.96** | | **-7606.71** | | **-7591.599** | |
| $log_{10}(\theta_{growth}/\theta_{B2})$ | 0<br>2 | **1.198** | 1.18<br>1.22 | **1.174** | 1.12<br>1.22 | **1.197** | 1.17<br>1.22 |
| $log_{10}(\theta_{present}/\theta_{growth})$ | 0<br>3 | **1.785** | 1.69<br>2.01 | **1.539** | 1.3<br>2.97 | **1.859** | 1.67<br>2.03 |
| $t_{c\_growth}$ **interval** | 0<br>1 | **0.869** | 0.85<br>0.88 | **0.883** | 0.82<br>0.94 | **0.87** | 0.85<br>0.88 |
| $log_{10}(\mu)$ | -10<br>-6 | | | | | **-7.944** | -8.18<br>-7.81 |
| $N_{e\ growth}$ | | **18.711** | 17.7K<br>19.5K | **8.198** | 7.3K<br>9.2K | **19.694** | 14.4K<br>34.5K |
| $N_{e\_present}$ | | **1.140.508** | 896K<br>1,951K | **283.829** | 151K<br>8,645K | **1.422.153** | 751K<br>2,836K |
| $t_{growth}$ **(years)** | | **5.712** | 5.283<br>6.211 | **3.146** | 2.329<br>3.989 | **5.924** | 4,402-10,402 |
| **growth rate (%)** | | **1.8** | 1.63<br>2.16 | **2.82** | 1.88<br>7.2 | **1.81** | 1<br>2.36 |
| $\mu$ | | | | | | **1.14 x $10^{-8}$** | 6.51 x $10^{-9}$<br>1.53 x $10^{-8}$ |

**Table S39** Free parameter estimates for our demographic model against the observed GoNL-Medieval Barbarian 2D-AFS excluding transitions under two fixed mutation rates and a variable mutation rate.

| **Parameter** | **Bounds** | **$\mu$ = 1.2 $x$ $10^{-8}$** | | **$\mu$ = 2.59 $x$ $10^{-8}$** | | **Variable $\mu$** | |
|---|---|---|---|---|---|---|---|
| | **Lower**<br>**Upper** | **Point** | **5% CI**<br>**95% CI** | **Point** | **5% CI**<br>**95% CI** | **Point** | **5% CI**<br>**95% CI** |
| **Log likelihood** | | **-4443.302** | | **-4446.69** | | **-4443.279** | |
| $log_{10}(\theta_{growth}/\theta_{B2})$ | 0<br>2 | **1.204** | 0.53<br>1.3 | **1.158** | 0.62<br>1.27 | **1.207** | 1.13<br>1.27 |
| $log_{10}(\theta_{present}/\theta_{growth})$ | 0<br>3 | **1.657** | 1.26<br>2.05 | **1.354** | 1.15-2.93 | **1.676** | 1.23<br>2.07 |
| $t_{c\_growth}$ **interval** | 0<br>1 | **0.833** | 0.05<br>0.87 | **0.806** | 0.1<br>0.92 | **0.835** | 0.77-0.89 |
| $log_{10}(\mu)$ | -10<br>-6 | | | | | **-8.451** | -8.82<br>-7.95 |
| $N_{e\ growth}$ | | **18.941** | 4.1K<br>22.9K | **7.908** | 2.3K<br>10.1K | **20.880** | 5.9K<br>53.0K |
| $N_{e\_present}$ | | **859.628** | 107K<br>2,212K | **178.666** | 50K<br>8,422K | **989.559** | 101K<br>4,128K |
| $t_{growth}$ **(years)** | | **6.865** | 5.577<br>31.719 | **4.221** | 2.616<br>14.176 | **7.316** | 3,286-15,600 |
| **growth rate (%)** | | **1.39** | 0.26<br>2.5 | **1.85** | 0.55<br>5.59 | **1.32** | 0.6<br>3.14 |
| $\mu$ | | | | | | **1.10 x $10^{-8}$** | 4.62 x $10^{-9}$<br>3.46 x $10^{-8}$ |

Note, $\mu$ is scaled based on titv of 2.1