

Tracing past human mobility and disease in western Eurasia by the genetic analysis of ancient human remains

Dissertation

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. Rer. Nat.)

Vorgelegt dem Rat der
Biologisch-Pharmazeutischen Fakultät der
Friedrich-Schiller-Universität Jena

Von Dipl. Genetik Michal Feldman
Geb. Am 17.06.1980 in Holon

Gutachter:

1. Prof. Dr. Johannes Krause (Max-Planck-Institut für Menschheitsgeschichte, Jena)
2. Prof. Dr. Martin Fischer (Friedrich-Schiller-Universität, Jena)
3. Dr. Maanasa Raghavan (University of Chicago, Chicago)

Dissertation eingereicht am: 24. Juni 2019

Tag der öffentlichen Verteidigung: 27. Januar 2020

Zugl.: Dissertation, Friedrich-Schiller-Universität Jena, [2020]

Table of Contents

1. Introduction.....	4
1.1. Looking back by moving forward – a short history of Archaeogenetics	4
1.2. Human mobility in ancient west Eurasia	7
1.2.1. Pots versus peoples – testing historical questions using aDNA.....	7
1.2.2. Genetic perspectives on the Neolithic transition in the Near East	9
1.2.3. The eastern Mediterranean Bronze Age collapse and the origins of the Philistines	11
1.3. Archaeogenetic perspectives on the Justinianic Plague.....	12
1.3.1. <i>Y. pestis</i> and the historic Plague pandemics	13
1.3.2. An historical overview on the Justinianic Plague	14
1.3.3. The natural history of <i>Y. pestis</i>	15
2. Aim of the thesis	18
3. Overview of Manuscripts and author’s Contribution	20
4. Manuscript A	26
5. Manuscript B.....	37
6. Manuscript C.....	48
7. Discussion	62
7.1. The role of aDNA in recording history	62
7.2. Challenges in DNA retrieval and interpretation	65
8. References.....	71
9. Summary	85
10. Zusammenfassung.....	88
11. Eigenständigkeitserklärung	91
12. Acknowledgments	92
13. Curriculum Vita.....	93
14. Appendix.....	95
14.1. Supplementary Materials of paper A.....	95
14.2. Supplementary Materials of paper B.....	156
14.3. Supplementary Materials of paper C.....	189

1. Introduction

In 1978, Dr. Mary Leakey discovered 3.6 million year old footprints of extinct hominids embedded in hardened volcanic ash in Laetoli, Tanzania. These footprints have led to the striking revelation that hominids could walk upright before they developed larger brains (Leakey and Hay 1979, 317-323). Like Mary Leakey, scholars have often turned to the past in the search of “footprints” to explain the processes that shaped humankind. Archaeologists have been doing so by investigating material culture, biologists have been using fossils to study human evolution and disease and historians have been analyzing ancient texts. Recent technological advances in the field of genetics have made it possible to study genetic material recovered from archaeological specimens, giving rise to a new field of research, often referred to as Archaeogenetics. Over the last decades, discoveries in Archaeogenetics have provided new insights on evolution, ancient diseases and human behavior (Green et al. 2010, 710-722; Reich et al. 2010, 1053; Bos et al. 2011, 506-510; Schuenemann et al. 2013, 179-183; Lazaridis et al. 2014, 409-413; Haak et al. 2015, 207-211).

This thesis focuses on three seminal events in human history that occurred over the last 15,000 years: 1) The Neolithic revolution in the Near East. 2) The 13th century BCE collapse of the Bronze Age eastern Mediterranean civilization that marked the transition into the Iron Ages. 3) The Justinianic Plague pandemic that inflected western Eurasia during the 6th century AD. By analyzing genomic material obtained from human remains of individuals contemporaneous to these events, this thesis attempts to shed light on the biological and the demographical processes preceding and coinciding with these events.

1.1. Looking back by moving forward – a short history of Archaeogenetics

Deoxyribonucleic acid (DNA) is the genetic material that holds the information essential for an organism’s development, function and reproduction. The term ancient DNA (aDNA) refers to DNA recovered from archaeological and paleontological remains (Hofreiter et al. 2001, 353). The most commonly used sources of aDNA are skeletal elements (bones and teeth) that preserve well compared to other organic material and thus are relatively abundant in the archaeological record. Nonetheless, aDNA has been successfully obtained from plants (Yoshida et al. 2013, e00731; Swarts et al. 2017, 512-515), coprolites (Poinar et al. 2003, 1150-1152), hair (Rasmussen

et al. 2010, 757), dental calculus (Warinner et al. 2014, 7104), mummified soft tissue (Salo et al. 1994, 2091-2094) and even from sediment (Slon et al. 2017, 605-608).

Genetic methods aimed to retrieve information on historical or evolutionary processes have been applied on living organisms. However, in contrast to aDNA, modern DNA can only offer indirect evidence and represents only extant lineages. The limitations of aDNA on the other hand, are mostly attributed to macromolecular damage accumulated due to degradation processes such as post-mortem enzymatic cleavage, microbial cleavage, hydrolytic depurination and hydrolytic deamination (Eglinton and Logan 1991, 315-328; Lindahl 1993, 709-715). As a result, only a fraction of the original DNA is preserved in an archaeological specimen (Pääbo 1989, 1939-1943) in a fragmented form. The fragments are typically shorter than 100 bp (Sawyer et al. 2012, e34131). The low amount of DNA is often insufficient for direct sequencing and poses a high risk of contamination by outside modern DNA that is environmentally abundant in much higher copy numbers (Zischler et al. 1995, 1192-1193).

The first aDNA studies that were published during the 1980's, overcame the low copy number hurdle by using bacterial cloning. This work led to the first reports of aDNA retrieval from an Egyptian mummy (Pääbo 1985, 644) and from an extinct Zebra-like animal called *quagga* (Higuchi et al. 1984, 282). However, due to challenging reproducibility, cloning was soon abandoned in favor of targeted DNA amplification by PCR (Polymerase Chain Reaction) (Saiki et al. 1985, 1350-1354; Pääbo and Wilson 1988, 387; Saiki et al. 1988, 487-491). While extremely powerful, PCR can only target a cohort of relatively long DNA fragments and since two ~20 bp long primer sequences are required for PCR amplification, molecules shorter than ~50 bp cannot even be amplified. This limits the availability of genetic information and biases the amplification for longer molecules that often represent contaminating DNA (Pääbo and Wilson 1988, 387). In addition, the lack of appropriate authentication methods has often ignited debates that casted doubts over the validity of some of the early findings (Cano et al. 1994, 2164-2167; Woodward, Weyand, and Bunnell 1994, 1229-1232; Cano and Borucki 1995, 1060-1064; Austin et al. 1997, 467-474). Efforts were made to set criteria aimed to avoid contamination such as the use of separate pre- and post- amplification facilities, negative controls and decontamination procedures (Pääbo 1989, 1939-1943; Cooper and Poinar 2000, 1139; Pääbo et al. 2004, 645-679). Nonetheless, contamination continued to be a threat, since the ancient specimens themselves usually contain a high amount of exogenous DNA.

Within three decades from the first aDNA publications, a new sequencing technology made its advent and has since revolutionized the field of Archaeogenetics. Next Generation sequencing (NGS) allows for the parallel sequencing of millions of untargeted DNA molecules (Margulies et al. 2005, 376; Bentley et al. 2008, 53; Mardis 2008, 387-402). By this, NGS not only dramatically increases sequencing throughputs, but also provides an overview of the metagenomic content within the sampled specimen and thus offers new ways to tackle the contamination issue.

When a DNA extract is prepared for NGS, a so called DNA library is constructed by attaching universal adaptors to the ends of the extracted DNA fragments. This library can subsequently be amplified by utilizing the adaptors as priming sites, resulting in an array of DNA molecules that can be re-amplified and sequenced as necessary (Meyer and Kircher 2010, pdb. prot5448). The adaptors can be tagged by unique short sequences called “indexes” in order to differentiate them from potential lab contamination and to enable the parallel sequencing of different libraries (Kircher, Sawyer, and Meyer 2011, e3-e3). The large amount of sequenced molecules can be subjected to statistical analysis, offering the chance to quantify contamination e.g. by quantifying specific base changes that typically occur at the ends of DNA molecules as a result of deamination processes over time (Briggs et al. 2007, 14616-14621) and are usually not found in contamination (Krause et al. 2010, 894; Sawyer et al. 2012, e34131) or by quantifying polymorphism in regions that are expected to be monomorphic such as the mitochondrial genome or the X chromosome in males (Krause et al. 2010, 894; Rasmussen et al. 2011, 94-98; Korneliussen, Albrechtsen, and Nielsen 2014, 356; Renaud et al. 2015, 224).

Another technological leap came with the development of DNA enrichment techniques. These are considered a “game changer” in the context of DNA retrieval from extremely old remains or remains from geographical regions unfavorable for DNA preservation as well as for the study of ancient pathogens that usually comprise only a miniscule portion of the DNA within an ancient sample. These enrichment assays are mostly based on hybridization techniques in which probes are designed to complement a known reference sequence and therefore capture the targeted DNA molecules while others are washed away (Burbano et al. 2010, 723-725; Maricic, Whitten, and Pääbo 2010, 499-503; Fu et al. 2013, 2223-2227).

In the following years, these technologies described above, together with improvements in sampling techniques have led to a boom in Archaeogenetic publications. Reports on mitochondrial and nuclear genome-wide data recovered from extinct hominines and ancient

modern humans have shed light on their phylogenetic relationships (Green et al. 2010, 710-722; Burbano et al. 2010, 723-725; Reich et al. 2010, 1053; Rasmussen et al. 2010, 757; Krause et al. 2010, 894; Fu et al. 2013, 2223-2227; Seguin-Orlando et al. 2014, 1113-1118; Fu et al. 2014, 445-449; Fu et al. 2015, 216-219). For the first time, whole genomes of ancient pathogens could be reconstructed and compared with present-day lineages (Bos et al. 2011, 506-510; Schuenemann et al. 2013, 179-183). New insights were revealed on the history of regions that were previously poorly explored genetically, such as the Near East (Lazaridis et al. 2016, 419; Broushaki et al. 2016, 499-503; Kilinc et al. 2017, 10.1098/rspb.2017.2064), the Americas (Raghavan et al. 2014, 87; Rasmussen et al. 2014, 225) and the southwest Pacific (Skoglund et al. 2016, 510).

The research reported in this thesis, utilized these technological advancements to reconstruct genome wide data of ancient humans that lived during the studied periods as well as of a 6th century *Y. pestis* genome, the infamous pathogen responsible for the Justinianic Plague.

1.2. Human mobility in ancient west Eurasia

1.2.1. Pots versus peoples – testing historical questions using aDNA

Traditionally, the study of human history was conducted by investigating ancient texts or material cultures. A biological perspective to the study of the human past was introduced when anatomical and pathological studies were applied on ancient human remains within the biological/physical anthropology discipline (Montagu 1960). The use of molecular techniques in this context began as early as the 1940's, with the development of radiocarbon dating (Libby 1954, 5-16) and continued with the use of isotope analysis to investigate the diet and mobility of ancient individuals (Beard and Johnson 2000, 1049-1061) as well as with the study of ancient proteins (Buckley, Melton, and Montgomery 2013, 531-538) and that of microfossils (Henry, Brooks, and Piperno 2011, 486-491).

The earliest Archaeogenetic human studies were mostly based on mitochondrial sequences which were more accessible due to their high copy number in the cell (Willerslev and Cooper 2004, 3-16). Though powerful in determining phylogenetic connections between species this approach is limited since the 16,500 bp long mitochondrial genome is a single locus solely passed through the maternal line and therefore does not represent the full historic or evolutionary picture (Chen and Li 2001, 444-456; Posth et al. 2017, 16046). Moreover, in many cases, mitochondrial genomes do not provide sufficient resolution to differentiate closely related

modern human populations (Novembre et al. 2008, 98). The advent of NGS allowed for genome-wide data to be reconstructed providing millions of analyzable loci that are inherited from both the maternal and paternal lines and thus better represent the probabilistic nature of inheritance (Rasmussen et al. 2010, 757; Green et al. 2010, 710-722; Fu et al. 2013, 2223-2227; Fu et al. 2014, 445-449; Prüfer et al. 2014, 43; Raghavan et al. 2014, 87; Fu et al. 2015, 216-219).

The analysis of human nuclear aDNA has been applied to address various questions, e.g., examining social structures by determining relatedness levels and the genetic sex of individuals (Mittnik et al. 2016, e0163019; Sikora et al. 2017, 659-662); detecting selection signals in search of evolutionary preferable traits; or characterizing phenotypic, metabolic and health related traits (Olalde et al. 2014, 225-228; Mathieson et al. 2015, 499-503).

The development of new frequency- and model- based computational analytic tools (Alexander, Novembre, and Lange 2009, 1655-1664; Durand et al. 2011, 2239-2252; Patterson et al. 2012, 1065-1093; Korneliussen, Albrechtsen, and Nielsen 2014, 356), enabled the application of ancient nuclear data to test theories on historical events related to human mobility by the genetic typing of human groups across time and space. Cultural transitions manifest in the archaeological record as changes in material culture, subsistence strategies, burial customs and artifact styles. Such changes might occur due to local technological or social developments or they might be imported from other human groups. In the latter case, two models can be hypothesized: 1) the new culture has been adopted by the local population following the exchange of knowledge and technology with other populations, a model often referred to as “cultural diffusion” (Frobenius 1897, 225-236, 262-267). 2) The cultural transition involves a large scale mobility of people that relocate to a new region and replace local cultural features with their own in a process termed “demic diffusion” (Cavalli-Sforza, Menozzi, and Piazza 1993, 639-646). These two models, often figuratively referred to as “pots” verses “people”, are indistinguishable when solely considering the material culture. Genetic analysis can in principal differentiate the two, since genetic continuity across the cultural transition supports cultural diffusion while external gene flows point to demic diffusion. The extent of these genetic influxes can potentially be quantified using model based analytical tools (Alexander, Novembre, and Lange 2009, 1655-1664; Patterson et al. 2012, 1065-1093).

Human genetic variation is in general small when compared to other non-human primates (Chen and Li 2001, 444-456). In fact, the average proportion of human genetic variation between individuals from different present-day human populations is only slightly higher than the

variation observed between unrelated individuals from the same population (Rosenberg et al. 2002, 2381-2385), therefore, detecting changes in a human population's genetic makeup can be challenging. Nonetheless, the process of genetic drift causes populations that were relatively isolated from each other over longer periods of time to exhibit different allele frequencies in positions throughout the genome called single nuclear polymorphic sites (SNPs) (International HapMap Consortium 2007, 851). These sites, therefore, are the more informative ones to distinguish between human populations and therefore are often ascertained either *in vitro* by DNA enrichment designed to capture a specific set of SNPs or *in silico*, during data processing, by genotyping (also referred to as variant calling) only a certain SNP set (Fu et al. 2013, 2223-2227; Haak et al. 2015, 207-211; Mathieson et al. 2015, 499-503; Mallick et al. 2016, 201).

The resulting genotyped data can then be used to determine the allele frequency differences between individuals or populations. This approach has yielded striking discoveries over the last decade. For example, sequencing of the Neanderthal and Denisovan genomes revealed that present day Eurasians and Oceanians inherited between 2% and 7% of their genome from these extinct hominines (Green et al. 2010, 710-722; Reich et al. 2010, 1053). Sequencing of prehistoric Europeans revealed that a substantial migration of Near Eastern farmers was responsible for the spread of agriculture into Europe (Skoglund et al. 2012, 466-469; Lazaridis et al. 2014, 409-413; Mathieson et al. 2015, 499-503) and that a massive population movement from the Pontic steppe likely introduced Indo-European languages into Europe during the Bronze Age (Haak et al. 2015, 207-211; Allentoft et al. 2015, 167). In manuscripts A and B of this thesis, I apply similar approaches to examine the genetic shifts between the hunter-gatherers and the first farmers of central Anatolia (manuscript A) as well as between Bronze- and Iron- Age Levantine populations (manuscript B).

1.2.2. Genetic perspectives on the Neolithic transition in the Near East

The Neolithic revolution was the gradual transition of hunting and gathering nomadic communities into a settled, agricultural, food producing subsistence strategy (Barker 2009). This transition began independently in several regions of the world. The earliest evidence for agriculture, estimated to around 11,000 years before present (yBP), is found in the Fertile Crescent of the Near East, a region that includes present-day Iraq, Iran, Syria, Lebanon, Jordan, Israel as well as parts of Egypt and the southeastern fringes of Anatolia (Barker 2009). Within around five millennia, domesticated plants and animals spread from the Fertile Crescent

throughout west Eurasia as farming gradually replaced local hunting and gathering practices (Barker 2009).

Genetic studies that recovered DNA from Mesolithic hunter-gatherers and early farmers of mainland Europe, detected genetic discontinuity between them, indicating a large scale replacement of the local gene pool by a foreign one (Bramanti et al. 2009, 137-140). The European farmers showed a higher affinity to Near Eastern present-day populations pointing to the Near East as the source of admixture (Brandt et al. 2013, 257-261; Lazaridis et al. 2014, 409-413). Genomic data recovered from Ceramic farmers of western Anatolia (~ 8,000 yBP, from the sites of Barçın and Mentese in present-day Turkey), revealed that these early farmers genetically cluster closely with the early European farmers. Moreover, the European farmers could be modeled as deriving their ancestry from the Anatolian farmers and the European hunter-gatherers, flagging the Anatolian farmers as the plausible Near Eastern source (Mathieson et al. 2015, 499-503). The body of evidence formed by these studies suggested that agriculture was introduced to mainland Europe by a demic expansion of early farmers from Anatolia.

The population structure and the genetic history of the Near Eastern first farmers, however, was at that point a mystery, mostly due to the low preservation of DNA in warm climates. Reports that the inner ear preserves DNA exceptionally well (Pinhasi et al. 2015, e0129102), have shifted sampling efforts to focus on the petrous portion of the temporal bone and made it possible to recover late Pleistocene and early Holocene genomes from early farming foci in the Fertile Crescent such as the southern Levant and the Zagros Mountains in Iran (Lazaridis et al. 2016, 419; Broushaki et al. 2016, 499-503). Analysis of these genomes painted a different picture than that observed in Europe. Strikingly, the three sampled Near Eastern Populations from Iran, the Levant and Anatolia were genetically highly differentiated from each other, suggesting long periods of relative isolation between them (Lazaridis et al. 2016, 419). Furthermore, in Iran and the Levant, a high level of genetic continuity was observed between the local Mesolithic hunter-gatherers and the early farmers, suggesting that the advent of agriculture in those regions did not follow the demic diffusion model as in Europe but was due to either cultural diffusion or it developed locally (Lazaridis et al. 2016, 419).

While the above studies have provided the first glimpse into the genetic map of the Near East at the dawn of the Neolithic revolution, many questions remained, mostly due to geographical and temporal gaps in the available data. Central Anatolia for example is a key region in understanding the early spread of farming because it harbors the first evidence of

farming outside the Fertile Crescent. However, the scarcity of human remains in the region has so far made it impossible to recover genome-wide data from Pleistocene hunter-gatherers. In manuscript A of this thesis the first genome-wide data from an Anatolian hunter-gatherer dated to 15,000 yBP and from seven early Near Eastern farmers are retrieved and analyzed to investigate the population structure in the Near East during the Pleistocene and the history and origin of the first farmers of central Anatolia.

1.2.3. The eastern Mediterranean Bronze Age collapse and the origins of the Philistines

The relatively high genetic differentiation between early Holocene and late Pleistocene human populations from the Levant, Iran and Anatolia, decreased in the following millennia as a result of gene exchange between these gene pools, evident from Genome-wide data reconstructed from later Chalcolithic and early Bronze Age Near Eastern populations (Lazaridis et al. 2016, 419; Lazaridis et al. 2017, 214; Harney et al. 2018, 3336; de Barros Damgaard et al. 2018, 10.1126/science.aar7711). Genetic data of Levantine people that postdate the Middle Bronze Age has so far not been available and therefore it is unknown how long this trend of homogenization has persisted. Some indication can be found in the archaeological and historical records that suggest that the Bronze- to Iron- Age transition in the region was a time of extended human mobility (Cline 2015).

The Bronze- to Iron Age transition in the Eastern Mediterranean occurred during the 12th century BCE and was marked by outstanding cultural turbulence (Cline 2015). The Archaeological record comprises of a destruction layer in almost all major eastern Mediterranean cities, some burned to the ground. Civilizations such as the Greek Mycenaeans, the Anatolian Trojans and Hittites, the Mesopotamian Babylonians and the Levantine Canaanites that for the preceding centuries have enjoyed prosperous economies, elaborate trade connections and developed writing systems collapsed within several decades leading to a period sometimes referred to as the “first dark ages” (Cline 2015).

A group of seafaring people referred to as the “Sea Peoples”, have been often suggested as the main culprits of these events (Roberts 2009; Cline 2015). Clues on their identity are found in ancient Egyptian texts from the reign of pharaoh Ramses III (Roberts 2009). These texts describe the Sea peoples as a diverse confederation of populations that left their island homes in a time of war, attacked existing kingdoms in the northeastern Mediterranean and ultimately headed toward Egypt itself (Roberts 2009). One of the mentioned populations is referred to as the

Peleset, a name historians have frequently associated with the “Philistines” mentioned in the Hebrew bible. It has been suggested that both sources describe a migration of this group across the Mediterranean basin and to the southern Levantine coast at the end of the Bronze Age (Dothan and Dothan 1992; Yasur-Landau 2010, 399-410; Cline 2015).

These textual studies motivated archaeological excavations at Ashkelon, Ashdod, Gath, and Ekron, located in present day Israel and identified as four of the five core cities mentioned as “Philistine” in the Hebrew Bible. Comparative analysis of the material culture revealed substantive cultural changes in these coastal cities during the 12th century BCE. The new cultural characteristics resembled those of 13th century Cypriot and Aegean. These led some scholars to suggest that a migration event from the Aegean introduced these cultural changes (Stager 1995, 332-348; Killebrew 1998, 379-405; Yasur-Landau 2010, 399-410). Others have argued for the cultural diffusion model or internal development of ideas to explain this transition (Sherratt 1992, 316-347; Bauer 1998, 149-168). Even among advocates of a migration, alternative geographic sources were proposed (Stager 1995, 332-348; Killebrew 1998, 379-405; Maeir and Hitchcock 2017, 149-162).

In paper B of this thesis, genome-wide data is reconstructed from human remains recovered from the ancient seaport of Ashkelon, identified as “Philistine” during the Iron Age (Killebrew 2013). The comparison of Bronze- and Iron- Age individuals is used to address the historical question of whether the cultural transition observed in the archaeological record is mirrored by a foreign genetic influx and the Ashkelon people throughout the studied period are genetically characterized.

1.3. Archaeogenetic perspectives on the Justinianic Plague

The advent of NGS and the development of DNA enrichment have opened up opportunities to explore the DNA of other organisms present in the host’s remains and a special interest was naturally devoted to human disease causing microbes (Bos et al. 2011, 506-510; Schuenemann et al. 2013, 179-183). Human pathogens have influenced human culture and dispersion along the course of history through epidemics events (Karlsson, Kwiatkowski, and Sabeti 2014, 379). Anthropogenic environmental changes or human mobility on the other hand have often shaped pathogen diversity and distribution (Bellwood 2005, 41-54; Bocquet-Appel and Bar-Yosef 2008). Investigating pathogen’s genetic diversity can therefore not only elucidate

questions related to the natural history of the pathogen but may also help address historical questions, e.g., regarding human communication networks or economic organization.

1.3.1. *Y. pestis* and the historic Plague pandemics

The bubonic plague is a disease caused by the Gram-negative bacterium *Y. pestis* (*Yersinia* Genus, *enterobacteriaceae* kingdom) (Perry and Fetherston 1997, 35-66). Although mostly known for its historical contexts, in present-day, hundreds of new cases of plague are reported each year in regions such as Africa, the FSU, the Americas and Asia (mondiale de la Santé, Organisation and World Health Organization 2016, 89-93).

Humans are not the primary hosts of *Y. pestis* that rather maintains a complex natural cycle which involves transmission between wild rodent species serving as a reservoir and their fleas that serve as the vector. Most human cases result from exposure to infectious fleas when the rodent population is depleted in a process termed epizootic infection (Treille and Yersin 1894, 310-311; Plague Research Commission 1907, 724-798). *Y. pestis* bacilli proliferate and form dense aggregates within the flea gut creating a biofilm that prevents the flea digestion of new blood, leading to more aggressive feeding which increases the transmission rates. Bacilli are refluxed through intradermal bites to their mammalian hosts from where they disseminate with affinity to lymph nodes (Bacot and Martin 1914, 423-439). The complex life cycle of *Y. pestis* with bacilli facing very different conditions when infecting the vector or the different hosts is mirrored in gene expression and epidemiology (Twigg 1984; Christie 1980).

Human plague can take three primary forms: bubonic, pneumonic or septicemic. The incubation period is typically two to six days and the onset is usually sudden and includes appearance of the characteristic bubo lesions accompanied by flu-like symptoms. Plague can be treated with antibiotics which are effective if given on time. Without treatment, bubonic plague has a mortality of around 50% with an average life expectancy in the lethal cases between three to six days (Christie 1980). The pneumonic form occurs in about 5% of the bubonic form infected patients when the bacteria infiltrate the lungs. In this case, sputum coughed out by the patient can be inhaled by others and infect them resulting in the pneumonic form. The pneumonic form is fatal in more than 90% of all cases with and without treatment (Christie 1980). The highly lethal septicemic form does not cause buboes and occurs when the bacteria is introduced directly into the blood stream (Hull, Montes, and Mann 1987, 113-118).

Y. pestis is considered to have been responsible for at least three historic pandemics. The most recent one started at China during the 19th century and within about twenty years has globally spread via all major ports of that time (Pollitzer 1954, 409-482). In 1894, Alexandre Yersin identified *Y. pestis* as the causative agent (Treille and Yersin 1894, 310-311; Zietz and Dunkelberg 2004, 165-178). The second plague pandemic began with the infamous Black Death between the years 1347 and 1351 and was estimated to have claimed the lives of 30-50% of the population in some European cities (Benedictow 2004). The pandemic that is traditionally thought of as the first plague pandemic was the so called Justinianic Plague. It was first recorded between the years 541 and 543 in Egypt and spread rapidly throughout the Byzantine Empire and neighboring regions (Little 2007). After this first wave, the epidemic returned in about eighteen waves, over a period of two hundred years until it finally disappeared from the region (Stathakopoulos 2004).

For many years it was generally accepted that similarly to the 19th century pandemic, the former historical pandemics were also caused by *Y. pestis*. This interpretation has been questioned during the mid-1980's when discrepancies were reported between epidemiologic features described in literary and non-literary sources compared with those described in medical records of modern outbreaks (Sallares 2007, 231). This ignited a long lasting debate over the nature of the responsible disease (Twigg 1984; Scott and Duncan 2001; Duncan and Scott 2005, 315-320; Cohn 2008, 74-100). During the early 2000's aDNA studies detected *Y. pestis* in skeletons of victims from the two historic pandemics (Wiechmann and Grupe 2005, 48-55; Haensch et al. 2010, e1001134; Harbeck et al. 2013, e1003349) and in 2011 the first *Y. pestis* genome was recovered from Black Death victims buried at the East Smithfield plague cemetery in London, providing strong evidence for the involvement of *Y. pestis* in the second pandemic (Bos et al. 2011, 506-510). Further support was gained during 2014, when another *Y. pestis* draft genome was recovered, this time, from a Justinianic Plague victim buried in southern Germany (Wagner et al. 2014, 319-326).

1.3.2. An historical overview on the Justinianic Plague

The first documented outburst of the Justinianic Plague was during the summer of 541 AD in the Egyptian small port city of Pelusium, located on the extreme eastern branch of the Nile (Dewing 1914). The actual geographic origin of this epidemic remains unclear with Indian or east African origins suggested based on written evidence (Sallares 2007, 231). At that time, Egypt

was a part of the Byzantine Empire under emperor Justinian that controlled most of the Mediterranean region (Treadgold 1997). From Pelusium, the plague rapidly spread towards the east along the coast to Gaza and towards the west to Alexandria arriving to the capital Constantinople by spring 542. It was reported to have spread to all of the lands bordering the Mediterranean and penetrated even further, as east as Persia and arguably as north as the British Isles (Little 2007).

Many government officials in the Byzantine Empire were struck by the plague including Justinian himself which later recovered from the disease (Treadgold 1997). The death of many soldiers and taxpayers all over the empire created major fiscal crises that were thought to have inhibited Justinian's plan to reestablish the Roman Empire and caused a return to a less powerful Byzantine state (Treadgold 1997). The first wave of the plague ended in 545 and plague outbursts continued to reoccur until the eighth century. The last wave started in the Islamic world, spreading from Egypt and northern Africa to Syria, Mesopotamia and Iraq. In 747 it reached Constantinople and raged in the city for almost a year. Many fled the city and even the emperor Constantine V himself moved to Nicomedeia (Stathakopoulos 2007, 99-118).

The historian Procopius of Caesarea described the plague as a "pestilence by which the whole human race came near to being annihilated" estimating more than ten thousand dead each day (Dewing 1914). The Greek Agathias reported a total of 400,000 dead in the city (Cameron 1970). These kinds of estimates should be viewed carefully as it is unlikely that means to count the dead during such crises existed at the time (Stathakopoulos 2007, 99-118). Bodies were being buried in mass graves due to lack of space and manpower and in certain cases even thrown over city walls or into the sea (Sarris 2007, 119-134).

The depopulation and extreme manpower shortage caused by the plague waves are thought to have contributed to the eventual demise of the Eastern Roman Empire. Estimated lower mortality in semi-nomadic and nomadic areas compared with higher mortality in urban areas of the empire and Persia, could have posed an advantage for Arabic and Berber tribes, even before their unification in Islam, providing favorable conditions for the rise of Islam, potentially contributing to the subsequent conquest of Egypt and Syria in the Arab invasions during the seventh century (Russell 1968, 174-184).

1.3.3. The natural history of *Y. pestis*

The genus *Yersinia* includes three pathogenic bacteria: *Yersinia enterocolitica*, *Yersinia pseudotuberculosis* and *Y. pestis* (Yang, Cui, and Zhou 2012, 211). In contrast to *Y. pestis*, *Yersinia pseudotuberculosis* and *Yersinia enterocolitica* are water-borne pathogens that cause gastroenteritis which is seldom lethal. With almost 100% homology in protein coding genes (Chain et al. 2004, 13826-13831), *Y. pestis* is considered a monomorphic clone that evolved from the more diverse *Yersinia pseudotuberculosis* (Achtman et al. 1999, 14043-14048; Zhou and Yang 2009, 2242-2250; Hershberg and Petrov 2010, e1001115). It possesses about 150 chromosomal pseudogenes compared with 62 in *Yersinia pseudotuberculosis*, mirroring the massive shift from an enteric to an intracellular lifestyle (Zinser et al. 2003, 1271-1277; Zhou and Yang 2009, 2242-2250).

The genome of *Y. pestis* consists of a 4.65 megabase (Mb) chromosome and three plasmids: the pCD1 (70 kb) virulence plasmid found in all pathogenic *Yersinia* and the pMT1 and pPCP1 plasmids that are unique to *Y. pestis* (100 and 9.5 kb respectively). Phylogenetic analysis on modern *Y. pestis* isolates shows a clear geographic clustering (Achtman et al. 2004, 17837-17842; Cui et al. 2013, 577-582). Isolates from China or the former Soviet Union (FSU) are distributed over the basal branch “0” as well as over branches “1” to “4”, while isolates from outside of China/FSU are only found on the more derived “1” to “4” branches, suggesting a east/central Asian origin for all present-day *Y. pestis* isolates followed by a global multiple radiation spread that created country-specific lineages (Morelli et al. 2010, 1140-1143).

The *Y. pestis* genome recovered from Black Death victims in London has provided the opportunity to compare the historical and modern *Y. pestis* diversities (Bos et al. 2011, 506-510) and revealed that the Black Death coincided with a phylogenetic radiation event, suggesting a fast expansion of *Y. pestis* diversity at that time (Cui et al. 2013, 577-582). Notably, no evidence was detected in the Black Death genome for increased virulence, therefore the question of what caused the high spread and increased mortality of this pandemic remains open (Bos et al. 2011, 506-510). In following years, additional second pandemic genomes from across Europe were reconstructed, shedding light on the dynamics leading to this pandemic, e.g. demonstrating that plague foci persisted in Europe from the initial Black Death outbreak and up to the 18th century, serving as reservoirs for the post- Black Death outbreaks (Spyrou et al. 2016, 874-881; Bos et al. 2016, e12994). In addition, a Black Death genome isolated from the Volga region in Russia has shown phylogenetic relatedness to the strains causing the 19th century pandemic, suggesting that the Black Death and modern plague are genetically associated (Spyrou et al. 2016, 874-881).

The earliest *Y. pestis* genomes to date were detected in Late Neolithic/Bronze Age Eurasians (Rasmussen et al. 2015, 571-582; Valtueña et al. 2017, 3683-3691. e8; Rascovan et al. 2019, 295-305. e10). The fact that these genomes lacked genetic elements coding for virulent factors that promote transmission via a flea vector have led researchers to suggest that during these early stages of *Y. pestis* evolution this mode of transmission was not possible (Rasmussen et al. 2015, 571-582). However, the sequencing of a contemporaneous ~3,800 year old *Y. pestis* genome containing these coding regions helped trace the origins of the bubonic plague to as early as the Bronze Age and demonstrated the parallel existence of *Y. pestis* bacteria potentially utilizing different forms of transmission (Spyrou et al. 2018, 2234).

The genetic data available for the first pandemic has so far been extremely scarce. Genomic data was reconstructed from only two individuals that died during the Justinianic Plague in southern Germany, producing a low covered partial *Y. pestis* genome (Wagner et al. 2014, 319-326). When placed into the context of the known phylogeny, this partial genome forms a branch with no deriving currently sampled representatives, suggesting this lineage to be extinct (Wagner et al. 2014, 319-326). The branch diverges from extant groups isolated from rodent populations in China and the FSU, supporting a south or central Asian origin for the Justinianic strain, similar to the suggested origin for the other historic pandemics and contrasting the historical African origin suggested on account of historical records (Wagner et al. 2014, 319-326). Recently, a *Y. pestis* genome from China dated to the second to third century AD was found to be basal to the Justinianic strains (de Barros Damgaard et al. 2018, 369). This finding has led to a hypothesis putting forward the Hun invasions as the source for the first pandemic (de Barros Damgaard et al. 2018, 369). This suggested origin is somewhat surprising as the Hun invasions occurred around three hundred years prior to the first documented outbreak of the Justinianic plague and is yet to be resolved, potentially with additional future data. The fact that the Justinianic branch diverges from present-day rodent strains suggests an independent emergence from rodents to humans, separate from the emergence of the strains ancestral to the Black Death (Wagner et al. 2014, 319-326).

In the third section of this thesis, I describe the reconstruction and analysis of a *Y. pestis* high coverage genome recovered from a sixth century plague victim.

2. Aim of the thesis

The aim of the thesis is to complement the historical and archaeological records by re-constructing and analyzing genetic material from archaeological human remains. For this purpose, three events that are considered to have greatly impacted human societies over the last 15,000 years are investigated, mostly in the context of human disease and mobility.

The earliest of the three was the Neolithic revolution that marked humanity's transition into a food producing subsistence strategy beginning around 11,000 years before present (yBP). The second was the collapse of the Bronze Age civilizations during the 13th century BCE in the eastern Mediterranean that marked the transition into the Iron Ages. The third is a catastrophic event, the Justinianic Plague, a pandemic that ravaged across western Eurasia during the 6th century AD and up to the 8th century AD claiming many lives and thought by some to have contributed to the eventual demise of the Eastern Roman empire.

By the Archaeogenetic investigation of contemporaneous human remains, this thesis attempts to address historical and biological questions regarding these past events.

Manuscript A:

- How did the Anatolian Epipaleolithic hunter-gatherers relate to the known human population genetic variation of west Eurasia?
- What are the demographic models best explaining the Neolithic transition in central Anatolia? Was farming adopted by local hunter-gatherers in a process facilitated by cultural diffusion or was farming brought to central Anatolia by a movement of people from earlier farming centers in a manner that can be explained by demic diffusion?
- Which ancient populations genetically impacted the central Anatolian gene pool and when did these interactions occur?

Manuscript B:

- How did the Bronze Age people of Ashkelon relate to the known genetic variation of the early- and middle- Bronze Age in the Levant?
- Did external gene flow events influence the local gene pool in Bronze Age Ashkelon and if so which ancient populations might have contributed to this interaction?

- Was the evident cultural change that marked the transition between the Bronze- and Iron-Age in Ashkelon mirrored by apparent shifts in the genetic variation? If so, which ancient populations might have been involved?
- How did the Late Iron Age population genetically relate to the earlier Ashkelon populations and to the known genetic variation in west Eurasia?

Manuscript C:

- Did the Justinianic Plague reach southern Germany?
- How does the Justinianic *Y. pestis* lineage relate to the present-day *Y. pestis* variation?
- Are there apparent substitutions or structural differences between the Justinianic *Y. pestis* strain and present-day ones?
- Are there apparent substitutions or structural differences between the available Justinianic *Y. pestis* strains?
- How fast did the Justinianic Plague spread in southern Germany?
- What can we learn from the phylogenetic placement of the reconstructed Justinianic genome regarding the origin and spread of this strain?

3. Overview of Manuscripts and author's Contribution

3.1 Manuscript A

“Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia”

Michal Feldman, Eva Fernández-Domínguez, Luke Reynolds, Douglas Baird, Jessica Pearson, Israel HersHKovitz, Hila May, Nigel Goring-Morris, Marion Benz, Julia Gresky, Raffaella A. Bianco, Andrew Fairbairn, Gökhan Mustafaoğlu, Philipp W. Stockhammer, Cosimo Posth, Wolfgang Haak, Choongwon Jeong & Johannes Krause

Published at Nature Communications (March 2019)

In manuscript A, we report for the first time, genome wide data from a Late Pleistocene hunter-gatherer from Anatolia (15,000 yBP). This prehistoric population has so far not been genetically described and is currently the oldest human Near-Eastern population from which DNA has been obtained. We report seven additional genomes from the earliest farmers of central Anatolia and the southern Levant (~10,000 yBP), by that we significantly increase the amount and geographical coverage of genome wide data from this time period in this region.

People began farming in the Fertile Crescent region of the Near East during the Early Holocene and within several Millennia farming spread throughout Europe and Asia replacing local hunting and gathering practices. Central Anatolia harbors some of the earliest evidence of farming outside the Fertile Crescent and thus is a key region in understanding the early spread of farming. As human remains from this time and region are scarce, the reported data spanning the early transition to farming provides us with an opportunity to address some of the long debated questions on the origin of the first farmers of central Anatolia.

We find a high level of genetic continuity between the hunter-gatherer and early farmers of Anatolia lasting for over seven millennia. This suggests that the local hunter-gatherers of Anatolia adopted agriculture and rejects the model in which colonizing groups of farmers from earlier farming centers introduced agriculture to central Anatolia.

In addition, we detect two distinct incoming ancestries: an early Iranian/Caucasus related one that can be framed in time to the late Pleistocene/early Holocene. The later one, which we

describe as a reciprocal gene flow, occurred during the Neolithic and is linked to the ancient Levant.

Finally, we observe a genetic link between southern Europe and the Near East predating 15,000 years ago and describing an interaction sphere that is currently poorly documented in the Archaeological record. Our results suggest a limited role of human migration in the emergence of agriculture in central Anatolia and a long term persistence of the local gene pool. Nonetheless, we reveal a complex pattern of genetic interactions with neighboring regions throughout the studied period. This study provides the first glimpse into the way that the Anatolian early farmer gene pool was formed. This gene pool later replaced most of the Pleistocene European gene pool and is currently the single largest ancestral component in modern-day Europeans.

Author's contributions:

J. Krause, E. Fernández-Domínguez, I. HersHKovitz, C. Jeong, W. Haak and M. Feldman conceived the study. D. Baird, A. Fairbairn, G. Mustafaoğlu, J. Pearson, I. HersHKovitz, H. May, N. Goring-Morris., M. Benz and J. Gresky provided archeological material. D. Baird, , J. Pearson, A. Fairbairn, G. Mustafaoğlu, I. HersHKovitz, H. May, N. Goring-Morris, M. Benz, J. Gresky and P.W. Stockhammer advised on the archeological background and interpretation. D. Baird, J. Pearson, E. Fernández-Domínguez, N. Goring-Morris, M. Benz and J. Gresky wrote the archeological and sample background sections. M. Feldman, L. Reynolds and R.A. Bianco performed the laboratory work. M. Feldman performed the nuclear data analysis under the guidance of C. Jeong and advice from C. Posth. E. Fernández-Domínguez, Reynolds and M. Feldman performed the mitochondrial data analysis. E. Fernández-Domínguez wrote the mitochondrial analysis supplementary section. M. Feldman wrote the paper under the supervision of C. Jeong and J. Krause and with input from all coauthors. In total, M. Feldman contributed 60% to the manuscript, laboratory work and analyses.

3.2 Manuscript B

“Ancient DNA sheds light on the genetic origins of early Iron Age Philistines”

Michal Feldman, Daniel M. Master, Raffaella A. Bianco, Marta Burri, Philipp W. Stockhammer, Alissa Mittnik, Adam J. Aja, Choongwon Jeong and Johannes Krause

In print, Science Advances (expected publication date: July 2019)

In manuscript B, we report genome-wide data of ten Bronze- and Iron- Age individuals from the ancient Mediterranean port city of Ashkelon. The Bronze to Iron Age transition (~12th century BC) in the Eastern Mediterranean was a time of great turbulence marked by the destruction of entire cities, the collapse of empires (e.g. the Mycenaean kingdom, the Hittite empire and the Kassite dynasty of Babylonia) and by economic turmoil. It has been suggested that mobile groups of seafarers, often called “Sea Peoples” were responsible for those dramatic events, contributing to the breakdown of the Eastern Mediterranean civilization. It has also been suggested that one of those groups was the biblical Philistines that appeared in coastal cities of the Eastern Mediterranean shores (e.g. Ashkelon) during that time as evident by the material culture in these cities. It has been long debated if the Philistines were indeed foreign to the region and where their potential place of origin might be. Due to the poor preservation of DNA in this region, no genetic data from individuals of this time period has so far been available in order to address these questions.

We find a high degree of genetic continuity between our newly reported late Bronze Age and earlier Bronze Age individuals. This continuity is apparent for over a millennium and across culturally and geographically distinct groups ranging from the inland southern Levant where present-day Jordan is located and along the coastal regions of present-day Israel and Lebanon.

In contrast, we find that the early Iron Age population received a European-related genetic admixture that was introduced in Ashkelon during the Bronze to Iron Age transition, supporting a migration event around the time of the Philistine appearance. We model southern European gene pools as plausible sources for this migration event.

Surprisingly, within no more than two centuries, the genetic signature of the above admixture is no longer detectable in the late Iron Age individuals despite the fact that they continue to be identified as Philistines in ancient texts, suggesting that the genetic signal was

diluted by the local Levantine related gene pool. We therefore conclude that the migration event did not leave a permanent detectable genetic footprint in Ashkelon.

Author's contributions:

J. Krause and D.M. Master conceived the study. J. Krause supervised the genetic work. D.M. Master and A.J. Aja provided archaeological material. D.M. Master, P.W. Stockhammer and A.J. Aja advised on the archaeological background and interpretation. D.M. Master and A.J. Aja wrote the archaeological and sample background section. M. Feldman, M. Burri, A. Mitnik and R.A. Bianco performed the laboratory work. M. Feldman performed the data analyses, with C. Jeong providing guidance. M. Feldman wrote the manuscript with critical comments from D.M. Master, C. Jeong and J. Krause and with input from all coauthors.

In total, M. Feldman contributed 70% to the manuscript, laboratory work and analyses.

3.3 Manuscript C

“A High-Coverage *Yersinia pestis* Genome from a Sixth-Century Justinianic Plague Victim”

Michal Feldman, Michaela Harbeck, Marcel Keller, Maria A. Spyrou, Andreas Rott, Bernd Trautmann, Holger C. Scholz, Bernd Pääffgen, Joris Peters, Michael McCormick, Kirsten Bos, Alexander Herbig and Johannes Krause

Published at Molecular Biology and Evolution (August 2016)

In manuscript C, we report the first high-coverage *Y. pestis* genome (~18 fold) reconstructed from a 6th century victim of the Justinianic Plague, buried in southern Germany. This historic pandemic was the first of three that were caused by *Y. pestis*. It was responsible for massive casualties and has been suggested to contribute to the eventual demise of the Eastern Roman Empire.

The coverage and quality of the data enabled the detection of novel substitutions and structural differences in this early *Y. pestis* strain, some of which were detected in genomic regions that have previously been suggested as plague virulence determinants. We confirm a previously suggested phylogenetic placement of the Justinianic strain with minor attenuations. In addition, we identify substitutions falsely typed in a previously published lower coverage *Y. pestis* genome from the same time period and region. The two strains are otherwise genetically identical, suggesting low genetic diversity of plague bacteria in Southern Germany during the 6th century plague pandemic.

These findings highlight the importance of following strict criteria of authenticity and genome quality. We establish criteria that include average coverage, evenness of coverage and heterozygosity inspection, as well as standard criteria often used by the ancient DNA community such as examination of DNA damage and DNA fragmentation patterns in order to ensure the authenticity and quality of ancient bacterial genomes.

This study demonstrated the presence of *Y. pestis* in a second early medieval rural site in southern Germany where no historical source records it, expanding the number of sites known to

have been afflicted by plague and reinforcing the time placement of this strain in the early waves of the 200-year-long pandemic. Future sampling from various periods and geographical locations is needed to deepen our understanding of the disease, its evolution and its human impact. In addition, the unique genetic features identified here call for functional studies that would explore their role in terms of *Y. pestis* physiology and adaptation.

Author's contributions:

J. Krause and M. Harbeck conceived the study. J. Krause supervised the genetic work. M. Harbeck, A. Rott, B. Trautmann, H.C. Scholz, B. Pääffgen and J. Peters provided archaeological material. M. Harbeck, M. Keller and A. Rott performed the initial PCR screening. M. McCormick advised on the historical background and interpretation. M. Feldman, M. Spyrou and K. Bos performed the laboratory work. M. Feldman performed the data analyses with A. Herbig providing guidance. M. Feldman wrote the manuscript with input from all coauthors.

In total, M. Feldman contributed 70% to the manuscript, laboratory work and analyses.




4. Manuscript A

ARTICLE

<https://doi.org/10.1038/s41467-019-09209-7>

OPEN

Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia

Michal Feldman¹, Eva Fernández-Domínguez², Luke Reynolds³, Douglas Baird⁴, Jessica Pearson⁴, Israel Hershkovitz^{5,6}, Hila May^{5,6}, Nigel Goring-Morris⁷, Marion Benz⁸, Julia Gresky⁹, Raffaella A. Bianco¹, Andrew Fairbairn¹⁰, Gökhan Mustafaoğlu¹¹, Philipp W. Stockhammer^{1,12}, Cosimo Posth ¹, Wolfgang Haak¹, Choongwon Jeong ¹ & Johannes Krause ¹

Anatolia was home to some of the earliest farming communities. It has been long debated whether a migration of farming groups introduced agriculture to central Anatolia. Here, we report the first genome-wide data from a 15,000-year-old Anatolian hunter-gatherer and from seven Anatolian and Levantine early farmers. We find high genetic continuity (~80–90%) between the hunter-gatherers and early farmers of Anatolia and detect two distinct incoming ancestries: an early Iranian/Caucasus related one and a later one linked to the ancient Levant. Finally, we observe a genetic link between southern Europe and the Near East predating 15,000 years ago. Our results suggest a limited role of human migration in the emergence of agriculture in central Anatolia.

¹Max Planck Institute for the Science of Human History (MPI-SHH), Kahlaische Strasse 10, 07745 Jena, Germany. ²Department of Archaeology, Durham University, Durham South Road, DH1 3LE, UK. ³School of Natural Sciences and Psychology, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK. ⁴Department of Archaeology, Classics and Egyptology, University of Liverpool, 8–14 Abercromby Square, Liverpool L69 7WZ, UK. ⁵Department of Anatomy and Anthropology, The Dan David Center for Human Evolution and Biohistory Research and The Shmunis Family Anthropology Institute, Sackler Faculty of Medicine, Tel Aviv University, Post Office Box 39040, Tel Aviv 6997801, Israel. ⁶The Steinhardt Museum of Natural History, Tel Aviv University, Post Office Box 39040, Tel Aviv 6997801, Israel. ⁷Department of Prehistory, Institute of Archaeology, The Hebrew University of Jerusalem, Jerusalem 919051, Israel. ⁸Department of Near Eastern Archaeology, Free University Berlin, Fabeckstrasse 23–25, 14195 Berlin, Germany. ⁹Department of Natural Sciences, German Archaeological Institute, Im Dol 2–6, 14195 Berlin, Germany. ¹⁰School of Social Science, The University of Queensland, Michie Building, St Lucia, Brisbane, QLD, Australia. ¹¹Department of Archaeology, Zonguldak Bülent Ecevit University, Incivez, 67100 Zonguldak, Turkey. ¹²Institut für Vor- und Frühgeschichtliche Archäologie und Provinzialrömische Archäologie Ludwig-Maximilians-Universität München München, Schellingstrasse 12, 80799 München, Germany. Correspondence and requests for materials should be addressed to E.F.-D. (email: eva.fernandez@durham.ac.uk) or to C.J. (email: jeong@shh.mpg.de) or to J.K. (email: krause@shh.mpg.de)

The practice of agriculture began in the Fertile Crescent of Southwest Asia as early as 10,000 to 9000 BCE. Subsequently, it spread across western Eurasia while increasingly replacing local hunting and gathering subsistence practices, reaching central Anatolia by c. 8300 BCE^{1–3}.

Recent genetic studies have shown that in mainland Europe, farming was introduced by an expansion of early farmers from Anatolia that replaced much of the local populations^{4,5}. Such mode of spread is often referred to as the demic diffusion model. In contrast, in regions of the Fertile Crescent such as the southern Levant and the Zagros Mountains (located between present-day eastern Iraq and western Iran), the population structure persists throughout the Neolithic transition⁶, indicating that the hunter-gatherers of these regions locally transitioned to a food-producing subsistence strategy.

Central Anatolia has some of the earliest evidence of agricultural societies outside the Fertile Crescent³ and thus is a key region in understanding the early spread of farming. While archaeological evidence points to cultural continuity in central Anatolia³, due to the lack of genetic data from pre-farming individuals, it remains an open question whether and to what scale the development of the Anatolian Neolithic involved immigrants from earlier farming centers admixing with the local hunter-gatherers.

Likewise, pre-farming genetic links between Near-Eastern and European hunter-gatherers are not well understood, partly due to the lack of hunter-gatherer genomes from Anatolia. Genetic studies have suggested that ancient Near-Eastern populations derived a substantial proportion of their ancestry from a common outgroup of European hunter-gatherers and East Asians^{4,6,7}. This deeply branching ancestry often referred to as Basal Eurasian likely diverged from other Eurasians before the latter received Neanderthal gene flow⁶. Interestingly, a previous study reported that European hunter-gatherers younger than 14,000 years ago tend to show an increased affinity with present-day Near Easterners compared to older European hunter-gatherers⁸, although how this affinity formed is not well understood.

Here, we report new genome-wide data from eight prehistoric humans (Fig. 1a, Table 1, and Supplementary Table 1), including the first Epipaleolithic Anatolian hunter-gatherer sequenced to date (labeled AHG; directly dated to 13,642–13,073 cal BCE, excavated from the site of Pınarbaşı, Turkey), five early Neolithic Aceramic Anatolian farmers (labeled AAF; c. 8300–7800 BCE, one directly dated to 8269–8210 cal BCE³, from the site of Boncuklu, Turkey), adding to previously published genomes from this site⁹, and two Early Neolithic (PPNB) farmers from the southern Levant (one labeled KFH2, directly dated to c. 7700–7600 cal BCE, from the site of Kfar HaHoresh, Israel; and the second labeled BAJ001, c. 7027–6685 cal BCE, from the site of Ba'ja, Jordan). These data comprise a genetic record stretching from the Epipaleolithic into the Early Holocene, spanning the advent of agriculture in the region.

We find that the AHG is genetically distinct from other reported late Pleistocene populations. We reveal that Neolithic Anatolian populations derive a large fraction of their ancestry from the Epipaleolithic Anatolian population, suggesting that farming was adopted locally by the hunter-gatherers of central Anatolia. We also detect distinct genetic interactions between the populations of central Anatolia and earlier farming centers to the east, during the late Pleistocene/early Holocene and describe a genetic link with European hunter-gatherers that predates 15,000 years ago.

Results

Genetic continuity and detected admixtures in Anatolia. We extracted DNA from the ancient human remains and prepared it for next-generation sequencing^{10,11}, which resulted in human

DNA yields lower than 2% (Supplementary Data 1), comparable with low DNA preservation previously reported in the region^{6,9}. To generate genome-wide data despite the low DNA yields, we performed in-solution DNA enrichment targeting 1.24 million genome-wide single-nucleotide polymorphisms (SNPs) (“1240k capture”)¹², which resulted in 129,406 to 917,473 covered SNPs per individual. We estimated low mitochondrial contamination levels for all eight individuals (1–6%; see Methods and Supplementary Table 2) and could further test the males for nuclear contamination, resulting in low estimates (0.05–2.23%; Supplementary Table 2). For population genetic analyses, we merged genotype data of the new individuals with previously published datasets from 587 ancient individuals and 254 present-day populations (Supplementary data 2).

To estimate how the ancient individuals relate to the known west Eurasian genetic variation, we projected them onto the top two dimensions (PC1, PC2) of present-day principal component analysis (PCA)⁶ (Fig. 1b). Strikingly, the AHG individual is positioned near both AAF and later Anatolian Ceramic farmers¹² (7000–6000 cal BCE). These three prehistoric Anatolian populations (AHG, AAF, and ACF), representing a temporal transect spanning the transition into farming, are positioned along PC1 between Mesolithic western European hunter-gatherers (WHG)^{4,7,12} who are at one extreme of PC1 and Levantine Epipaleolithic Natufians⁶ who are at the other. Along PC2, ancient Anatolians, WHG, and Natufians have similar coordinates. The newly reported Levantine Neolithic farmers (BAJ001 and KFH2) are positioned near the previously published Levantine Neolithic farmers⁶ (Supplementary Note 2). In ADMIXTURE analysis AHG, AAF, and ACF are inferred as a mixture of two components that are each maximized in Natufians and WHG, consistent with their intermediate positions along PC1 in PCA (Supplementary Figure 1).

Inspired by our qualitative observations in PCA and ADMIXTURE analyses, we applied formal statistical frameworks to describe the genetic profiles of the three Anatolian populations and to test and model genetic differences between them. We first characterized the ancestry of AHG. As expected from AHG's intermediate position on PCA between Epipaleolithic/Neolithic Levantines and WHG, Patterson's *D*-statistics¹³ of the form *D* (AHG, WHG; Natufian/Levant_N, Mbuti) ≥ 4.8 SE (standard error) and *D* (AHG, Natufian/Levant_N; WHG, Mbuti) ≥ 9.0 SE (Supplementary Table 3) indicate that AHG is distinct from both the WHG and Epipaleolithic/Neolithic Levantine populations and yet shares extra affinity with each when compared to the other. Then, we applied a *qpAdm*-based admixture modeling to integrate these *D*-statistics. *qpAdm* is a generalization of *Df₄*-statistics that test whether the target population and the admixture model (i.e., a linear combination of reference populations) are symmetrically related to multiple outgroups¹³. By doing so, it tests whether the proposed admixture model is adequate to explain the target gene pool and provides admixture coefficient estimates. We find an adequate two-way admixture model ($\chi^2 p = 0.158$), in which AHG derives around half of his ancestry from a Neolithic Levantine-related gene pool ($48.0 \pm 4.5\%$; estimate ± 1 SE) and the rest from the WHG-related one (Supplementary Tables 4, 5). While these results do not suggest that the AHG gene pool originated as a mixture of Levant_N and WHG, both of which lived millennia later than AHG, it still robustly supports that AHG is genetically intermediate between WHG and Levant_N. This cannot be explained without gene flow between the ancestral gene pools of those three groups. This supports a late Pleistocene presence of both Near-Eastern and European hunter-gatherer-related ancestries in central Anatolia. Notably, this genetic link with the Levant pre-dates the advent of farming in this region by at least five millennia.

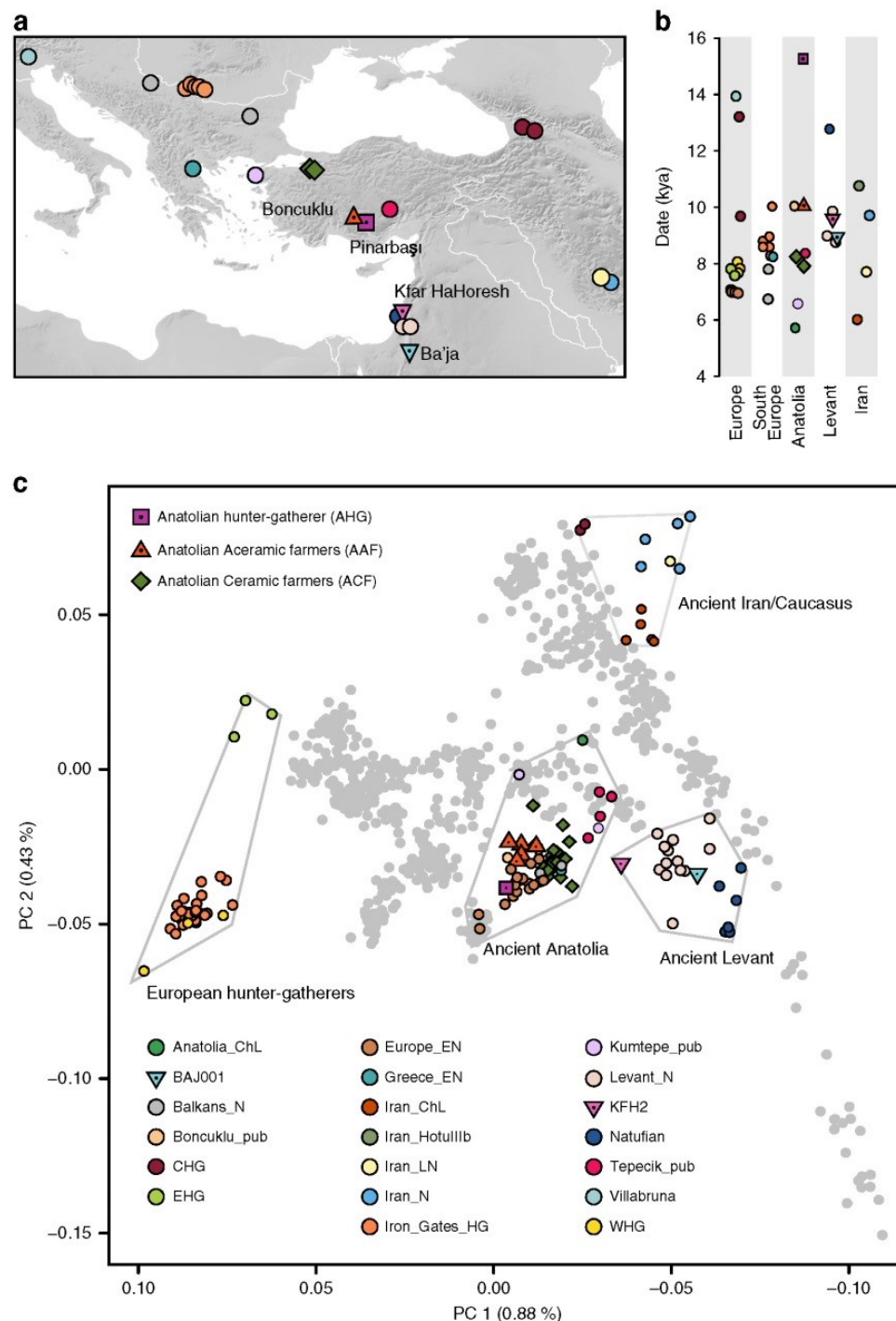


Fig. 1 Location, age, and principal component analysis (PCA) of analyzed individuals. **a** Locations of newly reported and selected published genomes. Archeological sites from which new data are reported are annotated. Symbols for the analyzed groups are annotated in **c**. **b** Average ages of ancient groups. **c** Ancient genomes (marked with color-filled symbols) projected onto the principal components computed from present-day west Eurasians (gray circles) (Supplementary Figure 8). The geographic location of each ancient group is marked in **a**. Ancient individuals newly reported in this study are additionally marked with a black dot inside the symbol. Source data are provided as a Source Data file

In turn, AAF are slightly shifted on PC2 compared to AHG, to the direction where ancient and modern Caucasus and Iranian groups are located. Likewise, when compared to AHG by $D(\text{AAF}, \text{AHG}; \text{test}, \text{Mbuti})$, the AAF early farmers show a marginal excess affinity with early Holocene populations from Iran or Caucasus and with present-day south Asians, who have also been genetically linked with Iranian/Caucasus ancestry^{14,15} (e.g., $D = 2.3$ and 2.7SE for CHG and Vishwabrahmin, respectively; Fig. 2a,

Supplementary Figures 2, 3, and Supplementary Data 3). Accordingly, a mixture of AHG and Neolithic Iranians provides a good fit to AAF in our $qpAdm$ modeling ($\chi^2 p = 0.296$), in which AAF derive most of their ancestry ($89.7 \pm 3.9\%$) from a population related to AHG (Supplementary Tables 4 and 6). A simpler model without contribution from Neolithic Iranians (i.e., modeling AAF as a sister clade of AHG) shows a significant reduction in model fit ($\chi^2 p = 0.014$). This suggests a long-term

Table 1 An overview of ancient genomes reported in this study

ID	Library name	Analysis group	Estimated date	Site	Sampled tissue	Total sequenced reads ($\times 10^6$)	Human DNA (%)	Mean coverage (fold)	Genetic sex	mt	Ychr
ZBC	IPB001.	AHG	13,642–13,073 cal BCE	Pınarbaşı	Intermediate phalanx	126.7	33	2.9	Male	K2b	C1a2
ZHAG	B/C0101	AAF	8300–7800 BCE	Boncuklu	Petrous	92.0	38	1.48	Female	N1a1a1	
ZMOJ	BON014.	AAF	8300–7800 BCE	Boncuklu	third molar	77.9	27	0.8	Male	K1a	C
ZKO	BON001.	AAF	8300–7800 BCE	Boncuklu	Petrous	84.8	31	0.9	Male	U3	G2a2b2b
ZHJ	BON024.	AAF	8300–7800 BCE	Boncuklu	Third molar	87.7	38	0.76	Female	U3	
ZHAJ	BON034.	AAF	8269–8210 cal BCE	Boncuklu	Petrous	75.4	30	0.69	Female	U3	
KFH2	KFH002.	Levant_Neol	7712–7589 cal BCE	Kfar HaHoresh	Petrous	342.0	8	0.16	Female	N1a1b	
BAJ001	BAJ001.	Levant_Neol	7027–6685 cal BCE	Ba'ja	Petrous	17.3	45	0.75	Female	N1b1a	

For each individual the analysis group is given (AHG = Anatolian hunter-gatherer; AAF = Anatolian Aceramic farmers; Levant_Neol = Levantine early farmer). When ^{14}C dating results are available, the date is given in cal BCE in 2-sigma range, otherwise a date based on the archeological context is provided (detailed dating information is provided in Supplementary Note 1 and Supplementary Table 1). The proportion of human DNA and the mean coverage on 1240k target sites in the “1240k” enriched libraries are given. Uniparental haplogroups (mt = mitochondrial; Ychr = Y chromosome) are listed. Detailed information on the uniparental analysis can be found in Supplementary Note 1 and Supplementary Data 6

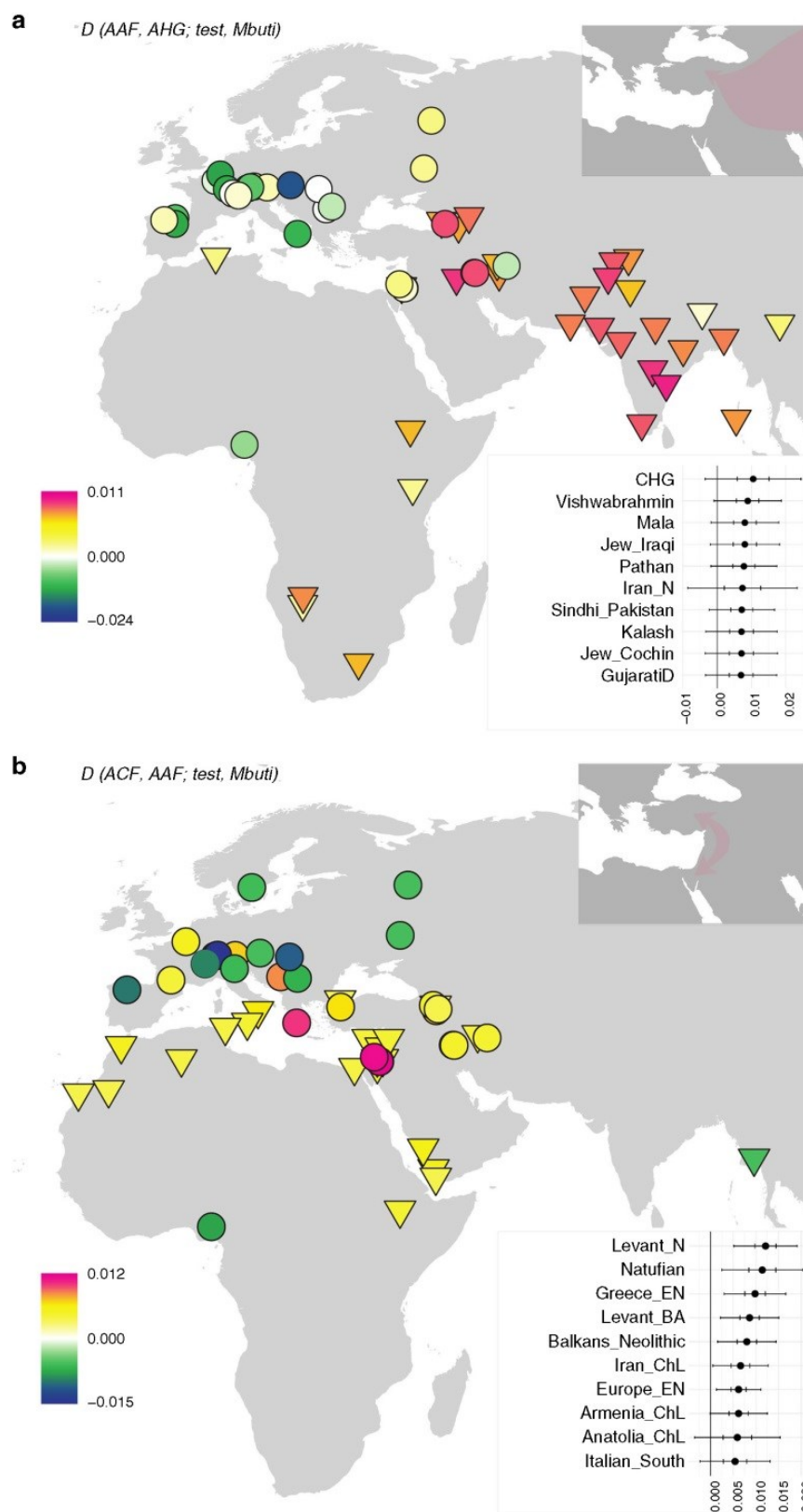
genetic stability in central Anatolia over five millennia despite changes in climate and subsistence strategy. The additional Neolithic Iranian-related ancestry ($10.3 \pm 3.9\%$) presumably diffused into central Anatolia during the final stages of the Pleistocene or early Holocene, most likely via contact through eastern Anatolia. This provides evidence of interactions between eastern and central Anatolia in the Younger Dryas or the first millennium of the Holocene, currently poorly documented archaeologically.

In contrast, we find that the later ACF individuals share more alleles with the early Holocene Levantines than AAF do, as shown by positive $D(\text{ACF}, \text{AAF}; \text{Natufian}/\text{Levant}_N, \text{Mbuti}) \geq 3.8$ SE (Fig. 2b, Supplementary Figures 4, 5, and Supplementary Data 3). Ancient Iran/Caucasus populations and contemporary South Asians do not share more alleles with ACF ($|D| < 1.3$ SE). Likewise, *qpAdm* modeling suggests that the AAF gene pool still constitutes more than 3/4 of the ancestry of ACF 2000 years later ($78.7 \pm 3.5\%$; Supplementary Tables 4 and 7) with additional ancestry well modeled by the Neolithic Levantines ($\chi^2 p = 0.115$) but not by the Neolithic Iranians ($\chi^2 p = 0.076$; the model estimated infeasible negative mixture proportions) (Supplementary Tables 4 and 7). These results suggest gene flow from the Levant to Anatolia during the early Neolithic. In turn, Levantine early farmers (Levant_Neol) that are temporally intermediate between AAF and ACF could be modeled as a two-way mixture of Natufians and AHG or AAF ($18.2 \pm 6.4\%$ AHG or $21.3 \pm 6.3\%$ AAF ancestry; Supplementary Tables 4 and 8 and Supplementary Data 4), confirming previous reports of an Anatolian-like ancestry contributing to the Levantine Neolithic gene pool⁶. These two distinct detected gene flows support a reciprocal genetic exchange between the Levant and Anatolia during the early stages of the transition to farming.

Genetic links between Pleistocene Europe and the Near East. AHGs experienced climatic changes during the last glaciation¹⁶ and inhabited a region that connects Europe to the Near East. However, pre-Neolithic interactions between Anatolia and Southeastern Europe are so far not well documented archaeologically. Interestingly, a previous genomic study showed that

present-day Near Easterners share more alleles with European hunter-gatherers younger than 14,000BP (“Later European HG”) than with older ones (“Earlier European HG”)⁸. With ancient genomic data available, we could directly compare the genetic affinity of European hunter-gatherers with Near-Eastern hunter-gatherers (AHG and Natufian) using the *D*-statistic of the form $D(\text{European hunter-gatherers}, \text{Kostenki14}; \text{AHG}/\text{Natufian}, \text{Mbuti})$. We compared the European hunter-gatherers to the 37 thousand-year-old individual *Kostenki14*^{8,17} representing the oldest available European genome with genetic affinity to later European hunter-gatherers (Fig. 3a and Supplementary Data 5). As is the case for present-day Near Easterners, this statistic is significantly positive for all European hunter-gatherers younger than 14,000BP. Most of the Later European HGs belong to a largely homogeneous gene pool referred to as the “Villabruna cluster,”⁸ named after its oldest available member from an Epigravettian site in northern Italy. Our results suggest that the non-Basal Eurasian ancestry of ancient Anatolians and Levantines derived from a gene pool related to the Villabruna cluster prior to its expansion within Europe observed after 14,000BP.

Among the Later European HG, recently reported Mesolithic hunter-gatherers from the Balkan peninsula, which geographically connects Anatolia and central Europe (“Iron Gates HG”)¹⁸, show the highest genetic affinity to AHG and the second highest one to Natufians, as shown in the positive statistic $D(\text{Iron Gates HG}, \text{European hunter-gatherers}; \text{AHG}/\text{Natufian}, \text{Mbuti})$ (Supplementary Figures 6 and 7). This affinity is surprising considering that Iron Gates HG have been previously modeled as a mixture of WHG (~85%) and eastern European hunter-gatherers (EHG; ~15%)¹⁸, the latter of which shares a much lower affinity with ancient Near Easterners in respect to other European HG (Fig. 3a). Since the previously reported WHG and EHG model did not fit well ($\chi^2 p = 0.0003$) and since Iron Gates HG harbored Near-Eastern-like mitochondrial groups, an affinity with Anatolians beyond the WHG + EHG model has been hypothesized¹⁸. Accordingly, we find that Iron Gates HG can be modeled as a three-way mixture of Near-Eastern hunter-gatherers ($25.8 \pm 5.0\%$ AHG or $11.1 \pm 2.2\%$ Natufian), WHG ($62.9 \pm 7.4\%$ or $78.0 \pm 4.6\%$, respectively) and EHG ($11.3 \pm 3.3\%$ or $10.9 \pm 3\%$,



respectively); ($\chi^2 p = 0.308$ and $\chi^2 p = 0.589$ respectively; Supplementary Tables 4 and 9).

To further test the model of Near-Eastern gene flow into the ancestors of Iron Gates HG as an explanation of the extra affinity between them, we utilized the Basal Eurasian ancestry that was widespread in early Holocene and late Pleistocene Near-Eastern

populations and their descendants but undetectable in European hunter-gatherers⁸, as a tracer for gene flow from the Near East. To estimate the Basal Eurasian ancestry proportion (“ α ”), we followed a previously established *qpAdm*-based approach that uses an African reference (the ancient Ethiopian *Mota* genome¹⁹) as a proxy⁶ (Supplementary Table 10). We estimated α to be 24.8

Fig. 2 Differences in genetic affinities between the ancient Anatolian populations. We plot the highest and lowest 40 values of $D(\text{population 1, population 2; test, Mbuti})$ on the map. Circles mark ancient populations and triangles present-day ones. “Test” share more alleles with *population 1* when values are positive and with *population 2* when negative. The detected gene flow direction is illustrated in the upper schematics; the illustrated rout represents the shortest one between the proximate source and the target and should not be interpreted as the historic rout of the gene flow. The statistics and SEs are found in Supplementary Figures 2–5 and Supplementary Data 3. **a** Early Holocene Iranian and Caucasus populations, as well as present-day South Asians, share more alleles with Aceramic Anatolian farmers (AAF) than with Anatolian hunter-gatherers (AHG), measured by positive $D(\text{AAF, AHG; test, mbuti})$. The top 10 values with ± 1 and $\pm 3\text{SE}$ are shown in the upper box. **b** Ancient Levantine populations share more alleles with Anatolian Ceramic farmers (ACF) than with AAF, measured by positive $D(\text{ACF, AAF; test, Mbuti})$. The top 10 values with ± 1 and $\pm 3\text{SE}$ are shown in the lower box. Source data are provided as a Source Data file

$\pm 5.5\%$ in AHG and $38.5 \pm 5.0\%$ in Natufians (Fig. 3b, Supplementary Table 10), consistent with previous estimates for the latter⁶. In turn, the Iron Gates HG could be modeled without any Basal Eurasian ancestry or with a non-significant proportion of $1.6 \pm 2.8\%$ when forced to have it as a third source (Fig. 3b and Supplementary Table 10). In contrast to the above direct estimate, the three-way admixture model of WHG + EHG + AHG predicts $\alpha = 6.4 \pm 1.9\%$ for Iron Gates, calculated as $(\% \text{ AHG in Iron Gates HG}) \times (\alpha \text{ in AHG})$, suggesting that unidirectional gene flow from the Near East to Europe alone may not be sufficient to explain the excess affinity between the Iron Gates HG and the Near-Eastern hunter-gatherers. Thus, it is plausible to assume that prior to 15,000 years ago there was either a bidirectional gene flow between populations ancestral to Southeastern Europeans of the early Holocene and those ancestral to Anatolians of the Late Glacial or a genetic influx from the populations ancestral to Southeastern Europeans into the Near East.

Uniparental markers and phenotypic analysis. The uniparental marker analysis placed AHG within the mitochondrial sub-haplogroup K2b and within the Y-chromosome haplogroup C1a2, both rare in present-day Eurasians (Table 1 and Supplementary Data 6). Mitochondrial Haplogroup K2 has so far not been found in Paleolithic hunter-gatherers²⁰. However, Y-haplogroup C1a2 has been reported in some of the earliest European hunter-gatherers^{8,17,21}. The early farmers belong to common early Neolithic mitochondrial (N1a, U3 and K1a) and Y chromosome types (C and G2a), with the exception of the Levantine BAJ001, which represents the earliest reported individual carrying the mitochondrial N1b group (Table 1 and Supplementary Data 6).

We examined alleles related to phenotypic traits in the ancient genomes (Supplementary Data 7). Notably, three of the AAF carry the derived allele for rs12193832 in the *HERC2* (hect domain and RLD2) gene that is primarily responsible for lighter eye color in Europeans²². The derived allele is observed as early as 14,000–13,000 years ago in individuals from Italy and the Caucasus^{8,23}, but had not yet been reported in early farmers or hunter-gatherers from the Near East.

Discussion

By analyzing genome-wide data from pre-Neolithic and early Neolithic Anatolians and Levantines, we describe the demographic developments leading to the formation of the Anatolian early farmer population that later replaced most of the European hunter-gatherers and represents the largest ancestral component in present-day Europeans^{4,5}.

We report a long-term persistence of the local AHG gene pool over seven millennia and throughout the transition from foraging to farming. This demographic pattern is similar to those previously observed in earlier farming centers of the Fertile Crescent⁶ and differs from the pattern of the demic diffusion-based spread of farming into Europe^{4,5}. Our results provide a genetic support

for archeological evidence³, suggesting that Anatolia was not merely a stepping stone in a movement of early farmers from the Fertile Crescent into Europe, but rather a place where local hunter-gatherers adopted ideas, plants, and technology that led to agricultural subsistence.

Interestingly, while we observe a continued presence of the AHG-related gene pool throughout the studied period, a pattern of genetic interactions with neighboring regions is evident from as early as the Late Pleistocene and early Holocene. In addition to the local genetic contribution from earlier Anatolian populations, Anatolian Aceramic farmers inherit about 10% of their genes from a gene pool related to the Neolithic Iran/Caucasus while later ACF derive about 20% of their genes from another distinct gene pool related to the Neolithic Levant.

Wide temporal gaps between available genomes currently limit our ability to distinguish the mode of transfer. Obtaining additional genomic data from these regions as well as from geographically intermediate populations of eastern Anatolia and the greater Mesopotamia region could help determine how these genetic changes happened in central Anatolia: for example, whether by a short-term massive migration or a low-level background gene flow in an isolation by distance manner.

To the west, we observe a genetic link between the Anatolian and European Pleistocene hunter-gatherers, which extends the temporal frame of the previously reported genetic affinity between late Pleistocene Europeans and present-day Near-Eastern populations⁸. Especially, Mesolithic Southeastern European hunter-gatherers (Iron Gates HG) show a strong genetic affinity with AHG. Our analysis on their Basal Eurasian ancestry proportions, although limited in resolution, suggests that a Near-Eastern gene flow from AHG into the ancestors of Iron Gates HG may not be sufficient to explain this affinity. Two additional scenarios, both involving gene flow from the ancestors of Iron Gates HG to the ancestors of AHG, can help explain the extra affinity between Iron Gates HG and AHG. One assumes a secondary gene flow from Southeastern Europe to Anatolia after the initial formation of the Near-Eastern gene pool as a mixture of the Basal Eurasian and the Villabruna-related gene pools. The other assumes that Iron Gates HG are indeed the most closely related group among European hunter-gatherers to the Villabruna-related ancestry in ancient Near Easterners. Further sampling in Anatolia and Southeastern Europe is needed to specify the spatiotemporal extent of the genetic interactions that we observe.

Methods

adNA analysis. We extracted and prepared DNA for next-generation sequencing in two different dedicated ancient DNA (adNA) facilities (Liverpool and Jena).

In Liverpool, UK, sampling and extraction steps for the individuals from Pınarbaşı and Boncuklu were carried out in the adNA labs at the Liverpool John Moores University. The outer layer of the bone was removed using powdered aluminum oxide in a sandblasting instrument. Then, the bone was ultraviolet (UV) irradiated for 10 min on each side and ground into fine powder using a cryogenic grinder Freezer/Mill. DNA was extracted from 100 mg of bone powder following an established protocol¹⁰. The extraction included incubation of the bone powder

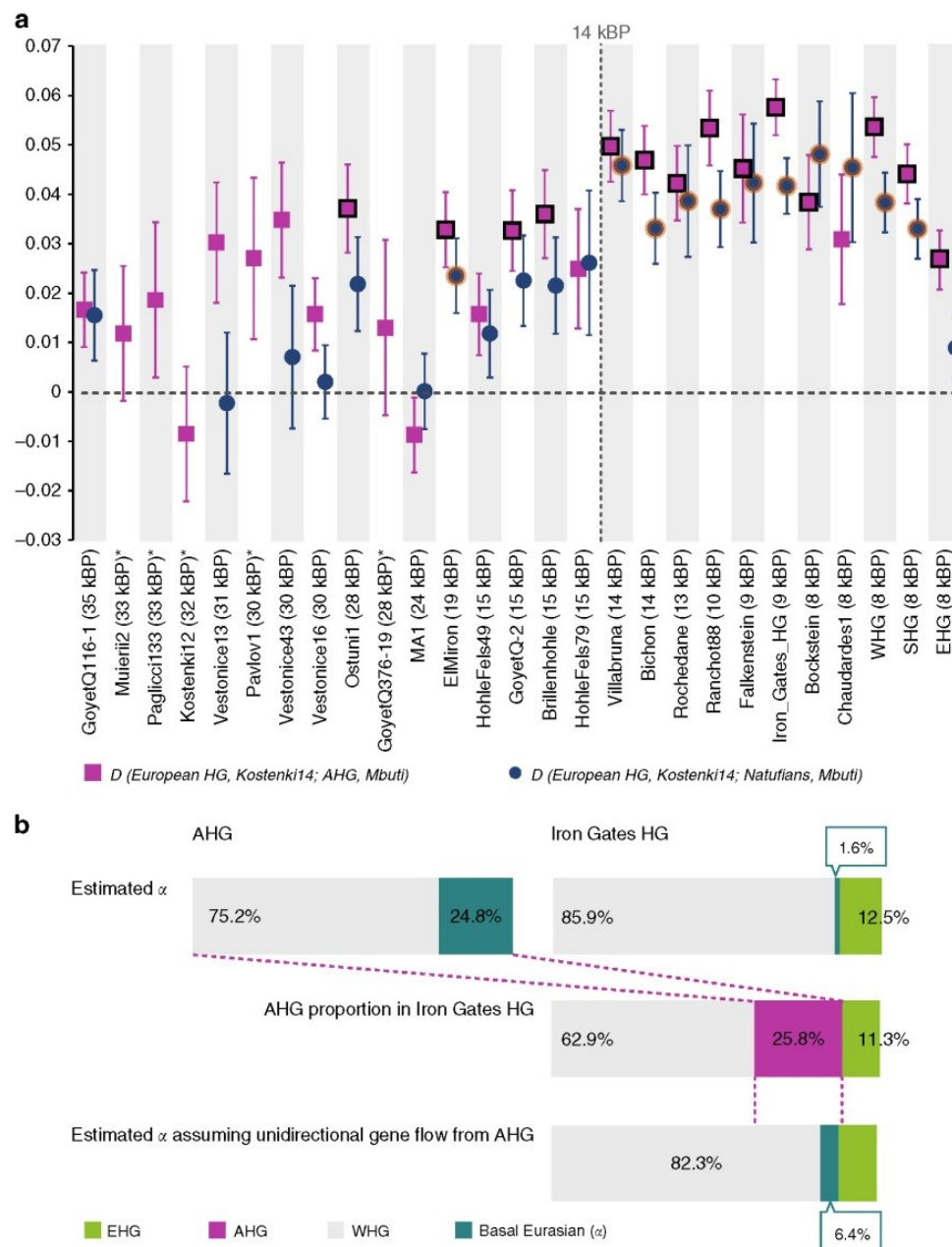


Fig. 3 Genetic links between Near-Eastern and European hunter-gatherers. **a** Genetic affinity between Near-Eastern and European hunter-gatherers increases after 14,000 years ago as measured by the statistic $D(\text{European HG}, \text{Kostenki14}; \text{Natufian/AHG}, \text{Mbuti})$. Vertical lines mark $\pm 1 \text{ SE}$. Data points for which $D > 3 \text{ SE}$ are outlined. *Kostenki14* serves here as a baseline for the earlier European hunter-gatherers. Statistics including all analyzed European hunter-gatherers are listed in Supplementary Data 5. Individuals marked with an asterisk did not reach the analysis threshold of over 30,000 single-nucleotide polymorphisms (SNPs) overlapping with *Natufian/AHG*. **b** Basal Eurasian ancestry proportions (α) as a marker for Near-Eastern gene flow. Mixture proportions inferred by qpAdm for the Anatolian hunter-gatherer (AHG) and the Iron Gates hunter-gatherers (Iron Gates HG) are schematically represented⁶. The lower schematic shows the expected α in Iron Gates HG under assumption of unidirectional gene flow, inferred from α in the AHG source population. The observed α for Iron Gates HG is considerably smaller than expected; thus, the unidirectional gene flow from the Near East to Europe is not sufficient to explain the affinity between Iron Gates HG and AHG. Source data are provided as a Source Data file

in 1 ml of extraction buffer (0.45 M EDTA, pH 8.0, and 0.25 mg ml⁻¹ proteinase K) at 37 °C for over a 12–16 h. Subsequently, DNA was bound to a silica membrane using a binding buffer containing guanidine hydrochloride and purified in combination with the High Pure Viral Nucleic Acid Large Volume Kit (Roche). DNA was eluted in 100 μ l of TET (10 mM Tris-HCl, 1 mM EDTA, pH 8.0, and 0.05% Tween-20). One extraction blank was taken along. The extracts were then shipped to Jena, Germany where downstream processing was performed.

In Jena, Germany, all pre-amplification steps were performed in dedicated aDNA facilities of the Max Planck Institute for the Science of Human History (MPI-SHH). The inner ear part of the petrous bones of the individuals from Kfar

HaHoresh and Ba'ja was sampled by drilling²⁴ and DNA was extracted from 76 to 109 mg of the bone powder. An extraction of ~100 mg pulverized bone from the Pınarbaşı individual ZBC was done in the Jena facility in addition to the Liverpool extraction (the sequenced data from the two extracts of individual ZBC were merged in downstream analysis after passing the quality control step). All extractions followed the same protocol as cited for Liverpool. A 20 μ l aliquot from each extract was used to prepare an Illumina double-stranded, double-indexed DNA library following established protocols^{11,25}. Deaminated cytosines that result from DNA damage were partially removed using uracil-DNA glycosylase and endonuclease VIII, but still retained in terminal read positions as a measure of

aDNA authentication²⁶. A negative library control (H₂O) was taken along for each experiment. Unique combinations of two indexes (8 bp length each) were assigned to each library. The indexes were then attached through a ten-cycle amplification reaction using the *Pfu Turbo Cx Hotstart DNA Polymerase* (Agilent), the PCR products purified using a Qiagen MinElute kit (Qiagen), and then eluted in TET (10 mM Tris-HCl, 1 mM EDTA, pH 8.0, and 0.05% Tween-20). Subsequently, indexed libraries were amplified using Herculase II Fusion DNA polymerase, following the manufacturer's protocol, to a total of 10¹³ DNA copies per reaction and again purified using a Qiagen MinElute kit (Qiagen) and eluted in TET (10 mM Tris-HCl, 1 mM EDTA, pH 8.0 and 0.05% Tween-20). Finally, all samples were diluted and pooled (10 nM) for sequencing. The indexed amplified libraries were also used for two previously published downstream in-solution enrichments: a protocol targeting 1,237,207 genome-wide SNPs ("1240k capture"¹²) and one targeting the entire human mitochondrial genome²⁷.

The "1240k capture" is an established in-solution enrichment assay based on hybridization of the indexed libraries to DNA probes^{12,13,27,28}. The targeted SNP panel is a combination of the two separate SNP sets first reported by Haak et al.¹³ and by Fu et al.²⁸ and further described by Mathieson et al.¹². For each of the ~1.2 million target SNPs, we used four distinct 52-bp-long probes: two flanking the target SNP from each side and the other two centered on the SNP matching with the reference and alternative allele, respectively²⁸. The capture was performed following the published protocol described in detail in the SI text sections 3.2–3.3 of Fu et al.²⁸ with modified hybridization conditions of 65 °C for about 24 h.

Both the initial shotgun and target-enriched libraries were single-end sequenced on an Illumina HiSeq 4000 platform (1 × 75 + 8 + 8 cycles). Sequenced reads were demultiplexed allowing one mismatch in each index and further processed using EAGER (v 1.92.54)²⁹. First, adapter sequences were clipped and reads shorter than 30 bp were discarded using AdapterRemoval (v 2.2.0)³⁰. Adapter-clipped reads were subsequently mapped with the BWA aln/samse programs (v 0.7.12)³¹ against the UCSC genome browser's human genome reference hg19 with a lenient stringency parameter ("–n 0.01"). We retained reads with Phred-scaled mapping quality scores ≥ 20 and ≥ 30 for the whole genome and the mitochondrial genome, respectively. Duplicate reads were subsequently removed using DeDup v 0.12.2²⁹. Pseudo-diploid genotypes were generated for each individual using pileupCaller, which randomly draws a high quality base (Phred-scaled base quality score ≥ 30) mapping to each targeted SNP position (<https://github.com/stschiff/sequenceTools>). To prevent false SNP calls due to retained DNA damage, two terminal positions in each read were clipped prior to genotyping. The genotyping produced between 129,406 and 917,473 covered targeted SNPs and a mean coverage ranging between 0.16 and 2.9 fold per individual (Table 1).

Dataset. We merged the newly reported ancient data and data reported by Mathieson et al. 2018¹⁸ with a dataset that has been described elsewhere⁶. This dataset includes 587 published ancient genomes^{6–9,12,14,17,23,32–35} and genomes from 2706 individuals, representing world-wide present-day populations^{6,36} that were genotyped on the Affymetrix Axiom™ Genome-Wide Human Origins 1 array⁴ ("HO dataset") with a total of 597,573 SNP sites in the merged dataset. To minimize bias from differences in analysis pipelines, we re-processed the raw read data deposited for previously published Neolithic Anatolian genomes⁹ (labeled Tepecik_pub and Boncuklu_pub) in the same way as described for the newly reported individuals.

aDNA authentication and quality control. We estimated authenticity of the ancient data using multiple measures. First, blank controls were included and analyzed for extractions as well as library preparations (Supplementary Data 8). Second, we assessed levels of DNA damage in the mapped reads using mapDamage (v 2.0)³⁷. Third, we estimated human DNA contamination on the mitochondrial DNA using schmutzi³⁸. Last, we estimated nuclear contamination in males with ANGSD (v 0.910)³⁹, which utilizes haploid X chromosome markers in males by comparing mismatch rates of polymorphic sites and adjacent ones (that are likely to be monomorphic). The genetic sex of the reported individuals was determined by comparing the genomic coverage of X and Y chromosomes normalized by the autosomal average coverage. To avoid bias caused by grouping closely related individuals into a population, we calculated the pairwise mismatch rates of the Boncuklu individuals following a previously reported method⁴⁰ (Supplementary Data 9).

Five of the 12 individuals reported here were excluded from the population genetic analysis: two due to a high genomic contamination level (>5%) and three due to low amount of analyzable data (<10,000 SNPs covered); (Supplementary Data 1).

Principal component analysis. We used the smartpca software from the EIGENSOFT package (v 6.0.1)⁴¹ with the lsqproject option to construct the principal components of 67 present-day west Eurasian groups and project the ancient individuals on the first two components (Supplementary Figure S8).

ADMIXTURE analysis. We used ADMIXTURE (v 1.3.0)⁴² to perform a maximum-likelihood unsupervised clustering of 3293 ancient and present-day individuals in the HO merged dataset, allowing the number of clusters (*k*) to range

between 2 and 20. Pruning for linkage disequilibrium (LD) was done by randomly removing one SNP from each pair with genotype $r^2 \geq 0.2$, using PLINK (v 1.90)^{43,44}, (–indep-pairwise 200 25 0.2). The analysis was replicated five times for each *k* value with random seeds and the highest likelihood replicate is reported (Supplementary Figures 1 and 9). Five-fold cross-validation errors were calculated for each run. Using the same settings, we additionally preformed the clustering on a smaller sample size of 983 ancient and modern west Eurasian individuals, which produced a clustering pattern comparable to that of the larger dataset.

D-statistics. To estimate allele frequency correlations between populations, *D*-statistics were computed using the *qpDstat* program (v 701) of the ADMIXTOOLS package⁴⁵ (v 4.1) with default parameters. *D*-statistics provide a robust and sensitive test of gene flow and are preferable for low quantity data analysis (typical of Archeogenetic studies) as they are insensitive to post-admixture drift, including artifactual drift due to a limited sample size⁴⁵. In order to determine whether a test population is symmetrically related to populations *X* and *Y*, the *D*-statistic *D*(*X*, *Y*; *Test*, *Outgroup*) was used. In particular, when comparing the affinity of different European hunter-gatherers to Near-Eastern ones in the *D*-statistic of the form *D*(*European HG1*, *European HG2*; *Near Eastern HG*, *Outgroup*), both the central African *Mbuti* and the Altai Neanderthal (*Altai_published.DG*) were used to check if the differing level of Neanderthal ancestry in these hunter-gatherers affects the results. Otherwise, *Mbuti* was used as the single outgroup. The above statistics are reported when more than 30,000 SNP positions were overlapping between the four analyzed populations. To further validate the *D*-statistics of the form *D*(*Anatolian I*, *Anatolian2*; *test*, *Mbuti*) beyond the jackknifing performed by *qpDstat*, we compared the inferred *D*-statistics based on the population mean to the distribution of the *D*-statistic when individuals are permuted between populations. We performed the permutation tests in the following settings: (1) for the *D*-statistics of the form *D*(*AAF**, *AHG**; *test*, *Mbuti*), we performed all five possible permutations. In each permutation, we placed one out of the five AAF individuals into the second position (*AHG**) while placing the other four individuals and the AHG individual into the first position (*AAF**) (Supplementary Data 11). To obtain the distribution within AAF we repeated the analysis, but now excluding AHG. The same set of global modern and ancient populations as in the original test was used as the "test." (2) For the *D*-statistics of the form *D*(*ACF**, *AAF**; *test*, *Mbuti*) a total of 1,000 permutations were performed, in addition to the original test, for each of the four "test" populations that had the most positive values in the original observed statistic (i.e., *Levant_N*, *Natufian*, *Greece_EN*, *Balkans_Neolithic*). In each test, we randomly chose five out of 30 individuals (5 AAF and 25 ACF) and placed them into the second position (*AAF**) while placing the rest into the first position (*ACF**). Empirical *P*-values were calculated by dividing the number of permutations with a *D*-statistic equal to or greater than the original observation by the total number of permutations (i.e., 1001).

Modeling ancestry proportions. We used the qpWave (v400) and qpAdm (v 632) programs of ADMIXTOOLS^{6,13} to test and model admixture proportions in a studied population from potential source populations (reference populations). As the explicit phylogeny is unknown, a diverse set of outgroup populations (Supplementary Notes 2–4) was used to distinguish the ancestry of the reference populations.

For estimating admixture proportions in the tested populations, we used a basic set of seven outgroups including present-day populations (Han, Onge, *Mbuti*, Mala, Mixe) that represent a global genetic variation and published ancient populations such as *Natufian*⁶, which represents a Levantine gene pool outside of modern genetic variation and the European Upper Paleolithic individual Kostenki14¹⁷. As a prerequisite for the admixture modeling of the target population, we tested whether the corresponding set of reference populations can be distinguished by the chosen outgroups using qpWave⁶ (Supplementary Note 3). In some cases, when a reference population did not significantly contribute to the target in the attempted admixture models, it was removed from the reference set and added to the basic outgroup set in order to increase the power to distinguish the references. In cases where "Natufian" was used as a reference population, we instead used the present-day Near-Eastern population "BedouinB" as an outgroup.

For estimations of Basal Eurasian ancestry, we followed a previously described qpAdm approach⁶ that does not require a proper proxy for the Basal Eurasian ancestry, which is currently not available in unadmixed form. This framework relies on the basal phylogenetic position of both Basal Eurasian and an African reference (the ancient Ethiopian *Mota* genome¹⁹) relative to other non-Africans. Thus, by using a set of outgroups that includes eastern non-African populations (Han; Onge; Papuan) and Upper Paleolithic Eurasian genomes (*Ust'-Ishim*⁴⁶, *Kostenki14*, *MA-1*⁴⁷), but neither west Eurasians with detectable basal Eurasian ancestry nor Africans, the mixture proportion computed for *Mota* (α) can be used indirectly to estimate the Basal Eurasian mixture proportion of west Eurasian populations.

Mitochondrial DNA analysis. The endogenous mitochondrial consensus sequences were inferred from the output of schmutzi³⁸, using its log2fasta program and a quality cutoff of 10. Mitochondrial haplotypes were established by aligning these consensus to rCRS⁴⁸ using the online tool haplosearch⁴⁹. The coverage of

each of the reported SNPs was confirmed by visually inspecting the bam pileup in Geneious (v11.0.4)⁵⁰. The resulting consensus sequences were then analyzed with HaploFind⁵¹ and Haplogrep⁵² to assign mitochondrial haplogroups and double-checked with the rCRS oriented version of Phylotree⁵³.

Y-chromosome analysis. To assign Y-chromosome haplogroups we used yHaplo⁵⁴. Each male individual was genotyped at 13,508 ISOGG consortium SNP positions (strand-ambiguous SNPs were excluded) by randomly drawing a single base mapping to the SNP position, using the same quality filters as for the HO dataset. In addition to the yHaplo automated haplogroup designations, we manually verified the presence of derived alleles supporting the haplogroup assignment.

Phenotypic traits analyses. We tested for the presence of alleles related to biological traits that could be of interest in the geographical and temporal context of the reported ancient populations, including lactase persistence^{55,56}, Malaria resistance^{57,58}, glucose-6-phosphate dehydrogenase deficiency^{59,60}, and skin pigmentation^{23,61,62}. The allele distribution for the SNP positions listed in Supplementary Data 7 was tabulated for each individual using Samtools mpileup (v 1.3).

Carbon dating. The phalanx bone from individual ZBC (Pinarbaşı) and the petrous bone from individual KFH2 (Kfar HaHoresh) were each sampled and directly radiocarbon dated at the CEZ Archaeometry gGmbH, Mannheim, Germany (Supplementary Table 1). Collagen was extracted from the bone samples, purified by ultrafiltration (fraction >30kDa), freeze-dried, and combusted to CO₂ in an elemental analyzer. CO₂ was converted catalytically to graphite. The dating was performed using the MICADAS-AMS of the Klaus-Tschira-Archäometrie-Zentrum. The resulting ¹⁴C ages were normalized to δ¹³C = −25‰⁶³ and calibrated using the dataset INTCAL13⁶⁴ and the software SwissCal 1.0⁶⁵.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All Alignment data (BAM) produced in this study is deposited in the European Nucleotide Archive (ENA) under the accession numbers (Study PRJEB24794). Other data supporting the findings of the study are available in this article and its Supplementary Information files, or from the corresponding authors upon request. The human skeletal specimens from Pinarbaşı are housed in the Karaman Museum, Turkey. They are available for study with the authorization of Professor Douglas Baird and the Directorate of Antiquities and Museums of Turkey. The human skeletal specimens from Boncuklu are housed in the Boncuklu excavation depot, under the purview of Konya Museum, Turkey. They are available for study with the authorization of Professor Douglas Baird and the Directorate of Antiquities and Museums of Turkey. The skeletal specimens of the Kfar HaHoresh individual are housed in the anthropological collection, the Dan David Center for Human Evolution and Biohistory Research, the Sackler faculty of medicine, Tel Aviv University, Israel and are available upon request and authorization of Prof. Nigel Goring-Morris, Prof. Israel Hershkovitz, and Dr. Hila May. The skeletal specimens of the Baj'a individual are housed in the Department of Anthropology, German Archeological Institute, Berlin, Germany and are available upon request and authorization of Dr. Julia Gresky.

Received: 20 June 2018 Accepted: 15 February 2019

Published online: 19 March 2019

References

- Bellwood, P. *First Farmers. The Origins of Agricultural Societies* (Wiley-Blackwell, Chichester, 2005).
- Baird, D. in *A Companion to the Archaeology of the Ancient Near East* (ed. Potts, D.) 431–465 (Wiley-Blackwell, Malden, 2012).
- Baird, D. et al. Agricultural origins on the Anatolian plateau. *Proc. Natl Acad. Sci. USA* **115**, E3077–E3086 (2018).
- Lazaridis, I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
- Bramanti, B. et al. Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* **326**, 137–140 (2009).
- Lazaridis, I. et al. Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016).
- Jones, E. R. et al. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015).
- Fu, Q. et al. The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
- Kılınç, G. M. et al. The demographic development of the first farmers in Anatolia. *Curr. Biol.* **26**, 2659–2666 (2016).
- Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. USA* **110**, 15758–15763 (2013).
- Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, 5448 (2010).
- Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
- Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
- Broushaki, F. et al. Early Neolithic genomes from the eastern Fertile Crescent. *Science* **353**, 499–503 (2016).
- Narasimhan, V. M. et al. The genomic formation of South and Central Asia. *bioRxiv* 292581 (2018).
- Gamble, C., Davies, W., Pettitt, P. & Richards, M. Climate change and evolving human diversity in Europe during the last glacial. *Philos. Trans. R. Soc. Lond. Ser. B* **359**, 243–253 (2004). discussion 253–4.
- Seguin-Orlando, A. et al. Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113–1118 (2014).
- Mathieson, I. et al. The genomic history of Southeastern Europe. *Nature* **555**, 197–203 (2018).
- Gallego Llorente, M. et al. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* **350**, 820–822 (2015).
- Posth, C. et al. Pleistocene mitochondrial genomes suggest a single major dispersal of non-Africans and a Late Glacial population turnover in Europe. *Curr. Biol.* **26**, 827–833 (2016).
- Sikora, M. et al. Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science* **358**, 659–662 (2017).
- Sturm, R. A. et al. A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. *Am. J. Human. Genet.* **82**, 424–431 (2008).
- Olalde, I. et al. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225–228 (2014).
- Pinhasi, R. et al. Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS ONE* **10**, e0129102 (2015).
- Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2011).
- Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. Ser. B* **370**, 20130624 (2015).
- Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
- Fu, Q. et al. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110**, 2223–2227 (2013).
- Peltzer, A. et al. EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).
- Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Prüfer, K. et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
- Hofmanova, Z. et al. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl Acad. Sci. USA* **113**, 6886–6891 (2016).
- Gallego-Llorente, M. et al. The genetics of an early Neolithic pastoralist from the Zagros, Iran. *Sci. Rep.* **6**, 31326 (2016).
- Omrak, A. et al. Genomic evidence establishes Anatolia as the source of the European Neolithic gene pool. *Curr. Biol.* **26**, 270–275 (2016).
- Mallick, S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
- Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
- Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015).
- Korneliusson, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinforma.* **15**, 356 (2014).
- Kennett, D. J. et al. Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat. Commun.* **8**, 14115 (2017).
- Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
- Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

43. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
44. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
45. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
46. Fu, Q. et al. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014).
47. Raghavan, M. et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
48. Andrews, R. M. et al. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147 (1999).
49. Fregel, R. & Delgado, S. HaploSearch: a tool for haplotype-sequence two-way transformation. *Mitochondrion* **11**, 366–367 (2011).
50. Kearse, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
51. Vianello, D. et al. HAPLOFIND: a new method for high-throughput mtDNA Haplogroup assignment. *Hum. Mutat.* **34**, 1189–1194 (2013).
52. Kloss-Brandstätter, A. et al. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat.* **32**, 25–32 (2011).
53. Van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, e386 (2009).
54. Poznik, G. D. Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. Preprint at <https://doi.org/10.1101/088716> (2016).
55. Ingram, C. J. et al. A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? *Hum. Genet.* **120**, 779–788 (2007).
56. Tishkoff, S. A. et al. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* **39**, 31–40 (2007).
57. Network, M. G. E. et al. Reappraisal of known malaria resistance loci in a large multicenter study. *Nat. Genet.* **46**, 1197 (2014).
58. Timmann, C. et al. Genome-wide association study indicates two novel resistance loci for severe malaria. *Nature* **489**, 443–446 (2012).
59. Sepúlveda, N. et al. Malaria host candidate genes validated by association with current, recent, and historical measures of transmission intensity. *J. Infect. Dis.* **216**, 45–54 (2017).
60. Maiga, B. et al. Glucose-6-phosphate dehydrogenase polymorphisms and susceptibility to mild malaria in Dogon and Fulani, Mali. *Malar. J.* **13**, 270 (2014).
61. Norton, H. L. et al. Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol. Biol. Evol.* **24**, 710–722 (2006).
62. Wilde, S. et al. Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proc. Natl Acad. Sci. USA* **111**, 4832–4837 (2014).
63. Stuiver, M. & Polach, H. A. Discussion reporting of ^{14}C data. *Radiocarbon* **19**, 355–363 (1977).
64. Reimer, P. J. et al. IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).
65. Christl, M. et al. The ETH Zurich AMS facilities: performance parameters and reference materials. *Nucl. Instrum. Methods Phys. Res. Sect. B* **294**, 29–38 (2013).

Acknowledgements

We thank G. Brandt, A. Wissgott, F. Aron, M. Burri, C. Freund, and R. Stahl (MPI-SHH) for their support in laboratory work, M. O'Reilly for graphic support, A. Gibson for help

in proofreading, and the members of the population genetics group in the Department of Archaeogenetics, SHH-MPI for their input and support. This work was supported by the Max Planck Society. Experimental work at LJM and LR stay at the MPI-SHH were funded with an internal grant from LJM Faculty of Science to EFD. The funding for the Ba'ja project was granted by the German Research Foundation (GZ: 80 1599/14–1) and ex oriente e.V., Berlin. Kfar HaHoresh fieldwork was supported (to N.G.-M.) by the Israel Science Foundation (Grants 840/01, 558/04, 755/07, 1161/10), The National Geographic Society (Grant 8625/09), and The Irene Levi Sala CARE Foundation. The Konya Plain fieldwork was funded by The British Institute in Ankara, British Academy (Research Development Award BR100077), a British Academy Large Research Grant LRG 35439, Australian Research Council (grants DP0663385 and DP120100969), National Geographic award GEFNE 1-11, University of Oxford (Wainwright Fund), Australian Institute for Nuclear Science and Engineering (AINSE awards AINGRA05051 and AINGRA10069), Wenner-Gren Foundation for Anthropological Research (Postdoctoral Research Grant 2008 The Origins Of Farming In The Konya Plain, Central Anatolia), Institute for Field Research.

Author contributions

J.K., E.F.-D., I.H., C.J., W.H. and M.F. conceived the study. D.B., A.F., G.M., J.P., I.H., H. M., N.G.-M., M.B. and J.G. provided archeological material. D.B., J.P., A.F., G.M., I.H., H. M., N.G.-M., M.B., J.G. and P.W.S. advised on the archeological background and interpretation. D.B., J.P., E.F.-D., N.G.-M., M.B. and J.G. wrote the archeological and sample background sections. M.F., L.R. and R.A.B. performed the laboratory work. M.F., E.F.-D., L.R., C.P. and C.J. performed data analysis. M.F., E.F.-D., C.J. and J.K. wrote the manuscript with input from all coauthors.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-019-09209-7>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Journal peer review information: *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

5. Manuscript B

ANTHROPOLOGY

Ancient DNA sheds light on the genetic origins of early Iron Age Philistines

Michal Feldman¹, Daniel M. Master^{2,3}, Raffaella A. Bianco¹, Marta Burri¹, Philipp W. Stockhammer^{1,4}, Alissa Mittnik^{1,5}, Adam J. Aja³, Choongwon Jeong^{1,6*}, Johannes Krause^{1*}

The ancient Mediterranean port city of Ashkelon, identified as “Philistine” during the Iron Age, underwent a marked cultural change between the Late Bronze and the early Iron Age. It has been long debated whether this change was driven by a substantial movement of people, possibly linked to a larger migration of the so-called “Sea Peoples.” Here, we report genome-wide data of 10 Bronze and Iron Age individuals from Ashkelon. We find that the early Iron Age population was genetically distinct due to a European-related admixture. This genetic signal is no longer detectible in the later Iron Age population. Our results support that a migration event occurred during the Bronze to Iron Age transition in Ashkelon but did not leave a long-lasting genetic signature.

INTRODUCTION

Within the history of the Eastern Mediterranean, the end of the Bronze Age and the beginning of the Iron Age were marked by exceptional cultural disarray that followed the demise of prosperous economies and cultures in Greece, Egypt, the Levant, and Anatolia (1). During the 12th century BCE, coincident with these events, substantive cultural changes appeared in the archeological record of Ashkelon, Ashdod, and Ekron, three of the five core cities mentioned as “Philistine” in the Hebrew bible (2–4). These settlements were distinct from neighboring sites in architectural traditions and material culture (2–4). Resemblances between the new cultural traits and 13th century patterns found in the Aegean have led some scholars to explain this so-called “Philistine phenomenon” by a migration from an Aegean-related source, potentially associated with the “Sea Peoples,” a population that is thought to have settled in other parts of the coastal Eastern Mediterranean (2, 3). This hypothesis has been challenged by those arguing that this cultural change was driven by a diffusion of knowledge or internal development of ideas (5–8) rather than by a large-scale movement of people. Even for those who do accept the idea of large-scale mobility, the homeland of the new arrivals is disputed with suggested alternatives including Cyprus or Cilicia (4), a mixture of non-Aegean east Mediterranean peoples (8), and mixed heterogeneous maritime groups, akin to pirates (9). Proposed links go as far as northern Italy where depopulation events have been suggested to trigger population movements throughout the Mediterranean (10).

Recent ancient DNA (aDNA) studies have reported a high degree of genetic continuity in the Levant during the late Pleistocene and early Holocene that was followed by an increasing population admixture with Anatolian- and Iranian-related populations at least up to the Middle Bronze Age (11–14). Genome-wide data from Late Bronze and Iron Age populations have, so far, not been available for this region.

Here, we report new genome-wide data from human remains excavated at the ancient seaport of Ashkelon, forming a genetic time series encompassing the Bronze to Iron Age transition (Fig. 1, A and B). We find that all three Ashkelon populations derive most of their ancestry from the local Levantine gene pool. The early Iron Age population was distinct in its high genetic affinity to European-derived populations and in the high variation of that affinity, suggesting that a gene flow from a European-related gene pool entered Ashkelon either at the end of the Bronze Age or at the beginning of the Iron Age. Of the available contemporaneous populations, we model the southern European gene pool as the best proxy for this incoming gene flow. Last, we observe that the excess European affinity of the early Iron Age individuals does not persist in the later Iron Age population, suggesting that it had a limited genetic impact on the long-term population structure of the people in Ashkelon.

RESULTS

Genome-wide data across the Bronze/Iron Age transition in Ashkelon

We extracted and sequenced DNA from 108 skeletal elements excavated in Ashkelon. In line with the low DNA preservation previously reported for the southern Levant (11–14), only 10 yielded sufficient amounts of human DNA (data file S1). Sequencing libraries for these 10 individuals were enriched for human DNA using an established in-solution DNA capture targeting 1.24 million genome-wide single-nucleotide polymorphisms (SNPs) (“1240 k capture”) (15–17). The earliest three individuals were excavated from a Bronze Age necropolis (18–20) and dated by the archeological context to the Middle Bronze IIC–Late Bronze Age II (labeled ASH_LBA; two of the individuals were directly carbon-dated to 1746 to 1542 cal BCE). Four early Iron Age infants were recovered from burials beneath late 12th century Iron Age I houses (labeled ASH_IA1; directly carbon-dated to 1379 to 1126 cal BCE), and three later Iron Age individuals were recovered from a cemetery adjacent to the city wall of ancient Ashkelon, which was dated to the 10th to 9th century BCE (labeled ASH_IA2; one individual directly carbon-dated to 1257 to 1042 cal BCE) (Fig. 1, A and B, Table 1, table S1, and text S1). While the chronological ranges estimated by the archeological context are approximately within the range of the radiocarbon assays conducted directly on the petrous bones sampled for DNA (table S1), we caution that they should be

Copyright © 2019
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

¹Department of Archaeogenetics, Max Planck Institute for the Science of Human History (MPI-SHH), Kahlaische Strasse 10, D-07745 Jena, Germany. ²Wheaton Archaeology Museum, Wheaton College, Wheaton, IL 60187, USA. ³Harvard Semitic Museum, Harvard University, Cambridge, MA 02138, USA. ⁴Institut für Vor- und Frühgeschichtliche Archäologie und Provinzialrömische Archäologie Ludwig-Maximilians-Universität München München, Schellingstrasse 12, D-80799 München, Germany. ⁵Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. ⁶School of Biological Sciences, Seoul National University, Seoul 08826, Republic of Korea.

*Corresponding author. Email: cwjeong@snu.ac.kr (C.J.); krause@shh.mpg.de (J.K.)

MS no:	Emp name:	Date / Time:	PE's:	AA's:	Comments:	Art no:
RAaax0061/AF/ANTHROPOLOGY	kmatro	6-3-2019 / 16:51				

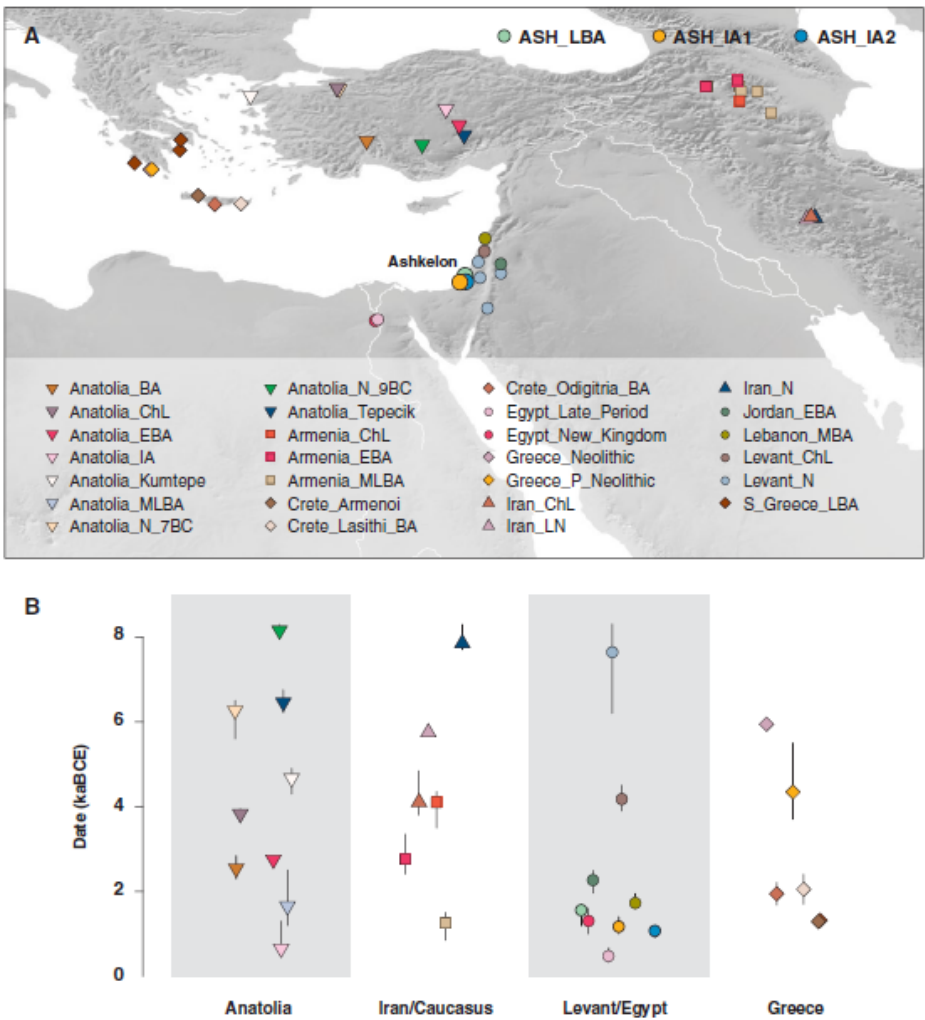


Fig. 1. Overview of locations and ages of analyzed individuals. (A) Locations of newly reported and other selected published genomes. The newly reported Ashkelon populations are annotated in the upper corner. (B) The range of average ages of the ancient groups.

taken as a rough estimate, since the presence of marine carbon in the diet could make the calibrated radiocarbon dates older than the true age.

To control for the quality of the dataset, we estimated exogenous DNA levels and relatedness. Mitochondrial contamination for all 10 individuals ranged between 0 and 9%, with only two of them above 5% (table S2). For males ($n = 4$), nuclear contamination was estimated to a maximum of 4.3% (table S2). None of the individuals were either first- or second-degree related to each other (fig. S1). We performed population genetic analysis on a merged dataset, including the genotype data of the newly reported individuals from this study and previously published datasets from 638 ancient individuals and 4943 individuals belonging to 298 present-day populations. (11, 12, 21). (An overview of the key ancient individuals analyzed is given in data file S2).

Persistence of the local gene pool in the Bronze Age Levant

To understand the genetic profile of the ancient Ashkelon individuals, we began by projecting them onto the first two axes of variation

(PC1 and PC2) of present-day west Eurasians, inferred from principal components analysis (PCA) (Fig. 2A and fig. S2) (12). The ASH_LBA individuals overlap with the cline of present-day Near Easterners and are close to earlier Bronze Age Levantines and Anatolians [Early Bronze Age individuals from ‘Ain-Ghazal, Jordan labeled “Jordan_EBA” (12); Middle Bronze Age individuals from Sidon, Lebanon labeled “Lebanon_MBA” (13); and Early to Late Bronze Age individuals from Central Anatolia labeled “Anatolia_EBA” (21) and “Anatolia_MLBA” (21), respectively]. Compared to earlier Levantines [Levantine early farmers from present-day Jordan and Israel labeled “Levant_N” (12) and Chalcolithic individuals from Peqi’in, Israel labeled “Levant_ChL” (11)], the Bronze Age individuals including ASH_LBA are all shifted along PC2 toward ancient Iranian and Caucasus individuals [e.g., the Caucasus Mesolithic hunter-gatherers labeled “CHG” (22); early farmers excavated from present-day Iran labeled “Iran_N” (12); and the Chalcolithic individuals excavated from present-day western Iran labeled “Iran_ChL” (12)], in agreement with previous observations (11–13). Unsupervised genetic clustering using ADMIXTURE (23) shows a similar pattern [with $K = 9$ clusters; Fig. 2B and fig. S3];

MS no:	Emp name:	Date / Time:	PE's:	AA's:	Comments:	Art no:
RAaax0061/AF/ANTHROPOLOGY	kmatro	6-3-2019 / 16:51				Fig 1

Table 1. An overview of the Ashkelon genomes reported in this study. For each individual, the analysis group is given (ASH_LBA, Ashkelon Late Bronze Age; ASH_IA1, Ashkelon Iron Age 1; ASH_IA2, Ashkelon Iron Age 2). ¹⁴C dating results are given in cal BCE in two-sigma range (detailed dating information is provided in text S1 and table S1). The proportion of human DNA and the mean coverage on 1240 K target sites in the “1240 K” enriched libraries are given. Uniparental haplogroups (mt, mitochondrial; Ychr, Y chromosome) are listed.

ID	Analysis group	Sampled tissue	Mean coverage (fold)	Carbon date	Archeological date	Total sequenced reads (x106)	Human DNA (%)	Genetic sex	mt	Ychr
ASH029.A0101	ASH_LBA	Petrous	0.14	1622–1522	MBIIC–LBII	86.6	6.9	F	H66a	
ASH033.A0101	ASH_LBA	Petrous	0.11	1746–1643	MBIIC–LBII	84.8	2	F	N	
ASH034.A0101	ASH_LBA	Petrous	0.42	NA	MBIIC–LBII	81.9	13.7	F	U3b1a	
ASH066.A0101	ASH_IA1	Petrous	0.18	1371–1129	Iron I–Post Ramses III	10.5	11.0	M	T2c1c	J
ASH067.A0101	ASH_IA1	Petrous	0.17	1379–1131	Iron I–Post Ramses III	10.9	11.2	M	H92	R1
ASH002/3.A0101	ASH_IA1	Petrous	0.08	1378–1134	Iron I–Post Ramses III	26.9	13	F	I1	
ASH068.A0101	ASH_IA1	Petrous	0.27	1284–1126	Iron I–Post Ramses III	11.2	17.6	F	T1a1	
ASH008.A0101	ASH_IA2	Petrous	0.7	1257–1042	Iron IIA	126.7	17.2	M	H2c	BT
ASH087.A0101	ASH_IA2	Petrous	0.22	NA	Iron IIA	12.8	15.6	M	H4a1c	L
ASH135.A0101	ASH_IA2	Molar	0.09	NA	Iron IIA	13.3	3.9	F	JT	

ASH_LBA is assigned the same major ancestral component as all earlier Levantine populations (shown in orange). Consistent with their PCA positions, a second component (shown in green) that is maximized in Iran_N is, on average, higher in ASH_LBA and in each of the earlier Bronze Age Levantines compared to all earlier Levantines (20 to 30% and 3 to 8%, respectively).

To formally test the qualitative observations based on the PCA and ADMIXTURE analyses, we used *f*₄-statistics (24). Consistent with the ASH_LBA positions on PC2, the *f*₄-statistic of the form *f*₄(ASH_LBA, Levant_N/Levant_ChL; test, Mbuti) (figs. S4 and S5) estimated excess allele sharing between ASH_LBA and ancient Iranian/Caucasus-related populations (such as CHG, Iran_N, and Iran_ChL) compared to the Neolithic/Chalcolithic Levantines (Levant_N and Levant_ChL), confirming previous reports of post-Neolithic gene flows from prehistoric populations related to Iran or the Caucasus into the Levant (11–13). Accordingly, modeling ASH_LBA as a two-way admixture using qpAdm (12, 17) produced a fitting model ($\chi^2P = 0.445$), in which ASH_LBA derives around 60% of their ancestry from Levant_ChL (60.0 ± 6.5%; estimate ± 1 SE) and the rest from Iran_ChL. The spatiotemporal origins of this post-Neolithic gene flow remain broadly defined because of alternative two-way models in which Anatolia_EBA/Anatolia_BA/Armenia_ChL are used as the second source, replacing Iran_ChL also fit albeit with smaller *P* values ($\chi^2P = 0.053$ to 0.060; table S3).

We then examined whether there are noticeable genetic differences between ASH_LBA and the earlier Bronze Age Levant populations by measuring *f*₄(ASH_LBA, Jordan_EBA/Lebanon_MBA; test, Mbuti/Chimp) (figs. S6 and S7). We find that ASH_LBA and both earlier Bronze Age populations to be symmetrically related to all test populations within our data’s resolution ($|f_4| < 3$ SE). However, we observe a marginal excess of affinity between ASH_LBA and populations genetically related to ancient Iran and the Caucasus (such as CHG and Steppe_Eneolithic) when compared to Jordan_EBA (fig. S7). Within our current resolution, we cannot distinguish whether this affinity is due to a small-scale gene flow from an Iranian/Caucasus-related population entering the Levant between the early and late Bronze Age or whether it reflects a certain population structure during this period.

Nonetheless, the apparent symmetry suggests a high degree of genetic continuity throughout at least a millennium between culturally and geographically distinct Bronze Age Levantine groups occupying a region that stretches from the inland southern Levant where present-day Jordan is located and along the coastal regions of present-day Israel and Lebanon.

Genetic discontinuity between the Bronze Age and the early Iron Age people of Ashkelon

In comparison to ASH_LBA, the four ASH_IA1 individuals from the following Iron Age I period are, on average, shifted along PC1 toward the European cline and are more spread out along PC1, overlapping with ASH_LBA on one extreme and with the Greek Late Bronze Age “S-Greece_LBA” (25) on the other (Fig. 2A). Similarly, genetic clustering assigns ASH_IA1 with an average of 14% contribution from a cluster maximized in the Mesolithic European hunter-gatherers labeled “WHG” (shown in blue in Fig. 2B) (15, 22, 26). This component is inferred only in small proportions in earlier Bronze Age Levantine populations (2 to 9%). In light of these observations, we formally tested the variation within each of the three Ashkelon populations by computing the variance of the square rooted statistic *f*₄(Ashkelon ind 1, Ashkelon ind 2; test, Mbuti) (fig. S8). Consistent with the wide spread of ASH_IA1 in the west Eurasian PCA, the variance is significantly higher in ASH_IA1 than in either ASH_LBA or ASH_IA2 ($P < 2.2 \times 10^{-16}$ for both by a Fligner-Killeen test) (table S4 and fig. S8), suggesting that the ASH_IA1 individuals were more heterogeneous in their genetic affinities to global populations in respect to both ASH_LBA and ASH_IA2.

We next formally quantified the genetic differences between ASH_IA1 and the earlier ASH_LBA by calculating *f*₄(ASH_IA1, ASH_LBA; test, Mbuti) (Fig. 3A). In agreement with the PCA and ADMIXTURE results, only European hunter-gatherers (including WHG) and populations sharing a history of genetic admixture with European hunter-gatherers (e.g., as European Neolithic and post-Neolithic populations) produced significantly positive *f*₄-statistics ($Z \geq 3$), suggesting that, compared to ASH_LBA, ASH_IA1 has additional European-related ancestry. To estimate the levels of the European-related ancestry in

MS no:	Emp name:	Date / Time:	PE's:	AA's:	Comments:	Art no:
RAaax0061/AF/ANTHROPOLOGY	kmatro	6-3-2019 / 16:51				

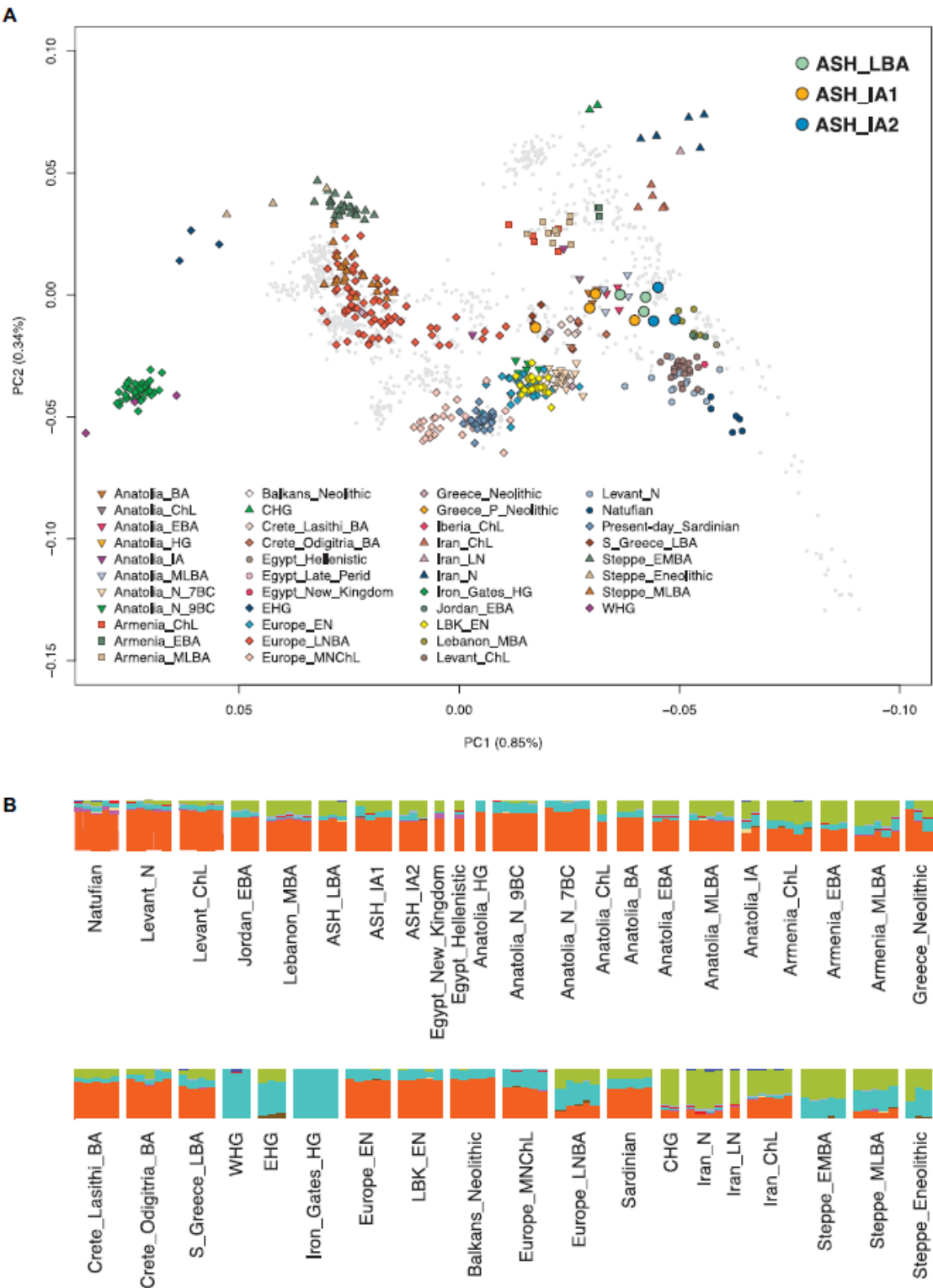


Fig. 2. PCA and ADMIXTURE analysis. (A) Ancient genomes (marked with color-filled symbols) projected onto the principal components inferred from present-day west Eurasians (gray circles) (fig. S2). The newly reported Ashkelon populations are annotated in the upper corner. (B) ADMIXTURE analysis. A selected set of ancient individuals (as well as present-day Sardinians) is plotted ($K=9$ was chosen since it is the cluster number that maximizes components correlated to the most differentiated populations in the west Eurasian PCA).

MS no:	Emp name:	Date / Time:	PE's:	AA's:	Comments:	Art no:
RAaax0061/AF/ANTHROPOLOGY	kmatro	6-3-2019 / 16:51				Fig 2

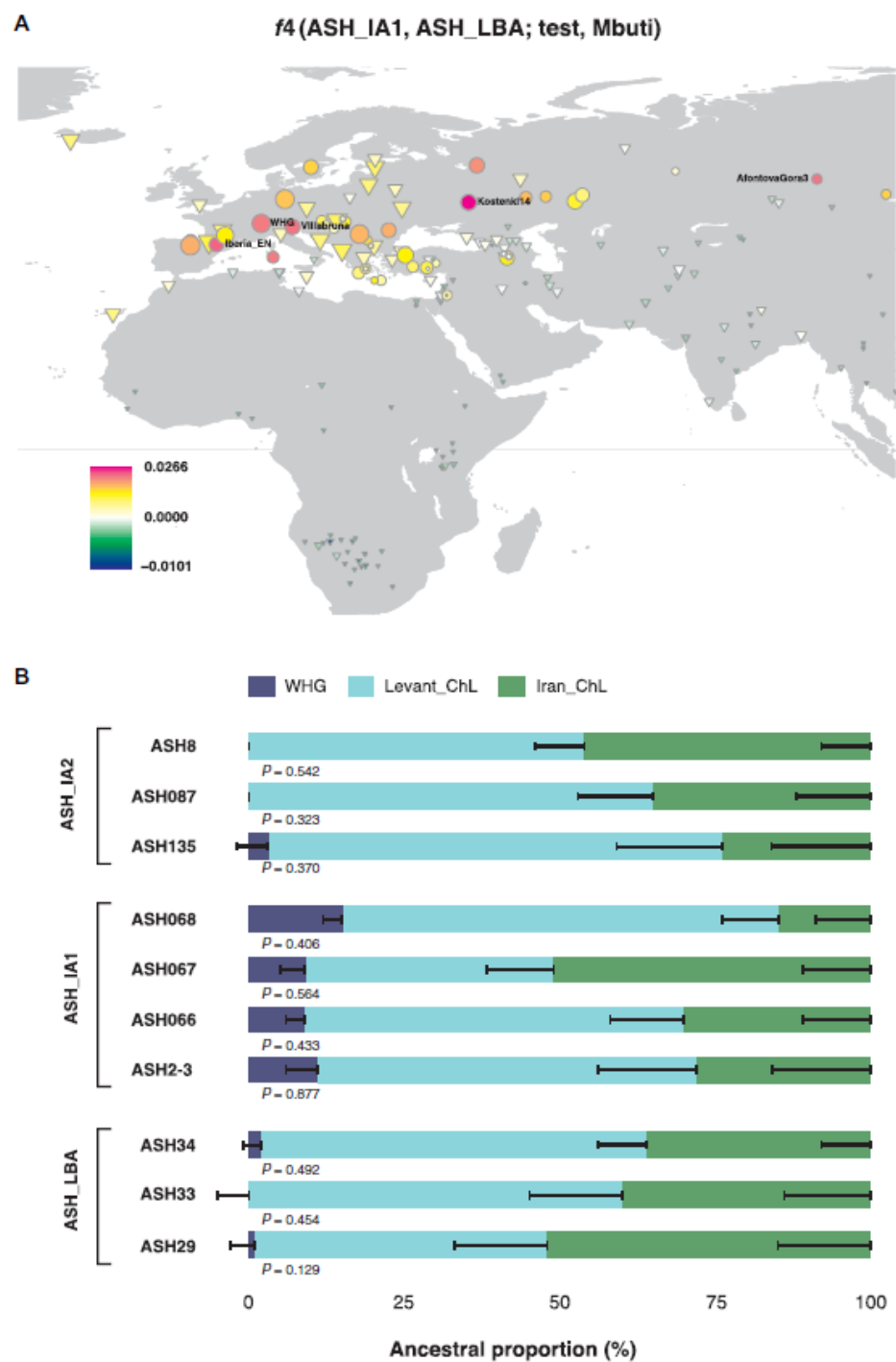


Fig. 3. ASH_IA1 received a European-related gene flow (A) ASH_IA1 shares access affinity with European-related populations compared to ASH_LBA. We plot the top and bottom 40 values of $f_4(\text{ASH_IA1}, \text{ASH_IA2}; X, \text{Mbuti})$ on the map. Circles mark the ancient populations and triangles the present-day ones. Z-scores calculated by 5-centimorgan block jackknifing are represented by the size of the symbols. "X" share more alleles with ASH_IA1 when values are positive and with ASH_IA2 when negative. The five groups with the most positive values are annotated on the map ($Z > 2.3$). (B) We plot the ancestral proportions of the Ashkelon individuals inferred by qpAdm using Iran_ChL, Levant_ChL, and WHG as sources ± 1 SEs. P values are annotated under each model. In cases when the three-way model failed ($\chi^2 P < 0.05$), we plot the fitting two-way model. The WHG ancestry is necessary only in ASH_IA1.

MS no:	Emp name:	Date / Time:	PE's:	AA's:	Comments:	Art no:
RAaax0061/AF/ANTHROPOLOGY	kmatro	6-3-2019 / 16:51				Fig 3

all Ashkelon populations, we compared qpAdm models of two-way mixtures (Levant_ChL and Iran_ChL; i.e., 0% contribution from WHG) to three-way ones, in which we add WHG as the third source (Fig. 3C and table S5). ASH_LBA and ASH_IA2 fit well with the two-way model ($\chi^2P = 0.445$ and $\chi^2P = 0.313$, respectively; table S5), whereas the three-way one infers small nonsignificant proportions (-2.3 ± 2.2 and $2.0 \pm 2.2\%$ for ASH_LBA and ASH_IA2, respectively; table S5). In contrast, for ASH_IA1, the two-way model is inadequate ($\chi^2P = 3.80 \times 10^6$; table S5) and the three-way one fits ($\chi^2P = 0.765$). Thus, of the three, additional WHG-like ancestry is only necessary to model ASH_IA1.

We find that the PC1 coordinates positively correlate with the proportion of WHG ancestry modeled in the Ashkelon individuals (fig. S9 and table S6), suggesting that WHG reasonably tag a European-related ancestral component within the ASH_IA1 individuals. However, these Mesolithic individuals are unlikely to be a good proxy for the true source in the much later early Iron Age. To examine more proximate sources, we compiled a set of chronologically and geographically relevant candidate populations, including populations that shared higher affinity with ASH_IA1 compared to ASH_LBA in the above f_4 -statistic. Subsequently, we modeled ASH_IA1 as two-way and three-way mixtures of ASH_LBA and combinations of the candidate populations (table S7). Of the 51 tested models, we find four plausible ones ($\chi^2P > 0.05$), all are two-way mixtures. The best supported one ($\chi^2P = 0.675$) infers that ASH_IA1 derives around 43% of ancestry from the Greek Bronze Age "Crete_Odigitria_BA" ($43.1 \pm 19.2\%$) and the rest from the ASH_LBA population. ASH_IA1 could also be modeled with either the modern "Sardinian" ($35.2 \pm 17.4\%$; $\chi^2P = 0.070$), the Bronze Age "Iberia_BA" ($21.8 \pm 21.1\%$; $\chi^2P = 0.205$), or the Bronze Age "Steppe_MLBA" ($15.7 \pm 9.1\%$; $\chi^2P = 0.050$) as the second source population to ASH_LBA. To check whether these results are due to the low coverage of ASH_LBA, we repeated this analysis, but this time, we modeled ASH_IA1 as a three-way mixture of each of the candidate populations, Levant_ChL and Iran_ChL. The two latter populations have higher genome coverage and can model ASH_LBA well in combination (table S3). In this analysis, only the models including "Sardinian," "Crete_Odigitria_BA," or "Iberia_BA" as the candidate population provided a good fit ($\chi^2P = 0.715$, $49.3 \pm 8.5\%$; $\chi^2P = 0.972$, $38.0 \pm 22.0\%$; and $\chi^2P = 0.964$, $25.8 \pm 9.3\%$, respectively). We note that, because of geographical and temporal sampling gaps, populations that potentially contributed the "European-related" admixture in ASH_IA1 could be missing from the dataset. Therefore, better proxies might be found in the future when more data is available. Nonetheless, the tested candidate populations from Anatolia, Egypt, and the Levant that did not produce well-fitting models can be excluded as potential sources of the admixture observed in ASH_IA1.

The transient impact of the "European-related" gene flow on the Ashkelon gene pool

The ASH_IA2 individuals are intermediate along PC1 between the ASH_LBA ones and the earlier Bronze Age Levantines (Jordan_EBA/Lebanon_MBA) in the west Eurasian PCA (Fig. 2A). Notably, despite being chronologically closer to ASH_IA1, the ASH_IA2 individuals position closer, on average, to the earlier Bronze Age individuals. Such a reduced affinity to the European populations is also apparent in the genetic clustering results, showing that the European hunter-gatherer-related ancestry contributes considerably less to ASH_IA2 than to ASH_IA1 (8 and 14%, respectively; Fig. 2B). To

test for differences in affinities between ASH_IA2 and ASH_IA1, we measured the f_4 -statistic of the form f_4 (ASH_IA2, ASH_IA1; test, Mbuti). This statistic produced no significantly positive results, and all significantly negative statistics ($Z \leq -3$) are with test populations from either Europe or Anatolia (fig. S10). This, again, highlights the increased European-related affinity of ASH_IA1.

The transient excess of European-related genetic affinity in ASH_IA1 can be explained by two scenarios. The early Iron Age European-related genetic component could have been diluted by either the local Ashkelon population to the undetectable level at the time of the later Iron Age individuals or by a gene flow from a population outside of Ashkelon introduced during the final stages of the early Iron Age or the beginning of the later Iron Age. Considering the symmetry measured between ASH_IA2 and ASH_LBA in the f_4 -statistic of the form f_4 (ASH_IA2, ASH_LBA; test, Mbuti) ($|Z| < 2.8$; fig. S11), the replacing population is likely to have stemmed from the Late Bronze Age Levantine gene pool, whether or not from Ashkelon. By modeling ASH_IA2 as a mixture of ASH_IA1 and earlier Bronze Age Levantines/Late Period Egyptian, we infer a range of 7 to 38% of contribution from ASH_IA1, although no contribution cannot be rejected because of the limited resolution to differentiate between Bronze Age and early Iron Age ancestries in this model (table S8).

DISCUSSION

By investigating genome-wide data from Ashkelon, we address long-pending historical questions regarding the demographic developments underlying the late Bronze Age to Iron Age cultural transformation. On a larger regional scale, these data begin to fill a temporal gap in the genetic map of the southern Levant, revealing persistence of the local Levantine gene pool throughout the Bronze Age for over a millennium. At the same time, by the "zoomed-in" comparative analysis of the Ashkelon genetic time transect, we find that the unique cultural features in the early Iron Age are mirrored by the distinct genetic composition we detect in ASH_IA1. Our analysis suggests that this genetic distinction is due to a European-related gene flow introduced in Ashkelon during either the end of the Bronze Age or the beginning of the Iron Age. This timing is in accord with estimates of the Philistines arrival to the coast of the Levant, based on archaeological and textual records (2–4). We find that, within no more than two centuries, this genetic footprint introduced during the early Iron Age is no longer detectable and seems to be diluted by a local Levantine-related gene pool.

The relatively rapid disappearance of this signal stresses the value of temporally dense genetic sampling for addressing historical questions. Transient gene flows, such as the one detected here, might be overlooked because of a lack of representative samples, potentially leading to erroneous conclusions. In geographic regions unfavorable to DNA preservation, obtaining such datasets requires exhaustive sampling and the utilization and further development of advanced technologies such as DNA enrichment techniques (15–17) and targeted sampling strategies (27).

We do not rule out that some gene flow occurred during the Bronze Age as low significance of the f_4 -statistics might be due to the limited statistical power of our data stemming from either insufficient coverage or a lack of appropriate contemporaneous proxy populations. Thus, additional sampling is needed to further investigate the question of the genetic diversity within the Levantine Bronze Age populations and to characterize the spatiotemporal extent of potential

MS no:	Emp name:	Date / Time:	PE's:	AA's:	Comments:	Art no:
RAaax0061/AF/ANTHROPOLOGY	kmatro	6-3-2019 / 16:51				

incoming gene flows. Similarly, a larger sample size might help to accurately infer the extent and magnitude of the early Iron Age gene flows and to identify more precisely the populations introducing the European-related component to Ashkelon. While our modeling suggests a southern European gene pool as a plausible source, future sampling in regions such as Cyprus, Sardinia, and the Aegean, as well as in the southern Levant, could better resolve this question.

AQ15 **MATERIALS AND METHODS**
Extraction of aDNA and preparation for next-generation sequencing

One hundred eight human skeletal elements excavated at the archaeological site of Ashkelon, Israel by the Leon Levy Expedition during 1997 to 2016 were sampled and screened for the presence of human aDNA (28 from the Bronze Age, 8 from the Iron Age I, and 72 from the Iron Age II). All preamplification procedures were performed in the dedicated aDNA facilities of the Max Planck Institute for the Science of Human History (MPI-SHH), Jena, Germany. Sampling of the petrous parts of temporal bones was done by drilling the inner ear part (27). Teeth were sampled by drilling the dental pulp. DNA was extracted from around 100 mg of pulverized bone following an established protocol (28). A double-stranded and double-indexed Illumina DNA library was prepared from a 20- μ l aliquot of each extract using published protocols (29, 30). To reduce potential errors in the obtained DNA sequences, we partially removed DNA damage resulting from cytosine deamination using uracil-DNA glycosylase and endonuclease VIII as previously reported (31). In this procedure, the damage is retained in the terminal positions of the DNA fragments and can be leveraged as a measure of aDNA authentication. Amplification of the indexed libraries was carried out using Herculase II Fusion DNA polymerase as in the manufacturer's protocol. All libraries were directly shotgun single-end sequenced on an Illumina HiSeq 4000 platform (1 \times 75 + 8 + 8 cycles).

DNA enrichment

A subset of libraries that exceeded 0.2% human DNA and 10% deamination damage at the terminal end of DNA fragments were subsequently used for two previously published hybridization-based in-solution DNA-enrichment assays: (i) The 1240 k capture (15–17, 32), which targets 1,237,207 genome-wide nuclear SNPs. The targeted SNP panel combines the sets reported by Haak *et al.* (17) and by Fu *et al.* (16) and is further described by Mathieson *et al.* (15); and (ii) The “Mitochondrial capture,” which targets the whole human mitochondrial genome (16). Both captures were carried out as described in the SI section of Fu *et al.* (16) with modified hybridization conditions of 65°C for about 24 hours. Enriched libraries were single-end sequenced on the same platform as the initial shotgun ones.

Postsequencing data processing

The binning of the sequenced reads (demultiplexing) allowed a maximum of one mismatch in each index. The demultiplexed libraries were processed and mapped using the EAGER (v 1.92.54) pipeline (33). The adapter sequences were clipped, and reads were filtered for the ones longer than 30 base pairs using AdapterRemoval (v2.2.0) (34). Adapter-clipped reads were aligned to the University of California, Santa Cruz genome browser's human genome reference hg19 using the BWA aln/samse programs (v0.7.12) (35) with a lenient stringency parameter (“-n 0.01”) and by retaining only reads with Phred-scaled

mapping quality scores ≥ 30 . Duplicate reads were removed with DeDup v0.12.2 (33). To eliminate genotyping bases with retained damage, the two terminal positions in each read were clipped, and subsequently, a pseudo-diploid genotype was reconstructed for each individual using pileupCaller by performing a random choice between high-quality bases (Phred-scaled base quality score ≥ 30) aligning to each targeted SNP position (<https://github.com/stschiff/sequenceTools>).

Dataset

The newly generated ancient data were merged with a previously described dataset (12) and with additional published datasets (11, 13–14, 21, 25, 36). The merged dataset includes 638 published ancient genomes (11–14, 21, 25, 36) and 4943 ones belonging to 298 worldwide present-day populations (12, 37) that were genotyped on the Affymetrix Axiom Genome-Wide Human Origins 1 array (38) (“HO dataset”) with a total of 593,124 SNP sites in the merged dataset.

aDNA authentication and quality control

We used multiple measures to authenticate the generated aDNA data. (i) To control for potential laboratory contamination, blank extractions and library preparations were included and analyzed for each sample batch (data file S3). (ii) Levels of DNA deamination damage in the mapped reads were estimated using mapDamage (v2.0) (39) and compared to the expected values in ancient skeletal element. (iii) We measured human mitochondrial DNA contamination using schmutzi (40). (iv) The genetic sex was inferred for each individual by calculating the ratio of the X and Y chromosome coverage (normalized by the autosomal average coverage). We next used ANGSD (v0.910) (41), which compares mismatch rates between polymorphic sites on the X chromosome to estimate nuclear contamination in males.

To avoid bias associated with including related individuals into analyzed populations, we calculated the pairwise allele mismatch rates of all newly reported individuals following a method described elsewhere (fig. S1) (42).

Principal components analysis

We constructed the principal components of 55 present-day west Eurasian groups and projected the ancient individuals onto the first two components (fig. S2) using the smartpca software from the EIGENSOFT package (v6.0.1) (43) with the lsqproject option.

ADMIXTURE analysis

We carried out a maximum likelihood unsupervised clustering of 5581 ancient and present-day individuals using ADMIXTURE (v1.3.0) (23) with a cluster number (k) ranging between 2 and 20. We used the --indep-pairwise option in PLINK (v1.90) (44, 45) to prune for linkage disequilibrium by identifying SNP pairs with genotype $r^2 \geq 0.2$ within a 200-SNP sliding window (advancing by 25 SNPs each time) and retaining only one randomly chosen SNP (--indep-pairwise 200 25 0.2). We performed five replicates for each k with random seeds and chose the highest likelihood replicate (fig. S3). Fivefold cross-validation (CV) errors were calculated for each run.

f-statistics

To test for gene flow between populations, we estimated their respective allele frequency correlations using f -statistics. Both f_4 - and f_3 -statistics were computed using the qpDstat program (v701) of the ADMIXTOOLS package (v4.1) (24) with default parameters.

MS no:	Emp name:	Date / Time:	PE's:	AA's:	Comments:	Art no:
RAaax0061/AF/ANTHROPOLOGY	kmatro	6-3-2019 / 16:51				

AQ25 To test the symmetry between populations X and Y, the f_4 -statistic
of the form $f_4(X \text{ and } Y; \text{ test, outgroup})$ was used. We report the
AQ26 f -statistics in which at least 20,000 SNP positions were overlapping
between the four tested populations.

Modeling ancestry proportions

We used the qpWave (v400) and qpAdm (v632) programs of ADMIXTOOLS (12, 17) to test and model admixture proportions from potential source populations (reference populations) in the tested populations. We used a basic set of seven outgroups including present-day populations (Han, Onge, Mbuti, Mala, Mixe) (37) that represent a global genetic variation and published ancient populations such as Natufian (12), which represents a Levantine gene pool outside of modern genetic variation and the European Upper Palaeolithic individual “Villabruna” (36).

The distribution of the squared f_4 -statistics of the form f_4 (Ashkelon individual 1, Ashkelon individual 2; test, Mbuti) was estimated for χ^2 distribution with 1 df using Q-Q plot for each analyzed group (ASH_LBA, ASH_IA1, and ASH_IA2) (fig. S8). The variance for each group was estimated by a Fligner-Killeen test, which is robust against departures from normality that can be caused by small sample size.

Mitochondrial DNA analysis

AQ27 We constructed the whole mitochondrial genome consensus sequence for each individual using the log2fasta program of schmutzi (40) with a quality cutoff of 10. The consensus sequences were inputted to HaploFind (46) and HaploGrep (47) for mitochondrial haplogroup assignment.

Y-chromosome analysis

AQ28 We assigned the Y-chromosome haplogroup for each male individual using yHaplo (48). The genotyping was carried out by a random draw of a single-base mapping to one of the 13,508 ISOGG consortium SNP positions. Strand-ambiguous SNPs were discarded, and quality filters were as those for the HO dataset. In addition, the automatically called derived alleles, which support the haplogroup assignment, were manually confirmed.

Phenotypic traits analyses

AQ29 We inspected the allele distribution in SNP positions associated with selected biological traits, including lactase persistence (49, 50), Malaria resistance (51, 52), glucose-6-phosphate dehydrogenase deficiency (53, 54), and skin pigmentation (26, 55–56). The allele distribution for the SNP positions listed in table S9 was tabulated for each individual using samtools mpileup (v1.3).

Carbon dating

AQ30 The petrous parts of the temporal bone of individuals ASH029, ASH033, ASH2-3, ASH066, ASH067, ASH068, and ASH008 were each sampled and directly radiocarbon-dated at the Curt-Engelhorn-Zentrum Archaeometry gGmbH, Mannheim, Germany (table S1). Collagen was extracted from the bone samples, purified by ultrafiltration (fraction, >30 kDa), freeze-dried, and combusted to CO₂ in an elemental analyzer. CO₂ was converted catalytically to graphite. The dating was performed using the MICADAS-AMS of the Klaus-Tschira-Archäometrie-Zentrum. The resulting ¹⁴C ages were normalized to $\delta^{13}\text{C} = -25\text{‰}$ (57) and calibrated using the dataset INTCAL13 (58) and the software SwissCal 1.0 (59).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/5/7/eaax0061/DC1>

Text S1. Description of individuals excavated in Ashkelon and archeological information Fig. S1. Pairwise mismatch rate.

Fig. S2. PCA plot of present-day west Eurasian populations.

Fig. S3. ADMIXTURE CV errors for each cluster number (K).

Fig. S4. The Bronze Age Ashkelon population shares a higher genetic affinity with Caucasus/Iranian-related populations when compared to the Neolithic Levantines.

Fig. S5. The Bronze Age Ashkelon population shares a marginal higher genetic affinity with populations related to ancient Caucasus/Iran when compared to Chalcolithic Levantines.

Fig. S6. The Bronze Age Ashkelon population is symmetrically related to Lebanon_MBA.

Fig. S7. The Bronze Age Ashkelon population is mostly symmetrically related to Jordan_EBA but shares a slightly higher genetic affinity with populations related to ancient Caucasus/Iran.

Fig. S8. Variation in differential affinities between the Ashkelon groups.

Fig. S9. The proportion of Mesolithic European WHG ancestry in ASH_IA1 individuals correlates with their position in the west Eurasian PCA.

Fig. S10. ASH_IA1 shares more alleles with European-related populations compared to ASH_IA2.

Fig. S11. ASH_IA2 is symmetrically related to ASH_LBA.

Fig. S12. Map of the Bronze Age necropolis.

Fig. S13. Relative stratigraphy of grid 38.

Fig. S14. Relative stratigraphy of NS. Square 14.

Fig. S15. Relative stratigraphy of NS. Square 24.

Table S1. ¹⁴C radiocarbon dating performed for this study.

Table S2. Nuclear and mitochondrial contamination estimates.

Table S3. qpAdm admixture models of the ASH_LBA group.

Table S4. Fligner-Killeen test based on $[f_4(\text{Ashkelon ind 1, Ashkelon ind 2; test, Mbuti})]^2$ statistics.

Table S5. Ashkelon individuals and groups modeled using qpAdm by “Levant_ChL,” “Iran_ChL,” and WHG as sources.

Table S6. Bronze Age and Iron Age individuals and groups modeled by the three maximized populations on the west Eurasian PCA (using qpAdm).

Table S7. qpAdm admixture models of the ASH_IA1 group.

Table S8. qpAdm admixture models of the ASH_IA2 group.

Table S9. Phenotypic analysis.

Table S10. Archeological context.

Data file S1. An overview of skeletal material screened for aDNA in this study.

Data file S2. An overview of the main analysis groups used in this study.

Data file S3. Sequencing statistics of negative controls.

References (28–66)

REFERENCES AND NOTES

1. E. H. Cline, 1177 BC: The Year Civilization Collapsed (Princeton Univ. Press, 2015).
2. A. Yasur-Landau, The Philistines and Aegean Migration at the End of the Late Bronze Age (Cambridge Univ. Press, 2014).
3. L. E. Stager, The impact of the Sea Peoples in Canaan (1185–1050 BCE), in The Archaeology of Society in the Holy Land, T. E. Levy, Ed. (Leicester Univ. Press, 1995), pp. 332–348.
4. A. E. Killebrew, Biblical peoples and ethnicity, in An Archaeological Study of Canaanites, Egyptians, Philistines, and Early Israel, 1300–1100 BCE. (Society of Biblical Literature, Atlanta, 2005), p. 9.
5. R. Drews, Medinet Habu: Oxcarts, ships, and migration theories. J. Near East. Stud. 59, 161–190 (2000).
6. E. S. Sherratt, Immigration and archaeology: Some indirect reflections, in Acta Cypria. Acts of an International Congress on Cypriot Archaeology, Göteborg om 22–24 August 1991, P. Åström, Ed. (Jonsered, 1992), vol. 2, pp. 316–347.
7. G. D. Middleton, Telling stories: The Mycenaean origins of the Philistines. Oxford J. Archaeol. 34, 45–65 (2015).
8. S. Sherratt, Sea Peoples and the economic structure of the late second millennium in the Eastern Mediterranean, in Mediterranean Peoples in Transition: Thirteenth to Early Tenth Centuries BCE, S. Gitin, A. Mazar, E. Stern, Eds. (Israel Exploration Society, 1998), pp. 292–313.
9. A. M. Maeir, L. A. Hitchcock, The appearance, formation and transformation of Philistine culture: New perspectives and new finds, in The Sea Peoples Up-to-Date: New Research on the Migration of Peoples in the 12th Century BCE, P. Fischer, T. Bürge, Eds. (Austrian Academy of Sciences, 2017), pp. 149–162.
10. R. Jung, Push and pull factors of the Sea Peoples between Italy and the Levant, in The Archaeology of Forced Migration (Presses universitaires de Louvain, 2018), p. 273.
11. E. Harnay, H. May, D. Shalem, N. Rohland, S. Mallick, I. Lazaridis, R. Sarig, K. Stewardson, S. Nordenfelt, N. Patterson, I. Hershkovitz, D. Reich, Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation. Nat. Commun. 9, 3336 (2018).

MS no:	Emp name:	Date / Time:	PE's:	AA's:	Comments:	Art no:
RAaax0061/AF/ANTHROPOLOGY	kmatro	6-3-2019 / 16:51				

12. I. Lazaridis, D. Nadel, G. Rollefson, D. C. Merrett, N. Rohland, S. Mallick, D. Fernandes, M. Novak, B. Gammar, K. Sirak, S. Connell, K. Stewardson, E. Harney, Q. Fu, G. Gonzalez-Fortes, E. R. Jones, S. A. Roodenberg, G. Lengyel, F. Bocquentin, B. Gasparian, J. M. Monge, M. Gregg, V. Eshed, A.-S. Mizrahi, C. Meiklejohn, F. Gerritsen, L. Bejani, M. Blüher, A. Campbell, G. Cavalleri, D. Comas, P. Froguel, E. Gilbert, S. M. Kerr, P. Kovacs, J. Krause, D. McGettigan, M. Merrigan, D. A. Merriwether, S. O'Reilly, M. B. Richards, O. Semino, M. Shamoon-Pour, G. Stefanesco, M. Stumvoll, A. Tönjes, A. Torrioni, J. F. Wilson, L. Yengo, N. A. Hovhannisyann, N. Patterson, R. Pinhasi, D. Reich, Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016).
13. M. Haber, C. Doumet-Serhal, C. Scheib, Y. Xue, P. Danecek, M. Mezzavilla, S. Youhanna, R. Martiniano, J. Prado-Martinez, M. Szpak, E. Matisoo-Smith, H. Schutkowski, R. Mikulski, P. Zalloua, T. Kivisild, C. Tyler-Smith, Continuity and admixture in the last five millennia of Levantine history from ancient Canaanite and present-day Lebanese genome sequences. *Am. J. Hum. Genet.* **101**, 274–282 (2017).
14. M. Feldman, E. Fernández-Domínguez, L. Reynolds, D. Baird, J. Pearson, I. Hershkovitz, H. May, N. Goring-Morris, M. Benz, J. Gresky, R. A. Bianco, A. Fairbairn, G. Mustafaoğlu, P. W. Stockhammer, C. Posth, W. Haak, C. Jeong, J. Krause, Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia. *Nat. Commun.* **10**, 1218 (2019).
15. I. Mathieson, I. Lazaridis, N. Rohland, S. Mallick, N. Patterson, S. A. Roodenberg, E. Harney, K. Stewardson, D. Fernandes, M. Novak, K. Sirak, C. Gamba, E. R. Jones, B. Llamas, S. Dryomov, J. Pickrell, J. L. Arsuaga, J. M. B. de Castro, E. Carbonell, F. Gerritsen, A. Khokhlov, P. Kuznetsov, M. Lozano, H. Meller, O. Mochalov, V. Moiseyev, M. A. R. Guerra, J. Roodenberg, J. M. Vergés, J. Krause, A. Cooper, K. W. Alt, D. Brown, D. Anthony, C. Lalueza-Fox, W. Haak, R. Pinhasi, D. Reich, Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
16. Q. Fu, M. Meyer, X. Gao, U. Stenzel, H. A. Burbano, J. Kelso, S. Pääbo, DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2223–2227 (2013).
17. W. Haak, I. Lazaridis, N. Patterson, N. Rohland, S. Mallick, B. Llamas, G. Brandt, S. Nordenfeldt, E. Harney, K. Stewardson, Q. Fu, A. Mittnik, E. Bánffy, C. Economou, M. Francken, S. Friederich, R. G. Pena, F. Hallgren, V. Khartanovich, A. Khokhlov, M. Kunst, P. Kuznetsov, H. Meller, O. Mochalov, V. Moiseyev, N. Nicklisch, S. L. Pichler, R. Risch, M. A. Rojo Guerra, C. Roth, A. Szécsényi-Nagy, J. Wahl, M. Meyer, J. Krause, D. Brown, D. Anthony, A. Cooper, K. W. Alt, D. Reich, Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
18. L. E. Stager, J. D. Schloen, D. M. Master, *Ashkelon 1: Introduction and Overview (1985–2006)* (Eisenbrauns, 2008).
19. D. M. Master, L. E. Stager, A. Yasur-Landau, Chronological observations at the dawn of the Iron Age in Ashkelon. *Agypte Und Levante/Egypt and the Levant* **21**, 261–280 (2011).
20. D. M. Master, A. J. Aja, The Philistine cemetery of Ashkelon. *Bull. Am. Schools Orient. Res.* **377**, 135–159 (2017).
21. P. de Barros Damgaard, R. Martiniano, J. Kamm, J. V. Moreno-Mayar, G. Kroonen, M. Peyrot, G. Barjamovic, S. Rasmussen, C. Zacho, N. Baimukhanov, V. Zaubert, V. Merz, A. Biddanda, I. Merz, V. Loman, V. Evdokimov, E. Usmanova, B. Hemphill, A. Seguin-Orlando, F. E. Yediyah, I. Ullah, K.-G. Sjögren, K. H. Iversen, J. Choin, C. de la Fuente, M. Ilardo, H. Schroeder, V. Moiseyev, A. Gromov, A. Polyakov, S. Omura, S. Y. Senyurt, H. Ahmad, C. McKenzie, A. Margaryan, A. Hameed, A. Samad, N. Gul, M. H. Khokhar, O. I. Goriunova, V. I. Bazaliiskii, J. Novembre, A. W. Weber, L. Orlando, M. E. Allentoft, R. Nielsen, K. Kristiansen, M. Sikora, A. K. Outram, R. Durbin, E. Willerslev, The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* **360**, eaar7711 (2018).
22. E. R. Jones, G. Gonzalez-Fortes, S. Connell, V. Siska, A. Eriksson, R. Martiniano, R. L. McLaughlin, M. Gallego Llorente, L. M. Cassidy, C. Gamba, T. Meshveliani, O. Bar-Yosef, W. Müller, A. Belfer-Cohen, Z. Matskevich, N. Jakeli, T. F. G. Higham, M. Currat, D. Lordkipanidze, M. Hofreiter, A. Manica, R. Pinhasi, D. G. Bradley, Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015).
23. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
24. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, D. Reich, Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
25. I. Lazaridis, A. Mittnik, N. Patterson, S. Mallick, N. Rohland, S. Pfrengle, A. Furtwängler, A. Peltzer, C. Posth, A. Vasilakis, P. J. P. McGeorge, E. Konsolaki-Yannopoulou, G. Korres, H. Martlew, M. Michalodimitrakaki, M. Özsaat, N. Özsaat, A. Papathanasiou, M. Richards, S. A. Roodenberg, Y. Tzedakis, R. Arnold, D. M. Fernandes, J. R. Hughey, D. M. Lotakis, P. A. Navas, Y. Maniatis, J. A. Stamatoyannopoulos, K. Stewardson, P. Stockhammer, R. Pinhasi, D. Reich, J. Krause, G. Stamatoyannopoulos, Genetic origins of the Minoans and Mycenaeans. *Nature* **548**, 214–218 (2017).
26. I. Olalde, M. E. Allentoft, F. Sánchez-Quintó, G. Santpere, C. W. K. Chiang, M. DeGiorgio, J. Prado-Martinez, J. A. Rodríguez, S. Rasmussen, J. Quilez, O. Ramirez, U. M. Marigorta, M. Fernández-Callejo, M. E. Prada, J. M. V. Encinas, R. Nielsen, M. G. Netea, J. Novembre, R. A. Sturm, P. Sabeti, T. Marqués-Bonet, A. Navarro, E. Willerslev, C. Lalueza-Fox, Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225–228 (2014).
27. R. Pinhasi, D. Fernandes, K. Sirak, M. Novak, S. Connell, S. Alpaslan-Roodenberg, F. Gerritsen, V. Moiseyev, A. Gromov, P. Raczyk, A. Anders, M. Pietruszewski, G. Rollefson, M. Jovanovic, H. Trinhhoang, G. Bar-Oz, M. Oxenham, H. Matsumura, M. Hofreiter, Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLOS ONE* **10**, e0129102 (2015).
28. J. Dabney, M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. Garcia, S. Paabo, J.-L. Arsuaga, M. Meyer, Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15758–15763 (2013).
29. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, pdb.prot5448 (2010).
30. M. Kircher, S. Sawyer, M. Meyer, Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
31. N. Rohland, E. Harney, S. Mallick, S. Nordenfeldt, D. Reich, Partial uracil–DNA–glycosylase screening for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130624 (2015).
32. Q. Fu, M. Hajdinjak, O. T. Moldovan, S. Constantin, S. Mallick, P. Skoglund, N. Patterson, N. Rohland, I. Lazaridis, B. Nickel, B. Viola, K. Prüfer, M. Meyer, J. Kelso, D. Reich, S. Pääbo, An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
33. A. Peltzer, G. Jäger, A. Herbig, A. Seitz, C. Knip, J. Krause, K. Nieselt, EAGER: Efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).
34. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
35. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
36. Q. Fu, C. Posth, M. Hajdinjak, M. Petr, S. Mallick, D. Fernandes, A. Furtwängler, W. Haak, M. Meyer, A. Mittnik, B. Nickel, A. Peltzer, N. Rohland, V. Slon, S. Talamo, I. Lazaridis, M. Lipson, I. Mathieson, S. Schiffels, P. Skoglund, A. P. Dereviak, N. Drozdov, V. Slavinsky, A. Tsybankov, R. G. Cremonesi, F. Mallegni, B. Gély, E. Vacca, M. R. G. Morales, L. G. Straus, C. Neugebauer-Maresch, M. Teschler-Nicola, S. Constantin, O. T. Moldovan, S. Benazzi, M. Peresani, D. Coppola, M. Lari, S. Ricci, A. Ronchitelli, F. Valentin, C. Thevenet, K. Wehrberger, D. Grigorescu, H. Rougier, I. Crevecoeur, D. Flis, P. Semal, M. A. Mannino, C. Cupillard, H. Bocherens, N. J. Conard, K. Harvati, V. Moiseyev, D. G. Drucker, J. Svoboda, M. P. Richards, D. Caramelli, R. Pinhasi, J. Kelso, N. Patterson, J. Krause, S. Pääbo, D. Reich, The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
37. S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, M. Zhao, N. Chennagiri, S. Nordenfeldt, A. Tandon, P. Skoglund, I. Lazaridis, S. Sankaraman, Q. Fu, N. Rohland, G. Renaud, Y. Erlich, T. Willems, C. Gallo, J. P. Spence, Y. S. Song, G. Poletti, F. Balloux, G. van Driem, P. de Knijff, I. G. Romero, A. R. Jha, D. M. Behar, C. M. Bravi, C. Capelli, T. Hervig, A. Moreno-Estrada, O. L. Posukh, E. Balanovska, O. Balanovsky, S. Karachanak-Yankova, H. Sahakyan, D. Toncheva, L. Yepiskoposyan, C. Tyler-Smith, Y. Xue, M. S. Abdullah, A. Ruiz-Linares, C. M. Beall, A. Di Rienzo, C. Jeong, E. B. Starikovskaya, E. Metspalu, J. Parik, R. Villesms, B. M. Henn, U. Hodoglugil, R. Mahley, A. Sajantila, G. Stamatoyannopoulos, J. T. S. Wee, R. Khushainova, E. Khushnutdinova, S. Litvinov, G. Ayodo, D. Comas, M. F. Hammer, T. Kivisild, W. Klitz, C. A. Winkler, D. Labuda, M. Bamshad, L. B. Jorde, S. A. Tishkoff, W. S. Watkins, M. Metspalu, S. Dryomov, R. Sukernik, L. Singh, K. Thangaraj, S. Pääbo, J. Kelso, N. Patterson, D. Reich, The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
38. I. Lazaridis, N. Patterson, A. Mittnik, G. Renaud, S. Mallick, K. Kirsanow, P. H. Sudmant, J. G. Schraiber, S. Castellano, M. Lipson, B. Berger, C. Economou, R. Bollongino, Q. Fu, K. I. Bos, S. Nordenfeldt, H. Li, C. de Filippo, K. Prüfer, S. Sawyer, C. Posth, W. Haak, F. Hallgren, E. Fomander, N. Rohland, D. Delsate, M. Francken, J.-M. Guinet, J. Wahl, G. Ayodo, H. A. Babiker, G. Bailliet, E. Balanovska, O. Balanovsky, R. Barrantes, G. Bedoya, H. Ben-Ami, J. Bene, F. Berrada, C. M. Bravi, F. Brisighelli, G. B. J. Busby, F. Cali, M. Churnosov, D. E. C. Cole, D. Corach, L. Damba, G. van Driem, S. Dryomov, J.-M. Dugoujon, S. A. Fedorova, I. Gallego Romero, M. Gubina, M. Hammer, B. M. Henn, T. Hervig, U. Hodoglugil, A. R. Jha, S. Karachanak-Yankova, R. Khushainova, E. Khushnutdinova, R. Kittles, T. Kivisild, W. Klitz, V. Kučinskas, A. Kushniarevich, L. Laredj, S. Litvinov, T. Loukidis, R. W. Mahley, B. Melegh, E. Metspalu, J. Molina, J. Mountain, K. Näkkäläjärvi, D. Nesheva, T. Nyambo, L. Osipova, J. Parik, F. Platonov, O. Posukh, V. Romano, F. Rothhammer, I. Rudan, R. Ruizbakiev, H. Sahakyan, A. Sajantila, A. Salas, E. B. Starikovskaya, A. Tarekne, D. Toncheva, S. Turdikulova, I. Uktvertye, O. Utevska, R. Vasquez, M. Villena, M. Voevodina, C. A. Winkler, L. Yepiskoposyan, P. Zalloua, T. Zemunik, A. Cooper, C. Capelli, M. G. Thomas, A. Ruiz-Linares, S. A. Tishkoff, L. Singh, K. Thangaraj, R. Villesms, D. Comas, R. Sukernik, M. Metspalu, M. Meyer, E. E. Eichler, J. Burger, M. Slatkin, S. Pääbo, J. Kelso, D. Reich, J. Krause, Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).

AQ41

AQ42

MS no:	Emp name:	Date / Time:	PE's:	AA's:	Comments:	Art no:
RAaax0061/AF/ANTHROPOLOGY	kmatro	6-3-2019 / 16:51				

39. H. Jönsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).

40. G. Renaud, V. Slon, A. T. Duggan, J. Kelso, Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015).

41. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).

42. D. J. Kennett, S. Plog, R. J. George, B. J. Culleton, A. S. Watson, P. Skoglund, N. Rohland, S. Mallick, K. Stewardson, L. Kistler, S. A. LeBlanc, P. M. Whiteley, D. Reich, G. H. Perry, Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat. Commun.* **8**, 14115 (2017).

43. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLOS Genet.* **2**, e190 (2006).

44. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

45. C. C. Chang, C. C. Chow, L. C. A. M. Tellier, S. Vattikuti, J. J. Lee, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

46. D. Vianello, F. Sevinci, G. Castellani, L. Lomartire, M. Capri, C. Franceschi, HAPLOFIND: A new method for high-throughput mtDNA haplogroup assignment. *Hum. Mutat.* **34**, 1189–1194 (2013).

47. A. Kloss-Brandstätter, D. Pacher, S. Schönherr, H. Weissensteiner, R. Binna, G. Specht, F. Kronenberg, HaploGrep: A fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat.* **32**, 25–32 (2011).

48. G. D. Poznik, Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. *bioRxiv* 088716 [Preprint] (19 November 2016).

49. C. J. E. Ingram, M. F. Elamin, C. A. Mulcare, M. E. Weale, A. Tarekegn, T. O. Raga, E. Bekele, F. M. Elamin, M. G. Thomas, N. Bradman, D. M. Swallow, A novel polymorphism associated with lactose tolerance in Africa: Multiple causes for lactase persistence? *Hum. Genet.* **120**, 779–788 (2007).

50. S. A. Tishkoff, F. A. Reed, A. Ranciaro, B. F. Voight, C. C. Babbitt, J. S. Silverman, K. Powell, H. M. Mortensen, J. B. Hirbo, M. Osman, M. Ibrahim, S. A. Omar, G. Lema, T. B. Nyambo, J. Ghoris, S. Bumpstead, J. K. Pritchard, G. A. Wray, P. Deloukas, Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* **39**, 31–40 (2007).

51. Malaria Genomic Epidemiology Network, Reappraisal of known malaria resistance loci in a large multicenter study. *Nat. Genet.* **46**, 1197–1204 (2014).

52. C. Timmann, T. Thye, M. Vens, J. Evans, J. May, C. Ehmen, J. Sievertsen, B. Muntau, G. Ruge, W. Loag, D. Ansong, S. Antwi, E. Asafo-Adjei, S. B. Nguah, K. O. Kwakye, A. O. Y. Akoto, J. Sylverken, M. Brendel, K. Schuldt, C. Loley, A. Franke, C. G. Meyer, T. Agbenyega, A. Ziegler, R. D. Horstmann, Genome-wide association study indicates two novel resistance loci for severe malaria. *Nature* **489**, 443–446 (2012).

53. N. Sepúlveda, A. Manjurano, S. G. Campino, M. Lemnge, J. Lusingu, R. Olomi, K. A. Rockett, C. Hubbard, A. Jeffreys, K. Rowlands, T. G. Clark, E. M. Riley, C. J. Drakeley; MalariaGEN Consortium, Malaria host candidate genes validated by association with current, recent, and historical measures of transmission intensity. *J. Infect. Dis.* **216**, 45–54 (2017).

54. B. Maiga, A. Dolo, S. Campino, P. Corran, K. A. Rockett, M. Troye-Blomberg, O. K. Doumbo, T. G. Clark, Glucose-6-phosphate dehydrogenase polymorphisms and susceptibility to mild malaria in Dogon and Fulani, Mali. *Malar. J.* **13**, 270 (2014).

55. H. L. Norton, R. A. Kittles, E. Parra, P. McKeigue, X. Mao, K. Cheng, V. A. Canfield, D. G. Bradley, B. McEvoy, M. D. Shriver, Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol. Biol. Evol.* **24**, 710–722 (2006).

56. S. Wilde, A. Timpson, K. Kirsanow, E. Kaiser, M. Kayser, M. Unterländer, N. Hoffelder, I. D. Potekhina, W. Schier, M. G. Thomas, J. Burger, Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4832–4837 (2014).

57. M. Stuiver, H. A. Polach, Discussion reporting of ¹⁴C data. *Radiocarbon* **19**, 355–363 (1977).

58. P. J. Reimer, E. Bard, A. Bayliss, J. W. Beck, P. G. Blackwell, C. B. Ramsey, C. E. Buck, H. Cheng, R. L. Edwards, M. Friedrich, P. M. Grootes, T. P. Guilderson, H. Hafflidason, I. Hajdas, C. Hatté, T. J. Heaton, D. L. Hoffmann, A. G. Hogg, K. A. Hughen, K. F. Kaiser,

B. Kromer, S. W. Manning, M. Niu, R. W. Reimer, D. A. Richards, E. M. Scott, J. R. Southon, R. A. Staff, C. S. M. Turney, J. van der Plicht, IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).

59. M. Christl, C. Vockenhuber, P. W. Kubik, L. Wacker, J. Lachner, V. Alfimov, H.-A. Synal, The ETH Zurich AMS facilities: Performance parameters and reference materials. *Nucl. Instrum. Methods Phys. Res. B* **294**, 29–38 (2013).

60. C. Berghoffen, Red lustrous wheelmade and coarse-ware spindle bottles from Ashkelon, in *Amilla: The Quest for Excellence Studies Presented to Guenter Kopcke in Celebration of His 75th Birthday*, (INSTAP Academic Press, 2013) pp. 281–292.

61. A. Cohen-Weinberger, "Petrography of Middle Bronze Age 2 pottery: Implications to understanding Egypto-Canaanite relations," thesis, Tel Aviv University, Tel Aviv, Israel (2008).

62. J. L. Baker, "The Middle and Late Bronze Age tomb complex at Ashkelon, Israel: The architecture and the funeral kit," thesis, Brown University, Providence, RI (2003).

63. D. M. Master, Iron I chronology at Ashkelon: Preliminary results of the Leon Levy Expedition, in *The Bible and Radiocarbon Dating: Archaeology, Text and Science* (Equinox Publishing, 2005) pp. 337–348.

64. K. Birney, B. R. Doak, Funerary iconography on an infant burial jar from Ashkelon. *Israel Explor. J.* **61**, 32–53 (2011).

65. N. Schreiber, *The Cypro-Phoenician Pottery of the Iron Age* (Brill, 2003).

66. P. Mountjoy, *Decorated Pottery in Cyprus and Philistia in the 12th century BC, Cypriot IIIC and Philistine IIIC* (Austrian Academy of Sciences Press, 2018).

Acknowledgments: We thank the late L.E. Stager who codirected the excavations in Ashkelon and has devoted much of his scientific career to understanding the origins of the Philistines. We thank M. McCormick for initial conversations with the late L.E. Stager that sparked this investigation. We thank G. Brandt, A. Wissgott, F. Aron, C. Freund, R. Stahl, and I. Kucukkalipci (MPI-SHH) for support in laboratory work. We thank A. Mötsch and S. Eisenmann for support in organization and sample management. We thank M. O'Reilly for graphic support. We thank M. Faerman, N. L. Shattah, P. Smith, S. Fox, K. Kalisher, and K. Marklein for the anthropological identification of skeletal remains used in this study. We thank the members of the population genetics group and the Max Planck–Harvard Research center for the Archaeoscience of the ancient Mediterranean (MHAAM) group in the Department of Archaeogenetics, MPI-SHH for their input and support. **Funding:** This work was supported by the Max Planck Society and the MHAAM. Excavations at Ashkelon were funded by grants from L. Levy, S. White, and the Leon Levy Foundation under license from the Israel Antiquity Authority and the Israel Nature and Parks Authority. **Author contributions:** J.K. and D.M.M. conceived the study. J.K. supervised the genetic work. D.M.M. and A.J.A. provided archeological material. D.M.M., P.W.S., and A.J.A. advised on the archeological background and interpretation. D.M.M. and A.J.A. wrote the archeological and sample background section. M.F., M.B., A.M., I.K., and R.A.B. performed the laboratory work. M.F. performed the data analyses, with C.J. providing guidance. M.F., D.M.M., C.J., and J.K. wrote the manuscript with input from all coauthors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and the Supplementary Materials. The genomic alignment data (BAM format) is available through the European Nucleotide Archive (ENA) under the accession number (Study PRJEB31035), and the Eigenstrat format 1240 K pulldown genotype data for ancient individuals newly reported in this study are available at the Edmond data repository of the Max Planck Society (<https://edmond.mpg.de/imeji/collection/NKBCLKpHR2hZQwb7q=>). The Ashkelon Iron Age skeletal material has been accessioned to the National Natural History Collections (NNHC) of the Hebrew University.

Submitted 13 February 2019
Accepted 23 May 2019
Published 3 July 2019
10.1126/sciadv.aax0061

Citation: M. Feldman, D. M. Master, R. A. Bianco, M. Burri, P. W. Stockhammer, A. Mittnik, A. J. Aja, C. Jeong, J. Krause, Ancient DNA sheds light on the genetic origins of early Iron Age Philistines. *Sci. Adv.* **5**, eaax0061 (2019).

MS no:	Emp name:	Date / Time:	PE's:	AA's:	Comments:	Art no:
RAaax0061/AF/ANTHROPOLOGY	kmatro	6-3-2019 / 16:51				

6. Manuscript C

A High-Coverage *Yersinia pestis* Genome from a Sixth-Century Justinianic Plague Victim

Michal Feldman,^{*,1,3} Michaela Harbeck^{*,2} Marcel Keller,^{1,2} Maria A. Spyrou,^{1,3} Andreas Rott,² Bernd Trautmann,² Holger C. Scholz,^{4,5} Bernd Pääffgen,⁶ Joris Peters,^{2,7} Michael McCormick,⁸ Kirsten Bos,^{1,3} Alexander Herbig,^{1,3} and Johannes Krause^{*,1,3}

¹Max Planck Institute for the Science of Human History, Jena, Germany

²SNSB, State Collection of Anthropology and Palaeoanatomy, Munich, Germany

³Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen, Germany

⁴Bundeswehr Institute of Microbiology, Munich, Germany

⁵German Center for Infection Research (DZIF), Munich, Germany

⁶Institute for Pre- and Protohistoric Archaeology and Archaeology of the Roman Provinces, Ludwig-Maximilian University Munich, Germany

⁷Institute of Palaeoanatomy, Domestication Research and the History of Veterinary Medicine, Ludwig-Maximilian University of Munich, Germany

⁸Department of History, Harvard University, Initiative for the Science of the Human Past

*Corresponding authors: E-mails: krause@shh.mpg.de; feldman@shh.mpg.de; michaela.harbeck@extern.lrz-muenchen.de.

Associate editor: Connie Mulligan

Abstract

The Justinianic Plague, which started in the sixth century and lasted to the mid eighth century, is thought to be the first of three historically documented plague pandemics causing massive casualties. Historical accounts and molecular data suggest the bacterium *Yersinia pestis* as its etiological agent. Here we present a new high-coverage (17.9-fold) *Y. pestis* genome obtained from a sixth-century skeleton recovered from a southern German burial site close to Munich. The reconstructed genome enabled the detection of 30 unique substitutions as well as structural differences that have not been previously described. We report indels affecting a *lacI* family transcription regulator gene as well as nonsynonymous substitutions in the *nrdE*, *fadJ*, and *pcp* genes, that have been suggested as plague virulence determinants or have been shown to be upregulated in different models of plague infection. In addition, we identify 19 false positive substitutions in a previously published lower-coverage *Y. pestis* genome from another archaeological site of the same time period and geographical region that is otherwise genetically identical to the high-coverage genome sequence reported here, suggesting low-genetic diversity of the plague during the sixth century in rural southern Germany.

Key words: *Yersinia pestis*, ancient DNA, justinianic plague, whole genome, reconstruction.

Introduction

The bacterium *Yersinia pestis* has been infecting humans for over 5,000 years (Rasmussen et al. 2015) and is thought responsible for at least three known historic plague pandemics. The first was the sixth- to AD eighth-century Justinianic pandemic, the second started with the infamous Black Death, claiming the lives of up to 50% of the European population during the 14th century (Benedictow 2004), and the last plague pandemic began in late 19th century China, seeding many of the plague foci that exist globally today (Pollitzer 1954; World Health Organization 2004). At present, plague is classified as a reemerging infectious disease in certain endemic regions and remains a public health problem with reservoirs on nearly every major continent (World Health Organization 2004).

Historical records suggest that the first known outbreak of the Justinianic Plague occurred between 541 and AD 543 in Egypt and spread throughout the eastern Roman Empire and its neighbors (Little 2007; Stathakopoulos 2004). Contemporary accounts indicate massive mortality caused by the disease that might have contributed to the weakening and the eventual decline of the eastern Roman Empire (Little 2007; Mitchell 2014). The epidemic itself returned in about 18 waves over a period of 200 years until it disappeared in Europe and the near East in the middle of the 8th century for yet unknown reasons (Stathakopoulos 2004).

Apparent discrepancies in epidemiological patterns between the modern and the historical pandemics have led scholars to suggest that etiological agents other than *Y. pestis* may have been responsible for the early and later medieval pandemics (Cohn 2008; Duncan and Scott 2005; Scott and

Duncan 2001; Twigg 1984). Molecular evidence obtained from ancient plague victims, however, has established *Y. pestis* as at least one of the causative agents for both historical pandemics (Bos et al. 2011; Haensch et al. 2010; Harbeck et al. 2013; Schuenemann et al. 2011; Wagner et al. 2014; Wiechmann and Grupe 2005). Nevertheless, the differences in epidemiology such as the apparently much faster geographical spread of the historical pandemics compared with the modern third pandemic (Christakos, et al. 2007; Cohn 2008; Kanaroglou and Delmelle 2015; Maddicott 1997) still need to be addressed. The geographic reach and mortality impact of individual waves as well as of the Justinianic pandemic as a whole remain unknown. Environmental and behavioral factors as well as genetic factors in the host, vector or pathogen have been known to alter the disease dynamics in modern plague outbreaks (Duplantier et al. 2005; Enscoe et al. 2002; Guiyoule et al. 1997; Keim and Wagner 2009; Parmenter et al. 1999; Schmid et al. 2015; Xu et al. 2014). The characterization of historical *Y. pestis* strains and a comparison to extant strains may well shed light on the role of the evolving bacterial genetic structure in forming these notable epidemiological differences. Moreover, a robust genomic description from the early stages of the Justinianic pandemic affords opportunities to trace and understand the evolution of one of humanity's most devastating pathogens over a period of deep time, with insights that may illuminate the evolutionary trajectories of other like organisms.

Despite their genetic resemblance *Y. pestis* and its closest relative the enteric soil- and water-borne *Yersinia pseudotuberculosis* differ greatly in pathogenicity and transmission (Achtman et al. 1999). Furthermore, the *Y. pestis* genome is characterized by structural variation caused by frequent intra-genomic rearrangements due to abundant insertion sequences, high expansion through horizontal gene transfer from other bacteria and bacteriophages and considerable gene decay that is evident from the large number of pseudogenes in the genome (Guiyoule et al. 1994; Parkhill et al. 2001; Zinser et al. 2003). Comparing the genome structure of historical *Y. pestis* strains to those of extant strains and to that of *Y. pseudotuberculosis* could offer insights into key evolutionary changes, such as those in virulence associated genes, through time.

A genome reconstructed from two sixth-century Justinianic Plague victims unearthed in Aschheim, southern Germany, estimated at 7.6-fold coverage, has recently been reported and used to characterize the *Y. pestis* strain responsible for the Justinianic Plague (Wagner et al. 2014). Phylogenetic analysis of this genome placed the Justinianic branch between two extant basal *Y. pestis* strains isolated from rodent populations in China, thus suggesting a Chinese origin for the Justinianic lineage. The Justinianic strain was found to have no extant descendants. In addition, nonsynonymous substitutions and one deletion were identified and were suggested to play a role in virulence changes (Wagner et al. 2014).

Here we present a high-coverage *Y. pestis* genome (17.9-fold) from the time of the Justinianic Plague, reconstructed from a victim buried during the sixth century in another early medieval settlement, Altenerding, southern Germany, approximately 20 km from Aschheim. The coverage and quality

of the data enabled a robust phylogenetic and structural analysis of the Justinianic *Y. pestis* strain. Comparison of the high-coverage genome to the previously published lower-coverage genome suggests that both represent a single strain. Our analysis confirms the phylogenetic placement previously suggested; however, we identify 19 out of 176 chromosomal substitutions previously reported in the lower-coverage genome as false positives. Fourteen of these false positives were purported as unique substitutions defining the Justinianic branch. In addition, our high-quality data permitted the identification of novel substitutions and structural polymorphisms that are unique to the Justinianic *Y. pestis* strain.

Results and Discussion

To identify Justinianic Plague victims that could potentially be used for *Y. pestis* genome reconstructions, teeth extracted from 20 skeletons unearthed from double burials in a sixth-century Bavarian burial ground in Altenerding, southern Germany were screened with a conventional polymerase chain reaction (PCR) assay (Seifert et al. 2013). Two adults (AE1175 female/AE1176 male) in the same double burial (Fig. 1, supplementary table S1, Supplementary Material online) generated amplification products of the expected size using primers specific for *Y. pestis*. An additional qPCR assay amplifying the *Y. pestis* specific plasminogen activator (*pla*) gene revealed that only individual AE1175, with 90 copies per μ l could be a candidate for subsequent *Y. pestis* genome capture (supplementary fig. S1, Supplementary Material online).

Radiocarbon dating suggests the two individuals died between 426 and AD 571 cal (95.4% probability at 2 sigma); grave goods point to a burial in or after ~530–570 (supplementary table S2, Supplementary Material online and Supplementary archaeological and historical information). According to historical sources, the Justinianic Plague struck Europe at least three times during this time period. The first known wave raged across the Mediterranean world from 541 to AD 543; further big waves occurred from 558 to ca. 590 (Stathakopoulos 2004). The incomplete written records do not mention plague in this region of Germany (see Supplementary archaeological and historical information). Since it is unclear whether the second big wave reached western Europe, these victims most likely died in the first or third wave of plague. In any case, the Altenerding genome dates to the first decades of the 200-year long pandemic. How the pathogen reached southern Germany is at present unknown: further evidence will be needed to clarify whether it traveled across the Alps from the Mediterranean as suggested for the Black Death (Carmichael 2014) or from France and western Germany. Alternatively, it could have traveled up the Danube from the east.

For genomic analysis, the DNA extract of individual AE1175 was converted into Illumina DNA double indexed libraries following established protocols (Kircher, et al. 2012; Meyer and Kircher 2010). Shotgun sequencing of the nonenriched library produced 5,024,997 reads of which only 0.91% mapped to the human genome (hg19) and 0.11% mapped to the *Y. pestis* reference chromosome (CO92), indicating that

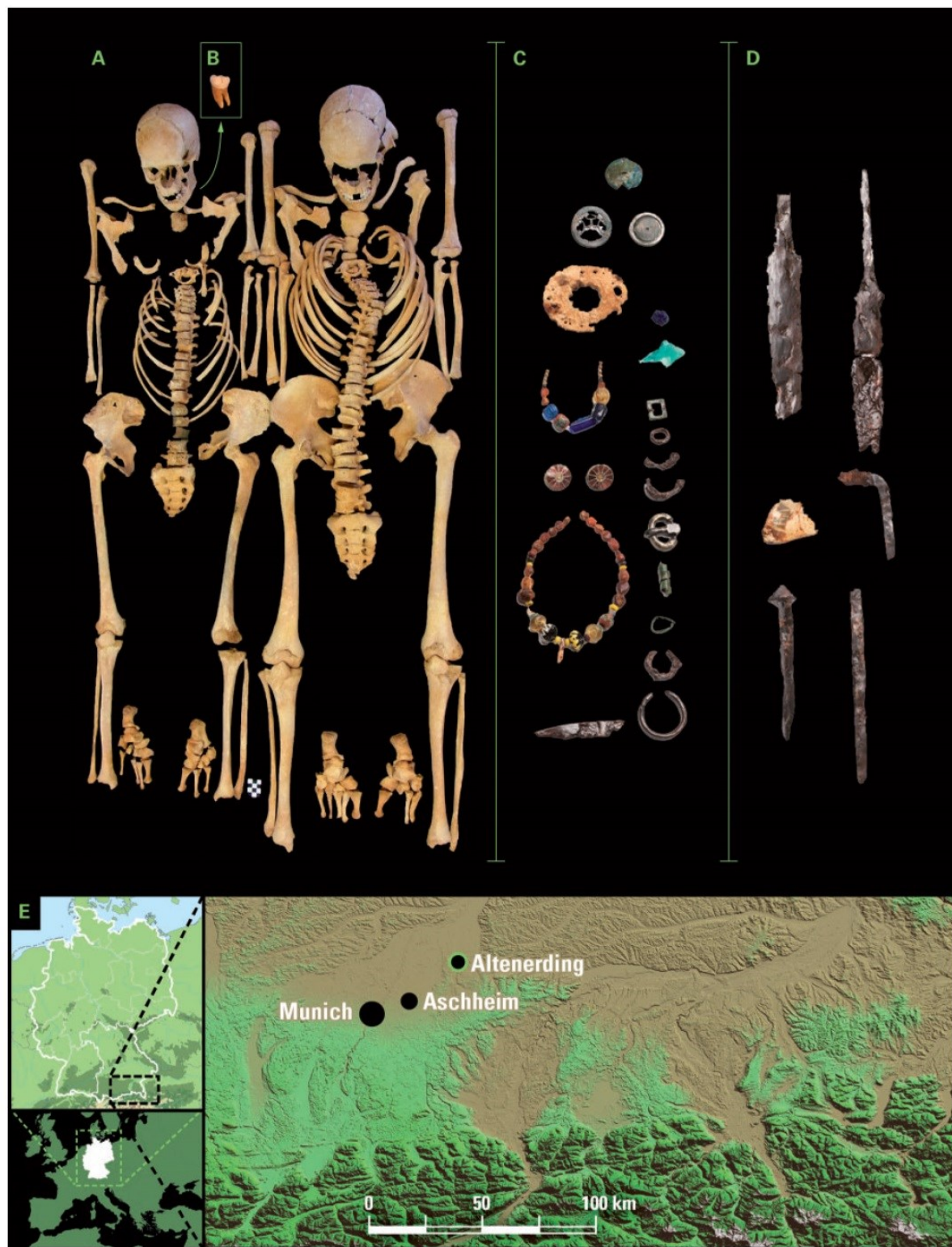


FIG. 1. The Altenerding plague burials. (A) Adult woman AE1175 (left) and adult man AE1176 (right) excavated at Altenerding and found positive for presence of *Y. pestis*. (B) The third molar sampled from individual AE1175 from which the Altenerding high-coverage genome was obtained. (C) Grave goods of individual AE1175 typical of the middle of the sixth century (© Archaeological State Collection Munich), Grave goods not shown true to scale (Detailed description in the historical and archaeological supplementary section). (D) Grave goods of individual AE1176 (© Archaeological State Collection Munich), Grave goods not shown true to scale (detailed description in the historical and archaeological supplementary section). (E) Geographical map specifying the location of the archaeological site “Altenerding”. The other site, where *Y. pestis* has been identified and its genome has been previously reconstructed (Wagner et al. 2014), is the cemetery of Aschheim, which is located approximately 20 km southwest of Altenerding.

targeted DNA enrichment would be required for genome-level analysis. Target enrichment was performed using DNA array hybridization. The DNA array capture probes were designed using the genome of *Y. pseudotuberculosis* as a template in order to avoid ascertainment bias for a specific *Y. pestis* strain. In addition, the *Y. pestis* (CO92) pCD1 and pMT1 plasmids were included. The array hybridization was performed following an established protocol (Hodges et al. 2009).

Sequencing of the enriched library produced in 62,126,715 total reads of which 11.3% mapped to the *Y. pestis* genome (9.69% to the chromosome), with an average length of 67 bp. The average genome coverage was 17.9-fold and 91.5% of the chromosome was covered at least 5-fold. Mapping to the pCD1 and to the pMT1 plasmid references resulted in an average 52-fold coverage with 94% of the positions covered at least 5-fold and an average of 77-fold coverage with 94.19% covered at least 5-fold, respectively (supplementary fig. S12, Supplementary Material online). The copy numbers of the plasmids can be roughly estimated by dividing the average coverage of plasmids by that of the chromosome. The estimated copy numbers were 3 for the pCD1 plasmid and 4.3 for the pMT1 plasmid. The following differences between the three loci may bias this estimate by affecting the average coverage of each locus: 1—mappability of the reference sequences, 2—base composition may alter the representation of sequenced reads (Benjamini and Speed 2012), 3—DNA degradation processes (Briggs et al. 2007) 4—average read length, and 5—DNA enrichment bias. The average read lengths are similar between the three loci and range between 61 to 68.7 bp (supplementary table S3, Supplementary Material online). To address the mappability and the base composition bias, artificial reads were generated from the CO92 genome reference sequence (with read length of 100 bp and tiling density of 1 bp). The artificial reads were mapped to the CO92 genome reference with the same mapping parameters used for the enriched library. Regions of 5,000 bp with 100% mappability (covered 100-fold) and a “G + C” content of around 48% were randomly chosen for each locus and an average coverage of the *Y. pestis* enriched reads in each of these three regions was calculated (supplementary table S3, Supplementary Material online). Dividing the average coverage of the corresponding regions in the plasmids by that of the chromosome resulted in a similar copy number estimate of 2.87 for the pCD1 plasmid and 2.55 for the pMT1 plasmid. Since DNA enrichment bias could affect the calculated copy number, it should be considered as a rough estimate testing whether the plasmid copy numbers in the sixth century strain are in general agreement with the previous estimates from modern *Y. pestis* strains (Parkhill et al. 2001).

Prior to array capture, the library was treated to prevent erroneous substitution assignments originating in DNA damage (Uracil DNA glycosylase [UDG] treatment) (Briggs et al. 2010). The DNA substitution plot of the library not treated with UDG mapped to the *Y. pestis* reference, displayed a pattern expected for ancient molecules (Briggs et al. 2007) with C to T substitution frequency of 21.3% at the first 5' position and G to A substitution of 20.7% at the first 3'

position (supplementary fig. S2, Supplementary Material online). Similar patterns were observed for endogenous human DNA thus supporting the authenticity of the ancient bacterium (supplementary fig. S3, Supplementary Material online). In addition, in-solution enrichment of human mitochondrial DNA was performed following established protocols (Maricic et al. 2010), resulting in 5,292-fold coverage. The mtDNA haplotype inferred from the obtained consensus sequence match the H1f1a haplogroup, which is common in European populations.

A mapping of 133 *Y. pestis* genomes to the CO92 reference genome resulted in 3,086 single nucleotide polymorphisms (SNPs) identified for the entire data set excluding the *Y. pseudotuberculosis* out-group. The number of SNPs ranged between 26 and 889 per genome (supplementary table S4, Supplementary Material online). For the ancient *Y. pestis* genome obtained from individual AE1175 (from now on referred to as the Altenerding genome) a total of 157 chromosomal SNPs were identified compared to the reference (supplementary table S5, Supplementary Material online) and 11 SNPs for the plasmids (supplementary table S6, Supplementary Material online). As input for phylogenetic analysis we used an alignment of 2,603 variable positions after removal of all positions in the alignment with missing data. The phylogenetic reconstructions (Fig. 2, supplementary figs. S4–S6, Supplementary Material online) resemble previously obtained phylogenies (Cui et al. 2013) and confirm the placement of the branch leading to the Justinianic strain on Branch 0 between two modern strains isolated from Chinese rodents (0.ANT1 and 0.ANT2) as reported previously (Wagner et al. 2014). However, our phylogeny indicates that the Justinianic strain was more genetically divergent than previously suggested. The genetic distance of the 14th-century Black Death strain and some of the extant strains isolated in China (e.g., 0.ANT1 and 0.ANT2) from the root of the tree is shorter than the much older Justinianic strain. This observation is counterintuitive, as ancient lineages should have fewer substitutions than their modern counterparts since they have had less time to accumulate genetic changes. Thus, the previously observed variation in mutation rates on different branches of the *Y. pestis* phylogeny (Cui et al. 2013; Wagner et al. 2014) is even more pronounced for the Justinianic strain than suggested before. The rate variation in *Y. pestis* has been attributed to an increased speed of SNP fixation as a result of higher replication rates of the bacteria per unit of time during epidemics and outbreaks compared with periods of enzootic disease (Cui et al. 2013).

The Aschheim cemetery and the Altenerding cemetery are located about 20 km apart and the two burials are dated to a similar time period (McCormick 2015; McCormick Forthcoming; Wagner et al. 2014). For ancient genome comparison, we reanalyzed the raw data from the Aschheim individual A120 genome (Wagner et al. 2014). When mapped to the *Y. pestis* reference chromosome, the reads from this reanalysis produced 3.9-fold average coverage with 31.08% of the CO92 reference chromosome covered at least 5-fold and 86% of the reference chromosome covered at least once; this is in contrast to an average of 7.6-fold coverage for 91.5% of

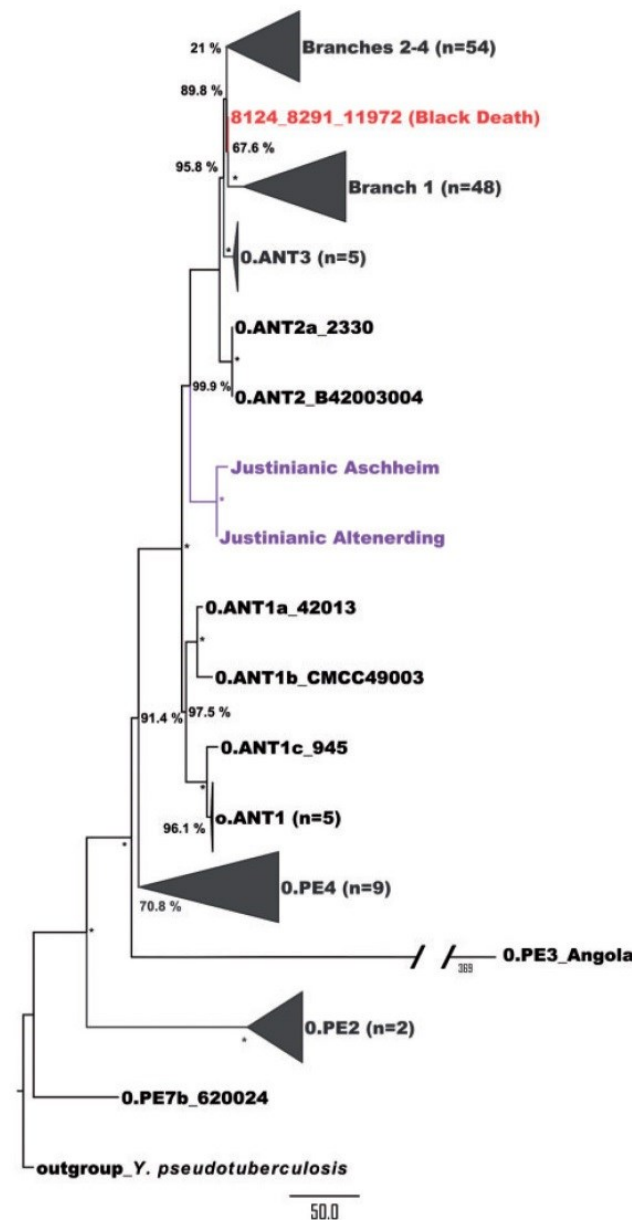


FIG. 2. Maximum parsimony tree: maximum parsimony analysis of 1418 nucleotide positions from genomes of 133 *Y. pestis* strains. All positions containing missing data were eliminated. Bootstrap values are indicated next to the nodes and bootstrap values of 100% are indicated by an asterisk. The tree is rooted using the genome of *Yersinia pseudotuberculosis* (strain IP32953). Branches leading to isolates from the historical pandemics are colored red and purple representing the second and first pandemics, respectively. Monophyletic nodes have been collapsed and are represented by a triangle. The number of isolates in a collapsed node is indicated in brackets. Our analysis suggests the specifically derived SNPs represented by the Aschheim branch to be false positives, supporting that the Aschheim and the Altenerding genomes likely represent the same bacterial strain.

the chromosome reported by Wagner et al. 2014. Only 55.51% of the reanalyzed Aschheim genome was covered at least three times when mapped to the CO92 reference, compared with 92.32% covered at least three times in the

Altenerding genome when mapped against the same reference with the same parameters. Of the 176 chromosomal SNPs reported by Wagner et al. 2014, 66 were not called in the Altenerding genome (supplementary table S7, Supplementary Material online). Of these, 47 were excluded from our analysis since they fell in noncore regions. Noncore regions, which have been defined elsewhere, contain deletions or duplications and are therefore more prone to erroneous mapping (Cui et al. 2013; Morelli et al. 2010). We inspected the remaining 19 SNPs to determine whether they indeed represent variation between the two genome sequences. Reads spanning a 150 bp flanking region on either side of each SNP from both the Altenerding genome and the reanalyzed Aschheim genome were considered (supplementary fig. S7, Supplementary Material online). Visual inspection revealed that these SNPs did not fulfill our criteria for SNP calling (coverage of more than five times and presence in at least 90% of the sequences obtained). The regions surrounding most of these positions show a pattern of abnormal coverage peaks (Fig. 3, supplementary fig. S8, Supplementary Material online) in which the SNP position is often heterozygous and located at the end of reads. A significant drop in coverage is often observed next to the SNP position, resulting in a lack of overlapping reads to support the SNP. These regions are therefore prone to erroneous SNP calls and the visual inspection suggests the reads containing the SNPs are not authentic to the *Y. pestis* ancient strain. The abnormal coverage peaks seem not to be artifacts from the SNP capture, since similar coverage peaks are also apparent in our reanalysis of the non-SNP captured data for these libraries (supplementary figs. S8 and S9 Supplementary Material online). In addition, our reanalysis revealed some of these 19 SNP positions to be identical to the reference sequence, whereas others had a coverage that is below the threshold set for SNP calling (supplementary fig. S7, Supplementary Material online). This analysis suggests that these 19 SNPs reported previously for the Aschheim genome are false positives and do not represent a true variation between the two genomes. Of these 19 SNPs, 14 were reported as derived SNPs, 2 of which were reported as nonsynonymous nucleotide changes (supplementary table S8, Supplementary Material online). In an additional analysis, we performed SNP calling for the reanalyzed Aschheim genome and identified 8 derived SNPs not shared with other *Y. pestis* genomes, including the Altenerding genome reported here (supplementary fig. S10, Supplementary Material online). Visualization of these SNPs revealed a similar pattern of abnormal coverage peaks and heterozygosity, suggesting that they too are likely false positives. Two of these positions (position 1,371,025 and 3,392,897) are among the 19 false positives described above, three positions (positions 362,357, 3,956,001 and 4,575,345) were manually removed from the final analysis in Wagner et al. 2014 following a visual inspection by the authors, and the remaining three were not identified in the published analysis of this genome (Wagner et al. 2014). This analysis reveals a repeated and problematic pattern in the Aschheim data set, which could in some cases, cause false SNP calls even under stringent SNP calling parameters. This pattern is

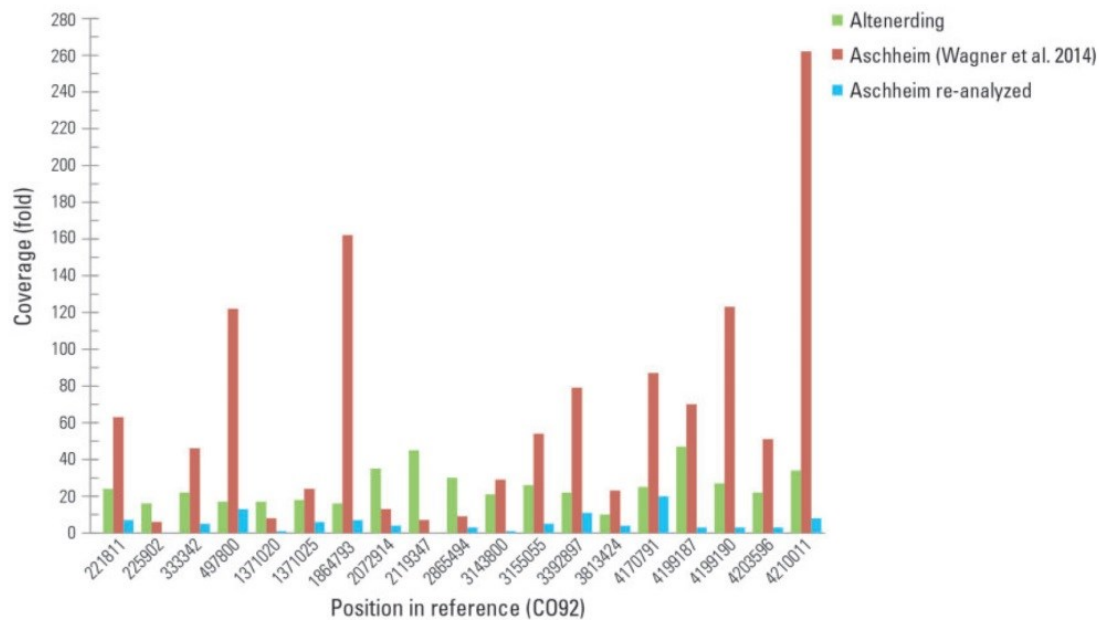


FIG. 3. Coverage of the false positive substitutions called for the Aschheim genome by Wagner et al., 2014 compared with the reanalyzed Aschheim genome and the Altenerding genome. Fold coverage was determined for each position where substitutions were erroneously reported in the Aschheim genome (Wagner et al. 2014). Values for the Altenerding genome, values published in Wagner et al. 2014 for the Aschheim genome and the values for the reanalyzed Aschheim genome are colored green, red, and blue, respectively. Most of the false positive positions in the Aschheim genome contain abnormal coverage peaks, significantly higher than the average coverage of the corresponding genome.

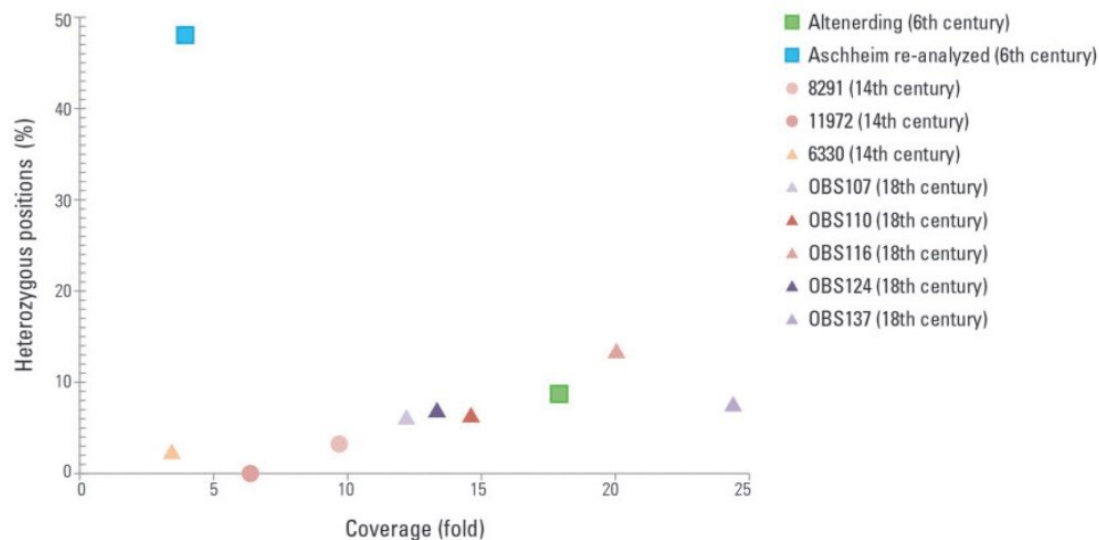


FIG. 4. Comparison of coverage and heterozygous positions between ancient *Y. pestis* genomes. The average coverage calculations and the calling of heterozygous positions were done against the CO92 chromosome (NC_003143.1) as a reference. The percentage of heterozygous positions was calculated out of the total number of SNPs called for each genome, using a mapping quality threshold of 30 and a coverage threshold of 5 reads. The raw data published in Wagner et al. 2014, Bos et al. 2011, and Bos et al. 2016 were reanalyzed and used for calculating the coverage and the percentage of heterozygous positions for the Aschheim Justinianic genome and the genomes from the second pandemic, respectively. The Black Death genome 8124 was excluded from this analysis since the coverage was not sufficient for calling of heterozygous positions. The Aschheim Justinianic genome contains a high percentage of heterozygous positions compared with the other ancient *Y. pestis* genomes. Heterozygous positions called for a genome of the haploid *Y. pestis* indicate nonendogenous reads.

exceedingly problematic for low-coverage genomes where nonendogenous reads could dominate positions where few or no endogenous reads mapped.

Our reanalysis of the Aschheim genome resulted furthermore in a high percentage of heterozygous positions. A position was considered heterozygous when reads mapping to

the position showed more than one allele, which is unexpected in a haploid organism. We identified a total of 48.05% heterozygous positions where SNPs were identified (111 heterozygous positions out of 231 SNPs) in the reanalyzed Aschheim genome compared with 8.72% (15 heterozygous positions out of 172 SNPs) in the Altenerding genome

Table 1. Derived nonsynonymous SNPs compared with CO92.

Position (CO92 chromosome)	CO92	Altenrading	Codon Change	AA Change	Gene ID	Gene Name	Gene Function
260148 ^a	C	T	CCG to TCG	P to S	YPO0257		Type III secretion protein
271114 ^a	C	A	CTT to ATT	L to I	YPO0270		Type III secretion system protein (iron-sulfur binding protein)
557841 ^a	C	T	CGT to CAT	R to H	YPO0517	<i>hepA (RapA)</i>	ATP-dependent helicase (transcription regulator)
727741	G	A	GAA to AAA	E to K	YPO0668	<i>parE</i>	DNA topoisomerase IV subunit B
1067966	C	A	GGC to TGC	G to C	YPO0966		Kinase
1296743 ^a	C	T	GTG to ATG	V to M	YPO1150	<i>bioA</i>	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase (part of the Biotin operon)
1434752 ^a	C	A	GAT to TAT	D to Y	YPO1275	<i>spr</i>	Outer membrane lipoprotein (Murein hydrolase)
1530658	C	A	AGA to ATA	R to I	YPO1363		Virulence factor
1609461 ^a	T	C	ACA to GCA	T to A	YPO1417		Iron-sulfur binding protein
1754708	C	T	CCA to CTA	P to L	YPO1539	<i>galU</i>	UTP-glucose-1-phosphate uridylyltransferase
2352174 ^a	T	G	GTG to GGG	V to G	YPO2071		DEAD box family helicase
2977542 ^a	C	A	AGC to ATC	S to I	YPO2649	<i>nrdE</i>	Ribonucleotide-diphosphate reductase subunit alpha
3078807 ^a	C	A	CGT to AGT	R to S	YPO2747	<i>fadJ (faoA)</i>	Multifunctional fatty acid oxidation complex subunit alpha
3360963 ^a	A	C	ACC to CCC	T to P	YPO3008		Two-component sensor histidine kinase (TCSs)
3360984 ^a	C	T	CAT to TAT	H to Y	YPO3008		TCSs
3500922	T	G	GTG to GGG	V to G	YPO3141	<i>tesB</i>	Acyl-CoA thioesterase
3535148 ^a	G	T	GCT to TCT	A to S	YPO3171	<i>apbA (panE)</i>	2-dehydropantoate 2-reductase (vitamin B5 biosynthesis)
3560088	G	A	CCA to TCA	P to S	YPO3199		Short chain dehydrogenase
3568597	C	T	GGA to GAA	G to E	YPO3205	<i>phoB</i>	Phosphate regulon transcriptional regulator
4066494 ^a	C	T	GTT to ATT	V to I	YPO3646	<i>pcp (pcpY, slyB)</i>	Outer membrane lipoprotein
4307755	G	A	GCG to GTG	A to V	YPO3839		Sugar transport system permease
4460688	C	T	CGG to CAG	R to Q	YPO3963		Sugar transport system permease
pCD1							
29959	A	G	AAC to AGC	N to S	YPCD1.41	<i>yscO</i>	Type III secretion apparatus component
50462	T	C	AAA to GAA	K to E	YPCD1.71c	<i>yopJ</i>	Targeted effector protein
66608	C	T	CTT to TTT	L to F	YPCD1.92		

^aDerived nonsynonymous SNPs detected in the Altenrading genome that were not called in Wagner et al. 2014 for the Aschheim genome.

(Fig. 4). The Aschheim enriched libraries were reported by Wagner et al., 2014 to be UDG treated, thus, damage that is typical of ancient DNA should be absent from the raw data and should not affect heterozygosity levels. The possibility that the A120 individual was infected with multiple strains of *Y. pestis* that exhibit a genetic distance as high as 48.05% from each other at polymorphic positions is unlikely. Furthermore, the genome-wide SNP allele frequency plot of the reanalyzed Aschheim genome does not show a bimodal pattern or any other pattern consistent with an infection with multiple strains as observed in other pathogens (Bos et al. 2014) (supplementary fig. S11, Supplementary Material online). It is, therefore more likely that the data contain a high percentage of reads that are either nontarget or are damaged in a nonstandard way for ancient molecules. There are several possible causes for a high percentage of heterozygous positions in a reconstructed ancient genome. One possibility is contamination of the sample by a phylogenetically close organism. Sequenced reads originating from the genome of such an organism could erroneously map to regions of the reference sequence that are conserved among bacterial species. It should be noted, however, that *Y. pestis* is the best match for these reads in the current Genbank database. Other explanations could be related to sample processing, including the accumulation of PCR errors due to extensive amplification of low-template libraries or artifacts created as a result of genome-wide capture processes. Identification of substitutions in low-coverage genomes necessitates a low-minimal coverage thus, making these analyses prone to false positive SNP calls. Awareness of the limitations arising from low-coverage data and attention to parameters such as heterozygosity are important for reliable future analyses.

Excluding the above putative false positive SNPs, no difference was found between the Aschheim genome and the Altenerding genome reported here, supporting that both genomes represent the same bacterial strain and that the infected individuals were victims of the same outbreak. Previous studies have reported the presence of multiple *Y. pestis* strains in modern plague outbreaks (Guiyoule et al. 1994; Shivaji, et al. 2000), for example, in the case of the 1980s and 1990s plague outbreak in the Ambositra region of Madagascar (Guiyoule et al. 1997). However, our findings point to a low-genetic diversity of the bacterium in this rural region of southern Germany during the sixth century. Future investigations of other Justinianic material may reveal the range of the outbreak, and allow us to comment on the paths traveled by the disease and its rate of transmission.

SNP calling for the Altenerding genome resulted in the detection of 63 substitutions unique to the Justinianic strain (supplementary table S9, Supplementary Material online), 30 of which were not previously reported. Of these novel unique SNPs, 14 are nonsynonymous (supplementary table S10, Supplementary Material online), 3 of which are located in genes that are thought to be associated with plague virulence: the *nrpE*, *fadJ*, and *pcp* genes (table 1). The *nrpE* gene codes for the subunit alpha of a protein catalyzing Deoxynucleotide Triphosphate (dNTP) synthesis. This gene is strongly transcriptionally induced in human plasma infected by *Y. pestis*

(Chauvaux et al. 2007) and is upregulated in the bubo during infection by *Y. pestis* in rats (Sebbane et al. 2006). The *fadJ* gene codes for the FadJ protein that is involved in fatty acid oxidation and is transcriptionally upregulated in *Y. pestis* infected cells at 37 °C (Chauvaux et al. 2011). The *pcp* gene (also referred to as *pcpY* and *slyB*) codes for an outer membrane lipoprotein. A study of the *Y. pestis* transcriptional profile has shown that a homolog of the *pcp* was transcribed 5.5-fold higher in a *lcrG* deleted strain of *Y. pestis* compared with the wild type strain. *lcrG* is a negative regulator for secretion of Yersinia outer membrane proteins (Yops), suggesting a role for *pcp* in the type 3 secretion system (Du et al. 2009) which is a known virulence determinant (Perry and Fetherston 1997).

Insertion and deletion analysis revealed 16 regions present in the Justinianic strain that were absent in the CO92 reference (supplementary table S11, Supplementary Material online). One such region is the 15.6 kb DFR4 locus (Radnedge et al. 2002) that was also present in other historic epidemic strains from the Black Death and post-Black Death periods including the previously published Justinianic genome (Bos et al. 2016; Wagner et al. 2014). Another such region is a 1122bp long locus that includes parts of the *glpFKX* operon that enables aerobic glycerol fermentation. Loss of capacity to ferment glycerol is shared by several *Y. pestis* “branch 1” strains, including CO92, due to a deletion in the *glpD* gene. Furthermore, additional deletions in the *glpFKX* operon are often present (Achtman et al. 1999; Motin et al. 2002). The aerobic glycerol fermentation pathway has been suggested to upregulate *Y. pestis* biofilm formation in the flea gut. In addition, a negative selection against *glpK* activity in the presence of dysfunctional *glpD* has been suggested (Willias, et al. 2014; Zinser et al. 2003). The fact that the *glpFKX* operon is pseudogenized in some *Y. pestis* strains but not all (Tong et al. 2005) indicates that its loss is not adaptive to the shift from an environmental and enteric life-style to one of specialized rodent-flea transmission, but is rather due to either genetic drift or a selective advantage specific to the *glpFKX* pseudogenized strains. The aerobic glycerol fermentation pathway is induced during different phases of the *Y. pestis* infectious cycle (Motin et al. 2004; Zhou et al. 2006). Thus, its loss in CO92 should be further explored to evaluate its potential role in regulating biofilm formation and its subsequent effect on the bacterial fitness and disease dynamics.

In an attempt to identify unique structural differences we compared the Altenerding genome to the phylogenetically close strains 0.ANT1b_CMCC49003 and 0.ANT2a_2330. Three deletions unique to the Justinianic strain were observed. A 1334 bp long unique deletion (positions 1010683–1012016 in the *Y. pseudotuberculosis* reference) in the metabolic *pheA* gene, which functions in L-phenylalanine biosynthesis, was detected. This deletion may account for an obligatory nutritional requirement for L-phenylalanine in the Justinianic strain similar to those exhibited in other *Y. pestis* strains (Brubaker 1991; Parkhill et al. 2001). Further we detected a 1074 bp unique deletion (positions 2563546–2564619 in the CO92 reference) interrupting the CO92 annotated gene YPO2283. YPO2283 is homologous to the *Escherichia coli* *galR* gene and has the characteristics of a

lacI family transcription regulator. However, the gene's function in *Y. pestis* is unknown (Carnoy et al. 2002; Marceau 2005). Finally, a unique deletion of 195 bp (positions 3462627–2462821 in the *Y. pseudotuberculosis* reference) was located in the range of the metabolic *celB* gene (also called *chbC*) that codes for the Phosphotransferase (PTS) system cellobiose-specific IIC component. This gene has been shown to be upregulated in the flea compared with in vitro conditions (Vadyvaloo et al. 2010). The presence of these unique deletions in the Justinianic strain call for further functional studies on the proteins coded by the above genes to examine whether and in what way the deletions affect bacterial physiology and disease infectious cycle.

In summary, we have demonstrated the presence of *Y. pestis* in a second early medieval rural site in southern Germany where no historical source records it, expanding the number of sites known to have been afflicted by plague and reinforcing the time placement of this strain in the early waves of the 200-year-long pandemic. Our reconstructed high-coverage genome permitted a more thorough and reliable analysis than was previously possible. Through this we identified novel substitutions and structural polymorphisms in this early *Y. pestis* strain, located in genomic regions such as the *glpFKX* operon (Motin et al. 2004; Zhou et al. 2006) as well as in genes such as *nrdE*, *fadJ*, and *pcp* that have been suggested as plague virulence factors (Chauvaux et al. 2007; Sebbane et al. 2006). Substitutions reported in the previously published Justinianic lower-coverage genome that were here found to be erroneous highlight the importance of using high-quality ancient pathogen genomes for future analysis when they are available, as well as following strict criteria of authenticity and genome quality. Such criteria include average coverage, evenness of coverage, and heterozygosity estimates in addition to the standard criteria such as DNA damage and fragmentation patterns. Exhaustive sampling of ancient material from suspected plague foci of various periods and geographical locations is needed to increase availability of high-quality data sets and to deepen our understanding of the disease, its evolution and its human impact. In addition, the unique genetic features identified here call for functional studies that would explore their role in terms of *Y. pestis* physiology and adaptation.

Materials and Methods

Prescreening of Samples

A total of 20 individuals from the Altenerding site, located in Bavaria, Germany, were included in the *Y. pestis* prescreening. The cemetery was used continuously between the first half of the fifth century and until the first half of the seventh century, with around 1,450 graves (Hakenbeck 2011; Losert and Pleterski 2003). The 20 individuals were buried in 10 double burials. Two teeth were sampled per individual and tested twice in independent PCRs for replication. The samples were prepared as described in Sofeso et al. 2012 (Sofeso et al. 2012) and DNA extraction was performed following the protocol C of Yang et al. 1998, modified by Wiechmann and Grupe 2005, using up to 0.40 g of tooth powder. All pre-PCR steps were performed in the aDNA facility of the ArchaeoBioCenter,

University of Munich (Seifert et al. 2013). Conventional PCR targeting 133 bp of the *Y. pestis* specific plasminogen activator (*pla*) gene on the pPCP1 plasmid was performed on both samples of each individual, respectively, as described in Seifert et al. 2013, using the primers *Y.-pest_F* and *Y.-pest_R2* (Tomaso et al. 2003). All PCRs were performed with extraction blanks and PCR blanks. Individuals AE1175 and AE1176 buried in a shared grave were found positive for *Y. pestis* DNA in the prescreening. Two new teeth were removed from each individual: AE1175 (lower- and upper- M3 teeth) and AE1176 (two molars from the mandible) for further screening.

Radiocarbon and Archaeological Dating

Metacarpal bones from individuals AE1175 and AE1176 were radiocarbon dated at the CEZ Archaeometry gGmbH, Mannheim, Germany. In addition, an archaeological date was assessed based on the grave furnishing (Supplementary archaeological and historical information).

DNA Extraction

DNA was sampled from each tooth by drilling the dental pulp. Approximately 50 mg of drilled powder was used for DNA extraction. DNA was extracted following established protocols for archaeological skeletal tissue (Rohland and Hofreiter 2007). Extraction was performed in a dedicated aDNA laboratory located at the University of Tuebingen. A negative control was included.

PLA Assay

Additional screening of all extracts was performed to assess preservation of *Y. pestis* DNA, using a previously described qPCR assay (*pla* assay) (Schuenemann et al. 2011) specific to the plasminogen activator (*pla*) gene that is located on the high copy pPCP1 plasmid of *Y. pestis* (Parkhill et al. 2001). The *pla* assay was performed using a Roche Lightcycler 480 machine. Standards were diluted in 10-fold serial dilutions to 2.23 copies/μl in TET buffer (10 mM Tris, 1 mM Ethylenediaminetetraacetic acid (EDTA), and 0.05% Tween). PCRs were performed in 20 μl reactions consisting of 1 unit of 10X PCR buffer II, 2.5 mM MgCl₂, 250 μM each dNTP, 5% Dimethyl sulfoxide (DMSO), 0.75 mg/ml Bovine serum Albumin (BSA), 300 nM each primer (Schuenemann et al. 2011), 1 unit of 20X EvaGreen dye (Biotium), 0.05 U/μl Amplitaq gold DNA polymerase (Applied Biosystems) and 2 μl of DNA extract.

Library Preparation for Sequencing

A 10 μl aliquot of each extract was converted into an Illumina double stranded (ds) DNA library. In addition, a 60 μl aliquot of each extract was converted into a concentrated ds Illumina DNA library, following established protocols (Meyer and Kircher 2010), with replacement of the final MinElute (Qiagen) purification by an 80 °C incubation for 20 minutes to denature the *Bst* polymerase. The concentrated libraries were pretreated with a UDG enzyme followed by endonuclease VIII to remove deaminated cytosines in this way preventing erroneous substitution assignments originating in DNA damage (Briggs et al. 2010).

A negative control for library preparation was included for each library type. Libraries were quantified using an IS7 and IS8 primer quantification assay (Lightcycler 480 Roche) (Meyer and Kircher 2010), using the DyNAmo SYBR Green qPCR kit (Thermo Scientific). Each library was double indexed following established protocols in 1 to 12 parallel 100 μ l reactions (Kircher, et al. 2012). Indexed products for each library were then pooled and purified over MinElute columns (Qiagen), eluted in 50 μ l TET buffer (10 mM Tris, 1 mM EDTA, and 0.05% Tween), and quantified using the above quantification assay with IS5 and IS6 primers (Meyer and Kircher 2010). Five microliters of MinElute purified product was subsequently amplified in 100 μ l reactions using AccuPrime Taq DNA polymerase (Life technologies) following the manufacturer's protocol with 0.3 μ M each primer. Products were again MinElute purified and quantified using the Agilent 2100 Bioanalyzer DNA 1000 chip. A 10 nM pool of all the libraries not treated with UDG was prepared for shotgun sequencing. The UDG treated concentrated libraries were used for downstream *Y. pestis* array capture enrichment. In addition, both UDG treated and non-UDG treated libraries were used for downstream mitochondrial in-solution target enrichment. The sequenced data generated from the human mitochondrial DNA enriched libraries was later used to estimate the preservation of human DNA in the sample and to assess the level of ancient human DNA damage as another means of ancient DNA authentication.

Mitochondrial Target Enrichment

Target enrichment of human mitochondrial DNA was performed by in-solution bead capture of the pooled libraries, following established protocols (Maricic, et al. 2010) and using baits generated from modern human mitochondrial DNA.

Y. Pestis Array Enrichment

Probes for the *Y. pestis* array capture were designed using an in-house probe design software. In order to avoid ascertainment bias for a specific *Y. pestis* strain, the genome of the *Y. pseudotuberculosis* outgroup (GenBank accession NC_006155) was used as a template for probe design. In addition, the *Y. pestis* (CO92) pCD1 (70 kb) and pMT1 (100 kb) plasmids were included (GenBank accession NC_003131 and NC_003134, respectively). Probes were designed with 5 bp tiling density across the *Y. pseudotuberculosis* template and 6 bp tiling density across the CO92 pMT1 and pCD1 plasmids template. A total of 976,658 probes were used on an Agilent 1-million feature array.

Array hybridization was performed following established protocols (Hodges et al. 2009), with two nights of incubation at 65 °C. Postharvest the 490 μ l eluate was reamplified to generate a 20 μ g pool of purified product for serial capture on an array identical to the one used for the first round of capture. Products were reamplified and purified as described above and made into a 10 nM pool for high-throughput sequencing. In addition, the purified product from the first round of capture was also made into a 10 nM pool and sequenced on in the same way.

Sequencing

Library pools were sequenced on the Illumina HiSeq 2500 platform using two index reads (2*100 + 7 + 7 cycles) following the manufacturer's protocol. Deindexing was performed by sorting all sequences by their P7 and P5 combinations, using the CASAVA software version 1.8. Adapters were clipped from all reads and forward and reverse reads were merged into single sequences if they overlapped by at least 11 bp (Kircher, et al. 2011). Reads that could not be merged were discarded. All reads were filtered for a length of at least 30 bp.

Analysis of Sequenced Reads

Merged and filtered reads from the two rounds of serial *Y. pestis* capture were pooled and mapped via BWA version 0.7.12 (Li and Durbin 2009) to the CO92 *Y. pestis* reference chromosome (GenBank accession NC_003143) and to the pCD1 plasmid (GenBank accession NC_003131) and the pMT1 plasmid (GenBank accession NC_003134). To avoid cross-mapping of reads from other organisms, the mapping stringency of BWA was increased to 0.1 ($-n$ parameter). Duplicate reads were removed and remaining reads were filtered for a minimal mapping quality of 37 using Samtools version 0.1.19. Mean coverage and coverage across the reference were calculated and plotted using QualiMap version 2.1. The same mapping tools and parameters were used for all data sets in the study. A set of published *Y. pestis* genomes that have been described elsewhere (Cui et al. 2013) was used, together with the published ancient Black Death *Y. pestis* genome (Bos et al. 2011), a reanalyzed A120 genome (Wagner et al. 2014), the new Altenerding genome, and the genome of strain IP32953 of *Y. pseudotuberculosis* as an outgroup (GenBank accession NC_006155). For complete genomes, artificial reads were generated by a tiling approach with read length of 100 bp and tiling density of 1 bp. Genomes used for analyses are listed in [supplementary table S4, Supplementary Material](#) online.

Ancient DNA Damage Assessment

The level of DNA damage was determined for the reads mapping to the human mitochondrial reference and for the reads mapping to the *Y. pestis* reference, using mapDamage version 2.0 (Jonsson et al. 2013) with standard parameters.

SNP Calling

Comparative SNP typing was performed on mapped data using the UnifiedGenotyper of the Genome Analysis Toolkit (GATK) (DePristo et al. 2011). *vcf* files (variant call format) were produced for the set of 133 genomes. The "EMIT_ALL_SITES" option was set to generate calls for both variant and nonvariant sites. A custom Java program (MultiVCFanalyzer) was applied to the generated *vcf* files to analyze and filter the detected SNPs and to produce a multiple sequence alignment of all positions for which a SNP was called in at least one of the strains in the complete data set. A SNP was called if GATK called a homozygous SNP, if the position was covered by at least 5 reads and the quality of the GATK call was at least 30. If GATK called a heterozygous

SNP, we called the SNP allele if it was supported by at least 5 reads with a quality of at least 30 and if the fraction of reads containing the SNP was at least 90%. If a SNP call was not possible, the reference base was called instead if the respective thresholds were reached. If neither a reference call nor a SNP call was possible based on these thresholds, the “N” character was inserted in the SNP alignment. We omitted genomic regions that were defined in previous studies as noncore regions (Cui et al. 2013; Morelli et al. 2010). In addition, we excluded tRNA regions, rRNAs, and annotated repetitive elements.

Comparison of SNPs called for the Aschheim genome by Wagner et al. 2014 to SNP Calls in the Altenerding Genome

SNPs called in Wagner et al. 2014 for the A120 Aschheim genome were compared with the SNPs called for the Altenerding genome and to SNPs called in the reanalyzed A120 Aschheim genome. In order to inspect discrepancies in SNP calling between the compared genomes, relevant positions were visualized using the IGV gene browser (Robinson et al. 2011).

Phylogenetic Analysis

For phylogenetic analysis, the SNP alignment generated from the *vcf* files was used as input for phylogenetic reconstruction software MEGA6 (Tamura et al. 2013). All alignment columns that contained missing values (i.e., “N”) were excluded from the analysis (complete deletion) in order to avoid any bias resulting from different genomic coverage between samples. Phylogenies were constructed from a total of either 2603 nucleotide positions (phylogenies excluding A120) or 1418 nucleotide positions (phylogenies including A120). Both maximum-likelihood and maximum-parsimony topologies were generated. Bootstrap support values were obtained over 1,000 replicates. The phylogenetic trees were edited in FigTree version 1.4.2.

SNP Effect Analysis

SNPs that were detected in the Altenerding Justinianic strain were investigated for their effect on protein-coding genes using *SnpEff* (Cingolani et al. 2012). Genes were annotated according to the annotation in the CO92 *Y. pestis* reference genome. *SnpEff* was applied to the complete set of SNPs using standard parameters, except that the size of the upstream and downstream regions of genes in which SNPs are classified was set to 100 nt.

Indel Analysis

In order to identify regions in the Altenerding genome that are missing in CO92, we performed a second mapping of the captured Altenerding reads to the CO92 reference as described above but with the mapping quality filtering parameter set to 0 ($-q$) in order to retain also unspecific mapping reads. We extracted all reads that did not map to CO92 and mapped them to the *Y. pseudotuberculosis* reference as described above. All covered regions longer than 250 bp were visually examined using the IGV gene browser.

In order to investigate regions that were uniquely deleted in the Altenerding Justinianic strain, noncovered regions in the mapping to *Y. pseudotuberculosis* reference were listed and cross-referenced with a list of noncovered regions in the mapping of two phylogenetically close strains to the *Y. pseudotuberculosis* reference: 0.ANT1b_CMCC49003 which is basal to the Justinianic strains and 0.ANT2a_2330 (Cui et al. 2013) which is more derived than the Justinianic strains. Regions that were covered in the two 0.ANT strains were filtered for length higher than 100 bp and visually inspected using the IGV gene browser. In addition, these regions were cross-referenced with the probes used for enrichment to make sure they are not missing due to capture bias.

Supplementary Material

Supplementary figures S1–S12 and tables S1–S11 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We thank Julian Parkhill and Sophia David from the Sanger Institute, Cambridge, UK for their useful comments on data analyses and on early drafts. We thank Dr Verena Schünemann and the other members of the Institute for Archaeological Sciences, Tuebingen University as well as the members of the Department of Archaeogenetics of the Max Planck Institute for the Science of Human History for their comments and advice. We thank Stephen Clayton, Alexander Immel, and Ofer Habushi for their help and support with data analysis. We thank Annette Günzel for graphical support. We thank the city of Erding for funding the prescreening of the samples from Altenerding. This work was supported by the European Research Council starting grant APGREID (to J.K.).

References

- Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A, Carniel E. 1999. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci USA* 96(24):14043–14048.
- Benedictow OJ. 2004. The Black Death, 1346–1353: the complete history. England: Boydell & Brewer.
- Benjamini Y, Speed TP. 2012. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* 40(10):e72.
- Bos KI, Herbig A, Sahl J, Waglechner N, Fourment M, Forrest SA, Klunk J, Schuenemann VJ, Poinar D, Kuch M. 2016. Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus. *eLife* 5:e12994.
- Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, Forrest SA, Bryant JM, Harris SR, Schuenemann VJ. 2014. Pre-Columbian mycobacterial genomes reveal seals as a source of new world human tuberculosis. *Nature* 514(7523):494–497.
- Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, Coombes BK, McPhee JB, DeWitte SN, Meyer M, Schmedes S, et al. 2011. A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* 478(7370):506–510.
- Briggs AW, Stenzel U, Johnson PL, Green RE, Kelso J, Prufer K, Meyer M, Krause J, Ronan MT, Lachmann M, et al. 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci.* 104(37):14616–14621.

- Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, Paabo S. 2010. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* 38(6):e87.
- Brubaker RR. 1991. Factors promoting acute and chronic diseases caused by *Yersinia*. *Clin Microbiol Rev.* 4(3):309–324.
- Carmichael AG. 2014. Plague persistence in western Europe: A hypothesis. *The Medieval Globe* 1(1):8.
- Carnoy C, Floquet S, Marceau M, Sebbane F, Haentjens-Herwegh S, Devalckenaere A, Simonet M. 2002. The superantigen gene *ypm* is located in an unstable chromosomal locus of *Yersinia pseudotuberculosis*. *J Bacteriol.* 184(16):4489–4499.
- Chauvaux S, Dillies M, Marceau M, Rosso M, Rousseau S, Moszer J, Simonet M, Carniel E. 2011. In silico comparison of *Yersinia pestis* and *Yersinia pseudotuberculosis* transcriptomes reveals a higher expression level of crucial virulence determinants in the plague bacillus. *Int J Med Microbiol.* 301(2):105–116.
- Chauvaux S, Rosso ML, Frangeul L, Lacroix C, Labarre L, Schiavo A, Marceau M, Dillies MA, Foulon J, Coppee JY, et al. 2007. Transcriptome analysis of *Yersinia pestis* in human plasma: An approach for discovering bacterial genes involved in septicemic plague. *Microbiology* 153(Pt 9):3112–3124.
- Christakos G, Olea RA, Yu H. 2007. Recent results on the spatiotemporal modeling and comparative analysis of Black Death and Bubonic Plague epidemics. *Public Health* 121(9):700–720.
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6(2):80–92.
- Cohn SK. 2008. Epidemiology of the Black Death and successive waves of plague. *Med Hist.* 52:74–100.
- Cui Y, Yu C, Yan Y, Li D, Li Y, Jombart T, Weinert LA, Wang Z, Guo Z, Xu L, et al. 2013. Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proc Natl Acad Sci.* 110(2):577–582.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 43(5):491–498.
- Du Z, Tan Y, Yang H, Qiu J, Qin L, Wang T, Liu H, Bi Y, Song Y, Guo Z. 2009. Gene expression profiling of *Yersinia pestis* with deletion of *lcrG*, a known negative regulator for Yop secretion of type III secretion system. *Inter J Med Microb* 299(5):355–366.
- Duncan CJ, Scott S. 2005. What caused the black death? *Postgrad Med J.* 81(955):315–320.
- Duplantier J, Duchemin J, Chanteau S, Carniel E. 2005. From the recent lessons of the Malagasy foci towards a global understanding of the factors involved in plague reemergence. *Vet Res.* 36(3):437–453.
- Ensore RE, Biggerstaff BJ, Brown TL, Fulgham RE, Reynolds PJ, Engelthaler DM, Levy CE, Parmenter RR, Monteneri JA, Cheek JE. 2002. Modeling relationships between climate and the frequency of human plague cases in the southwestern United States, 1960–1997. *Am J Trop Med Hyg.* 66(2):186–196.
- Guiyoule A, Grimont F, Iteman I, Grimont PA, Lefevre M, Carniel E. 1994. Plague pandemics investigated by ribotyping of *Yersinia pestis* strains. *J Clin Microbiol.* 32(3):634–641.
- Guiyoule A, Rasoamanana B, Buchrieser C, Michel P, Chanteau S, Carniel E. 1997. Recent emergence of new variants of *Yersinia pestis* in Madagascar. *J Clin Microbiol.* 35(11):2826–2833.
- Haensch S, Bianucci R, Signoli M, Rajerison M, Schultz M, Kacki S, Vermunt M, Weston DA, Hurst D, Achtman M. 2010. Distinct clones of *Yersinia pestis* caused the Black Death. *PLoS Pathogens* 6(10):e1001134.
- Hakenbeck S. 2011. Local, regional and ethnic identities in early Medieval cemeteries in Bavaria. Firenze: All'Insegna del Giglio.
- Harbeck M, Seifert L, Hänsch S, Wagner DM, Birdsell D, Parise KL, Wiechmann I, Grupe G, Thomas A, Keim P. 2013. *Yersinia pestis* DNA from skeletal remains from the 6th century AD reveals insights into Justinianic Plague. *PLoS Pathogens* 9(5):e1003349.
- Hodges E, Rooks M, Xuan Z, Bhattacharjee A, Gordon DB, Brizuela L, McCombie WR, Hannon CJ. 2009. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nat. Protoc.* 4(6):960–974.
- Jonsson H, Ginolhac A, Schubert M, Johnson PL, Orlando L. 2013. mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29(13):1682–1684.
- Kanaroglou P, Delmelle E. 2015. Spatial analysis in health geography. United Kingdom: Ashgate Publishing.
- Keim PS, Wagner DM. 2009. Humans and evolutionary and ecological forces shaped the phylogeography of recently emerged diseases. *Nature Rev. Microbiol.* 7(11):813–821.
- Kircher M, Heyn P, Kelso J. 2011. Addressing challenges in the production and analysis of Illumina sequencing data. *BMC Genomics* 12:382. 2164–12–382.
- Kircher M, Sawyer S, Meyer M. 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* 40(1):e3.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
- Little LK. 2007. Plague and the end of antiquity: The pandemic of 541–750. UK: Cambridge University Press in association with the American Academy in Rome.
- Losert H, Pleterski A. 2003. Altenerding in Oberbayern. Struktur Des frühmittelalterlichen gräberfeldes und “Ethnogenese” der Bajuwaren. Berlin, Bamberg, Ljubljana. Berlin: Sripvaz/Založba ZRC.
- Maddicott J. 1997. Plague in seventh-century England. *Past and Present* 156(1):7–54.
- Marceau M. 2005. Transcriptional regulation in *Yersinia*: An update. *Curr Issues Mol Biol.* 7(2):151–178.
- Maricic T, Whitten M, Pääbo S. 2010. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* 5(11):e14004.
- McCormick M. 2015. Tracking mass death during the fall of rome's empire (I). *J Roman Archaeol.* 28:325–357.
- McCormick M. (forthcoming). Tracking mass death during the fall of rome's empire (II). *J Roman Archaeol.* 29.
- Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc.* 2010(6):pdb. prot5448.
- Mitchell S. 2014. A history of the later Roman Empire, AD 284–641. United Kingdom: John Wiley & Sons.
- Morelli G, Song Y, Mazzoni CJ, Eppinger M, Roumagnac P, Wagner DM, Feldkamp M, Kusecek B, Vogler AJ, Li Y, et al. 2010. *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat Genet.* 42(12):1140–1143.
- Motin VL, Georgescu AM, Elliott JM, Hu P, Worsham PL, Ott LL, Slezak TR, Sokhansanj BA, Regala WM, Brubaker RR, et al. 2002. Genetic variability of *Yersinia pestis* isolates as predicted by PCR-based IS100 genotyping and analysis of structural genes encoding glycerol-3-phosphate dehydrogenase (*glpD*). *J Bacteriol.* 184(4):1019–1027.
- Motin VL, Georgescu AM, Fitch JP, Gu PP, Nelson DO, Mabery SL, Garnham JB, Sokhansanj BA, Ott LL, Coleman MA, et al. 2004. Temporal global changes in gene expression during temperature transition in *Yersinia pestis*. *J Bacteriol.* 186(18):6298–6305.
- Parkhill J, Wren BW, Thomson NR, Titball RW, Holden MTG, Prentice MB, Sebahia M, James KD, Churcher C, Mungall KL, et al. 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* 413(6855):523–527.
- Parmenter RR, Yadav EP, Parmenter CA, Ettestad P, Gage KL. 1999. Incidence of plague associated with increased winter-spring precipitation in New Mexico. *Am J Trop Med Hyg.* 61(5):814–821.
- Perry RD, Fetherston JD. 1997. *Yersinia pestis*-etiologic agent of plague. *Clin Microbiol Rev.* 10(1):35–66.
- Pollitzer R. 1954. Plague. Geneva: WHO. 409–482.
- Radnedge L, Agron PG, Worsham PL, Andersen GL. 2002. Genome plasticity in *Yersinia pestis*. *Microbiology* 148(Pt 6):1687–1698.
- Rasmussen S, Allentoft ME, Nielsen K, Orlando L, Sikora M, Sjögren K, Pedersen AG, Schubert M, Van Dam A, Kapel CMO. 2015. Early

- divergent strains of *Yersinia pestis* in Eurasia 5,000 years ago. *Cell* 163(3):571–582.
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative Genomics Viewer. *Nat Biotechnol*. 29(1):24–26.
- Rohland N, Hofreiter M. 2007. Ancient DNA extraction from bones and teeth. *Nat Protoc*. 2(7):1756–1762.
- Schmid BV, Buntgen U, Easterday WR, Ginzler C, Walloe L, Bramanti B, Stenseth NC. 2015. Climate-driven introduction of the Black Death and successive plague reintroductions into Europe. *Proc Natl Acad Sci USA*. 112(10):3020–3025.
- Schuenemann VJ, Bos K, DeWitte S, Schmedes S, Jamieson J, Mittnik A, Forrest S, Coombes BK, Wood JW, Earn DJ, et al. 2011. Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of *Yersinia pestis* from victims of the Black Death. *Proc Natl Acad Sci USA*. 108(38):E746–E752.
- Scott S, Duncan CJ. 2001. Biology of plagues: evidence from historical populations. United Kingdom: Cambridge University Press.
- Sebbane F, Jarrett CO, Gardner D, Long D, Hinnebusch BJ. 2006. Role of the *Yersinia pestis* plasminogen activator in the incidence of distinct septicemic and bubonic forms of flea-borne plague. *Proc Natl Acad Sci USA*. 103(14):5526–5530.
- Seifert L, Harbeck M, Thomas A, Hoke N, Zöller L, Wiechmann I, Grupe G, Scholz HC, Riehm JM. 2013. Strategy for sensitive and specific detection of *Yersinia pestis* in skeletons of the Black Death pandemic. *PLoS One* 8(9):e75742.
- Shivaji S, Bhanu NV, Aggarwal R. 2000. Identification of *Yersinia pestis* as the causative organism of plague in India as determined by 16S rDNA sequencing and RAPD-based genomic fingerprinting. *FEMS Microbiol Lett*. 189(2):247–252.
- Sofeso C, Vohberger M, Wisnosky A, Paffgen 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). In: Burger J, Kaiser E, Schier W, editors. Population dynamics in pre- and Early History. New approaches by using stable isotopes and genetics. proceedings of the conference in Berlin, March 24–26, 2010. Berlin/Boston: De Gruyter. p. 113–130.
- Stathakopoulos DC. 2004. Famine and pestilence in the late Roman and early Byzantine Empire: A systematic survey of subsistence crises and epidemics. United Kingdom: Ashgate.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 30(12):2725–2729.
- Tomaso H, Reisinger EC, Dahouk S, Frangoulidis D, Rakin A, Landt O, Neubauer H. 2003. Rapid detection of *Yersinia pestis* with multiplex real-time PCR assays using fluorescent hybridisation probes. *FEMS Immunol Med Microbiol*. 38(2):117–126.
- Tong Z, Zhou D, Song Y, Zhang L, Pei D, Han Y, Pang X, Li M, Cui B, Wang J. 2005. Pseudogene accumulation might promote the adaptive microevolution of *Yersinia pestis*. *J Med Microbiol*. 54(3):259–268.
- Twigg G. 1984. The Black Death: a biological reappraisal. London: Batsford Academic and Educational.
- Vadyvaloo V, Jarrett C, Sturdevant DE, Sebbane F, Hinnebusch BJ. 2010. Transit through the flea vector induces a pretransmission innate immunity resistance phenotype in *Yersinia pestis*. *PLoS Pathogens* 6(2):e1000783.
- Wagner DM, Klunk J, Harbeck M, Devault A, Waglechner N, Sahl JW, Enk J, Birdsell DN, Kuch M, Lumibao C. 2014. *Yersinia pestis* and the Plague of Justinian 541–543 AD: a genomic analysis. *Lancet Infect Dis*. 14(4):319–326.
- Wiechmann I, Grupe G. 2005. Detection of *Yersinia pestis* DNA in two early Medieval skeletal finds from Aschheim (upper Bavaria, 6th century AD). *Am J Phys Anthropol*. 126(1):48–55.
- Willias SP, Chauhan S, Motin VL. 2014. Functional characterization of *Yersinia pestis* aerobic glycerol metabolism. *Microb Pathog*. 76:33–43.
- World Health Organization 2004. Human plague in 2002 and 2003. *Wkly Epidemiol Rec*. 79(301):6.
- Xu L, Stige LC, Kausrud KL, Ben Ari T, Wang S, Fang X, Schmid BV, Liu Q, Stenseth NC, Zhang Z. 2014. Wet climate and transportation routes accelerate spread of human plague. *Proc Biol Sci*. 281(1780):20133159.
- Yang DY, Eng B, Wayne JS, Dudar JC, Saunders SR. 1998. Technical note: Improved DNA extraction from ancient bones using silica-based spin columns. *Am J Phys Anthropol*. 105(4):539–543.
- Zhou D, Han Y, Qiu J, Qin L, Guo Z, Wang X, Song Y, Tan Y, Du Z, Yang R. 2006. Genome-wide transcriptional response of *Yersinia pestis* to stressful conditions simulating phagolysosomal environments. *Microb Infect*. 8(12):2669–2678.
- Zinser ER, Schneider D, Blot M, Kolter R. 2003. Bacterial evolution through the selective loss of beneficial genes. trade-offs in expression involving two loci. *Genetics* 164(4):1271–1277.

7. Discussion

7.1. The role of aDNA in recording history

By analyzing genomes recovered from ancient human remains spanning around 15,000 years, this thesis demonstrates how ancient DNA can help record history and shed light on natural history. Genetic data, such as reported here, are imperative when there are gaps in the traditional sources used for studying the past. In the context of prehistoric periods, traditional evidences are mostly gained through comparative analysis of material cultures, burial practices and metallurgy. In later periods, textual records are at times available, for example, documents authored by contemporaneous historians or archival material such as funerary inscriptions or legal documents. The literary sources are very valuable since they were composed by eye witnesses of the events but they are also strongly influenced by the point of view of the author. Thus, rigorous scientific investigation could also be advantageous when studying relatively well recorded historic events.

In manuscript A of this thesis, a 7,000 year long human genetic time series in central Anatolia was reconstructed. In this region, only few archaeological sites have been excavated, most of which did not yield human remains (Baird 2012, 431-465). This genetic time series encompasses the Neolithic transition in this region. Archaeological evidence of cultural continuity between the hunter-gatherers and the first farmers of central Anatolia has been suggested based on burial practices and artifact styles, however, these evidences are insufficient to resolve a long lasting debate on whether agriculture was brought to the region by a large movement of people who settled there, while adopting some local traditions or whether it was the local hunter-gatherers that adopted an agricultural subsistence strategy. The lack of Epipaleolithic wild crop findings in the region, suggests it is unlikely that cultivation independently developed in central Anatolia (Baird et al. 2018, E3077-E3086). Therefore, the second scenario would have involved a cultural process in which knowledge, plants and technology are exchanged through contact with earlier farming centers (Baird et al. 2018, E3077-E3086). The analysis described in paper A, revealed a high degree of genetic continuity between the hunter-gatherers of the region and the first farmers, suggesting a limited role for human mobility in introducing farming into central Anatolia. These findings comply with a cultural diffusion model and differ from the demic diffusion process, known to have facilitated the introduction of farming into Europe (Bramanti et al. 2009, 137-140; Haak et al. 2010, e1000536. doi:10.1371/journal.pbio.1000536;

Lazaridis et al. 2014, 409-413). Similar genetic studies in additional regions that played a role in the early spread of farming can help accumulate knowledge on the demographic characterizations of this process across the globe and flag regions in which farming developed locally. These regions would be key for addressing the question of why this subsistence strategy emerged.

In addition to the observed long-term persistence of the local gene pool in central Anatolia throughout the studied period, incoming gene flows were detected as well (Figure 1). An earlier one from a population related to ancient Iran/Caucasus and a later one related to Neolithic Levant replacing around 10% and 20% of the local gene pool, respectively. The earlier influx most likely involved eastern Anatolia, suggesting connections between these two regions, currently poorly documented by archaeology. Similarly, a detected genetic link with southern Europe that predated 15,000 yBP revealed a connection with southern Europe that was missing from the archaeological record.

The advantage of aDNA in tracing human mobility was only recently recognized. In fact, up to the last decade, the popular notion held by population geneticists was that prior to cross oceanic travel, human mobility was of a minor scale and thus had limited influence on the geographical distribution of human genetic variation. According to this perception, populations that were not involved in the migration events which postdated the 15th century AD properly represent the genetic makeup of the ancient people that lived in the same region (Cavalli-Sforza 2005, 333). This idea has been laid to rest by aDNA studies that detected substantive admixture events caused by massive human mobility which occurred as early as the Neolithic and the Bronze Age (Bramanti et al. 2009, 137-140; Skoglund et al. 2012, 466-469; Lazaridis et al. 2014, 409-413; Haak et al. 2015, 207-211).

In manuscript B of this thesis, human mobility large enough to replace up to a third of the local gene pool in the ancient port city of Ashkelon was detected. This detected gene flow coincided with cultural changes observed in the archaeological record during the Bronze- to Iron-Age transition, referred to as the “Philistine phenomenon” and was inferred to be south European related (Figure 1C). The “Philistine phenomenon” describes how in several coastal Levantine sites that were identified by textual sources as Philistine during the Iron Age, artifact and architecture styles began to resemble Bronze Age Aegean ones (Stager 1995, 332-348). Interestingly, this genetic signature is no longer detectable in the later Iron Age (although the late Iron Age people of Ashkelon are still culturally identified as Philistine) and appears to have been diluted by the local Levantine gene pool. The transient nature of the detected admixture signal

highlights the limitation in using recent DNA to infer past events as well as the need for dense sampling through time to make inferences based on aDNA. Nuanced and transient genetic shifts, such as the one reported in manuscript B, might be overlooked if the temporal gaps are too wide.

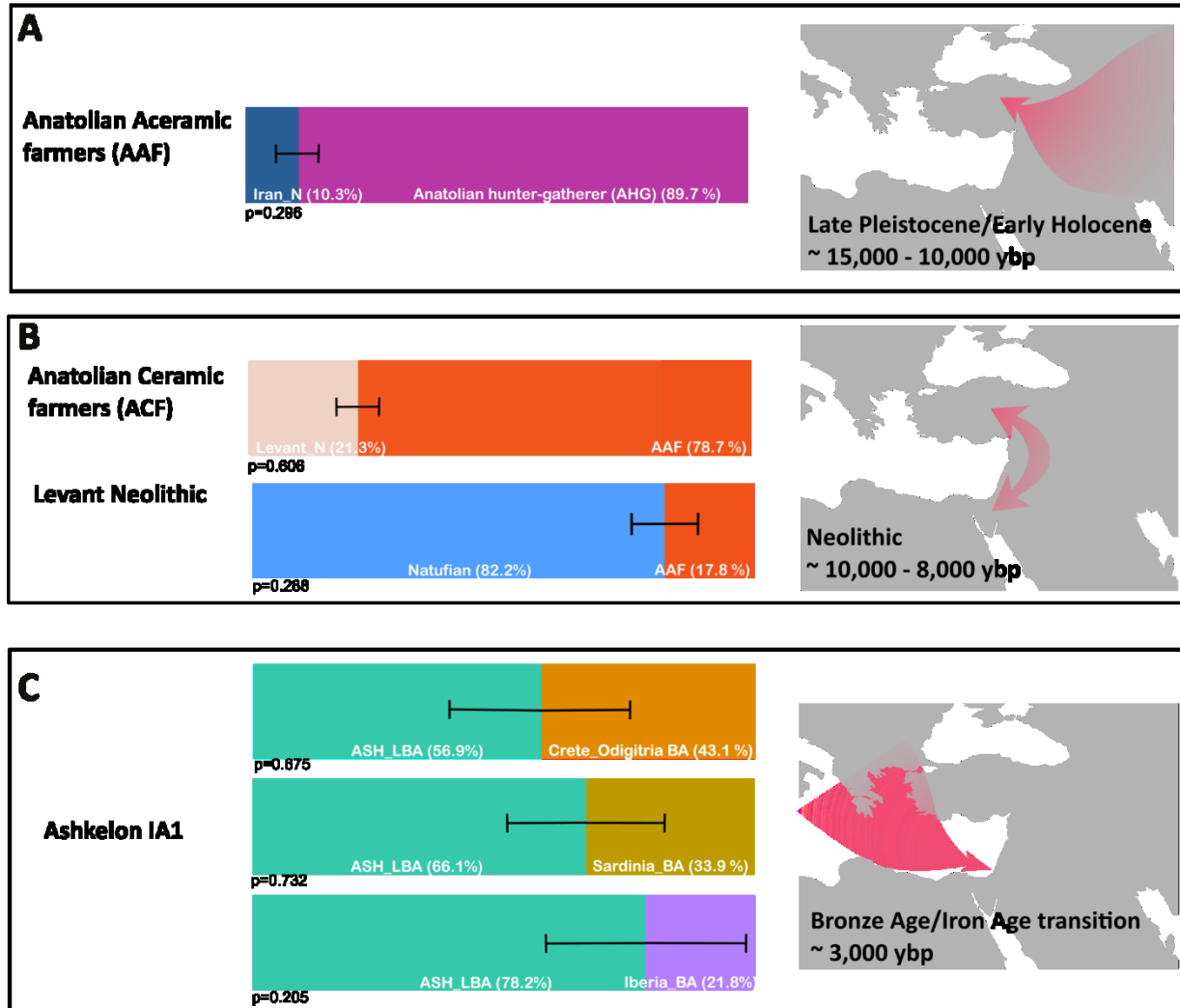


Figure 2: A summary of gene flows detected in this thesis and their proportion estimated by qpAdm modeling (Manuscripts A and B). The arrows schematically represent the direction of gene flow and not the actual rout of migration. The modeled population is annotated left of the models in each panel.

In the context of plague (Manuscript C), the archaeological record can mostly provide negative evidences such as abandonment of structures and marked breaks in construction and expansion (Little 2007). In historical times, literary sources and archival material can be useful but are absent in many cases (Little 2007). In manuscript C of this thesis, a genome of *Y. pestis*, the causative agent of the Justinianic Plague was recovered from a 6th century human skeleton,

buried in the town of Altenerding in southern Germany. There was no historical record of the Justinianic Plague ever reaching this region, stressing the importance of molecular evidence in tracing disease. The phylogenetic assignment of this historical strain confirmed this lineage lacked any sampled present-day descendants and supported the previously suggested south or central Asian origin suggested for this lineage (Wagner et al. 2014, 319-326).

Y. pestis has a distinct geographical clustering pattern, observed for both modern and ancient lineages, a characteristic that is useful for human history inference. Neither the natural hosts nor vectors of *Y. pestis* typically travel in a radius of more than few kilometers, thus long distance displacement of this pathogen can mostly be attributed to human mobility. The bacteria might be carried by infected people or the commensal animals that accompany them, such as the rat which has played a central role in spreading the historic plagues (McCormick 2007, 290-313).

In terms of aDNA authentication, *Y. pestis* holds an advantage over other pathogens since neither the pathogen nor its closely related organisms are abundant in soil and therefore the risk of environmentally derived erroneous sequence assignments is very low. In light of these characteristics, the systematic mapping and typing of *Y. pestis* genetic diversity through time and space could provide valuable information on human mobility, common trade routes and on links between historical populations (McCormick 2007, 290-313). Thus, widening the currently limited temporal and geographical available data by extensive sampling targeting *Y. pestis* as well as other pathogens that have similar characteristics could be useful not only to learn about the pathogen's natural history but also to help record history.

7.2. Challenges in DNA retrieval and interpretation

Our ability to retrieve DNA and reconstruct genome wide data has tremendously improved following the advent of methodological advancements such as NGS and enrichment techniques. Yet, this ability is limited by the availability of human remains, the total amount of DNA preserved in a sample and the level of intruding exogenous DNA. To date, the oldest genome wide data reported was reconstructed from a ~700,000 yBP horse preserved in permafrost conditions (Orlando et al. 2013, 74) and the oldest hominine genome was recovered from a Neanderthal, excavated in the Sima de los Huesos cave in Spain (Meyer et al. 2016, 504). In vitro experiments have suggested that DNA can survive only up to several hundred-thousands of years (Lindahl 1993, 709-715) while frozen conditions have been estimated to allow more than a million years of preservation (Allentoft et al. 2012, 4724-4733). One can therefore speculate

that while future technological advancements might improve our ability to better exhaust the retrieval of DNA preserved in a sample, we might be close to reaching the chronological barrier for aDNA recovery. DNA degradation is accelerated in conditions of high temperature, humidity and acidity (Lindahl 1993, 709-715). Accordingly, subtropical regions such as the southern Levant are less favorable for DNA preservation compared to temperate regions. In fact, the prospect of obtaining genome wide data from this region seemed unlikely up to the last three years (Lazaridis et al. 2016, 419; Haber et al. 2017, 274-282; Harney et al. 2018, 3336).

As part of this thesis (manuscripts A and B) 227 human skeletal elements were screened for human DNA retrieval (Figure 2). Around 8% of them yielded sufficient DNA for genome-wide data analysis. This proportion remains similar throughout the tested time periods (Figure 2A). However, since some periods are not fully represented in the geographical range of our data, further samples are needed to draw conclusions on the correlation of time and DNA preservation in the different regions. In accordance with previous reports (Pinhasi et al. 2015, e0129102), the *pars-petrosa* of the temporal bone show the highest DNA yields (~15%) compared to the pulp chamber of teeth (~3%) and other bones (~1%) (Figure 2B). As expected in higher latitude regions, the proportion of human aDNA in the Anatolian samples was on average higher than in the southern Levant (32.8 and 13.7% respectively). The fact that the southern Levantine samples were younger on average (Figure 2C) further supports that environmental conditions might play a bigger role than age in DNA preservation (Lindahl 1993, 709-715).

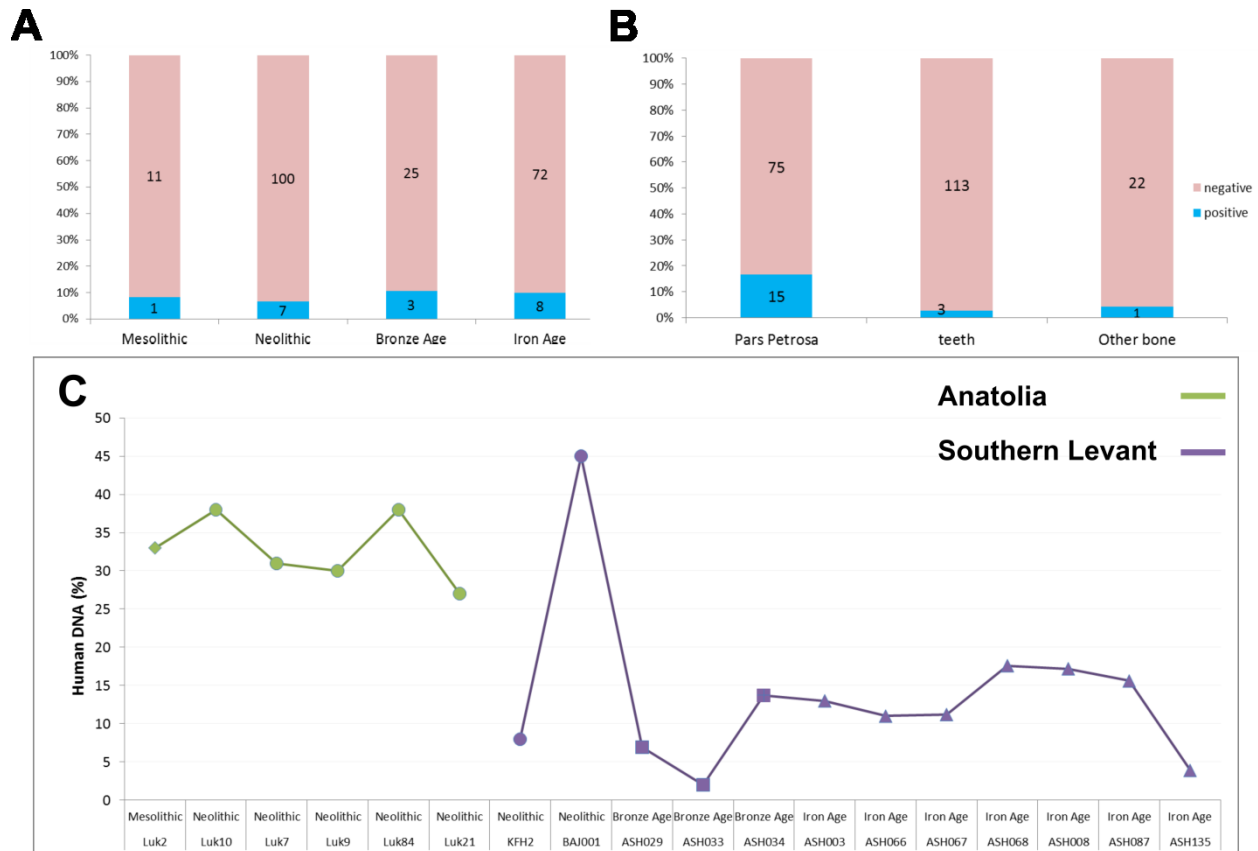


Figure 2: Human DNA preservation in 227 screened skeletal elements from the Near East and neighboring regions (A) The proportion of skeletal remains found positive for presence of human ancient DNA by period (skeletal remains were determined as positive when the shotgun sequenced corresponding partially UDG treated DNA library harbored at least 0.1% human DNA and 5' deamination levels higher than around 7%; the absolute number of skeletal elements is marked by the inner labels). **(B)** The proportion of skeletal remains found positive for presence of human ancient DNA by tissue type. In accordance to previous reports (Pinhasi et al. 2015, e0129102), out of the tested elements, the *pars petrosa* of the temporal bone shows the highest proportion of human DNA preservation. **(C)** The proportion of “on-target” human DNA in positive DNA libraries enriched for 1.24 million SNPs (‘1240K capture’) (Fu et al. 2013, 2223-2227; Haak et al. 2015, 207-211; Mathieson et al. 2015, 499-503). Skeletal elements from Central Anatolia show a higher on average human DNA proportion than contemporaneous southern and even later Levantine ones.

When attempting ancient pathogen genomic retrieval, the picture becomes more complex since often there is no clear indication that the pathogen was ever present in the studied tissue. In rare cases, diseases leave skeletal lesions that can be identified on the remains but mostly, as is the case for *Y. pestis*, there are no morphological traces left to indicate that the investigated individual was indeed infected. Some clues can be found in the written sources, if they exist, or by the burial manner, for example a multiple or mass burial might point to a catastrophic event.

However, usually, a broad screening approach is needed to detect samples that could potentially be used for whole genome retrieval. In manuscript C of this thesis, a PCR approach has been used as a screening tool to detect specific short DNA sequences of *Y. pestis* prior to enrichment and NGS. The choice of tissue for *Y. pestis* screening is also important, while high yields of human DNA are found in the inner ear part of the petrous bone, *Y. pestis* and other blood borne pathogens are unlikely to reach this tissue that is characterized with low blood supply (Margaryan et al. 2018, 3534-3542). The most commonly used tissue for this purpose is the pulp chamber of the tooth which is rich in blood supply as well as relatively shielded from the outside environment within the tooth. Even when there is high certainty that an individual was indeed infected with *Y. pestis*, viral loads might differ between tissues as well as between different elements of the same tissue type. In manuscript C, four teeth, sampled from two individuals (two teeth from each), screened positive for presence of *Y. pestis* but only one out of two teeth of one individual yielded sufficient data for genome reconstruction, demonstrating that the micro-environment can considerably affect DNA retrieval.

The limitations in aDNA retrieval described here, often lead to smaller sample sizes used in Archaeogenetic studies in comparison to modern ones. In many cases only few individuals or even only one are available to represent a human or a bacterial lineage. Indeed, a high number of genomes can increase the resolution and scope of possible analyses. Nonetheless, in the case of the human genome for example, recombination processes which mix maternal and paternal alleles over generations to the extent that every genome can be perceived as the gene pool of their recent ancestors. To fairly sample this gene pool, hundreds of thousands of markers across the genome can be ascertained and statistically tested (Patterson et al. 2012, 1065-1093; Fu et al. 2013, 2223-2227; Haak et al. 2015, 207-211).

In manuscripts A and B of this thesis, a limited availability of human remains and low preservation respectively, has allowed a relatively low number of individuals to be analyzed. Even so, clear signals of genetic continuity as well as of incoming gene flows could be detected.

Both manuscripts A and B, call for further sampling that would fill gaps in the genetic record of the studied regions and periods. To better understand the population structure in the Epipaleolithic and Neolithic (manuscript A), more genomes from central Anatolia would be of value as well as from surrounding regions such as eastern and southern Anatolia, the northern Levant, the Balkans, the Aegeans, Iran and the greater Mesopotamia. Of particular importance would be genomes that would fill the current 5,000 year gap between the Epipaleolithic hunter-

gatherer and the earliest Anatolian farmers. Such data would also help in determining the mode of transfer for the gene flows detected during the Epipaleolithic/Neolithic, for example whether short term massive movements of people carried these genes or long term isolation by distance events. Currently, the 15,000 yBP Anatolian hunter-gatherer genome is the oldest recovered in the Near East, older genomes could reveal the demographic processes involved in the formation of the late Pleistocene Near Eastern population structure.

In manuscript B, Late Bronze- to Late Iron- Age genomes from the southern Levant are reported and analyzed for the first time. To the resolution of our data, the Late Bronze Age genomes show high genetic similarity with earlier Bronze Age populations from the region but it cannot be excluded that analysis of more genomes that might be available with future sampling, could reveal subtle signals of gene flow. In contrast to the long term persistence of the Bronze Age gene pool, a genetic discontinuity is detected between the Bronze- and Iron- Age populations in the site of Ashkelon, caused by a gene flow of a European-related ancestry. Exhaustive sampling from additional regions in the Eastern Mediterranean in regions such as Cyprus, Sardinia and the Aegean are needed to compile a contemporaneous reference dataset that might identify more precisely the source of this incoming ancestry. Extensive sampling increases the chances of obtaining higher quality genomes potentially preserved due to favorable micro-environments. These could help perform more refined inferences compared to those possible so far, such as on relatedness levels and phenotypic traits within Near Eastern populations (Olalde et al. 2014, 225-228; Mittnik et al. 2016, e0163019).

The *Y. pestis* genome reconstructed as part of the study reported in manuscript C of this thesis is of relatively high coverage (18 fold average coverage and over 91 % of the genome covered more than 5 times) serving as an example of cases in which a high coverage allows for a more precise and detailed analysis. The high quality of the genome enabled the detection of previously undescribed substitutions and structural differences between the Justinianic strain and later strains as well as some erroneous variants previously reported for another Justinianic *Y. pestis* draft genome (Wagner et al. 2014, 319-326). The novel unique features require empirical testing to determine their significance in *Y. pestis* physiology. The corrections and augmentations of the above mentioned draft genome helped in establishing working criteria for analysis in the field that would prevent erroneous conclusions due to low quality data (Key et al. 2017, 508-520; Warinner et al. 2017, 321-356). In addition, the published high quality genome data from

manuscript C, serves as a valuable resource for the research community when combined with other data for analyses such as molecular dating (Spyrou et al. 2016, 874-881; Valtueña et al. 2017, 3683-3691. e8; Spyrou et al. 2018, 2234).

Notably, only two Justinianic *Y. pestis* genomes were published to date, both recovered from skeletons buried in southern Germany (Wagner et al. 2014, 319-326; Feldman et al. 2016, 2911-2923). Therefore, the diversity of *Y. pestis* during this pandemic has so far barely been explored. Furthermore, a south Asian sampling bias is apparent in the present day isolates, hence future sampling that would expand both geographical and temporal scopes including those of modern genomes might elucidate questions regarding the source of the pandemic, the mode of transfer and the reason it ceased during the 8th century AD.

Availability of ancient DNA data has been dramatically increasing over the last decade due to the technological advancements described in this thesis and due to laboratory automation processes that further increase the efficiency and cost effectiveness of these methods. Thus, it might be feasible to gradually fill the described sampling gaps in the near future. With increasing amounts of data we are faced with new frontiers regarding the computational processing and storage of these so called “big data” (Papageorgiou et al. 2018). Models and tools adopted from the field of data science may further improve our ability to exhaust the potential of the sequenced data. At the same time, basic science and method developments are still crucial to better target micro-environments and tissues favorable for DNA preservation (Pinhasi et al. 2015, e0129102).

8. References

- Achtman, M., G. Morelli, P. Zhu, T. Wirth, I. Diehl, B. Kusecek, A. J. Vogler, et al. 2004. "Microevolution and History of the Plague Bacillus, *Y. pestis*." *Proceedings of the National Academy of Sciences of the United States of America* 101 (51): 17837-17842.
- Achtman, M., K. Zurth, G. Morelli, G. Torrea, A. Guiyoule, and E. Carniel. 1999. "*Y. pestis*, the Cause of Plague, is a Recently Emerged Clone of *Yersinia Pseudotuberculosis*." *Proceedings of the National Academy of Sciences of the United States of America* 96 (24): 14043-14048.
- Alexander, D. H., J. Novembre, and K. Lange. 2009. "Fast Model-Based Estimation of Ancestry in Unrelated Individuals." *Genome Research* 19 (9): 1655-1664.
- Allentoft, Morten E., Matthew Collins, David Harker, James Haile, Charlotte L. Oskam, Marie L. Hale, Paula F. Campos, Jose A. Samaniego, M. Thomas P. Gilbert, and Eske Willerslev. 2012. "The Half-Life of DNA in Bone: Measuring Decay Kinetics in 158 Dated Fossils." *Proceedings of the Royal Society B: Biological Sciences* 279 (1748): 4724-4733.
- Allentoft, Morten E., Martin Sikora, Karl-Göran Sjögren, Simon Rasmussen, Morten Rasmussen, Jesper Stenderup, Peter B. Damgaard, Hannes Schroeder, Torbjörn Ahlström, and Lasse Vinner. 2015. "Population Genomics of Bronze Age Eurasia." *Nature* 522 (7555): 167.
- Armelagos, George J., Kathleen C. Barnes, and James Lin. 1996. "Disease in Human Evolution: The Reemergence of Infectious Disease in the Third Epidemiological Transition."
- Austin, Jeremy J., Andrew J. Ross, Andrew B. Smith, Richard A. Fortey, and Richard H. Thomas. 1997. "Problems of Reproducibility—does Geologically Ancient DNA Survive in Amber—preserved Insects?" *Proceedings of the Royal Society of London. Series B: Biological Sciences* 264 (1381): 467-474.
- Bacot, A. W. and C. J. Martin. 1914. "LXVII. Observations on the Mechanism of the Transmission of Plague by Fleas." *The Journal of Hygiene* 13 (Suppl): 423-439.
- Baird, Douglas. 2012. "The Late Epipaleolithic, Neolithic, and Chalcolithic of the Anatolian Plateau, 13,000–4000 BC." *A Companion to the Archaeology of the Ancient Near East*: 431-465.
- Baird, D., A. Fairbairn, E. Jenkins, L. Martin, C. Middleton, J. Pearson, E. Asouti, et al. 2018. "Agricultural Origins on the Anatolian Plateau." *Proceedings of the National Academy of Sciences of the United States of America* 115 (14): E3077-E3086.
- Barker, Graeme. 2009. *The Agricultural Revolution in Prehistory: Why did Foragers Become Farmers?* Oxford University Press on Demand.

- Bauer, Alexander A. 1998. "Cities of the Sea: Maritime Trade and the Origin of Philistine Settlement in the Early Iron Age Southern Levant." *Oxford Journal of Archaeology* 17 (2): 149-168.
- Beard, Brian L. and Clark M. Johnson. 2000. "Strontium Isotope Composition of Skeletal Material can Determine the Birth Place and Geographic Mobility of Humans and Animals." *Journal of Forensic Science* 45 (5): 1049-1061.
- Bellwood, Peter. 2005. "Examining the Farming/Language Dispersal Hypothesis in the East Asian Context." In *The Peopling of East Asia*, 41-54: Routledge.
- Benedictow, Ole Jørgen. 2004. *The Black Death, 1346-1353: The Complete History*. England: Boydell & Brewer.
- Bentley, David R., Shankar Balasubramanian, Harold P. Swerdlow, Geoffrey P. Smith, John Milton, Clive G. Brown, Kevin P. Hall, Dirk J. Evers, Colin L. Barnes, and Helen R. Bignell. 2008. "Accurate Whole Human Genome Sequencing using Reversible Terminator Chemistry." *Nature* 456 (7218): 53.
- Bocquet-Appel, Jean-Pierre and Ofer Bar-Yosef. 2008. *The Neolithic Demographic Transition and its Consequences* Springer Science & Business Media.
- Bos, Kirsten I., Verena J. Schuenemann, G. B. Golding, Hernan A. Burbano, Nicholas Waglechner, Brian K. Coombes, Joseph B. McPhee, et al. 2011. "A Draft Genome of *Y. pestis* from Victims of the Black Death." *Nature* 478 (7370): 506-510.
- Bos, Kirsten I., Alexander Herbig, Jason Sahl, Nicholas Waglechner, Mathieu Fourment, Stephen A. Forrest, Jennifer Klunk, Verena J. Schuenemann, Debi Poinar, and Melanie Kuch. 2016. "Eighteenth Century *Y. pestis* Genomes Reveal the Long-Term Persistence of an Historical Plague Focus." *eLife*: e12994.
- Bramanti, B., M. G. Thomas, W. Haak, M. Unterlaender, P. Jores, K. Tambets, I. Antanaitis-Jacobs, et al. 2009. "Genetic Discontinuity between Local Hunter-Gatherers and Central Europe's First Farmers." *Science (New York, N.Y.)* 326 (5949): 137-140.
- Brandt, G., W. Haak, C. J. Adler, C. Roth, A. Szecsenyi-Nagy, S. Karimnia, S. Moller-Rieker, et al. 2013. "Ancient DNA Reveals Key Stages in the Formation of Central European Mitochondrial Genetic Diversity." *Science (New York, N.Y.)* 342 (6155): 257-261.
- Briggs, A. W., U. Stenzel, P. L. Johnson, R. E. Green, J. Kelso, K. Prufer, M. Meyer, et al. 2007. "Patterns of Damage in Genomic DNA Sequences from a Neandertal." *Proceedings of the National Academy of Sciences of the United States of America* 104 (37): 14616-14621.
- Broushaki, F., M. G. Thomas, V. Link, S. Lopez, L. van Dorp, K. Kirsanow, Z. Hofmanova, et al. 2016. "Early Neolithic Genomes from the Eastern Fertile Crescent." *Science (New York, N.Y.)* 353 (6298): 499-503.

- Brubaker, Robert R. 1972. "The Genus *Yersinia*: Biochemistry and Genetics of Virulence with 3 Figures." In *Current Topics in Microbiology and Immunology*, 111-158: Springer.
- Buckley, Michael, Nigel D. Melton, and Janet Montgomery. 2013. "Proteomics Analysis of Ancient Food Vessel Stitching Reveals > 4000-year-old Milk Protein." *Rapid Communications in Mass Spectrometry* 27 (4): 531-538.
- Burbano, H. A., E. Hodges, R. E. Green, A. W. Briggs, J. Krause, M. Meyer, J. M. Good, et al. 2010. "Targeted Investigation of the Neandertal Genome by Array-Based Sequence Capture." *Science (New York, N.Y.)* 328 (5979): 723-725.
- Cameron, Averil. 1970. *Agathias* Clarendon Press.
- Cano, R. J. and M. K. Borucki. 1995. "Revival and Identification of Bacterial Spores in 25- to 40-Million-Year-Old Dominican Amber." *Science (New York, N.Y.)* 268 (5213): 1060-1064.
- Cano, R. J., M. K. Borucki, M. Higby-Schweitzer, H. N. Poinar, G. O. Poinar Jr, and K. J. Pollard. 1994. "Bacillus DNA in Fossil Bees: An Ancient Symbiosis?" *Applied and Environmental Microbiology* 60 (6): 2164-2167.
- Cavalli-Sforza, L. Luca. 2005. "The Human Genome Diversity Project: Past, Present and Future." *Nature Reviews Genetics* 6 (4): 333.
- Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza. 1993. "Demic Expansions and Human Evolution." *Science (New York, N.Y.)* 259 (5095): 639-646.
- Chain, P. S., E. Carniel, F. W. Larimer, J. Lamerdin, P. O. Stoutland, W. M. Regala, A. M. Georgescu, et al. 2004. "Insights into the Evolution of *Y. pestis* through Whole-Genome Comparison with *Yersinia Pseudotuberculosis*." *Proceedings of the National Academy of Sciences of the United States of America* 101 (38): 13826-13831.
- Chen, Feng-Chi and Wen-Hsiung Li. 2001. "Genomic Divergences between Humans and Other Hominoids and the Effective Population Size of the Common Ancestor of Humans and Chimpanzees." *The American Journal of Human Genetics* 68 (2): 444-456.
- Christie, Andrew Barnett. 1980. *Infectious Diseases: Epidemiology and Clinical Practice*. Churchill Livingstone, Robert Stevenson House, 1-3 Baxter's Place, Leith Walk, Edinburgh EH1 3AF.
- Cline, Eric H. 2015. *1177 BC: The Year Civilization Collapsed* Princeton University Press.
- Cohn, Samuel, K. 2008. "Epidemiology of the Black Death and Successive Waves of Plague." *Medical History* 52: 74-100.
- Cooper, A. and H. N. Poinar. 2000. "Ancient DNA: Do it Right Or Not at All." *Science (New York, N.Y.)* 289 (5482): 1139.

- Cui, Y., C. Yu, Y. Yan, D. Li, Y. Li, T. Jombart, L. A. Weinert, et al. 2013. "Historical Variations in Mutation Rate in an Epidemic Pathogen, *Y. pestis*." *Proceedings of the National Academy of Sciences of the United States of America* 110 (2): 577-582.
- Dawkins, Richard and John Richard Krebs. 1979. "Arms Races between and within Species." *Proceedings of the Royal Society of London. Series B. Biological Sciences* 205 (1161): 489-511.
- de Barros Damgaard, P., R. Martiniano, J. Kamm, J. V. Moreno-Mayar, G. Kroonen, M. Peyrot, G. Barjamovic, et al. 2018. "The First Horse Herders and the Impact of Early Bronze Age Steppe Expansions into Asia." *Science (New York, N.Y.)* 360 (6396): 10.1126/science.aar7711. Epub 2018 May 9.
- de Barros Damgaard, Peter, Nina Marchi, Simon Rasmussen, Michaël Peyrot, Gabriel Renaud, Thorfinn Korneliussen, J. Víctor Moreno-Mayar, Mikkel Winther Pedersen, Amy Goldberg, and Emma Usmanova. 2018. "137 Ancient Human Genomes from Across the Eurasian Steppes." *Nature* 557 (7705): 369.
- Procopius, Henry Bronson Dewing, and Robert Emory Glanville Downey. *Procopius. With an English Translation by HB Dewing. (vol. 7. With the Collaboration of Glanville Downey.) [With Plates.]*. London; Macmillan Company: New York, 1914.
- Dothan, Trude and Moshe Dothan. 1992. *People of the Sea: The Search for the Philistines*. Macmillan New York.
- Duncan, C. J. and S. Scott. 2005. *What Caused the Black Death?*. Vol. 81.
- Durand, Eric Y., Nick Patterson, David Reich, and Montgomery Slatkin. 2011. "Testing for Ancient Admixture between Closely Related Populations." *Molecular Biology and Evolution* 28 (8): 2239-2252.
- Eglinton, Geoffrey and Graham A. Logan. 1991. "Molecular Preservation." *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 333 (1268): 315-328.
- Feldman, Michal, Michaela Harbeck, Marcel Keller, Maria A. Spyrou, Andreas Rott, Bernd Trautmann, Holger C. Scholz, Bernd Pääfgen, Joris Peters, and Michael McCormick. 2016. "A High-Coverage *Y. pestis* Genome from a Sixth-Century Justinianic Plague Victim." *Molecular Biology and Evolution* 33 (11): 2911-2923.
- Frobenius, Leo. 1897. "Der Westafrikanische Kulturkreis." *Petermanns. Geogr. Mitt.* 43: 225-236, 262-267.
- Fu, Qiaomei, Mateja Hajdinjak, Oana Teodora Moldovan, Silviu Constantin, Swapan Mallick, Pontus Skoglund, Nick Patterson, Nadin Rohland, Iosif Lazaridis, and Birgit Nickel. 2015. "An Early Modern Human from Romania with a Recent Neanderthal Ancestor." *Nature* 524 (7564): 216-219.

- Fu, Qiaomei, Heng Li, Priya Moorjani, Flora Jay, Sergey M. Slepchenko, Aleksei A. Bondarev, Philip LF Johnson, Ayinuer Aximu-Petri, Kay Prüfer, and Cesare de Filippo. 2014. "Genome Sequence of a 45,000-Year-Old Modern Human from Western Siberia." *Nature* 514 (7523): 445-449.
- Fu, Q., M. Meyer, X. Gao, U. Stenzel, H. A. Burbano, J. Kelso, and S. Paabo. 2013. "DNA Analysis of an Early Modern Human from Tianyuan Cave, China." *Proceedings of the National Academy of Sciences of the United States of America* 110 (6): 2223-2227.
- Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, et al. 2010. "A Draft Sequence of the Neandertal Genome." *Science (New York, N.Y.)* 328 (5979): 710-722.
- Haak, W., O. Balanovsky, JJ Sanchez, S. Koshel, and V. Zaporozhchenko. 2010. "Ancient DNA from European Early Neolithic Farmers Reveals their Near Eastern Affinities." *PLoS Biol* 8 (11): e1000536. doi:10.1371/journal.pbio.1000536.
- Haak, W., I. Lazaridis, N. Patterson, N. Rohland, S. Mallick, B. Llamas, G. Brandt, et al. 2015. "Massive Migration from the Steppe was a Source for Indo-European Languages in Europe." *Nature* 522 (7555): 207-211.
- Haber, Marc, Claude Doumet-Serhal, Christiana Scheib, Yali Xue, Petr Danecek, Massimo Mezzavilla, Sonia Youhanna, Rui Martiniano, Javier Prado-Martinez, and Michał Szpak. 2017. "Continuity and Admixture in the Last Five Millennia of Levantine History from Ancient Canaanite and Present-Day Lebanese Genome Sequences." *The American Journal of Human Genetics* 101 (2): 274-282.
- Haensch, Stephanie, Raffaella Bianucci, Michel Signoli, Minoarisoa Rajerison, Michael Schultz, Sacha Kacki, Marco Vermunt, Darlene A. Weston, Derek Hurst, and Mark Achtman. 2010. "Distinct Clones of *Y. pestis* Caused the Black Death." *PLoS Pathogens* 6 (10): e1001134.
- Harbeck, Michaela, Lisa Seifert, Stephanie Hänsch, David M. Wagner, Dawn Birdsell, Katy L. Parise, Ingrid Wiechmann, Gisela Grupe, Astrid Thomas, and Paul Keim. 2013. "*Y. pestis* DNA from Skeletal Remains from the 6th Century AD Reveals Insights into Justinianic Plague." *PLoS Pathogens* 9 (5): e1003349.
- Harney, Éadaoin, Hila May, Dina Shalem, Nadin Rohland, Swapan Mallick, Iosif Lazaridis, Rachel Sarig, Kristin Stewardson, Susanne Nordenfelt, and Nick Patterson. 2018. "Ancient DNA from Chalcolithic Israel Reveals the Role of Population Mixture in Cultural Transformation." *Nature Communications* 9 (1): 3336.
- Henry, A. G., A. S. Brooks, and D. R. Piperno. 2011. "Microfossils in Calculus Demonstrate Consumption of Plants and Cooked Foods in Neanderthal Diets (Shanidar III, Iraq; Spy I and II, Belgium)." *Proceedings of the National Academy of Sciences of the United States of America* 108 (2): 486-491.

- Hershberg, Ruth and Dmitri A. Petrov. 2010. "Evidence that Mutation is Universally Biased Towards AT in Bacteria." *PLoS Genetics* 6 (9): e1001115.
- Higuchi, Russell, Barbara Bowman, Mary Freiberger, Oliver A. Ryder, and Allan C. Wilson. 1984. "DNA Sequences from the Quagga, an Extinct Member of the Horse Family." *Nature* 312 (5991): 282.
- Hinnebusch, J. and T. G. Schwan. 1993. "New Method for Plague Surveillance using Polymerase Chain Reaction to Detect *Y. pestis* in Fleas." *Journal of Clinical Microbiology* 31 (6): 1511-1514.
- Hofreiter, Michael, David Serre, Hendrik N. Poinar, Melanie Kuch, and Svante Pääbo. 2001. "Ancient DNA." *Nature Reviews Genetics* 2 (5): 353.
- Hollingsworth, T. H. 1969. *Historical Demography*. NY: Cornell University Press.
- Hull, H. F., J. M. Montes, and J. M. Mann. 1987. "Septicemic Plague in New Mexico." *The Journal of Infectious Diseases* 155 (1): 113-118.
- International HapMap Consortium. 2007. "A Second Generation Human Haplotype Map of Over 3.1 Million SNPs." *Nature* 449 (7164): 851.
- Karlsson, Elinor K., Dominic P. Kwiatkowski, and Pardis C. Sabeti. 2014. "Natural Selection and Infectious Disease in Human Populations." *Nature Reviews Genetics* 15 (6): 379.
- Key, Felix M., Cosimo Posth, Johannes Krause, Alexander Herbig, and Kirsten I. Bos. 2017. "Mining Metagenomic Data Sets for Ancient DNA: Recommended Protocols for Authentication." *Trends in Genetics* 33 (8): 508-520.
- Kilinc, G. M., D. Koptekin, C. Atakuman, A. P. Sumer, H. M. Donertas, R. Yaka, C. C. Bilgin, et al. 2017. "Archaeogenomic Analysis of the First Steps of Neolithization in Anatolia and the Aegean." *Proceedings Biological Sciences* 284 (1867): 10.1098/rspb.2017.2064.
- Killebrew, Ann E. 1998. "Ceramic Typology and Technology of Late Bronze II and Iron I Assemblages from Tel Mique-Ekron: The Transition from Canaanite to Philistine Culture." *Mediterranean Peoples in Transition: Thirteenth to Early Tenth Centuries BCE*: 379-405.
- Killebrew, Ann E. 2013. *The Philistines and Other "sea Peoples" in Text and Archaeology*. Society of Biblical Lit.
- Kircher, Martin, Susanna Sawyer, and Matthias Meyer. 2011. "Double Indexing Overcomes Inaccuracies in Multiplex Sequencing on the Illumina Platform." *Nucleic Acids Research* 40 (1): e3-e3.
- Korneliussen, Thorfinn Sand, Anders Albrechtsen, and Rasmus Nielsen. 2014. "ANGSD: Analysis of Next Generation Sequencing Data." *BMC Bioinformatics* 15 (1): 356.

- Krause, Johannes, Qiaomei Fu, Jeffrey M. Good, Bence Viola, Michael V. Shunkov, Anatoli P. Derevianko, and Svante Pääbo. 2010. "The Complete Mitochondrial DNA Genome of an Unknown Hominin from Southern Siberia." *Nature* 464 (7290): 894.
- Lazaridis, Iosif, Alissa Mittnik, Nick Patterson, Swapan Mallick, Nadin Rohland, Saskia Pfrenkle, Anja Furtwängler, Alexander Peltzer, Cosimo Posth, and Andonis Vasilakis. 2017. "Genetic Origins of the Minoans and Mycenaeans." *Nature* 548 (7666): 214.
- Lazaridis, Iosif, Dani Nadel, Gary Rollefson, Deborah C. Merrett, Nadin Rohland, Swapan Mallick, Daniel Fernandes, Mario Novak, Beatriz Gamarra, and Kendra Sirak. 2016. "Genomic Insights into the Origin of Farming in the Ancient Near East." *Nature* 536 (7617): 419.
- Lazaridis, I., N. Patterson, A. Mittnik, G. Renaud, S. Mallick, K. Kirsanow, P. H. Sudmant, et al. 2014. "Ancient Human Genomes Suggest Three Ancestral Populations for Present-Day Europeans." *Nature* 513 (7518): 409-413.
- Leakey, Mary D. and Richard L. Hay. 1979. "Pliocene Footprints in the Laetoli Beds at Laetoli, Northern Tanzania." *Nature* 278 (5702): 317-323.
- Libby, Willard F. 1954. "Radiocarbon Dating." *Endeavour* 13 (49): 5-16.
- Lindahl, Tomas. 1993. "Instability and Decay of the Primary Structure of DNA." *Nature* 362 (6422): 709-715.
- Little, Lester K. 2007. *Plague and the End of Antiquity: The Pandemic of 541–750* Cambridge University Press in association with the American Academy in Rome.
- Maeir, Aren M. and Louise A. Hitchcock. 2017. "The Appearance, Formation and Transformation of Philistine Culture: New Perspectives and New Finds." *The Sea Peoples Up-to-Date: New Research on the Migration of Peoples in the 12th Century BCE*: 149-162.
- Mallick, Swapan, Heng Li, Mark Lipson, Iain Mathieson, Melissa Gymrek, Fernando Racimo, Mengyao Zhao, Niru Chennagiri, Susanne Nordenfelt, and Arti Tandon. 2016. "The Simons Genome Diversity Project: 300 Genomes from 142 Diverse Populations." *Nature* 538 (7624): 201.
- Mardis, Elaine R. 2008. "Next-Generation DNA Sequencing Methods." *Annu.Rev.Genomics Hum.Genet.* 9: 387-402.
- Margaryan, Ashot, Henrik B. Hansen, Simon Rasmussen, Martin Sikora, Vyacheslav Moiseyev, Alexandr Khoklov, Andrey Epimakhov, Levon Yepiskoposyan, Aivar Kriiska, and Liivi Varul. 2018. "Ancient Pathogen DNA in Human Teeth and Petrous Bones." *Ecology and Evolution* 8 (6): 3534-3542.

- Margulies, Marcel, Michael Egholm, William E. Altman, Said Attiya, Joel S. Bader, Lisa A. Bembien, Jan Berka, Michael S. Braverman, Yi-Ju Chen, and Zhoutao Chen. 2005. "Genome Sequencing in Microfabricated High-Density Picolitre Reactors." *Nature* 437 (7057): 376.
- Maricic, Tomislav, Mark Whitten, and Svante Pääbo. 2010. "Multiplexed DNA Sequence Capture of Mitochondrial Genomes using PCR Products." *PloS One* 5 (11): e14004.
- Mathieson, Iain, Iosif Lazaridis, Nadin Rohland, Swapan Mallick, Nick Patterson, Songül Alpaslan Roodenberg, Eadaoin Harney, Kristin Stewardson, Daniel Fernandes, and Mario Novak. 2015. "Genome-Wide Patterns of Selection in 230 Ancient Eurasians." *Nature* 528 (7583): 499-503.
- McCormick, M. 2007. "Toward a Molecular History of the Justinianic Pandemic." In *Plague and the End of Antiquity, the Pandemic of 541-750*, edited by L. Little, 290-313: Cambridge University Press.
- McCormick, Michael. 2003. "Rats, Communications, and Plague: Toward an Ecological History." *Journal of Interdisciplinary History* 34 (1): 1-25.
- Meyer, Matthias, Juan-Luis Arsuaga, Cesare de Filippo, Sarah Nagel, Ayinuer Aximu-Petri, Birgit Nickel, Ignacio Martínez, Ana Gracia, de Castro, José María Bermúdez, and Eudald Carbonell. 2016. "Nuclear DNA Sequences from the Middle Pleistocene Sima De Los Huesos Hominins." *Nature* 531 (7595): 504.
- Meyer, Matthias and Martin Kircher. 2010. "Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing." *Cold Spring Harbor Protocols* 2010 (6): pdb.prot5448.
- Mitnik, Alissa, Chuan-Chao Wang, Jiří Svoboda, and Johannes Krause. 2016. "A Molecular Approach to the Sexing of the Triple Burial at the Upper Paleolithic Site of Dolní Věstonice." *PloS One* 11 (10): e0163019.
- mondiale de la Santé, Organisation and World Health Organization. 2016. "Plague Around the World, 2010–2015." *Weekly Epidemiological Record= Relevé Épidémiologique Hebdomadaire* (8): 89-93.
- Montagu, Montague Francis. 1960. "An Introduction to Physical Anthropology."
- Morelli, Giovanna, Yajun Song, Camila J. Mazzoni, Mark Eppinger, Philippe Roumagnac, David M. Wagner, Mirjam Feldkamp, et al. 2010. "*Y. pestis* Genome Sequencing Identifies Patterns of Global Phylogenetic Diversity." *Nature Genetics* 42 (12): 1140-1143.
- Novembre, John, Toby Johnson, Katarzyna Bryc, Zoltán Kutalik, Adam R. Boyko, Adam Auton, Amit Indap, Karen S. King, Sven Bergmann, and Matthew R. Nelson. 2008. "Genes Mirror Geography within Europe." *Nature* 456 (7218): 98.

- Olalde, Inigo, Morten E. Allentoft, Federico Sánchez-Quinto, Gabriel Santpere, Charleston WK Chiang, Michael DeGiorgio, Javier Prado-Martinez, Juan Antonio Rodríguez, Simon Rasmussen, and Javier Quilez. 2014. "Derived Immune and Ancestral Pigmentation Alleles in a 7,000-Year-Old Mesolithic European." *Nature* 507 (7491): 225-228.
- Orlando, Ludovic, Aurélien Ginolhac, Guojie Zhang, Duane Froese, Anders Albrechtsen, Mathias Stiller, Mikkel Schubert, Enrico Cappellini, Bent Petersen, and Ida Moltke. 2013. "Recalibrating Equus Evolution using the Genome Sequence of an Early Middle Pleistocene Horse." *Nature* 499 (7456): 74.
- Pääbo, Svante. 1989. "Ancient DNA: Extraction, Characterization, Molecular Cloning, and Enzymatic Amplification." *Proceedings of the National Academy of Sciences* 86 (6): 1939-1943.
- Pääbo, Svante. 1985. "Molecular Cloning of Ancient Egyptian Mummy DNA." *Nature* 314 (6012): 644.
- Pääbo, Svante, Hendrik Poinar, David Serre, Viviane Jaenicke-Després, Juliane Hebler, Nadin Rohland, Melanie Kuch, Johannes Krause, Linda Vigilant, and Michael Hofreiter. 2004. "Genetic Analyses from Ancient DNA." *Annu.Rev.Genet.* 38: 645-679.
- Pääbo, Svante and Allan C. Wilson. 1988. "Polymerase Chain Reaction Reveals Cloning Artefacts." *Nature* 334 (6181): 387.
- Papageorgiou, Louis, Picas Eleni, Sofia Raftopoulou, Meropi Mantaïou, Vasileios Megalooikonomou, and Dimitrios Vlachakis. 2018. "Genomic Big Data Hitting the Storage Bottleneck." *EMBnet.Journal* 24.
- Patterson, N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, and D. Reich. 2012. "Ancient Admixture in Human History." *Genetics* 192 (3): 1065-1093.
- Perry, R. D. and J. D. Fetherston. 1997. "*Y. pestis*-Etiologic Agent of Plague." *Clinical Microbiology Reviews* 10 (1): 35-66.
- Pinhasi, Ron, Daniel Fernandes, Kendra Sirak, Mario Novak, Sarah Connell, Songül Alpaslan-Roodenberg, Fokke Gerritsen, Vyacheslav Moiseyev, Andrey Gromov, and Pál Raczky. 2015. "Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone." *PloS One* 10 (6): e0129102.
- Plague Research Commission. 1907. "The Epidemiological Observations made by the Commission in Bombay City." *Journal of Hygiene* 7: 724-798.
- Poinar, Hendrik, Melanie Kuch, Gregory McDonald, Paul Martin, and Svante Pääbo. 2003. "Nuclear Gene Sequences from a Late Pleistocene Sloth Coprolite." *Current Biology* 13 (13): 1150-1152.

- Poland, Jack D. and David T. Dennis. 1998. "Plague." In *Bacterial Infections of Humans*, 545-558: Springer.
- Pollitzer, Richard. 1954. "Plague." *Geneva: WHO*: 409-482.
- Posth, Cosimo, Christoph Wißing, Keiko Kitagawa, Luca Pagani, Laura van Holstein, Fernando Racimo, Kurt Wehrberger, Nicholas J. Conard, Claus Joachim Kind, and Hervé Bocherens. 2017. "Deeply Divergent Archaic Mitochondrial Genome Provides Lower Time Boundary for African Gene Flow into Neanderthals." *Nature Communications* 8: 16046.
- Prüfer, Kay, Fernando Racimo, Nick Patterson, Flora Jay, Sriram Sankararaman, Susanna Sawyer, Anja Heinze, Gabriel Renaud, Peter H. Sudmant, and Cesare De Filippo. 2014. "The Complete Genome Sequence of a Neanderthal from the Altai Mountains." *Nature* 505 (7481): 43.
- Raghavan, Maanasa, Pontus Skoglund, Kelly E. Graf, Mait Metspalu, Anders Albrechtsen, Ida Moltke, Simon Rasmussen, Thomas W. Stafford Jr, Ludovic Orlando, and Ene Metspalu. 2014. "Upper Palaeolithic Siberian Genome Reveals Dual Ancestry of Native Americans." *Nature* 505 (7481): 87.
- Rascovan, Nicolás, Karl-Göran Sjögren, Kristian Kristiansen, Rasmus Nielsen, Eske Willerslev, Christelle Desnues, and Simon Rasmussen. 2019. "Emergence and Spread of Basal Lineages of *Yersinia Pestis* during the Neolithic Decline." *Cell* 176 (1-2): 295-305. e10.
- Rasmussen, Morten, Yingrui Li, Stinus Lindgreen, Jakob Skou Pedersen, Anders Albrechtsen, Ida Moltke, Mait Metspalu, Ene Metspalu, Toomas Kivisild, and Ramneek Gupta. 2010. "Ancient Human Genome Sequence of an Extinct Palaeo-Eskimo." *Nature* 463 (7282): 757.
- Rasmussen, M., X. Guo, Y. Wang, K. E. Lohmueller, S. Rasmussen, A. Albrechtsen, L. Skotte, et al. 2011. "An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia." *Science (New York, N.Y.)* 334 (6052): 94-98.
- Rasmussen, Morten, Sarah L. Anzick, Michael R. Waters, Pontus Skoglund, Michael DeGiorgio, Thomas W. Stafford Jr, Simon Rasmussen, Ida Moltke, Anders Albrechtsen, and Shane M. Doyle. 2014. "The Genome of a Late Pleistocene Human from a Clovis Burial Site in Western Montana." *Nature* 506 (7487): 225.
- Rasmussen, Simon, Morten Erik Allentoft, Kasper Nielsen, Ludovic Orlando, Martin Sikora, Karl-Göran Sjögren, Anders Gorm Pedersen, Mikkel Schubert, Alex Van Dam, and Christian Moliin Outzen Kapel. 2015. "Early Divergent Strains of *Y. pestis* in Eurasia 5,000 Years Ago." *Cell* 163 (3): 571-582.
- Reich, David, Richard E. Green, Martin Kircher, Johannes Krause, Nick Patterson, Eric Y. Durand, Bence Viola, Adrian W. Briggs, Udo Stenzel, and Philip LF Johnson. 2010. "Genetic History of an Archaic Hominin Group from Denisova Cave in Siberia." *Nature* 468 (7327): 1053.

- Renaud, Gabriel, Viviane Slon, Ana T. Duggan, and Janet Kelso. 2015. "Schmutzi: Estimation of Contamination and Endogenous Mitochondrial Consensus Calling for Ancient DNA." *Genome Biology* 16 (1): 224.
- Roberts, R. Gareth. 2009. *The Sea Peoples and Egypt*.
- Rosenberg, Noah A., Jonathan K. Pritchard, James L. Weber, Howard M. Cann, Kenneth K. Kidd, Lev A. Zhivotovsky, and Marcus W. Feldman. 2002. "Genetic Structure of Human Populations." *Science* 298 (5602): 2381-2385.
- Russell, Josiah C. 1968. "That Earlier Plague." *Demography* 5 (1): 174-184.
- Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1988. "Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase." *Science* 239 (4839): 487-491.
- Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim. 1985. "Enzymatic Amplification of Beta-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia." *Science (New York, N.Y.)* 230 (4732): 1350-1354.
- Sallares, Robert. 2007. "Ecology, Evolution, and Epidemiology of Plague." *Plague and the End of Antiquity*: 231.
- Salo, Wilmar L., Arthur C. Aufderheide, Jane Buikstra, and Todd A. Holcomb. 1994. "Identification of Mycobacterium Tuberculosis DNA in a Pre-Columbian Peruvian Mummy." *Proceedings of the National Academy of Sciences* 91 (6): 2091-2094.
- Sarris, Peter. 2007. "Bubonic Plague in Byzantium: The Evidence of Non-Literary Sources." In *Plague and the End of Antiquity: The Pandemic of 541-750*, edited by L. Little, 119-134: Cambridge University Press.
- Sawyer, Susanna, Johannes Krause, Katerina Guschanski, Vincent Savolainen, and Svante Pääbo. 2012. "Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA." *PloS One* 7 (3): e34131.
- Schuenemann, Verena J., Pushpendra Singh, Thomas A. Mendum, Ben Krause-Kyora, Günter Jäger, Kirsten I. Bos, Alexander Herbig, et al. 2013. "Genome-Wide Comparison of Medieval and Modern Mycobacterium Lepae." *Science (New York, N.Y.)* 341 (6142): 179-183.
- Scott, Susan and Christopher J. Duncan. 2001. *Biology of Plagues: Evidence from Historical Populations* Cambridge University Press.
- Seguin-Orlando, A., T. S. Korneliussen, M. Sikora, A. S. Malaspinas, A. Manica, I. Moltke, A. Albrechtsen, et al. 2014. "Paleogenomics. Genomic Structure in Europeans Dating Back at Least 36,200 Years." *Science (New York, N.Y.)* 346 (6213): 1113-1118.

- Sherratt, E. Susan. 1992. "Immigration and Archaeology: Some Indirect Reflections." *Acta Cypria* 2: 316-347.
- Sikora, M., A. Seguin-Orlando, V. C. Sousa, A. Albrechtsen, T. Korneliussen, A. Ko, S. Rasmussen, et al. 2017. "Ancient Genomes show Social and Reproductive Behavior of Early Upper Paleolithic Foragers." *Science (New York, N.Y.)* 358 (6363): 659-662.
- Skoglund, Pontus, Cosimo Posth, Kendra Sirak, Matthew Spriggs, Frederique Valentin, Stuart Bedford, Geoffrey R. Clark, Christian Reepmeyer, Fiona Petchey, and Daniel Fernandes. 2016. "Genomic Insights into the Peopling of the Southwest Pacific." *Nature* 538 (7626): 510.
- Skoglund, P., H. Malmstrom, M. Raghavan, J. Stora, P. Hall, E. Willerslev, M. T. Gilbert, A. Gotherstrom, and M. Jakobsson. 2012. "Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe." *Science (New York, N.Y.)* 336 (6080): 466-469.
- Slon, Viviane, Charlotte Hopfe, Clemens L. Weiß, Fabrizio Mafessoni, Marco De la Rasilla, Carles Lalueza-Fox, Antonio Rosas, Marie Soressi, Monika V. Knul, and Rebecca Miller. 2017. "Neandertal and Denisovan DNA from Pleistocene Sediments." *Science* 356 (6338): 605-608.
- Søe, Martin Jensen, Peter Nejsum, Brian Lund Fredensborg, and Christian Moliin Outzen Kapel. 2015. "DNA Typing of Ancient Parasite Eggs from Environmental Samples Identifies Human and Animal Worm Infections in Viking-Age Settlement." *The Journal of Parasitology* 101 (1): 57-63.
- Spyrou, Maria A., Rezeda I. Tukhbatova, Michal Feldman, Joanna Drath, Sacha Kacki, Julia Beltrán de Heredia, Susanne Arnold, Airat G. Sitdikov, Dominique Castex, and Joachim Wahl. 2016. "Historical Y. Pestis Genomes Reveal the European Black Death as the Source of Ancient and Modern Plague Pandemics." *Cell Host & Microbe* 19 (6): 874-881.
- Spyrou, Maria A., Rezeda I. Tukhbatova, Chuan-Chao Wang, Aida Andrades Valtueña, Aditya K. Lankapalli, Vitaly V. Kondrashin, Victor A. Tsybin, Aleksandr Khokhlov, Denise Kühnert, and Alexander Herbig. 2018. "Analysis of 3800-Year-Old Y. pestis Genomes Suggests Bronze Age Origin for Bubonic Plague." *Nature Communications* 9 (1): 2234.
- Stager, Lawrence E. 1995. "The Impact of the Sea Peoples in Canaan (1185–1050 BCE)." *The Archaeology of Society in the Holy Land*: 332-348.
- Stathakopoulos, D. 2007. "Crime and Punishment: The Plague in the Byzantine Empire, 541-749." In *Plague and the End of Antiquity, The Pandemic of 541–750*, edited by K. Lester Little, 99-118: Cambridge University Press.
- Stathakopoulos, D. C. 2004. *Famine and Pestilence in the Late Roman and Early Byzantine Empire: A Systematic Survey of Subsistence Crises and Epidemics*. UK: Ashgate.

- Straley, S. C., E. Skrzypek, G. V. Plano, and J. B. Bliska. 1993. "Yops of *Yersinia* Spp. Pathogenic for Humans." *Infection and Immunity* 61 (8): 3105-3110.
- Swarts, K., R. M. Gutaker, B. Benz, M. Blake, R. Bukowski, J. Holland, M. Kruse-Peebles, et al. 2017. "Genomic Estimation of Complex Traits Reveals Ancient Maize Adaptation to Temperate North America." *Science (New York, N.Y.)* 357 (6350): 512-515.
- Treadgold, W. T. 1997. *A History of the Byzantine State and Society* Stanford University Press.
- Treille, Georges-Félix and Alexandre Yersin. 1894. "La Peste Bubonique À Hong Kong." *Pesti konyvnyomda-részvénytársaság*.
- Twigg, Graham. 1984. *The Black Death: A Biological Reappraisal*. London: Batsford Academic and Educational.
- Valtueña, Aida Andrades, Alissa Mittnik, Felix M. Key, Wolfgang Haak, Raili Allmäe, Andrej Belinskij, Mantas Daubaras, Michal Feldman, Rimantas Jankauskas, and Ivor Janković. 2017. "The Stone Age Plague and its Persistence in Eurasia." *Current Biology* 27 (23): 3683-3691. e8.
- Wagner, David M., Jennifer Klunk, Michaela Harbeck, Alison Devault, Nicholas Waglechner, Jason W. Sahl, Jacob Enk, Dawn N. Birdsell, Melanie Kuch, and Candice Lumibao. 2014. "*Y. pestis* and the Plague of Justinian 541–543 AD: A Genomic Analysis." *The Lancet Infectious Diseases* 14 (4): 319-326.
- Warinner, Christina, Jessica Hendy, Camilla Speller, Enrico Cappellini, Roman Fischer, Christian Trachsel, Jette Arneborg, Niels Lynnerup, Oliver E. Craig, and Dallas M. Swallow. 2014. "Direct Evidence of Milk Consumption from Ancient Human Dental Calculus." *Scientific Reports* 4: 7104.
- Warinner, Christina, Alexander Herbig, Allison Mann, James A. Fellows Yates, Clemens L. Weiß, Hernán A. Burbano, Ludovic Orlando, and Johannes Krause. 2017. "A Robust Framework for Microbial Archaeology." *Annual Review of Genomics and Human Genetics* 18: 321-356.
- Whittow, Mark. 1996. *The Making of Byzantium, 600-1025* Univ of California Press.
- Wiechmann, Ingrid and Gisela Grupe. 2005. "Detection of *Y. pestis* DNA in Two Early Medieval Skeletal Finds from Aschheim (Upper Bavaria, 6th Century AD)." *American Journal of Physical Anthropology* 126 (1): 48-55.
- Willerslev, Eske and Alan Cooper. 2004. "Ancient Dna." *Proceedings of the Royal Society B: Biological Sciences* 272 (1558): 3-16.
- Woodward, S. R., N. J. Weyand, and M. Bunnell. 1994. "DNA Sequence from Cretaceous Period Bone Fragments." *Science (New York, N.Y.)* 266 (5188): 1229-1232.

- Yang, Ruifu, Yujun Cui, and Dongsheng Zhou. 2012. "Rapid Evolution of the Plague Pathogen." *Rapidly Evolving Genes and Genetic Systems*: 211.
- Yasur-Landau, Assaf. 2010. "On Birds and Dragons: A Note on the Sea Peoples and Mycenaean Ships." *Pax Hethitica. Studies on the Hittites and their Neighbours in Honor of Itamar Singer*: 399-410.
- Yoshida, Kentaro, Verena J. Schuenemann, Liliana M. Cano, Marina Pais, Bagdevi Mishra, Rahul Sharma, Chirsta Lanz, Frank N. Martin, Sophien Kamoun, and Johannes Krause. 2013. "The Rise and Fall of the Phytophthora Infestans Lineage that Triggered the Irish Potato Famine." *Elife* 2: e00731.
- Zhou, D. and R. Yang. 2009. "Molecular Darwinian Evolution of Virulence in *Y. pestis*." *Infection and Immunity* 77 (6): 2242-2250.
- Zietz, Björn P. and Hartmut Dunkelberg. 2004. "The History of the Plague and the Research on the Causative Agent *Y. pestis*." *International Journal of Hygiene and Environmental Health* 207 (2): 165-178.
- Zinser, E. R., D. Schneider, M. Blot, and R. Kolter. 2003. "Bacterial Evolution through the Selective Loss of Beneficial Genes. Trade-Offs in Expression Involving Two Loci." *Genetics* 164 (4): 1271-1277.
- Zischler, Hans, Matthias Hoss, Oliva Handt, A. Von Haeseler, AC Van der Kuyl, and J. Goudsmit. 1995. "Detecting Dinosaur DNA." *Science* 268 (5214): 1192-1193.

9. Summary

The emerging field of Archaeogenetics has experienced a dramatic increase in data production over the past decade, mostly attributed to advances in sequencing technology. As a result, the field has been claiming a pivotal role in ongoing efforts to reconstruct the human past. This thesis makes use of Archaeogenetic approaches to retrieve genomic material from historical and archaeological human remains, comprising “molecular fossils” that harbor direct evidence from specific points in time. The papers comprising this thesis are focused around three of the major events that have shaped human history over the last 15,000 years, attempting to shed light on some of the processes that have led to these events as well as on their long term genetic impact.

First, I retrieved genome-wide data from a ~ 15,000 year old Epipaleolithic hunter-gatherer from central Anatolia as well as from seven ~ 8,000-10,000 year old Neolithic early farmers from central Anatolia and the Levant (**Manuscript A**), comprising a genetic dataset that stretches over seven millennia and includes the period of the Neolithic transition in central Anatolia. I analyzed this dataset together with published ancient and modern ones to address the long standing debate of whether the transition into farming in central Anatolia was introduced by colonizing farmers from earlier farming centers in the Fertile Crescent or whether it developed locally through a cultural process in which the local hunter-gatherers adopted an agricultural subsistence strategy. In addition, I examined potential genetic interactions between the people of central Anatolia and neighboring regions. The results show a high degree of genetic continuity between the Anatolian hunter-gatherer and early farmers, suggesting a limited effect of migration during the transition into farming in central Anatolia. Despite this long-term stability of the local gene pool, gene flow from two distinct sources is revealed: an earlier one that occurred during either the late Pleistocene or early Holocene and derived from an ancient Iranian/Caucasus related ancestry and a later bidirectional one, that occurred during the Neolithic and derived from an ancestry related to the ancient Levant. A third genetic link that predates 15,000 years ago is detected between the ancestors of southern Europe hunter-gatherers and of Near Eastern ones. These results provide insights on the early spread of farming and shed light on the formation of the Anatolian early farmer gene pool that later replaced most of the European hunter-gatherer gene pool and became the single largest ancestral component in present-day Europeans.

Second, I reconstructed genome-wide data from ten Bronze- to Iron- Age individuals (~15th – 10th century BCE) from the ancient Mediterranean port city of Ashkelon, located in present-day Israel and identified by ancient texts to be “Philistine” during the Iron Ages (**Manuscript B**). These studied people have lived in a time of great political and economic turmoil as well as extensive human mobility marked by the collapse of the eastern Mediterranean civilizations (~12th century BCE) and the transition between the Bronze- and Iron- Ages. Coincident with these events, there was a major change in the material culture in coastal cities of the Eastern Mediterranean (e.g. Ashkelon). The new material culture of the Iron Ages (defined in archaeology as “philistine”) resembled that found in Bronze Age Aegean sites. I analyzed the retrieved genome-wide data and combined it with published data, to address the long debated question on the origins of the Philistines. The results show that the early Iron Age population in Ashkelon harbored a European-related gene flow that entered Ashkelon either during the end of the Bronze- or the beginning of the Iron- Age, coinciding with the time of the Philistine estimated arrival in Ashkelon. Genetic modeling suggests southern European gene pools as plausible sources for this genetic influx. Notably, the later Iron Age individuals, culturally identified as “Philistines” were missing the genetic signal of this European influx, suggesting that within no more than two centuries, this genetic footprint was diluted by the local Levantine gene pool, thus, the European genetic influx did not make a long term genetic impact on the Ashkelon gene pool.

Finally, I retrieved a high-coverage *Y. pestis* genome (~18 fold) from the remains of a 6th century victim of the Justinianic Plague, the first of three historic pandemics caused by *Y. pestis* (**Manuscript C**). Only one low coverage Justinianic genome had been previously available. Both the new and published genomes were isolated from southern Germany. The relatively high quality of the new genome allowed for the detection of previously undescribed unique features of the Justinianic *Y. pestis* lineage, some of which located in genomic regions previously suggested to be involved in plague virulence. The results mostly confirm the previous phylogenetic placement of this lineage that points to a south Asian origin. In addition, erroneous substitutions reported for the lower covered *Y. pestis* genome were identified. The two strains were otherwise genetically identical, suggesting low genetic diversity of the bacteria in Southern Germany during this plague pandemic. In terms of the disease’s history, the results demonstrate *Y. pestis* was present in a second early medieval rural site in southern Germany where no historical source records it and place this strain in the early waves of the 200-year-long pandemic.

In conclusion, this thesis demonstrates how Archaeogenetic studies can help record history by tracing human mobility as well as human disease. These kinds of evidences become increasingly important when reliable written records and archaeological findings are scarce. Some of the genomic data produced in the course of this thesis were retrieved from challenging environments, highlighting how applied methodologies might assist in similar future work. Exhaustive genetic sampling of a wide temporal and geographical scope has the potential to address further pending questions regarding the human past.

10. Zusammenfassung

Das immer weiter an Bedeutung gewinnende Forschungsfeld der Archäogenetik hat in den letzten Jahren, dank fortschrittlicher Sequenzieretechnologien, einen spektakulären Anstieg in der Datengewinnung erlebt. Durch diese Fortschritte konnte es eine Schlüsselrolle in den Bemühungen um die Erfassungen der Vergangenheit einnehmen.

Die hier vorgelegte Arbeit bedient sich der Archäogenetik um genomisches Material aus historischen und archäologischen Menschenfunden zu gewinnen, eine Art “molekulare Fossilien”, Zeugnisse aus vergangenen Zeiten, die Erkenntnisse über diesen Zeitpunkt enthalten. Die in dieser Arbeit zusammengefassten Veröffentlichungen behandeln drei große Ereignisse, die die Menschheitsgeschichte der letzten 15000 Jahre geprägt haben. Sie erfasst einige Prozesse die zu diesen Ereignissen führten und beleuchtet deren langfristigen genetischen Einfluss.

Zunächst gewann ich genomweite Daten eines ca. 15000 Jahre alten, Epipaläolithischen Jäger und Sammlers aus Zentralanatolien sowie von sieben ca. 8000 – 10000 Jahre alten früh-neolithischen Ackerbauern aus Zentralanatolien und der Levante (**Manuskript A**), die sich zusammengefasst über einen Zeitraum von sieben Jahrtausenden erstrecken und den Zeitraum des Überganges zum Ackerbau in Zentralanatolien beleuchten. Die Analyse dieses Datensets zusammen mit bereits publizierten alten und rezenten Individuen behandelt die lange diskutierte Frage, ob dieser Übergang zum Ackerbau in Zentralanatolien durch einwandernde Menschen aus dem fruchtbaren Halbmond ausgelöst wurde, oder ob sich der Ackerbau lokal durch einen kulturellen Prozess entwickelte bei dem die dort heimischen Jäger und Sammler den eine neue Lebensweise annahmen. Zusätzlich betrachtete ich die genetischen Interaktionen zwischen den Menschen Zentralanatoliens und den angrenzenden Regionen. Diese Vergleiche zeigen einen hohen Grad der genetischen Kontinuität zwischen den anatolischen Jägern und Sammlern und den darauffolgenden Ackerbauern. Diese Beobachtung lässt auf einen nur kleinen Anteil der Migration an dem Übergang zu einer neuen Lebensweise schließen. Trotz dieser langen Stabilität des Genpools wurden zwei eigenständige Abstammungskomponenten deutlich: Eine erste, im späten Pleistozän oder frühen Holozän eingebrachte, dessen Abstammung im Zusammenhang mit alten Populationen des Iran und Kaukasus steht, sowie eine Zweite die während des Neolithikums aus der Levante eingebracht wurde und anschließend im Austausch zwischen Anatolien und der Levante stand. Eine weitere entdeckte genetische Verbindung zwischen den Vorfahren der südeuropäischen Jäger und Sammler und denen des Nahen Ostens bestand im Paläolithikum noch

vor 15 000 vor heute. Diese Ergebnisse liefern neue Einsichten in die Ausbreitung der sesshaften, ackerbaubetreibenden Lebensweise und beleuchten die Bildung des Genpools der ersten anatolischen Ackerbauern, welcher später den europäischen Jäger und Sammler verdrängte und dadurch zur größten Abstammungskomponente in heutigen Europäern wurde.

Anschließend rekonstruierte ich genomweite Daten von zehn Individuen aus der antiken Mittelmeer-Hafenstadt Ashkelon im heutigen Israel, die in der Bronze bis Eisenzeit (15. – 10. Jahrhundert vor Christus) lebten. In altertümlichen Texten wurden diese während der Bronzezeit als „Philister“ bezeichnet (**Manuskript B**). Die untersuchten Völker lebten zu einer Zeit politischen und ökonomischen Aufruhrs und starker Wanderungen die zum Kollaps großer Zivilisationen im östlichen Mittelmeerraum (ca. 12.Jhdt.v.Ch.) führten und den Übergang der Bronze- zur Eisenzeit prägten. Zusammenfallend mit diesen Ereignissen lässt sich eine Änderung der materiellen Kultur in den Städten des östlichen Mittelmeerraumes, beispielsweise Ashkelon, beobachten. Diese neue materielle Kultur der Eisenzeit wird im archäologischen Befund als „Philister“ definiert und ähnelt der zur Bronzezeit in der Ägäis beschriebenen materiellen Kultur. Ich analysierte die gewonnenen genomweiten Daten zusammen mit bereits publizierten Individuen um die langjährige Frage nach der Herkunft der Philister zu klären. Die Ergebnisse zeigen in der frühen Eisenzeit einen europäisch geprägten genetischen Einfluss der die in Ashkelon lebende Bevölkerung entweder zum Ende der Bronzezeit oder zum Beginn der Eisenzeit erreichte und so mit der geschätzten Ankunft der Philister zusammenfällt. In genetischen Modellierungen stellt sich ein südeuropäischer Genpool als wahrscheinlichste Quelle für diesen Einfluss heraus. Beachtenswerter Weise zeigen die eisenzeitlichen, kulturell als Philister identifizierten, Individuen diesen Einfluss nicht mehr. Diese Modelle lassen den Schluss zu, dass in weniger als zwei Jahrhunderten das genetische Erbe der „Philister“ von dem lokalen levantinischen Genpool letztlich ersetzt wurde und so die europäische Herkunft keinen langfristigen genetischen Einfluss auf die Bevölkerung Ashkelons hatte.

Abschließend untersuchte ich die sterblichen Überreste eines Opfers der Justinianischen Pest aus dem 6. Jahrhundert n. Ch. und konnte ein hochqualitatives Genom des Erregers der Pest, *Yersinia pestis*, gewinnen (**Manuskript C**). Die Justinianische Pest war die erste von drei historischen, durch *Y. pestis* ausgelösten Pandemien. Bisher war nur eine weitere Sequenz aus diesem Kontext verfügbar welche wie die neu gewonnene aus Süddeutschland stammte. Die relative hohe Qualität der neu isolierten Sequenz erlaubte es bisher unbekannte, einzigartige Eigenschaften des mit der Justinianischen Pest in Verbindung stehenden *Y.pestis* Stammes zu

ermitteln. Einige dieser liegen in Bereichen des Genoms, die zuvor mit besonderer Virulenz in Zusammenhang gebracht wurden. Die Ergebnisse bestätigen die phylogenetische Einordnung des Stammes die auf eine Herkunft in Asien hindeutend. Zusätzlich konnten fehlerhafte Substitutionen im weniger qualitativen ersten Genom festgestellt werden. Abgesehen von diesen stellten sich die beiden Sequenzen als genetisch identisch heraus, was auf eine niedrige genetische Diversität der Bakterien in Süddeutschland während dieses Pestausbruchs schließen lässt. Des Weiteren geben uns die Ergebnisse Einblicke in die Geschichte der Justinianischen Pest. *Y. pestis* war, von keiner historischen Quelle erwähnt, in einer weiteren mittelalterlichen Fundstätte im ländlichen Süddeutschland aufgetreten und der isolierte Stamm kann den frühen Wellen der 200 Jahre anhaltenden Pandemie zugeordnet werden.

Zusammenfassend zeigt die hier vorgelegte Arbeit wie die Archäogenetik helfen kann die Menschheitsgeschichte aufzuzeichnen und menschliche Mobilitätsmuster und Krankheiten zu rekonstruieren. Die hier verwendeten Methoden werden wichtiger, je weniger zuverlässige schriftliche Quellen verfügbar sind und wo archäologische Befunde rar sind. Einige der in dieser Arbeit verwendeten genomischen Daten wurden aus geographischen Regionen gewonnen die eine schlechte DNA Erhaltung aufweisen und unterstreichen wie die hier angewendeten Methoden bei zukünftigen Arbeiten helfen können Material aus solchen Kontexten zu bearbeiten. Zusätzliches Beprobieren von menschlichen Überresten aus einem weiten zeitlichen und geographischen Umfang hat das Potential noch andere ausstehende Fragen der Menschheitsgeschichte zu beleuchten.

11. Eigenständigkeitserklärung

Entsprechend §5 Abs. 4 der Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität Jena, erkläre ich, dass mir die geltende Promotionsordnung der Fakultät bekannt ist. Ich bezeuge, dass ich die vorliegende Dissertation selbst angefertigt habe und keine Textabschnitte eines Dritten oder eigener Prüfungsarbeiten ohne kennzeichnung übernommen und alle von mir benutzten Hilfsmittel, persönliche Mitteilungen sowie Quellen in meiner vorliegenden Arbeit angegeben habe. Zudem habe ich alle Personen, die mir bei der Auswahl und Auswertung sowie bei der Erstellung der Manuskripte unterstützt haben, in der Auflistung der Manuskripte und den entsprechenden Danksagungen namentlich erwähnt. Zudem verichere ich, dass ich die Hilfe eines Promotionsberaters nicht in Anspruch genommen haben und auch Dritten von mir keine unmittelbaren sowie mittelbaren geldwerte Leistungen für Arbeiten, die im Zusammenhang mit dieser Dissertaton stehen, erhalten haben. Die vorliegende Promotion wurde zuvor weder für eine staatliche oder andere wissenschaftliche Prüfung eingereicht, also auch einer anderen Hochschule als Dissertation vorgelegt.

Jena, den 15.04.2019

Michal Feldman

12. Acknowledgments

I would like to thank the following people that helped me along this journey.

First and foremost, to my doctorate supervisor Johannes Krause, for his support and generosity, for the opportunity to engage in this fascinating research, for being an outstanding scientific role model and at the same time providing me with the freedom to grow as an independent researcher.

To Choongwon Jeong who guided me through my first steps in population genetics and taught me to be my own toughest critic. To Alexander Herbig for his guidance and support through the ancient pathogen analyses and to Verena Schuenemann for showing me the ropes around the lab.

To Israel HersHKovitz who first introduced me to the scientific study of the past and has continuously been a mentor to me.

To Raffaella Bianco, for being my lab assistant and doing an excellent job of it.

To Wolfgang Haak, Stephan Schiffels and all the participants of our weekly Popgen meetings for their input and advice, always given in a friendly and supportive atmosphere.

To Philipp Stockhammer and all the members of the MHAAM group for their support and helpful discussions as well as to all the members of the Archaeogenetics department at the MPI-SHH for creating such a lovely dynamic working environment.

To my colleagues and very dear friends Maria Spyrou, Cosimo Posth, Eirini Skourtanioti, Alex Immel and Kathrin Naegle who were always there offering an open ear and helping hand in both professional and personal matters. A special thanks to Kathrin for translating the summary of the thesis into German.

To my family, my husband Ofer who made this possible by supporting and believing in me throughout this intense experience. Many times taking over home responsibilities through a series of never-ending deadlines. To our son Jonathan for understanding that Ima has to work and for making me smile at the end of the day. To my in-laws, Eva and Moshe for supporting me despite the difficult geographic distance and for their uplifting genuine enthusiasm to learn about my research. Last but not least, to my brother Giora and my parents Sofia and the late Yevgeni, for their love and sacrifices throughout the years and for encouraging my curiosity and passion for learning.

13. Curriculum Vita

Personal Information	
Name	Michal Feldman
Date of Birth	June 17th, 1980
School and scientific Education	
2015-2019	Continuation of studies towards a PhD in Archaeogenetics under the supervision of Prof. Johannes Krause. The Max Planck institute for the Science of human History, Jena, Germany. Title of thesis: <i>“Tracing past human mobility and disease in western Eurasia by the genetic analysis of ancient human remains”</i>
2014-2015	Studies towards a PhD in Archaeogenetics under the supervision of Prof. Johannes Krause, Institut für Naturwissenschaftliche Archäologie (INA), The University of Tuebingen, Germany. Title of thesis: <i>“Tracing past human mobility and disease in western Eurasia by the genetic analysis of ancient human remains”</i>
2010-2013	MSc in Medical Sciences, The Sackler Faculty of Medicine, Tel-Aviv University. Thesis titled <i>“Detection and characterization of human disease-related genes from ancient human mummified tissue”</i> under the supervision of Prof. Israel HersHKovitz, Dr. Rina Arbesfeld and Dr. Ella Sklan Dep. Of Anatomy and Anthropology, The Sackler Faculty of Medicine, Tel-Aviv University
2006-2009	BSc <i>Pharm</i> in Pharmacy, The school of pharmacy, Faculty of Medicine, The Hebrew University of Jerusalem.
2001-2004	BSc in Life Sciences, The George S. Wise Faculty of Life Sciences, Tel-Aviv University
1998	High school diploma from Kiriat-Sharet high school, Holon, Israel
Employment	
2012-2013	Teaching assistant (dissection instructor and assistant in CT-based classes) in the “Anatomy of the Human Body” classes for medicine and dentistry students, The Sackler Faculty of Medicine, Tel-Aviv University
2010-2012	Community Pharmacist in the “Maccabi national health care services”, Ramat-gan, Israel
2009	Pharmacy Internship in the Pharmaceutical research unit, Research and development department, “Teva pharmaceutical industries LTD”
List of Publications	
Feldman M., Master D.M., Bianco R.A., Burri M., Stockhammer P. W., Mitnik A., Aja A.J., Jeong C., and Krause J. “Ancient DNA sheds light on the genetic origins of early Iron Age Philistines.” <i>Science Advances</i> (2019).	
Feldman M., Fernández-Domínguez E., Reynolds L., Baird D., Pearson J., HersHKovitz I., May H., Goring-Morris N., Benz M., Gresky J., Bianco R. A., Fairbairn A. , Mustafaoğlu G., Stockhammer P. W., Posth C., Haak W., Jeong C., and Krause J. “Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia.” <i>Nature Communications</i> . 10, 1218 (2019).	
Mitnik A., Wang C.C., Pfrengle S., Daubaras M., Zariņa G., Hallgren F., Allmäe R., Khartanovich V., Moiseyev V., Tõrv M., Furtwängler A., Valtueña A.A., <u>Feldman M.</u> ,	

Economou C., Oinonen M., Vasks A., Balanovska E., Reich D., Jankauskas R., Haak, W., Schiffels S., Krause J. "The genetic prehistory of the Baltic Sea region." <i>Nature communications</i> , 9(1), p.442 (2018).
Valtueña A.A., Mitnik A., Key F.M., Haak W., Allmäe R., Belinskij A., Daubaras M., <u>Feldman M.</u> , Jankauskas R., Massy I. K., Novak M., Pfrengle S., Reinhold S., Šlaus M., Spyrou A.M., Szécsényi-Nagy A., Törv M., Hansen S., Bos K.I., Stockhammer P.W., Herbig A., Krause J. "The stone age plague and its persistence in Eurasia." <i>Current biology</i> 27, no. 23 (2017): 3683-3691.
<u>Feldman M.</u> , Harbeck M, Keller M, Spyrou MA, Rott A, Trautmann B, Scholz HC, Pääffgen B, Peters J, McCormick M, Bos K., Herbig A., Krause J. "A high-coverage <i>Yersinia pestis</i> genome from a sixth-century Justinianic plague victim." <i>Molecular biology and evolution</i> 30;33, no. 11 (2016): 2911-23.
Spyrou M.A., Tukhbatova R.I., <u>Feldman M.</u> , Drath J., Kacki S., de Heredia J.B., Arnold S., Sitdikov A.G., Castex D., Wahl J., Gazimzyanov I.R., Nurgaliev D.K., Herbig A., Bos K., Krause J. "Historical <i>Y. pestis</i> genomes reveal the European Black Death as the source of ancient and modern plague pandemics." <i>Cell host & microbe</i> 8, no. 19(6) (2016):874-81.
<u>Feldman M.</u> , HersHKovitz I., Sklan E.H., Kahila Bar-Gal G., Pap I., Szikossy I., Rosin-Arbesfeld R. "Detection of a tumor suppressor gene variant predisposing to colorectal cancer in an 18 th century Hungarian mummy." <i>PloS one</i> 11, no. 2 (2016): e0147217.
HersHKovitz I., Donoghue H.D., Minnikin D.E., May H., Lee O.Y.C., <u>Feldman M.</u> , Galili E., Spigelman M., Rothschild B.M., Kahila Bar-Gal G. "Tuberculosis origin: the Neolithic scenario." <i>Tuberculosis</i> 95 (2015): S122-S126.

Michal Feldman

14. Appendix

14.1 Supplementary materials of paper A

Supplementary Information

Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia

Feldman et al.

Supplementary Notes

1. Description of samples and archaeological information
2. Genetic analysis of Neolithic Levantines
3. Admixture modeling of the Ancient Anatolian populations
4. Investigating genetic links between Near-Eastern and European hunter-gatherers
5. Mitochondrial DNA analysis

Supplementary Figures 1-14

Supplementary Tables 1-10

Supplementary Notes

Supplementary Note 1: Description of samples and archaeological information

Pınarbaşı – archaeological information

Pınarbaşı is situated 33.4 km southeast of Boncuklu on the eastern edge of the western Konya basin (37° 29'N, 33°02'E) at the end of the Bozdağ limestone hills, northwest of the Karadağ mountain, and represents the only excavated Epipaleolithic site on the central Anatolian plateau¹. The site contains a series of rock shelters and caves, the most northerly of which was subjected to excavation, as Area B. The site was initially excavated in 1994-95 by Professor Trevor Watkins of Edinburgh University². Excavations recommenced in 2003 under the direction of Professor Douglas Baird from the University of Liverpool and resulted in the detection and excavation of the Epipaleolithic deposits. Occupation of the site is evidenced by a long settlement sequence that commences by c. 13,500 cal BC in one of the rock shelters¹. At the beginning of the Holocene, c. 9600 cal BC, the site saw the emergence of a sedentarising community on a small mound c. 100m west of the Epipaleolithic rock shelter³. The mound settlement was excavated in Areas A and D³.

The Epipaleolithic population represented at Pınarbaşı was probably highly mobile and low density, ranging over a large area of the central Anatolian plateau and quite possibly south of the Taurus peaks in winter¹. At Pınarbaşı they hunted local wild caprines and cattle, wetland birds and fish, with no clear evidence for any significant exploitation of local plants. Notably ancestors of the first cultivated cereals are absent¹. Strong links to other Epipaleolithic groups especially the Natufian of the Levant but also to groups on the Mediterranean coast of Turkey have also been documented through similarities in and exchanges of material culture, especially chipped stone, obsidian and sea shell beads, as well as technological, ritual and social practices^{1, 4, 5}. The 10th-9th millennium cal BC saw the development of more sedentary practices at Pınarbaşı but this

occupation lacks evidence of exploitation of wild ancestors of early cultivated plants, cultivation of cereals and legumes and animal herding³.

Pınarbaşı - sample description

Five individuals were selected for ancient DNA analysis, skeletons ZBC and ZBD from the Epipaleolithic rock shelter and ZDS, ABM and ZAN from the open air 10th-9th millennia site (Supplementary Data 1). Only skeleton ZBC, found within Grave 13, provided sufficient genomic data. This burial was placed in an oval cut, early in the Epipaleolithic sequence at the site and thus overlaid by c. 1 m of Epipaleolithic occupation deposit and rock shatter from the rock-shelter wall. It was thus very well stratified within the site sequence. Anthropological analyses by Dr Kirsi Lorentz identified this burial as belonging to a c. 25-29 year old male¹. The body was found fully articulated, extended in supine position, with both hands resting at the pelvis area and was lacking the cranium (Supplementary Figure 10). The presence of maxillary teeth within the grave together with the absence of evidence for major disturbance, suggests that skull removal practice characteristic of Aceramic Neolithic communities was already practised and is one feature that suggests interactions with Levantine Natufian communities where it is sporadically attested within Natufian mortuary practices¹.

Two radiocarbon dates are available for this specific articulated skeleton, ZBC. The original date of 14,209-13,121 cal BCE (2 sigma range 95.4%) (OxA 16536)¹ was confirmed with a second date of 13,646-13,284 cal BCE (2 sigma range 95.4%) (MAMS 31616) (Supplementary Table 1) from the same phalanx bone that was subjected to ancient DNA analysis. This second date, confirms the original date but usefully has a shorter range, which supports the view that this burial probably predates the Natufian or possibly overlaps with its very earliest phases, and certainly predates the Bolling-Allerod/GI 1 interstadial¹.

Boncuklu - archaeological information

Boncuklu is situated on the Konya Plain (37° 45'N 32°52'E) and lies 33.4 km northwest of Pınarbaşı and 10.2 km northeast from Çatalhöyük. The site was discovered during the archaeological surveys of the Konya Plain between 1994-2002 under the direction of Professor Douglas Baird from the University of Liverpool. Excavations began in 2006 and continue at the present time³. Occupation of the site is documented from 8300-7600 cal BC directly through radiocarbon dating. However, stratigraphic and material evidence suggest a slightly longer occupational span^{3, 5, 6}.

The Boncuklu community seems to have relied on the exploitation of wild resources to a large degree, especially wild cattle and boar, fish and wetland birds along with nuts and fruits from surrounding hill areas^{3, 6}. To these resources were added small-scale cultivation of wheat, lentils and peas³. The chipped stone industry was microlithic, in significant contrast to broadly contemporary Levantine PPNB and northern Fertile Crescent assemblages and thus shows significant continuities with the earlier local Epipaleolithic and the earlier 10th/early 9th millennium BC community at Pınarbaşı in technological and raw material exploitation traditions^{3, 6}. There is thus strong archaeological evidence of continuities from Epipaleolithic and early Holocene forager communities with the community at Boncuklu. Thus by 8300 cal BC it appears local foragers adopted domestic plants from areas to the south and east and fitted them into their traditional wetland exploitation practices³. They were presumably introduced to the region as a consequence of the far reaching and continuous interactions with neighbouring regions from the Epipaleolithic through the 10th-early 9th millennia cal BC, as also documented at earlier and contemporary Pınarbaşı³.

The site possessed a number of sub-oval domestic buildings with mudbrick walls. The Boncuklu houses underwent repeated continuous reconstruction over multiple generations in the same location, a pattern similar to other certain Aceramic Neolithic sites in the surrounding

regions, for example, to the north east at Aşıklı from 8300 cal BC⁷, just to the south at Çatalhöyük from 7100 cal BC⁸, in the Levant at PPNA Jericho⁹ and in PPNB Tell Halula¹⁰.

Primary inhumations were buried under the houses during their occupation, a common practice across the Near-Eastern Neolithic, but there were also primary burials and burials of deliberately disarticulated human remains including skulls in open areas between buildings. More than 37 Neolithic burials, plus 274 individual bones and 129 isolated finds of human remains have been found in the site so far¹¹.

Boncuklu – sample description

Skeletal samples from 31 individuals from areas H, K, M and Q were selected for genetic analysis (Supplementary Data 1), from which five, described in detail below, provided sufficient genomic data (Supplementary Data 1). Three of these burials (ZHAJ, ZHAG and ZHJ) were all articulated primary inhumations stratified within a sequence of 4 buildings in Area H, all were securely stratified under long sequences of plaster floors, two were buried during the ongoing occupation of one of the buildings. One of these, ZHJ, was directly dated by C14 to 8269-8210 cal BC in a Bayesian model³. ZHAJ and ZHAG were securely stratified earlier than this and thus definitively predate 8200 cal BC. ZKO was a primary fully articulated inhumation buried under the plaster floor of a building, within a stratified sequence of six buildings in Area K. It was therefore securely stratified, overlaid by a large number of plaster floors. Unpublished C14 dates from this sequence of buildings clearly indicate it falls within the main sequence dated at the site to 8300-7800 cal BC. ZMOJ is a primary inhumation, although with some elements disturbed by animal burrows, deeply stratified in a sequence of midden deposits in Area M. Currently this sequence is not directly dated, although associated artifacts suggest it overlaps with the other

dated excavation sequences in the main excavated phases of occupation at the site, within the date ranges outlined above.

ZHAJ (Area H, Grave 27). This is a primary single inhumation of a middle age adult buried in a sub-oval cut. The individual was found lying tightly flexed on their left side, positioned east-west with the head towards the west and facing north. Considerable damage from bioturbation had disturbed the arms and pelvis¹¹. Initial anthropological analysis based on the skull in the absence of pelvis suggested a possible male. However, ancient DNA analysis has determined that this individual was a female (Table 1).

ZHAG (Area H, Grave 18). Grave 18 contained a double inhumation of a middle age adult female (ZHAF) and a perinatal baby (ZHAG) found in an oval cut larger than average. The adult individual (ZHAF) was found lying tightly flexed on their left side and positioned with a northeast-southwest orientation with the head towards the northwest. The perinatal individual was articulated and found with the head on top of the adult pelvis¹¹. The female sex of the adult individual (genome labelled as Bon005_pub) could be confirmed by ancient DNA¹². In the present study we could also determine the sex of the perinatal individual as a female (Table 1).

ZHJ (Area H, Grave 15). This is a primary single old adult inhumation found in a sub-oval cut. The individual was found in a crouched position lying on its right side and positioned north-south with the head orientated towards the south (Supplementary Figure 11). The bones were relatively well preserved compared with other graves although burrowing animals have destroyed parts of the skull and axial skeleton, including the left foot¹¹. Morphological sexing was difficult because the remains were gracile, probably as a result of the ageing process. Ancient DNA analyses allowed us to establish that this individual was a female (Table 1).

ZKO (Area K, Grave 12). This is a single inhumation of an old adult male in an oval cut. The individual was found lying tightly flexed on their left side and orientated east-west with the

head towards the east. Scattered fragments of at least two infants (ZKM, ZKR, ZKQ, KQE), probably from earlier disturbed primary burials were also recovered from the grave fill. The bones were generally well preserved, but rodent burrowing activity caused significant disturbance to the ribs, scapula and vertebrae¹¹. We were able to confirm through DNA analysis the sex of this individual as a male (Table 1).

ZMOJ (Area M, Grave 49, associated unit MAKR). A primary but heavily disturbed burial of a young adult in a sub-circular grave. The individual was orientated east-west with head to the west and facing north. The skull was found at one end of the grave and many of the other bones had been moved by animal action, so their anatomical position was not maintained¹¹. We could determine through ancient DNA that this individual was male (Table 1).

Kfar HaHoresh - archaeological information

Kfar HaHoresh (32°42'13.3"N 35°16'13.3"E, 375 m above sea level) is a small, 0.75 ha, Pre-Pottery Neolithic B (PPNB) site on the western flanks of the Nazareth hills in lower Galilee, Israel. It is situated within a rock escarpment embayment in the uppermost reaches of a small tributary wadi that flows to the Jezreel valley and thence to the Mediterranean coast 25 km distant. During seventeen excavation seasons (1991-2012) a total of 500 m² were excavated¹³⁻¹⁵. Three principle stratigraphic phases were identified, broadly corresponding to the Early, Middle and Late PPNB, dating from ca. 8,600 cal BCE to after 7,500 cal BCE¹⁶. Occupation intensity increased through the sequence.

The earliest occupation is dominated by a massive, walled and lime plaster-surfaced, quadrilateral podium (Locus 1604), ca. 22 x 10 m, with at least three architectural sub-phases and a hearth molded into the earliest plastered surface¹⁵. Plaster curling up at the edge of the walls likely indicate the use of a mudbrick parapet (not preserved). In the center of the podium a grave

(Locus 1005) was dug into sterile sediments below the lowermost plaster surface containing a partially articulated but headless adult human male that also included remains of a herd of eight aurochs (*Bos primigenius*), likely representing a funeral feast^{17, 18}. In open areas south of the podium at least two more pits containing mostly aurochs remains were documented¹⁹, as were headless single inhumations, one articulated, the other an unusual secondary burial.

The later phases are characterized by smaller, mostly quadrilateral but also oval plaster-surfaced structures (one painted red), none exceeding 5 x 5 m, sometimes in association with one or more retaining walls on the upslope side, likely representing foundations of mudbrick retaining walls and/or parapets. The structures are accompanied by other terrace/dividing walls, platforms, cists, monoliths, postholes (seemingly non-architectural), and a numerous and diverse array of combustion features: hearths, ovens, kilns²⁰ and midden deposits, including knapping pits and caches^{21, 22}.

Later phase burials are more numerous and varied; they include single and multiple, fully articulated, with or without skulls, as well as secondary burials, skull caches and more common isolated human remains than before^{15, 16, 23}. Remains of three plastered skulls were recovered, all of young adult males, one painted with cinnabar from the Taurus Mountains in southern Turkey²⁴. Many graves either directly underlie architectural plastered surfaces or are covered by chalky plaster. Though including males and females and all age cohorts from neonates to elderly, the demographic profile is unusual, with an emphasis on young adult males, 20-29 years old²⁵. Of note are two graves (Locus 1003 and Locus 1155 complex) under different plastered surfaces, each with minimum number of individuals (MNIs) of 17, and both containing mixtures of articulated and secondary remains (but few cranial elements), the bones having been carefully arranged^{23, 26, 27}. Morphological analyses on the teeth are rather heterogeneous, but clearly show one cluster (mostly from L1003) belonging to a quite homogenous group, suggesting close

biological relations between females and sub-adults that may indicate matrilocal residence patterns²⁸. Grave-goods include animal (aurochs, fox, goat and gazelle) remains as well as projectile points, sickles, ground-stone tools, marine molluscs, exotic minerals, ochre and clay tokens²⁹.

The fauna at Kfar HaHoresh indicate that wild ungulates (aurochs and gazelle) were preferentially selected (in contrast to coeval sites in the region), as well as evidence for increasing goat management through the sequence³⁰⁻³². Among smaller species fox, hare, tortoise, cat, birds and fish are notable. Preservation of palaeobotanical remains is almost non-existent, though a seed of *Vicia faba* was identified.

Abundant and varied small finds categories were recovered. The huge chipped stone assemblages were made on-site using three knapping technologies - an *ad hoc* blade/flake approach, one for serial blade production from bidirectional (naviform) cores, and one for bifacial tools^{14, 33}. Tools include sickle blades, projectile points, perforators, burins and chamfered items, axes and knives. Bone tools were present in some quantity. Passive ground-stone tools include querns and workslabs, while active items include numerous pounders, hammerstones, abraders, polishers, grooved items and minute polished pebbles. Abundant baked clay items include tokens and figurines, while a small ceramic assemblage was also identified³⁴. Marine and freshwater molluscs are common, with most deriving from the Mediterranean though also from the Red Sea³⁵. Colourful exotic minerals, in the form of lumps, pendants and beads include obsidian, malachite, amazonite, jet, bitumen and carnelian; sources range from south/central Turkey, northern Syria, the Rift valley, the Negev/Sinai, and southern Transjordan³⁶. Animal and human figurines are made on clay and stone, with the only gendered items relating to phallic imagery.

Based upon its modest size, unusual and secluded setting, the lack of adjacent arable land, and the nature of the recovered finds, Kfar HaHoresh is interpreted as a local cult and funerary

locality that was probably only occupied on a periodic basis³⁷. It may have served neighbouring lowland village communities, such as Yiftahel³⁸, situated 7 km to the northwest, or Mishmar HaEmeq, 15 km to the southwest³⁹.

Kfar HaHoresh – sample description

Among three petrous bones from different graves that were sampled (KFH 1-3) only KFH2 had sufficient genomic data for subsequent genetic analysis (Supplementary Data 1).

KFH2 derives from square J53 (elevation 390/400cm below datum; L1003 catalogue #1050) in the multiple grave, Locus 1003, underlying the corner of a quadrilateral plastered surface, L1001 (Supplementary Figure 12). It is the nearly complete cranium of a 0-3 year old infant, genetically identified here as a female. The same petrous bone that was used for the DNA analysis was radiocarbon dated to 7,712-7,589 cal BCE (2 sigma range 95.4%) (MAMS 30693) (Supplementary Table 1). The C14 date and stratigraphic considerations indicate the grave dates to the transition from the Middle to Late PPNB phase at the site.

Ba'ja - archaeological information

The Neolithic site of Ba'ja (35°27'45" E / 30°24'55" N) was discovered during a survey in 1983 by M. Lindner and identified as a late Pre-Pottery Neolithic site by Hans Georg K. Gebel⁴⁰. Three soundings were carried out in 1984. Large scale excavations started in 1997 (co-directed by Bienert and Gebel) and were continued from 1999 until 2007 directed by Gebel⁴¹. In 2008, 2010 and 2012 special investigations at the site and in the region were carried out⁴². Besides several test units, five large areas have been excavated so far (B-South, B-North, C, D, F). In 2016, during a pilot project, new burials were discovered⁴¹, including that of BAJ001, the

individual genetically analyzed in this study. Based on the promising results of the test phase, a new 3-years project started in 2018.

Ba‘ja is famous for its extraordinary location in a naturally secluded setting (altitude: 1140-1175 m a.s.l.), surrounded by steep slopes (Supplementary Figure 13B). Access is possible through an up to 70 m deep gorge, the *Siq al-Ba‘ja*. Though the intramontane basin on which the village rests is only about 1.5 ha large, the site is considered a *mega-site*⁴³, since it resembles in some aspects other sites of this phenomenon, such as Beidha⁴⁴, Basta and other middle and late Pre-Pottery Neolithic sites of the southern Levant⁴⁰.

Radiocarbon dates confirm the typological dating to the late Pre-Pottery Neolithic B (second half of the 8th millennium BCE)⁴⁵. Preliminary analyses of archaeobotanical and archaeozoological analyses were presented by Neef (1997) and von den Driesch et al. (2004)⁴⁶. Beside domesticated cereals (especially *Triticum dicoccum*), wild fruits such as pistachio, hawthorn, and fig were collected; charcoal analyses comprise juniper and pistachio, but no remains of oak have been discovered so far. Animal husbandry was dominated by ovicaprines, but hunting also played an important role (for meat and fur). Taken together the species suggest a year-round occupation of the site. This is corroborated by the elaborate architecture. In every trench, densely packed clusters of buildings were discovered with at least two-storeyed terraced buildings with cellar-type substructures, indicating that the whole plateau was once occupied in a pueblo-like manner⁴⁵. “Ba‘ja’s final occupation, interrupted by at least one earthquake, is reconstructed as a densely built village without open spaces and lanes, with houses/ rooms accessible from roof tops or lower roofs, representing the settlement’s communal space”⁴¹.

Flint industries (thoroughly studied by^{42, 47}) as well as the production of various other objects, above all sandstone rings⁴⁸ and beads, provide valuable information on social identities, exchange and development supported by immaterial values. As suggested by Hans Georg Gebel

depositions of objects related to households and burials beneath floors might relate to practices of “avoidance, strengthening, fear, commodification and recommodification”^{41, 49}.

The dead were buried either in abandoned houses, in between houses or beneath floors. One primary burial in the most western part of the site and three collective burials had been discovered in Area C and D during earlier excavations⁵⁰⁻⁵². Results of palaeoanthropological work were presented by Schultz et al.^{53, 54}. Two new burials were discovered during the 2016 season in Area C, Room CR35 (Loci 405 and 408). These burials and further observations in adjacent rooms of Area C suggest that this area had been used as an intramural burial ground between the earliest and later architectural phases⁴¹.

Loc. 405 is a double burial of two infants aged 0.5-1 year and 3-4 years (labeled here as BAJ001 and BAJ002 respectively). They were buried in a crouched position, squeezed in a rather small pit. BAJ001 oriented E-W slightly above BAJ002. BAJ002 was oriented W-E. The two were facing each other.

Loc. 408 is a single primary burial which is outstanding in several respects. The grave construction as well as the burial ritual was very complex (for a detailed description see⁴¹). Moreover, the young adult individual (labeled here BAJ003) was buried with two categories of “grave goods”: seemingly personal items such as beads of various exotic raw-materials, arm rings on each upper arm still in situ and a deliberately destroyed “mace head” near the left shoulder. Additionally more objects were embedded in the grave cover.

Individual BAJ003 was lying on its left side with the legs in a crouched position. The orientation of the skeleton was SW-NE with the orientation of the face remaining unknown because the skull had fallen onto the chest, the mandible being turned upside down. Taphonomic

processes indicate that there must have been a void into which the sand had penetrated only after decay and that the head was originally slightly elevated⁴¹.

Similar to the collective burial in the same room, both graves were covered with stones slabs, which were then buried with up-to-fist sized stones and reused plaster/limestone fragments from an ancient floor.

A charcoal sample from the upper filling of the double infant burial was dated to 7027-6685 BCE (2 sigma range; MAMS 3015: 7928± 27 BP).

Out of the three sampled individuals (BAJ001 – BAJ003) only BAJ001 had sufficient genomic data for subsequent genetic analysis (Supplementary Data 1).

Ba'ja - Sample description

The double infant burial of Locus 405 comprised two complete skeletons of young infants (BAJ001 and BAJ002). The bones are of brittle consistency, inhibiting some morphological measurements. The preservation of the bone surfaces in both skeletons is rather good.

Age at death was estimated using the dental development, lengths of long bones, development of the cranium, carpals, and vertebrae⁵⁵. Applying these methods, age at death of the younger infant (BAJ001) was estimated between 0.5-1 year and of the older infant between 3-4 years (BAJ002).

Sex was roughly estimated by the markers of the mandible which suggested the BAJ001 individual was a male and BAJ002 a female. However, the genetic sexing determined BAJ001 to be a female (Table 1). Both individuals are of strong and short stature (When compared to the expected stature from the level of teeth development). The younger infant shows porosity and newly built bone plaques on the internal lamina of the skull, possibly being remnants of severe bleeding.

Supplementary Note 2: Genetic analysis of Neolithic Levantines

The Near-Eastern Levantine corridor, a narrow strip of land parallel to the Mediterranean Sea extending from the Sinai Peninsula to the north of Syria was one of the earliest centers of farming and cultivation⁵⁶. Recently reported early farmer genomes from the southern Levant showed considerable genetic continuity with Epipaleolithic Natufian individuals. However, they harbored additional admixture from an Anatolian-Neolithic related gene pool⁵⁷, providing a first glimpse at the demographic history throughout the Neolithic transition in this area. However, due to the poor DNA preservation in the region, the available genomic data is mostly of low coverage and limited to a handful of sites. Therefore, additional genomes that would fill temporal and geographical gaps in the available data could shed light on the demographic structure and heterogeneity of both the Levantine hunter-gatherers and early farming Levantine populations. We compared the newly produced genome-wide data of two individuals, a ca 9,600 ya Levantine PPNB early farmer (KFH2) from the site of Kfar HaHoresh in northern Israel and another Levantine PPNB early farmer (BAJ001) from the site of Ba'ja, Jordan (Table 1), to the previously published genome-wide data of contemporaneous individuals from the sites of Motza, Israel and 'Ain-Ghazal, Jordan⁵⁷ (grouped together and labeled Levant_N, following a previous labeling system²). We estimated an average coverage of 0.16 and 0.75 fold with 67,535 and 254,565 covered SNPs overlapping with the Human Origins dataset for KFH2 and BAJ001 respectively. While the coverage of KFH2 exceeds our threshold for analysis, we note that it is relatively low and provides limited statistical power and resolution. We determined the genetic sex of both KFH2 and BAJ001 as females and therefore we could not estimate their nuclear contamination rate based on X chromosome. However, based on the mitochondrial contamination estimate, both genomes were suitable for analysis with 0-6 % contamination estimates (Supplementary Table 2).

We first projected the ancient samples onto the first two dimensions of PCA (PC1, PC2) calculated for present-day west Eurasians (Materials and Methods and Fig. 1B). KFH2, BAJ001 and the published Levantine early farmers all fall in the vicinity of the Natufian cluster, shifted to the direction of Anatolian Neolithic populations, along both PC1 and PC2; among the Neolithic Levant individuals, KFH2 is further shifted from the others to this direction. In ADMIXTURE analysis ($K = 10$), BAJ001, KFH2 and the published Levant_N are all modeled as a mixture of a component maximized in Natufians (84.1 %, 89.8 % and 67.8 – 89.6 % for KFH2, BAJ001 and Levant_N, respectively) and a second component maximized in Mesolithic western hunter-gatherers (WHG) (Supplementary Figure 1).

We formally tested the diversity visualized on PCA and ADMIXTURE between the new and published Levantine early farmers by the D -statistic of the form $D(KFH2/BAJ001, Levant_N/KFH2; test, Mbuti)$ using both ancient and present-day worldwide populations as “test” (Supplementary Figure 14 A-D and Supplementary Data 4). We found the new and published individuals to be symmetrically related to most test populations within our data’s resolution. The only exceptions were a slight additional affinity of Levant_N with Early Neolithic individuals from the Peloponnese region in Greece⁵⁸ (labeled Greece_EN) and with KFH2 when compared to BAJ001 ($D = -4.2$ and -3.2 SE, respectively; Supplementary Figure 14D).

In accordance with this result, KFH2 can be modeled with qpWave as one stream of Levant_N ancestry ($\chi^2 p=0.473$ for rank=0), whereas two-way admixture models of KFH2 with Anatolian early farmer ancestry on top of Levant_N did not fit (Supplementary Table 8). In addition, modeling KFH2 as a two-way mixture of Levant_N and WHG lacked resolution to detect whether KFH2 has additional WHG related ancestry compared to Levant_N. Similarly,

when Levant_N was modeled with additional Anatolian early farmer ancestry on top of BAJ001, the model lacked resolution to determine whether Levant_N comprised a higher proportion of Anatolian early farmer like ancestry or Early Neolithic Peloponnese one compared to BAJ001 and the two-way model of BAJ001 and WHG did not fit (Supplementary Table 8).

5 KFH2, BAJ001 and Levant_N could all be separately modeled as two-way mixtures of around 75 – 85 % Natufian related ancestry and the rest from Anatolian early farmer ancestry (Supplementary Table 8). This result confirms the previously reported Levantine Neolithic ancestral mixture and indicates that the here reported Kfar HaHoresh and Ba’ja individuals share a similar ancestral composition with the published Motza and ‘Ain-Ghazal ones. We do not rule
10 out the possibility that the non-significant D-statistics are due to limited statistical power of our data. Further sampling is needed to investigate the question of the genetic diversity within the Levant Neolithic populations.

 The published and new Levantine early farmers could be grouped into one population (labeled Levant_Neol) and modeled as a mixture of Natufians and AHG or AAF (18.2 ± 6.4 %
15 AHG or 21.3 ± 6.3 % AAF ancestry; Supplementary Tables 4 and 8 and Supplementary Data 4), supporting a previously reported gene flow from an Anatolian Neolithic like population to the Levantine Neolithic gene pool⁵⁷. Moreover, we find ACF have additional genetic affinity compared to the earlier AAF that is best represented by the ancient Levantine gene pool (Fig. 2B), suggesting that the described genetic exchange between the Neolithic Levantine and
20 Anatolian gene pools was bidirectional.

Supplementary Note 3: Admixture modeling of the Ancient Anatolian populations

While we observe a long-term persistence of the local hunter-gatherer gene pool in Anatolia throughout the Neolithic (Fig. 1C, Supplementary Table 4), PCA and formal f -statistics suggest that the Anatolian hunter-gatherer (AHG), Anatolian Aceramic farmers (AAF) and Anatolian Ceramic farmers (ACF) differ in affinities to certain modern and ancient populations, likely due to differences in external genetic contributions to each of these two early farmer populations. To trace the ancestral sources of these ancient Anatolian populations we used qpAdm-based admixture modeling⁵⁹ that tests and models admixture proportions from potential source populations (“reference” populations herein) without assuming an explicit phylogeny.

For estimating admixture proportions in AHG, AAF and ACF we defined a basic set of seven outgroups, comprised of the following ancient and present-day populations.

‘Basic set’ = Han; Onge; Mbuti; Mala; Mixe; Natufian⁵⁷; Kostenki14⁶⁰

These outgroups were chosen to distinguish the ancestry of the reference populations since they broadly represent the known global genetic diversity and are unlikely to harbor recent gene flow with the target or reference populations either due to geographical/temporal distance or based on their genetic clustering in ADMIXTURE and PCA analysis⁵⁹. The modern outgroups (Han; Onge; Mbuti; Mala; Mixe) represent a global genetic variation outside west Eurasia. The Levantine Natufian⁵⁷ population (ca 12,000 years ago) and the European Upper Palaeolithic Kostenki14⁶⁰ (ca 37,000 years ago) both represent a gene pool outside of modern genetic variation. In some cases, when a reference population did not significantly contribute to the target in the attempted admixture models, it was removed from the reference set and added to the basic outgroup set in order to increase statistical power to distinguish the references.

As a prerequisite, we tested whether each set of reference populations can be distinguished by the chosen outgroups using qpWave⁵⁷. The chosen outgroups clearly distinguished the corresponding references in all tests we performed ($\chi^2 p \leq 7.70 \times 10^{-33}$). For both qpWave and qpAdm we use a significance level of $p=0.05$ for rejecting models.

To increase statistical power, individuals were grouped together under the analysis labels: AHG, AAF and ACF. To test differential affinities within the AAF individuals, we performed the *D*-statistic of the form *D* (*ind1*, *ind2*; *test*, *Mbuti*) which resulted in non-significant results for all tested pairs to the exception of individual ZHAJ that showed slightly higher affinities ($-3.50 < Z < -3.01$) than other AAF individuals with some Asian related populations (Supplementary Data 10).

The Anatolian hunter-gatherer (AHG)

As expected from the PCA results (Fig. 1C) and as reflected by the *D*-statistics of the form *D* (*AHG*, *pop1*; *pop2*, *Mbuti*) (Supplementary Table 3), AHG does not form a clade with Late Pleistocene or early Holocene Near-Easterners (Natufian, Levant_N or Iran_N) nor with Mesolithic hunter-gatherers from Europe (WHG and EHG). We therefore used the above populations, which are maximally differentiated in the PCA as potential sources of the AHG ancestry (Supplementary Table 5). For this analysis, Levant_N was chosen as a proxy for the Levantine late Pleistocene gene pool.

All two-way models were rejected except for the two-way admixture ($\chi^2 p = 0.158$) of a Neolithic Levantine-related gene pool (48.0 ± 4.5 %; estimate ± 1 SE) and a WHG-related gene pool (52.0 ± 4.5 %; estimate ± 1 SE). A three-way model including EHG as the third source did not increase the fit in comparison to the simpler nested two-way Levant_N + WHG model ($\chi^2 p$

= 0.717) and one with Iran_N as the third source only marginally increased the fit ($\chi^2 p = 0.081$; $11.9 \pm 6.9 \%$; estimate ± 1 SE).

While these results do not suggest AHG received direct gene flow from the tested sources (which are younger than AHG), they clearly support the presence of both Levantine and European hunter-gatherer related ancestries in central Anatolia during the Pleistocene.

Anatolian Aceramic farmers (AAF)

Inspired by the observed genetic similarity between the Anatolian hunter-gatherer and farmers as visualized in PCA and ADMIXTURE (Fig. 1C and Supplementary Figure 1) as well as the cultural continuity evidenced by the archaeological findings^{1, 3}, we attempted to estimate the contribution of the endogenous AHG gene pool in AAF. Furthermore, we traced potential external genetic contributions using *D*- statistics and estimated their proportion with qpAdm.

Compared to AHG, AAF have a slight excess affinity with early Holocene populations from Iran or Caucasus and with present-day south Asians, which have also been genetically linked with ancient Iranian/Caucasus ancestry⁶¹⁻⁶² as shown by *D*(AAF, AHG; test, Mbuti) (Fig. 2A, Supplementary Figures 2 –3 and Data tables S3 – S11). We therefore attempted to model AAF using Iran_N and AHG as two source populations. We also tested other combinations of the four reference populations mentioned above (Supplementary Table 6).

Using the basic outgroup set, the two-way model of AHG and Iran_N provided a good fit but with a rather big standard error estimate of ancestry proportion ($\chi^2 p = 0.054$; $8.6 \pm 7.2\%$ Iran_N ancestry). To increase model resolution, we added Levant_N, EHG and WHG to the outgroup set; here we found a well- consistent model with a smaller standard error estimate ($\chi^2 p = 0.296$; $10.3 \pm 3.9\%$ Iran_N ancestry). Modeling AAF as a sister clade of AHG (one way model

without contribution from Neolithic Iranians) results in a significantly reduced fit ($\chi^2 p = 0.014$). In the better fitting model the AHG gene pool comprises most of the AAF ancestry (89.7 ± 3.9 %), suggesting a high degree of genetic continuity in central Anatolia from the Epipaleolithic to the Neolithic past the emergence of farming.

Our results also suggest that the additional Neolithic Iran or Caucasus related ancestry (10.3 ± 3.9 %) diffused into central Anatolia during the same 5,000-year period, although for now we cannot narrow it down further due to lack of ancient genomes between AHG and AAF. Genome-wide data from additional AHG and AAF individuals could also help to increase our resolution and more accurately quantify the differences in ancestry between the two populations.

Anatolian Ceramic farmers (ACF)

Using a similar approach as for AAF, we estimated the contribution of the AAF gene pool in ACF and used *D*-statistics to detect potential external genetic contributions and estimate their proportion.

ACF share excess affinity with the early Holocene Levantines compared with AAF, as shown by positive *D* (*ACF*, *AAF*; *test*, *Mbuti* when “test” has Levantine related ancestry) (Fig. 2B, Supplementary Figures 4-5 and Supplementary Data 3). When the “test” populations are ancient Iran/Caucasus related populations and contemporary South Asians excess allele sharing with ACF is not observed ($Z < 1.3$).

We can model ACF as a mixture of Neolithic Levantines and AAF ($\chi^2 p = 0.606$; Supplementary Table 7). All the other tested models with AAF as a source are either rejected or produce infeasible proportions. When replacing AAF with AHG as the source population the three-way model with Levant_N and Iran_N as additional sources works well, confirming the

two sources of gene flow entering central Anatolia between the Epipaleolithic and the Ceramic Neolithic. The two-way model with only AHG and Levant_N also fits well ($\chi^2 p = 0.115$; table S7). However, this is likely due to limited power of our data to detect such a small contribution in a complex three-way model scenario.

- 5 When ACF is modeled by AAF and Levant_N, the AAF gene pool still comprises more than 3/4 of the ancestry in ACF ($78.7 \pm 3.5 \%$), suggesting that the hunter-gatherer gene pool persisted in the region for at least 2,000 years more and indicating limited influence from external gene pools during the Neolithic.

Supplementary Note 4: Investigating genetic links between Near-Eastern and European hunter-gatherers

The Anatolian Epipaleolithic hunter-gatherer (AHG) and the Mesolithic European hunter-gatherers (WHG and EHG) show a considerable degree of genetic differentiation in PCA (Fig. 1C). Nonetheless, central Anatolia geographically connects Europe to the Near East and with major climatic changes affecting the region during the last glaciation⁶³ it is not unlikely that Anatolia was the ground for East and West genetic exchange during the Palaeolithic. A recent study reported an affinity between modern Near-Easterners and European hunter-gatherers post-dating 14,000 years ago compared to earlier ones⁶⁴. With ancient genetic data available, we could directly compare the Near-Eastern hunter-gatherers (AHG and Natufian; labeled “*Near-Eastern HGs*”) and the European hunter-gatherers by $D(\text{European HG}, \text{Kostenki14}; \text{AHG/Natufians}, \text{Mbuti/Altai_published.DG})$. We used the 37 thousand-year-old *Kostenki14*^{60, 64} individual which is the oldest available European genome with genetic affinity to later European hunter-gatherers as a base line representing European HG pre-dating 14,000 years ago. This statistic resulted in significantly positive values for almost all individuals post-dating 14,000 years ago (“*later European HG*”) when positioned in “*European HG*” while for earlier ones (“*earlier European HG*”) the statistic was less positive on average and reached significance in only some individuals (Fig. 3A and Supplementary Data 5). These results suggest increased genetic affinity of later European HGs with the Near-Eastern HGs compared to the earlier ones as previously observed for modern Near-Eastern populations.

Interchanging the central African *Mbuti* with the Altai Neanderthal (*Altai_published.DG*) as an outgroup did not significantly alter the results, confirming that the observed affinities are not

caused by differing levels of Neanderthal ancestry in the tested hunter-gatherers (Supplementary Data 5).

One particular population among the later European HGs, the recently reported Mesolithic hunter-gatherers from the Balkan peninsula ('Iron Gates HG')⁵⁸, shows the most allele sharing with AHG in $D(\text{Iron_Gates_HG}, \text{European HGs}; \text{AHG}, \text{Mbuti/Altai})$; Fig. 3A, Supplementary Figures 6-7 and Supplementary Data 5). The Iron Gates HG population was previously modeled as a mixture of WHG, EHG and a third unknown ancestral component⁵⁸.

The geographic location of the 'Iron Gates' site within the natural corridor connecting the Near East, through Anatolia, with continental Europe as well as the genetic affinities observed in the above D -statistic, motivated us to consider Near Eastern HGs as potential sources for the Iron Gates ancestry. We modeled Iron Gates HG as a three-way mixture of Near-Eastern HGs or the Iran_N population (used as a proxy for Iranian hunter-gatherer ancestry) (Supplementary Table 9). The model in which Iran_N is used as the third source population was rejected. However, we can model the Iron Gates HG as a three-way mixture of AHG or Natufian (25.8 ± 5.0 % or 11.1 ± 2.2 % respectively), WHG (62.9 ± 7.4 % or 78.0 ± 4.6 % respectively) and EHG (11.3 ± 3.3 % or 10.9 ± 3 % respectively) (Supplementary Tables 4 and 9).

It should be noted that the published individuals from the Iron Gates region date several millennia later than AHG and include individuals that have been reported to be migrants from Anatolia showing northwestern Anatolian Neolithic-like ancestry⁵⁸. We excluded from our analysis the outlier individuals showing the above Anatolian farmer ancestry to avoid signals related to Neolithic interactions that postdate the formation of the observed Pleistocene genetic link between the Near East and Europe (Supplementary Data 2).

We tested whether a model in which a gene flow from a Near-Eastern ancestry is introduced into the ancestors of Iron Gates could sufficiently explain the excess affinity we observe between the two populations. For this purpose, we exploited the fact that Near-Eastern populations harbor a Basal Eurasian ancestry component (α) which is undetectable in European hunter-gatherers⁶⁴.

Therefore the Basal Eurasian ancestry could serve as a marker for Near-Eastern gene flow. We assessed the Basal Eurasian ancestry proportion by following a previously described approach of *qpAdm* modeling⁵⁷. This framework relies on the basal phylogenetic position of both the Basal Eurasian ghost population and an African reference (the ancient Ethiopian Mota genome⁶⁵) relative to other non-Africans. Therefore, by using a set of outgroups that includes eastern non-African populations (Han; Onge; Papuan) and Upper Palaeolithic Eurasian genomes (Ust_Ishim⁶⁶; Kostenki14; Malta_cluster⁶⁷) but neither west Eurasians with detectable basal Eurasian ancestry nor Africans, the mixture proportion computed for Mota (α) can be used indirectly to estimate the Basal Eurasian mixture proportion of west Eurasian populations (Supplementary Table 10).

We estimated α to be 24.8 ± 5.5 % in AHG (Fig. 3b and Supplementary Table 9) and 38.5 ± 5.0 % for Natufian, which is consistent with previous estimates. If we assume an Anatolia to Europe gene flow, we can use our estimate for AHG derived ancestry in the Iron Gates HG to calculate the expected proportion of Basal Eurasian ancestry in Iron Gates HG ($\% \text{ AHG in Iron Gates HG} \times (\alpha \text{ in AHG})$) resulting in an expected α of 6.4 %. Yet, we could model Iron Gates HG without any Basal Eurasian ancestry or when forcing “Mota” into the model, as comprising a non-significant 1.6 ± 2.8 % (Supplementary Table 10), suggesting that it is unlikely that unidirectional gene flow from the Near East to Europe alone can account for the Iron Gates HG and the Near-Eastern HG affinity. We propose a plausible scenario in which a genetic exchange

between populations ancestral to southeastern Europeans of the early Holocene and Anatolians of the late glacial occurred before 15,000 years ago (the age of AHG).

Supplementary Note 5: Mitochondrial DNA analysis

The Anatolian hunter-gatherer (AHG)

AHG (The Pinarbaşı Epipaleolithic individual ZBC) displays 31 polymorphisms from the rCRS and can be confidently assigned to mitochondrial haplogroup K2b.

5 Haplogroup K2 is a sub-clade of the major haplogroup K, which according to Maximum Likelihood estimates based on complete mtDNA sequences arose ~25-29 kya during the cooling period preceding the Late Glacial Maximum^{68, 69}. Bayesian estimates using several internal calibration points within haplogroup U have however provided a more recent date of ~18.5 ka (14.5-23.3 ka, 95% CI)^{69, 70}. Subclade K2b detected in Pinarbaşı has been dated by Maximum
10 Likelihood method to the Late Glacial ~17–18 ka⁶⁹. To date, K2 and its basal sub-clusters K2a, K2b and K2c have been almost exclusively detected in modern-day Europeans, which was used as an argument for a European origin of the whole sub-clade and recent back-migration from Europe into the Near East to explain the spurious presence of K2 haplotypes in the Near East⁶⁹. However, no K2 haplogroups have been found in pre-Neolithic Europe⁷⁰ and so far only one
15 ancient sample with haplogroup K2b has been reported in a Corded Ware individual from Esperstedt (Germany, 2500-2050 BCE)⁷¹. Most modern DNA reported mitogenomes belong to sub-clusters K2b1 or in less frequency to K2b2 within K2, with just two sequences at the root of K2.

20 While the Pinarbaşı individual postdates the average time node of sub-cluster K2 by a couple of millennia, the presence of haplogroup K2b in Epipaleolithic Anatolia raises the possibility that this sub-clade could have a Near-Eastern origin. Interestingly, the Pinarbaşı individual lacks one of the terminal mutations characteristic of the K2b, a transition at position

14067. While this could represent a back-mutation event, it is also possible that it mutation emerged in the Near East after 13,000 BCE and was carried into Europe afterwards. The lack of other contemporaneous representatives of K2b does not allow distinguishing between both possibilities.

5

The Anatolian Aceramic farmers (AAF) and the Levantine early farmers

Boncuklu individuals **ZHAJ**, **ZHJ** and **ZKO** all display the 19 diagnostic mutations of mitochondrial haplogroup U3. Individual ZKO shows four extra differences from rCRS, including an insertion in the HVII poly-C stretch. Individuals ZHAJ and ZHJ share the same
10 mitochondrial haplotype, one mutational step away from ZKO (+4820A).

According to Maximum Likelihood estimates of modern mtDNA haplotype diversity, haplogroup U3 originated in the Near East during the Upper Palaeolithic ca. 32ka^{68, 72}. In modern populations this haplogroup is primarily found in the Near East and the Caucasus, while it is present in lower frequency or even absent in western European populations^{68, 72}.

15 The oldest report of haplogroup U3 corresponds to Boncuklu (this publication,¹²). In the Near East it is also present in two Ceramic Neolithic individuals (ACF) from Barçın in the Marmara region dated back to 6500-6000 cal BCE⁷¹, but it is absent in Tepecic Çiftik (5500-7800 cal BCE) and in contemporaneous PPNB-PPNC populations from the southern and northern Levant^{12,57,73}. The presence of this haplogroup in Aceramic and Ceramic Anatolian
20 Neolithic is in agreement with the genetic continuity between Anatolian pre-pottery and pottery Neolithic inferred from whole genome analyses.

In the Early European Neolithic haplogroup U3 only appears in two individuals belonging to the Starčevo (Hungary) and LBK (Germany) archaeological cultures⁷⁴. In Europe it is otherwise detected in the Middle Neolithic Salzmünde culture in Germany (N= 5) and in the Middle, Late Neolithic and Chalcolithic periods in Spain (N=5)^{57, 75-78}. Our results are in agreement with a concomitant spread of a few members of this haplogroup with the Neolithic, which however did not have a substantial demographic impact due to genetic drift.

Boncuklu individual **ZHAG** belongs to the subclade N1a1a1 within mitochondrial haplogroup N1a. The oldest report of this haplogroup corresponds also to Boncuklu (this publication,⁷) and as described for haplogroup U3, there are no contemporaneous parallels of subclade N1a1a1 in the PPNB Levantine populations whereas it was reported in later Ceramic Anatolian populations (Barçın and Mentese)⁷⁵. Sub-clade N1a1a1 and its derived cluster N1a1a1a (+16320T) are ubiquitously present in considerable frequencies in Early Neolithic European cultures (Starčevo, LBK, Epicardial), probably as a result of a founder effect following the spread of the Neolithic from Anatolia^{59,74}.

Individual **KFH2** from the PPNB archaeological site of Kfar Hahoresh is also classified as N1a, albeit from the sub-branch N1a1b. The present SNPs together with the absence of the five diagnostic positions leading to the more widely distributed sub-branch N1a1b1 places this haplotype at the root of N1a1b. Therefore, KFH2 represents the first reported prehistoric member of the N1a1b node that, according to modern phylogeographic mitochondrial data, originated ~28ka most probably in the Near East⁷⁹⁻⁸⁰. It is important to note that four of the diagnostic positions of the haplogroup are not covered and an additional 14 have a coverage ≤ 5 (Supplementary Data 6).

Individual **BAJ001** from the PPNB archaeological site of Ba'ja in Jordan harbours all the diagnostic SNPs characteristic of haplogroup N1b1a with the exception of mutation 1703, plus one extra transition in position 16519 and an extra C insertion in the HVRII poly-C tract. A back-mutation in position 1703, together with a T insertion in 455 and a transition in 8084, define the sub-branch N1b1a1 within N1b1a. The absence of 1703 in BAJ001 therefore suggests that this back-mutation emerged before 9,000 BP and preceded the other two substitutions.

Haplogroup N1b is extremely rare in Neolithic and post-Neolithic Near East and Europe, and has been reported so far in just two ancient individuals, one belonging to the sub-clade N1b2 from Ivanovo, Bulgaria, dated back to the Middle Chalcolithic (4,725-4,605 cal BCE)⁵⁸ and interestingly, one classified also as N1b1a in the Anatolian Ceramic Neolithic site of Barcin (6500-6200 BCE)⁷⁵. In modern populations, haplogroup N1b1 is found primarily in the Near East, with minor branches in Europe and North Africa. It reaches maximum frequencies in the southern Levant and in Ashkenazi Jewish groups⁸⁰. Sub-cluster N1b1a has been dated to ~13-14ka⁸⁰. In the same study, two scenarios of expansion have been postulated based on HVRI data: 1. during the Neolithic or 2. during the Late Glacial period⁸⁰. The presence of this sub-clade in Early Neolithic Ba'ja together with its scarceness during and after the Neolithic are more in agreement with the latter.

Boncuklu individual **ZMOJ** can be assigned to mitochondrial haplogroup K1a, the SNPs found in this sample place it at the root of this sub-clade. Until recently, haplogroup K as a whole had only been detected among farmers, however recent analyses have reported haplogroup K1 in eleven hunter-gatherer individuals, two from the Mesolithic site of Theopetra in Greece (7,605-7,529 cal BCE and 7,288-6,771 cal BCE)⁸¹, one from Măgura Buduiasca in Romania (6061-5985 cal BCE) and eight from several sites across the Iron Gates region in Romania and Serbia (ca.

5800-9000 cal BCE)⁵⁸. Moreover one hunter-gatherer from Satsurblia in Georgia (11,430-11,180 cal BCE) carried haplogroup K3⁸².

With the exception of one individual from Padina (6,061-5,841 cal BCE) with admixed hunter-gatherer and Anatolian Neolithic ancestry, none of these hunter-gatherers have been classified as belonging to sub-cluster K1a (K3, K1, K1c and K1f). Therefore, current evidence restricts this clade to Neolithic and post-Neolithic Near-Eastern and European individuals.

With the exception of individual ZHF from Boncuklu, who also carries the root haplotype of sub-clade K1a¹², Early farmers from the Levant and Anatolia belong to derived K1a sub-lineages (mainly K1a18, K1a2, K1a3, K1a4 and K1a12). The whole cluster reaches a high frequency (33%) among Anatolian ceramic farmers (Barcin, Mentese and Tepeçik Ciftik)^{12, 75} and it is present in almost all early Neolithic cultures at frequencies between 10 and 20%^{59,74, 76, 83}.

According to modern phylogeographic studies on the diversity and distribution of K1a mitotypes, the coalescence age of K1a has been estimated in ~20 ka⁶⁹. The absence of pre-farming representatives of K1a in Europe point out a more probable Near-Eastern origin for this sub-clade, and the spread of the Neolithic as the main source of its dispersal and diversification.

Supplementary References

1. Baird, D. *et al.* Juniper smoke, skulls and wolves' tails. The Epipalaeolithic of the Anatolian plateau in its South-west Asian context; insights from Pınarbaşı. *Levant* **45**, 175-209 (2013).
2. Watkins, T. Excavations at Pınarbaşı: The early stages. *On the Surface: Çatalhöyük* **1995**, 47-57 (1993).
3. Baird, D. *et al.* Agricultural origins on the Anatolian plateau. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E3077-E3086 (2018).
4. Baird, D. Pınarbaşı; from Epipalaeolithic campsite to sedentarising village in central Anatolia. *The Neolithic in Turkey* **3**, 181-218 (2012).

5. Baird, D. The Late Epipaleolithic, Neolithic, and Chalcolithic of the Anatolian Plateau, 13,000–4000 BC. *A companion to the archaeology of the Ancient Near East*, 431-465 (2012).
6. Baird, D., Fairbairn, A., Martin, L. & Middleton, C. in Archaeology and Art Publications. *The Boncuklu Project: the origins of sedentism, cultivation and herding in central Anatolia* (2012).
- 5 7. Özbaşaran, M. Aşıklı. *The Neolithic in Turkey* **3**, 135-158 (2012).
8. Hodder, I. Çatalhöyük: the leopard's tale: revealing the mysteries of Turkey's ancient 'town'. *Thames & Hudson* (2006).
9. Byrd, B. F. Reassessing the emergence of village life in the Near East. *Journal of Archaeological Research* **13**, 231-290 (2005).
- 10 10. Kuijt, I., Guerrero, E., Molist, M. & Anfruns, J. The changing Neolithic household: Household autonomy and social segmentation, Tell Halula, Syria. *Journal of Anthropological Archaeology* **30**, 502-522 (2011).
11. Pearson, J. in Boncuklu; first farmers in central Anatolia and the antecedents of Çatalhöyük. From foragers to farmers in central Anatolia, D. Baird, A. Fairbairn, Eds. (forthcoming), vol. 1.
- 15 12. Kılınç, G. M. *et al.* The demographic development of the first farmers in Anatolia. *Current Biology* **26**, 2659-2666 (2016).
13. Goring-Morris, A. N. A PPNB settlement at Kfar Hahores in lower Galilee: A preliminary report of the 1991 season. *Journal of the Israel Prehistoric Society* **24**, 77-101 (1991).
- 20 14. Goring-Morris, A. N., Goren, Y., Horwitz, L. K., Bar-Yosef, D. & HersHKovitz, I. Investigations at an early Neolithic settlement in the lower Galilee: Results of the 1991 season at Kfar Hahores. *Atiqot* **27**, 37-62 (1995).
15. Goring-Morris, A. N. *et al.* The 2007–8 excavation seasons at pre-pottery Neolithic B Kfar HaHores, Israel. *Antiquity* **82**, 1151ff (2008).
16. Goring-Morris, N. in *Life in Neolithic farming communities* 103-136 (Springer, 2002).
- 25 17. Horwitz, L. K. & Goring-Morris, N. Animals and ritual during the Levantine PPNB: a case study from the site of Kfar Hahores, Israel. *Anthropozoologica* **39**, 165-178 (2004).
18. Goring-Morris, N. & Horwitz, L. K. Funerals and feasts during the Pre-Pottery Neolithic B of the Near East. *antiquity* **81**, 902-919 (2007).
- 30 19. Meier, J. S., Goring-Morris, A. N. & Munro, N. D. Depositional histories of faunal remains from the Neolithic cultic site of Kfar HaHores, Israel. *Journal of Anthropological Archaeology* **48**, 233-249 (2017).

20. Goren, Y. & Goring-Morris, A. Early pyrotechnology in the Near East: Experimental lime-plaster production at the Pre-Pottery Neolithic B site of Kfar HaHoresh, Israel. *Geoarchaeology* **23**, 779-798 (2008).
- 5 21. Barzilai, O. & Goring-Morris, A. N. Bidirectional blade and tool caches and stocks in the PPNB of the southern Levant. *Technical Systems and Near Eastern PPN Communities*, 227-294 (2007).
- 10 22. Davidzon, A. & Goring-Morris, N. Knapping in the graveyard: a refitted naviform sequence from Kfar HaHoresh, Lower Galilee, Israel. *Systemes techniques et communautés du Néolithique précéramique au Proche-Orient Technical Systems and Near Eastern PPN Communities*. Antibes: Editions APDCA, 295-309 (2007).
23. Goring-Morris, A. N. in *Life, death and the emergence of differential status in the Near Eastern Neolithic: evidence from Kfar HaHoresh, Lower Galilee, Israel* (2005).
- 15 24. Goren, Y., Goring-Morris, A. N. & Segal, I. The technology of skull modelling in the Pre-Pottery Neolithic B (PPNB): regional variability, the relation of technology and iconography and their archaeological implications. *Journal of Archaeological Science* **28**, 671-690 (2001).
25. Eshed, V., HersHKovitz, I. & Goring-Morris, A. N. A re-evaluation of burial customs in the Pre-Pottery Neolithic B in light of paleodemographic analysis of the human remains from Kfar HaHoresh, Israel. *Paléorient*, 91-103 (2008).
- 20 26. Goring-Morris, A. N. *et al.* The 1997 season of excavations at the mortuary site of Kfar HaHoresh, Galilee, Israel. *Neo-lithics* **3**, 1-4 (1998).
27. HersHKovitz, I. *et al.* Remedy for an 8500 year-old plastered human skull from Kfar HaHoresh, Israel. *Journal of Archaeological Science* **22**, 779-788 (1995).
- 25 28. Alt, K. W., Benz, M., Vach, W., Simmons, T. L. & Goring-Morris, A. N. Insights into the social structure of the PPNB site of Kfar HaHoresh, Israel, based on dental remains. *PloS one* **10**, e0134528 (2015).
29. Simmons, T., Goring-Morris, N. & Horwitz, L. K. IN THE PRE-POTTERY NEOLITHIC B Mortuary COMPLEX AT KFAR HAHORESH, ISRAEL. *Faces from the Past: Diachronic Patterns in the Biology of Human Populations from the Eastern Mediterranean: Papers in Honour of Patricia Smith* **1603**, 100 (2007).
- 30 30. Makarewicz, C. A., Horwitz, L. K. & Goring-Morris, A. N. Local adoption of animal husbandry in the Southern Levant: an isotopic perspective from the Pre-Pottery Neolithic B funerary site of Kfar HaHoresh. *Environmental Archaeology* **21**, 199-213 (2016).
31. Meier, J. S., Goring-Morris, A. N. & Munro, N. D. Provisioning the Ritual Neolithic Site of Kfar HaHoresh, Israel at the Dawn of Animal Management. *PloS one* **11**, e0166573 (2016).

32. Meier, J. S., Goring-Morris, A. N. & Munro, N. D. Aurochs bone deposits at Kfar HaHoresh and the southern Levant across the agricultural transition. *antiquity* **91**, 1469-1483 (2017).
33. Barzilai, O. & Goring-Morris, A. N. Bidirectional blade production at the PPNB site of Kfar HaHoresh: the techno-typological analysis of a workshop dump. *Paléorient*, 5-34 (2010).
- 5 34. Biton, R., Goren, Y. & Goring-Morris, A. N. Ceramics in the Levantine Pre-Pottery Neolithic B: evidence from Kfar HaHoresh, Israel. *Journal of Archaeological Science* **41**, 740-748 (2014).
35. Mayer, D. E. B. Shell ornaments and artifacts in Neolithic Cyprus and correlations with other Mediterranean regions. *Quaternary International* (2017).
- 10 36. Bar-Yosef Mayer, D. E. & Porat, N. Green stone beads at the dawn of agriculture. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 8548-8551 (2008).
37. Birkenfeld, M. & Goring-Morris, A. Out of sight: The role of Kfar HaHoresh within the PPNB landscape of the lower Galilee, Israel. *Archaeological Review from Cambridge* **30**, 95-14 (2015).
- 15 38. Garfinkel, Y. *et al.* in *The pre-pottery Neolithic B village of Yiftahel: The 1980s and 1990s excavations* (Hebrew University of Jerusalem, Jerusalem, Israel, 2012).
39. Barzilai, O. & Getzov, N. The 2010 excavation season at Mishmar Ha'emeq in the Jezreel Valley. *Neolithics* **1**, 19-22 (2011).
- 20 40. Bienert, H. & Gebel, H. G. K. Summary on Ba'ja 1997, and insights from the later seasons. *Central Settlements in Jordan. Studies in Early Near Eastern Production, Subsistence, and Environment* **5**. *Ex oriente, Berlin*, 119-144 (2004).
41. Gebel, H. G. K. *et al.* Household and Death. Preliminary results of the 2016 11th Season at Late PPNB Ba'ja, Southern Jordan. *Neo-Lithics* **1/17**, 18-36 (2017).
- 25 42. Purschwitz, C. Die lithische Ökonomie von Feuerstein im Frühneolithikum der Größeren Petra-Region. *Studies in Early Near Eastern Production, Subsistence, and Environment* **19**. *Ex Oriente, Berlin* (2017).
43. Gebel, H. G. K. Central to what? Remarks on the Settlement Patterns of the LPPNB Mega-Sites in Jordan. *Central Settlements in Jordan. Studies in Early Near Eastern Production, Subsistence, and Environment* **5**. *Ex Oriente, Berlin*, 1-20 (2004).
- 30 44. Byrd, B. F. *Early village life at Beidha, Jordan: Neolithic spatial organization and vernacular architecture: the excavations of Mrs. Diana Kirkbride-Helbæk* (Oxford University Press Oxford, 2005).

45. Kinzel, M. Am Beginn des Hausbaus: Studien zur PPNB-Architektur von Shkārat Msaied und Ba'ja in der Petra-Region, SüdJordanien. *Studies in Early Near Eastern Production, Subsistence, and Environment* **17**. Ex Oriente, Berlin (2014).
- 5 46. Von den Driesch, A., Cartajena, H. Manharth, The Late PPNB site of Ba'ja, Jordan: the faunal remains (1997 season). *Central Settlements in Neolithic Jordan. Studies in Early Near Eastern Production, Subsistence, and Environment* **5**. Ex Oriente, Berlin 271-288 (2004).
47. Purschwitz, C. Ba'ja 2012: Abiotic resources and Early Neolithic raw material procurement in the Greater Petra area (ARGPA). Research aims and first results. *Neo-Lithics* **1/13**, 3-10 (2013).
- 10 48. Michiels, T., Al-Souliman, A. S. & Gebel, H. G. K. Stage 3 manufacturing traces of the Ba'ja LPPNB sandstone rings. *Neo-lithics* **2/12**, 41-50 (2012).
49. Gebel, H. G. K. Commodification and the formation of early Neolithic social identity. The issues as seen from the southern Jordanian Highlands. The principle of sharing. Segregation and construction of social identities at the transition from foraging to farming. *Studies in Early Near Eastern Production, Subsistence, and Environment* **14**, 35-80 (2010).
- 15 50. Gebel, H. G. K., Hermansen, B. & Kinzel, M. Ba'ja 2005: A Two-Storied Building and Collective Burials. Results of the 6th Season of Excavation. *Neo-Lithics* **1**, 12-19 (2006).
51. Gebel, H. G. K. & Hermansen, B. D. LPPNB Ba'ja 2001. A Short Note. *Neo-lithics* **2/01**, 15-20 (2001).
- 20 52. Gebel, H. G. K. & Hermansen, B. D., Ba'ja 2003. Summary of the 5th season of excavation. *Neo-Lithics* **2/04**, 15-18 (2004).
53. Schultz, M., Schmidt-Schultz, T. H., Gresky, J., Kreutz, K. & Berner, M. *Die Menschen von Basta und Ba'ja im akeramischen Neolithikum*. In: B. Salje, N. Riedl, G. Schauerte and R.-B. Wartke (eds.), *Gesichter des Orients. 10000 Jahre Kunst und Kultur aus Jordanien*. Mainz: Verlag Philipp von Zabern 57-60 (2004).
- 25 54. Schultz, M., Schmidt-Schultz, T. H., Gresky, J., Kreutz, K. & Berner, M. Morbidity and mortality in the late PPNB populations from Basta and Ba'ja (Jordan). *BAR International Series*. **1603**, 82 (2007).
55. Beck, L. A. Standards for data collection from human skeletal remains. Edited by Jane E. Buikstra and Douglas H. Ubelaker. 272 pp. Fayetteville: Arkansas Archeological Survey Research Series No. 44, 1994. *Am. J. Hum. Biol.* **7**, 672-672 (1995).
- 30 56. Zohary, D., Hopf, M. & Weiss, E. in *Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin* (Oxford University Press on Demand, 2012).

57. Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419-424 (2016).
58. Mathieson, I. *et al.* The Genomic History of Southeastern Europe. **555**, 197-203 *Nature* (2018).
- 5 59. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207-211 (2015).
60. Seguin-Orlando, A. *et al.* Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113-1118 (2014).
- 10 61. Broushaki, F. *et al.* Early Neolithic genomes from the eastern Fertile Crescent. *Science* **353**, 499-503 (2016).
62. Narasimhan, V. M. *et al.* The Genomic Formation of South and Central Asia. *bioRxiv*, 292581 (2018).
63. Gamble, C., Davies, W., Pettitt, P. & Richards, M. Climate change and evolving human diversity in Europe during the last glacial. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **359**, 243-53; discussion 253-4 (2004).
- 15 64. Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200-205 (2016).
65. Gallego Llorente, M. *et al.* Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* **350**, 820-822 (2015).
- 20 66. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445-449 (2014).
67. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87-91 (2014).
68. Behar, D. M. *et al.* A “Copernican” reassessment of the human mitochondrial DNA tree from its root. *The American Journal of Human Genetics* **90**, 675-684 (2012).
- 25 69. Costa, M. D. *et al.* A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages. *Nature communications* **4**, 2543 (2013).
70. Pereira, L. *et al.* Population expansion in the North African late Pleistocene signaled by mitochondrial DNA haplogroup U6. *BMC evolutionary biology* **10**, 390 (2010).
- 30 71. Posth, C. *et al.* Pleistocene mitochondrial genomes suggest a single major dispersal of non-Africans and a Late Glacial population turnover in Europe. *Current Biology* **26**, 827-833 (2016).
72. Soares, P. *et al.* The archaeogenetics of Europe. *Current Biology* **20**, R174-R183 (2010).

73. Fernández, E. *et al.* Ancient DNA analysis of 8000 BC near eastern farmers supports an early Neolithic pioneer maritime colonization of Mainland Europe through Cyprus and the Aegean Islands. *PLoS genetics* **10**, e1004401 (2014).

5 74. Szecsenyi-Nagy, A. *et al.* Tracing the genetic origin of Europe's first farmers reveals insights into their social organization. *Proc. Biol. Sci.* **282**, 10.1098/rspb.2015.0339 (2015).

75. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503 (2015).

76. Brandt, G. *et al.* Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science* **342**, 257-261 (2013).

10 77. Szecsenyi-Nagy, A. *et al.* The maternal genetic make-up of the Iberian Peninsula between the Neolithic and the Early Bronze Age. *Scientific reports* **7**, 15644 (2017).

78. Gómez-Sánchez, D. *et al.* Mitochondrial DNA from El Mirador cave (Atapuerca, Spain) reveals the heterogeneity of Chalcolithic populations. *PLoS One* **9**, e105105 (2014).

15 79. Olivieri, A. *et al.* Mitogenomes from two uncommon haplogroups mark late glacial/postglacial expansions from the near east and Neolithic dispersals within Europe. *PloS one* **8**, e70492 (2013).

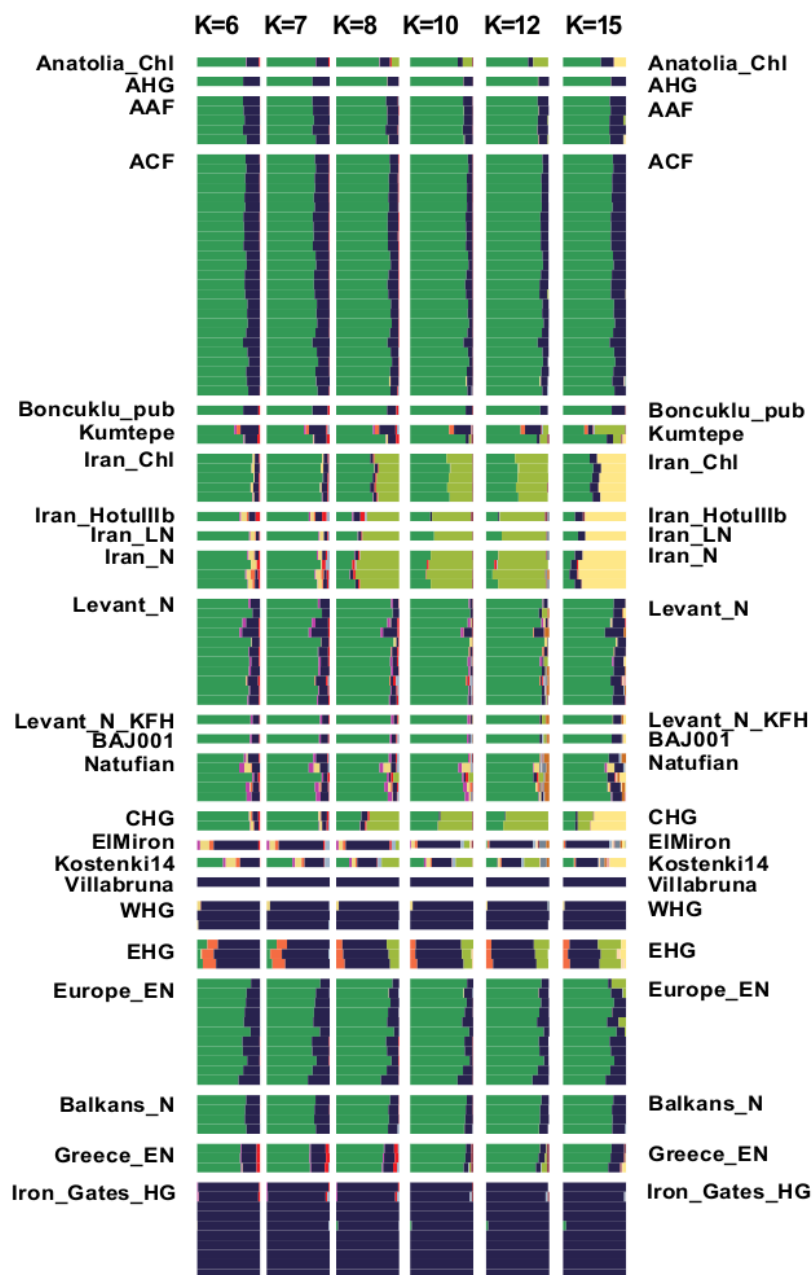
80. Fernandes, V. *et al.* The Arabian cradle: mitochondrial relicts of the first steps along the southern route out of Africa. *The American Journal of Human Genetics* **90**, 347-355 (2012).

20 81. Hofmanova, Z. *et al.* Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 6886-6891 (2016).

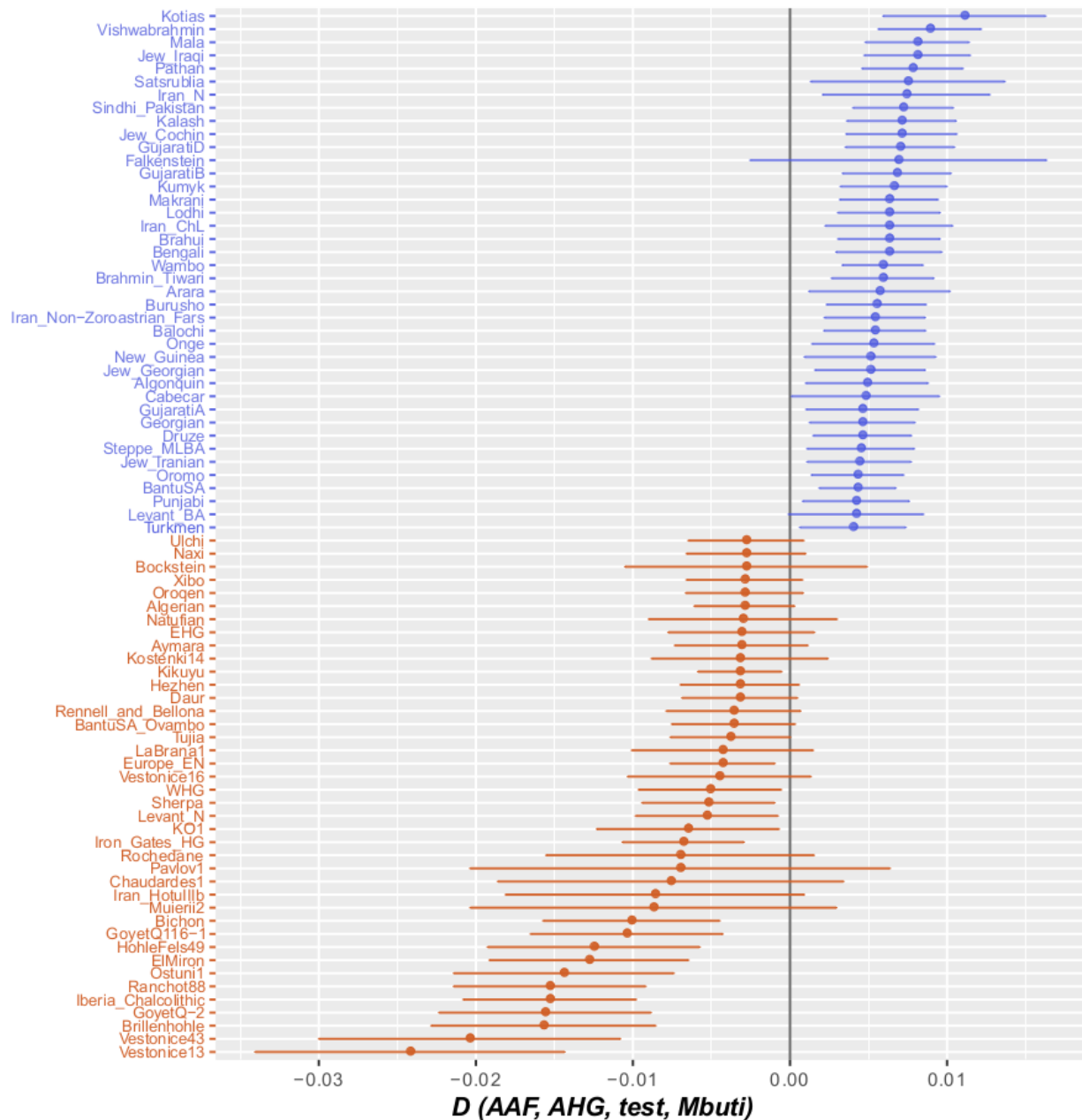
82. Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015).

83. Olalde, I. *et al.* A common genetic origin for early farmers from Mediterranean Cardial and Central European LBK cultures. *Mol. Biol. Evol.* **32**, 3132-3142 (2015).

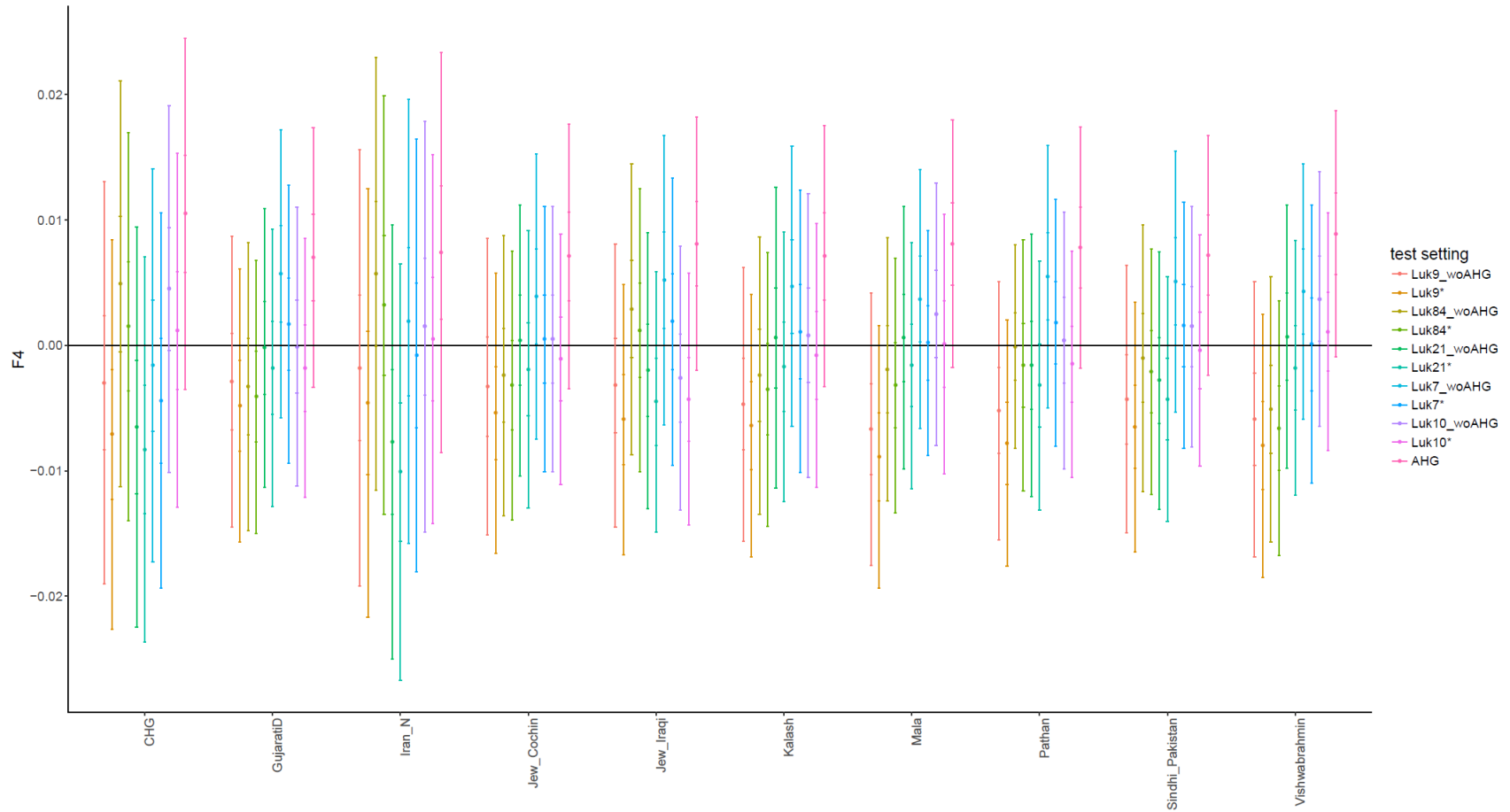
Supplementary Figures



Supplementary Figure 1. ADMIXTURE analysis. ADMIXTURE analysis (Materials and Methods) computed from 2,706 present-day and 594 ancient individuals is shown (K = 6-8; 10; 12; 15). A selected set of ancient individuals are plotted. For ADMIXTURE plot of all individuals (K=10) see Supplementary Figure 9. AHG, AAF and ACF compose of similar main ancestry components, maximized in Natufians (green) and WHG (blue). Source data are provided as a Source Data file.



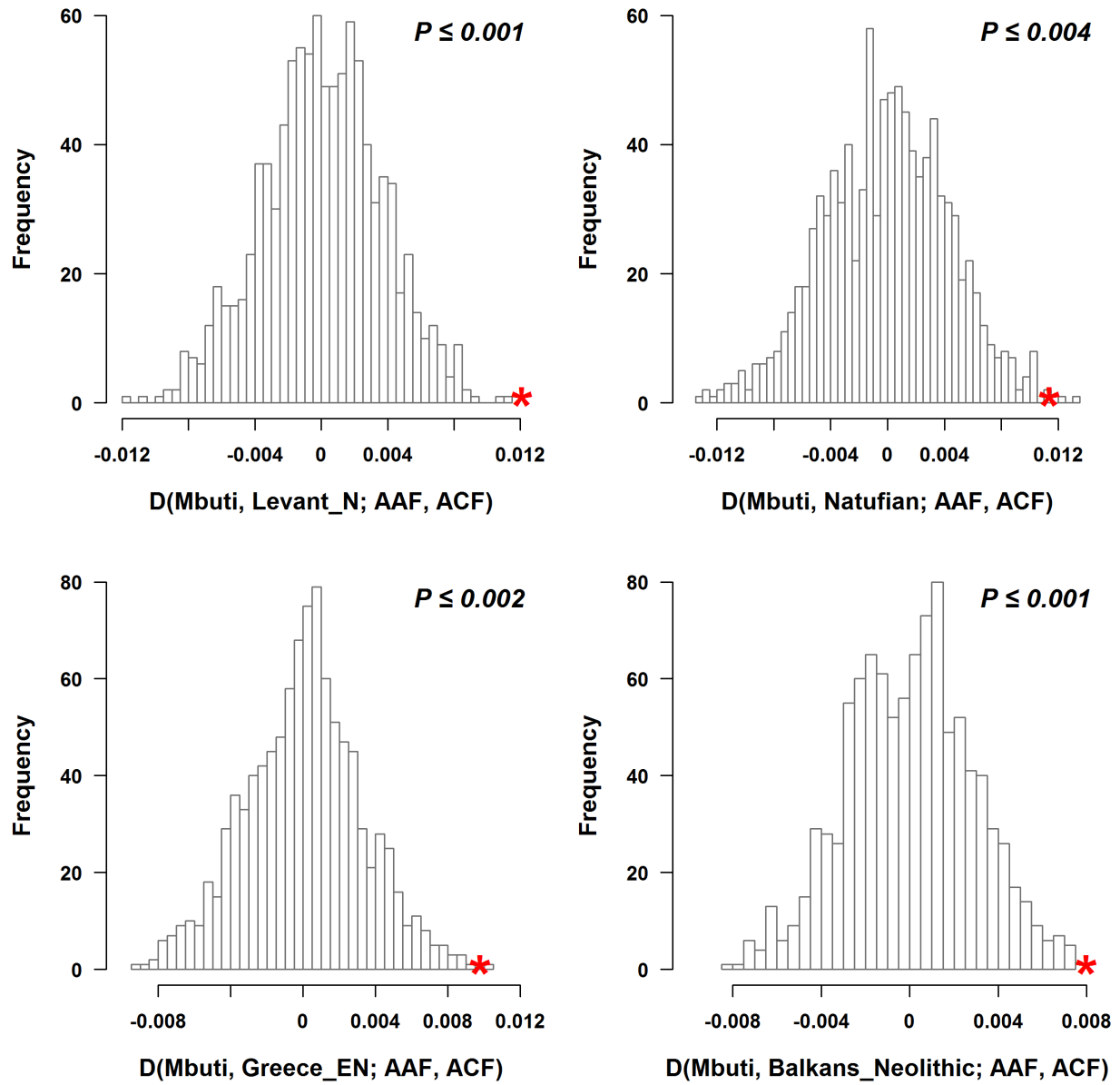
Supplementary Figure 2. AAF have excess allele sharing with Iranian/Caucasus related populations compared with AHG. We plot the populations with the 40 most positive (blue) and 40 most negative values (orange) of $D(AAF, AHG; test, Mbuti)$ with ± 1 standard errors estimated by 5 cm block jackknifing (represented by the horizontal bars). Positive values indicate that “test” shares more alleles with AAF than with AHG, and negative values that it shares more with AHG than AAF (analysis was restricted to individuals > 30,000 SNPs).



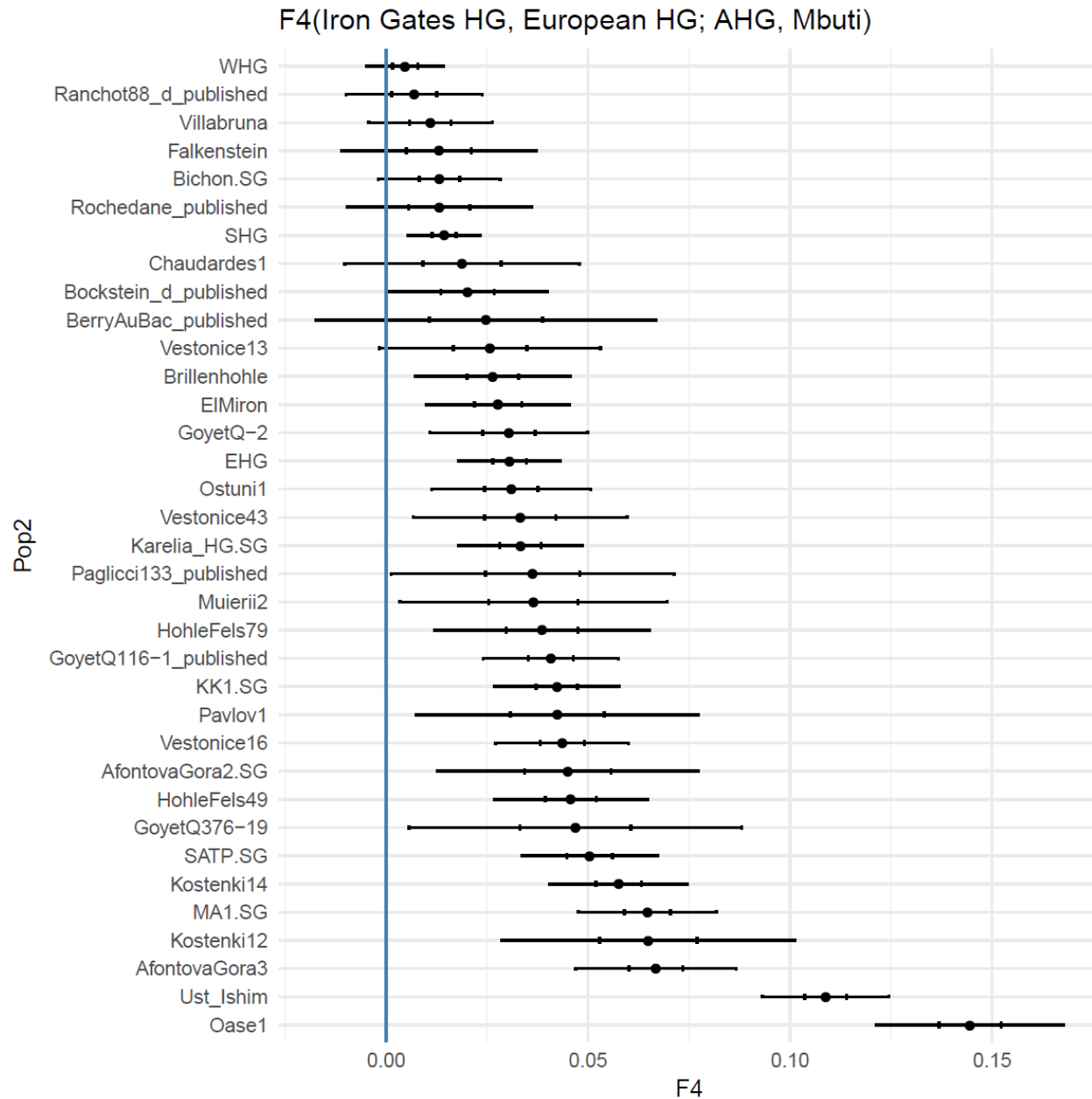
Supplementary Figure 3. Permutation test of the D statistic of the form $D(AAF^*, AHG^*; test, Mbuti)$. We plot the populations with the 10 most positive values of $D(AAF, AHG; test, Mbuti)$ and the permutation values for the same populations with ± 1 and ± 3 standard errors estimated by 5 cM block jackknifing (represented by the inner and outer vertical bars respectively). In each test setting “ AAF^* ” either includes AHG and all AAF individuals except the individual marked with an asterisk or all AAF individuals (and not AHG) except the individual marked with “_woAHG”. The test with the original population labels is indicated in pink (analysis was restricted to individuals $> 30,000$ SNPs). All permuted settings are consistent with a negative D-score or 0 within ± 1 standard errors while the original D-scores for these populations resulted in a positive D-score within ± 1 standard errors.



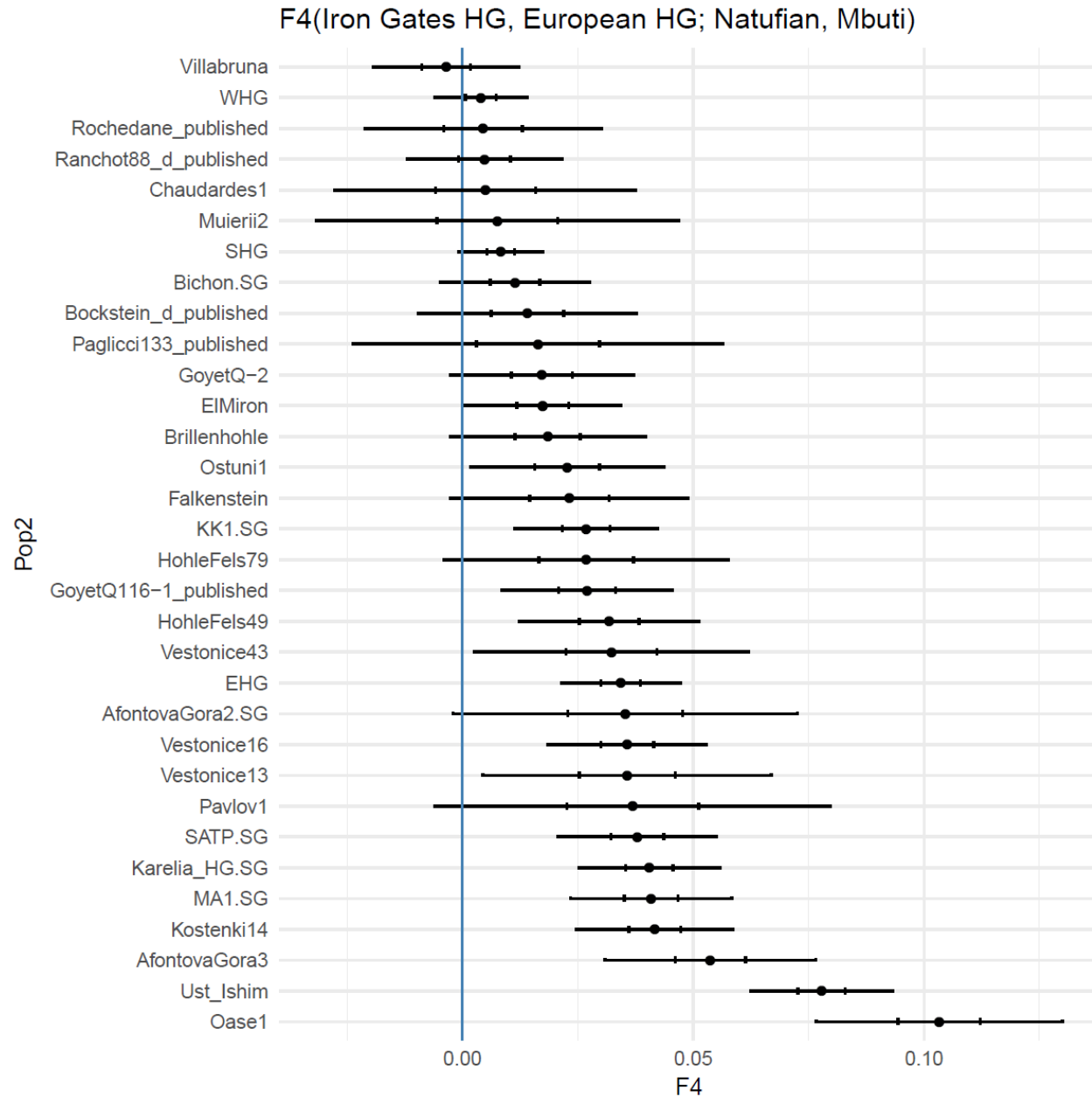
Supplementary Figure 4. ACF have excess allele sharing with Levantine related populations compared with AAF. We plot the populations with the 40 most positive (blue) and 40 most negative values (orange) of $D(ACF, AAF; test, Mbuti)$ with ± 1 standard errors estimated by 5 cM block jackknifing (represented by the horizontal bars). Positive values indicate that “test” shares more alleles with *ACF* than with *AAF*, and negative values that it shares more with *AAF* than *ACF* (analysis was restricted to individuals > 30,000 SNPs).



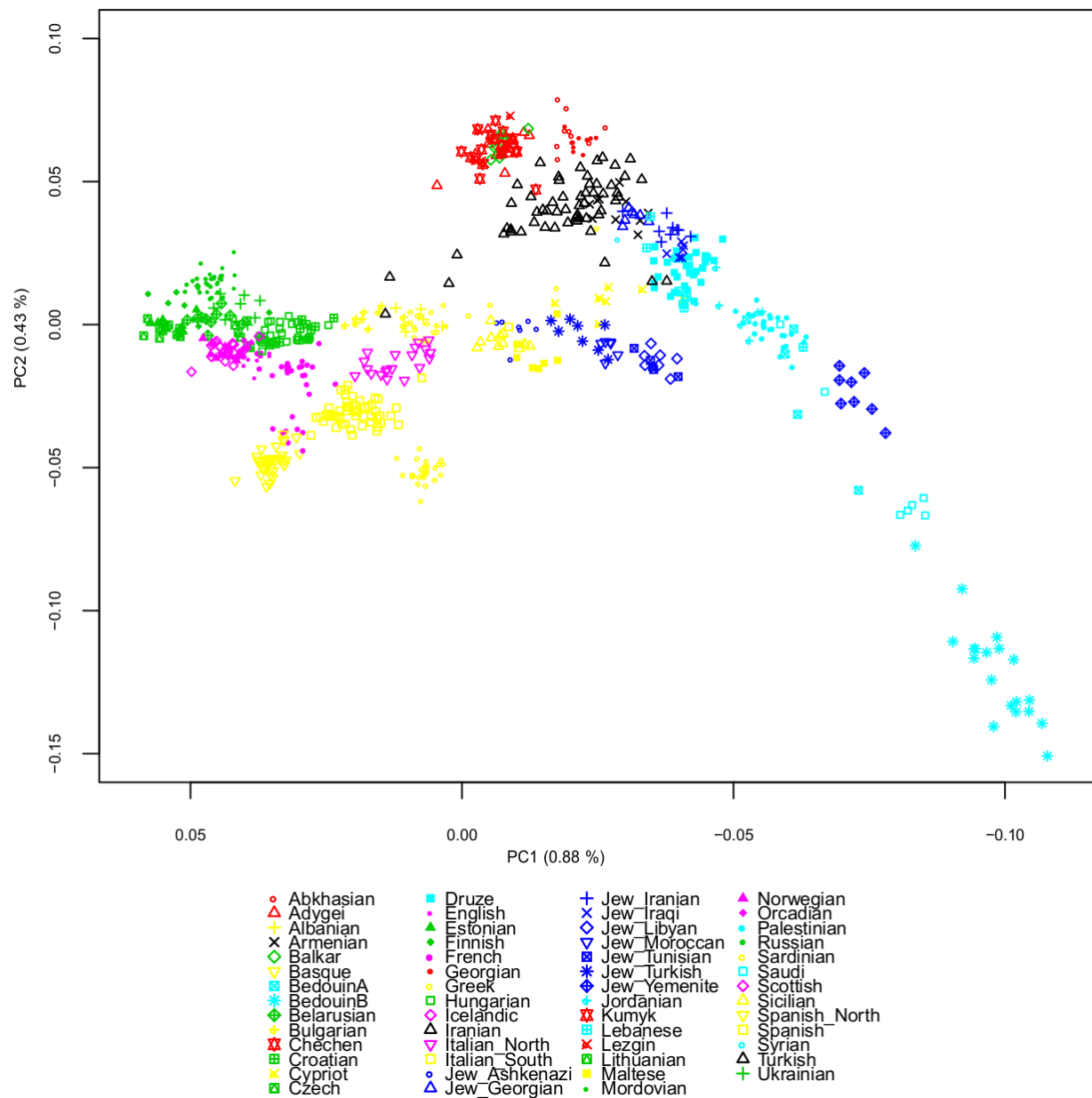
Supplementary Figure 5. Permutation test of the D -statistic of the form $D(ACF^*, AAF^*; test, Mbuti)$. We plot the empirical null distribution of D -statistics based on 1000 permutation tests performed for each of the four “test” populations that had the most positive values in the original observed statistic (*Levant_N*, *Natufian*, *Greece_EN*, *Balkans_Neolithic*). In each test individuals were randomly shuffled between the ACF^* and AAF^* groups. The D -statistic with the original population labels is marked with an asterisk. Empirical P -values are indicated for each plot and were calculated by dividing the number of permutations that resulted in a value equal to or greater than the original observation by the total number of permutations. In all cases, the observed value is at the top $< 1\%$ tail.



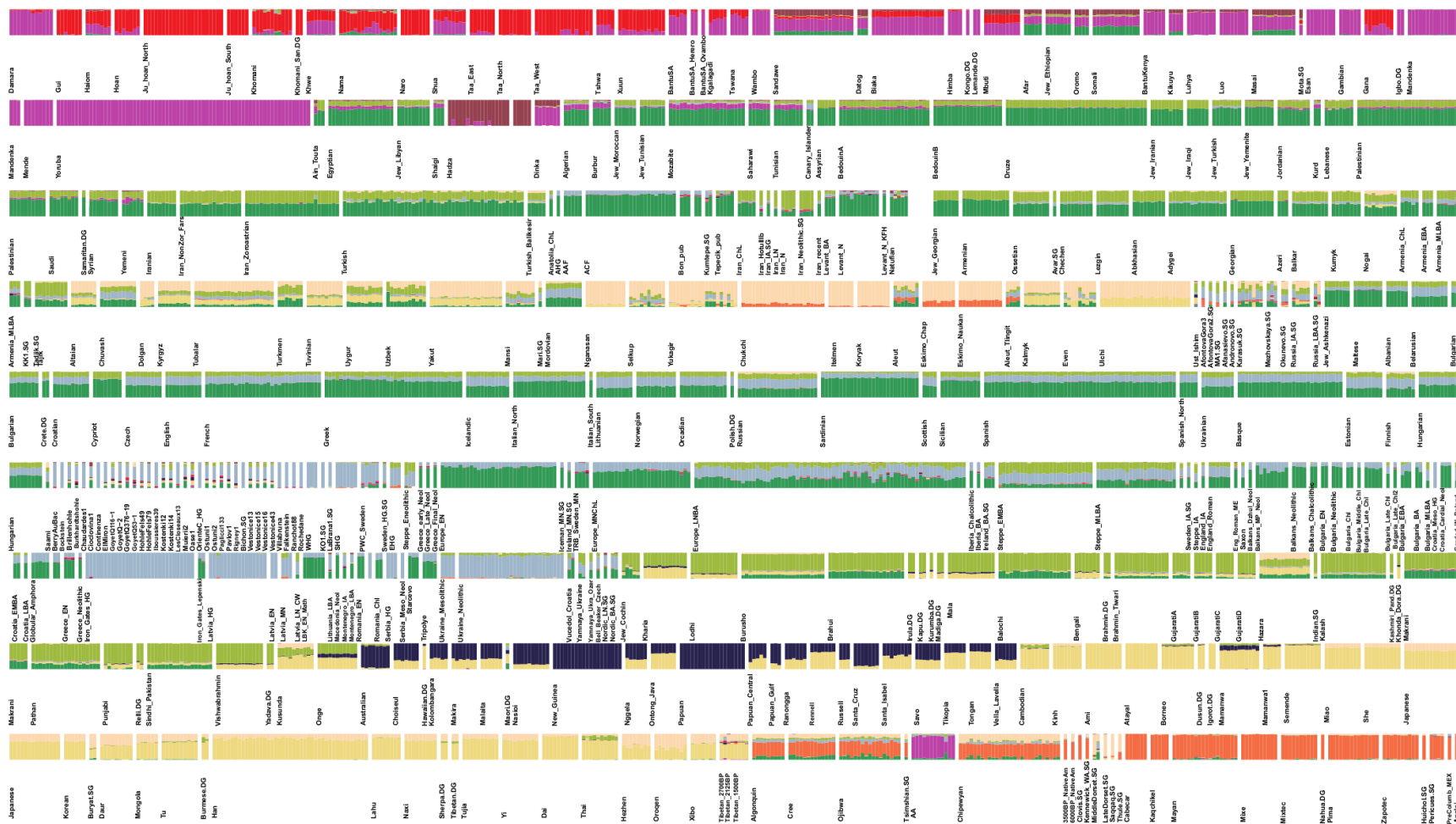
Supplementary Figure 6. Iron Gates HG show higher genetic affinity to AHG than all the other European HG. We plot the f_4 values of $D(\text{Iron Gates HG}, \text{European HG}; \text{AHG}, \text{Mbuti})$ with ± 1 and ± 3 standard errors estimated by 5 cM block jackknifing (represented by the inner and outer horizontal bars respectively). Positive values indicate that “AHG” shares more alleles with *Iron Gates HG* than with the tested *European HG* (analysis was restricted to individuals > 30,000 SNPs).



Supplementary Figure 7. Iron Gates HG show a higher genetic affinity to Natufians than all the other European HG but Villabruna. We plot the f_4 values of $D(\text{Iron Gates HG}, \text{European HG}; \text{Natufian}, \text{Mbuti})$ with ± 1 and ± 3 standard errors estimated by 5 cM block jackknifing (represented by the inner and outer horizontal bars respectively). Positive values indicate that “Natufian” shares more alleles with *Iron Gates HG* than with the tested *European HG* (analysis was restricted to individuals > 30,000 SNPs).



Supplementary Figure 8. PCA plot of present-day west Eurasian populations. The two first principal components computed for 67 published present day west Eurasian populations are shown. An estimated 0.88 % of the variation is explained by the first principal component (PC1) and 0.43 % is explained by the second principal component (PC2). Source data are provided as a Source Data file.



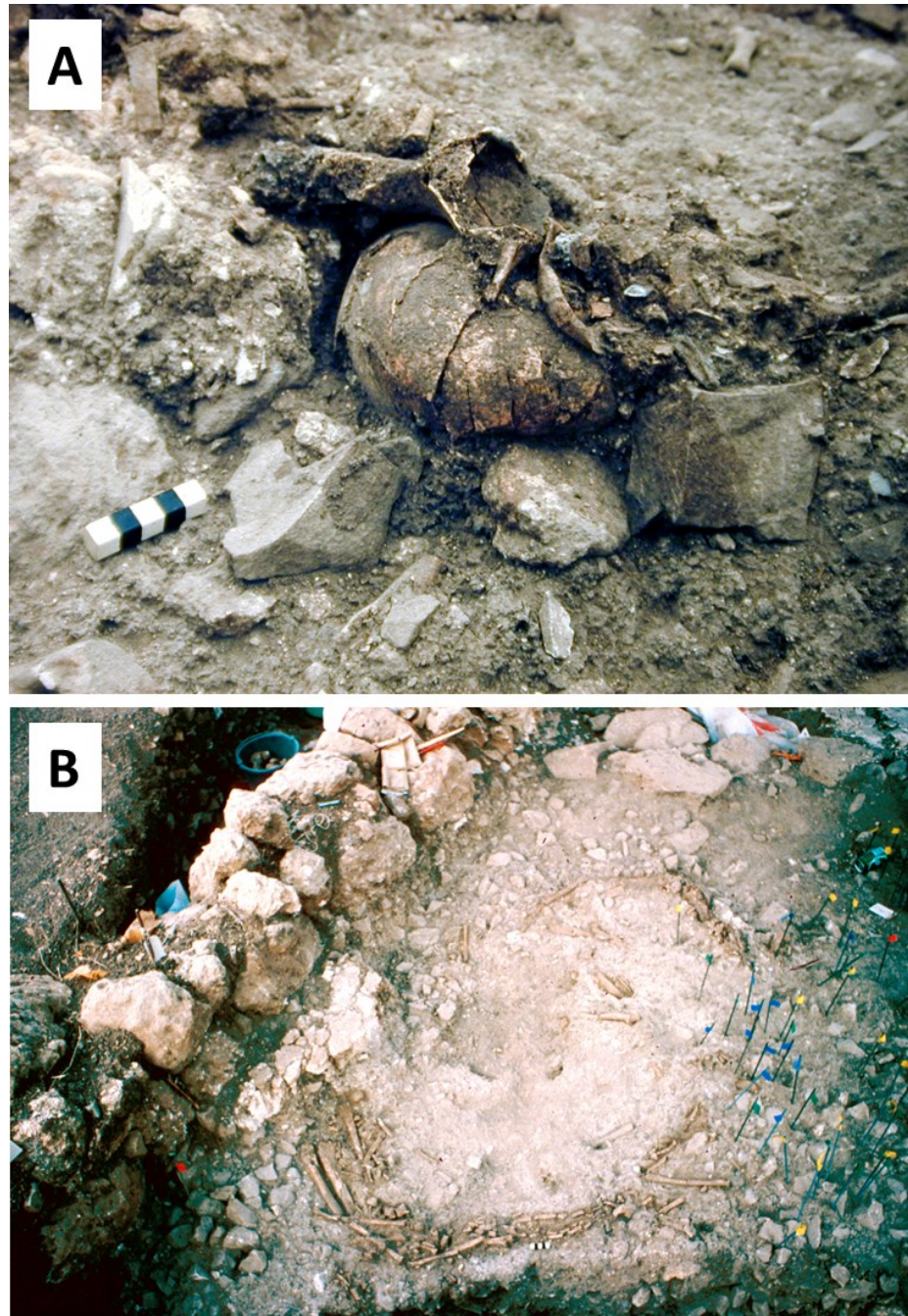
Supplementary Figure 9. ADMIXTURE plot of all individuals analyzed. ADMIXTURE analysis (Methods) computed from 2,706 present-day and 594 ancient individuals is shown (K = 10). Source data are provided as a Source Data file.



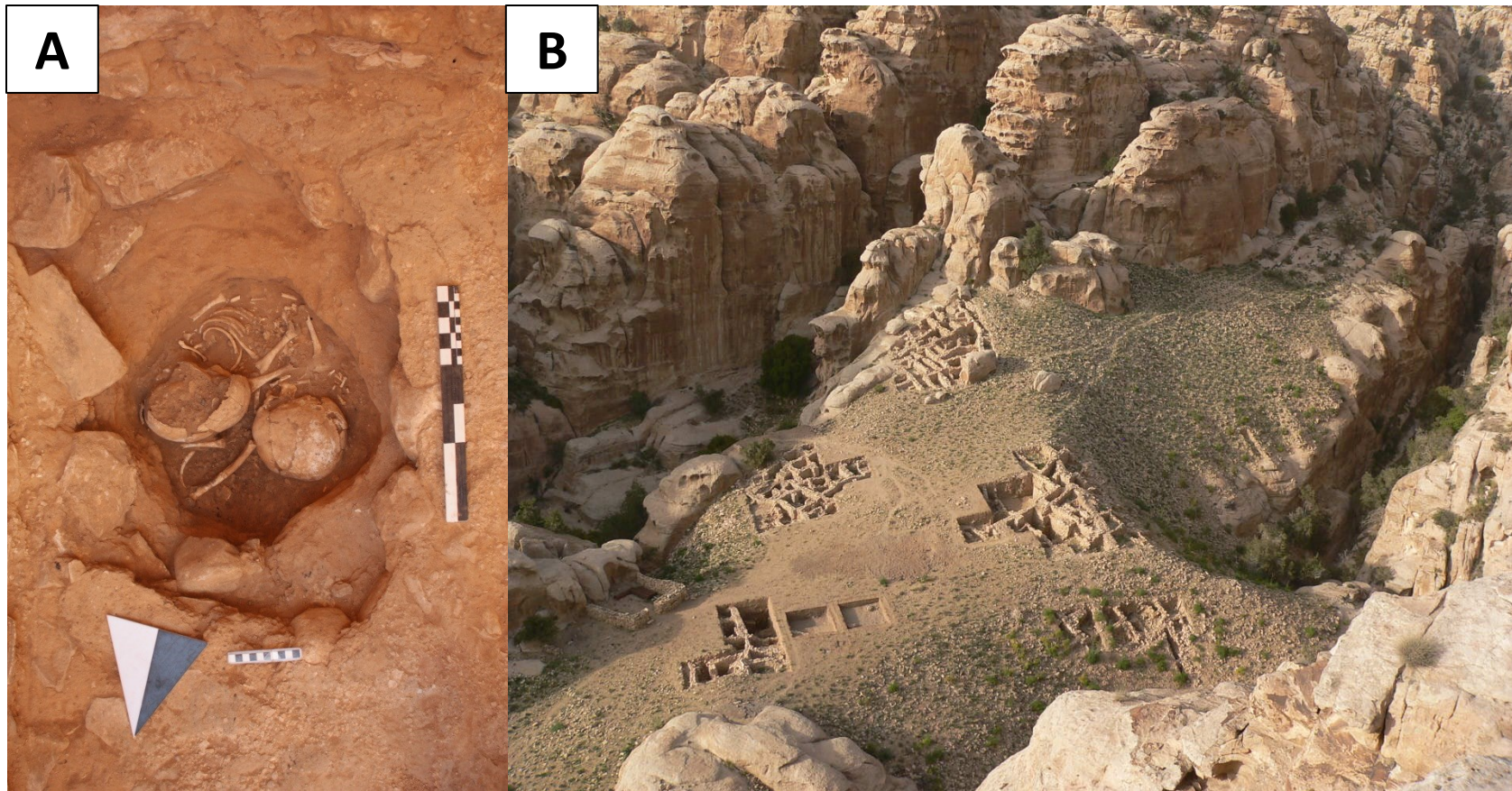
Supplementary Figure 10. View of the sampled Anatolian Epipaleolithic hunter-gatherer (AHG/ZBC). Excavated from grave 13 in Pınarbaşı. Photo by Douglas Baird.



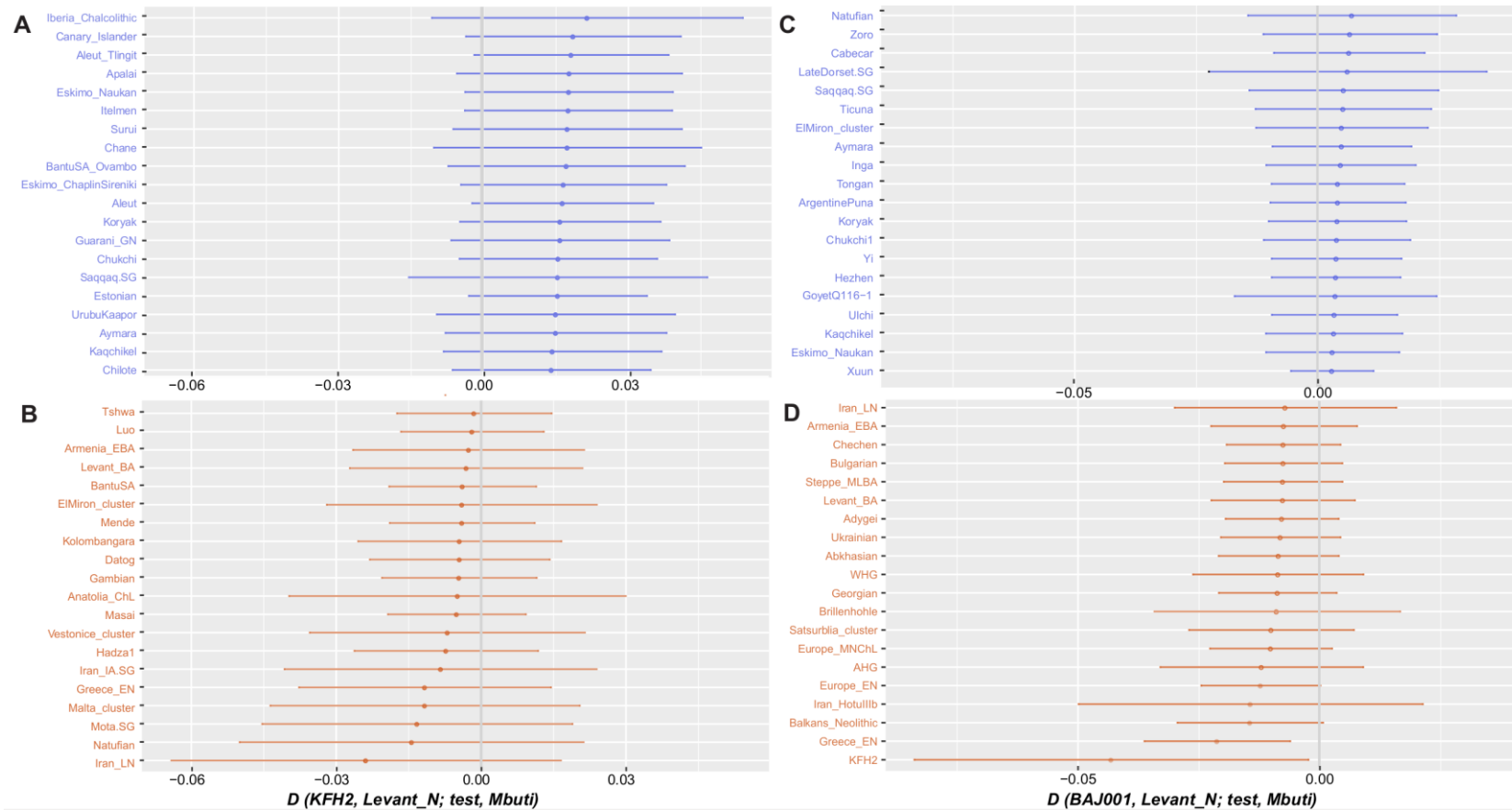
Supplementary Figure 11. View of sampled AAF individual ZHJ. Excavated from grave 15, Boncuklu¹¹. Photo by Douglas Baird/Boncuklu project.



Supplementary Figure 12. Kfar HaHoresh site and sampled early farmer KFH2 (A) View of sampled infant skull KFH2 *in situ* in L1003 of Kfar HaHoresh. Scale 5cm. **(B)** View of upper levels of multiple grave in L1003 (underlying L1001), showing intentional arrangement of human bones. Scale 5cm.



Supplementary Figure 13. The site of Ba'ja and the double infant burial. (A) View of the double infant burial (Loc. 405 of Room 35; individuals BAJ001 and BAJ002); Photo by Benz. (B) View on the Late Pre-Pottery Neolithic B site of Ba'ja (southern Jordan) from the top of the mountains. Ba'ja Neolithic Project, Photo by Borowski.



Supplementary Figure 14. Symmetry testing of the newly reported KFH2 and BAJ001 individuals with previously reported Levantine Neolithic individuals (grouped and labeled Levant_N). We plot the 20 most positive (blue) and the 20 most negative values (orange) of the D -statistic of the form $D(KFH2/BAJ001, Levant_N; test, mbuti)$ with ± 3 standard errors (represented by the horizontal lines) estimated by 5 cM block jackknifing. “test” populations include versatile global ancient and modern populations. Positive values indicate that “test” shares more alleles with $KFH2/BAJ001$ than with $Levant_N$, and negative values that it shares more with $Levant_N$ than $KFH2/BAJ001$. Results for all tested quadruples can be found in Supplementary Data 4 (analysis was restricted to individuals $> 30,000$ SNPs). (A) The 20 most positive values of the D statistic of the form $D(KFH2, Levant_N; test, mbuti)$ are plotted. (B) The 20 most negative values of the D statistic of the form $D(KFH2, Levant_N; test, mbuti)$ are plotted. (C) The 20 most positive values of the D statistic of the form $D(BAJ001, Levant_N; test, mbuti)$ are plotted. (D) The 20 most negative values of the D statistic of the form $D(BAJ001, Levant_N; test, mbuti)$ are plotted.

Supplementary Tables

Supplementary Table 1. ^{14}C radiocarbon dating performed for this study. Carbon dated at the CEZ Archaeometry gGmbH, Mannheim, Germany. The ^{14}C ages are given in BP (before present; meaning years before 1950). The calibrated dates are shown in columns “Cal 1-sigma” and “Cal 2-sigma” using the 1-sigma and 2-sigma uncertainty of the ^{14}C ages, respectively. The $\delta^{13}\text{C}$ value was obtained from the isotope determination in the AMS system with a typical uncertainty of 2%. This value may be influenced by isotope fractionation in the ion source and during graphitization and is only used for fractionation correction. Hence, this value is not comparable to the one obtained in a stable isotope IRMS and should not be used for further data interpretation.

Individual	Experiment number (MAMS)	C14 date (BP)	SE	Sigma 13C AMS [‰]	Cal 1-sigma (BCE)	Cal 2-sigma (BCE)	C:N	C (%)	Collagen (%)	Tissue
KFH2	31616	8,638	24	-9.1	7,647-7,594	7,712-7,589	3.3	40.2	1.5	Petrous bone
ZBC	30693	12,890	40	-16.7	13,530-13,335	13,647-13,284	3.3	37.7	0.8	Phalanx bone

Supplementary Table 2. Nuclear and mitochondrial contamination estimates. For each individual newly reported and analyzed in this study mitochondrial contamination estimates calculated with schmutzi (Materials and Methods) are given. For genetic males the nuclear contamination estimate is provided (Materials and Methods). The levels of DNA damage are given as the deamination level at the 5' terminal position of the mapped reads.

Individual ID	Site	Nuclear contamination estimate	SE	Mitochondrial contamination estimate	Deamination at 5' terminal position (%)
ZBC	Pınarbaşı	0.005	0.002	0.01 (0.00 -0.02)	11
KFH2	Kfar HaHoresh	NA	NA	0.06 (0.04-0.08)	26
BAJ001	Ba'ja	NA	NA	0.01(0.00-0.02)	28
ZHAG	Boncuklu	NA	NA	0.01 (0.00-0.02)	11
ZMOJ	Boncuklu	0.009	0.006	0.03 (0.02-0.04)	22
ZKO	Boncuklu	0.022	0.007	0.01 (0.00-0.02)	16
ZHJ	Boncuklu	NA	NA	0.01 (0.00-0.02)	9
ZHAJ	Boncuklu	NA	NA	0.03 (0.02-0.04)	16

Supplementary Table 3. Test of cladeness between AHG and late Pleistocene/early Holocene populations. The D -statistics of the form $D(AHG, pop1; pop2, Mbuti)$ is shown where “ $pop1$ ” and “ $pop2$ ” are late Pleistocene/early Holocene groups from Europe or the Near East. Positive D -values ($Z > 3$) indicate that “ $pop2$ ” shares more alleles with AHG compared to “ $pop1$ ” and negative D -values ($Z < -3$) indicate that “ $pop2$ ” shares more alleles with “ $pop1$ ” compared to AHG. The Z scores were calculated from a 5 cM block jackknifing standard error. The number of SNP positions covered in all four tested populations is given in column ‘nSNPs’.

Pop1	Pop2	D	Z	nSNPs
WHG	Levant_N	0.0458	9.953	384,901
WHG	Natufian	0.0285	4.822	236,731
WHG	Iran_N	0.0087	1.613	390,457
WHG	EHG	-0.0693	-14.094	453,196
Levant_N	WHG	0.0439	9.038	384,901
Levant_N	Natufian	-0.0365	-5.56	215,370
Levant_N	Iran_N	0.0103	1.779	341,358
Levant_N	EHG	0.0448	8.965	374,568
Natufian	WHG	0.0545	9.585	236,731
Natufian	Levant_N	-0.0084	-1.364	215,370
Natufian	Iran_N	0.0331	4.766	214,816
Natufian	EHG	0.057	8.804	230,917
Iran_N	WHG	0.075	13.793	390,457
Iran_N	Levant_N	0.0784	14.371	341,358
Iran_N	Natufian	0.0725	10.499	214,816
Iran_N	EHG	0.0394	6.598	379,907
EHG	WHG	-0.043	-8.546	453,196
EHG	Levant_N	0.0691	14.148	374,568
EHG	Natufian	0.0568	8.846	230,917
EHG	Iran_N	-0.0041	-0.692	379,907

5

Supplementary Table 4. Summary of best fitting qpAdm admixture models of key ancient populations. For each target population the proportions estimated by the best fitting admixture models ($P_{\text{val}} > 0.05$) are given with their standard errors estimated by 5 cM block jackknifing. ‘Ref’ 1-3 indicate the ancestral sources (reference) used to model the target populations. The abbreviations of the population names are listed in Supplementary Data 2. We define the ‘Basic set’ of outgroups as: *Han*; *Onge*; *Mbuti*; *Natufian*; *Kostenki14*; *Mala*; *Mixe*. When *Natufian* was used as a source population, the present-day *BedouinB* Near-Eastern population was used as an outgroup in the ‘Basic set’ instead.

Target	Ancestral sources			Mixture proportions (%)			Standard errors (%)			Pval (rank - 1)	Outgroups
	Ref1	Ref2	Ref3	Ref1	Ref2	Ref3	Ref1	Ref2	Ref3		
AHG	Levant_N	WHG		47.9	52.1		4.5	4.5		0.158	Basic set
AHG	Levant_N	WHG	Iran_N	46.4	41.7	11.9	4.6	7.1	6.9	0.296	Basic set
AAF	AHG	Iran_N		89.7	10.3		3.9	3.9		0.296	Basic set; WHG; EHG
ACF	AAF	Levant_N		78.7	21.3		3.5	3.5		0.606	Basic set
Iron_Gates_HG	AHG	WHG	EHG	25.8	62.9	11.3	5.0	7.4	3.3	0.308	Basic set
Iron_Gates_HG	Natufian	WHG	EHG	11.1	78.0	10.9	2.2	4.6	3.0	0.589	Basic set
Levant_Neol	AAF	Natufian		17.8	82.2		6.4	6.4		0.288	Basic set; WHG; EHG
Levant_Neol	AHG	Natufian		16.3	83.7		6.7	6.7		0.074	Basic set; WHG; EHG

5

Supplementary Table 5. qpADM Admixture models of the Anatolian hunter-gatherer (AHG). The proportions estimated for each ancestral source (Ref1-4) used to model the target population AHG are given. Fitting models (Pval > 0.05; and admixture proportions are feasible) are highlighted in green. When resolution is lacking to determine whether a mixture proportion is required the source population, proportion and Standard error are highlighted in lighter green. The fitting models with minimal waves of ancestry are marked in bold. The standard errors were estimated by 5 cM block jackknifing. The abbreviations of the population names are listed in Supplementary Data 2. The ‘Basic set’ of outgroup populations (*Han*; *Onge*; *Mbuti*; *Natufian*; *Kostenki14*; *Mala*; *Mixe*) was used.

Ancestral sources				Mixture proportions (%)				Standard errors (%)				Pval (rank-1)
Ref1	Ref2	Ref3	Ref4	Ref1	Ref2	Ref3	Ref4	Ref1	Ref2	Ref3	Ref4	
Levant_N	WHG			47.9	52.1			4.5	4.5			1.58E-01
Levant_N	EHG			68.7	31.3			3.1	3.1			1.15E-05
Levant_N	Iran_N			55.7	44.3			5.1	5.1			3.31E-07
Iran_N	WHG			27.1	72.9			15.9	15.9			4.59E-21
Iran_N	EHG			110.0	-10.0			8.7	8.7			2.27E-29
Levant_N	WHG	EHG		48.2	49.5	2.4		5.2	10.4	6.5		1.31E-01
Levant_N	WHG	Iran_N		46.4	41.7	11.9		4.6	7.1	6.9		2.96E-01
Levant_N	EHG	Iran_N		59.0	21.3	19.7		4.8	4.7	7.4		1.16E-03
Iran_N	WHG	EHG		28.0	110.7	-38.7		11.4	14.6	7.5		8.95E-12
Levant_N	WHG	Iran_N	EHG	46.7	39.7	11.5	2.2	5.4	10.7	7.0	6.6	2.35E-01

Supplementary Table 6. qpADM Admixture models of the Anatolian Aceramic farmers (AAF). The proportions estimated for each ancestral source (Ref1-3) used to model the target population AAF are given. Fitting models (Pval > 0.05; and admixture proportions are feasible) are highlighted in green and in bold letters. The standard errors were estimated by 5 cM block jackknifing. The abbreviations of the population names are listed in Supplementary Data 2. We define the ‘Basic set’ of outgroup populations as *Han*; *Onge*; *Mbuti*; *Natufian*; *Kostenki14*; *Mala*; *Mixe*.

Ancestral sources			Mixture proportions (%)			Standard errors (%)			Pval (rank – 1)	Outgroups
Ref1	Ref2	Ref3	Ref1	Ref2	Ref3	Ref1	Ref2	Ref3		
AHG	Iran_N		91.4	8.6		7.2	7.2		5.38E-02	Basic set
AHG	Iran_N		89.7	10.3		3.9	3.9		2.96E-01	Basic set; Levant_N;WHG;EHG
AHG	EHG		98.7	1.3		4.0	4.0		1.37E-02	Basic set
AHG	Levant_N		108.1	-8.1		9.8	9.8		5.76E-02	Basic set
AHG	WHG		103.7	-3.7		9.9	9.9		1.07E-02	Basic set
WHG	EHG		144.4	-44.4		8.5	8.5		5.71E-13	Basic set
AHG	Levant_N	Iran_N	93.8	-2.6	8.8	19.2	11.6	9.5	9.62E-02	Basic set

Supplementary Table 7. qpADM Admixture models of the Anatolian Ceramic farmers (ACF). The proportions estimated for each ancestral source (Ref1-3) used to model the target population ACF are given. Fitting models (Pval > 0.05; and admixture proportions are feasible) are highlighted in green. When resolution is lacking to determine whether a mixture proportion is required the source population, proportion and Standard error are highlighted in lighter green. The fitting models with minimal waves of ancestry are marked in bold. The standard errors were estimated by 5 cM block jackknifing. The abbreviations of the population names are listed in Supplementary Data 2. We define the ‘Basic set’ of outgroup populations as *Han*; *Onge*; *Mbuti*; *Natufian*; *Kostenki14*; *Mala*; *Mixe*.

Ancestral sources			Mixture proportions (%)			Standard errors (%)			Pval (rank – 1)	Outgroups
Ref1	Ref2	Ref3	Ref1	Ref2	Ref3	Ref1	Ref2	Ref3		
AAF	Levant_N		78.7	21.3		3.5	3.5		6.06E-01	Basic set
AAF	EHG		110.3	-10.3		2.4	2.4		2.69E-02	Basic set
AAF	Iran_N		117.2	-17.2		4.6	4.6		4.21E-02	Basic set
AAF	WHG		122.9	-22.9		5.8	5.8		5.49E-02	Basic set
AAF	Levant_N	Iran_N	86.9	17.8	-4.7	9.2	5.3	4.8	5.66E-01	Basic set
AHG	Levant_N		84.3	15.7		6.8	6.8		1.15E-01	Basic set
AHG	EHG		108.0	-8.0		3.9	3.9		7.49E-02	Basic set
AHG	Iran_N		106.5	-6.5		7.0	7.0		7.65E-02	Basic set
AHG	WHG		124.1	-24.1		9.8	9.8		3.98E-01	Basic set
AHG	Levant_N	Iran_N	77.9	18.0	4.2	13.3	8.2	6.5	2.16E-01	Basic set
AHG	Levant_N	Iran_N	71.0	22.3	6.7	5.6	4.7	3.2	3.50E-01	Basic set; WHG; EHG

Supplementary Table 8. qpAdm Admixture models of Levantine early farmers. The proportions estimated for each ancestral source (Ref1-2) used to model each target population are given with their standard errors estimated by 5 cM block jackknifing. Fitting models (Pval > 0.05; and admixture proportions are feasible) are highlighted in green. When resolution is lacking to determine whether a mixture proportion is required the source population, proportion and Standard error are highlighted in lighter green. The fitting models with minimal waves of ancestry are marked in bold. The abbreviations of the population names are listed in Supplementary Data 2. We define the ‘Basic set’ of outgroups as: Han; Onge; Mbuti; Natufian; Kostenki14; Mala; Mixe. When Natufian was used as a source population, the present-day BedouinB Near-Eastern population was used as an outgroup in the ‘Basic set’ instead. The previously published individuals from Motza, Israel and ‘Ain-Ghazal, Jordan⁵⁷ are grouped together and labeled as ‘Levant_N’. When the published (Levant_N) and newly reported Levantine early farmers (BAJ001, KFH2) are grouped into one population they are labeled ‘Levant_Neol’.

Target	Ancestral sources		Mixture proportions (%)		Standard errors (%)		Pval (rank – 1)	Outgroups
	Ref1	Ref2	Ref1	Ref2	Ref1	Ref2		
KFH2	ACF	Levant_N	-5.4	105.4	25.6	25.6	0.377	Basic set; EHG; WHG
KFH2	AAF	Levant_N	-2.6	102.6	16.5	16.5	0.313	Basic set; EHG; WHG
KFH2	AHG	Levant_N	-2.2	102.2	16.5	16.5	0.322	Basic set; EHG; WHG
KFH2	WHG	Levant_N	6.1	93.9	6.3	6.3	0.435	Basic set
KFH2	ACF	Natufian	24.9	75.1	16.6	16.6	0.525	Basic set; EHG; WHG
KFH2	AAF	Natufian	23.3	76.7	13.5	13.5	0.622	Basic set; EHG; WHG
Levant_N	ACF	BAJ001	9.8	90.2	14.5	14.5	0.625	Basic set
Levant_N	AAF	BAJ001	3.4	96.6	10	10	0.510	Basic set
Levant_N	AHG	BAJ001	-4.8	95.2	15.3	15.3	0.580	Basic set
Levant_N	WHG	BAJ001	-2.4	102.4	4.5	4.5	0.607	Basic set
Levant_N	Greece_EN	BAJ001	7.6	92.4	15.4	15.4	0.496	Basic set
Levant_N	ACF	Natufian	30.4	69.6	7.2	7.2	0.575	Basic set; EHG; WHG
Levant_N	AAF	Natufian	21.3	78.7	6.3	6.3	0.375	Basic set; EHG; WHG
BAJ001	ACF	Natufian	15.8	84.2	12.5	12.5	0.128	Basic set; EHG; WHG
BAJ001	AAF	Natufian	8.3	91.7	10.3	10.3	0.098	Basic set; EHG; WHG
Levant_Neol	AAF	Natufian	17.8	82.2	6.4	6.4	0.288	Basic set; EHG; WHG
Levant_Neol	AHG	Natufian	16.3	83.7	6.7	6.7	0.074	Basic set; EHG; WHG

Supplementary Table 9. Admixture models of the Iron Gates hunter-gatherers (Iron Gates HG). The proportions estimated for each ancestral source (Ref1-3) used to model Iron Gates HG population are given with their standard errors estimated by 5 cM block jackknifing. Fitting models ($P_{\text{val}} > 0.05$; and admixture proportions are feasible) are highlighted in green. When resolution is lacking to determine whether a mixture proportion is required the source population, proportion and Standard error are highlighted in lighter green. The fitting models with minimal waves of ancestry are marked in bold. The abbreviations of the population names are listed in Supplementary Data 2. We define the ‘Basic set’ of outgroups as: *Han*; *Onge*; *Mbuti*; *Natufian*; *Kostenki14*; *Mala*; *Mixe*. When *Natufian* was used as a source population, the present-day *BedouinB* Near-Eastern population was used as an outgroup in the ‘Basic set’ instead.

Ancestral sources			Mixture proportions (%)			Standard errors (%)			Pval (rank-1)	Outgroups
Ref1	Ref2	Ref3	Ref1	Ref2	Ref3	Ref1	Ref2	Ref3		
AHG	WHG	EHG	25.8	62.9	11.3	5.0	7.4	3.3	3.08E-01	Basic set
AAF	WHG	EHG	21.8	69.0	9.2	4.1	5.8	2.8	5.41E-01	Basic set
Iran_N	WHG	EHG	6.2	90.3	3.5	2.9	4.4	3.0	1.10E-02	Basic set
Natufian	WHG	EHG	11.1	78.0	10.9	2.2	4.6	3.0	5.89E-01	Basic set

Supplementary Table 10. Basal Eurasian proportion estimates. The proportions estimated for each ancestral source (Ref1-3) used to model the target population are given with their standard errors estimated by 5 cM block jackknifing. The mixture proportion of ‘Mota.SG’ listed in bold IS used as a proxy for the proportion of Basal Eurasian ancestry (α) in the target population⁵⁷. Fitting models (Pval > 0.05; and admixture proportions are feasible) are highlighted in green. The abbreviations of the population names are listed in Supplementary Data 2. The set of outgroups used for this analysis includes: *Han*; *Onge*; *Mbuti*; *Natufian*; *Kostenki14*; *Mala*; *Mixe*. For models where Mota.SG was added to this set, the target is marked with an asterisk.

Test	Ancestral sources			Mixture proportions (%)			Standard errors (%)			Pval (rank-1)
	Ref1	Ref2	Ref3	Ref1	Ref2	Ref3	Ref1	Ref2	Ref3	
AHG	Mota.SG	WHG		24.8	75.2		5.5	5.5		2.72E-02
Natufian	Mota.SG	WHG		38.5	61.5		5.0	5.0		2.36E-01
Natufian	Mota.SG	EHG		66.2	33.8		3.7	3.7		8.03E-07
Natufian	Mota.SG	WHG	EHG	36.2	69.0	-5.2	6.2	12.4	7.7	1.16E-01
Iron_Gates_HG	Mota.SG	WHG	EHG	1.6	85.9	12.5	2.8	5.3	3.3	6.80E-02
Iron_Gates_HG	WHG	EHG		88.4	11.6		2.9	2.9		1.12E-01
Iron_Gates_HG*	WHG	EHG		88.6	11.4		2.9	2.9		1.57E-01

14.2 Supplementary materials of paper B

Supplementary Materials for

Ancient DNA sheds light on the genetic origins of early Iron Age Philistines

Michal Feldman, Daniel M. Master, Raffaella A. Bianco, Marta Burri, Philipp Stockhammer,
Alissa Mittnik, Adam J. Aja, Choongwon Jeong and Johannes Krause

Correspondence to: cwjeong@snu.ac.kr (CJ); krause@shh.mpg.de (JK)

This PDF file includes:

Supplementary Text S1
Figs. S1 to S15
Tables S1 to S10

Other Supplementary Materials for this manuscript include the following:

Data file S1: An overview of skeletal material screened for ancient DNA in this study
Data file S2: An overview of the main analysis groups used in this study
Data file S3: Sequencing statistics of negative controls

Supplementary Text

Supplementary Text S1. Description of individuals excavated in Ashkelon and archaeological information

Bronze Age

Ashkelon's Bronze Age necropolis included 16 chambers containing more than 200 individuals (18). The tombs contained numerous ceramic vessels from Bronze Age trade with the Aegean, Cyprus, Lebanon, and Egypt. Comparative ceramic studies provide a date from Middle Bronze IIA through the Late Bronze Age II (Figure 15.74-5 in Stager *et al.* (18) and (60, 61)).

Two samples were recovered from individuals immediately adjacent to one another in Chamber 14 (ASH029 and ASH033), and one sample was recovered from an individual in Chamber 13 (ASH034). Both chambers were part of a single tomb complex (fig. S12); Baker argues that this complex began to be used in MBIIC and was used through the Late Bronze Age (page 56 in Baker *et al.* (62)). Initial morphological analysis of the bones was conducted by Patricia Smith and Leslie Dawson.

Iron Age I

Four samples (ASH2-3, ASH066, ASH067 and ASH068) were taken from infant burials found under the dirt floors of Iron Age I houses excavated on the central mound in Ashkelon (Grid 38) (fig. S13). The Iron Age I houses were built upon the architectural debris of the Bronze Age, and rebuilt or replaced several times over the Iron Age I. The Iron Age I house sequence consists of several superimposed architectural complexes (Architectural Phase 20-17) each of which underwent minor changes (subphases a, b). On a floor of the earliest phase (Architectural Phase 20b), a scarab of Ramses III forms an early 12th century BCE *terminus post quem* for the entire sequence (19, 63).

Within these houses, at least 9 infants were buried beneath the floors of the buildings, typically in a simple pit or in a storage jar set into such a pit e.g. (64). Samples were recovered from four infants buried under the floors of the Phase 19a and Phase 18b buildings. These remains were identified by Marina Faerman, Netta Lev-Tov Shattah and Patricia Smith. Houses within both of these architectural phases (19-18) contained ceramics known as Philistine Bichrome pottery (18).

Iron Age IIA

Three samples were recovered from an Iron IIA cemetery located outside, but adjacent to, the city wall of ancient Ashkelon (20). More than two hundred individuals were so far identified in the cemetery. Analysis of this cemetery is ongoing under the direction of Sherry Fox, Rachel Kalisher and Kathryn Marklein. 72 samples were submitted for genetic analysis in this study, and three successfully yielded DNA.

Sample ASH008 /Burial 226 is part of succession of stratigraphically sequenced pit graves. An earlier pit grave (Burial 236) included a complete Black-on-Red Ware juglet. This ceramic form is generally accepted to be not earlier than the mid-tenth century (65) and forms a *terminus post quem* for the date of Burial 226 (fig. S14).

Sample ASH087 was taken from a complete articulation (Burial 260) which is the latest in a succession of pit graves. Earlier burials in this sequence are accompanied by Black-on-Red Ware juglets (fig. S15).

ASH135.A0101 comes from a disarticulated tooth uncovered as part of Burial 343 in square 13. The earliest material from this region of the cemetery is cremation Burial 345 potentially dated as early as the late eleventh century (pages 144, 146, Figure 16 in (20)).

Supplementary Figures

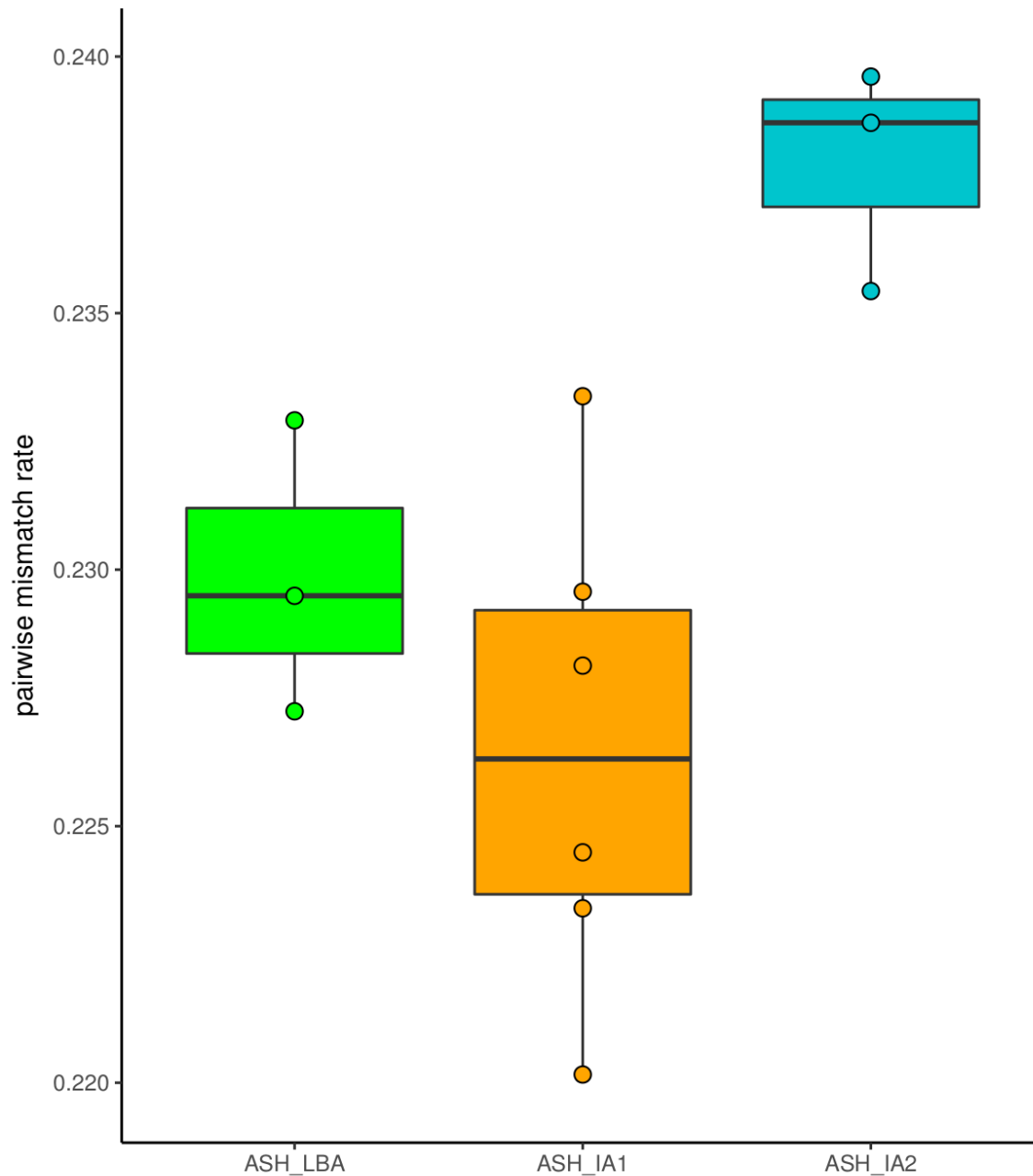


Fig. S1. Pairwise mismatch rate. Pairwise SNP mismatch rates (the proportion of mismatching SNPs out of the total number of overlapping SNPs) between the individuals of the three Ashkelon analyzed groups are marked with circles. Quartiles are indicated by the horizontal lines. None of the individuals are estimated to be first degree related.

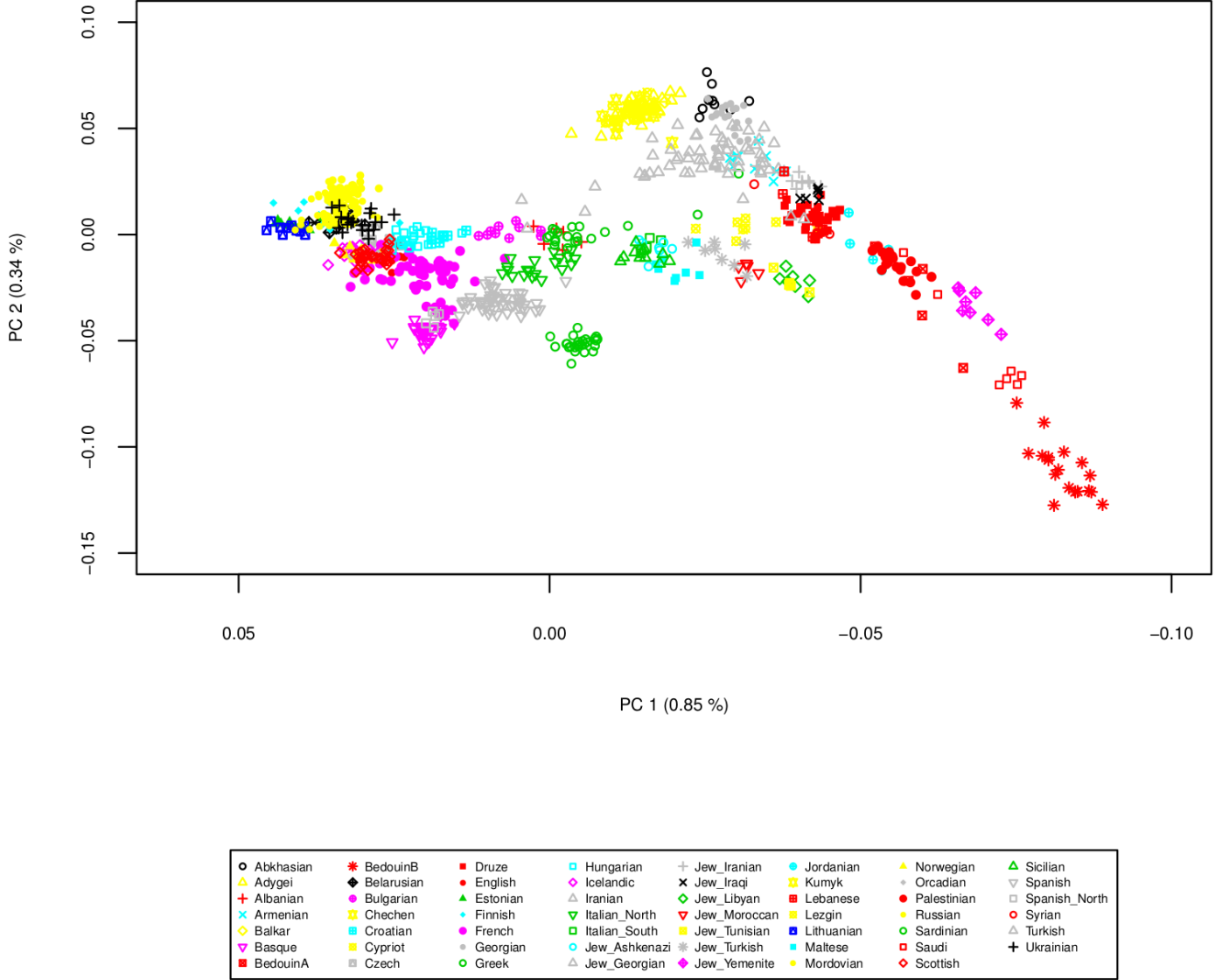


Fig. S2: PCA plot of present-day west Eurasian populations. The two highest principal components calculated for 55 published present-day west Eurasian populations are plotted. An estimated 0.85 % of the variation is explained by the first principal component (PC1) and 0.34 % is explained by the second (PC2).

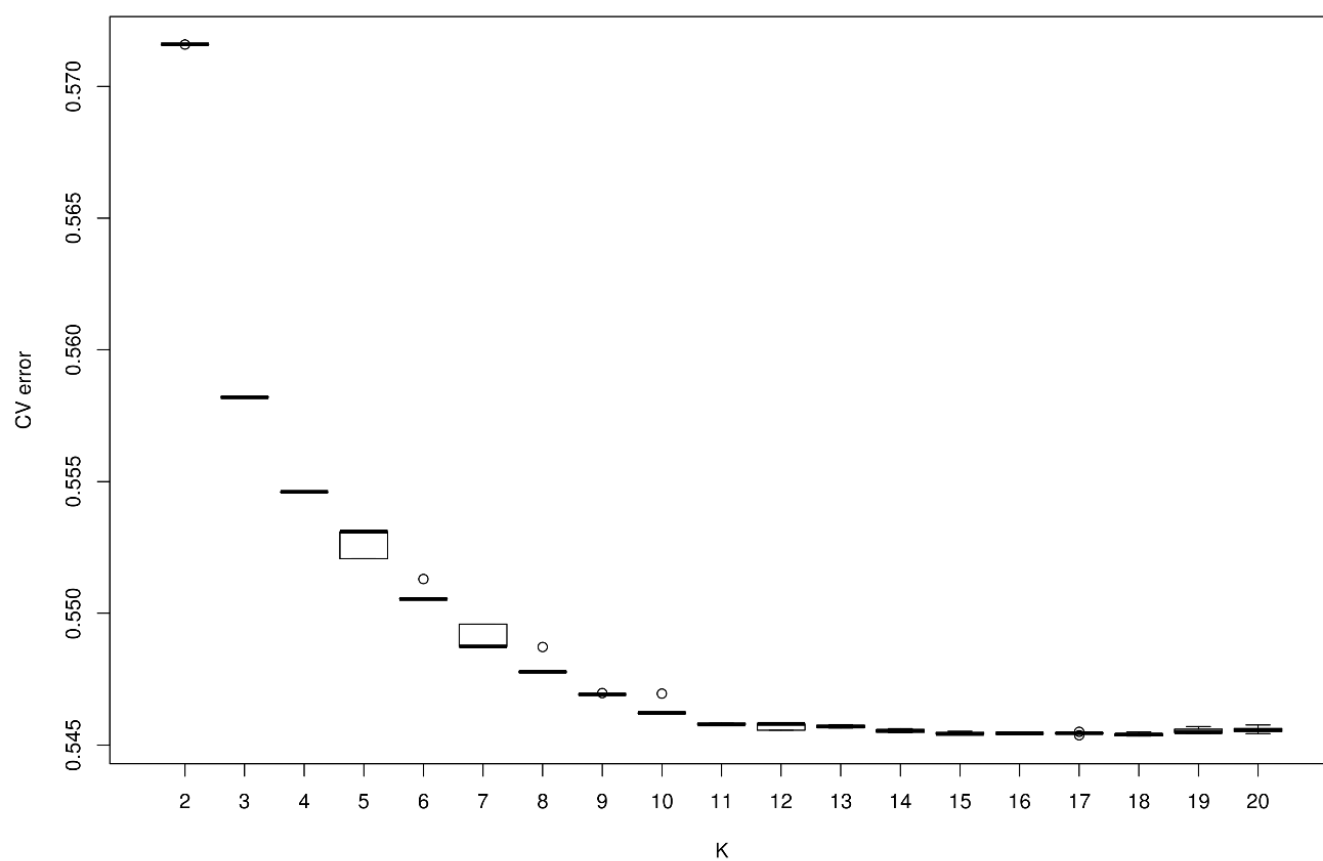


Fig. S3. ADMIXTURE cross-validation (CV) errors for each cluster number (K). Five replicates of the clustering using ADMIXTURE were performed for each K using random seeds. The horizontal lines indicate the quartiles and circles indicate outliers.

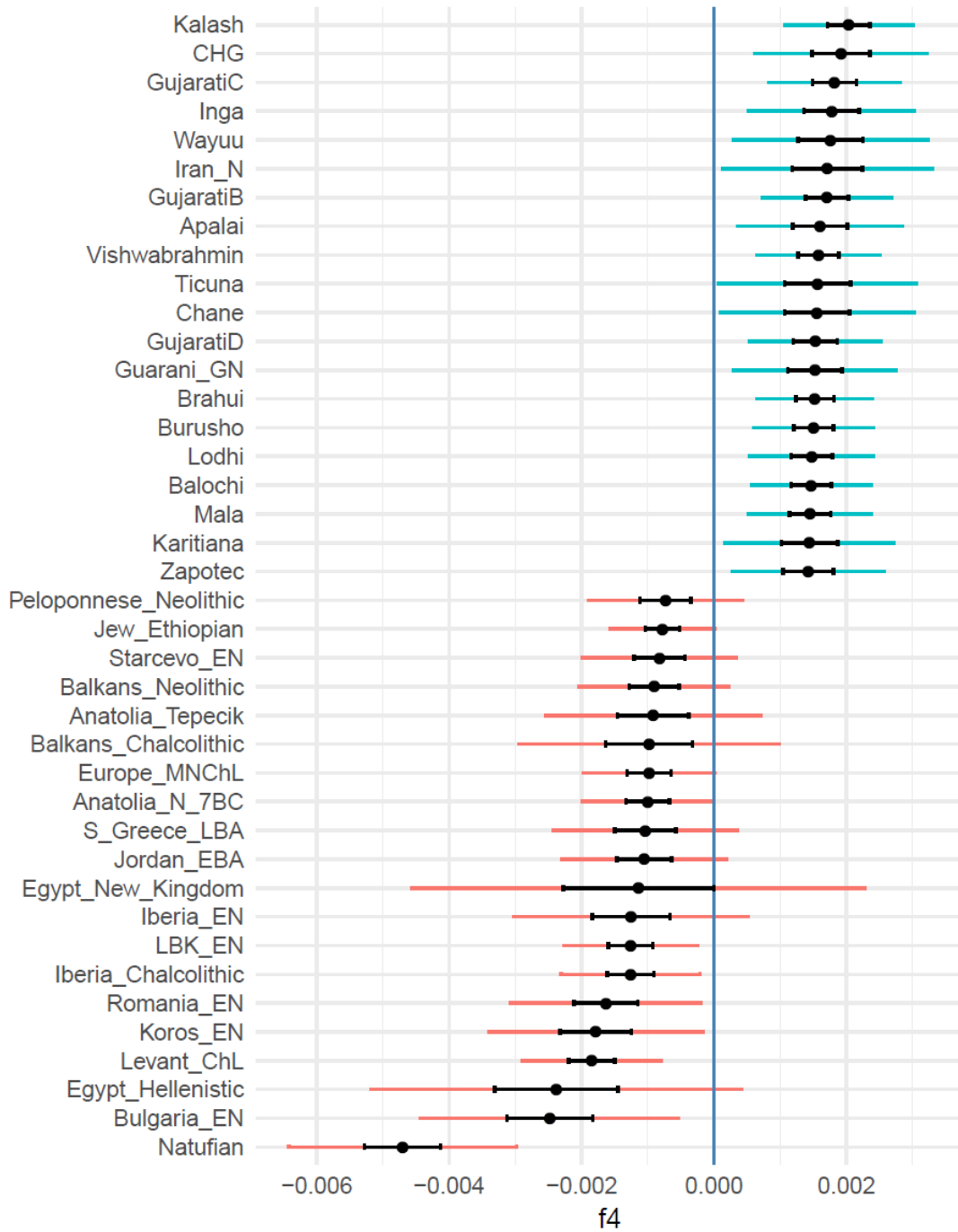


Fig. S4. The Bronze Age Ashkelon population shares a higher genetic affinity with Caucasus/Iranian related populations when compared to the Neolithic Levantines. We plot the 20 highest (blue) and 20 lowest (orange) values of the f_4 -statistic of the form $f_4(\text{ASH_LBA}, \text{Levant_N}; \text{test}, \text{Mbuti})$ with ± 3 standard errors (represented by the colored horizontal lines) and ± 1 standard errors (represented by the black horizontal lines) both estimated by 5 cM block jackknifing. “test” populations include versatile global ancient and modern populations. Positive values indicate that “test” shares more alleles with ASH_LBA than with Levant_N, and negative values that it shares more with Levant_N than ASH_LBA.

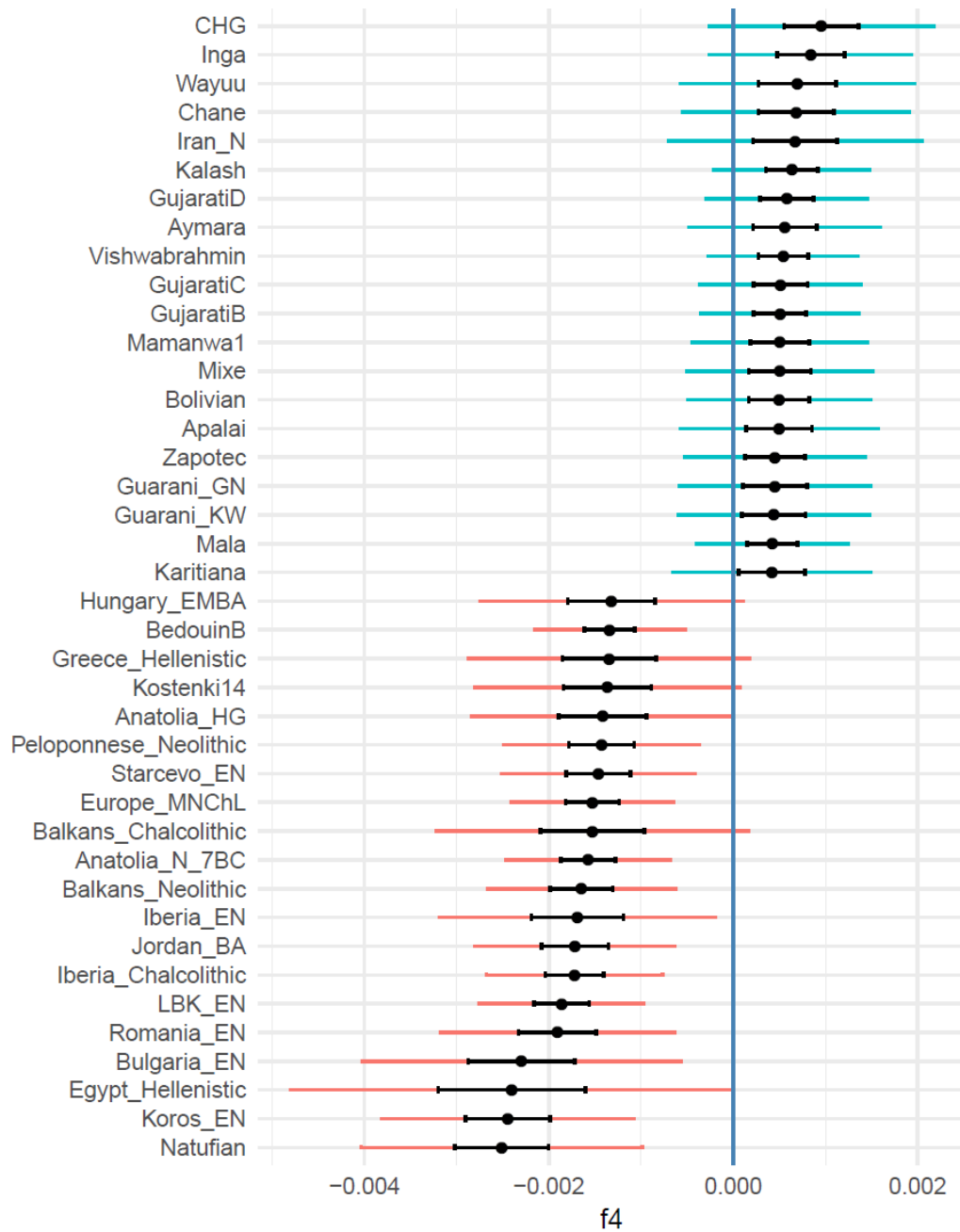


Fig. S5. The Bronze Age Ashkelon population shares a marginal higher genetic affinity with populations related to ancient Caucasus/Iran when compared to Chalcolithic Levantines. We plot the 20 highest (blue) and 20 lowest (orange) values of the f_4 -statistic of the form $f_4(\text{ASH_LBA}, \text{Levant_ChL}; \text{test}, \text{Mbuti})$ with ± 3 standard errors (represented by the colored horizontal lines) and ± 1 standard errors (represented by the black horizontal lines) both estimated by 5 cM block jackknifing. “test” populations include versatile global ancient and modern populations. Positive values indicate that “test” shares more alleles with ASH_LBA than with Levant_ChL, and negative values that it shares more with Levant_ChL than ASH_LBA. While Levant_ChL shares more alleles with earlier Levantine and Anatolian populations, ASH_LBA shows a slight excess of affinity with populations related to ancient Caucasus/Iran.

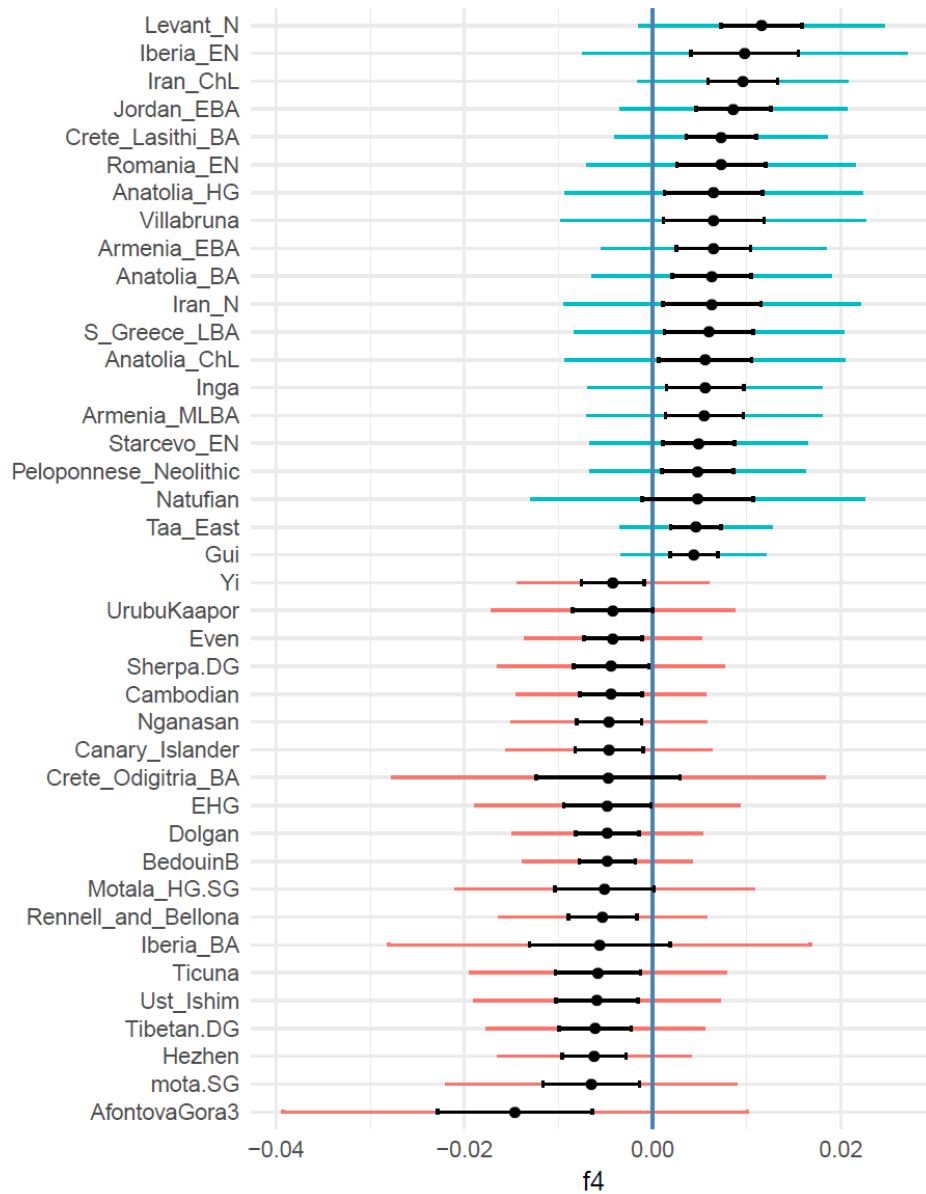


Fig. S6. The Bronze Age Ashkelon population is symmetrically related to Lebanon_MBA. We plot the 20 highest (blue) and 20 lowest (orange) values of the f_4 -statistic of the form $f_4(\text{ASH_LBA}, \text{Lebanon_MBA}; \text{test}, \text{Mbuti})$ with ± 3 standard errors (represented by the colored horizontal lines) and ± 1 standard errors (represented by the black horizontal lines) both estimated by 5 cM block jackknifing. “test” populations include versatile global ancient and modern populations.

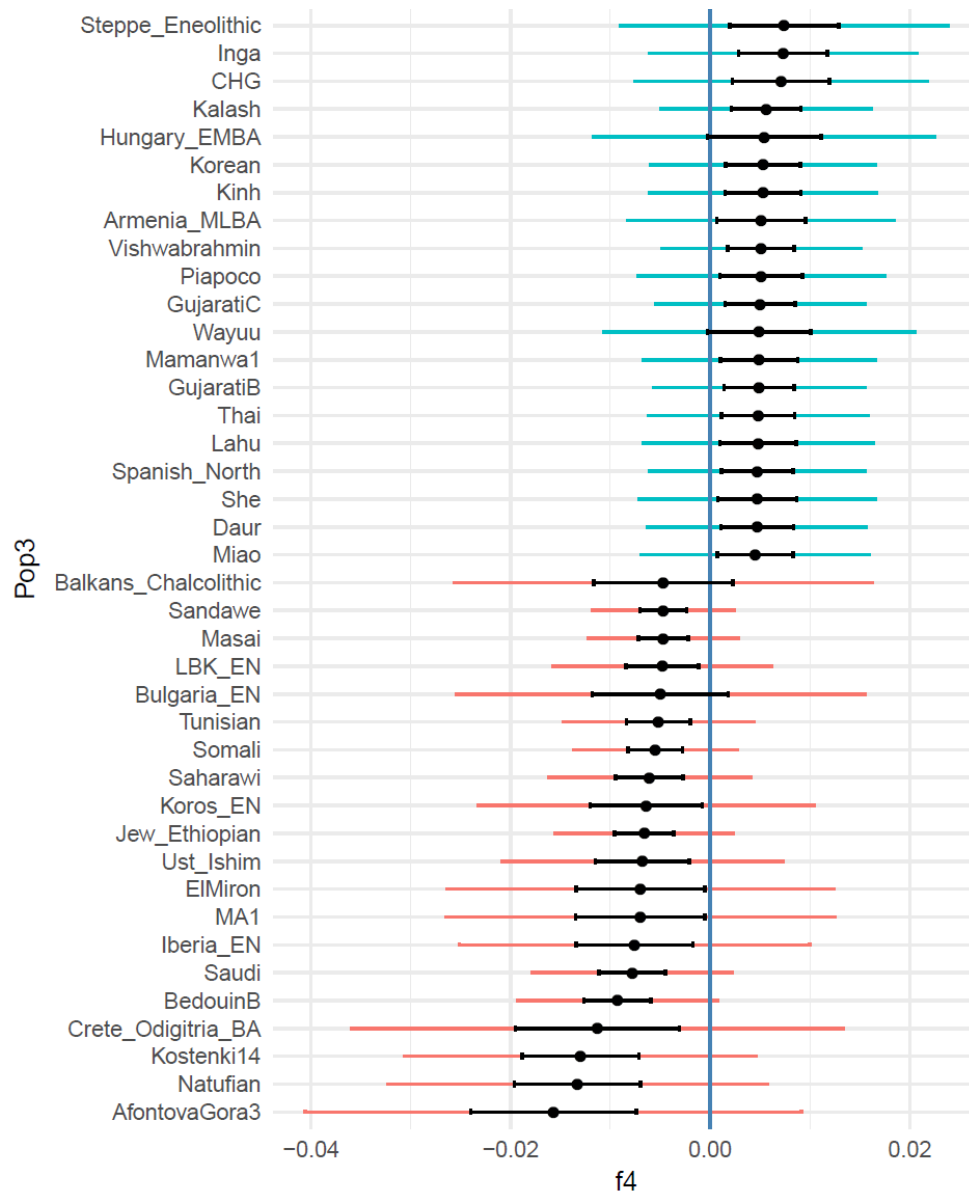


Fig. S7. The Bronze Age Ashkelon population is mostly symmetrically related to Jordan_EBA but shares a slightly higher genetic affinity with populations related to ancient Caucasus/Iran. We plot the 20 highest (blue) and 20 lowest (orange) values of the f_4 -statistic of the form $f_4(\text{ASH_LBA}, \text{Jordan_EBA}; \text{test}, \text{Mbuti})$ with ± 3 standard errors (represented by the colored horizontal lines) and ± 1 standard errors (represented by the black horizontal lines) both estimated by 5 cM block jackknifing. “test” populations include versatile global ancient and modern populations.

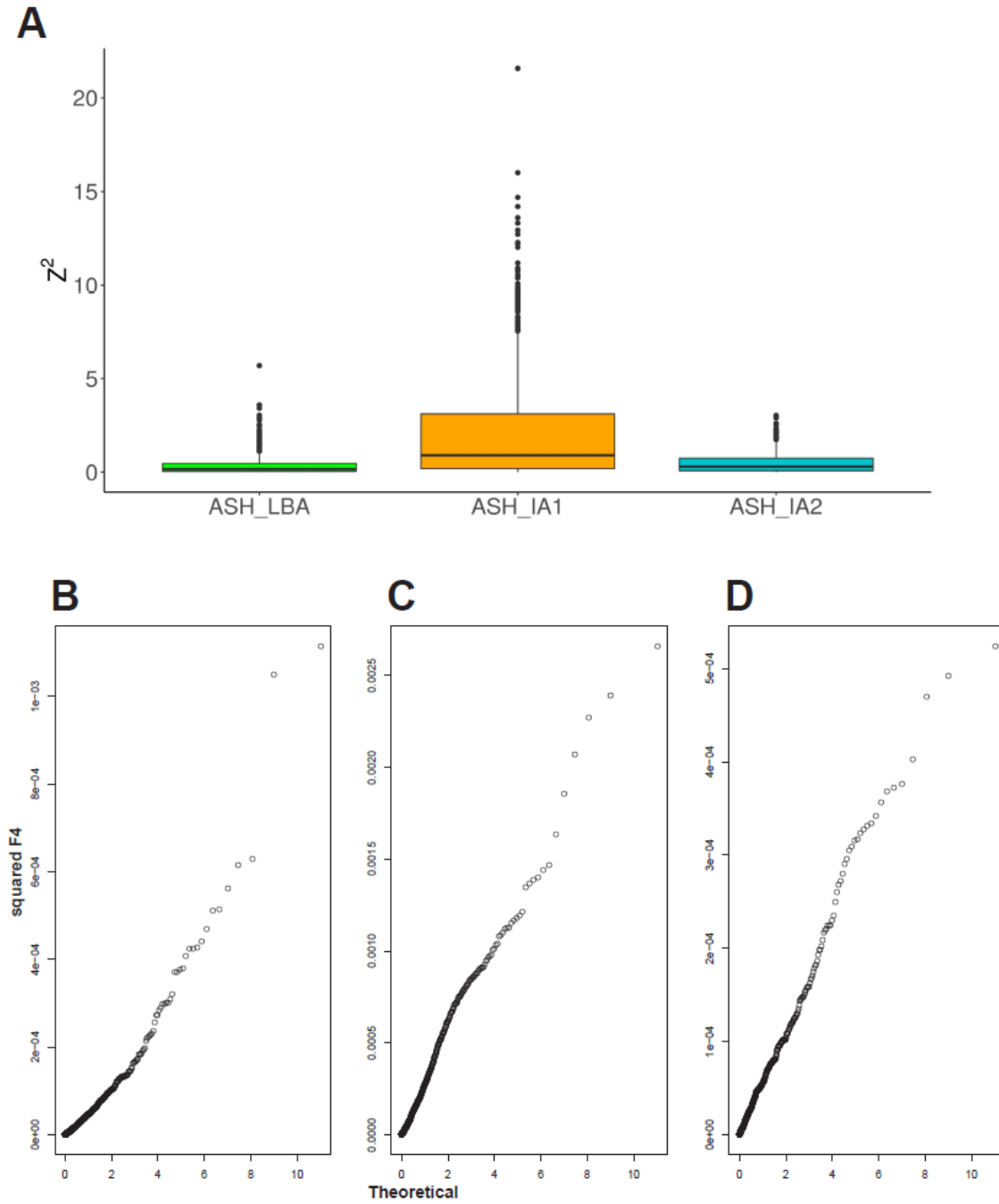


Fig. S8. Variation in differential affinities between the Ashkelon groups. (A) ASH_IA1 show the widest distribution of differential affinities. We plot the squared Z-score values of the f_4 -statistic of the form f_4 (Individual1, Individual 2; *test*, Mbuti) estimated by 5 cM block jackknifing. “test” populations include versatile global ancient and modern populations. Analysis was restricted to quadruples with at least 20,000 overlapping SNPs. The horizontal lines indicate the quartiles and black dots indicate outliers. **(B-D)** Q-Q plots used to estimate the correlation of the squared f_4 -values calculated by f_4 (Individual1, Individual 2; *test*, Mbuti) for ASH_LBA, ASH_IA1 and ASH_IA2 respectively to a Chi-squared distribution with 1 degree of freedom.

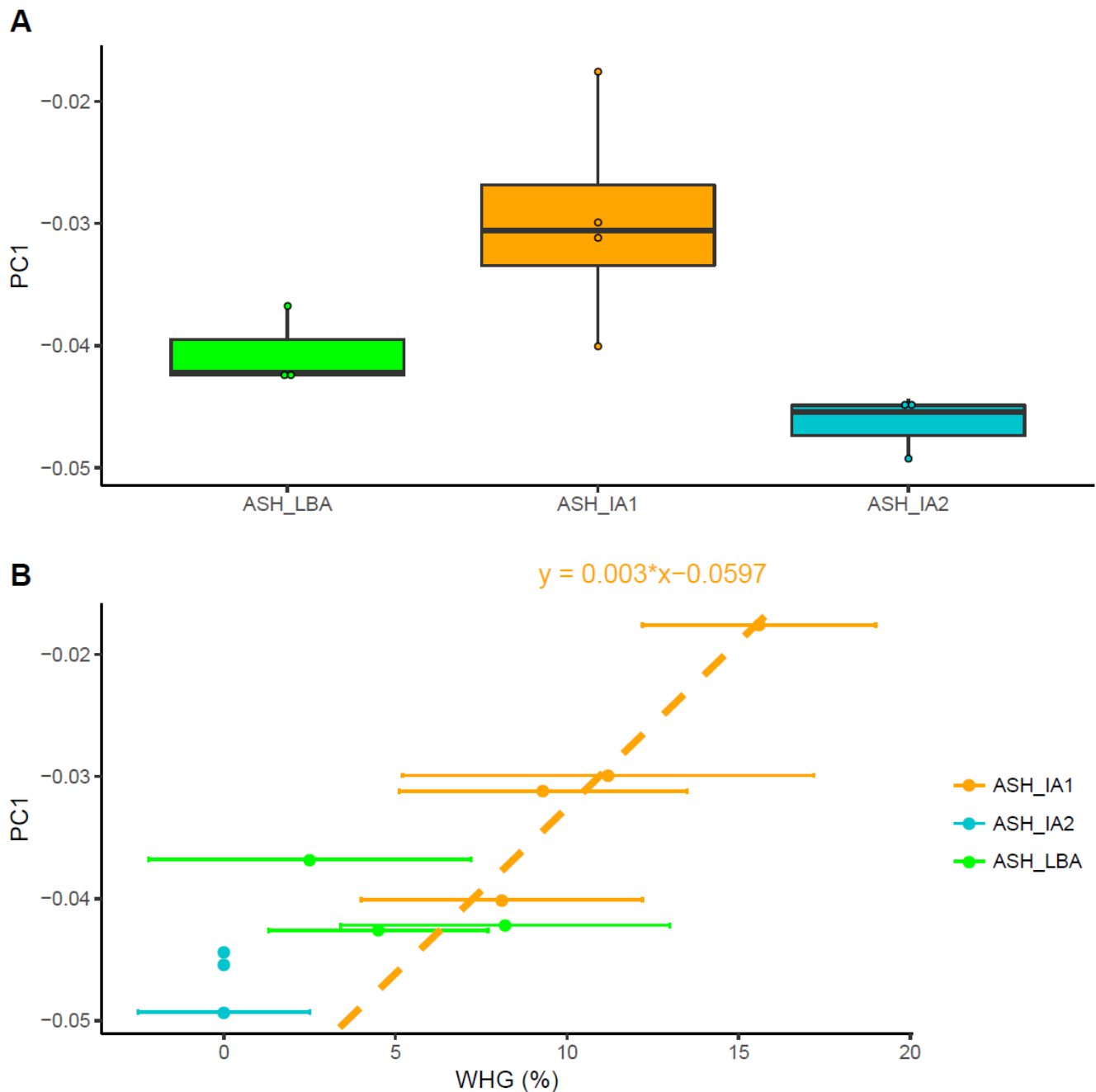


Fig. S9. The proportion of Mesolithic European ‘WHG’ ancestry in ASH_IA1 individuals correlates with their position in the west Eurasian PCA. (A) We plot the PC1 values of the Ashkelon individuals calculated in the west Eurasian PCA. The horizontal lines indicate the quartiles. **(B)** The PC1 values described in (A) plotted against the proportion of ‘WHG’ ancestry modeled in the Ashkelon individuals by a 3-way model including ‘Levant_N’, ‘Iran_N’ and ‘WHG’ as sources using qpAdm (table S6). The regression line is shown for the ASH_IA1 individuals.

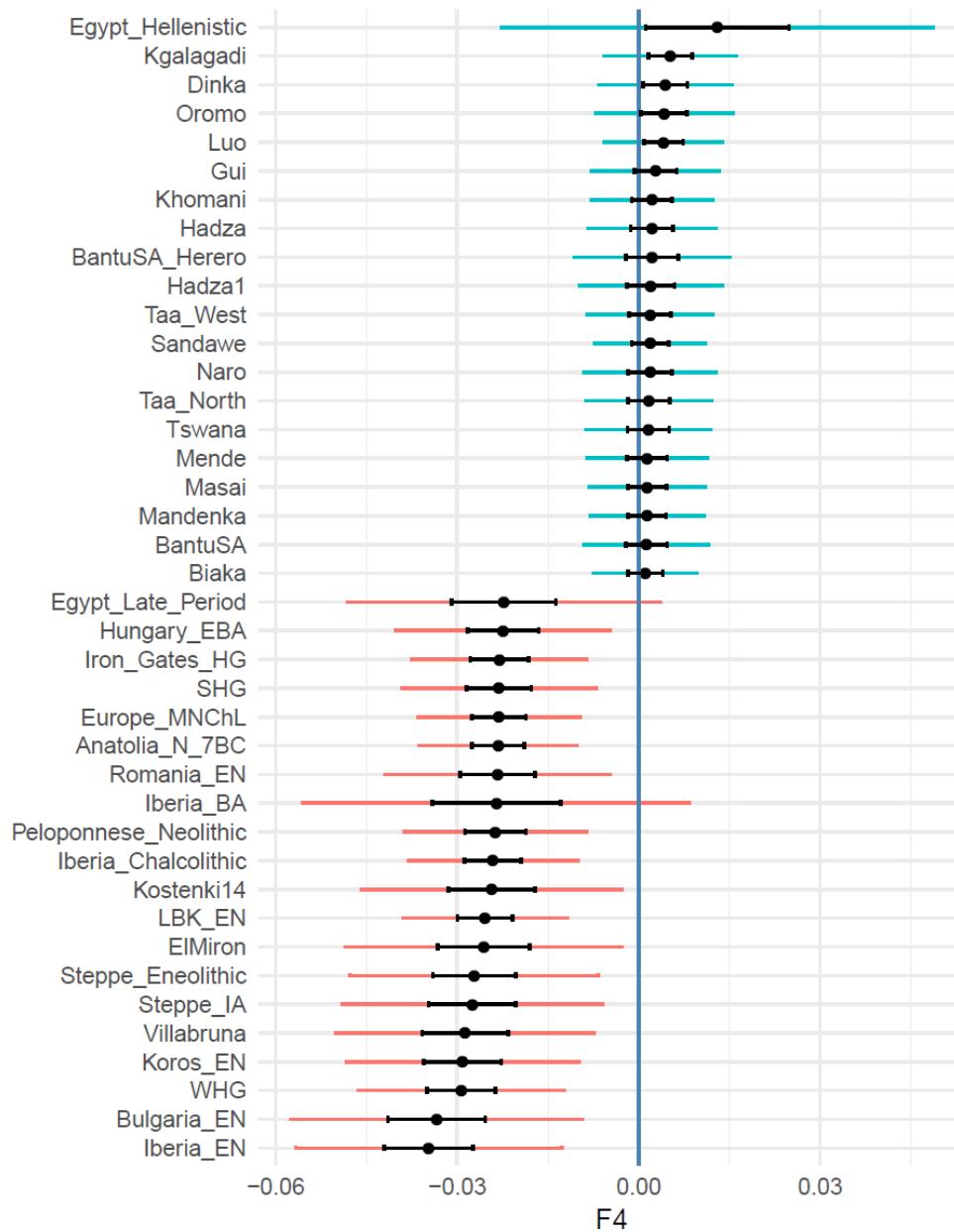


Fig. S10. ASH_IA1 shares more alleles with European- related populations compared to ASH_IA2. We plot the 20 highest (blue) and 20 lowest (orange) values of the f_4 -statistic of the form $f_4(\text{ASH_IA2}, \text{ASH_IA1}; \text{test}, \text{Mbuti})$ with ± 3 standard errors (represented by the colored horizontal lines) and ± 1 standard errors (represented by the black horizontal lines) both estimated by 5 cM block jackknifing. “test” populations include versatile global ancient and modern populations. Positive values indicate that “test” shares more alleles with ASH_IA2 than with ASH_IA1, and negative values that it shares more with ASH_IA1 than ASH_IA2.

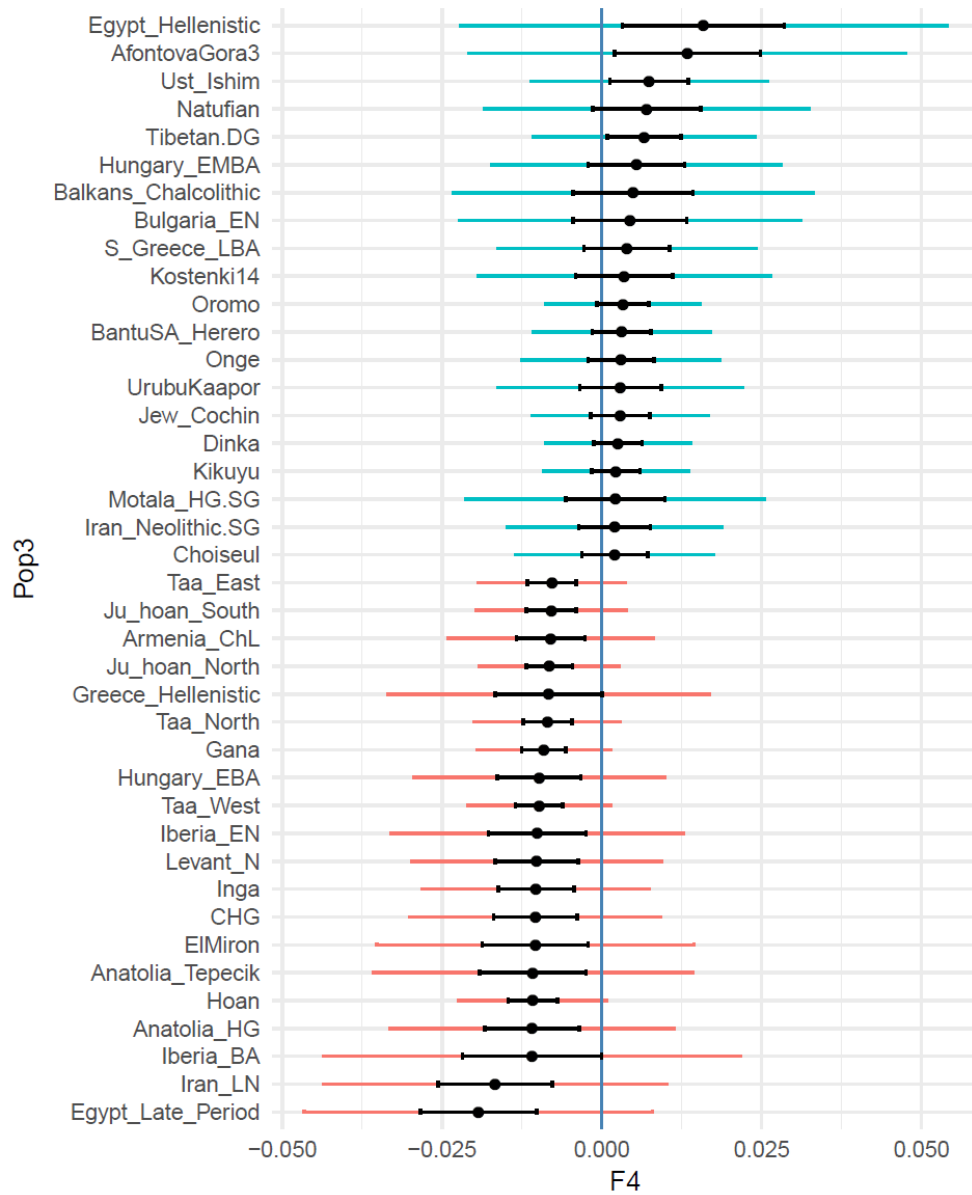


Fig. S11. ASH_IA2 is symmetrically related to ASH_LBA. We plot the 20 highest (blue) and 20 lowest (orange) values of the f_4 -statistic of the form $f_4(\text{ASH_IA2}, \text{ASH_LBA}; \text{test}, \text{Mbuti})$ with ± 3 standard errors (represented by the colored horizontal lines) and ± 1 standard errors (represented by the black horizontal lines) both estimated by 5 cM block jackknifing. “test” populations include versatile global ancient and modern populations.

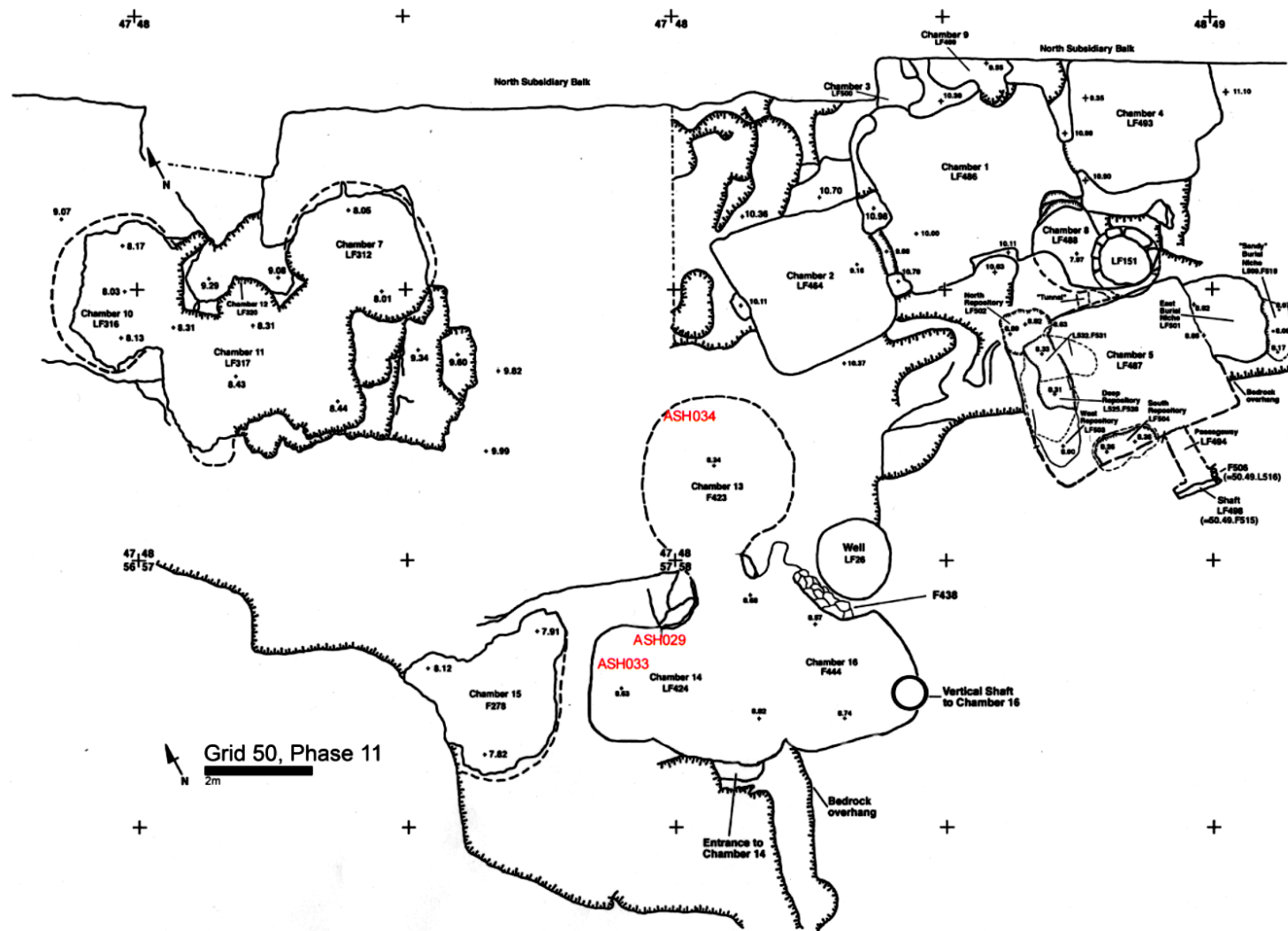
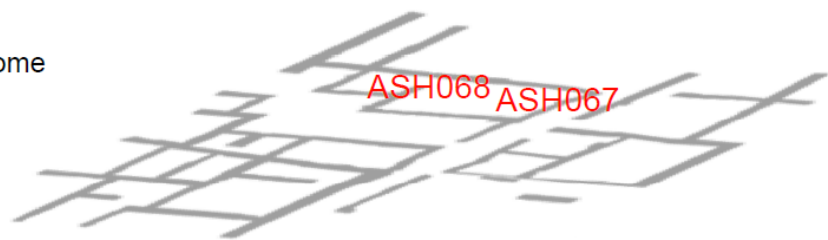
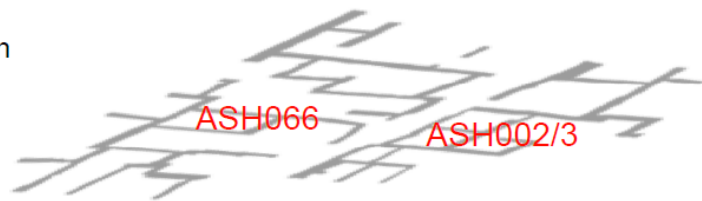


Fig. S12. Map of the Bronze Age Necropolis. Shown is the location of sampled individuals.

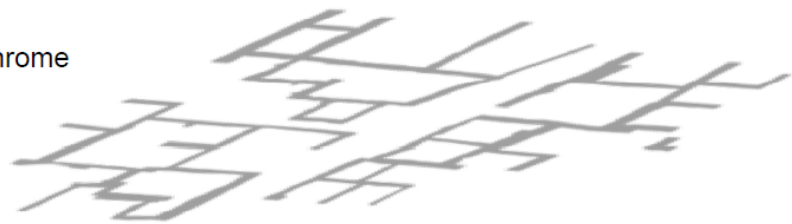
Phase 18b
Fully Developed Bichrome



Phase 19a
Bichrome more common



Phase 19b
Phil IIIC Middle- First Bichrome



Phase 20a
Phil IIIC Early 2- Middle



Phase 20b
Phil IIIC Early 1-2

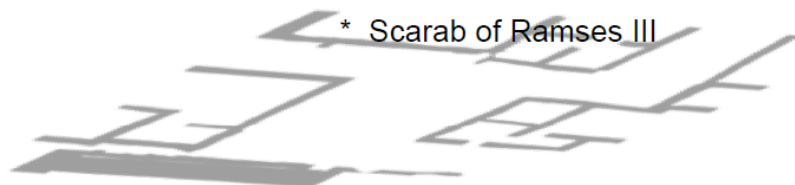


Fig. S13. Relative Stratigraphy of Grid 38. The plot includes the dominant ceramic assemblage in each phase. Location of scarab of Ramses III is also marked. Early Iron Age ceramic horizons after Mountjoy 2018 (66).

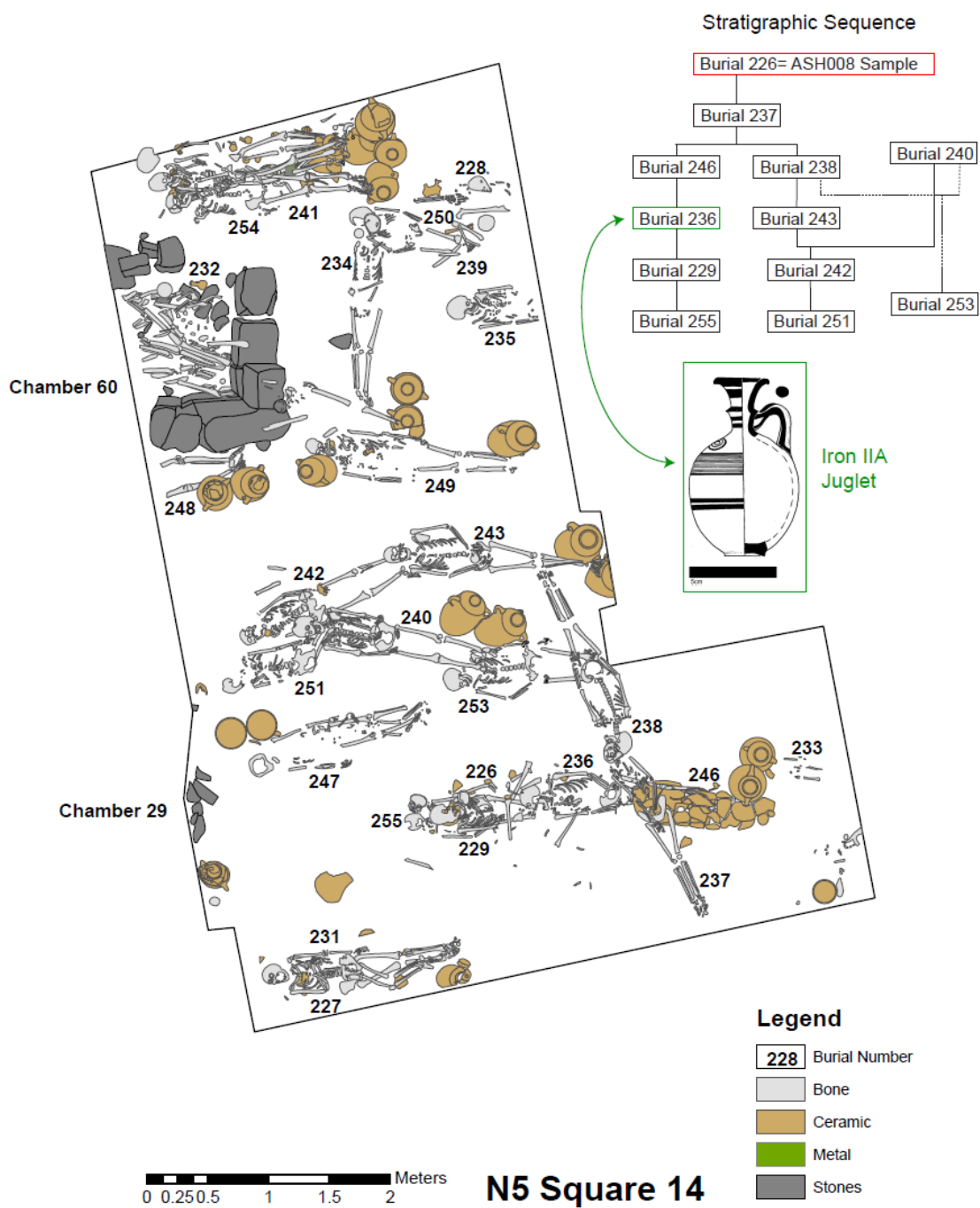


Fig. S14. Relative Stratigraphy of N5, Square 14. The location of Sample ASH008 is shown relative to well-dated Iron IIA ceramic forms. For further details see Master and Aja 2017 (20).

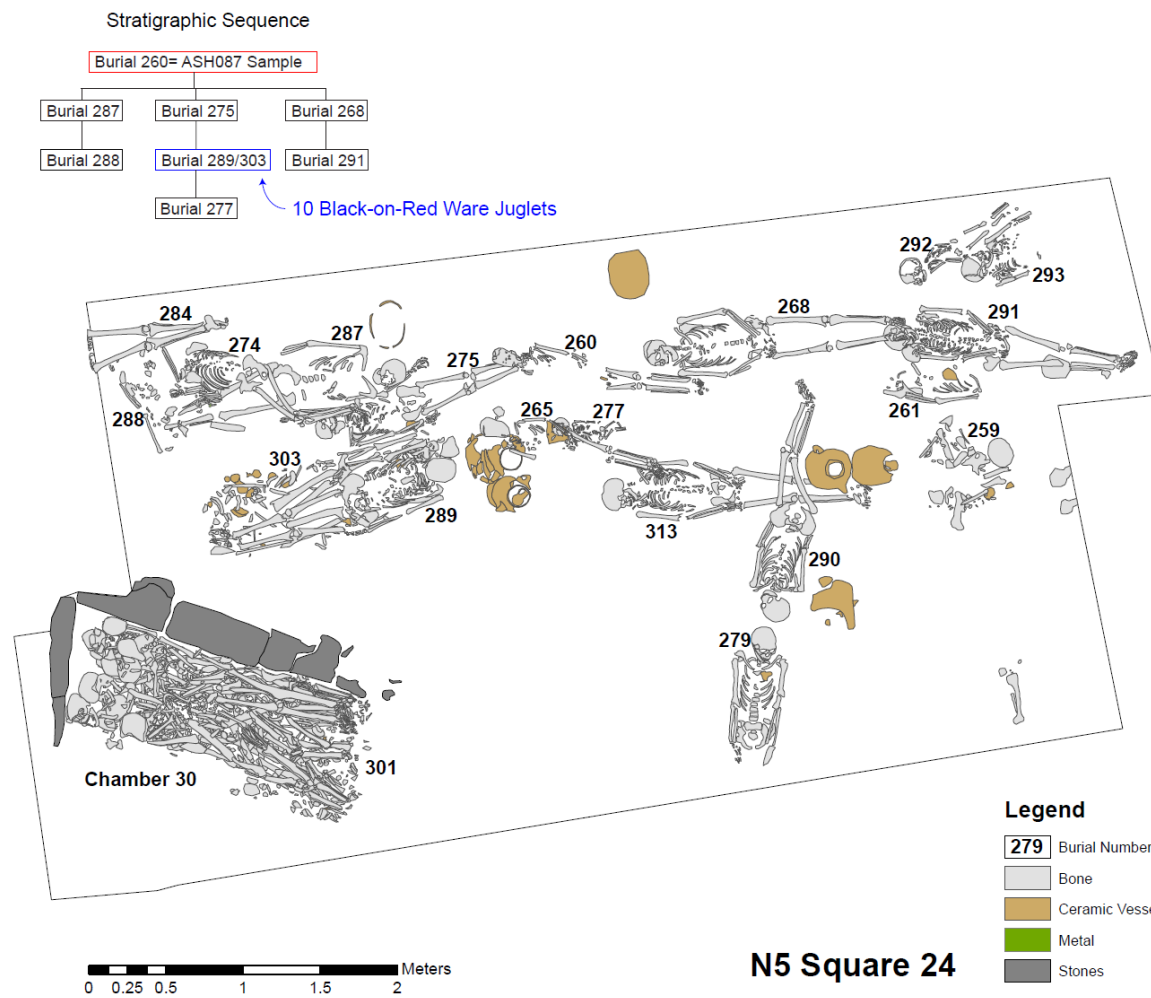


Fig. S15. Relative Stratigraphy of N5. Square 24. The location of Sample ASH087 is shown relative to well-dated Iron IIA ceramic forms. For further detail see Master and Aja 2017 (20).

Supplementary tables

Table S1. ^{14}C radiocarbon dating performed for this study. Carbon dated at the CEZ Archaeometry gGmbH, Mannheim, Germany. The ^{14}C ages are given in BP (before present; meaning years before 1950). The calibrated dates are shown in columns “Cal 1-sigma” and “Cal 2-sigma” using the 1-sigma and 2-sigma uncertainty of the ^{14}C ages, respectively. The $\delta^{13}\text{C}$ value was obtained from the isotope determination in the AMS system with a typical uncertainty of 2‰. This value may be influenced by isotope fractionation in the ion source and during graphitization and is only used for fractionation correction. Hence, this value is not comparable to the one obtained in a stable isotope IRMS and should not be used for further data interpretation.

Individual	Experiment number (MAMS)	C14 date (BP)	SE	Sigma	13C AMS [‰]	Cal 1-sigma (BCE)	Cal 2-sigma (BCE)	C:N	C (%)	Collagen (%)	Tissue
ASH029	30689	3,296	18	-10.2		1,611-1,535	1,622-1,522	3.2	40.3	2.6	Petrous bone
ASH033	30691	3,403	19	-17.9		1,740-1,666	1,746-1,643	3.2	29.4	0.7	Petrous bone
ASH2-3	31617	3,013	21	-15.5		1,280-1,221	1,378-1,134	3.2	39.5	2.6	Petrous bone
ASH066	33789	2,998	23	-17.0		1,275-1,136	1,371-1,129	3.2	41.6	3.0	Petrous bone
ASH067	33792	3,009	25	-14.9		1,284-1,212	1,379-1,131	3.2	41.8	5.4	Petrous bone
ASH068	33790	2,989	24	-17.9		1,262-1,133	1,284-1,126	3.2	41.1	2.8	Petrous bone
ASH008	30688	2,941	31	-45.0		1,213-1,095	1,257-1,042	3.4	35.3	0.5	Petrous bone

Table S2. Nuclear and mitochondrial contamination estimates. For each individual newly reported and analyzed in this study mitochondrial contamination estimates calculated with schmutzi (40) (Materials and Methods) are given. For genetic males the nuclear contamination estimate is provided (Materials and Methods). The levels of DNA damage are given as the deamination level at the 5' terminal position of the mapped reads.

Individual ID	Analysis group	Nuclear contamination estimate	SE	Mitochondrial contamination estimate	Deamination at 5' terminal position (%)
ASH029	ASH_LBA	NA	NA	0.09(0.08-0.10)	21.4
ASH033	ASH_LBA	NA	NA	0.01(0.00-0.02)	20.1
ASH034	ASH_LBA	NA	NA	0.05(0.04-0.06)	22.2
ASH2-3	ASH_IA1	NA	NA	0.04(0.03-0.05)	25.2
ASH066	ASH_IA1	0.043	0.031	0.04(0.03-0.05)	20.1
ASH067	ASH_IA1	NA	NA	0.01(0.00-0.02)	23.7
ASH068	ASH_IA1	0.027	0.028	0.06(0.05-0.07)	24.2
ASH008	ASH_IA2	0.006	0.005	0.04(0.03-0.05)	21.0
ASH087	ASH_IA2	-0.002	0.002	0.01(0.00-0.02)	22.4
ASH135	ASH_IA2	NA	NA	0.01(0.00-0.02)	28.0

Table S3. qpADM Admixture models of the Ashkelon Late Bronze Age group (ASH_LBA). The proportions estimated for each ancestral source (Ref1-3) used to model the target population ASH_LBA are given. Fitting models (Pval > 0.05; and admixture proportions are feasible) are highlighted in green. When resolution is lacking to determine whether a mixture proportion is required the source population, proportion and standard error are highlighted in lighter green. The best fitting models with minimal waves of ancestry are marked in bold. The standard errors were estimated by 5 cM block jackknifing. The abbreviations of the population names are listed in Data table S2. The ‘Basic set’ of outgroup populations (*Han*; *Onge*; *Mbuti*; *Natufian*; *Villabruna*; *Mala*; *Mixe*) was used.

Ancestral sources			Mixture proportions (%)			Standard errors (%)			Pval (rank-1)
Ref1	Ref2	Ref3	Ref1	Ref2	Ref3	Ref1	Ref2	Ref3	
Levant_ChL	Iran_ChL		60.0	40.0		6.5	6.5		0.445
Levant_ChL	Armenia_ChL		57.7	42.3		7.6	7.6		0.059
Levant_ChL	Anatolia_BA		41.2	58.8		11.9	11.9		0.060
Levant_ChL	Anatolia_EBA		41.5	58.5		10.8	10.8		0.053
Levant_ChL	Anatolia_ChL		37.2	62.8		14.5	14.5		0.044
Levant_ChL	Europe_MNChL		83.1	16.9		7.7	7.7		1.11E-06
Levant_ChL	Iran_ChL	Europe_MNChL	55.6	38.3	6.2	8.2	6.7	7.2	0.397
Levant_ChL	Armenia_ChL	Europe_MNChL	60.6	46.6	-7.2	8.3	9.4	8.8	0.042
Levant_ChL	Iran_ChL	Anatolia_BA	74.0	64.2	-38.2	18.2	30.7	46.7	0.429
Levant_ChL	Iran_ChL	Anatolia_EBA	52.6	28.3	19.1	30.2	62.8	91.9	0.047
Levant_ChL	Iran_ChL	Anatolia_ChL	55.1	36.9	8.1	12.5	14.8	24	0.268

Table S4: Fligner-Killeen test based on $[f_4(\text{Ashkelon ind 1}, \text{Ashkelon ind 2}; \text{test}, \text{Mbuti})]^2$ statistics

Population 1	Population 2	Variation 1	Variation 2	median chi-squared	df	p-value
ASH_LBA	ASH_IA1	1.22e-08	1.37e-07	374.3	1	< 2.2e-16
ASH_IA1	ASH_IA2	1.37e-07	6.31e-09	383.33	1	< 2.2e-16
ASH_IA2	ASH_LBA	6.31e-09	1.22e-08	8.10	1	0.004

Table S5. Ashkelon individuals and groups modeled using qpADM by ‘Levant_ChL’, ‘Iran_ChL’ and ‘WHG’ as sources. The proportions estimated for each ancestral source (Ref1-3) used to model the target populations are given. Fitting models (Pval > 0.05; and admixture proportions are feasible) are highlighted in green. When resolution is lacking to determine whether a mixture proportion is required the source population, proportion and standard error are highlighted in lighter green. The standard errors were estimated by 5 cM block jackknifing. The abbreviations of the population names are listed in Data table S2. The ‘Basic set’ of outgroup populations (*Han*; *Onge*; *Mbuti*; *Natufian*; *Villabruna*; *Mala*; *Mixe*) was used.

Target	Ancestral sources			Mixture proportions (%)			Standard errors (%)			Pval (rank-1)
	Ref1	Ref2	Ref3	Ref1	Ref2	Ref3	Ref1	Ref2	Ref3	
ASH_LBA	Levant_ChL	Iran_ChL	WHG	58.4	39.7	2.0	6.7	6.4	2.2	0.394
ASH29	Levant_ChL	Iran_ChL	WHG	46.9	51.6	1.5	15.1	15.0	4.3	0.129
ASH33	Levant_ChL	Iran_ChL	WHG	60.0	39.9	0.2	14.9	14.1	4.6	0.454
ASH34	Levant_ChL	Iran_ChL	WHG	61.3	36.3	2.3	8.4	8.1	2.9	0.492
ASH_IA1	Levant_ChL	Iran_ChL	WHG	56.3	32.3	11.4	6.2	5.7	2.0	0.765
ASH2-3	Levant_ChL	Iran_ChL	WHG	61.2	28.1	10.7	15.6	15.5	5.3	0.877
ASH066	Levant_ChL	Iran_ChL	WHG	60.7	30.3	9.0	11.9	11.2	3.4	0.433
ASH067	Levant_ChL	Iran_ChL	WHG	39.6	51.8	8.6	11.4	10.7	3.7	0.564
ASH068	Levant_ChL	Iran_ChL	WHG	70.0	14.8	15.2	9.3	8.7	3.1	0.406
ASH_IA2	Levant_ChL	Iran_ChL	WHG	64.9	37.5	-2.3	6.5	6.2	2.2	0.317
ASH135	Levant_ChL	Iran_ChL	WHG	73.1	23.6	3.2	17.4	16.3	5.2	0.370
ASH087	Levant_ChL	Iran_ChL	WHG	71.8	33.9	-5.6	11.8	11.2	3.6	0.528
ASH8	Levant_ChL	Iran_ChL	WHG	54.7	46.2	-0.9	8.0	8.0	2.8	0.668

Table S6. Bronze Age and Iron Age individuals and groups modeled by the three maximized populations on the west Eurasian PCA (using qpADM). The proportions estimated for each ancestral source (Ref1-3) used to model the target populations are given. The standard errors were estimated by 5 cM block jackknifing. The abbreviations of the population names are listed in Data table S2. The ‘Basic set’ of outgroup populations (*Han; Onge; Mbuti; Natufian; Villabruna; Mala; Mixe*) was used.

Target	Ancestral sources			Mixture proportions (%)			Standard errors (%)			Pval (rank-1)
	Ref1	Ref2	Ref3	Ref1	Ref2	Ref3	Ref1	Ref2	Ref3	
ASH_LBA	Levant_N	Iran_N	WHG	51.2	43.6	5.2	4.7	4.2	2.6	0.341
ASH29	Levant_N	Iran_N	WHG	55.3	42.3	2.5	8.6	7.7	4.7	0.331
ASH33	Levant_N	Iran_N	WHG	45.9	45.9	8.2	8.1	7.6	4.8	0.982
ASH34	Levant_N	Iran_N	WHG	53.6	42.0	4.5	5.9	5.5	3.2	0.148
ASH_IA1	Levant_N	Iran_N	WHG	50.0	38.3	11.7	4.1	3.7	2.3	0.021
ASH2-3	Levant_N	Iran_N	WHG	52.7	36.1	11.2	8.7	7.9	6.0	0.900
ASH066	Levant_N	Iran_N	WHG	53.4	38.5	8.1	8.3	7.1	4.1	0.338
ASH067	Levant_N	Iran_N	WHG	39.2	51.4	9.3	7.9	7.1	4.2	0.143
ASH068	Levant_N	Iran_N	WHG	57.2	27.2	15.6	5.8	5.3	3.4	0.048
ASH_IA2	Levant_N	Iran_N	WHG	60.0	40.0	0.0	4.4	3.8	2.5	0.053
ASH135	Levant_N	Iran_N	WHG	68.3	32.6	-0.9	9.7	8.2	6.1	0.836
ASH087	Levant_N	Iran_N	WHG	62.4	41.9	-4.3	7.3	6.3	4.0	0.442
ASH8	Levant_N	Iran_N	WHG	55.2	42.6	2.2	5.2	4.8	2.9	0.177
Jordan_EBA	Levant_N	Iran_N	WHG	65.2	34.3	0.5	3.1	2.8	1.8	0.509
I1705	Levant_N	Iran_N	WHG	67.0	32.4	0.5	4.3	3.9	2.4	0.454
I1706	Levant_N	Iran_N	WHG	64.9	32.6	2.5	4.2	4.0	2.4	0.073
I1730	Levant_N	Iran_N	WHG	64.6	36.1	-0.7	4.4	4.0	2.6	0.998
Lebanon_MBA	Levant_N	Iran_N	WHG	45.5	51.5	3.0	2.8	2.6	1.7	0.056
ERS1790729	Levant_N	Iran_N	WHG	42.7	55.7	1.6	6.6	6.3	3.7	0.267
ERS1790730	Levant_N	Iran_N	WHG	45.3	53.8	0.9	4.5	4.2	2.6	0.390
ERS1790731	Levant_N	Iran_N	WHG	48.9	47.0	4.1	4.3	3.9	2.6	0.075
ERS1790732	Levant_N	Iran_N	WHG	38.0	56.4	5.5	4.4	4.1	2.8	0.232

ERS1790733	Levant_N	Iran_N		50.8	49.2	NA	4.5	4.5	NA	0.151
ERS1790733	Levant_N	Iran_N	WHG	52.5	50.0	-2.5	4.8	4.6	2.8	0.125

Table S7. qpADM Admixture models of the Ashkelon Iron Age 1 group (ASH_IA1). The proportions estimated for each ancestral source (Ref1-3) used to model the target population ASH_IA1 are given. Fitting models (Pval > 0.05; and admixture proportions are feasible) are highlighted in green. When resolution is lacking to determine whether a mixture proportion is required the source population, proportion and standard error are highlighted in lighter green. The best fitting models with minimal waves of ancestry are marked in bold. The standard errors were estimated by 5 cM block jackknifing. The abbreviations of the population names are listed in Data table S2. We define the ‘Basic set’ of outgroup populations (*Han; Onge; Mbuti; Natufian; Villabruna; Mala; Mixe*).

Ancestral sources			Mixture proportions (%)			Standard errors (%)			Pval (rank-1)
Ref1	Ref2	Ref3	Ref1	Ref2	Ref3	Ref1	Ref2	Ref3	
ASH_LBA	Anatolia_BA		71.9	28.1	NA	47.8	47.8	NA	0.025
ASH_LBA	Anatolia_EBA		181.4	-81.4	NA	304.9	304.9	NA	0.044
ASH_LBA	Anatolia_MLBA		8160.2	-8060.2	NA	6445390.5	6445390.5	NA	0.191
ASH_LBA	Egypt_Late_Period		217.2	-117.2	NA	220.2	220.2	NA	0.499
ASH_LBA	Egypt_New_Kingdom		250.6	-150.6	NA	521.8	521.8	NA	0.302
ASH_LBA	Jordan_EBA		394.4	-294.4	NA	723.8	723.8	NA	0.119
ASH_LBA	Crete_Lasithi_BA		-19.0	119.0	NA	93.1	93.1	NA	0.105
ASH_LBA	Crete_Odigitria_BA		56.9	43.1	NA	19.2	19.2	NA	0.675
ASH_LBA	Crete_Armenoi		69.0	31.0	NA	32.1	32.1	NA	0.929
ASH_LBA	S_Greece_LBA		-124.6	224.6	NA	338.5	338.5	NA	0.271
ASH_LBA	Sardinian		64.8	35.2	NA	17.4	17.4	NA	0.070
ASH_LBA	Lebanon_MBA		1751.1	-1651.1	NA	9490.0	9490.0	NA	0.636
ASH_LBA	Armenia_ChL		80.2	19.8	NA	23.6	23.6	NA	0.020
ASH_LBA	Iberia_BA		78.2	21.8	NA	21.1	21.1	NA	0.205
ASH_LBA	Steppe_EMBA		90.5	9.5	NA	7.9	7.9	NA	0.030
ASH_LBA	Steppe_MLBA		84.3	15.7	NA	9.1	9.1	NA	0.050
ASH_LBA	Anatolia_BA	Steppe_MLBA	69.6	16.9	13.5	36.0	39.9	9.7	0.031
ASH_LBA	Armenia_ChL	Anatolia_BA	-43.2	-169.2	312.4	416.1	666.6	1077.5	0.008
ASH_LBA	Armenia_ChL	Iberia_BA	87.2	-23.3	36.1	31.3	54.1	42.5	0.114

ASH_LBA	Armenia_ChL	S_Greece_LBA	-242.4	-132.1	474.4	744.6	362.4	1094.2	0.357
ASH_LBA	Armenia_ChL	Steppe_EMBA	186.8	-158.1	71.4	468.7	772.6	304.5	0.017
ASH_LBA	Armenia_ChL	Steppe_MLBA	100.7	-28.5	27.9	44.1	73.6	32.3	0.025
ASH_LBA	Iberia_BA	Anatolia_BA	43.5	6.3	50.2	35.6	21.1	45.4	0.241
ASH_LBA	Iberia_BA	Steppe_MLBA	64.9	107.4	-72.3	27.8	89.2	70.4	0.427
ASH_LBA	Crete_Odigitria_BA	Anatolia_BA	61.5	51.7	-13.3	26.5	36.5	47.7	0.554
ASH_LBA	Crete_Odigitria_BA	Armenia_ChL	57.0	71.8	-28.8	21.8	47.8	43.2	0.615
ASH_LBA	Crete_Odigitria_BA	Iberia_BA	77.4	-2.4	25.0	102.3	199.7	107.6	0.457
ASH_LBA	Crete_Odigitria_BA	S_Greece_LBA	89.2	42.1	-31.3	46.7	35.3	61.7	0.954
ASH_LBA	Crete_Odigitria_BA	Steppe_EMBA	55.2	51.5	-6.8	19.7	25.6	12.4	0.589
ASH_LBA	Crete_Odigitria_BA	Steppe_MLBA	56.8	52.1	-8.9	19.5	27.1	17.1	0.586
ASH_LBA	S_Greece_LBA	Anatolia_BA	-104.1	297.9	-93.7	278.3	418.1	168.8	0.336
ASH_LBA	S_Greece_LBA	Iberia_BA	-380.4	479.7	0.7	10545.4	10442.7	195.8	0.822
ASH_LBA	S_Greece_LBA	Steppe_MLBA	-609.4	865.2	-155.8	72489.4	90731.5	18242.5	0.401
ASH_LBA	Steppe_EMBA	Anatolia_BA	79.0	7.6	13.3	48.2	10.9	55.7	0.017
ASH_LBA	Steppe_EMBA	Iberia_BA	70.1	-26.1	56.0	25.7	30.8	50.7	0.218
ASH_LBA	Steppe_EMBA	S_Greece_LBA	-224.6	-38.3	362.9	870.9	136.2	1003.7	0.303
Levant_ChL	Iran_ChL	Anatolia_BA	1.4	-43.6	142.2	26.0	27.3	50.6	0.185
Levant_ChL	Iran_ChL	Anatolia_EBA	16.9	-54.4	137.5	20.6	34.7	51.0	0.134
Levant_ChL	Iran_ChL	Anatolia_MLBA	10.2	-55.3	145.1	23.1	30.2	49.6	0.375
Levant_ChL	Iran_ChL	Armenia_ChL	51.1	-23.6	72.5	7.4	13.5	16.1	0.216
Levant_ChL	Iran_ChL	Egypt_Late_Period	496.6	110.4	-506.9	1246.5	253.3	1493.8	0.036
Levant_ChL	Iran_ChL	Egypt_New_Kingdom	-111.6	34.8	176.8	456.3	44.8	482.9	0.020
Levant_ChL	Iran_ChL	Iberia_BA	62.4	11.8	25.8	9.6	11.6	9.3	0.964
Levant_ChL	Iran_ChL	Jordan_EBA	979.4	222.6	-1102.0	1956.5	413.5	2365.1	0.221
Levant_ChL	Iran_ChL	Crete_Lasithi_BA	-6.0	-15.6	121.6	19.6	13.7	29.8	0.500
Levant_ChL	Iran_ChL	Crete_Odigitria_BA	47.5	14.5	38.0	19.2	10.6	22.0	0.972

Levant_ChL	Iran_ChL	Crete_Armenoi	4.8	41.6	53.6	30.3	25.5	21.8	0.787
Levant_ChL	Iran_ChL	S_Greece_LBA	8.0	-16.5	108.5	19.6	16.4	30.9	0.322
Levant_ChL	Iran_ChL	Sardinian	39.5	11.3	49.3	7.6	6.5	8.5	0.715
Levant_ChL	Iran_ChL	Lebanon_MBA	-170.1	-190.0	460.1	256.4	243.7	498.3	0.038
Levant_ChL	Iran_ChL	Steppe_EMBA	75.0	-4.4	29.4	5.9	8.8	5.4	0.312
Levant_ChL	Iran_ChL	Steppe_MLBA	62.9	5.3	31.8	5.7	7.1	5.5	0.760

Table S8. qpADM Admixture models of the Ashkelon Iron Age 2 group (ASH_IA2). The proportions estimated for each ancestral source (Ref1-3) used to model the target population ASH_IA2 are given. Fitting models (Pval > 0.05; and admixture proportions are feasible) are highlighted in green. When resolution is lacking to determine whether a mixture proportion is required the source population, proportion and standard error are highlighted in lighter green. The best fitting models with minimal waves of ancestry are marked in bold. The standard errors were estimated by 5 cM block jackknifing. The abbreviations of the population names are listed in Data table S2. The ‘Basic set’ of outgroup populations (*Han*; *Onge*; *Mbuti*; *Natufian*; *Villabruna*; *Mala*; *Mixe*) was used.

Ancestral sources		Mixture proportions (%)		Standard errors (%)		Pval (rank-1)
Ref1	Ref2	Ref1	Ref2	Ref1	Ref2	
ASH_IA1	Anatolia_BA	271.5	-171.5	358.0	358.0	0.101
ASH_IA1	Anatolia_IA	110.9	-10.9	22.3	22.3	0.048
ASH_IA1	Jordan_EBA	7.3	92.7	47.1	47.1	0.292
ASH_IA1	ASH_LBA	31.2	68.8	45.0	45.0	0.688
ASH_IA1	Egypt_Late_Period	38.2	61.8	25.1	25.1	0.591
ASH_IA1	Egypt_New_Kingdom	112.7	-12.7	43.5	43.5	0.615
ASH_IA1	Crete_Lasithi_BA	-9887.2	9987.2	3675379.2	3,675,379.2	0.730
ASH_IA1	Crete_Odigitria_BA	-8764.3	8864.3	3054312.5	3,054,312.5	0.764
ASH_IA1	S_Crete_LBA	-345.6	445.6	853.8	853.8	0.486
ASH_IA1	Sardinian	183.2	-83.2	39.5	39.5	0.611
ASH_IA1	Lebanon_MBA	8.6	91.4	39.9	39.9	0.341

Table S9. Phenotypic analysis. Allele counts in SNP positions related to phenotypic traits. For each reference SNP (SNP ID) we provide the corresponding chromosome (Chr), the position on the Hg19 human reference (Position), the reference allele (Ref) and the alternative one (Alt), both on the positive strand, the gene name (Gene) and the alternative allele phenotype. The counts of sequenced reads covering each “Ref” and “Alt” allele are given for each of the reported individuals. The counts are in bold when positions were covered more than once.

SNP ID	Chr	Position	Ref	Alt	Gene	Alternative allele phenotype	ASH34		ASH33		ASH29		ASH2-3		ASH066		ASH067		ASH068		ASH8		ASH087		ASH135	
							Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt
snp_2_136608642	2	136,608,642	A	T	LCT	Lactase persistence	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0
rs41525747	2	136,608,643	G	C	LCT	Lactase persistence	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0
rs4988236	2	136,608,644	G	A	LCT	Lactase persistence	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0
rs4988235	2	136,608,646	G	A	LCT	Lactase persistence	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0
rs41456145	2	136,608,649	A	G	LCT	Lactase persistence	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0
rs41380347	2	136,608,651	A	C	LCT	Lactase persistence	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0
rs1541255	1	203,652,141	G	A	ATP2B4	No increased risk of severe Malaria	0	2	0	0	0	0	0	0	0	0	0	1	0	2	0	0	0	2	0	0
rs8176746	9	136,131,322	G	T	ABO	increased risk of severe malaria	0	0	0	0	0	1	0	0	0	0	0	0	1	0	2	0	0	1	0	0
rs762516	X	153,764,663	C	T	G6PD	G6PD deficiency	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0
rs149902811	X	153,773,160	A	C	G6PD	G6PD deficiency	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
rs3827760	2	109,513,601	A	G	EDAR	Thick straight hair	0	0	0	0	0	0	1	0	0	0	3	0	2	0	0	0	0	0	0	0
rs1042602	11	88,911,696	C	A	TYR	freckles	1	0	0	0	0	0	0	0	0	0	0	2	0	1	0	0	1	0	0	0
rs1800404	15	28,235,773	C	T	OCA2	lighter skin	2	1	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0

						pigmentation																				
rs12913832	15	28,365,618	A	G	HERC2	lighter eye color	1	0	0	0	0	0	0	0	1	0	1	0	0	0	3	0	0	0	0	0
rs1426654	15	48,426,484	A	G	SLC24A5	darker skin color	2	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0

Table S10. Archaeological context

ID	Context	Architectural Context	Archaeological Date
ASH029.A0101	50.58 LF424 Burial 110	Chamber 14	MBIIC- LBII
ASH033.A0101	50.58 LF424 Burial 111	Chamber 14	MBIIC- LBII
ASH034.A0101	50.58 F423 Burial 128	Chamber 13	MBIIC- LBII
ASH066.A0101	38.74.F1056.L1066	Phase 19A, Building 957, Room 957	Iron I- Post Ramses III
ASH002/3.A0101	38.75 L386	Phase 19A, Building 577, Room 392	Iron I- Post Ramses III
ASH067.A0101	38.74.F962.L963	Phase 18B, Building 905, Room 905	Iron I- Post Ramses III
ASH068.A0101	38.74.L985	Phase 18B, Building 905, Room 905	Iron I- Post Ramses III
ASH135.A0101	N5.13.Burial 343	N/A	Iron IIA
ASH008.A0101	N5.14.Burial 226	N/A	Iron IIA
ASH087.A0101	N5.24.Burial 260	N/A	Iron IIA

14.3 Supplementary materials of paper C

Supplementary material

Supplementary Tables

Table S1: Results of conventional PCR and sequencing pre-screening on the *Y. pestis* specific *p/a* fragment (133 bp)

Individual	Tooth sample 1	Tooth sample 2
AE96	-	-
AE97	-	-
AE127	-	-
AE128	-	-
AE349	-	-
AE350	-	-
AE468	-	-
AE469	-	-
AE887	-	-
AE888	-	-
AE1004	-	-
AE1005	-	-
AE1154	-	-
AE1155	-	-
AE1175	-	+
AE1176	+	-
AE1184	-	-
AE1185	-	-
AE1223	-	-
AE1241	-	-

Table S2: Carbon-dating

Individual	Sample	C14 date	+/-	Cal 1 sigma	Cal 2 sigma	C:N	%C	percent collagen
AE1175	metacarpal	1541	19	Cal AD 434-556	Cal AD 428-571	3.3 ^a	37.8 ^a	13.8 ^a
AE1176	metacarpal	1563	20	Cal AD 430-538	Cal AD 426-545	3.3 ^a	36.1 ^a	3.4 ^a

^a The C:N relation, percent C and percent collagen values are in the normal range and show a good collagen quality.

Table S3: Copy number estimation of plasmids

	Chromosome	pCD1	pMT1
Reference sequence (CO92)	NC_003143.1	NC_003131.1	NC_003134.1
Average read length	67.41 bp	68.7 bp	61 bp
Selected region of full mappability (positions in reference)	2306929 - 2311929	24206-29205	2000-6999
"G+C" content in region of full mappability	48.04 %	47.98 %	47.96 %
Average coverage in region of full mappability	21.2 X	60.8 X	54 X
Estimated copy number ^a	-	2.87	2.55

^a The copy number of the plasmids was estimated on the average coverage in the region of full mappability divided by that of the chromosome, and is likely to be affected by ascertainment bias in the array capture.

Table S4: Genomes used for SNP calling and phylogeny

Strain ID	SNP Calls	Coverage (fold)	Coverage (percent)	accession
0.ANT1a_42013	162	63.19	93.93	ADPG000000000 47685
0.ANT1b_CMCC49003	172	66.88	94.59	ADQX000000000 47685
0.ANT1c_945	185	50.49	92.05	ADPV000000000 47685
0.ANT1d_164	179	47.76	93.37	ADOW000000000 47685
0.ANT1e_CMCC8211	188	55.41	94.27	ADRD000000000 47685
0.ANT1f_42095	185	47.24	93.98	ADPJ000000000 47685
0.ANT1g_CMCC42007	185	53.09	93.43	ADQV000000000 47685
0.ANT1h_CMCC43032	182	38.54	93.05	ADQW000000000 47685
0.ANT2_B42003004	90	94.52	95.25	NZ_AAYU000000000
0.ANT2a_2330	90	66.53	92.44	ADOY000000000 47685
0.ANT3a_CMCC38001	93	67.2	93.74	ADQU000000000 47685
0.ANT3b_A1956001	91	66.62	93.98	ADPX000000000 47685
0.ANT3c_42082	92	27.29	92.06	ADPH000000000 47685
0.ANT3d_CMCC21106	99	42.78	91.74	ADQP000000000 47685
0.ANT3e_42091b	98	85.13	93.89	ADPI000000000 47685

0.PE2_PEST-F	507	92.06	93.07	NC_009381
0.PE2b_G8786	567	51.87	91.33	ADSG00000000 47685
0.PE3_Angola	889	89.73	91.04	NC_010159
0.PE4_Microtus91001	351	93.27	94.25	NC_005810
0.PE4Aa_12	267	49.15	92.27	ADOV00000000 47685
0.PE4Ab_9	264	46.84	93.04	ADPT00000000 47685
0.PE4Ba_PestoidesA	315	93.44	94.42	NZ_ACNT00000000
0.PE4Ca_CMCCN01002 5	325	61.42	92.9	ADRT00000000 47685
0.PE4Cb_M0000002	316	49.02	93.05	ADST00000000 47685
0.PE4Cc_CMCC18019	322	35.08	93.45	ADQO00000000 47685
0.PE4Cd_CMCC93014	331	52.98	94.06	ADRM00000000 47685
0.PE4Ce_CMCC91090	333	87.7	93.18	ADRJ00000000 47685
0.PE7b_620024	444	38.79	92.68	ADPM00000000 47685
1.ANT1_Antiqua	156	93.95	94.9	NC_008150
1.ANT1_UG05-0454	163	91.51	92.34	NZ_AAYR00000000
1.IN1a_CMCC11001	48	52.46	93.59	ADQK00000000 47685
1.IN1b_780441	51	54.96	94.66	ADPS00000000 47685
1.IN1c_K21985002	62	60.58	92.21	ADSS00000000 47685
1.IN2a_CMCC640047	43	45.14	92.33	ADRA00000000 47685
1.IN2b_30017	46	89.81	94.69	ADPC00000000 47685
1.IN2c_CMCC31004	46	56.78	92.84	ADQR00000000 47685
1.IN2d_C1975003	40	49.75	92.45	ADPZ00000000 47685
1.IN2e_C1989001	50	47.25	95.07	ADQB00000000 47685
1.IN2f_710317	49	42.58	94.17	ADPP00000000 47685
1.IN2g_CMCC05013	50	24.84	93.45	ADQF00000000 47685
1.IN2h_5	49	36.98	93.36	ADPK00000000 47685
1.IN2i_CMCC10012	53	59.62	93.98	ADQG00000000 47685
1.IN2j_CMCC27002	54	46.36	92.45	ADQQ00000000 47685
1.IN2k_970754	55	61.24	93.04	ADPW00000000 47685
1.IN2l_D1991004	52	38.21	94.56	ADRX00000000 47685
1.IN2m_D1964002b	53	76.63	94.7	ADRV00000000 47685
1.IN2n_CMCC02041	56	62.95	94.48	ADQC00000000 47685
1.IN2o_CMCC03001	54	41.46	92.42	ADQD00000000 47685
1.IN2p_D1982001	50	47.49	93.11	ADRW00000000 47685
1.IN2q_D1964001	52	33.69	92.63	ADRU00000000 47685
1.IN3a_F1954001	48	60.11	95.31	ADSC00000000 47685
1.IN3b_E1979001	44	40.72	94.92	NZ_AAYV00000000
1.IN3c_CMCC84038b	45	70.76	94.72	ADRF00000000 47685
1.IN3d_YN1683	46	32.63	93.19	ADTD00000000 47685
1.IN3e_YN472	44	57.37	95.34	ADTH00000000 47685
1.IN3f_YN1065	45	60.74	94.36	ADTC00000000 47685

1.IN3g_E1977001	45	50.42	93.47	ADRY00000000 47685
1.IN3h_CMCC84033	44	38.97	94	ADRE00000000 47685
1.IN3i_CMCC84046	48	66.26	95.35	ADRG00000000 47685
1.ORI1_CA88	13	94.97	95.69	NZ_ABCD00000000
1.ORI1_CO92	0	95.06	95.74	NC_003143
1.ORI1a_CMCC114001	26	44.58	92.37	ADQL00000000 47685
1.ORI1b_India195	28	92.74	93.47	NZ_ACNR00000000
1.ORI1c_F1946001	34	42.44	94.74	ADSB00000000 47685
1.ORI2_F1991016	40	94.5	95.2	NZ_ABAT00000000
1.ORI2a_YN2179	41	40.78	92.33	ADTE00000000 47685
1.ORI2c_YN2551b	42	44.84	94.5	ADTF00000000 47685
1.ORI2d_YN2588	40	48.06	91.99	ADTG00000000 47685
1.ORI2f_CMCC87001	38	59.25	94.71	ADRH00000000 47685
1.ORI2g_F1984001	40	42.89	92.63	ADSD00000000 47685
1.ORI2h_YN663	38	56.2	92.37	ADTI00000000 47685
1.ORI2i_CMCC100001a	38	70.87	95.36	ADRR00000000 47685
1.ORI2i_CMCC110001b	39	50.54	92.56	ADRS00000000 47685
1.ORI3_IP275	42	94.47	95.21	NZ_AAOS00000000
1.ORI3_MG05-1020	34	94.73	95.36	NZ_AAYS00000000
1.ORI3a_EV76	30	50.95	91.01	ADSA00000000 47685
2.ANT1_Nepal516	138	91.87	92.7	NZ_ACNQ00000000
2.ANT1a_34008	122	61.73	92.57	ADPD00000000 47685
2.ANT1b_34202	122	53.84	93.97	ADPE00000000 47685
2.ANT2a_2	117	52.78	93.05	ADOX00000000 47685
2.ANT2b_351001	121	40.92	92.98	ADPF00000000 47685
2.ANT2c_CMCC347001	119	32.57	93.87	ADQS00000000 47685
2.ANT2d_G1996006	117	56.48	94.12	ADSE00000000 47685
2.ANT2e_G1996010	121	68.03	94.21	ADSF00000000 47685
2.ANT2f_CMCC348002	120	57.17	94.3	ADQT00000000 47685
2.ANT3a_CMCC92010	128	60.52	91.48	ADRL00000000 47685
2.ANT3b_CMCC95001	126	49.74	92.38	ADRN00000000 47685
2.ANT3c_CMCC96001	127	74.21	94.22	ADRO00000000 47685
2.ANT3d_CMCC96007	129	69.28	92.53	ADRP00000000 47685
2.ANT3e_CMCC67001	137	62.64	91.29	ADRB00000000 47685
2.ANT3f_CMCC104003	126	71.42	93.99	ADQH00000000 47685
2.ANT3g_CMCC51020	127	50.32	93.12	ADQY00000000 47685
2.ANT3h_CMCC106002	126	68.1	93.94	ADQI00000000 47685
2.ANT3i_CMCC64001	133	56.49	91.43	ADQZ00000000 47685
2.ANT3j_H1959004	124	51.89	93.69	ADSI00000000 47685
2.ANT3k_5761	131	52.31	91.38	ADPL00000000 47685
2.ANT3l_735	135	49.37	92.53	ADPR00000000 47685
2.MED1b_2506	158	70.57	93.99	ADPA00000000 47685

2.MED1c_2654	157	29	93.57	ADPB00000000 47685
2.MED1d_2504	158	37.87	92.65	ADOZ00000000 47685
2.MED2_KIM10	171	93.81	94.65	NC_004088
2.MED2b_91	131	77.47	91.46	ADPU00000000 47685
2.MED2c_K11973002	129	59.49	93.09	NZ_AAYT00000000
2.MED2d_A1973001	129	31.97	91.78	ADPY00000000 47685
2.MED2e_7338	132	62.53	90.23	ADPQ00000000 47685
2.MED3a_J1963002	134	51.58	93.19	ADSP00000000 47685
2.MED3b_CMCC125002 b	134	49.53	93.86	ADQN00000000 47685
2.MED3c_I1969003	133	73.87	93.97	ADSK00000000 47685
2.MED3d_J1978002	122	60.65	94.71	ADSQ00000000 47685
2.MED3f_I1970005	139	46.96	93.47	ADSL00000000 47685
2.MED3g_CMCC99103	136	41.71	93.66	ADRQ00000000 47685
2.MED3h_CMCC90027	136	50.58	91.2	ADRI00000000 47685
2.MED3i_CMCC92004	140	42.18	91.35	ADRK00000000 47685
2.MED3j_I2001001	137	78.49	94.82	ADSO00000000 47685
2.MED3k_CMCC12003	138	51.11	91.09	ADQM00000000 47685
2.MED3l_I1994006	141	62.9	93.9	ADSN00000000 47685
2.MED3m_SHAN11	142	51.92	92.72	ADTA00000000 47685
2.MED3n_SHAN12	142	42.92	93.19	ADTB00000000 47685
2.MED3o_I1991001	140	48.81	91.82	ADSM00000000 47685
2.MED3p_CMCC107004	141	100.25	94.78	ADQJ00000000 47685
3.ANT1a_7b	96	65.01	94.49	ADPN00000000 47685
3.ANT1b_CMCC71001	95	42	92.33	ADRC00000000 47685
3.ANT1c_C1976001	92	59.96	95.27	ADQA00000000 47685
3.ANT1d_71021	94	44.74	94.2	ADPO00000000 47685
3.ANT2a_MGJZ6	112	54.85	91.3	ADSX00000000 47685
3.ANT2b_MGJZ7	107	40.09	91.68	ADSY00000000 47685
3.ANT2c_MGJZ9	111	56.22	93.67	ADSZ00000000 47685
3.ANT2d_MGJZ11	113	53.19	94.25	ADSU00000000 47685
3.ANT2e_MGJZ3	111	63.11	93.09	ADSW00000000 47685
4.ANT1a_MGJZ12	94	55.06	93.32	ADSV00000000 47685
BD 8124_8291_11972	65	16.7	88.07	SRR341963, SRR341961, SRR341962
Altenerding	157	17.9	91.5	This study
Aschheim	120	3.94	30.24	SRP033879
Y. pseudotuberculosis (outgroup)	1238 9	88.08	91.16	NC_006155

Table S5: 157 chromosomal SNPs called for the Altenerding genome compared to the CO92 reference (5X coverage and SNP allele frequency of at least 90%)

Position	CO92	Altenerding	Coverage in Altenerding	Ancestral/Derived	SNP type
74539	C	T	31	Anc	non-synonymous
86824*	A	G	8	der	intergenic
130643	G	A	24	Anc	non-synonymous
155747*	A	G	20	Anc	synonymous
189227*	C	T	19	Anc	non-synonymous
189912*	A	G	18	der	intergenic
228268	T	G	31	Anc	synonymous
260148*	C	T	15	der	non-synonymous
271114*	C	A	19	der	non-synonymous
286528*	T	A	17	Anc	synonymous
325836	T	C	24	Anc	synonymous
341720	A	G	19	Anc	non-synonymous
399533	C	A	17	Anc	non-synonymous
420208*	G	T	7	Anc	intergenic
485976*	C	T	15	der	synonymous
545488	T	C	25	Anc	stop lost
547131	T	G	18	Anc	synonymous
549767*	T	C	10	Anc	intergenic
557841*	C	T	14	der	non-synonymous
699494	A	G	22	Anc	synonymous

699647	T	C	22	Anc	synonymous
727741	G	A	14	der	non-synonymous
779365*	C	T	17	der	synonymous
877258	T	C	26	Anc	non-synonymous
898980	A	T	11	der	intergenic
918790	C	T	13	Anc	non-synonymous
1017647	T	C	16	Anc	synonymous
1025278	T	G	20	Anc	non-synonymous
1051913	A	G	26	Anc	non-synonymous
1067966	C	A	20	der	non-synonymous
1098675*	A	C	15	Anc	non-synonymous
1102174	A	G	14	Anc	non-synonymous
1178178	T	C	9	Anc	intergenic
1178459	T	C	28	Anc	synonymous
1211729*	A	C	18	der	synonymous
1251046	T	C	10	Anc	non-synonymous
1263337	G	A	5	Anc	intergenic
1272559	T	C	16	Anc	non-synonymous
1296743*	C	T	27	der	non-synonymous
1306718	T	C	17	Anc	intergenic
1385780	T	C	23	Anc	synonymous
1387701*	C	T	13	der	intergenic
1387756*	A	G	16	der	intergenic

1413031*	C	A	8	der	intergenic
1434752*	C	A	16	der	non-synonymous
1489055	C	T	31	der	synonymous
1512930	A	G	17	Anc	non-synonymous
1530658	C	A	12	der	non-synonymous
1540754	A	G	17	Anc	synonymous
1609461*	T	C	14	der	non-synonymous
1705810	A	C	14	Anc	non-synonymous
1735263	A	C	26	Anc	synonymous
1749443	T	C	15	Anc	synonymous
1754708	C	T	22	der	non-synonymous
1804559	C	T	12	Anc	synonymous
1808946	T	C	19	Anc	intergenic
1859946*	T	C	15	Anc	non-synonymous
1868678	G	T	11	der	intergenic
1871476	G	A	10	Anc	intergenic
1956162	T	C	9	der	intergenic
2012524	T	G	19	Anc	non-synonymous
2022335	A	C	16	Anc	intergenic
2092152*	C	T	21	der	synonymous
2097520	G	T	22	der	synonymous
2098628	T	C	16	Anc	non-synonymous
2262577	T	G	31	Anc	non-synonymous

2277583	G	A	8	Anc	intergenic
2278317	A	G	20	Anc	non-synonymous
2300659	T	G	23	Anc	non-synonymous
2352174*	T	G	17	der	non-synonymous
2356003	T	A	20	Anc	non-synonymous
2419529*	G	A	25	der	synonymous
2495165*	C	A	8	der	intergenic
2508389	T	C	22	Anc	non-synonymous
2655012	C	T	13	Anc	intergenic
2656129*	T	G	20	Anc	synonymous
2684793	A	G	18	Anc	intergenic
2721828	C	A	16	Anc	synonymous
2725715	C	T	11	der	intergenic
2739149	C	A	13	Anc	synonymous
2744933	A	G	9	Anc	non-synonymous
2753572*	A	T	14	der	synonymous
2773647	A	G	18	Anc	non-synonymous
2797988*	A	G	13	Anc	synonymous
2812384	G	T	18	Anc	non-synonymous
2829833	A	G	32	Anc	non-synonymous
2903882	T	G	18	Anc	non-synonymous
2934972	C	G	22	Anc	synonymous
2936268	G	A	18	Anc	non-synonymous

2950954	G	A	11	Anc	non-synonymous
2958327	C	T	23	Anc	intergenic
2977542*	C	A	26	der	non-synonymous
2995771	A	G	8	Anc	intergenic
3078807*	C	A	21	der	non-synonymous
3085079	A	G	16	Anc	non-synonymous
3096319	G	A	17	Anc	non-synonymous
3145523	A	C	22	Anc	non-synonymous
3179828*	C	A	30	der	synonymous
3190399	A	G	35	Anc	non-synonymous
3210101	A	G	18	Anc	synonymous
3223359*	C	A	5	Anc	intergenic
3244204	A	G	26	Anc	non-synonymous
3267118*	A	G	16	Anc	non-synonymous
3274298*	A	T	8	der	synonymous
3324959	A	G	18	Anc	intergenic
3360963*	A	C	14	der	non-synonymous
3360984*	C	T	6	der	non-synonymous
3362591	A	G	18	Anc	non-synonymous
3397040*	A	G	16	Anc	synonymous
3398153	G	A	12	der	synonymous
3409414	T	C	5	der	intergenic
3421335	A	G	10	Anc	non-synonymous

3426560*	A	G	20	Anc	synonymous
3442617*	A	T	15	Anc	non-synonymous
3500922	T	G	20	der	non-synonymous
3535148*	G	T	11	der	non-synonymous
3560088	G	A	31	der	non-synonymous
3564026	C	T	24	Anc	non-synonymous
3568597	C	T	23	der	non-synonymous
3571531	A	G	30	Anc	synonymous
3616733	A	G	17	Anc	non-synonymous
3645151*	C	G	20	Anc	non-synonymous
3658233*	T	G	12	Anc	non-synonymous
3667806	A	G	13	Anc	non-synonymous
3726726	A	G	18	Anc	stop lost
3750736	G	A	16	der	intergenic
3755861	C	T	7	der	intergenic
3767613	C	T	11	Anc	intergenic
3806677	C	T	12	Anc	non-synonymous
3843195	C	A	5	der	synonymous
3892488*	C	T	21	Anc	non-synonymous
3973746	C	T	27	Anc	non-synonymous
4066494*	C	T	22	der	non-synonymous
4080579	T	C	19	Anc	non-synonymous
4081612	T	C	17	Anc	intergenic

4082562*	T	C	21	Anc	synonymous
4083536	A	G	19	Anc	intergenic
4087224	T	C	28	Anc	intergenic
4173149	A	C	6	Anc	intergenic
4194600	G	A	21	Anc	non-synonymous
4243823	A	T	9	Anc	non-synonymous
4307755	G	A	25	der	non-synonymous
4339366	T	G	21	Anc	non-synonymous
4371886*	A	G	19	Anc	synonymous
4399470	A	G	29	Anc	non-synonymous
4412624*	A	G	12	der	intergenic
4421633	T	C	36	Anc	non-synonymous
4421689	A	G	32	Anc	non-synonymous
4423366*	G	A	38	der	synonymous
4460688	C	T	18	der	non-synonymous
4465967	C	A	24	der	synonymous
4518401	G	A	18	Anc	synonymous
4527483	A	G	14	Anc	intergenic
4579183	A	G	20	Anc	non-synonymous
4628496	C	A	13	der	synonymous
4629169	G	A	33	der	intergenic
4634287	A	G	13	Anc	non-synonymous

* Chromosomal SNPs detected in the Altenerding genome that were not called in the Wagner et al. Aschheim genome.

Table S6: 11 SNPs called for plasmids pCD1 and pMT1 (5X coverage and SNP frequency of 90% minimum).

Plasmid	Position	CO92	Altenerding	Coverage
pCD1 (NC_003131.1)	23564	T	G	67
	29959	A	G	40
	50462	T	C	33
	54237	T	G	44
	55839	G	A	53
	66608	C	T	83
pMT1 (NC_003134.1)	12976*	A	C	70
	32569	T	G	82
	47365	C	T	79
	62994	T	C	90
	82435	C	T	99

* pMT1 SNP detected in the Altenerding genome and not called in the Wagner et al. Aschheim genome.

Table S7: SNPs published by Wagner et al. for the Aschheim genome that were not called in the Altenerding genome

Position in CO92	CO92	Altenerding	Coverage in Altenerding	Aschheim genome published in Wagner et al.	Coverage in the Aschheim genome published in Wagner et al.	Re-analyzed Aschheim genome	Coverage in the re-analyzed Aschheim genome	Variant frequency in the re-analyzed Aschheim genome (%)	Reason SNP was not called in the Altenerding genome
105187	A	C	17	C	19	C	14	100	Excluded non-core region
107738	A	G	14	G	13	G	10	100	Excluded non-core region
221811	A	A	24	G	63	G	7	70	Potential false positive in Wagner et al.: heterogeneous position with abnormal cover

									peak in region
225902	A	A	16	T	6	N/A	0	0	Potential false positive in Wagner et al.: position not covered in the re-analysis of Wagner et al.
333342	A	A	22	G	46	G	5	63	Potential false positive in Wagner et al.: heterogeneous position with abnormal cover peak in region
417323	G	A	3	A	30	A	18	100	Excluded non-core region
442439	T	C	19	C	46	C	33	100	Excluded non-core region
497800	T	T	17	A	122	A	13	72	False positive in Wagner et al.: heterogeneous position with abnormal cover peak in region
567757	C	A	2	A	6	A	6	100	Excluded non-core region
698477	C	A	19	A	21	A	15	100	Excluded non-core region
754287	T	C	13	C	18	C	13	100	Excluded non-core region
773110	T	C	17	C	24	C	13	100	Excluded non-core region
809132	A	G	15	G	13	G	9	100	Excluded non-core region
1044488	A	N/A	0	G	23	G	17	100	Less than 5X coverage in Altnerding
1137603	G	T	13	T	6	T	2	100	Excluded non-core region
1138676	G	T	22	T	6	T	5	83	Excluded non-core region
1237756	C	N/A	0	T	8	T	3	100	Less than 5X coverage in Altnerding

1371020	C	C	17	A	8	C	1	100	Potential false positive in Wagner et al.: SNP does not exist in the re-analysis of Wagner et al.
1371025	C	C	18	T	24	T	6	86	False positive in Wagner et al.: heterogeneous position with abnormal cover peak in region
1440494	A	C	29	C	12	C	11	100	Excluded non-core region
1444672	A	G	2	G	9	G	7	88	Less than 5X coverage in Altenerding
1796044	T	C	14	C	16	C	13	93	Excluded non-core region
1864793	A	A	16	G	162	G	7	54	Potential false positive in Wagner et al.: heterogeneous position with abnormal cover peak in region
1895361	C	A	22	A	14	A	10	100	Excluded non-core region
1914093	T	C	19	C	14	C	8	100	Excluded non-core region
1982740	A	C	11	C	9	C	8	89	Excluded non-core region
2072914	G	G	35	A	13	A	4	80	Potential false positive in Wagner et al.: heterogeneous position with abnormal cover peak in region
2117516	G	A	12	A	22	A	15	100	Excluded non-core region
2119347	T	T	45	A	7	N/A	0	0	Potential false positive in Wagner et al.: position not covered in the re-

									analysis of Wagner et al.
2141322	T	C	27	C	5	C	3	60	Excluded non- core region
2141910	C	A	18	A	10	A	6	100	Excluded non- core region
2218046	G	T	17	T	5	T	1	100	Excluded non- core region
2235109	T	C	1	C	12	C	11	100	Less than 5X coverage in Altenerding
2281856	A	N/A	0	C	10	C	7	100	Less than 5X coverage in Altenerding
2304950	A	G	11	G	27	G	14	100	Excluded non- core region
2317730	A	C	10	C	18	C	9	100	Excluded non- core region
2453454	A	G	14	G	11	G	7	100	Excluded non- core region
2548551	G	T	20	T	19	T	12	100	Excluded non- core region
2575152	G	A	4	A	33	A	14	100	Less than 5X coverage in Altenerding
2577686	A	G	28	G	24	G	12	100	Excluded non- core region
2607034	C	T	23	T	18	T	15	100	Excluded non- core region
2619611	T	G	15	G	9	G	5	100	Excluded non- core region
2787770	T	G	25	G	13	G	2	100	Excluded non- core region
2865494	A	A	30	C	9	C	3	100	Potential false positive in Wagner et al.: less than 5X coverage in the re-analysis of Wagner et al.
2894703	T	N/A	0	C	16	C	10	100	Excluded non- core region
2896636	A	G	10	G	41	G	21	88	Excluded non- core region
3143800	G	G	21	T	29	G	1	100	Potential false positive in

									Wagner et al.: SNP does not exist in the re- analysis of Wagner et al.
3155055	G	G	26	C	54	C	5	71	Potential false positive in Wagner et al.: heterogeneous position with abnormal cover peak in region
3248223	T	C	9	C	10	C	7	100	Excluded non- core region
3358603	G	T	9	T	30	T	23	100	Excluded non- core region
3392897	A	A	22	G	79	G	11	85	Potential false positive in Wagner et al.: heterogeneous position with abnormal cover peak in region
3403167	G	T	6	T	16	T	10	100	Excluded non- core region
3472427	G	A	17	A	18	A	16	100	Excluded non- core region
3725545	T	C	14	C	19	C	12	100	Excluded non- core region
3732919	A	G	15	G	20	G	13	100	Excluded non- core region
3737968	G	A	13	A	19	A	11	100	Excluded non- core region
3739401	C	A	18	A	9	A	6	100	Excluded non- core region
3813424	C	C	10	A	23	A	4	100	Potential false positive in Wagner et al.: less than 5X coverage in the re-analysis of Wagner et al.; abnormal cover peak in region
3909258	T	C	8	C	10	C	10	100	Excluded non- core region

4170791	A	A	25	G	87	G	20	80	Potential false positive in Wagner et al.: heterogeneous position with abnormal cover peak in region
4199187	A	A	47	G	70	A	3	75	Potential false positive in Wagner et al.: heterogeneous position with abnormal cover peak in region
4199190	T	T	27	C	123	C	3	38	Potential false positive in Wagner et al.: heterogeneous position with abnormal cover peak in region
4203596	G	G	22	T	51	G	3	100	Potential false positive in Wagner et al.: SNP does not exist in the re-analysis of Wagner et al.
4210011	T	T	34	C	262	T	8	100	Potential false positive in Wagner et al.: SNP does not exist in the re-analysis of Wagner et al.
4232217	C	N/A	0	T	5	T	5	100	Less than 5X coverage in Altenerding
4542642	A	G	8	G	9	G	7	100	Excluded non-core region

Table S8: 19 potential false positive SNPs called by Wagner et al.

Position in CO92	SNP type defined in Wagner et al.	Gene
221811	Derived, non-synonymous	<i>rpsJ</i>
225902	Ancestral	
333342	Derived, synonymous	<i>uvrA</i>
497800	Derived, synonymous	<i>dnaK</i>
1371020	Ancestral	
1371025	Derived, synonymous	<i>gyrA</i>
1864793	Derived, synonymous	<i>mnmA</i>
2072914	Derived, synonymous	<i>fliI</i>
2119347	Ancestral	
2865494	Ancestral	
3143800	Derived, non-synonymous	<i>ynbD</i>
3155055	Derived, synonymous	<i>purI</i>
3392897	Derived, synonymous	<i>napA</i>
3813424	Derived, synonymous	<i>acnB</i>
4170791	Derived, synonymous	<i>aceA</i>
4199187	Derived, synonymous	<i>rpoC</i>
4199190	Derived, synonymous	<i>rpoC</i>
4203596	Derived, synonymous	<i>rpoB</i>
4210011	Ancestral	

Table S9: 63 Unique SNPs in the Altenerding genome cross-referenced against all *Y. pestis* genomes in the data set (excluding the Aschheim genome)

Position in CO92	SNP type	Ancestral AA	Derived AA	Gene ID	Gene name
Chromosom					
86824*	intergenic	N/A	N/A		
189912*	intergenic	N/A	N/A		
260148*	non-synonymous	P	S	YPO0257	
271114*	non-synonymous	L	I	YPO0270	
485976*	synonymous	A	A	YPO0460	<i>thrB</i>
557841*	non-synonymous	R	H	YPO0517	<i>hepA</i>
727741	non-synonymous	E	K	YPO0668	<i>parE</i>
779365*	synonymous	S	S	YPO0717	<i>flil</i>
898980	intergenic	N/A	N/A		
1067966	non-synonymous	G	C	YPO0966	
1211729*	synonymous	V	V	YPO1068	<i>proS</i>
1296743*	non-synonymous	V	M	YPO1150	<i>bioA</i>
1387701*	intergenic	N/A	N/A		
1387756*	intergenic	N/A	N/A		
1413031*	intergenic	N/A	N/A		
1434752*	non-synonymous	D	Y	YPO1275	<i>spr</i>
1489055	synonymous	Q	Q	YPO1322	<i>deoR</i>
1530658	non-synonymous	R	I	YPO1363	
1609461*	non-synonymous	T	A	YPO1417	
1754708	non-synonymous	P	L	YPO1539	<i>galU</i>
1868678	intergenic	N/A	N/A		
1956162	intergenic	N/A	N/A		
2092152*	synonymous	G	G	YPO1847	<i>yecS</i>
2097520	synonymous	A	A	YPO1851	<i>putA</i>
2352174*	non-synonymous	V	G	YPO2071	
2419529*	synonymous	L	L	YPO2150	
2495165*	intergenic	N/A	N/A		

2725715	intergenic	N/A	N/A		
2753572*	synonymous	P	P	YPO2455	
2977542*	non-synonymous	S	I	YPO2649	<i>nrdE</i>
3078807*	non-synonymous	R	S	YPO2747	<i>fadJ</i>
3179828*	synonymous	S	S	YPO2847	<i>yegM</i>
3274298*	synonymous	A	A	YPO2930	<i>pdxJ</i>
3360963*	non-synonymous	T	P	YPO3008	
3360984*	non-synonymous	H	Y	YPO3008	
3398153	synonymous	L	L	YPO3043	
3409414	intergenic	N/A	N/A		
3500922	non-synonymous	V	G	YPO3141	<i>tesB</i>
3535148*	non-synonymous	A	S	YPO3171	<i>apbA</i>
3560088	non-synonymous	P	S	YPO3199	
3568597	non-synonymous	G	E	YPO3205	<i>phoB</i>
3750736	intergenic	N/A	N/A		
3755861	intergenic	N/A	N/A		
3843195*	synonymous	V	V	YPO3438	<i>intB</i>
3892488*	non-synonymous	A	T	YPO3483	
4066494*	non-synonymous	V	I	YPO3646	<i>pcp</i>
4307755	non-synonymous	A	V	YPO3839	
4412624*	intergenic	N/A	N/A		
4423366*	synonymous	A	A	YPO3938	<i>glgP</i>
4460688	non-synonymous	R	Q	YPO3963	
4465967	synonymous	S	S	YPO3966	
4628496	synonymous	A	A	YPO4107	<i>yieG</i>
4629169	intergenic	N/A	N/A		
pCD1 plasmid					
23564	synonymous	A	A	YPCD1.33c	<i>lcrR</i>
29959	non-synonymous	N	S	YPCD1.41	<i>yscO</i>
50462	non-synonymous	K	E	YPCD1.71c	<i>yopJ</i>

54237	intergenic	N/A	N/A		
55839	intergenic	N/A	N/A		
66608	non-synonymous	L	F	YPCD1.92	
pMT1 plasmid					
32569	synonymous	T	T	YPMT1.30	
47365	synonymous	R	R	YPMT1.44	
62994	intergenic	N/A	N/A		
82435	intergenic	N/A	N/A		

*30 unique SNPs detected in the Altenerding genome that were not called in *Wagner et al.*

Table S10: Non-synonymous SNPs compared to CO92

Position in CO92	CO92	Altenerding	Ancestral /Derived	Codon Change	Amino Acid Change	Gene ID	Gene name	Gene function
Chromosome								
74539	C	T	Anc	GCC to ACC	A to T	YPO0063		
130643	G	A	Anc	ACC to ATC	T to I	YPO0122	<i>glpE</i>	thiosulfate sulfurtransferase
189227*	C	T	Anc	CGT to TGT	R to C	YPO0169	<i>pabA</i>	para-aminobenzoate synthase component II
260148*	C	T	der	CCG to TCG	P to S	YPO0257		type III secretion protein
271114*	C	A	der	CTT to ATT	L to I	YPO0270		type III secretion system protein (iron-sulfur binding protein)
341720	A	G	Anc	CAC to CGC	H to R	YPO0332	<i>rhaS</i>	transcriptional activator
399533	C	A	Anc	TTC to TTA	F to L	YPO0383	<i>aidB</i>	isovaleryl CoA dehydrogenase
557841*	C	T	der	CGT to CAT	R to H	YPO0517	<i>hepA</i> (<i>RapA</i>)	ATP-dependent helicase (transcription regulator)
727741	G	A	der	GAA to AAA	E to K	YPO0668	<i>parE</i>	DNA to poimerase IV subunit B
877258	T	C	Anc	TTT to CTT	F to L	YPO0797	<i>lysR</i>	LysR family transcriptional regulator
918790	C	T	Anc	CCA to CTA	P to L	YPO0839	<i>kduD2</i>	2-deoxy-D-gluconate 3- dehydrogenase
1025278	T	G	Anc	TCT to GCT	S to A	YPO0932		
1051913	A	G	Anc	AGG to GGG	R to G	YPO0953		
1067966	C	A	der	GGC to TGC	G to C	YPO0966		kinase
1098675*	A	C	Anc	AGC to CGC	S to R	YPO0989	<i>iucA</i>	pseudo gene

1102174	A	G	Anc	ACA to GCA	T to A	YPO0993	<i>iucD</i>	siderophore biosynthesis protein
1251046	T	C	Anc	ATC to ACC	I to T	YPO1107	<i>GrpE</i>	heat shock protein
1272559	T	C	Anc	GTT to GCT	V to A	YPO1126		I-pal system protein YbgF
1296743*	C	T	der	GTG to ATG	V to M	YPO1150	<i>bioA</i>	adenosylmethionine-8-amino-7-oxononanoate aminotransferase (part of the Biotin operon)
1434752*	C	A	der	GAT to TAT	D to Y	YPO1275	<i>spr</i>	outer membrane lipoprotein (Murein hydrolase)
1512930	A	G	Anc	ACG to GCG	T to A	YPO1348		
1530658	C	A	der	AGA to ATA	R to I	YPO1363		virulence factor
1609461*	T	C	der	ACA to GCA	T to A	YPO1417		iron-sulfur binding protein
1705810	A	C	Anc	GTG to GGG	V to G	YPO1502		alcohol dehydrogenase
1754708	C	T	der	CCA to CTA	P to L	YPO1539	<i>galU</i>	UTP-glucose-1-phosphate uridylyltransferase
1859946*	T	C	Anc	AGT to GGT	S to G	YPO1634	<i>phoP</i>	DNA-binding transcriptional regulator
2012524	T	G	Anc	TTG to GTG	L to V	YPO1767	<i>hpaI (hpcH)</i>	2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase
2098628	T	C	Anc	AAT to AGT	N to S	YPO1851	<i>putA</i>	trifunctional transcriptional regulator/proline dehydrogenase/pyrroline-5-carboxylate dehydrogenase
2262577	T	G	Anc	CTG to CGG	L to R	YPO1990		
2278317	A	G	Anc	GTT to GCT	V to A	YPO2005		
2300659	T	G	Anc	GAC to GCC	D to A	YPO2029		
2352174*	T	G	der	GTG to GGG	V to G	YPO2071		DEAD box family helicase
2356003	T	A	Anc	AAT to AAA	N to K	YPO2074	<i>fadD</i>	long chain fatty acid CoA ligase
2508389	T	C	Anc	ACA to GCA	T to A	YPO2234	<i>cstA</i>	carbon starvation protein A
2744933	A	G	Anc	ATT to GTT	I to V	YPO2446		2-deoxyglucose-6-phosphatase
2773647	A	G	Anc	GTC to GCC	V to A	YPO2472		
2812384	G	T	Anc	CCG to CAG	P to Q	YPO2502	<i>gutB</i>	zinc-binding dehydrogenase
2829833	A	G	Anc	TGC to CGC	C to R	YPO2519		SAM-dependent methyltransferase
2903882	T	G	Anc	TGT to GGT	C to G	YPO2582		sugar transport ATP-binding protein
2936268	G	A	Anc	CTT to TTT	L to F	YPO2614	<i>gltJ</i>	glutamate/Faspartate transport system permease

2950954	G	A	Anc	GCG to GTG	A to V	YPO2625	<i>nagC</i>	N-acetylglucosamine regulatory protein
2977542*	C	A	der	AGC to ATC	S to I	YPO2649	<i>nrdE</i>	ribonucleotide-diphosphate reductase subunit alpha
3078807*	C	A	der	CGT to AGT	R to S	YPO2747	<i>fadJ</i> (<i>faoA</i>)	multifunctional fatty acid oxidation complex subunit alpha
3085079	A	G	Anc	ACT to GCT	T to A	YPO2752	<i>mepA</i>	penicillin-insensitive murein endopeptidase
3096319	G	A	Anc	GGA to GAA	G to E	YPO2762		AraC family transcriptional regulator
3145523	A	C	Anc	CTC to CGC	L to R	YPO2814	<i>ynbD</i>	
3190399	A	G	Anc	CAG to CGG	Q to R	YPO2853	<i>baeR</i>	DNA-binding transcriptional regulator
3244204	A	G	Anc	GTC to GCC	V to A	YPO2901		
3267118*	A	G	Anc	CTG to CCG	L to P	YPO2921	<i>purL</i>	phosphoribosylformylglycine midine synthase
3360963*	A	C	der	ACC to CCC	T to P	YPO3008		two-component sensor histidine kinase (TCSs)
3360984*	C	T	der	CAT to TAT	H to Y	YPO3008		two-component sensor histidine kinase (TCSs)
3362591	A	G	Anc	AGC to GGC	S to G	YPO3009		two-component response regulator
3421335	A	G	Anc	ATG to GTG	M to V	YPO3064	<i>bcp</i>	thioredoxin-dependent thiol peroxidase
3442617*	A	T	Anc	AGA to AGT	R to S	YPO3086	<i>copA</i>	copper exporting ATPase
3500922	T	G	der	GTG to GGG	V to G	YPO3141	<i>tesB</i>	acyl-CoA thioesterase
3535148*	G	T	der	GCT to TCT	A to S	YPO3171	<i>apbA</i> (<i>panE</i>)	2-dehydropantoate 2-reductase (vitamin B5 biosynthesis)
3560088	G	A	der	CCA to TCA	P to S	YPO3199		short chain dehydrogenase
3564026	C	T	Anc	TGT to TAT	C to Y	YPO3201	<i>proY</i>	permease
3568597	C	T	der	GGA to GAA	G to E	YPO3205	<i>phoB</i>	phosphate regulon transcriptional regulator
3616733	A	G	Anc	CTA to CCA	L to P	YPO3247	<i>hmwA</i>	adhesin
3645151*	C	G	Anc	GCC to GGC	A to G	YPO3272	<i>yfiQ</i>	acetyltransferase
3658233*	T	G	Anc	AGA to AGC	R to S	YPO3277	<i>rluD</i>	23S rRNA pseudouridine synthase D
3667806	A	G	Anc	AAA to GAA	K to E	YPO3287	<i>yehT</i>	two-component response-regulatory protein

3806677	C	T	Anc	AGG to AAG	R to K	YPO3408	<i>hpt</i>	hypoxanthine phosphoribosyltransferase
3892488*	C	T	Anc	GCT to ACT	A to T	YPO3483		multidrug efflux protein
3973746	C	T	Anc	ACT to ATT	T to I	YPO3559		
4066494*	C	T	der	GTT to ATT	V to I	YPO3646	<i>pcp</i> (<i>pcpY</i> , <i>slyB</i>)	outer membrane lipoprotein
4080579	T	C	Anc	CAC to CGC	H to R	YPO3659	<i>accB</i>	acetyl-CoA carboxylase biotin carboxyl carrier protein subunit
4194600	G	A	Anc	GTT to ATT	V to I	YPO3742	<i>thiG</i>	thiazole synthase
4243823	A	T	Anc	TTT to TAT	F to Y	YPO3781	<i>ubiE</i>	ubiquinone,menaquinone biosynthesis methyltransferase
4307755	G	A	der	GCG to GTG	A to V	YPO3839		
4339366	T	G	Anc	AGT to CGT	S to R	YPO3865	<i>wzzE</i>	lipopolysaccharide biosynthesis protein
4399470	A	G	Anc	AGC to GGC	S to G	YPO3917		dihydrolipoamide dehydrogenase
4421633	T	C	Anc	GTG to GCG	V to A	YPO3937	<i>glpD</i>	glycerol-3-phosphate dehydrogenase
4421689	A	G	Anc	ACA to GCA	T to A	YPO3937	<i>glpD</i>	glycerol-3-phosphate dehydrogenase
4460688	C	T	der	CGG to CAG	R to Q	YPO3963		sugar transport system permease
4579183	A	G	Anc	AGC to GGC	S to G	YPO4060	<i>fdhD</i>	formate dehydrogenase accessory protein
4634287	A	G	Anc	TCG to CCG	S to P	YPO4113	<i>phoU</i>	transcriptional regulator
pCD1								
29959	A	G	der	AAC to AGC	N to S	YPCD1.41	<i>yscO</i>	type III secretion apparatus component
50462	T	C	der	AAA to GAA	K to E	YPCD1.71 c	<i>yopJ</i>	targeted effector protein
66608	C	T	der	CTT to TTT	L to F	YPCD1.92		
pMT1								
12976*	A	C	anc	TAT to GAT	Y to D	YPMT1.09 c		minor tail fiber protein L

*SNPs detected in the Altenerding genome that were not called in the *Wagner et al.* Aschheim genome.

Table S11: Regions missing in CO92 but present in the Altenerding genome. Positions and annotated genes refer to the *Yersinia pseudotuberculosis* reference genome (GenBank accession NC_006155)

Start Position	End Position	Region Length (bp)	Mean Coverage	ID of genes in region	Name of genes in region	Function
97811	98932	1122	17.37	YPTB0085	<i>glpX</i>	type II fructose 1,6-bisphosphatase (part of the glpFKX operon that enables aerobic glycerol fermentation).
				YPTB0086	<i>glpK</i>	glycerol kinase (part of the glpFKX operon that enables aerobic glycerol fermentation).
263822	264118	297	7.34	YPTB0221	<i>ftsY</i>	cell division protein
369602	369881	280	12.98	YPTB0305	<i>ytaP</i>	
408757	409016	260	3.2	YPTB0343		
				YPTB0344		coproporphyrinogen III oxidase
799654	799932	279	5.24	YPTB0668	<i>setA (yadM)</i>	major facilitator superfamily transporter sugar efflux pump
807313	807587	275	5.88	Intergenic		
994244	994536	293	2.26	Intergenic		
1433136 (DFR4)	1448749	15614	20.85	YPTB1202	<i>xapB</i>	major facilitator superfamily xanthosine permease
				YPTB1203	<i>zraP</i>	zinc resistance protein
				YPTB1204		two component Histidine kinase sensor
				YPTB1205	<i>hydG</i>	transcriptional regulator
				YPTB1206	<i>morB</i>	morphinone reductase
				YPTB1207		LysR family transcriptional regulator
				YPTB1208		
				YPTB1209		multidrug ABC transporter permease
				YPTB1210		multidrug ABC transporter permease
				YPTB1211		ABC transporter ATPase
				YPTB1212		
				YPTB1213		DNA binding transcriptional regulator
				YPTB1214	<i>rhIE</i>	ATP dependent RNA helicase

1753403	1753691	289	12.94	YPTB1458	<i>helD</i>	DNA helicase IV
2222016	2222305	290	6.05	YPTB1880		
2603581	2604597	1017	13.44	YPTB2210	<i>tauB</i>	taurine/sulfonate transporter (part of the tauABC operon)
				YPTB2211	<i>amn</i>	AMP nucleosidase
3877157	3879821	2665	18.66	YPTB3285		Va autotransporter
				YPTB3286		Va autotransporter
3975571	3975841	271	10.76	YPTB3344		
				YPTB3345	<i>flil</i>	flagellum specific ATP synthase
4373203	4373468	266	12.02	YPTB3659		transferase
4490444	4490698	255	15.12	YPTB3782	<i>glpD</i>	glycerol 3-phosphate dehydrogenase
4510659	4510920	262	22.16	YPTB3789		Ig-like domain containing protein

Supplementary figures

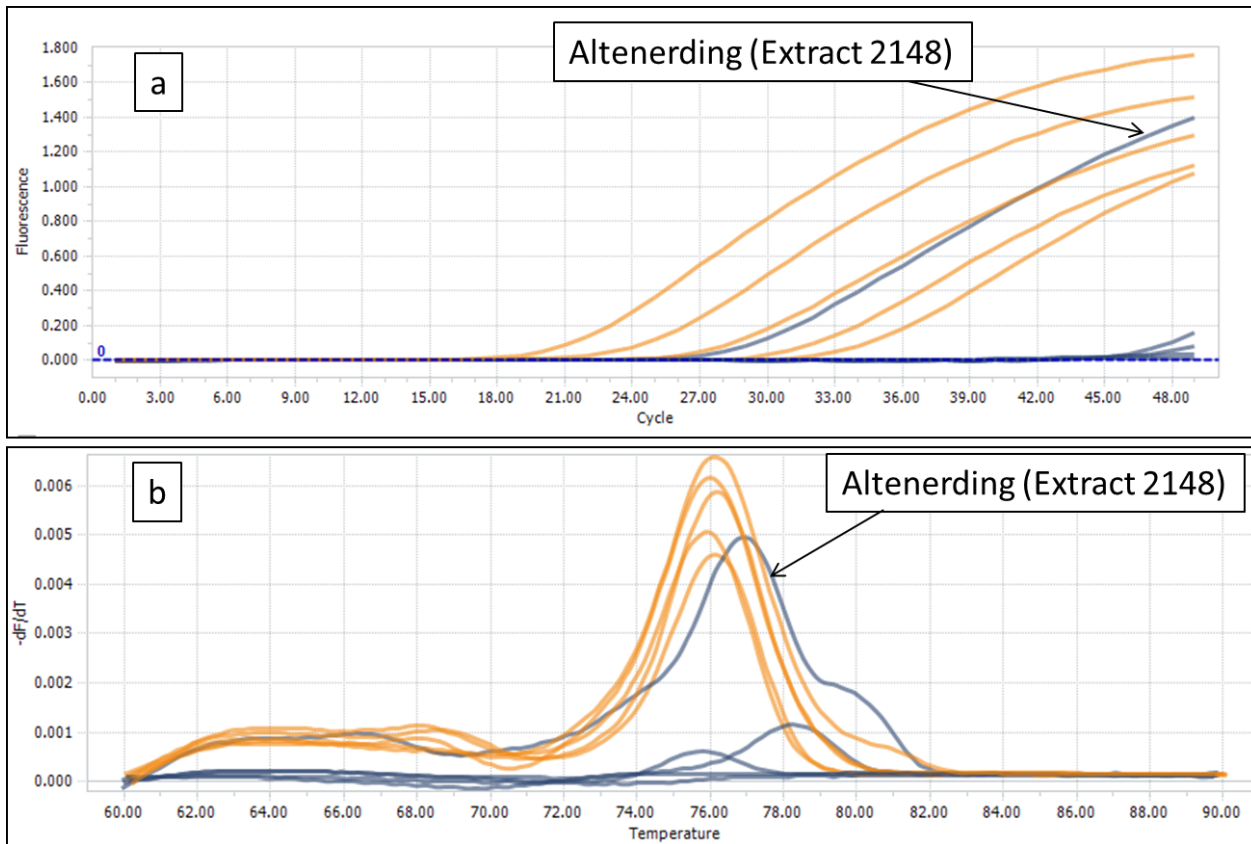


Fig.S1: *p/a* gene qPCR amplification and melting curves used for sample screening. The standards of known copy number (22,300, 2230, 223, 22.3 and 2.23 copies/uL) are in duplicates, shown in yellow and grey. (a) Amplification curves. (b) Melting peaks.

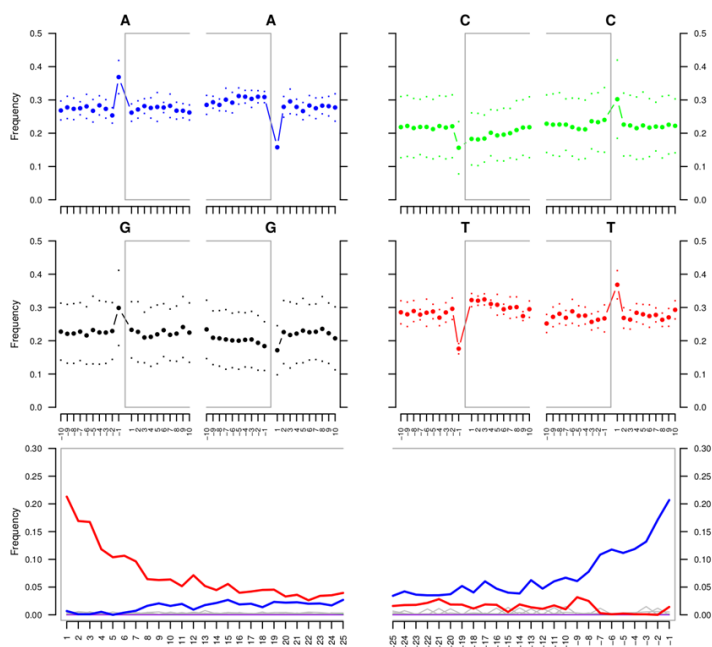


Fig.S2: mapDamage fragment misincorporation plot based on the non-UDG treated Altenerding library mapped to the CO92 chromosome reference (5510 fragments).

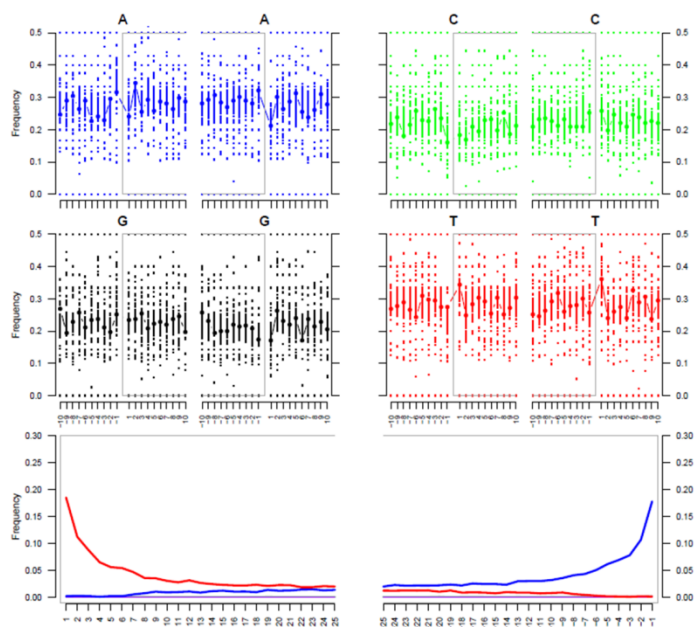


Fig.S3: mapDamage fragment misincorporation plot based on the non-UDG treated Altenerding library mapped to the human hg19 reference (45,359 fragments).

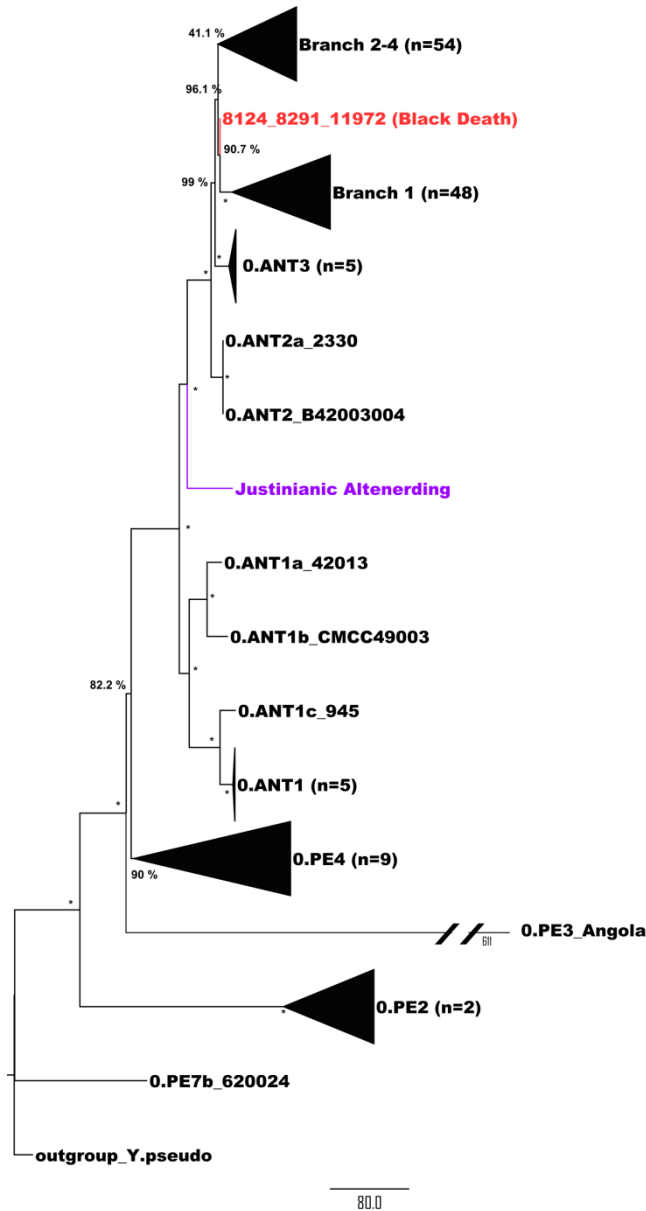


Fig.S4: Maximum Parsimony tree excluding the Aschheim genome. Maximum Parsimony analysis of 2603 nucleotide positions from genomes of 132 *Y. pestis* strains (the Aschheim genome was excluded from this analysis). All positions containing missing data were eliminated. Bootstrap values are next to nodes and bootstrap values of 100 % are indicated by an asterisk. The tree is rooted using the genome of *Y. pseudotuberculosis* (strain IP32953). Branches leading to isolates from the historical pandemics are colored red and purple representing the second and first pandemics respectively. Number of isolates in a collapsed node is indicated in brackets.

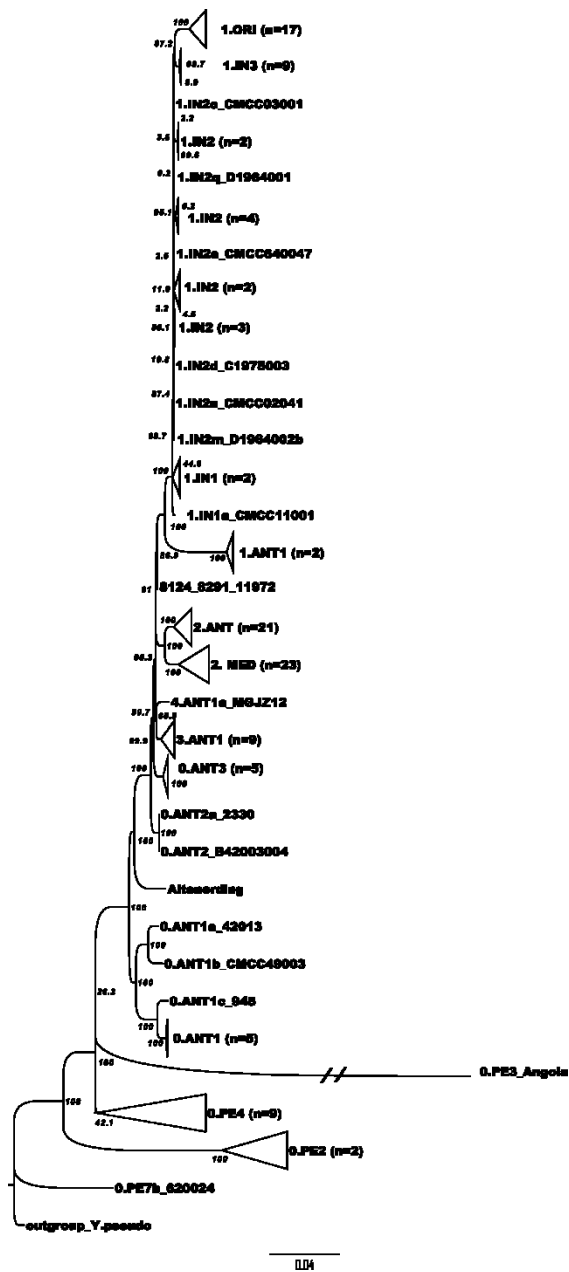


Fig.S5: Maximum Likelihood tree excluding the Aschheim genome. Maximum Likelihood analysis of 2603 nucleotide positions from genomes of 132 *Y. pestis* strains (the Aschheim genome was excluded from this analysis). All positions containing gaps and missing data were eliminated. Bootstrap values in italics. The tree is rooted using the genome of *Y. pseudotuberculosis* (strain IP32953). Number of isolates in a collapsed node is indicated in brackets.

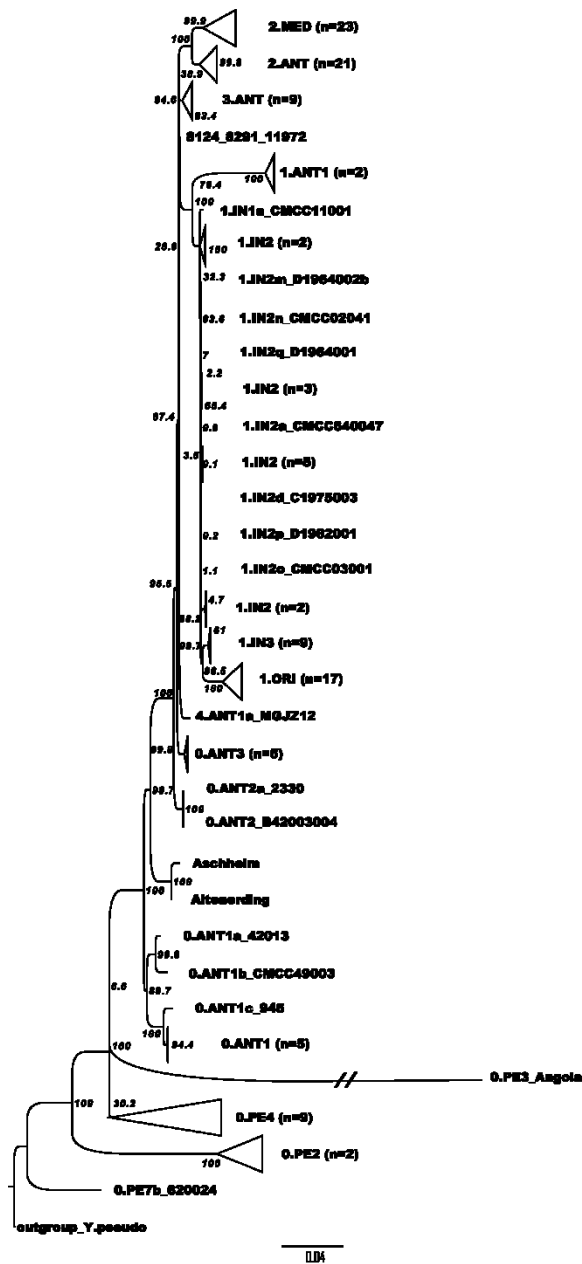
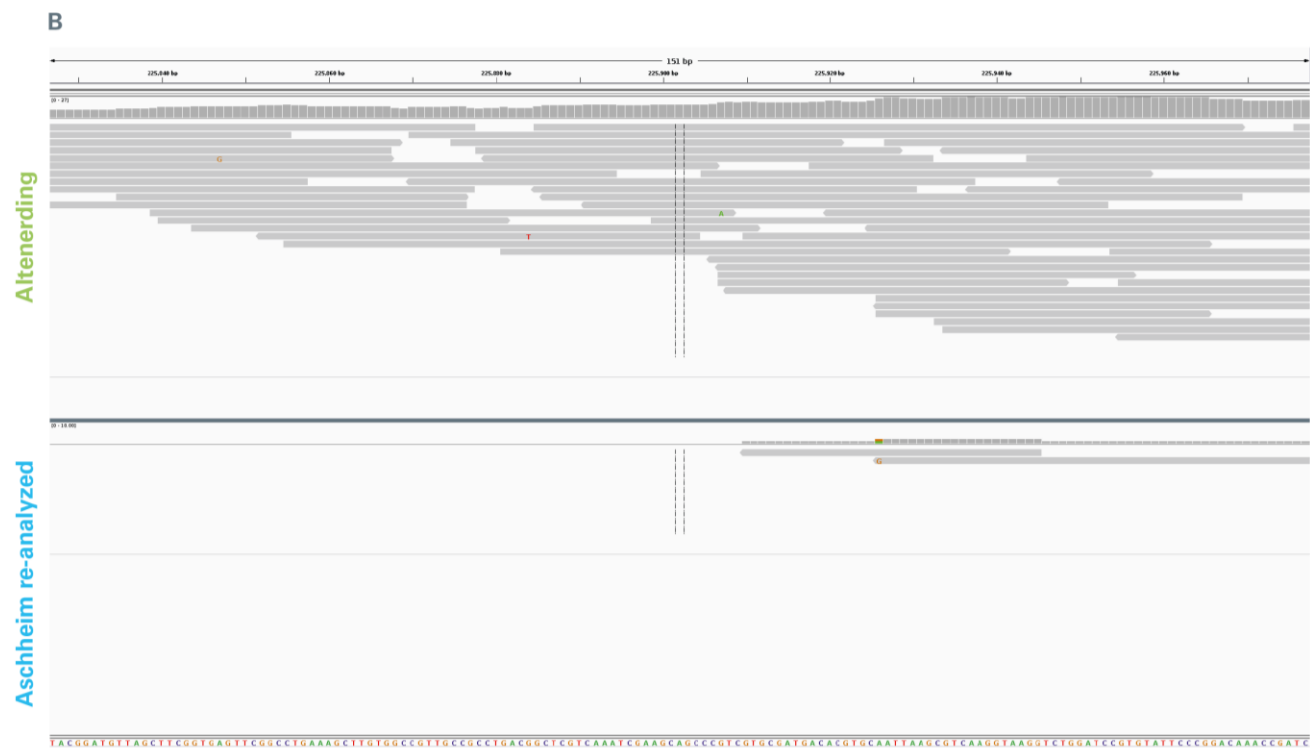
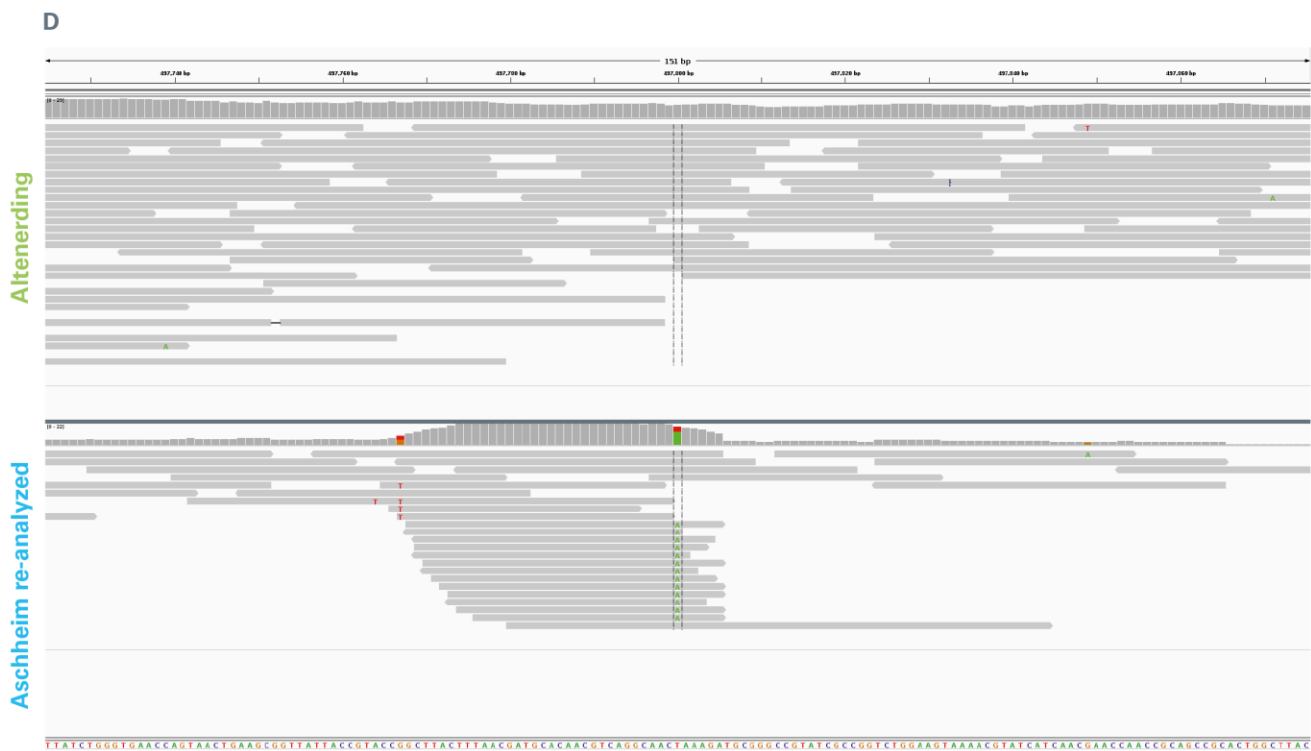
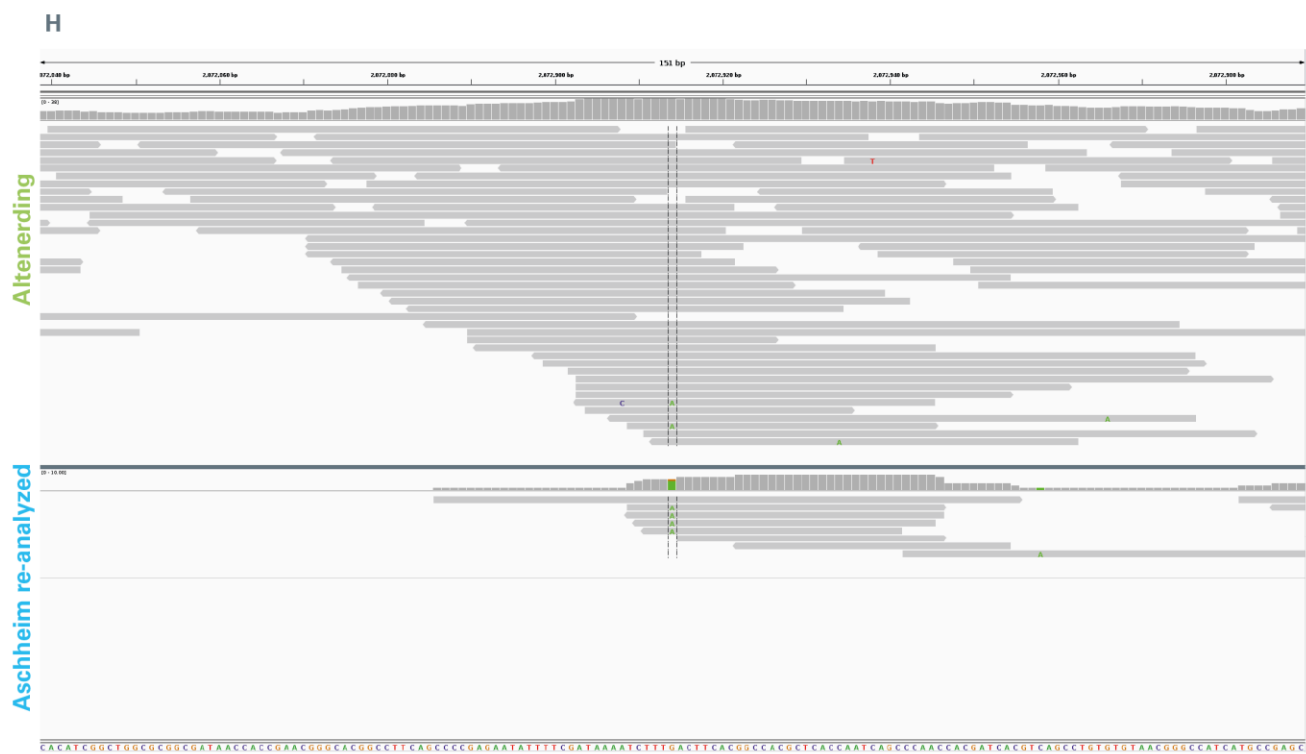


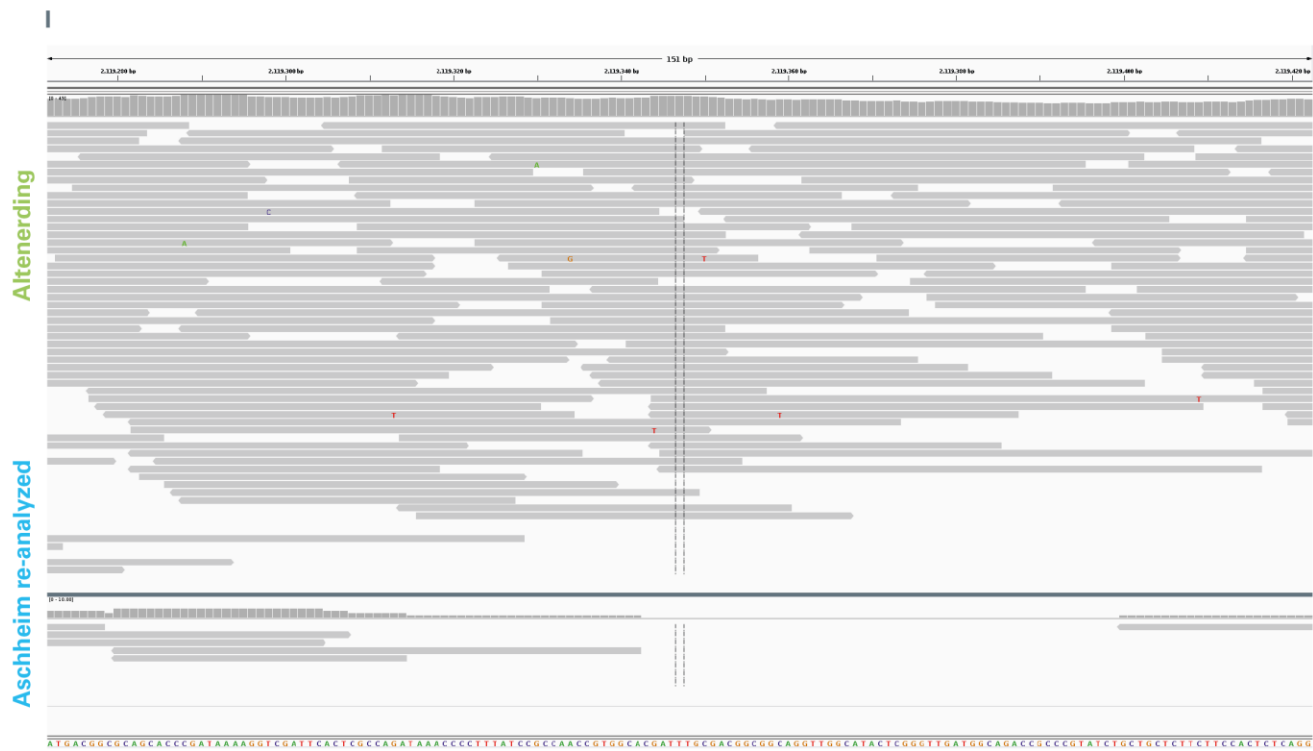
Fig.S6: Maximum Likelihood tree. Maximum Likelihood analysis of 1418 nucleotide positions from genomes of 133 *Y. pestis* strains. All positions containing gaps and missing data were eliminated. Bootstrap values in italics. The tree is rooted using the genome of *Y. pseudotuberculosis* (strain IP32953). Number of isolates in a collapsed node is indicated in brackets.















0



P







Fig.S7 (A-S): Visualization of positions containing potential false positive SNPs called by *Wagner et al. 2014*. Reads were mapped to the CO92 reference with sensitivity of 0.1 and minimum mapping quality of 30 and visualized on IGV gene browser. Upper bend shows coverage plot for the region corresponding to the genome beneath. Upper scale shows position (bp) in reference sequence. Bottom sequence shows the 150 bp in the CO92 reference. Dotted line marks the potential false positive SNP.

(A) Position 221811: Re-analysis of the Aschheim raw data shows 70 % variant frequency for the “A” to “G” variant called as SNP by Wagner et al. Variant frequency is lower than the minimum variant frequency of 95 % set by Wagner et al. The “G” variants are located at end of reads in a region with an abnormal cover peak. The Altenerding mapping shows even 24 fold coverage in the region and a 96 % variant frequency that supports the “A” variant (identical to reference).

(B) Position 225902: Re-analysis of the Aschheim raw data shows no coverage in position 225902 called as an “A” to “T” SNP by Wagner et al., in contrast to a 6 fold coverage in the original analysis (Wagner et al. 2014). The Altenerding mapping shows even 16 fold coverage in the region and a 100 % variant frequency that supports the “A” variant (identical to reference).

(C) Position 333342: Re-analysis of the Aschheim raw data shows 63 % variant frequency for the “A” to “G” variant called as SNP by Wagner et al. Variant frequency is lower than the minimum variant frequency of 95 % set by Wagner et al. The “G” variants are located at end of reads in a region with an abnormal peak in coverage. The Altenerding mapping shows even 22 fold coverage in the region and a 100 % variant frequency that supports the “A” variant (identical to reference).

(D) Position 497800: Re-analysis of the Aschheim raw data shows 72 % variant frequency for the “T” to “A” variant called as SNP by Wagner et al. Variant frequency is lower than the minimum variant frequency of 95 % set by Wagner et al. The “A” variants are located at end of reads in a region with an abnormal peak in coverage. The Altenerding mapping shows even 17 fold coverage in the region and a 100 % variant frequency that supports the “T” variant (identical to reference).

(E) Position 137020: Re-analysis of the Aschheim raw data shows 1 fold coverage in position 137020 called as a “C” to “A” SNP by Wagner et al., in contrast with a 8 fold coverage in the original analysis (Wagner et al. 2014). The one read covering the position shows the “C” variant (identical to reference). The Altenerding mapping shows even 17 fold coverage in the region and a 100 % variant frequency that supports the “C” variant (identical to reference).

(F) Position 137025: Re-analysis of the Aschheim raw data shows 86 % variant frequency for the “C” to “T” variant called as SNP by Wagner et al. Variant frequency is lower than the minimum variant frequency of 95 % set by Wagner et al. The “T” variants are located at end of reads in a region with an abnormal peak in coverage. The Altenerding mapping shows even 18 fold coverage in the region and a 100 % variant frequency that supports the “C” variant (identical to reference).

(G) Position 1864793: Re-analysis of the Aschheim raw data shows 54 % variant frequency for the “A” to “G” variant called as SNP by Wagner et al. Variant frequency is lower than the minimum variant frequency of 95 % set by Wagner et al. The reads containing the “G” variants are located in a region with an abnormal peak in coverage. The Altenerding mapping shows even 16 fold coverage in the region and a 100 % variant frequency that supports the “A” variant (identical to reference).

(H) Position 2072914: Re-analysis of the Aschheim raw data shows 80 % variant frequency for the “G” to “A” variant called as SNP by Wagner et al. Variant frequency is lower than the minimum variant frequency of 95 % set by Wagner et al. The “A” variants are located at end of reads in a region with an abnormal peak in coverage. The Altenerding mapping shows 37 fold coverage in the region and a 95% variant frequency that supports the “G” variant (identical to reference).

(I) Position 2119347: Re-analysis of the Aschheim raw data shows no coverage in position 2119347 called as a “T” to “A” SNP by Wagner et al., in contrast with a 7 fold coverage in the original analysis (Wagner et al. 2014). The Altenerding mapping shows even 45 fold coverage in the region and a 100 % variant frequency that supports the “T” variant (identical to reference).

(J) Position 2865494: Re-analysis of the Aschheim raw data shows 3 fold coverage in position 2865494 called as an “A” to “C” SNP by Wagner et al., in contrast with a 9 fold coverage in the original analysis (Wagner et al. 2014). Coverage is lower than the 5 fold minimum set by Wagner et al. The Altenerding mapping shows even 30 fold coverage in the region and a 100 % variant frequency that supports the “A” variant (identical to reference).

(K) Position 3143800: Re-analysis of the Aschheim raw data shows 1 fold coverage in position 3143800 called as a “G” to “T” SNP by Wagner et al., in contrast with a 29 fold coverage in the original analysis (Wagner et al. 2014). The one read covering the position shows the “G” variant (identical to reference). The Altenerding mapping shows even 21 fold coverage in the region and a 100 % variant frequency that supports the “G” variant (identical to reference).

(L) Position 3155055: Re-analysis of the Aschheim raw data shows 71 % variant frequency for the “G” to “C” variant called as SNP by Wagner et al. Variant frequency is lower than the minimum variant frequency of 95 % set by Wagner et al. The “C” variants are located at end of reads in a region with an abnormal peak in coverage. The Altenerding mapping shows even 27 fold coverage in the region and a 100 % variant frequency that supports the “G” variant (identical to reference).

(M) Position 3392897: Re-analysis of the Aschheim raw data shows 85 % variant frequency for the A to G variant called as SNP by Wagner et al. Variant frequency is lower than the minimum variant frequency of 95 % set by Wagner et al. The reads containing the “G” variants are located

in a region with an abnormal peak in coverage. The Altenerding mapping shows even 22 fold coverage in the region and a 100 % variant frequency that supports the “A” variant (identical to reference).

(N) Position 3813424: Re-analysis of the Aschheim raw data shows 4 fold coverage in position 3813424 called as a “C” to “A” SNP by Wagner et al., in contrast with a 23 fold coverage in the original analysis (Wagner et al. 2014). Coverage is lower than the 5 fold minimum set by Wagner et al. The A variants are located at end of reads in a region with an abnormal peak in coverage. The Altenerding mapping shows 10 fold coverage in the region and a 100 % variant frequency that supports the “C” variant (identical to reference).

(O) Position 4170791: Re-analysis of the Aschheim raw data shows 80 % variant frequency for the “A” to “G” variant called as SNP by Wagner et al. Variant frequency is lower than the minimum variant frequency of 95 % set in Wagner et al. The reads containing the “G” variants are located in a region with an abnormal peak in coverage. The Altenerding mapping shows even 25 fold coverage in the region and a 100 % variant frequency that supports the “A” variant (identical to reference).

(P) Position 4199187: Re-analysis of the Aschheim raw data shows 75 % variant frequency and 3 fold coverage for the “A” to “G” variant called as SNP by Wagner et al. Variant frequency is lower than the minimum variant frequency of 95 % and coverage is lower than the 5 fold minimum set by Wagner et al. The only “G” variant is located at the end of a read in a region with an abnormal peak in coverage. The Altenerding mapping shows 47 fold coverage in the region and a 100 % variant frequency that supports the “A” variant (identical to reference).

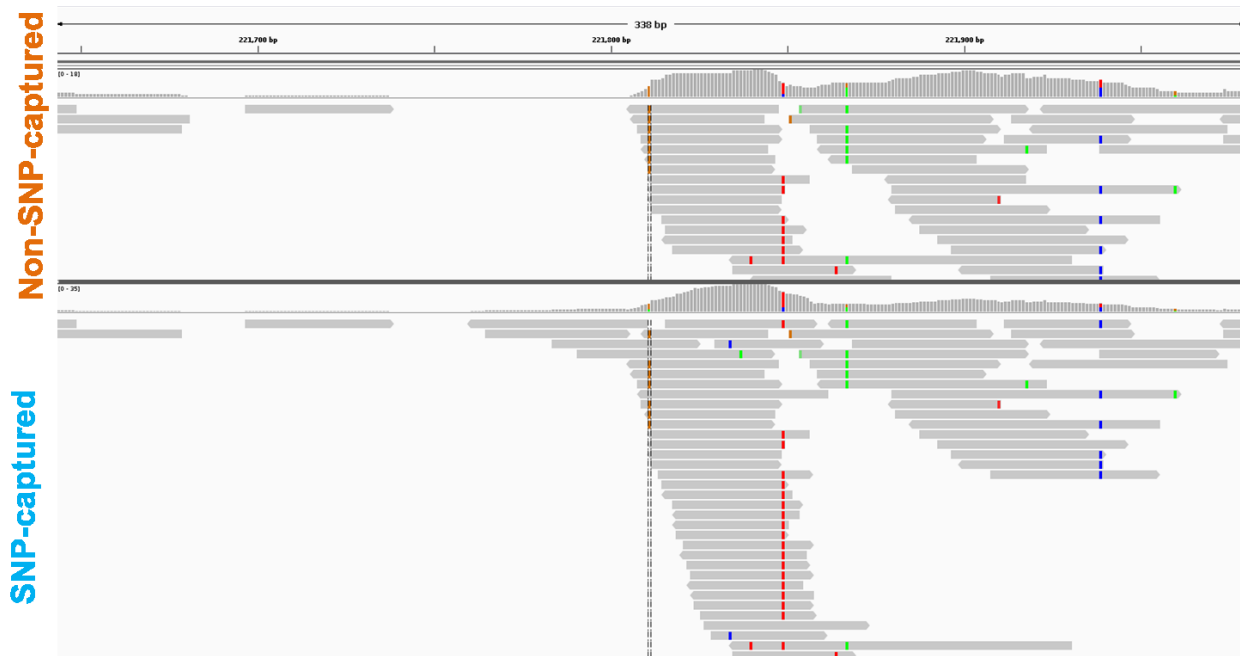
(Q) Position 4199190: Re-analysis of the Aschheim raw data shows 38 % variant frequency for the “T” to “C” variant called as SNP by Wagner et al. Variant frequency is lower than the minimum variant frequency of 95 % set by Wagner et al. The “C” variants are located at end of reads in a region with an abnormal peak in coverage. The Altenerding mapping also shows a cover peak in this region with 50 fold coverage and a 54% variant frequency of the “T” variant (identical to reference).

(R) Position 4203596: Re-analysis of the Aschheim raw data shows 3 fold coverage in position 4203596 called as a “G” to “T” SNP by Wagner et al., in contrast with a 51 fold coverage in the

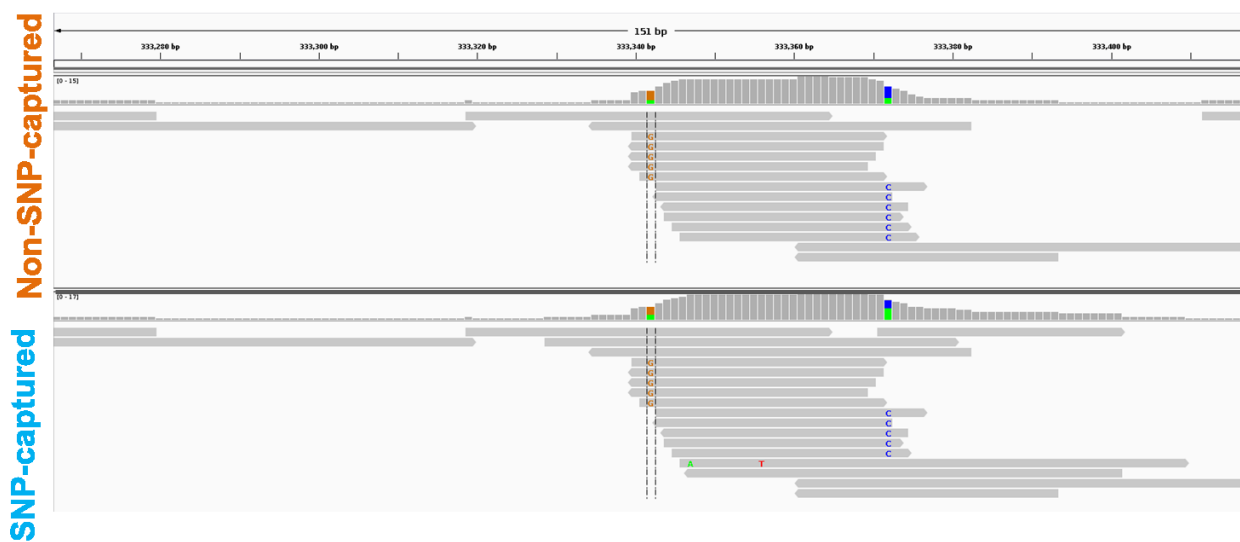
original analysis (Wagner et al. 2014). All 3 reads contain the “G” variant (identical to reference). The Altenerding mapping shows even 22 fold coverage in the region and a 100 % variant frequency that supports the “G” variant (identical to reference).

(S) Position 4210011: Re-analysis of the Aschheim raw data shows 8 fold coverage in position 4210011 called as a “T” to “C” SNP by Wagner et al., in contrast with a 262 fold coverage in the original analysis (Wagner et al. 2014). All 8 reads contain the “T” variant (identical to reference). The Altenerding mapping shows 40 fold coverage in the position and an 85 % variant frequency of the “T” variant (identical to reference).

A



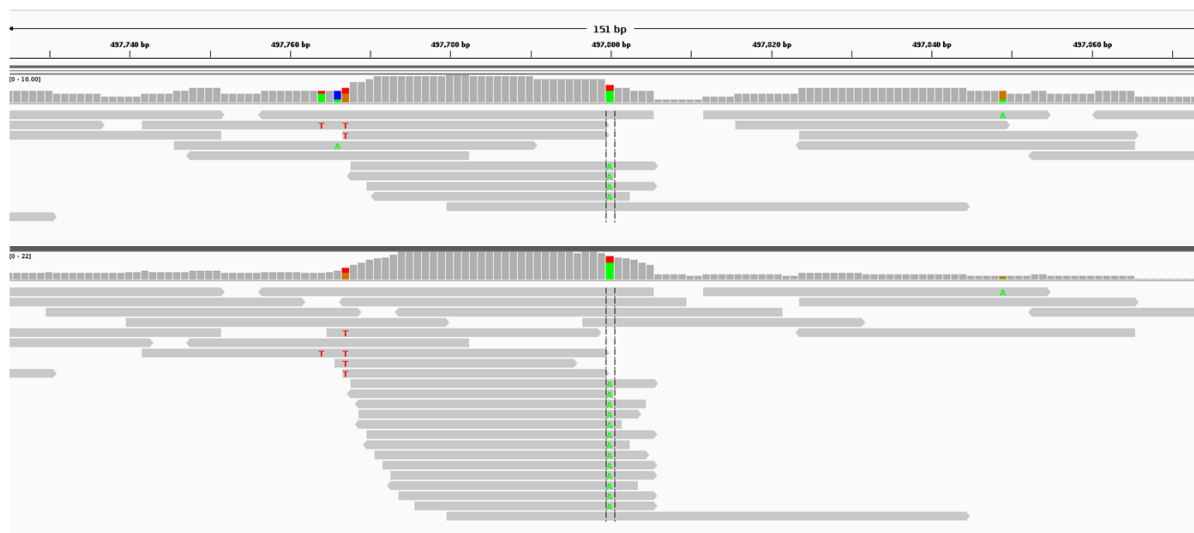
B



C

Non-SNP-captured

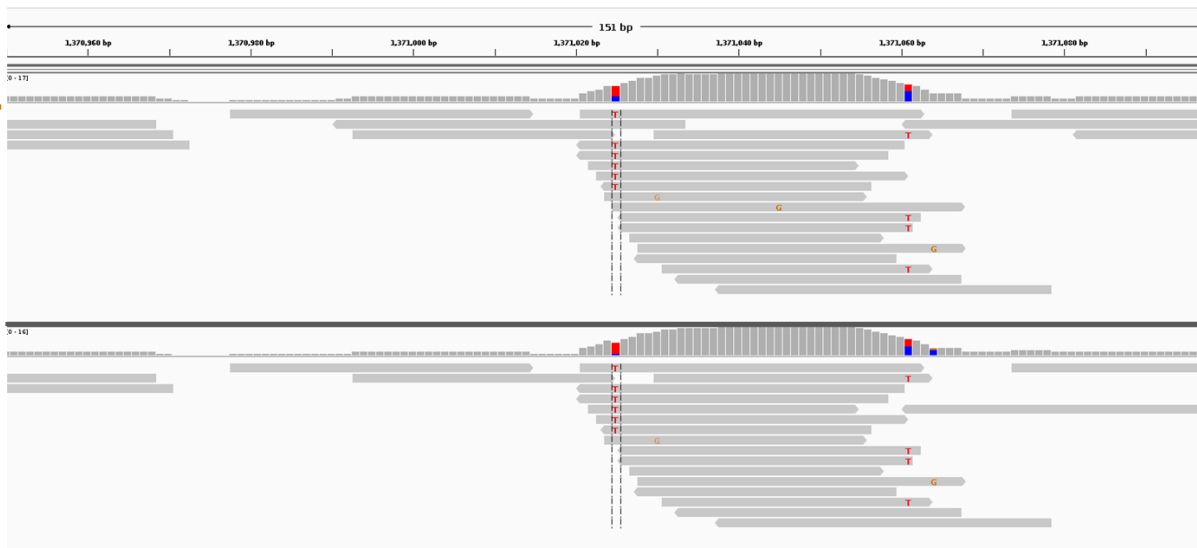
SNP-captured



D

Non-SNP-captured

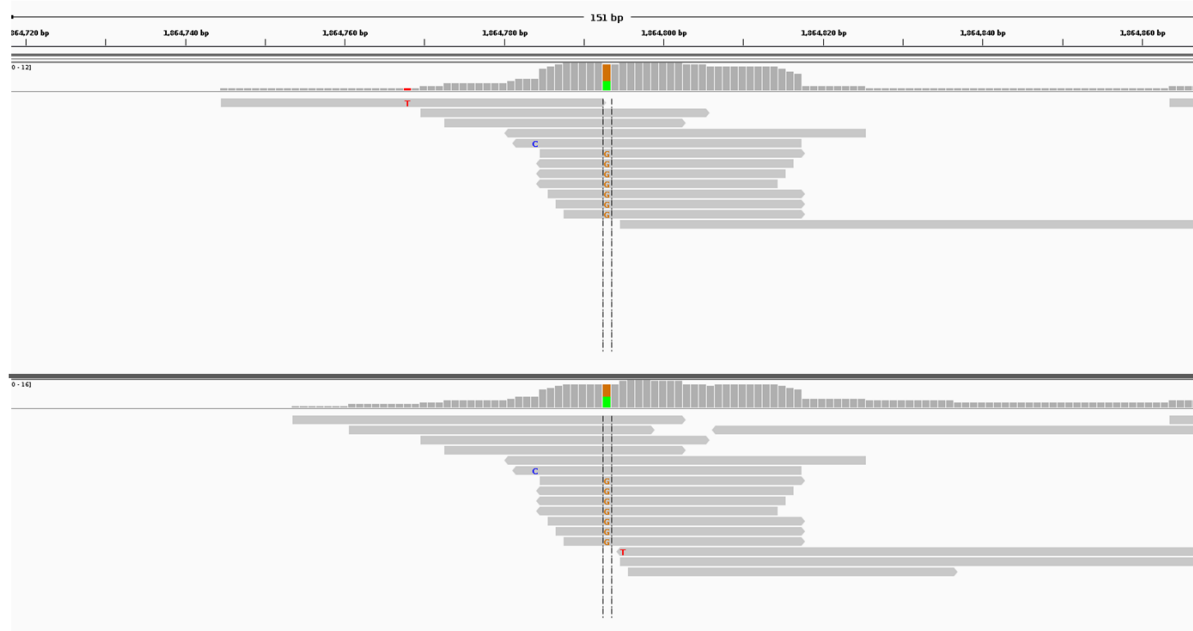
SNP-captured



E

Non-SNP-captured

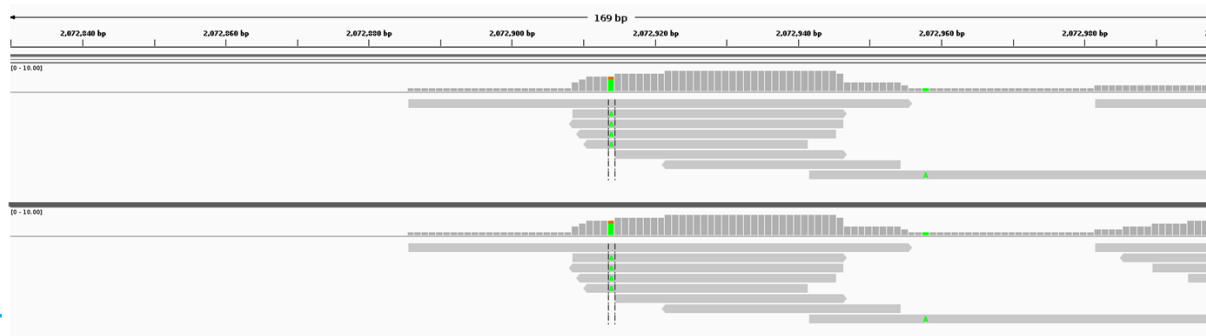
SNP-captured



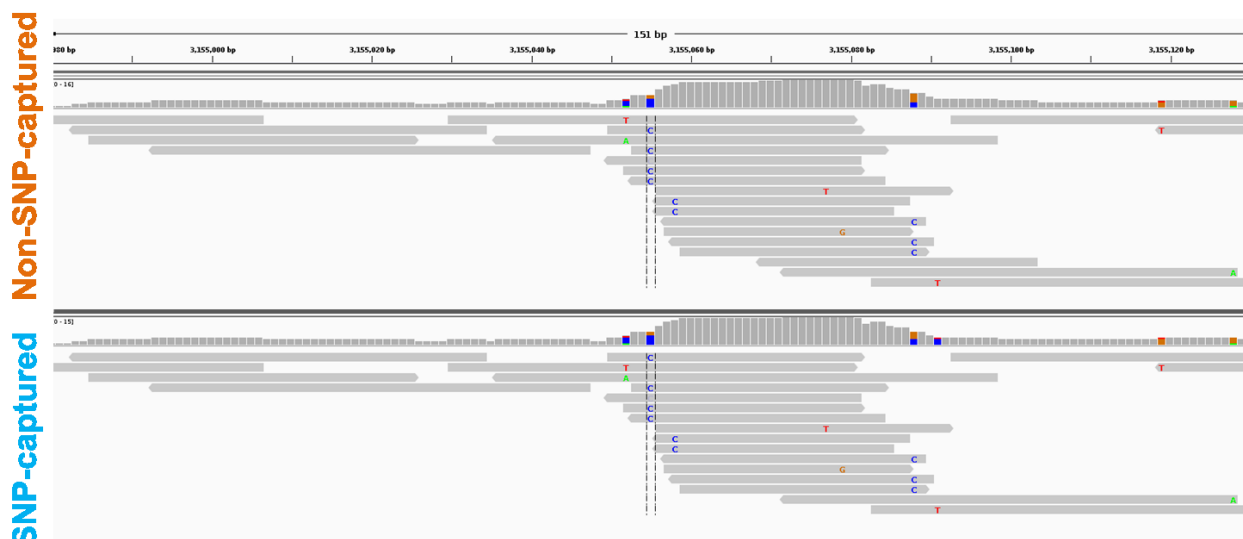
F

Non-SNP-captured

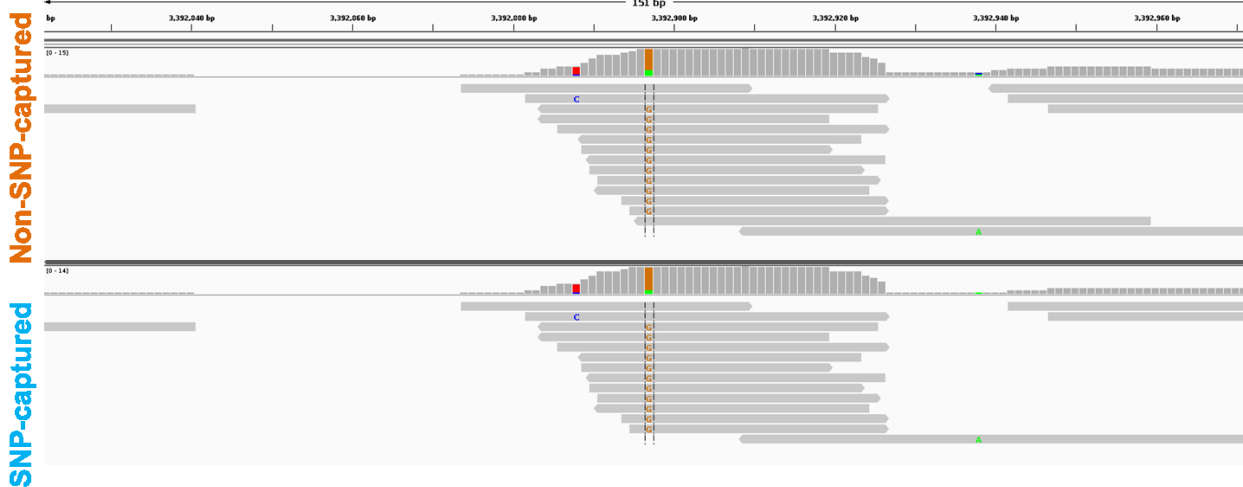
SNP-captured

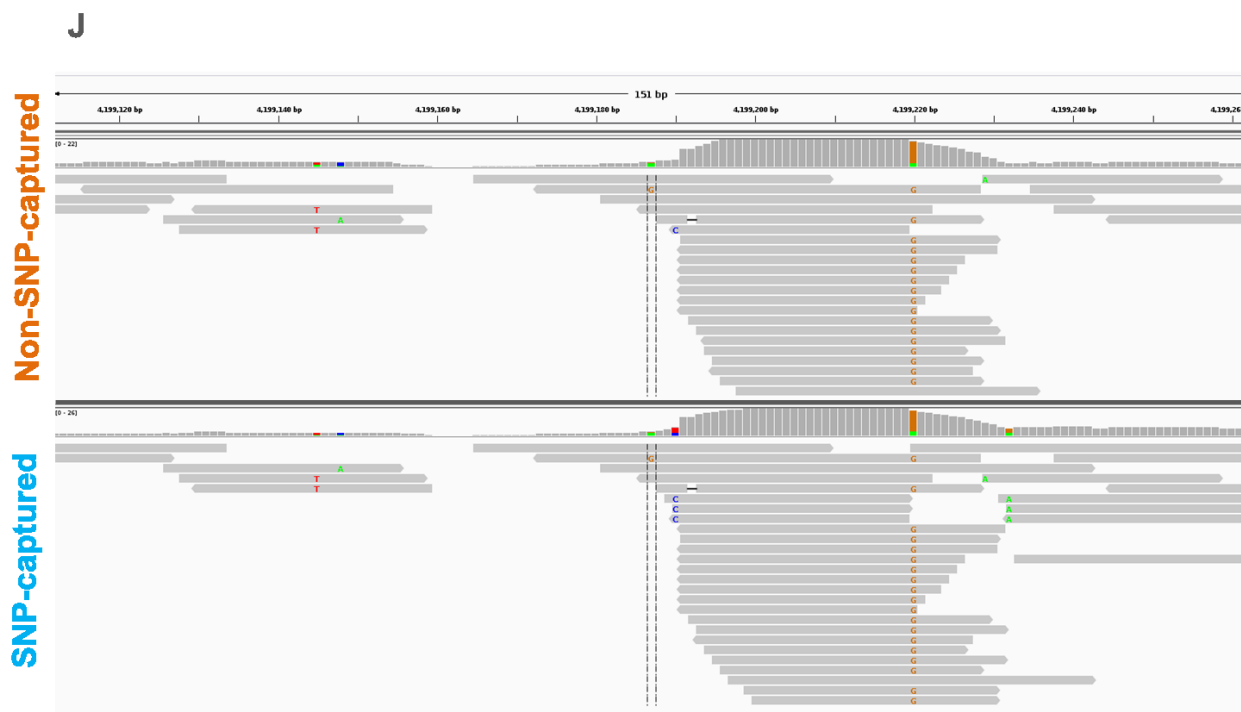


G



H





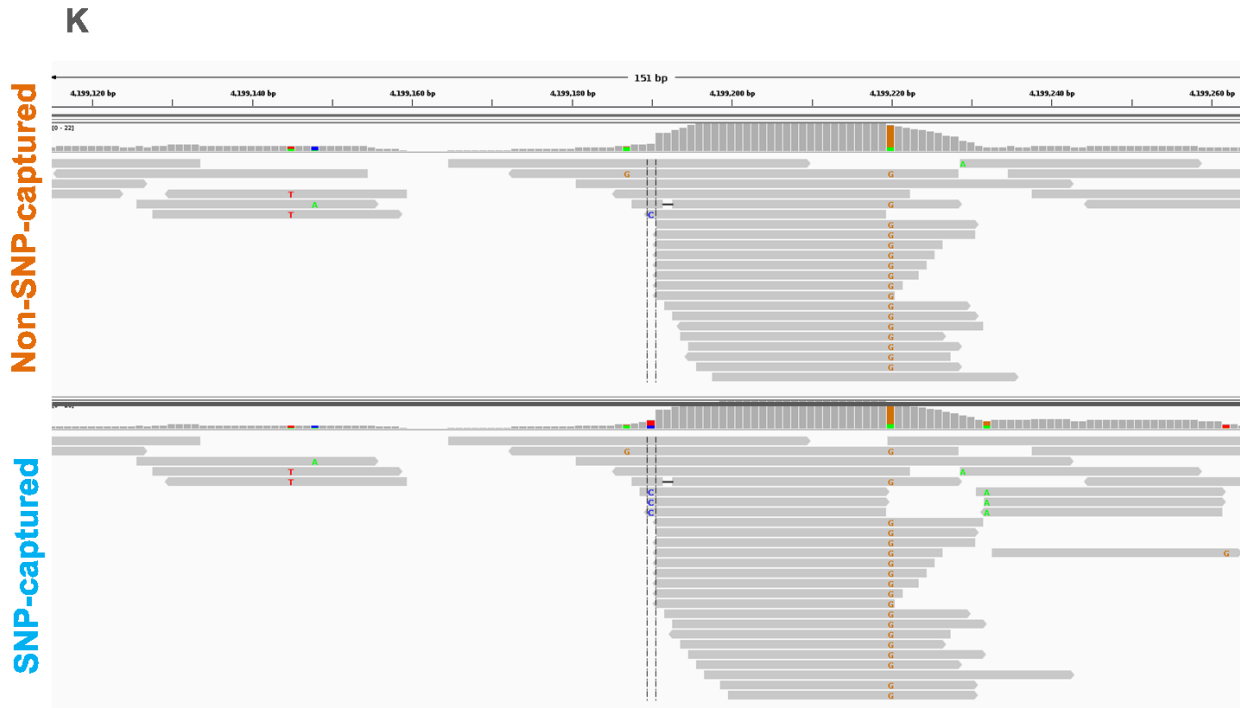
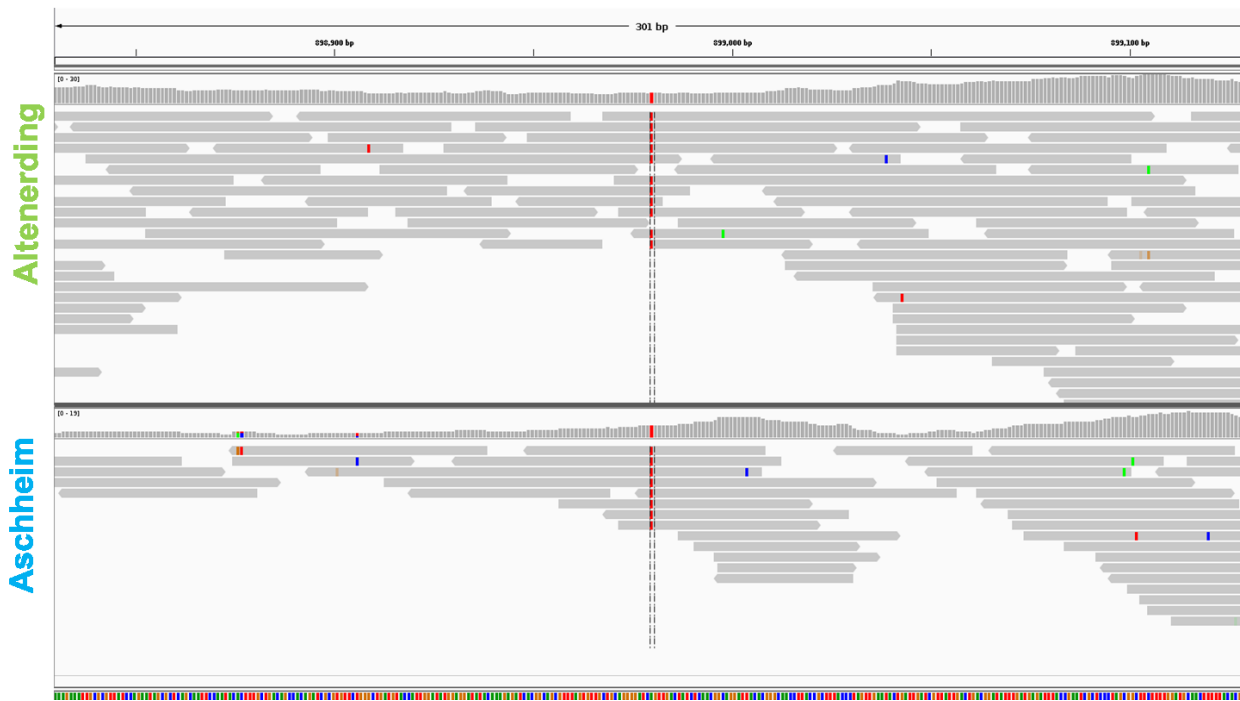
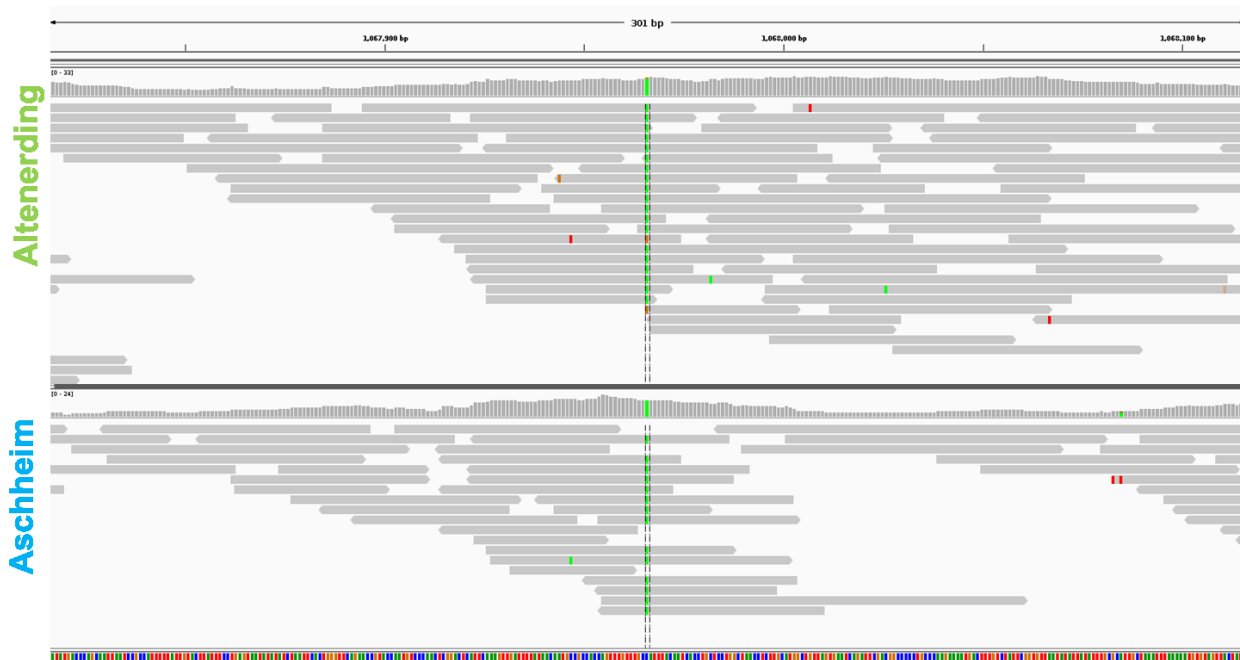


Fig. S8 (A-K): Visualization of the abnormal coverage peaks in the re-analyzed Aschheim SNP enriched data and non-SNP enriched data, containing potential false positive SNPs called by *Wagner et al. 2014*. Reads were mapped to the CO92 reference with sensitivity of 0.1 and minimum mapping quality of 30 and visualized on IGV gene browser. Upper bend shows coverage plot for the region corresponding to the genome beneath. Upper scale shows position (bp) in reference sequence. Dotted line marks the SNP.

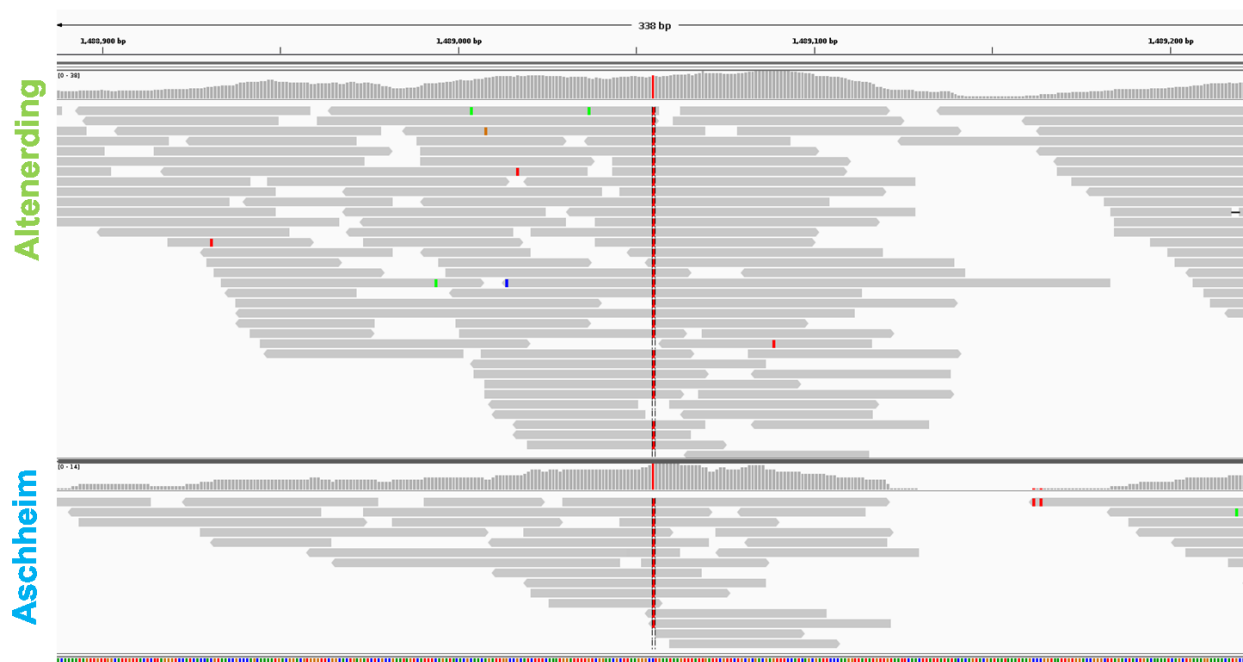
A



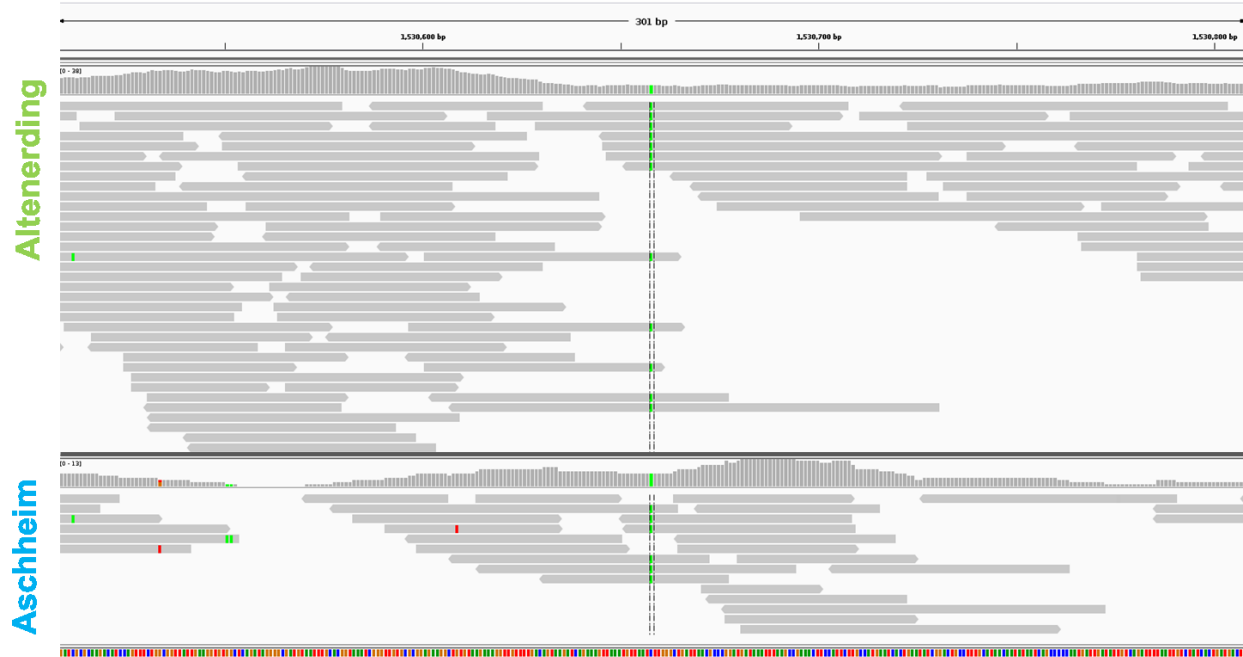
B



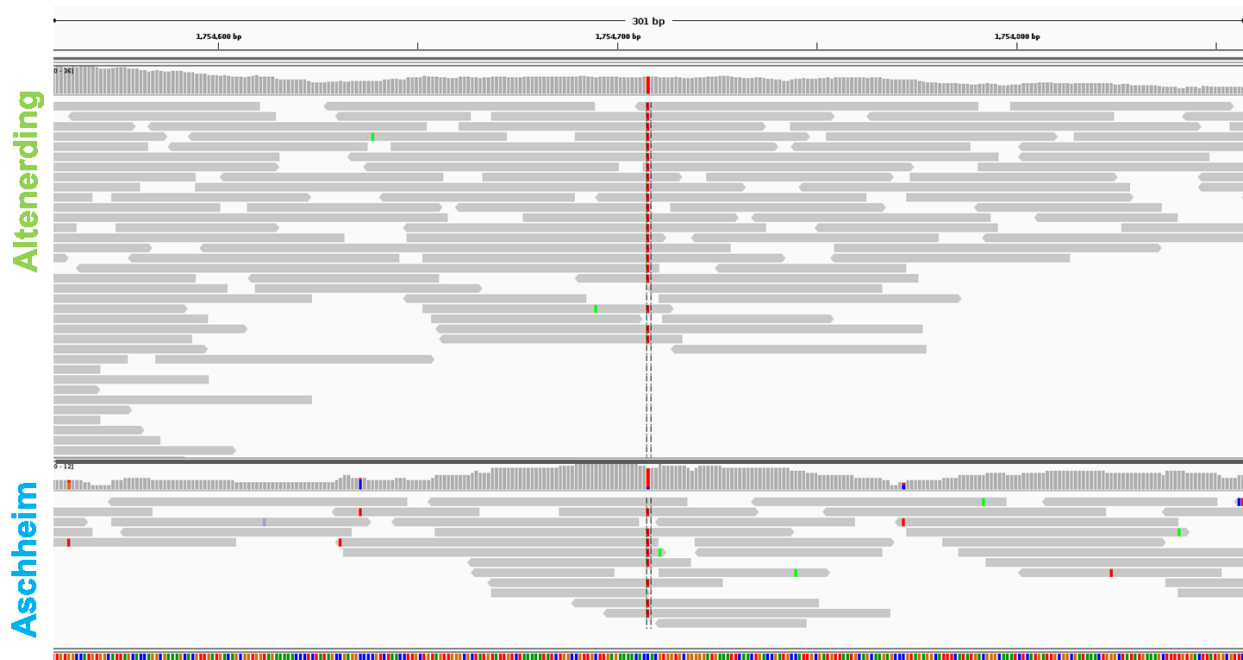
C



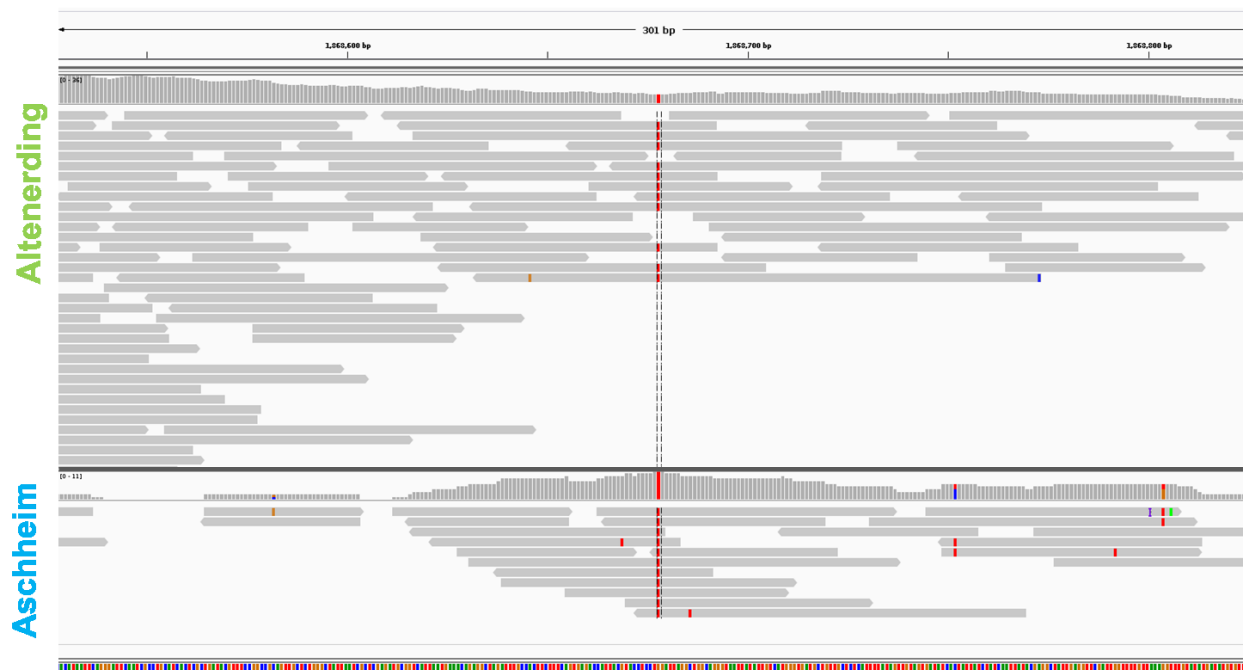
D



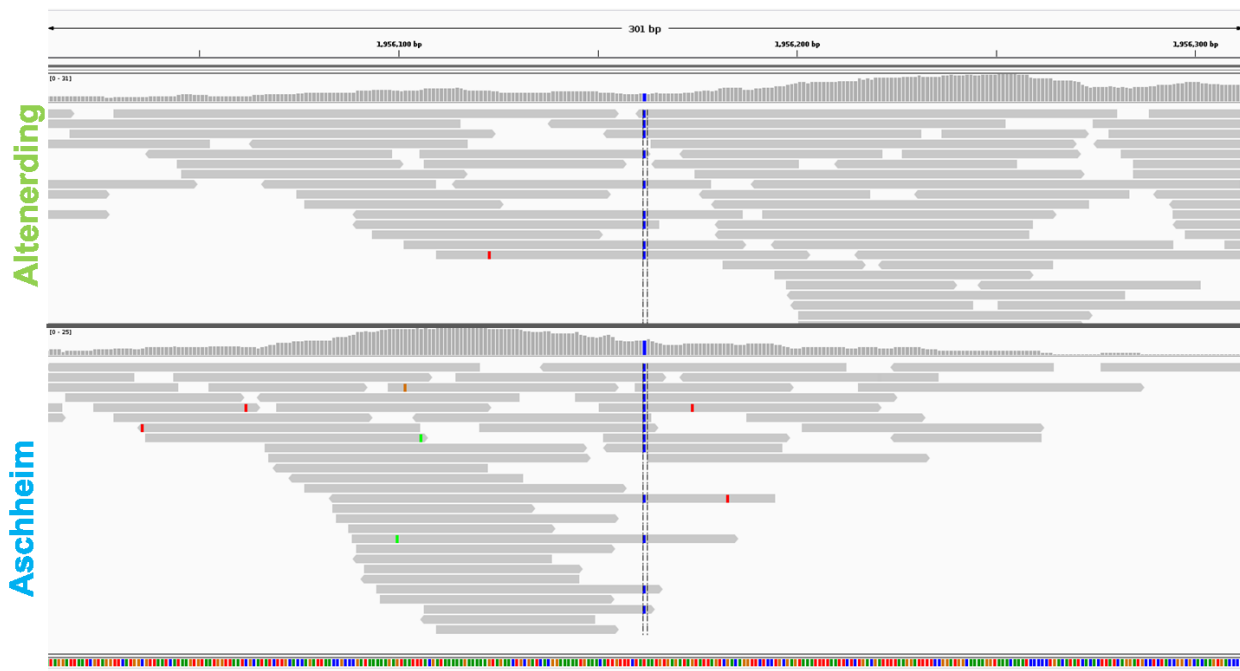
E



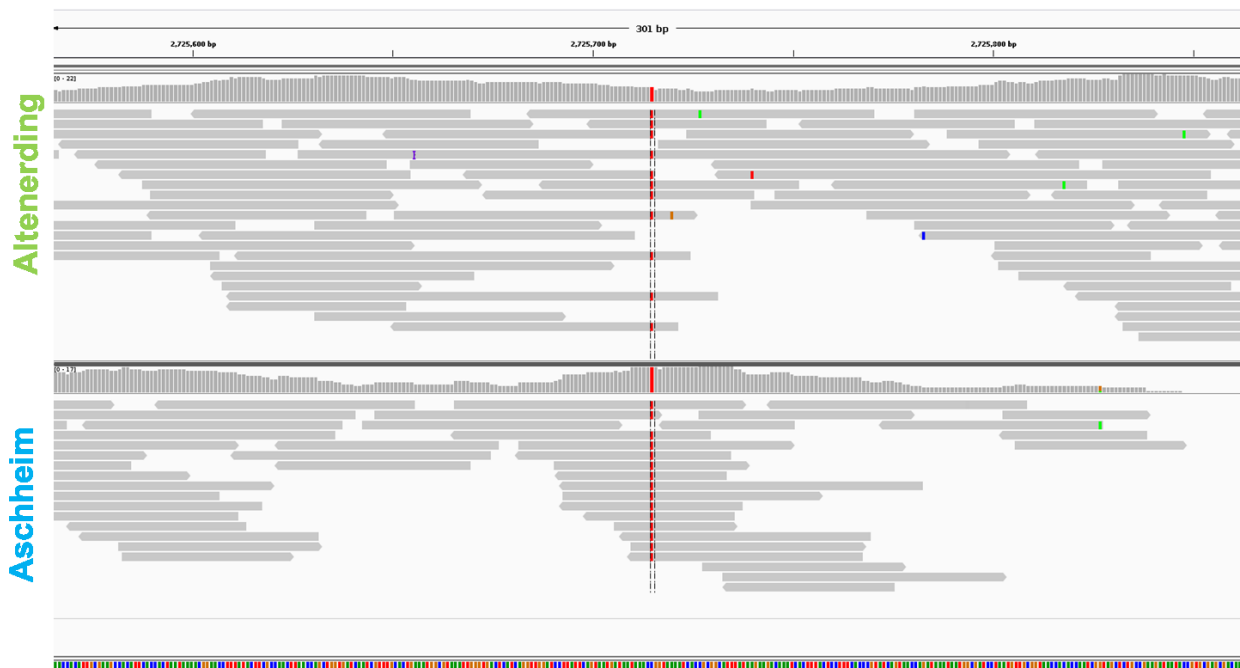
F



G



H



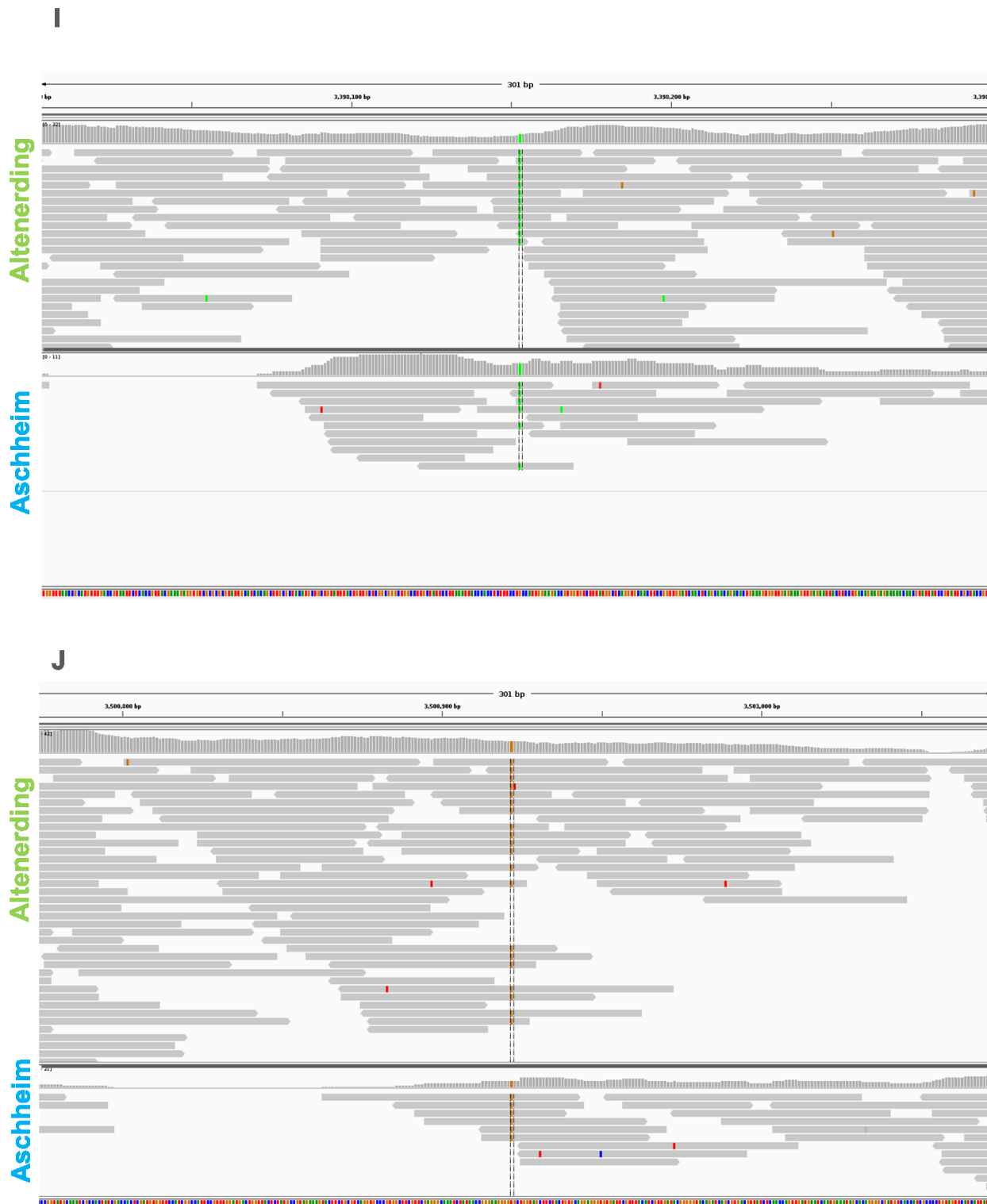
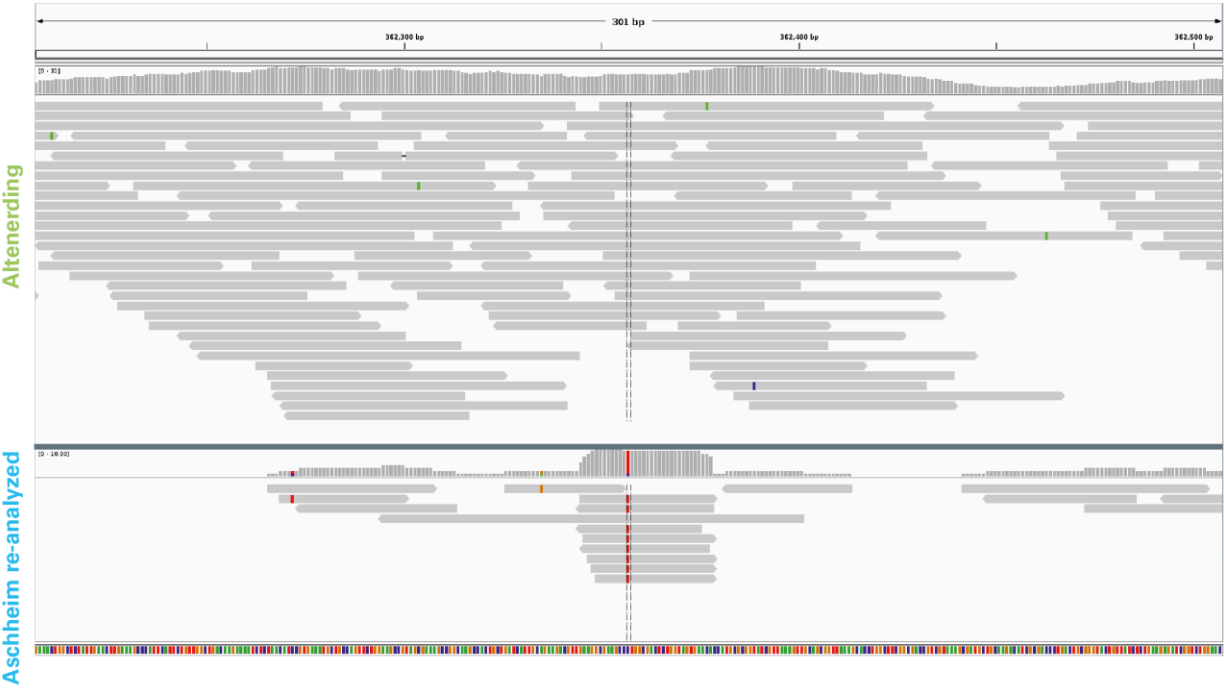


Fig. S9 (A-J): Visualization of 10 positions containing true SNPs called for the re-analyzed Aschheim genome as well as for the Altenerding genome. The positions were randomly picked to represent visual patterns consistent with the set criteria for SNP calling and with a

relative even coverage, in contrast with the pattern of the abnormal coverage peaks shown in figures S4 and S5. Reads were mapped to the CO92 reference with sensitivity of 0.1 and minimum mapping quality of 30 and visualized on IGV gene browser. Upper bend shows coverage plot for the region corresponding to the genome beneath. Upper scale shows position (bp) in reference sequence. Bottom sequence shows the 150 bp in the CO92 reference. Dotted line marks the SNP. **A:** position 898980 (A to T) **B:** position 1067966 (C to A) **C:** position 1489055 (C to T) **D:** position 1530658 (C to A) **E:** position 1754708 (C to T) **F:** position 1868678 (G to T) **G:** position 1956162 (T to C) **H:** position 2725715 (C to T) **I:** position 3398153 (G to A) **J:** position 3500922 (T to G).

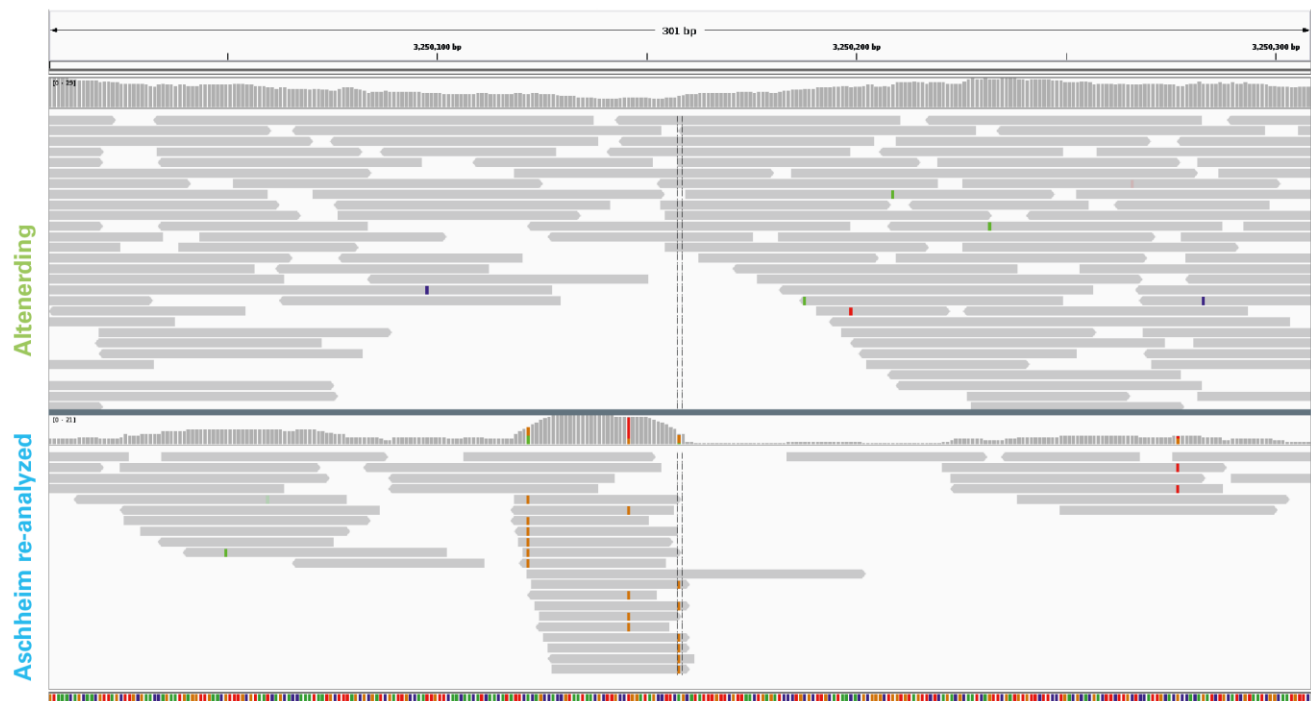
A



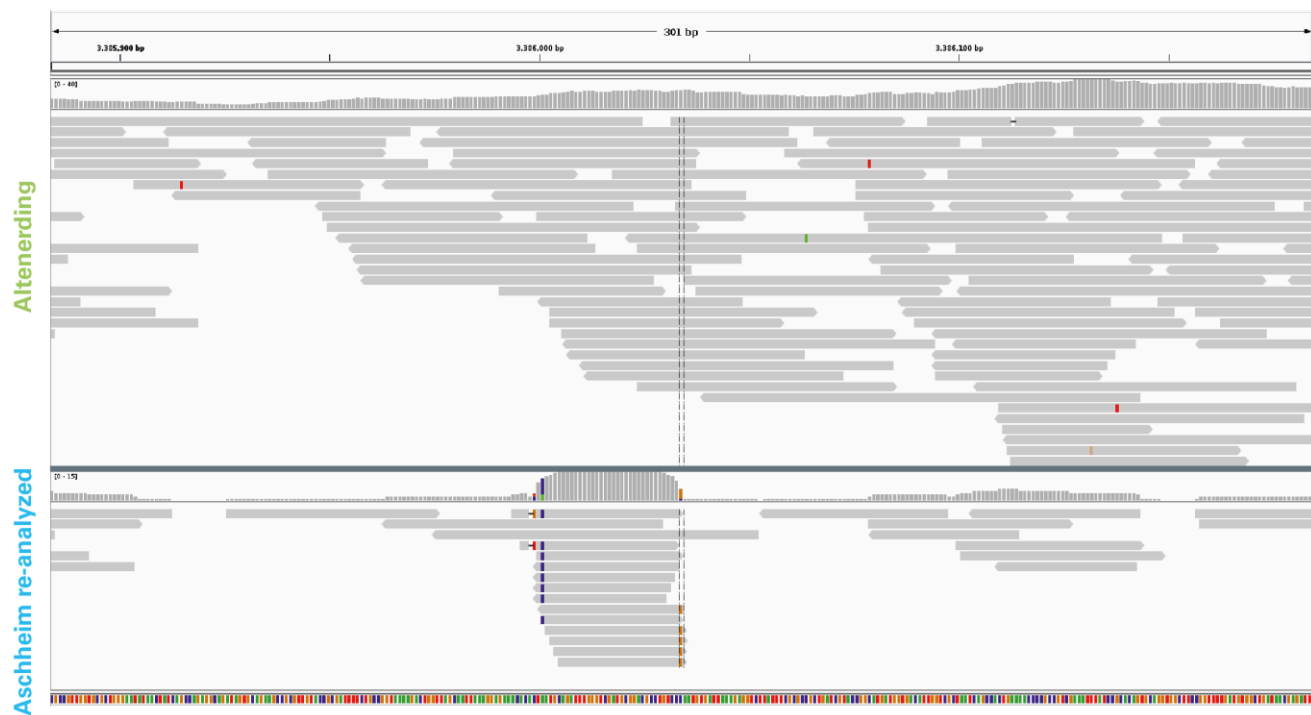
B

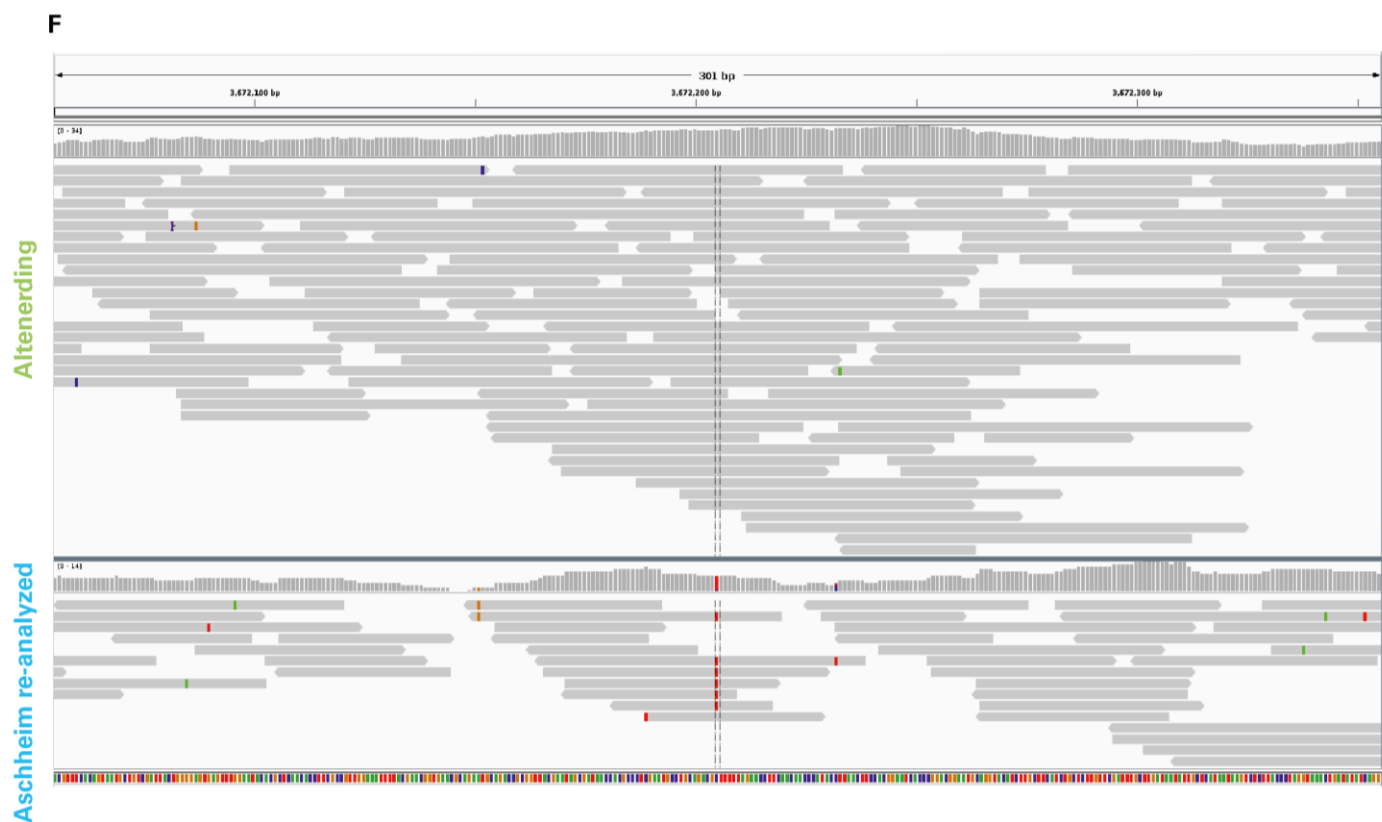
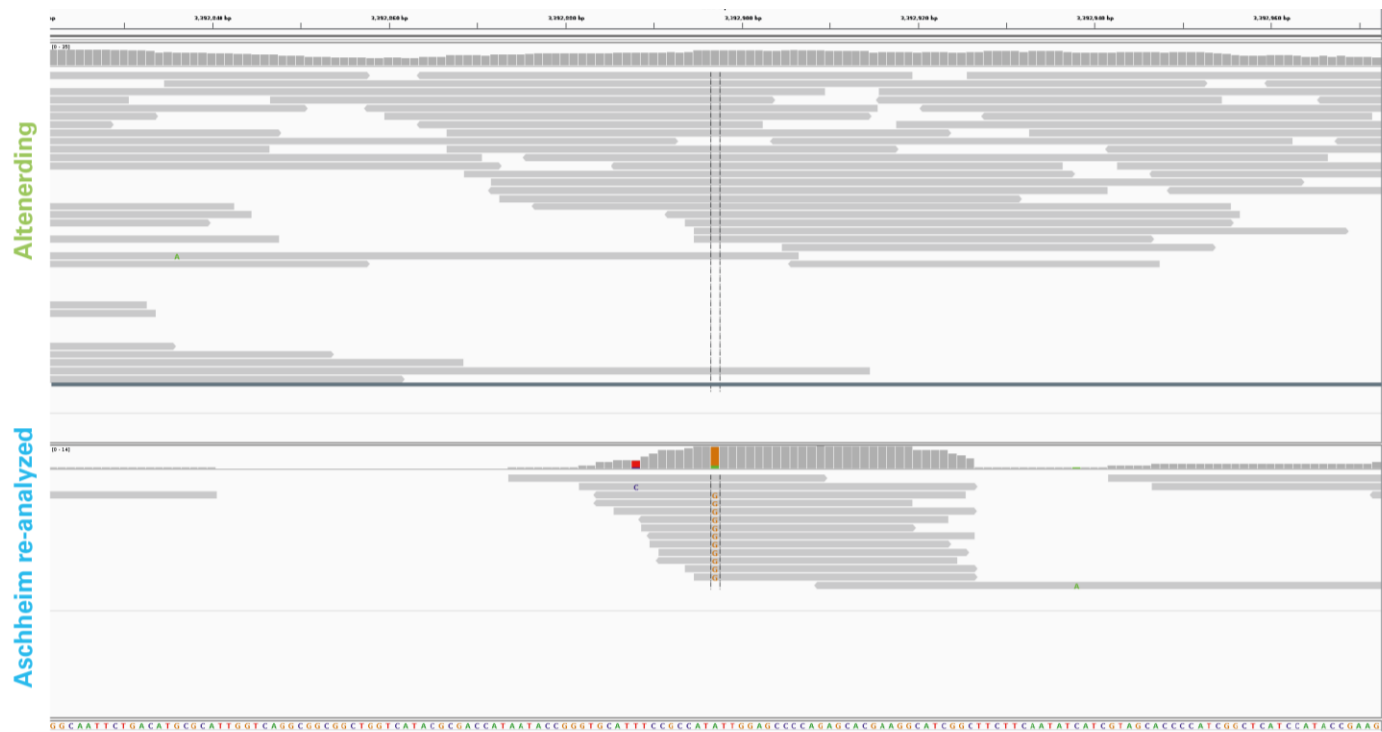


C



D





G



H



Fig.S10 (A-H): Visualization of positions containing potential false positive SNPs specifically derived in the re-analyzed Aschheim. Reads were mapped to the CO92 reference with sensitivity of 0.1 and minimum mapping quality of 30 and visualized on IGV gene browser. Upper bend shows coverage plot for the region corresponding to the genome beneath. Upper scale shows position (bp) in reference sequence. Bottom sequence shows the 150 bp in the CO92 reference. Dotted line marks the false positive SNP.

(A) Position 362357: Re-analysis of the Aschheim raw data shows 89 % variant frequency for the “C” to “T” variant at the position. The “T” variants are located in a region with an abnormal cover peak. The Altenerding mapping shows even 21 fold coverage in the region and a 100 % variant frequency that supports the “C” variant (identical to reference). This SNP was removed from final analysis by *Wagner et al.*, 2014 following a visual inspection.

(B) Position 1371025: Re-analysis of the Aschheim raw data shows 86 % variant frequency for the “C” to “T” variant called as SNP by Wagner et al. Variant frequency is lower than the minimum variant frequency of 95 % set by Wagner et al. The “T” variants are located at end of reads in a region with an abnormal peak in coverage. The Altenerding mapping shows even 18 fold coverage in the region and a 100 % variant frequency that supports the “C” variant (identical to reference).

(C) Position 3250158: Re-analysis of the Aschheim raw data shows 86 % variant frequency for the “A” to “G” variant at the position. The “G” variants are located at end of reads in a region with an abnormal cover peak. The Altenerding mapping shows even 11 fold coverage in the region and a 100 % variant frequency that supports the “A” variant (identical to reference). This SNP was not called by *Wagner et al.*, 2014.

(D) Position 3386034: Re-analysis of the Aschheim raw data shows 83 % variant frequency for the “C” to “G” variant at the position. The “G” variants are located at end of reads in a region with an abnormal cover peak. The Altenerding mapping shows even 24 fold coverage in the region and a 100 % variant frequency that supports the “C” variant (identical to reference). This SNP was not called by *Wagner et al.*, 2014.

(E) Position 3392897: Re-analysis of the Aschheim raw data shows 85 % variant frequency for the A to G variant called as SNP by Wagner et al. Variant frequency is lower than the minimum variant frequency of 95 % set by Wagner et al. The reads containing the “G” variants are located in a region with an abnormal peak in coverage. The Altenerding mapping shows even 22 fold coverage in the region and a 100 % variant frequency that supports the “A” variant (identical to reference).

(F) Position 3672205: Re-analysis of the Aschheim raw data shows 86 % variant frequency for the “C” to “T” variant at the position. The “T” variants are located in a region with an abnormal cover peak. The Altenerding mapping shows even 28 fold coverage in the region and a 100 % variant frequency that supports the “C” variant (identical to reference). This SNP was not called by *Wagner et al.*, 2014.

(G) Position 3956001: Re-analysis of the Aschheim raw data shows 83 % variant frequency for the “T” to “A” variant at the position. The “A” variants are located between two abnormal cover peaks, in a region with high variability. The Altenerding mapping shows even 18 fold coverage in the region and a 100 % variant frequency that supports the “A” variant (identical to reference). This SNP was removed from final analysis by Wagner et al., 2014 following a visual inspection.

(H) Position 4575345: Re-analysis of the Aschheim raw data shows 90 % variant frequency for the “A” to “G” variant at the position. The “G” variants are located in a region with an abnormal cover peak. The Altenerding mapping shows even 23 fold coverage in the region and a 100 % variant frequency that supports the “A” variant (identical to reference). This SNP was removed from final analysis by Wagner et al., 2014 following a visual inspection.

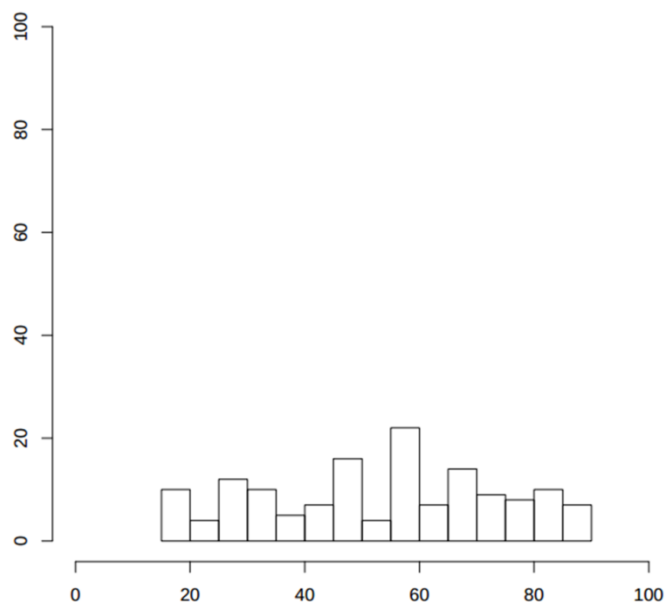


Fig.S11: Genome-wide SNP allele frequency plot of the re-analyzed Aschheim draft genome. The x axis indicates the frequency of reads covering a SNP position in which the SNP was detected in the re-analyzed Aschheim draft genome. The y axis indicates the number of SNP calls with the respective frequency. The observed frequencies are not showing any bimodal pattern or any other pattern that could indicate an infection with multiple strains.

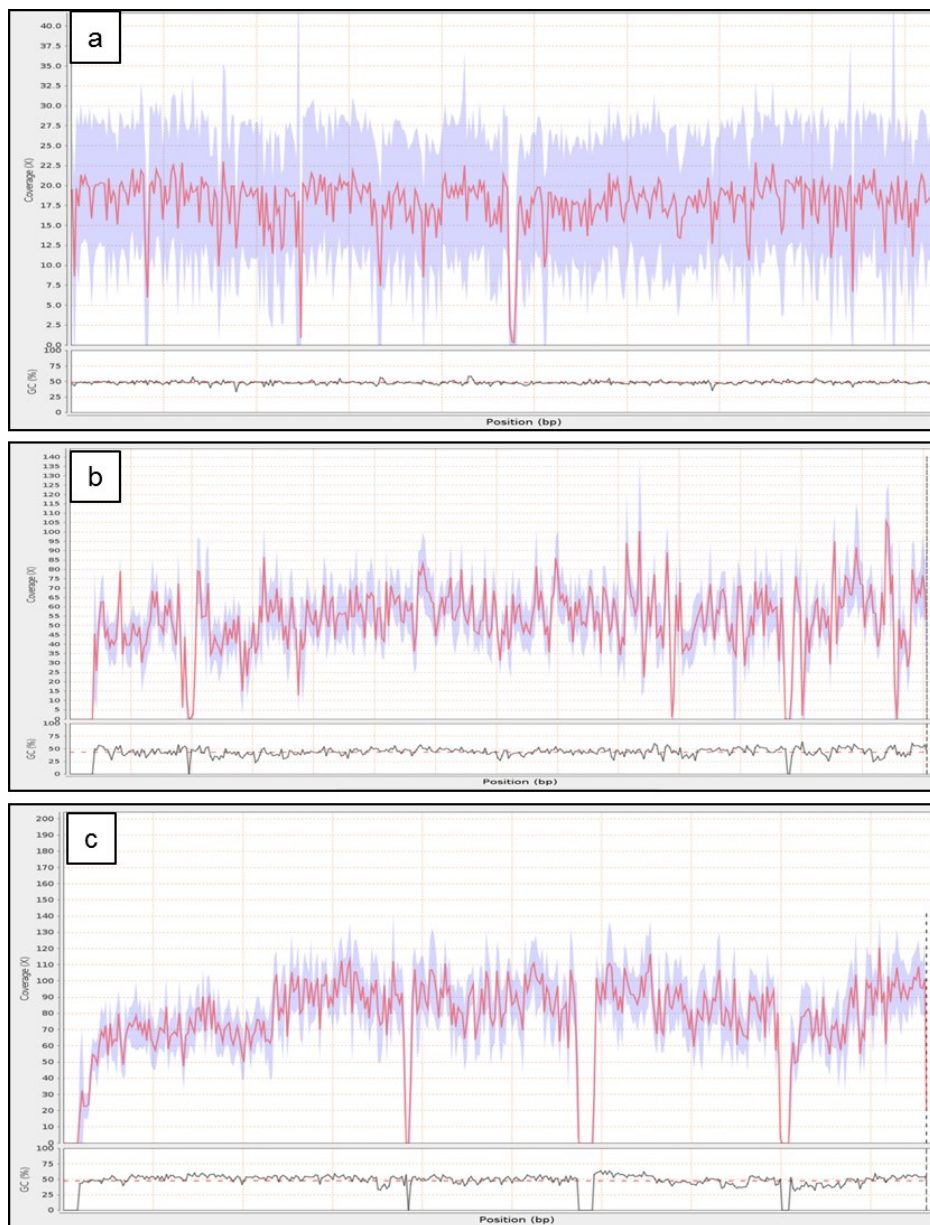


Fig.S12: Coverage plots across the CO92 reference for the Altenerding genome. Coverage across the reference was plotted using QualiMap version 2.1. (a) Coverage across the CO92 chromosome (NC_003143.1). (b) Coverage across the pCD1 plasmid (NC_003131.1). (c) Coverage across the pMT1 plasmid (NC_003134.1). Coverage in red, percent GC content in grey and mean GC content in dashed line.

Supplementary archaeological and historical information

The early medieval cemetery from Altenerding (also called Altenerding/Klettham) is located near Munich in southern Germany. It contains around 1450 inhumations and is therefore one of the largest early medieval cemeteries in Central Europe. Altenerding was excavated from 1966 to 1973 (Sage 1984) and was used from the second half of the fifth century until the seventh century AD (Losert and Pleterski 2003).

Within the cemetery 16 double burials and no multiple burials could be identified. Ten double burials were chosen randomly for screening of *Y. pestis* presence (Table S9). All of them contained grave furnishings, mostly a set of brooches and other dress ornaments for women and a combination of weapons for men, as is typical to this time period (Sage 1984). This is also true for the plague-positive individuals in the double-burial 1175/1176. Here, a 25- to 30-year old woman was laid to rest together with a 20 to 25 year old male. The dead woman (1175) was buried with a variety of clothes and jewels (Fig. 1C) typical of the middle of the 6th century, including an iron arm ring; a belt with bronze belt buckle; a chatelaine with antler pendant; Roman brooches; iron keys; chained links and scales; a knife; a fragment of La Tène glass arm ring; a necklace of glass and amber beads and fragments of a blue Roman glass vessel. She wore a pair of garnet disk fibulas whose typology has been dated between ~530 and ~570 (Vielitz 2003) and among other ornaments she was equipped with a so-called Hercules- or Donar club amulet, which probably expresses a special hope for growth and fertility (Losert and Pleterski 2003). Due to her young age only minor expressions of degenerative lesions in the great joints and the spine are visible on the bones. However, the orbital roofs show porotic lesions on the bone surface (*cribra orbitalia*). These kinds of lesions are rather unspecific and can occur in a variety of diseases (Walker et al. 2009).

The young man was buried without weapons, but a bag hanging on a belt could be reconstructed as containing an iron knife, a lighter and nails (Fig. 1D). Despite the young age of the individual some degenerative lesions are visible in the spine including Schmorl's nodes. The right orbital roof shows porotic lesions on the bone surface (the left orbital roof is missing). Both tibiae exhibit an extensive inflammation of the bone

surface (periostitis). Both symptoms are rather unspecific and can be connected to different kinds of infectious diseases or anemia.

Wooden traces indicate the existence of two coffins or wooden planking in the grave. This is a further sign that the dead were carefully arranged. Furthermore, both individuals were buried with rather expensive clothes and jewels. This indicates that the victims had been dressed and prepared carefully for their funeral. Burial rites, which probably also included washing and public laying out of the body seem to have been conducted also for these plague victims. The same has been noted in the neighboring Aschheim cemetery (Gutsmiedl-Schümann et al. 2010).

No historical record has yet been adduced that mentions the impact of the Justinianic Pandemic in this region. In fact, an 8th-century historian who used some reliable early sources, with respect to the wave dated ~565-571 states explicitly that this outbreak went as far as this region, but stayed within "Italy": "In his [Narses] time, the greatest plague emerged, particularly in the province of Liguria.... And what is more, these evils occurred only within Italy up to the region of the Alamannian and Bavarian peoples, to the Romans alone." (Bethmann and Waitz 1878)

References

- Bethmann L and Waitz G. 1878. Paul the Deacon: History of the Lombards 2.4. Monumenta Germaniae historica, Scriptores rerum Langobardicarum: Hanover 74.3-26 (translation by McCormick).
- Gutsmiedl-Schümann D, Greipl EJ, Sommer S, für Denkmalpflege BL. 2010. Das frühmittelalterliche gräberfeld Aschheim-bajuwarenring. Lassleben.
- Losert H and Pleterski A. 2003. Das frühmittelalterliche gräberfeld von Altenerding in Oberbayern und die "ethnogenese" der Bajuwaren. Scîpvaz-Verlag.
- Sage W. 1984. Das reihengräberfeld von Altenerding in Oberbayern: Katalog der Anthropologischen und Archäologischen funde und befunde. Gebrüder Mann Verlag.
- Vielitz K. 2003. Die granatscheibenfibeln der merowingerzeit. M. Mergoil.

- Wagner DM, Klunk J, Harbeck M, Devault A, Waglechner N, Sahl JW, Enk J, Birdsell DN, Kuch M, Lumibao C. 2014. *Yersinia pestis* and the Plague of Justinian 541–543 AD: a genomic analysis. *The Lancet Infectious Diseases* 14 (4):319-26.
- Walker PL, Bathurst RR, Richman R, Gjerdrum T, Andrushko VA. 2009. The causes of porotic hyperostosis and *Cribra Orbitalia*: A reappraisal of the iron deficiency anemia hypothesis. *Am J Phys Anthropol* 139(2):109-25.