Genomic Insights into Late Neolithic and Early Bronze Age Populations of Modern-Day Switzerland

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Anja Furtwängler
aus Villingen-Schwenningen

Tübingen
2020

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	09.10.2020
Stellvertretender Dekan:	Prof. Dr. József Fortágh
1. Berichterstatter:	Prof. Dr. Johannes Krause
2. Berichterstatter:	Prof. Nicholas J. Conard, PhD

Acknowledgments

I would like to express my thanks for all the support, advice, and help that I have received during my time as a Ph.D. student.

First and foremost, I would like to express my gratitude to Prof. Johannes Krause for his guidance, constant support, and understanding supervision.

To Verena Schünemann for all the helpful advice.

To Prof. Nick Conard for supporting me during my Ph.D. and for kindly agreeing to supervise my dissertation.

My sincere thanks to Dr. Sandra Lösch, Prof. Dr. Albert Hafner, Inga Siebke, and Noah Steuri for a wonderful collaboration on the Swiss material. And to Stephan Schiffels and Wolfgang Haak for providing invaluable advice on the population genetic analyses.

A special thanks to Antje Wissgott and Guido Brandt and all the technicians at the MPI-SHH for performing the pathogen in-solution captures and the sequencing.

To all my colleagues in Tübingen for creating such a lovely working environment.

And to Ella, for deep friendship and support in hard times.

Table of Contents

Abbreviations

aDNA	ancient DNA
ANE	Ancient North Eurasian
bp	base pairs
BP	before present
C	cytosine
CWC	Corded Ware Complex
DNA	deoxyribonucleic acid
LGM	last glacial maximum
MALT	MEGAN Alignment Tool
mtDNA	mitochondrial DNA
nDNA	nuclear DNA
NGS	Next-Generation Sequencing
PCR	Polymerase Chain Reaction
PEC	Primer Extension Capture
POPRES	Population Reference Sample
qPCR	quantitative PCR
RNA	ribonucleic acid
SNP	Single Nucleotide Polymorphism
T	thymine
UDG	Uracil DNA glycosylase
UV	ultraviolet

Summary

Research on aDNA from human remains provides great opportunities to study past populations regarding their genetic history and processes that shaped modern-day populations, such as migration and admixture. In this dissertation, I further improved specific techniques for aDNA analysis, such as specific target enrichment and authentication of the retrieved sequences, and used these improved methods to investigate the population history of modern-day Switzerland.

In the first study, I address issues in estimating contamination which represents a major problem in aDNA research. Although the level of contaminating DNA in the mtDNA can easily be assessed it might not be representative in case of extreme excess of mtDNA compared to nDNA. In this study, it is shown that the ratio between mitochondrial to nuclear DNA is the smallest in petrous bones, meaning that they contain more nuclear DNA in relation to the mitochondrial DNA and that this causes the mitochondrial and nuclear contamination rates to be most similar.

In the second paper, I compare different methods for specific target enrichment to determine the most efficient one. I use statistics to compare their performance regarding the DNA yield, their specificity and their reproducibility. Results indicate that the commercial myBaits® kit utilising RNA baits is the most suitable for the work on aDNA and that in-solution approaches are advantageous.

The third paper focuses on the genetic history of past human populations from present-day Switzerland and surrounding regions such as Southern Germany and the Alsace. The temporal focus is on the genetic transitions detected in the 5th millennium BP. The main aspects of the study are ancestry components, admixture dates and social structure. In this study, results show that a steppe-like ancestry component arrives in Switzerland in around 4700 BP and that the relative amount decreases after a sudden steep increase. Furthermore, it was possible to identify female individuals with relatively young radiocarbon dates but with zero steppe ancestry who were likely causing this decline.

Zusammenfassung

Untersuchungen an aDNA menschlicher Überreste bieten vielfältige Möglichkeiten, vergangene Population hinsichtlich ihrer genetischen Geschichte und der Prozesse, die heutige Populationen formten, wie Migration und Admixture, zu erforschen. In dieser Dissertation habe ich spezifische Techniken zur aDNA-Analyse und Authentifizierung der aus Überresten gewonnen Sequenzen weiter verbessert und die daraus gewonnenen Erkenntnisse zur genetischen Untersuchung der Bevölkerungsgeschichte der heutigen Schweiz verwendet.

Der erste Artikel beschäftigt sich mit Problemen bei der Bestimmung der Kontamination, die ein Hauptproblem in der aDNA-Forschung darstellt. Obwohl der Gehalt an kontaminierender DNA in der mtDNA leicht beurteilt werden kann, ist er im Falle eines extremen Überschusses an mtDNA im Vergleich zu nDNA möglicherweise nicht repräsentativ. In dieser Studie konnte gezeigt werden, dass das Verhältnis zwischen mitochondrialer zu Kern-DNA in Felsenbeinknochen am niedrigsten ist, was bedeutet, dass sie im Verhältnis zur mitochondrialen DNA mehr nukleare DNA enthalten und dass dies dazu führt, dass die Anteile der mitochondrialen und der nuklearen Kontamination am wenigsten voneinander abweichen.

In der zweiten Arbeit verglich ich verschiedene Methoden zur spezifischen Anreicherung von Zielregionen. Ich verwendte Statistiken zum Vergleich ihrer Leistung hinsichtlich der DNA-Ausbeute, ihrer Spezifität und ihrer Reproduzierbarkeit. Die Ergebnisse zeigen, dass das kommerzielle myBaits® Kit für die Arbeit an aDNA am besten geeignet ist und in-solution Methoden generell von Vorteil sind.

Der dritte Artikel enthält eine detaillierte Untersuchung des genetischen Umbruchs, der im 5. Jahrtausend vor heute festgestellt wurde. Die Studie konzentriert sich auf die Region der heutigen Schweiz und umliegende Regionen wie Süddeutschland und das Elsass. Hauptaspekte der Studie sind Abstammungskomponenten, Admixture Dates und die soziale Struktur der

Bevölkerung. In dieser Studie zeigen die Ergebnisse, dass die Abstammungskomponente aus der Pontischen Steppe in der Schweiz um 4700 Jahre vor heute ankommt und dass die relative Menge dieser Komponente nach einem plötzlichen steilen Anstieg abnimmt. Darüber hinaus konnten weibliche Personen mit relativ jungen Radiokarbondaten, ohne Steppenvorfahren identifiziert werden, die wahrscheinlich diesen Rückgang verursachten.

List of Publications

The following three publications are included and discussed in this dissertation:

I. **Furtwängler, A.**, Reiter E., Neumann G.U., Siebke I., Steuri N., Hafner A., Lösch S., Anthes N., Schuenemann V.J., Krause J. "Ratio of mitochondrial to nuclear DNA affects contamination estimates in ancient DNA analysis." *Scientific reports* 8.1 (2018): 14075.

II. **Furtwängler A.**, Neukamm J., Böhme L., Reiter E., Vollstedt M., Arora N., Singh P., Cole S.T., Knauf S., Calvignac-Spencer S., Krause-Kyora B., Krause J., Schuenemann V.J., Herbig A. "Comparison of target enrichment strategies for pathogen studies in ancient DNA." *Manuscript*

III. **Furtwängler A.**, Rohrlach A.B., Lamnidis T.C., Papac L., Neumann G.U., Siebke I., Reiter E., Steuri N., Hald J., Denaire A., Schnitzler B., Wahl J., Ramstein M., Schuenemann V.J., Stockhammer P.W., Hafner A., Lösch S., Haak W., Schiffels S., Krause J. "Ancient genomes reveal social and genetic structure of Late Neolithic Switzerland." *Nature Communications* (2020)

Additionally, I am a co-author on the following articles published during my graduate studies:

Schuenemann VJ, Peltzer A, Welte B, van Pelt WP, Molak M, Wang CC, **Furtwängler A**, Urban C, Reiter E, Nieselt K, Teßmann B, Francken M, Harvati K, Haak W, Schiffels S, Krause J. Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. *Nat Commun*. 2017 May 30;8:15694. doi: 10.1038/ncomms15694.

Key FM., Posth C., Esquivel-Gomez LR., Hübler R., Spyrou MA., Neumann GU., **Furtwängler A**., Sabin S., Burri M., Wissgott A., Lankapalli AK., Vågene ÅJ., Meyer M., Nagel S., Tukhbatova R., Khokhlov A., Chizhevsky A., Hansen S., Belinsky AB., Kalmykov A., Kantorovich AR., Maslov VE., Stockhammer PW., Vai S., Zavattaro M., Riga A., Caramelli D., Skeates R., Beckett J., Gradoli MG., Steuri N., Hafner A., Ramstein M., Siebke I., Lösch S., Selim Erdal Y., Nabil-Fareed Alikhan, Zhou Z., Achtman M., Bos K., Reinhold S., Haak W., Kühnert D., Herbig A., Krause J. Emergence of human-adapted Salmonella enterica is linked to the Neolithization process. *Nature Ecology & Evolution*. 2020.

Mittnik A, Massy K, Knipper C, Wittenborn F, Friedrich R, Pfrengle S, Burri M, Carlichi-Witjes N, Deeg H, **Furtwängler A**, Harbeck M, von Heyking K, Kociumaka C, Kucukkalipci I, Lindauer S, Metz S, Staskiewicz A, Thiel A, Wahl J, Haak W, Pernicka E, Schiffels S, Stockhammer PW, Krause J. Kinship-based social inequality in Bronze Age Europe. *Science*. 2019 Nov 8;366(6466):731-734.

Posth C, Nakatsuka N, Lazaridis I, Skoglund P, Mallick S, Lamnidis TC, Rohland N, Nägele K, Adamski N, Bertolini E, Broomandkhoshbacht N, Cooper A, Culleton BJ, Ferraz T, Ferry M, **Furtwängler A**, Haak W, Harkins K, Harper TK, Hünemeier T, Lawson AM, Llamas B, Michel M, Nelson E, Oppenheimer J, Patterson N, Schiffels S, Sedig J, Stewardson K, Talamo S, Wang CC, Hublin JJ, Hubbe M, Harvati K, Nuevo Delaunay A, Beier J, Francken M, Kaulicke P, Reyes-Centeno H, Rademaker K, Trask WR, Robinson M, Gutierrez SM, Prufer KM, Salazar-García DC, Chim EN, Müller Plumm Gomes L, Alves ML, Liryo A, Inglez M, Oliveira RE, Bernardo DV, Barioni A, Wesolowski V, Scheifler NA, Rivera MA, Plens CR, Messineo PG, Figuti L, Corach D, Scabuzzo C, Eggers S, DeBlasis P, Reindel M, Méndez C, Politis G, Tomasto-Cagigao E, Kennett DJ, Strauss A, Fehren-Schmitz L, Krause J, Reich D. Reconstructing the Deep Population History of Central and South America. *Cell*. 2018 Nov 15;175(5):1185-1197.e22.

Mittnik A, Wang CC, Pfrengle S, Daubaras M, Zariņa G, Hallgren F, Allmäe R, Khartanovich V, Moiseyev V, Tõrv M, **Furtwängler A**, Andrades Valtueña A, Feldman M, 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. *Nat Commun*. 2018 Jan 30;9(1):442.

Lazaridis I, Mittnik A, Patterson N, Mallick S, Rohland N, Pfrengle S, **Furtwängler A**, Peltzer A, Posth C, Vasilakis A, McGeorge PJP, Konsolaki-Yannopoulou E, Korres G, Martlew H, Michalodimitrakis M, Özsait M, Özsait N, Papathanasiou A, Richards M, Roodenberg SA, Tzedakis Y, Arnott R, Fernandes DM, Hughey JR, Lotakis DM, Navas PA, Maniatis Y, Stamatoyannopoulos JA, Stewardson K, Stockhammer P, Pinhasi R, Reich D, Krause J, Stamatoyannopoulos G. Genetic origins of the Minoans and Mycenaeans. *Nature*. 2017 Aug 10;548(7666):214-218.

Gutaker RM, Reiter E, **Furtwängler A**, Schuenemann VJ, Burbano HA. Extraction of ultrashort DNA molecules from herbarium specimens. *Biotechniques*. 2017 Feb 1;62(2):76-79.

Rougier H, Crevecoeur I, Beauval C, Posth C, Flas D, Wißing C, **Furtwängler A**, Germonpré M, Gómez-Olivencia A, Semal P, van der Plicht J, Bocherens H, Krause J. Neandertal cannibalism and Neandertal bones used as tools in Northern Europe. *Sci Rep*. 2016 Jul 6;6:29005.

Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, **Furtwängler A**, Haak W, Meyer M, Mittnik A, Nickel B, Peltzer A, Rohland N, Slon V, Talamo S, Lazaridis I, Lipson M, Mathieson I, Schiffels S, Skoglund P, Derevianko AP, Drozdov N, Slavinsky V, Tsybankov A, Cremonesi RG, Mallegni F, Gély B, Vacca E, Morales MR, Straus LG, Neugebauer-Maresch C, Teschler-Nicola M, Constantin S, Moldovan OT, Benazzi S, Peresani M, Coppola D, Lari M, Ricci S, Ronchitelli A, Valentin F, Thevenet C, Wehrberger K, Grigorescu D, Rougier H, Crevecoeur I, Flas D, Semal P, Mannino MA, Cupillard C, Bocherens H, Conard NJ, Harvati K, Moiseyev V, Drucker DG, Svoboda J, Richards MP, Caramelli D, Pinhasi R, Kelso J, Patterson N, Krause J, Pääbo S, Reich D. The genetic history of Ice Age Europe. *Nature*. 2016 Jun 9;534(7606):200-5.

Posth C, Renaud G, Mittnik A, Drucker DG, Rougier H, Cupillard C, Valentin F, Thevenet C, **Furtwängler A**, Wißing C, Francken M, Malina M, Bolus M, Lari M, Gigli E, Capecchi G, Crevecoeur I, Beauval C, Flas D, Germonpré M, van der Plicht J, Cottiaux R, Gély B, Ronchitelli A, Wehrberger K, Grigorescu D, Svoboda J, Semal P, Caramelli D, Bocherens H, Harvati K, Conard NJ, Haak W, Powell A, Krause J. Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe. *Curr Biol*. 2016 Mar 21;26(6):827-33.

Author's Own Contributions

I. I performed all laboratory work consisting of DNA extraction, generation of DNA sequencing libraries and mitochondrial capture. I subsequently performed all the analyses of shotgun and capture data. I carried out the statistical analysis with extensive support from Nils Anthes. I drafted the main manuscript with input from all co-authors and wrote the supplement. My work was carried out under the supervision of Johannes Krause.

II. I performed all laboratory work with the help of Ella Reiter, starting with the preparation of existing sequencing libraries of the samples for target enrichment. I then conducted all array and RNA bait in-solution capture experiments. I processed the raw data and performed the subsequent statistical analysis with support from Alexander Herbig. I drafted the main manuscript, generated all figures and wrote the supplement. My work was carried out under the supervision of Johannes Krause and Verena Schünemann.

III. I conducted laboratory work for all the samples with help from Ella Reiter and the student assistant Gunnar Neumann. I performed DNA extraction, generated sequencing libraries for 81 of the reported samples and conducted in-solution capture for all. I analysed the shotgun and mtDNA capture data and processed the genome-wide in-solution capture data. I performed population genetic analysis under the supervision of Stephan Schiffels with input from Thiseas Lamnidis, Luka Papac, Ben Rohrlach and Wolfgang Haak. I drafted the main manuscript, generated all the figures and wrote the supplement with input from all co-authors. My work was carried out under the supervision of Johannes Krause.

1. Introduction

This dissertation concerns detailed genetic studies on Late and Final Neolithic human individuals from the region of present-day Switzerland. It includes investigations regarding methodological improvements as well as the subsequent examination of the human and pathogen DNA found in the archaeological remains. The following sections contain background information about working with aDNA and genetic transitions in Central Europe.

1.1 The field of ancient DNA research

The research field of ancient DNA delves into the genetic background of archaeological and palaeontological events such as human migration, animal domestication, past species dispersal and extinction (**Hofreiter et al. 2001**). It is a powerful tool to investigate evolutionary processes and can be applied to shed light on archaeological research questions and prehistoric events.

Its beginnings go back to the 1980s, when the DNA of an extinct mammal belonging to the genus *Equus*, the quagga, was first analysed (**Higuchi et al. 1984**). Shortly after, the DNA of a human – an Egyptian mummy to be precise – was also reconstructed (**Pääbo et al. 1985**). Both studies presented rather short sequences retrieved via bacterial cloning.

Similar attempts to obtain genetic information from archaeological and palaeontological material were then made easier by the invention of the polymerase chain reaction (PCR) – a method used to amplify DNA (**Pääbo et al. 1989, Krause et al. 2010a, Hagelberg et al. 2015**). This approach makes it possible to multiply even trace amounts of genetic material left in fossils, and numerous studies on aDNA were published in the decades after its invention. Hopes were high that DNA could be extracted from long-extinct organisms and there were even attempts to obtain DNA from dinosaur remains (**Woodward et al. 1994**). The retrieved sequences of the latter were, however, later shown to be human contamination (**Hedges et al. 1995**).

Intensive work on the survival of DNA after the death of an organism soon put an end to these illusions. Several studies showed that DNA undergoes strong fragmentation and chemical modification *post-mortem* and its survival is therefore limited to a maximum of one million years (**Lindahl 1993, Sawyer et al. 2012, Allentoft et al. 2012, Dabney et al. 2013**). At present the oldest authentic sequences originate from a Middle Pleistocene horse found in the Yukon in Canada dating to 560 – 780 kyr BP, which was preserved under permafrost conditions (**Orlando et al. 2013**).

The next important invention that revolutionised the field of aDNA also allowing the abovementioned studies about DNA survival, was that of the Next-Generation Sequencing (NGS) approach (**Krause 2010a**). The necessary process of generating sequencing libraries by the ligation of artificial adapter sequences to the DNA fragment ends proved to be ideal for the work with aDNA. Now, the highly fragmented nature of aDNA was no longer such a disadvantage and chemical modification could now be investigated more easily and then even be used as an authentication criterion. Furthermore, specific enrichment of genomic regions of interest was now possible. Both procedures of authentication and specific target enrichment are described in more detail in the following sections.

1.2 aDNA authentication

Soon after starting the analysis of aDNA via PCR, it was realised that this method was prone to false positive results and contamination (**Pääbo et al. 1989**). The characteristics of aDNA, such as high fragmentation, chemical modifications and its abundance in only trace amounts, constitute obstacles to this otherwise so practicable method, and in PCR-based approaches there is no possibility of authenticating any of the retrieved sequences (**Krause 2010a**). Therefore, a list of strict criteria was established in 2000 concerning how to proceed when working with aDNA, including the rule of replicating the experiments in independent laboratories (**Cooper and Poinar 2000**).

In contrast to PCR, using NGS-based approaches has certain advantages. First, by ligating artificial adapters to the fragment ends it is possible to obtain the entire fragment and not only the parts in between primer binding sites, as is the case in PCR-based approaches. These fragment ends are of particular importance. Because it has been shown that the most frequently occurring chemical modification, namely the deamination of cytosine, later read as uracil by polymerases causing substitutions of C by T, increases in frequency towards the ends of the fragments (**Briggs et al. 2007, Sawyer et al. 2012, Dabney et al. 2013**). This leads to a characteristic pattern which can be used to determine the ancient origin of retrieved DNA sequences. The inspection of this pattern typical of aDNA is the first step of authentication of all aDNA sequences from various source organisms. A software that is widespread and very practicable for the visualisation of this damage pattern is *mapDamage* (**Ginolhac et al. 2011**).

The second step of authentication is to determine that the sequences originate from one single individual only (**Krause et al. 2010b**). To do this, all the individual reads sequenced are aligned to a reference genome. Then, heterozygosity can be measured. The only possible regions for this procedure are the mtDNA and the X and Y chromosome in males, since they are the only haploid regions of the human genome. The first methods invented simply counted the occurrence of heterozygous positions on the mtDNA and the number of reads with the alternative alleles at these positions (**Krause et al. 2010b**). Soon after, more sophisticated approaches were developed to estimate the contamination rate using a Bayesian method (**Fu et al. 2013**). Currently, the software s*chmutzi* is widely used to determine mitochondrial contamination. This program not only estimates the contamination but also computes consensus calling. Furthermore, it uses an iterative approach considering deamination patterns, fragment length and a database of the worldwide distribution of mtDNA SNPs to separate the putative contaminating fragments from the endogenous ones before the consensus generation (**Renaud et al. 2015**), resulting in highly reliable reconstructions of mitochondrial genomes.

The amount of contamination in the nuclear DNA can also be estimated in male individuals, since their X and Y chromosomes are haploid as well. Estimates are usually performed on the X chromosome, since it is larger, and X chromosomal reads are contained in the contamination of both males and females. The proportion of contaminating reads is estimated using ANGSD (**Rasmussen et al. 2011**). This includes two steps: the first is to calculate the occurrence of the four possible alleles in a list of polymorphic sites which is provided with the software. In the second step, Fisher's exact test is used to test whether the resulting value is significantly different from the ones determined from the value obtained from adjacent sites. In the same step, the relative amount of contamination is estimated.

The described approach for estimating nuclear contamination rates is only possible for male individuals, of course. For female individuals, nuclear contamination cannot be estimated this way since they do not have haploid regions on their nuclear genome. For females, mtDNA contamination rates are often used to extrapolate nDNA contamination rates, or their datasets are reduced to reads carrying aDNA-specific damage at the fragment ends, leading to an extensive loss of data (**Fu et al. 2016**).

1.3 Specific target enrichment for aDNA sequencing libraries

Not only is aDNA highly fragmented and damaged, it is also only present in trace amounts. In modern tissue samples, approximately 1 μg per gram of tissue can be found. In archaeological and palaeontological samples, the amount is usually well over 10,000 times less than that (**Higgins et al. 2015**).

Furthermore, the vast majority of the retrieved sequences are not from the target organism itself but originated from the environment, such as soil bacteria or fungi and/or flesh-decomposing microorganisms (**Green et al. 2006**). The proportion of endogenous DNA usually represents less than one per cent of this environmental background, which means that a large proportion of the sequenced reads is useless for analysis (**Briggs et al. 2007, Sawyer et al. 2012, Allentoft et al. 2012**). Furthermore,

the sample must be sequenced to a certain depth to obtain enough information from this one per cent. Shotgun sequencing of ancient samples to sufficient coverage is therefore very cost-intensive (**Krause 2010a**).

Specific target enrichment strategies have been developed to circumvent this (**Hodges et al. 2007, Burbano et al. 2010, Maricic et al. 2010, Fu et al. 2013**). With these approaches, it is possible to increase the proportion of endogenous DNA before sequencing. Different procedures are possible to do this and there are some commercial kits available which are all based on hybridisation capture, meaning that artificial probes with sequences complementary to the target region are used to fish out the fragments of interest. These kits have not been exclusively developed for the research on aDNA but can also be applied in modern genetics – in the case of uncultivable bacteria such as *Mycobacterium leprae* or *Treponema pallidum*, for example. For these species, the increase of genetic material through bacterial growth is not possible or very difficult (**Schuenemann et al. 2013, Arora et al. 2016)**. For the application on archaeological and palaeontological material, some of these approaches must be adapted to the specific nature of aDNA.

First, "arrays" were used for these enrichments. On these arrays, several million artificial probes are printed on small glass slides. The hybridisation happens in the liquid between the printed glass slide and a second slide. Afterwards, the unbound fragments are washed away, and the specific fragments hybridised to the complementary probes are eluted (**Hodges et al. 2007, Burbano et al. 2010**).

An in-solution hybridisation capture can be performed by cleaving the probes from the glass slide (**Fu et al. 2013**). One of the main advantages of this improvement, compared to the arrays, is that the probes can be transformed into a bait library, and are therefore immortalised and can be reused indefinitely, which highly reduces the costs. Furthermore, freely moving probes might also increase the enrichment efficiency. After the hybridisation happening in the liquid phase, the probes together with the sample's bound target fragments are immobilised on magnetic beads via streptavidin-biotin bonds and can be washed and subsequently eluted.

The third widespread method, at least for aDNA from pathogens, is the in-solution capture with RNA baits. The best known application for aDNA is the commercial myBaits® kit from Arbor Bioscience. In this hybridisation capture kit, the probes are made of RNA instead of DNA. Like the in-solution approach mentioned above, it involves magnetic beads to immobilise the hybridisation compound. This kit is available for whole genomes of a variety of organisms and for custom-made bait designs.

In-solution approaches for specific targeted DNA enrichment are also used for applications on human samples allowing detailed investigations about the genetic history of past populations.

1.4 Genetic transitions in Europe

Investigating human history using genetic data from human remains is especially challenging, as modern human DNA is ubiquitous. Nevertheless, using the methods described above including contamination tests and enrichment strategies, researchers have succeeded in reconstructing various ancient human genomes and have investigated the genetic history of different parts of the world including Europe.

Early modern humans contributing to the modern non-African populations appeared in Europe approximately 45,000 BP (**Hublin et al. 2020**). These hunter-gatherer societies successfully inhabited Europe for several millennia and even persisted during the last glacial maximum (LGM). Genetic studies have shown them to be quite a homogeneous population over large distances (**Fu et al. 2016, Mathieson et al. 2015**).

After this long period of hunter-gatherer tradition in Central Europe, a new subsistence strategy appears starting from around 8,000 BP. The humans start to become sedentary and cultivate crops (**Whittle and Cummings 2007**). For quite some time there was debate in the archaeological community as to whether these changes happened through cultural diffusion, by local people adapting new ideas or

technologies, for instance, or whether migration occurred and the new material culture was brought into Central Europe with this influx of people. Here, genetic investigations hold the key to answering this question. From the very start of aDNA research with analysis restricted to partial and complete mitochondrial genomes, it already became obvious that new people arrived in Central Europe carrying mtDNA haplogroups which had not been present in the area before (**Haak et al. 2005, Haak et al. 2010, Brandt et al. 2013, Brotherton et al. 2013**). Nuclear studies confirmed these findings soon after (**Skoglund et al. 2012**, **Lazaridis et al. 2014**), and it was furthermore possible to determine the origin of these newcomers. Starting from Western Anatolia, their population expanded into Central Europe and admixed with the hunter-gatherer groups, which had been indigenous until that point.

Investigations of mtDNA haplogroups throughout the entire Neolithic have also shown that there are changes in the haplogroup composition of Central European populations in the 5th millennium BP (**Brandt et al. 2013**). This coincides with the emergence and spread of the Corded Ware Complex in Central Europe. With regard to this transition in material culture, there was also debate as to whether this cultural change was also associated with a transition in the population or whether the new cultural practices were adopted by local people.

In 2014, genetic comparisons of modern Europeans with Neolithic Early Farmers – represented by one individual of the Linear Band Ware culture in Southern Germany – and Mesolithic Hunter-Gatherer individuals from Luxembourg and Sweden showed that modern-day Europeans carry not two ancestry components but three (**Lazaridis et al. 2014**). The third component, besides those derived from the Early Farmers from Anatolia and the Western Hunter-Gatherer, was also found in present-day Native Americans. Comparing this third component with ancient genomes published at that time, it was most closely related to one upper Palaeolithic Hunter-Gatherer individual from Siberia, the MA1 (Mal'ta boy, **Raghavan et al. 2014**). This led to the hypothesis of an Ancient North Eurasian (ANE) metapopulation which contributed to both Native Americans and modern-day Europeans after the Neolithic (**Lazaridis et al. 2014**).

Subsequent publications showed that this third component can be found in Central Europe starting from the 5th millennium BP (**Haak et al. 2015, Allentoft et al. 2015**). These studies showed individuals from the Yamnaya culture in the Pontic steppe to be of the best fitting origin. This late Copper Age and Early Bronze Age culture could be found in the region between the Southern Bug, Dniester and Ural rivers (**Šišlina 2008**). Their subsistence strategy was that of nomadic pastoralists. Genetically, these Steppe Herders were a mixture of Eastern Hunter-Gatherers with a close affinity to the ANE, and a population of Iranian Neolithic ancestry. This Near Eastern ancestry came into these Steppe Herders by Neolithic Farmers from what is now Iran spreading northwards into the Eurasian Steppe (**Wang et al. 2019, Lazaridis et al. 2016**).

This genetic component formed in the Pontic steppe suddenly appears in Central Europe in the 5th millennium BP. Its sudden and massive appearance is indicated by the earliest admixed individuals associated with the Corded Ware Complex being genetically the closest to the Yamnaya individuals, while younger individuals have less affinity to the Steppe Herders (**Haak et al. 2015**). This extreme immigration appears to follow a previous population decline. This decline is reflected by a lack of skeletons and the demise of the mega settlements formed during the Neolithic. An epidemic event has been suggested to be partially responsible of this severe population decline, with *Yersinia pestis* being the causative agent. DNA of this bacterium was found in several Final Neolithic individuals (**Rasmussen et al. 2015, Andrades Valtuena et al. 2017**).

Besides the abovementioned population decline at the end of the Neolithic period, one other important aspect of these far-reaching migrations of members of the Yamnaya culture or their close relatives both westwards into Central Europe and eastwards could also be the domestication of the horse or, more precisely, the skill of horse-riding greatly increasing the mobility of the members of this culture (**Anthony 2007**).

With this large population turnover and therefore a large number of newcomers, it has been suggested that the Indo-European languages also came to

Central Europe around that time (**Haak et al. 2015, Allentoft et al. 2015**). This steppe hypothesis concerning the origin of the Indo-European languages was substantially supported by these genetic studies. Originating in Northern Iran this language form spread northwards into the Eurasian steppe and, at the end of the Neolithic, into Central Europe together with the steppe-related genetic component (**Meller et al. 2017**). The Indo-European languages also left the Pontic steppe eastwards and reached what is now India, where genetic steppe-related ancestry can also be found in today's population (**Narasimhan et al. 2019**).

This language family is strongly associated with the Y chromosomal haplogroup R1, which is the most common haplogroup in Europe today. In Western Europe, the subclade R1b is most common, and in Eastern Europe, the subclade R1a (**Myres et al. 2011, Underhill et al. 2015**).

Being located at the very centre of Central Europe, present-day Switzerland provides a great opportunity to study transitions in the populations caused by migrating people, as the region might have served as a corridor for migrating populations from and to Eastern, Western and Southern Europe. The region's topography also shows a wide range of landscapes, including some of Europe's highest mountains, which could also have been a barrier to human migration. This contrast makes this region particularly interesting for the study of human migrations in the past.

1.5 The end of the Neolithic in what is now Switzerland

At the beginning of this dissertation only four genomes were published from the region that forms Switzerland today. One Upper Palaeolithic Hunter-Gatherer from the Bichon cave (**Jones et al. 2015**) and three Bell Beaker individuals from the Dolmen of Petit-Chasseur (**Olalde et al. 2018**). That the numbers were that low might also be related to the bad bone preservation from the Neolithic to the Bronze Age as described in the following section.

Rich archaeological assemblages from the Neolithic period can be found in Switzerland. These findings are dominated by lakeshore and bog settlement sites, inner alpine sites of the Rhône valley and high mountain pass sites (**Hafner et al. 2014**). Besides numerous remains of settlements, there are only a few burials to be found and even less that have a good skeletal preservation. These comprise Chamblandes-type stone cist graves and a few dolmen burials towards the end of the Neolithic, such as the megalithic burial sites of Oberbipp, Sion and Aesch (**Hafner et al. 2003, Hafner et al. 2014, Siebke et al. 2018, Siebke et al. 2019**). The pile dwellings and wetland settlements are particularly worth mentioning, with wooden parts often in an excellent state of preservation, providing one of the best dated dendrochronological records in prehistoric Europe (**Hafner et al. 2003**).

Towards the end of the Neolithic, the Corded Ware Complex appears in Switzerland. Material associated with the Corded Ware Complex can be found exclusively from sites on the banks of large pre-alpine lakes. In the region of Lake Zurich in Eastern Switzerland and the Three Lakes Region in western Switzerland especially, they can be found in high numbers, with the sites on Lake Neuchâtel being some of the most southwestern ones in the entire Corded Ware Complex. While the new material culture was quickly adopted in Eastern Switzerland, the process in Western Switzerland took several centuries. Dendrochronological data obtained from wooden parts of the buildings from Corded Ware Complex settlements provide the opportunity for absolute dating (**Suter et al. 2017**).

Unlike the numerous Neolithic and Early Bronze Age settlements, there are only a few graves associated with Corded Ware Complex settlements. This might be because the occurrence of Chamblandes-type stone cist tombs as they had been used in the 7th millennium BP declines around 5,800 BP, which is exactly the time when the number of wetland settlements on the lakeshores increase.

A high concentration of Early Bronze Age burials is found in inner alpine regions (Rhône valley, Lake Thun area, foothills of the Alps) where no lakeside settlements from this period can be found. Probably for taphonomic reasons, there are no settlements in Switzerland in periods when graves were numerous, and, vice

versa, in periods with a high number of settlements, the corresponding burials are scarce due to bad preservation conditions.

2. Aims and Objectives

The aim of this dissertation was to improve specific techniques widely used in aDNA for the retrieval of genetic material from archaeological remains and authentication of these retrieved sequences, and to apply these improved techniques to investigate the genetic history of present-day Switzerland.

This aim was achieved with three studies, each of which completes one of the following subgoals: 1) improving the authentication method for data of human female individuals, 2) determining the most successful target-enrichment strategy for aDNA, and 3) using the results from 1) and 2) to investigate in detail the genetic transition in the 5th millennium BP in what is now Switzerland.

In paper I, new data from diverse parts of the skeleton, such as bones, petrous bones and teeth, was generated and combined with data from previously published studies to obtain a large dataset to enable a statistics-based analysis of the relationship between the mitochondrial to nuclear DNA ratio and the accuracy of the extrapolation from mtDNA contamination to the nuclear level.

In paper II, 30 modern and ancient samples of *Mycobacterium leprae* and *Treponema pallidum* were enriched using three different protocols that have been widely used in aDNA research. The capture experiments for each protocol were repeated independently to enable detailed statistical comparisons to detect the most efficient and specific target enrichment strategy that produces the most constant results.

In paper III, skeletons of almost 100 individuals from what are now Switzerland, Southern Germany and the Alsace region in France dating between approximately 6500 BP and 2000 BP were screened for their aDNA content, and later enriched for human mtDNA and nDNA. The resulting sequence data was then used to investigate the genetic transition in the Final Neolithic with the focus on timing and duration, as well as the dominant social structure in the populations studied before and after the transition.

3. Methods

Precautions for working with aDNA

When working with ancient DNA, very specific techniques are applied. The degraded nature and the contamination with environmental DNA that often occurs make aDNA vulnerable and it must be strictly protected from further contamination during the laboratory workflow. Therefore, very strict measures are applied, and all pre-amplification work is carried out in cleanroom facilities (**Knapp et al. 2012**). Sample preparation, DNA extraction and the generation of DNA libraries were carried out in a pre-PCR cleanroom facility dedicated to ancient DNA work at the University of Tübingen. Precautions against contamination include full-body suits, face masks, hairnets, protective footwear and two to three pairs of gloves. All surfaces are routinely UV-irradiated and cleaned with a diluted solution of bleach to minimise DNA cross-contamination between samples. Chemicals that are not sensitive to UV light were UV-irradiated to remove contaminant DNA. The human remains were also irradiated with UV light prior to sampling in order to inactivate surface DNA.

DNA extraction and generation of sequencing libraries

Where possible, both petrous bones and teeth were sampled for all ancient individual studied during this dissertation. Petrous bones were chosen for the human DNA analysis since they have been shown to contain the highest proportions of endogenous DNA (**Gamba et al 2014, Pinhasi et al. 2015**). Teeth, on the other hand, were picked and analysed to detect any DNA of blood-borne pathogens. For both teeth and petrous bones, specific regions were chosen for sampling. For petrous bones, the cochlea was targeted. Teeth were cut at the cementoenamel junction, and dentine was sampled from the inside of the crown.

Samples of both skeletal elements were processed with the same workflow for the initial screening. After sampling, DNA extraction was performed in line with **Dabney et al.** (2013), a protocol specifically designed for the extraction of very short DNA fragments. In this approach, the mineral matrix of the bone and teeth powder

is dissolved with an extraction buffer containing EDTA and the DNA is released into the solution. It is then bound to silica membranes in commercial DNA clean-up columns. In this way, the DNA can be washed with ethanol-based wash buffers and subsequently be eluted from the membrane. Portions of this purified DNA are then used to build double-stranded sequencing libraries in line with **Meyer and Kircher** (2010) by the ligation of artificial adapters to both ends of the DNA fragments. These libraries are then marked with a sample-specific barcode, which means that a unique index sequence of 8 bp is added at each end (**Kircher et al. 2012**). This allows the samples to be pooled for the sequencing.

To preserve aDNA-specific damage patterns resulting from the deamination of cytosine in all amplification steps, polymerases were used which were able to read these deaminated cytosine bases as uracil and therefore incorporated thymine, such as Pfu Turbo Cx Hotstart Polymerase (Agilent Technologies). To review the success of library generation and the subsequent indexing PCR, qPCR was performed. Subsequently, the libraries were amplified to sufficient concentration for the sequencing, which was carried out at the Max Planck Institute for the Science of Human History in Jena.

Shotgun sequencing and mtDNA enrichment

For the initial screening, all libraries were prepared for shotgun sequencing. For the libraries of the petrous bones, a specific target enrichment for human mitochondrial DNA was also performed after a slight modification of the protocol of **Maricic et al.** (2010) as described in **Furtwängler et al.** (2018). The modifications mainly affect the generation of the baits from modern DNA. As described in the original version, the human mtDNA is divided into three parts and amplified via long-range PCR. The products are subsequently shared with ultrasound. The resulting fragments are then transformed into a bait library by adding artificial adapters similar in sequence to the adapter used in **Fu et al.** (2013) to the fragment ends. The hybridisation is performed prior to the binding of the baits to the magnetic beads. After binding the hybridisation compounds to the beads via streptavidin-biotin bonds and washing them, the enriched fragments are used for amplification

without prior elution from the baits to minimise the loss of sequences. The performance of these modifications was compared empirically to the version of **Maricic et al.** (2010), and the results were comparable in terms of DNA yield and enrichment efficiency. Applied to organisms other than humans, such as uncultivable bacteria like *Mycobacterium leprae*, where the supply of modern DNA extracts for bait generation is limited, the generation of immortalised bait libraries enables indefinite use.

Extracts that tested positive for relatively high amounts of endogenous human DNA, high frequencies of typical aDNA damage, completely reconstructed mitochondrial genomes and low levels of mtDNA contamination were made into two additional libraries including treatment with UDG (**Rohland et al 2015**) to remove the damage from inside the fragments while preserving the terminal C to T substitutions.

Genome-wide enrichment for pathogen DNA

For a detailed comparison of the methods for the enrichment for pathogen DNA, existing libraries of samples that tested positively for *Mycobacterium leprae* and *Treponema pallidum* were amplified and captured with array capture, in-solution capture with immortalised DNA bait libraries originally cleaved from arrays, and the myBaits® kit involving RNA baits. All capture experiments were performed with three independent replicates. The in-solution capture with probes derived from arrays was performed at the MPI in Jena. The first replicate of myBaits® capture was conducted at the post-amplification aDNA laboratory at Kiel University, while array capture and the two subsequent replicates of the myBaits® capture were done in the post-amplification laboratory in Tübingen.

Genome-wide enrichment for human DNA

The initial library with non-UDG treatment and the two additional libraries with partial UDG treatment of the Neolithic individuals in paper III were then used for genome-wide capture at the University of Tübingen. For this enrichment, 1.2 million SNPs across the entire human genome are selected (**Patterson et al. 2012**).

The procedure of this in-solution capture is described in **Fu et al.** (2013). This capture method utilises probes originally cleaved from Agilent arrays which were subsequently transformed into a bait library. The enriched libraries were sent to the MPI in Jena for sequencing.

Raw data processing

The raw data from sequencing was processed with the GUI-based EAGER pipeline (**Peltzer et al. 2016**). The EAGER pipeline was used in all publications presented regarding the pre-processing of the reads (**Peltzer et al. 2016**), including mapping with BWA (**Li & Durbin, 2009**), duplicate removal with DeDup (**Peltzer et al. 2016**) and generation of damage plots with mapDamage (**Ginolhac et al. 2011**). The only exception is that a newer version of EAGER was used in paper II, based on *nextflow* programming (https://github.com/nf-core/eager, **Fellows-Yates et al. 2020**). Analyses specific to the individual studies are described in the respective supplementary information in the Appendix.

4. Results

4.1 New strategy in authenticating genomic DNA from ancient female individuals

1. **Furtwängler, A**., Reiter E., Neumann G.U., Siebke I., Steuri N., Hafner A., Lösch S., Anthes N., Schuenemann V.J., Krause J. "Ratio of mitochondrial to nuclear DNA affects contamination estimates in ancient DNA analysis." *Scientific reports* 8.1 (2018): 14075.

Synopsis

Paper I presents a statistical analysis of the differences in the ratio of mitochondrial to nuclear DNA in diverse tissues typically used for ancient DNA retrieval. Furthermore, the impact of these differences on the contamination estimates is investigated. The aim of this statistical comparison was to find a strategy for evaluating the validity of the extrapolation from mitochondrial contamination estimates to the nuclear genome in female individuals where the estimation of X chromosomal contamination is not possible.

To achieve this, the ratio between mtDNA and nDNA was analysed in 317 samples. To obtain a dataset of this size, new data was generated and combined with data from previous publications. The results were that the ratio of mitochondrial to nuclear DNA was the smallest in petrous bones, meaning that these contain large amounts of nuclear DNA relative to their mtDNA content. The ratio was significantly larger in teeth and other bony skeletal parts than in the petrous bones.

A comparison of the mitochondrial DNA contamination estimates and the contamination estimates derived from X chromosomes in the samples from male individuals showed that the values of the estimates were the most similar in cases where the ratio was small. Therefore, large amounts of nDNA relative to mtDNA make the extrapolation from mtDNA contamination estimates to nDNA more reliable. The main conclusion of the paper, therefore, was that in order to use the

mtDNA contamination estimates from female individuals as the sole authentication criterion besides damage patterns, the mtDNA to nDNA ratio should be taken into consideration.

4.2 Identification of the best method for target enrichment in ancient DNA

2. **Furtwängler A.**, Neukamm J., Böhme L., Reiter E., Vollstedt M., Arora N., Singh P, Cole S.T., Knauf S., Calvignac-Spencer S., Krause-Kyora B., Krause J., Schuenemann V.J., Herbig A. "Comparison of target enrichment strategies for ancient pathogen DNA" *Manuscript*

Synopsis

Paper II presents a detailed statistical analysis of three different methods for the specific target enrichment of ancient pathogen DNA. A comparison was made of array capture, enrichment with DNA probes cleaved from arrays, and the myBaits® kit utilising RNA probes from Arbor Biosciences. For this purpose, existing libraries of published studies on ancient and modern DNA from *Mycobacterium leprae* and *Treponema pallidum* were captured with these three different enrichment protocols in three independent replicates. The resulting dataset enabled statistical analysis regarding DNA yield, specificity and reproducibility.

The DNA yield – the efficiency of the enrichment – was assessed by comparing features such as mean coverage, enrichment factor and the percentage of the genome covered at least fivefold. The commercial myBaits® kit with two repetitions of the hybridisation shows good performance in these features for ancient and modern samples of both *M. leprae* and *T. pallidum*. In most cases, the myBaits® kit with two repetitions also shows the highest percentage of *M. leprae* and *T. pallidum* reads when compared to general mycobacterial and treponemal reads and therefore the highest specificity. Values of the percentage of specific reads were in average 1.5 times higher than in the other methods. For the reproducibility of the capture performance, a high throughput approach with robot assistance and little manual

handling seems to be advantageous as performed for the in-solution capture with DNA probes derived from arrays resulting in the least differences between independent replicates.

Overall, the myBaits® kit with two rounds of hybridisation can be recommended due to the high DNA yield and high specificity for the work with ancient pathogen DNA. Besides its good performance, its major disadvantages are higher costs especially in case of large numbers of samples that need to be enriched. In these cases, the in-solution capture with DNA baits can be recommended.

4.3 Population genetics analysis of Late and Final Neolithic individuals from what is now Switzerland

3. **Furtwängler A.**, Rohrlach A.B., Lamnidis T.C., Papac L., Neumann G.U., Siebke I., Reiter E., Steuri N., Hald J., Denaire A., Schnitzler B., Wahl J., Ramstein M., Schuenemann V.J., Stockhammer P.W., Hafner A., Lösch S., Haak W., Schiffels S., Krause J. "Ancient genomes reveal social and genetic structure of Late Neolithic Switzerland." *Nature Communications* (2020)

Synopsis

Paper III is about a detailed population genetic analysis of Neolithic and Early Bronze Age individuals from what is now Switzerland. The focus of these analyses is the genetic turnover in the Central European population around 4,700 BP. To supplement the picture, samples from the Hegau region in Southern Germany and the Alsace region in France were also included.

This genetic turnover can be observed throughout Europe at relatively similar times. It is marked by the arrival of an additional ancestry component from the Pontic steppe, which can be found in almost all present-day Europeans in addition to components of the Western Hunter-Gatherer and Early Farmers from Western Anatolia. Large population genetics studies have investigated this turnover in areas

like the Iberian Peninsula or the British Isles, while initial studies provided data for regions such as the Middle Elbe-Saale region in Eastern Germany. In this dissertation, we therefore present data for a region connecting these different points to complete the picture.

Furthermore, through a very dense sampling over the entire time span of this turnover, further insights into the detailed processes of this population change were gathered, and it was revealed that this period of genetic change lasted over several generations. We were also able to identify female individuals with zero steppe ancestry dating to approximately one thousand years after the arrival of this component in Central Europe. Furthermore, hints of a non-local origin of these women suggest that female exogamy involving women with zero or little steppe ancestry results in the detected decrease in the relative amount of steppe ancestry seen in the centuries after its arrival.

5. Discussion

In the following section, the results of the three studies presented are discussed in connection with each other and with similar studies, some of which I also co-authored (see list of publications). This section will first discuss how the methodological guidelines established during this dissertation can improve the efficiency of DNA recovery and authentication, and then address how the genome-wide data of Neolithic and Early Bronze Age individuals from present-day Switzerland and the surrounding regions which results from the application of these improved methods contributes to our understanding of the genetic transition in the Final Neolithic in Central Europe.

5.1 DNA preservation and target enrichment (papers I, II & III)

The strong importance that specific target enrichment represents in aDNA research is caused by the small amount of DNA that is left after several centuries or even millennia. Furthermore, large quantities of background DNA originating from the environment, such as soil-inhabiting or flesh-decomposing microorganisms, cover the slight signal of the endogenous DNA of the target organism. The specific enrichment before sequencing is therefore often inevitable.

Before enrichment, samples are screened for the preservation of DNA. Even though this is relatively low in most aDNA samples, differences in the amount of endogenous DNA are detectable. As expected from observations published previously (**Gamba et al. 2014, Pinhasi et al. 2015**), the petrous bones from the skeletons investigated in papers I and III show relatively good DNA preservation, especially compared to teeth and other bone fragments from the same individuals. The lowest values of percentages of endogenous DNA in teeth are far below 0.1%; highest values were achieved in petrous bones from the site of Niederried, with 84% of endogenous DNA. This individual dates to 6,458 – 6,362 BP and is therefore one of the oldest skeletons in this entire dissertation. These observations show the great

importance of the petrous bone for aDNA studies, as it allows sufficient DNA retrieval from very old specimens.

These differences in the abundance of DNA throughout the skeleton (**Gamba et al. 2014, Pinhasi et al. 2015, Hansen et al. 2017**) were investigated further during this dissertation. In paper I, teeth and petrous bone were compared in terms of the amount and composition of their human DNA content. Here, it was important to compare different tissue types, such as different skeletal parts, from the same individuals otherwise site-specific differences would have biased this analysis. Such samples from different tissue types were given in the newly generated data of the Swiss samples, especially from the individuals from the Oberbipp site. Furthermore, by including data from other studies, such as **Gamba et al.** (2014), **Günther et al.** (2015), **Allentoft et al.** (2015), **Broushaki et al.** (2016), **Gallego-Llorente et al.** (2016) and **Schuenemann et al.** (2017), and conducting a detailed statistical analysis of the large dataset thus generated, amongst other results, the findings of higher endogenous DNA content in petrous bones than in other skeletal parts could be confirmed and therefore match those of the studies mentioned above.

After an initial screening using shotgun sequencing and enrichment for mtDNA, samples with more than one percent endogenous DNA and the full reconstructed mitochondrial genome were selected for the nuclear enrichment, as in these cases, chances of retrieving high quality data after nuclear capture are good. Subsequently, only individuals that covered more than 56,000 of the 1.2 million SNPs, enriched for, were used for population genetic analysis, to ensure that the differences detected were statistically significant. The generation of such a high-quality dataset of ancient genomes was mainly facilitated by the use of petrous bones as a source together with specific target enrichment.

Despite the differences detected in the amount of endogenous DNA in the different sample types available, specific enrichments of target regions of the genomes are expedient in all cases. There are different approaches available for target enrichment in aDNA, most of which are based on hybridisation reactions. The first methods invented, such as primer extension capture (PEC), were restricted to small

regions of the genome (**Briggs et al. 2009**). Protocols using probes generated via long-range PCR are likewise limited to only parts of the genome, like bacterial plasmids or mitochondrial genomes (**Maricic 2010**). Therefore, more sophisticated approaches for genome-wide capture were developed (**Hodges et al. 2007, Burbano et al. 2010, Fu et al. 2013**).

Since the efficiency of the target enrichment later influences the mean coverage and the portion of the genome that can be reconstructed, it therefore directly affects the quality of the retrieved data. To prevent cost-intensive repetitions of the capture experiments, it is therefore crucial to determine the best-suited approaches for the different sample types. Furthermore, in ancient DNA it is often a stroke of luck to find positive samples and the material is therefore limited, making endless repetitions of experiments difficult. These precious samples therefore require not only the most efficient method but also the most reliable one. Methods involving robot assistance and less manual work have been shown to provide the most consistent results as shown in paper II.

In paper II, the results further indicate that the myBaits® approach is highly suitable because it is very efficient for the enrichment of ancient pathogen DNA. In this commercial kit, the probes are made of RNA. Compared to the array capture, the probes of the myBait® kit are longer and more movable. This could increase the chances of the baits finding complementary fragments of the library because 1) there is a greater chance of overlapping bases between bait and target fragment, and 2) if the probes are more mobile, they have a greater chance of encountering them.

For most capture methods based on hybridisation, two successive rounds are recommended for use in aDNA. In the first systematic and statistic comparison in Paper II, it was shown that this does indeed increase the efficiency and thus the yield of sequences, such as the mean coverage, for example. However, this procedure causes a loss of library complexity. Therefore, with regard to aDNA and not only in the case of modern samples, it must be deliberated whether this second round of hybridisation and the associated increase in mean coverage compensate for this loss of complexity. For well-preserved samples in particular, a single round of capture

and slightly deeper sequencing may be sufficient, and also preserves the library complexity, enabling a greater variety of downstream analysis.

Besides DNA yield and reproducibility of target enrichment methods, another important characteristic is the capture's specificity. This is especially important for microbial DNA. One very good example is the genus of mycobacteria. Many studies on ancient DNA of the human pathogenic species *M. leprae* and *M. tuberculosis* struggled with the contamination of mycobacteria inhabiting the soil. If these environmental species are too closely related to the target organism, their DNA will be enriched as well, and, furthermore, cannot always be filtered out during bioinformatics processing by applying stringent mapping criteria (**Vågene et al. 2018**). If this is the case, these contaminating fragments can influence downstream analysis. Of the three methods tested in paper II, the myBaits® capture exhibits the highest proportion of reads specific to the target organisms *T. pallidum* and *M. leprae* as determined using the MALT algorithm (**Vågene et al. 2018**). However, this does not exclude the fact that unspecific reads are still present and that further authentication criteria must be applied to the data during bioinformatics analysis, for instance, determining whether numbers of heterozygous positions are in reasonable magnitudes throughout the reconstructed genomes as an indicator for the presence of more than one strain.

High enrichment efficiency and constant, reliable results in the case of robot-assisted approaches are therefore the main advantages of in-solution capture methods. As shown in paper II, methods based on in-solution hybridisation are advantageous, especially for the enrichment of aDNA. In-solution capture with DNA baits is used throughout papers I and III presented in this dissertation to ensure the retrieval of sufficient DNA to generate genome-wide data. Furthermore, in-solution capture with DNA baits was chosen since it allows the usage of the largest number of unique probes in one experiment. In the case of the ~1.2 M SNPs enriched for in papers I and III 4.8 M unique probes are used. The standard probe panels of the myBaits® capture are limited to 200,000 unique probes.

Applied to Neolithic and Bronze Age samples from Eurasia, the method of enrichment with probes derived from arrays enabled the detection of *Salmonella enterica* subsp. *enterica* with an age up to 6,500 years (**Key et al. 2020**). One tooth in this study which tested positive originated from the megalithic burial site at Oberbipp in Switzerland, dating to the 6th millennium BP. More teeth, bone fragments and petrous bones from this burial site were also used in papers I and III for analysis of the human DNA. The subspecies identified in the study of **Key et al.** (2020) contains many strains pathogenic to humans and animals. The strains in the ancient samples are all from a branch that has not yet been characterised, while younger strains cluster with the modern *S. enterica* paratyphi C, providing a great opportunity to study the evolutionary history and trajectory of this pathogen. This has been made possible by specific target enrichment. Finding DNA of this bacteria in the Oberbipp individual indicates that this individual suffered from sickness and likely died due to the infection.

In paper III, in-solution enrichment with probes derived from arrays is also used, but here it is applied to enrich for human DNA. In this capture method, 1.2 million SNP positions spanning the whole human genome are targeted (**Fu et al., 2013, Mathieson et al. 2015**). These SNPs were selected because they are informative for population genetics analysis and can therefore be used to reconstruct human population history and admixture events (**Patterson et al. 2012**).

As such approaches are started with well-preserved samples with high amounts of endogenous DNA, the capture efficiency can be expected to be low. Since it cannot be enriched over one hundred per cent, the enrichment factor is consequently limited. The enrichment factor is hereby calculated by dividing the percentage of endogenous DNA after enrichment by the percentage of endogenous DNA from the shotgun sequencing. For particularly well-preserved samples with endogenous DNA portions above about fifty per cent, target enrichment makes little sense, especially for pathogen DNA (**Schuenemann et al. 2013**). Shotgun sequencing would be preferable in such cases, since target enrichment can introduce a bias towards the reference used for probe design.

Paper III also identified samples with high percentages of endogenous DNA, some revealing over eighty per cent endogenous human DNA in the initial screening. However, it is still advisable to carry out a target enrichment for human DNA, because the special protocol used does not enrich the entire genome but only certain positions – namely the 1.2 million SNPs mentioned above. The proportion of the fragments and thus the mean coverage at these target positions can be increased by the capture even in well-preserved samples. This would be particularly advantageous if diploid genotype calls were required for some analyses and the pseudo-haploid calls that are often used in aDNA were insufficient. The enrichment increases the coverage on the selected positions and is very advantageous if analyses are based on these positions. On the other hand, one does of course lose information between these regions that are enriched. If close related populations should be studied the 1.2 million SNPs would not reflect enough differences for significant results, only rare variants that occur with around one per cent in the populations could potentially be used (**Schiffels et al. 2016**). For these analyses only shotgun sequencing without enrichment can be used since such shared rare variants that vary from population to population are not included in the capture. However, for the research questions addressed in paper III the 1.2 million SNPs are sufficient and the differences that can be detected with them are informative.

Despite the very good level of overall preservation, there were a few samples that failed during the aDNA analysis. These included the femora from the Oberbipp burial site which were identified by systematic radiocarbon dating to belong to a second later phase of use. Dating to exactly the time when the Steppe ancestry first appeared (**Steuri et al. 2019**), these individuals are likely to have been admixed already. Due to the low DNA preservation, an investigation of those individuals was not possible but would probably have enabled a direct comparison of the same subpopulation over time, and before and after the transition.

5.2 Authentication and assessment of data quality (papers I & III)

When generating genetic data from archaeological material, gaining enough DNA and therefore high coverage is not the only aspect concerning the quality of the data. If conclusions from this genetic data are drawn, it must furthermore be ensured that the sequences are authentic. Besides characteristics typical of aDNA, such as damage, which are used to determine whether the genetic material is old, relative portions of DNA from other individuals are estimated in human samples.

Authentication usually starts with the initial screening, when aDNA-specific damage is checked. The older the samples, the higher the frequency of deamination on the fragment ends, which was shown to be time-dependent (**Sawyer et al. 2012**). The majority of the samples analysed during this dissertation date to the 5th to 3rd millennia BP and are therefore several thousand years old. Hence, the damage rate caused by deamination is expected to reach between at least fifteen and twenty per cent of all fragments. This threshold is reached by all samples with relatively high amounts of endogenous human DNA, that is, most of the petrous bones but only a few of the teeth, which tend to contain only little amounts of endogenous human DNA. On the one hand, these high frequencies of damage typical of aDNA are favourable, as it proves the age of the sequences, but on the other hand, these modifications of the fragment hinder downstream analysis.

To reduce the impact of these base substitutions on the analysis, the damage is already corrected for during library generation before sequencing. During the UDG treatment, the deaminated cytosines are cleaved and the resulting abasic sites are removed (**Briggs et al. 2010**). To still allow authentication and filtering for ancient reads a version of this protocol is available that only removes the damage inside the fragments and leaves it at the ends (**Rohland et al. 2015**). For screening, all the samples in papers I and III were non-UDG treated, and libraries which were solely for the purpose of population genetics analysis received partial UDG treatment, in order to always be able to ensure that the damage pattern was detectable.

In addition to checking and correcting for damage typical of aDNA, mitochondrial and nuclear contamination rates on the X chromosomes were assessed. Based on the appearance of heterozygous positions on haploid regions, the amount of DNA from additional individuals was estimated. An acceptable threshold of five per cent is set for both nuclear and mitochondrial contamination. Below these limits, the wrong basecalls caused by the contaminating DNA are so rare that they can be neglected.

Nuclear contamination estimates based on heterozygosity on haploid regions can only be assessed in male individuals. Datasets of female individuals are therefore often reduced to sequences carrying aDNA-specific damage, resulting in a great loss of data. Using only the mtDNA contamination rates for females to assess the quality of the nuclear data carries the risk of under- or overestimating the actual contamination in the nuclear sequences. One factor affecting the differences between the mitochondrial and nuclear contamination rates is clearly the ratio of the amount of mitochondrial to nuclear DNA, as shown in paper I. For this reason, the mitochondrial to nuclear ratio provides an opportunity to further assess the quality of the sequence data in the case of female individuals where only mtDNA contamination rates are available.

Furthermore, this ratio differs in samples of different skeletal parts. The lowest and hence most suitable values for using the mtDNA contamination rates as the sole criterion for assessing contamination in female individuals can be found in petrous bones. This fact was demonstrated with statistical significance from the samples in paper I and was also confirmed in the dataset of the individuals from paper III.

Paper I followed the population genetics study of **Schuenemann et al.** (2017) and also used data published within this article. In which, more than 151 mummified human remains from the archaeological site Abusir-el Meleq in Egypt were screened for human DNA content. In 90 of these cases it was possible to reconstruct entire mitochondrial genomes, suggesting relatively good DNA preservation. Furthermore, mitochondrial contamination rates were detected to be in an acceptable range –

below five per cent, for most samples. These results made this set of individuals promising for large-scale population genetics analysis based on nuclear DNA. However, all samples were retrieved from teeth and skull bones other than petrous bones, and the mitochondrial to nuclear DNA ratio was high in the data of all individuals. After enrichment for nuclear DNA, the contamination estimation based on the X chromosome was far above acceptable thresholds for the vast majority of those individuals. These observations make this study a perfect example of the application of the mitochondrial to nuclear ratio to assess the quality of the data after an initial screening utilising shotgun sequencing and mitochondrial capture.

Starting with these considerations for DNA retrieval and authentication, which form the basis for successful studies on human remains, investigations of the demographic history of present-day Switzerland were begun. The undisturbed megalithic burial site of Oberbipp in the Swiss canton of Bern, which was recently excavated using the newest techniques, not only provided material for studying the suitability of different skeletal elements for aDNA analysis and methodological improvements, but also provided the basis for detailed investigations of population history (**Ramstein et al. 2012, Ramstein et al. 2014, Siebke et al. 2018**). One of the reasons why the human remains collected are special is that they were retrieved under almost ideal conditions with the excavators wearing face masks and gloves to prevent contamination with their own DNA. Furthermore, the high number of petrous bones retrieved makes this anthropological assemblage precious for palaeogenetics studies.

5.3 Population history of present-day Switzerland (paper III)

The region of present-day Switzerland, located right in the middle of Central Europe and spanning various landscapes, such as parts of the Alps, the Swiss plateau and the Jura mountains, is relevant when studying past population movements within and into Central Europe.

In addition to the dolmen of Oberbipp, excavated in 2012 and contributing almost 30 individuals to the analysis (**Ramstein et al. 2012, Ramstein et al. 2014**), it is mainly skeletons excavated at the beginning of the 20th century which were studied here. For these remains, however, the archaeological context is often lost. Nevertheless, by combining these early finds with the few more recently excavated finds and applying systematic radiocarbon dating (**Steuri et al. 2019, Siebke et al. 2019**), it was possible to obtain a dataset that spans the period from approximately 6,000 to 3,500 BP with a relatively high density of individuals. Such a dataset is exceptional because it allows fine-scale investigation of population history and therefore provides new insights into the genetic history of past populations.

The time span investigated is of particular interest since initial studies from 2015 revealed a sweeping transition in the population of Central Europe (**Haak et al. 2015, Allentoft et al. 2015**). An entirely new genetic component arrives in Western Eurasia, originating from the Pontic-Caspian steppe – a component that is found in almost all European populations today.

Haak et al. (2015) postulate a sudden migration wave based on the strong genetic affinity of Final Neolithic individuals associated with the Corded Ware Complex to individuals from the Pontic steppe associated with the Yamnaya culture, whereas later-dating individuals associated with the Bell Beaker phenomenon show still strong but fewer similarities to these steppe individuals. Therefore, the individuals between the Late to Final Neolithic did not slowly accumulate genetic similarities to the steppe herders, but this genetic connection is likely to have come rather suddenly.

Similar patterns were observed in the individuals studied in paper III. While modelling all individuals in the study as a mixture of the three components which present-day Europeans possess, namely Western Hunter-Gatherers, Early Farmers from Anatolia and ancestry from steppe pastoralists, differences in the genetic makeup can be observed over time. There is a particularly sharp increase in the steppe ancestry component around 4,700 BP. This rapid increase is followed by a decline in the relative amount of steppe ancestry which continued for several

centuries, probably caused by the further admixture of individuals with steppe ancestry and individuals who did not yet carry this component. **Mittnik et al**. (2019) have already proposed that admixture between groups with differences in the relative amount of steppe-like ancestry are causing this decline.

In Mittnik et al. 2019, the social system in the Lech valley in Southern Germany was investigated. The individuals included in this study date slightly younger than those in paper III – to the Early Bronze Age, for example. Based on the family structures, mating habits and stable isotope analysis, **Mittnik et al.** (2019) propose a patrilocal society with females being more mobile than males. The time span investigated lies exactly in the period of the decline in steppe ancestry. It can therefore be suggested that these mobile women carried less of the steppe component, and may have caused the decline of this ancestry in the population, a fact which is also confirmed by a sex-bias detected by **Mittnik et al.** (2019) in early individuals as well as in the Final Neolithic populations from Switzerland in paper III. This hypothesis of mobile women with less steppe-like ancestry was strongly supported in paper III, where women without any steppe ancestry dating approximately one thousand years after its arrival were identified.

As with the population of the Lech valley which was studied, for the individuals from Switzerland and surrounding regions in paper III, a patrilocal society can be suggested based on the reconstructed family trees. These women without steppe ancestry were therefore most likely not local. Strontium isotope analysis does not contradict this, and in one case, individual 3 (MX193) from the Spreitenbach burial site associated with the Corded Ware Complex even confirms the foreign origin of this women. However, no concrete conclusions can be drawn as to where they originated from. The Strontium isotopic landscape of Switzerland is so varied that the location of their origin cannot be pinpointed. For this reason, neither can it be ruled out with any certainty that the women even migrated into Switzerland from farther away, such as regions in Italy or the Iberian Peninsula. Nevertheless, a patrilocal society and mobile women were identified as the cause of the decline in steppe ancestry. Moreover, the original population of these women must have been

largely without steppe ancestry. This alone can explain the complete absence of the steppe component in these females. Somewhere in the vicinity of the populations studied in paper III, a population with the genetic makeup of Early European Farmers must therefore have remained unadmixed for several centuries after the arrival of the steppe component in Central Europe. Besides regions in Italy or the Iberian Peninsula, a likely location would be alpine mountain valleys where isolated populations with strong genetic differentiation can be found even today (**Pichler et al. 2009, Berger et al. 2013**).

The pattern of a sudden appearance and steep increase in steppe-like ancestry seems to be the same all over Central Europe, as a comparison of the individuals from Switzerland with those from Great Britain, Iberia and the Middle Elbe-Saale region in Eastern Germany has shown in paper III. Despite the similarities in the process, the timing of the appearance of the new component is different in these regions. It appears rather contemporaneous in Eastern Germany (**Haak et al 2015**) and Switzerland and then later in Great Britain (~ 4450 BP onwards, **Olalde et al. 2018**) and even later in Iberia (~4500–4000 BP, **Olalde et al. 2019**). Close genetic similarities between individuals associated with the Corded Ware Complex from the Swiss region and the Middle Elbe-Saale region in Eastern Germany suggest that people from this cultural context constitute a relatively homogeneous population genetically across large parts of Central Europe.

To investigate the processes forming this population further, dating methods for the admixture between the steppe component and the indigenous Early Farmers were applied, and a pattern was revealed that was more complex than the single pulse-like migration wave previously suggested (**Haak et al. 2015**). If the arrival of the steppe component was one single pulse, individuals with younger radiocarbon dates would show older admixture dates. No such trend was visible in the data, however, therefore the input from the Pontic-Caspian steppe was still ongoing when the relative amount of the component in the Central European individuals was already starting to decline. Here, paper III clearly contributed to our understanding of the dynamics of this major population transition in Central Europe.

Initially, the steppe component was mainly brought to Central Europe by males. Very distinctive patterns of the dispersal of specific haplogroups can therefore be observed on the Y chromosome. Starting from the Middle and Late Neolithic,most male individuals carry Y chromosomal haplogroups belonging to the macro-haplogroup G2. Haplogroups belonging to this clade were widespread in Central Europe during the Neolithic. In the Final Neolithic, in the only burial site associated with the Corded Ware Complex, that is, the collective burial site of Spreitenbach, all males carry haplogroups belonging to I2. Only in the later burial sites, like Singen in the Hegau region of Southern Germany, do haplogroups belonging to R1b appear, and one male with R1a (approximately 3000 – 2500 BP) was even identified. This supports the hypothesis that these haplogroups became the dominant Y chromosome lineages after the studied time span probably during the period of the Bell Beaker phenomenon (**Olalde et al 2018**). Furthermore, it indicates ongoing gene flow from the Pontic Steppe into Central Europe after the initial appearance of steppe-related ancestry in the region.

Another great opportunity to study the population history of present-day Switzerland further is presented by the POPRES dataset, which also includes the subjects of the CoLause study (**Nelson et al. 2008**). This population-based study of modern Europeans includes almost 3000 individuals, of which around 1000 are of Swiss origin. Back in 2008, **Novembre et al.** already used this dataset to show that the plotting of genetic distances between modern European individuals can be used to reproduce the map of Europe, and, furthermore, that strong genetic differentiation can be found between the three language groups in Switzerland. The individuals in the POPRES dataset are genotyped using the Affymetrix 500K SNP panel (**Nelson et al. 2008**) and therefore at least partially overlap with the 1.2 million SNPs for which the libraries of the ancient individuals were enriched. In paper III, the ancient genomes from Switzerland are compared with the modern individuals from the POPRES. As expected, due to various historical events, such as the migration period and the globalisation within the last century, for example, simple population continuity between the populations of the Final Neolithic and today can be rejected,

showing clearly that the history of Europe is one of migration and admixture. To close the gap left between 3500 BP and today samples, from later periods than studied in paper III are needed and probably will help to understand the processes that led to the present genetic structure of this region.

Overall, paper III contributed to the understanding of the dynamics of this transition in the Final Neolithic to the onset of the Early Bronze Age.

6. Outlook

The area of aDNA research has changed a great deal over the past few years. Technological improvements have paved the way for genome-wide studies on ancient human remains with hundreds of individuals and for obtaining DNA from human remains from regions where conditions are not favourable for DNA preservation. As a result, it has been possible to find answers to many questions about the peopling of large parts of the world and the processes that shaped the genetic structure of modern populations.

Today, the trend is moving towards fine-scale analysis of microregions such as the Early Bronze Age population of the Lech valley, as studied by **Mittnik et al.** (2019), or that of the Final Neolithic population of what is now Switzerland, as examined in paper III. As a result, detailed questions about the social structure of past populations can be addressed. Furthermore, by focusing on small genomic differences – shared rare variants –, more recent demographic changes involving populations that are genetically quite similar can be reconstructed.

In the same way, the transition at the end of the Neolithic is studied using a large dataset with dense temporal sampling in paper III, also the period before the Late and Final Neolithic provides opportunities for further research. From this time period, present-day Switzerland provides a rich material culture and several Chamblandes-type burial sites. In addition to reconstructing social structure, questions about how and when Anatolian ancestry arrived in what is now Switzerland might also be addressed.

Even though the improvement of specific techniques used in aDNA is ongoing, further methodological developments could help to obtain DNA from even older samples or those that are poorly preserved due to environmental conditions, and therefore provide meaningful insights into the ancestors of humans, animals and their pathogens.

References

Allentoft, Morten E.; Collins, Matthew; Harker, David; Haile, James; Oskam, Charlotte L.; Hale, Marie L. et al. (2012): The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. In: *Proceedings. Biological sciences* 279 (1748), S. 4724–4733. DOI: 10.1098/rspb.2012.1745.

Allentoft, Morten E.; Sikora, Martin; Sjögren, Karl-Göran; Rasmussen, Simon; Rasmussen, Morten; Stenderup, Jesper et al. (2015): Population genomics of Bronze Age Eurasia. In: *Nature* 522 (7555), S. 167–172. DOI: 10.1038/nature14507.

Andrades Valtueña, Aida; Mittnik, Alissa; Key, Felix M.; Haak, Wolfgang; Allmäe, Raili; Belinskij, Andrej et al. (2017): The Stone Age Plague and Its Persistence in Eurasia. In: *Current biology : CB* 27 (23), 3683-3691.e8. DOI: 10.1016/j.cub.2017.10.025.

Anthony, David W. (2007): The horse, the wheel, and language. How Bronze-Age riders from the Eurasian steppes shaped the modern world. Princeton, N.J: Princeton University Press. Online verfügbar unter http://www.jstor.org/stable/10.2307/j.ctt7sjpn.

Arora, Natasha; Schuenemann, Verena J.; Jäger, Günter; Peltzer, Alexander; Seitz, Alexander; Herbig, Alexander et al. (2016): Origin of modern syphilis and emergence of a pandemic Treponema pallidum cluster. In: *Nature microbiology* 2, S. 16245. DOI: 10.1038/nmicrobiol.2016.245.

Berger, Burkhard; Niederstätter, Harald; Erhart, Daniel; Gassner, Christoph; Schennach, Harald; Parson, Walther (2013): High resolution mapping of Y haplogroup G in Tyrol (Austria). In: *Forensic science international. Genetics* 7 (5), S. 529–536. DOI: 10.1016/j.fsigen.2013.05.013.

Brandt, Guido; Haak, Wolfgang; Adler, Christina J.; Roth, Christina; Szécsényi-Nagy, Anna; Karimnia, Sarah et al. (2013): Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. In: *Science (New York, N.Y.)* 342 (6155), S. 257–261. DOI: 10.1126/science.1241844.

Briggs, Adrian W.; Good, Jeffrey M.; Green, Richard E.; Krause, Johannes; Maricic, Tomislav; Stenzel, Udo; Pääbo, Svante (2009): Primer extension capture: targeted sequence retrieval from heavily degraded DNA sources. In: *Journal of visualized experiments : JoVE* (31), S. 1573. DOI: 10.3791/1573.

Briggs, Adrian W.; Stenzel, Udo; Johnson, Philip L. F.; Green, Richard E.; Kelso, Janet; Prüfer, Kay et al. (2007): Patterns of damage in genomic DNA sequences from a Neandertal. In: *Proceedings of the National Academy of Sciences of the United States of America* 104 (37), S. 14616–14621. DOI: 10.1073/pnas.0704665104.

Briggs, Adrian W.; Stenzel, Udo; Meyer, Matthias; Krause, Johannes; Kircher, Martin; Pääbo, Svante (2010): Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. In: *Nucleic acids research* 38 (6), e87. DOI: 10.1093/nar/gkp1163.

Brotherton, Paul; Haak, Wolfgang; Templeton, Jennifer; Brandt, Guido; Soubrier, Julien; Jane Adler, Christina et al. (2013): Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. In: *Nature communications* 4, S. 1764. DOI: 10.1038/ncomms2656.

Broushaki, Farnaz; Thomas, Mark G.; Link, Vivian; López, Saioa; van Dorp, Lucy; Kirsanow, Karola et al. (2016): Early Neolithic genomes from the eastern Fertile Crescent. In: *Science (New York, N.Y.)* 353 (6298), S. 499–503. DOI: 10.1126/science.aaf7943.

Burbano, Hernán A.; Hodges, Emily; Green, Richard E.; Briggs, Adrian W.; Krause, Johannes; Meyer, Matthias et al. (2010): Targeted investigation of the Neandertal genome by array-based sequence capture. In: *Science (New York, N.Y.)* 328 (5979), S. 723–725. DOI: 10.1126/science.1188046.

Cooper, A.; Poinar, H. N. (2000): Ancient DNA: do it right or not at all. In: *Science (New York, N.Y.)* 289 (5482), S. 1139. DOI: 10.1126/science.289.5482.1139b.

Dabney, Jesse; Knapp, Michael; Glocke, Isabelle; Gansauge, Marie-Theres; Weihmann, Antje; Nickel, Birgit et al. (2013): Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. In: *Proceedings of the National Academy of Sciences of the United States of America* 110 (39), S. 15758–15763. DOI: 10.1073/pnas.1314445110.

Dabney, Jesse; Meyer, Matthias; Pääbo, Svante (2013): Ancient DNA damage. In: *Cold Spring Harbor perspectives in biology* 5 (7). DOI: 10.1101/cshperspect.a012567.

Fellows Yates, James A.; Lamnidis, Thiseas C.; Borry, Maxime; Valtueña, Aida Andrades; Fagernäs, Zandra; Clayton, Stephen et al. (2020): Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager.

Fu, Qiaomei; Meyer, Matthias; Gao, Xing; Stenzel, Udo; Burbano, Hernán A.; Kelso, Janet; Pääbo, Svante (2013): DNA analysis of an early modern human from Tianyuan Cave, China. In: *Proceedings of the National Academy of Sciences of the United States of America* 110 (6), S. 2223–2227. DOI: 10.1073/pnas.1221359110.

Fu, Qiaomei; Posth, Cosimo; Hajdinjak, Mateja; Petr, Martin; Mallick, Swapan; Fernandes, Daniel et al. (2016): The genetic history of Ice Age Europe. In: *Nature* 534 (7606), S. 200–205. DOI: 10.1038/nature17993.

Gallego-Llorente, M.; Connell, S.; Jones, E. R.; Merrett, D. C.; Jeon, Y.; Eriksson, A. et al. (2016): The genetics of an early Neolithic pastoralist from the Zagros, Iran. In: *Scientific reports* 6, S. 31326. DOI: 10.1038/srep31326.

Gamba, Cristina; Jones, Eppie R.; Teasdale, Matthew D.; McLaughlin, Russell L.; Gonzalez-Fortes, Gloria; Mattiangeli, Valeria et al. (2014): Genome flux and stasis in a five millennium transect of European prehistory. In: *Nature communications* 5, S. 5257. DOI: 10.1038/ncomms6257.

Ginolhac, Aurelien; Rasmussen, Morten; Gilbert, M. Thomas P.; Willerslev, Eske; Orlando, Ludovic (2011): mapDamage: testing for damage patterns in ancient DNA sequences. In: *Bioinformatics (Oxford, England)* 27 (15), S. 2153–2155. DOI: 10.1093/bioinformatics/btr347.

Green, Richard E.; Krause, Johannes; Ptak, Susan E.; Briggs, Adrian W.; Ronan, Michael T.; Simons, Jan F. et al. (2006): Analysis of one million base pairs of Neanderthal DNA. In: *Nature* 444 (7117), S. 330–336. DOI: 10.1038/nature05336.

Günther, Torsten; Valdiosera, Cristina; Malmström, Helena; Ureña, Irene; Rodriguez-Varela, Ricardo; Sverrisdóttir, Óddny Osk et al. (2015): Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. In: *Proceedings of the National Academy of Sciences of the United States of America* 112 (38), S. 11917–11922. DOI: 10.1073/pnas.1509851112.

Haak, Wolfgang; Balanovsky, Oleg; Sanchez, Juan J.; Koshel, Sergey; Zaporozhchenko, Valery; Adler, Christina J. et al. (2010): Ancient DNA from European early neolithic farmers reveals their near eastern affinities. In: *PLoS biology* 8 (11), e1000536. DOI: 10.1371/journal.pbio.1000536.

Haak, Wolfgang; Forster, Peter; Bramanti, Barbara; Matsumura, Shuichi; Brandt, Guido; Tänzer, Marc et al. (2005): Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. In: *Science (New York, N.Y.)* 310 (5750), S. 1016–1018. DOI: 10.1126/science.1118725.

Haak, Wolfgang; Lazaridis, Iosif; Patterson, Nick; Rohland, Nadin; Mallick, Swapan; Llamas, Bastien et al. (2015): Massive migration from the steppe was a source for Indo-European languages in Europe. In: *Nature* 522 (7555), S. 207–211. DOI: 10.1038/nature14317.

Hafner, Albert; Heitz, Caroline; Stapfer, Regine Barbara (2014): Pile-dwellings of the Neolithic and the Bronze Age in Switzerland. Long-term research and future tasks. In: *von Carnap-Bornheim, Claus (Hg.) Quo vadis? Status and Future Perspectives of Long-Term Excavations in Europe. Schriften des Archäologischen Landesmuseum Schleswig* 10, S. 59–83.

Hafner, Albert; Suter, Peter (2003): Das Neolithikum in der Schweiz. In: *Journal of Neolithic Archaeology* 5.

Hagelberg, Erika; Hofreiter, Michael; Keyser, Christine (2015): Introduction. Ancient DNA: the first three decades. In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 370 (1660), S. 20130371. DOI: 10.1098/rstb.2013.0371.

Hansen, Henrik B.; Damgaard, Peter B.; Margaryan, Ashot; Stenderup, Jesper; Lynnerup, Niels; Willerslev, Eske; Allentoft, Morten E. (2017): Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum. In: *PloS one* 12 (1), e0170940. DOI: 10.1371/journal.pone.0170940.

Hedges, S. B.; Schweitzer, M. H. (1995): Detecting dinosaur DNA. In: *Science (New York, N.Y.)* 268 (5214), 1191-2; author reply 1194. DOI: 10.1126/science.7761839.

Higgins, Denice; Rohrlach, Adam B.; Kaidonis, John; Townsend, Grant; Austin, Jeremy J. (2015): Differential nuclear and mitochondrial DNA preservation in post-mortem teeth with implications for forensic and ancient DNA studies. In: *PloS one* 10 (5), e0126935. DOI: 10.1371/journal.pone.0126935.

Higuchi, R. G.; Wrischnik, L. A.; Oakes, E.; George, M.; Tong, B.; Wilson, A. C. (1987): Mitochondrial DNA of the extinct quagga: relatedness and extent of postmortem change. In: *Journal of molecular evolution* 25 (4), S. 283–287. DOI: 10.1007/bf02603111.

Hodges, Emily; Xuan, Zhenyu; Balija, Vivekanand; Kramer, Melissa; Molla, Michael N.; Smith, Steven W. et al. (2007): Genome-wide in situ exon capture for selective resequencing. In: *Nature genetics* 39 (12), S. 1522–1527. DOI: 10.1038/ng.2007.42.

Hofreiter, M.; Serre, D.; Poinar, H. N.; Kuch, M.; Pääbo, S. (2001): Ancient DNA. In: *Nature reviews. Genetics* 2 (5), S. 353–359. DOI: 10.1038/35072071.

Hublin, Jean-Jacques; Sirakov, Nikolay; Aldeias, Vera; Bailey, Shara; Bard, Edouard; Delvigne, Vincent et al. (2020): Initial Upper Palaeolithic Homo sapiens from Bacho Kiro Cave, Bulgaria. In: *Nature* 581 (7808), S. 299–302. DOI: 10.1038/s41586-020-2259-z.

Jones, Eppie R.; Gonzalez-Fortes, Gloria; Connell, Sarah; Siska, Veronika; Eriksson, Anders; Martiniano, Rui et al. (2015): Upper Palaeolithic genomes reveal deep roots of modern Eurasians. In: *Nature communications* 6, S. 8912. DOI: 10.1038/ncomms9912.

Kircher, Martin; Sawyer, Susanna; Meyer, Matthias (2012): Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. In: *Nucleic acids research* 40 (1), e3. DOI: 10.1093/nar/gkr771.

Knapp, Michael; Clarke, Andrew C.; Horsburgh, K. Ann; Matisoo-Smith, Elizabeth A. (2012): Setting the stage - building and working in an ancient DNA laboratory. In: *Annals of anatomy = Anatomischer Anzeiger : official organ of the Anatomische Gesellschaft* 194 (1), S. 3–6. DOI: 10.1016/j.aanat.2011.03.008.

Krause, Johannes (2010a): From Genes to Gennomes: What is New in Ancient DNA? In: *Mitteilungen der Gesellschaft für Urgeschichte* 19, S. 11–33.

Krause, Johannes; Briggs, Adrian W.; Kircher, Martin; Maricic, Tomislav; Zwyns, Nicolas; Derevianko, Anatoli; Pääbo, Svante (2010b): A complete mtDNA genome of an early modern human from Kostenki, Russia. In: *Current biology : CB* 20 (3), S. 231–236. DOI: 10.1016/j.cub.2009.11.068.

Lazaridis, Iosif; Nadel, Dani; Rollefson, Gary; Merrett, Deborah C.; Rohland, Nadin; Mallick, Swapan et al. (2016): Genomic insights into the origin of farming in the ancient Near East. In: *Nature* 536 (7617), S. 419–424. DOI: 10.1038/nature19310.

Lazaridis, Iosif; Patterson, Nick; Mittnik, Alissa; Renaud, Gabriel; Mallick, Swapan; Kirsanow, Karola et al. (2014): Ancient human genomes suggest three ancestral populations for present-day Europeans. In: *Nature* 513 (7518), S. 409–413. DOI: 10.1038/nature13673.

Li, Heng; Durbin, Richard (2009): Fast and accurate short read alignment with Burrows-Wheeler transform. In: *Bioinformatics (Oxford, England)* 25 (14), S. 1754–1760. DOI: 10.1093/bioinformatics/btp324.

Lindahl, T. (1993): Instability and decay of the primary structure of DNA. In: *Nature* 362 (6422), S. 709–715. DOI: 10.1038/362709a0.

Maricic, Tomislav; Whitten, Mark; Pääbo, Svante (2010): Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. In: *PloS one* 5 (11), e14004. DOI: 10.1371/journal.pone.0014004.

Mathieson, Iain; Lazaridis, Iosif; Rohland, Nadin; Mallick, Swapan; Patterson, Nick; Roodenberg, Songül Alpaslan et al. (2015): Genome-wide patterns of selection in 230 ancient Eurasians. In: *Nature* 528 (7583), S. 499–503. DOI: 10.1038/nature16152.

Meller, Harald; Daim, Falko; Krause, Johannes; Risch, Roberto (Hg.) (2017): Migration und Integration von der Urgeschichte bis zum Mittelalter. 9. Mitteldeutscher Archäologentag vom 20. bis 22. Oktober 2016 in Halle (Saale) = Migration and integration from prehistory to the Middle Ages ; 9th Archaeological Conference of Central Germany October 20-22, 2016 in Halle (Saale). Mitteldeutscher Archäologentag; Archaeological Conference of Central Germany. Halle (Saale): Landesamt für Denkmalpflege und Archäologie Sachsen-Anhalt Landesmuseum für Vorgeschichte (Tagungen des Landesmuseums für Vorgeschichte Halle, Band 17).

Mendum, Tom A.; Schuenemann, Verena J.; Roffey, Simon; Taylor, G. Michael; Wu, Huihai; Singh, Pushpendra et al. (2014): Mycobacterium leprae genomes from a British medieval leprosy hospital: towards understanding an ancient epidemic. In: *BMC genomics* 15, S. 270. DOI: 10.1186/1471-2164-15-270.

Meyer, Matthias; Kircher, Martin (2010): Illumina sequencing library preparation for highly multiplexed target capture and sequencing. In: *Cold Spring Harbor protocols* 2010 (6), pdb.prot5448. DOI: 10.1101/pdb.prot5448.

Mittnik, Alissa; Massy, Ken; Knipper, Corina; Wittenborn, Fabian; Friedrich, Ronny; Pfrengle, Saskia et al. (2019): Kinship-based social inequality in Bronze Age Europe. In: *Science (New York, N.Y.)* 366 (6466), S. 731–734. DOI: 10.1126/science.aax6219.

Myres, Natalie M.; Rootsi, Siiri; Lin, Alice A.; Järve, Mari; King, Roy J.; Kutuev, Ildus et al. (2011): A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. In: *European journal of human genetics : EJHG* 19 (1), S. 95–101. DOI: 10.1038/ejhg.2010.146.

Narasimhan, Vagheesh M.; Patterson, Nick; Moorjani, Priya; Rohland, Nadin; Bernardos, Rebecca; Mallick, Swapan et al. (2019): The Formation of Human Populations in South and Central Asia. In: *Science (New York, N.Y.)* 365 (6457). DOI: 10.1126/science.aat7487.

Nelson, Matthew R.; Bryc, Katarzyna; King, Karen S.; Indap, Amit; Boyko, Adam R.; Novembre, John et al. (2008): The Population Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics research. In: *American journal of human genetics* 83 (3), S. 347–358. DOI: 10.1016/j.ajhg.2008.08.005.

Novembre, John; Johnson, Toby; Bryc, Katarzyna; Kutalik, Zoltán; Boyko, Adam R.; Auton, Adam et al. (2008): Genes mirror geography within Europe. In: *Nature* 456 (7218), S. 98–101. DOI: 10.1038/nature07331.

Olalde, Iñigo; Brace, Selina; Allentoft, Morten E.; Armit, Ian; Kristiansen, Kristian; Booth, Thomas et al. (2018): The Beaker Phenomenon and the Genomic Transformation of Northwest Europe. In: *Nature* 555 (7695), S. 190–196. DOI: 10.1038/nature25738.

Olalde, Iñigo; Mallick, Swapan; Patterson, Nick; Rohland, Nadin; Villalba-Mouco, Vanessa; Silva, Marina et al. (2019): The genomic history of the Iberian Peninsula over the past 8000 years. In: *Science (New York, N.Y.)* 363 (6432), S. 1230–1234. DOI: 10.1126/science.aav4040.

Orlando, Ludovic; Ginolhac, Aurélien; Zhang, Guojie; Froese, Duane; Albrechtsen, Anders; Stiller, Mathias et al. (2013): Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. In: *Nature* 499 (7456), S. 74–78. DOI: 10.1038/nature12323.

Pääbo, S. (1985): Molecular cloning of Ancient Egyptian mummy DNA. In: *Nature* 314 (6012), S. 644–645. DOI: 10.1038/314644a0.

Pääbo, Svante, Russell G. Higuchi, and Allan C. Wilson (1989): Ancient DNA and the polymerase chain reaction: the emerging field of molecular archaeology (Minireview). In: *he Journal of biological chemistry* 264.17, S. 9709–9712.

Patterson, Nick; Moorjani, Priya; Luo, Yontao; Mallick, Swapan; Rohland, Nadin; Zhan, Yiping et al. (2012): Ancient admixture in human history. In: *Genetics* 192 (3), S. 1065–1093. DOI: 10.1534/genetics.112.145037.

Peltzer, Alexander; Jäger, Günter; Herbig, Alexander; Seitz, Alexander; Kniep, Christian; Krause, Johannes; Nieselt, Kay (2016): EAGER: efficient ancient genome reconstruction. In: *Genome biology* 17, S. 60. DOI: 10.1186/s13059-016-0918-z.

Pichler, Irene; Mueller, Jakob C.; Stefanov, Stefan A.; Grandi, Alessandro de; Volpato, Claudia Beu; Pinggera, Gerd K. et al. (2009): Genetic structure in contemporary South Tyrolean isolated populations revealed by analysis of Y-chromosome, mtDNA, and Alu polymorphisms. 2006. In: *Human biology* 81 (5-6), S. 875–898. DOI: 10.3378/027.081.0629.

Pinhasi, Ron; Fernandes, Daniel; Sirak, Kendra; Novak, Mario; Connell, Sarah; Alpaslan-Roodenberg, Songül et al. (2015): Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. In: *PloS one* 10 (6), e0129102. DOI: 10.1371/journal.pone.0129102.

Raghavan, Maanasa; Skoglund, Pontus; Graf, Kelly E.; Metspalu, Mait; Albrechtsen, Anders; Moltke, Ida et al. (2014): Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. In: *Nature* 505 (7481), S. 87–91. DOI: 10.1038/nature12736.

Ramstein, Marianne; Rüttimann, Domenic; Lösch, Sandra (2012): Oberbipp, Steingasse. Ein neolithischer Dolmen. Bern: ArchBE. Archäologie Bern.

Ramstein, Marianne; Schimmelpfennig, Dirk; Lösch, Sandra (2014): Ein neolithischer Dolmen an der Steingasse in Oberbipp. In: *Archäologie Schweiz / Archéologie Suisse / Archeologia Svizzera* 37(3).

Rasmussen, Morten; Guo, Xiaosen; Wang, Yong; Lohmueller, Kirk E.; Rasmussen, Simon; Albrechtsen, Anders et al. (2011): An Aboriginal Australian genome reveals separate human dispersals into Asia. In: *Science (New York, N.Y.)* 334 (6052), S. 94–98. DOI: 10.1126/science.1211177.

Rasmussen, Simon; Allentoft, Morten Erik; Nielsen, Kasper; Orlando, Ludovic; Sikora, Martin; Sjögren, Karl-Göran et al. (2015): Early divergent strains of Yersinia pestis in Eurasia 5,000 years ago. In: *Cell* 163 (3), S. 571–582. DOI: 10.1016/j.cell.2015.10.009.

Renaud, Gabriel; Slon, Viviane; Duggan, Ana T.; Kelso, Janet (2015): Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. In: *Genome biology* 16, S. 224. DOI: 10.1186/s13059-015-0776-0.

Rohland, Nadin; Harney, Eadaoin; Mallick, Swapan; Nordenfelt, Susanne; Reich, David (2015): Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 370 (1660), S. 20130624. DOI: 10.1098/rstb.2013.0624.

Sawyer, Susanna; Krause, Johannes; Guschanski, Katerina; Savolainen, Vincent; Pääbo, Svante (2012): Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. In: *PloS one* 7 (3), e34131. DOI: 10.1371/journal.pone.0034131.

Schiffels, Stephan; Haak, Wolfgang; Paajanen, Pirita; Llamas, Bastien; Popescu, Elizabeth; Loe, Louise et al. (2016): Iron Age and Anglo-Saxon genomes from East England reveal British migration history. In: *Nature communications* 7, S. 10408. DOI: 10.1038/ncomms10408.

Schuenemann, Verena J.; Peltzer, Alexander; Welte, Beatrix; van Pelt, W. Paul; Molak, Martyna; Wang, Chuan-Chao et al. (2017): Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. In: *Nature communications* 8, S. 15694. DOI: 10.1038/ncomms15694.

Schuenemann, Verena J.; Singh, Pushpendra; Mendum, Thomas A.; Krause-Kyora, Ben; Jäger, Günter; Bos, Kirsten I. et al. (2013): Genome-wide comparison of medieval and modern Mycobacterium leprae. In: *Science (New York, N.Y.)* 341 (6142), S. 179–183. DOI: 10.1126/science.1238286.

Siebke, Inga; Campana, Lorenzo; Ramstein, Marianne; Furtwängler, Anja; Hafner, Albert; Lösch, Sandra (2018): The application of different 3D-scan-systems and photogrammetry at an excavation — A Neolithic dolmen from Switzerland. In: *Digital Applications in Archaeology and Cultural Heritage* 10, e00078. DOI: 10.1016/j.daach.2018.e00078.

Siebke, Inga; Steuri, Noah; Furtwängler, Anja; Ramstein, Marianne; Arenz, Gabriele; Hafner, Albert et al. (2019): Who lived on the Swiss Plateau around 3300 BCE? Analyses of commingled human skeletal remains from the dolmen of Oberbipp. In: *Int J Osteoarchaeol* 29 (5), S. 786–796. DOI: 10.1002/oa.2791.

Šišlina, Natal'ja I. (2008): Reconstruction of the Bronze Age of the Caspian steppes. Life styles and life ways of pastoral nomads. Oxford: Archaeopress (BAR international series, 1876).

Skoglund, Pontus; Malmström, Helena; Raghavan, Maanasa; Storå, Jan; Hall, Per; Willerslev, Eske et al. (2012): Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. In: *Science (New York, N.Y.)* 336 (6080), S. 466–469. DOI: 10.1126/science.1216304.

Steuri, Noah; Siebke, Inga; Furtwängler, Anja; Szidat, Sönke; Krause, Johannes; Lösch, Sandra; Hafner, Albert (2019): Multiple Radiocarbon Dating of Human remains: Clarifying the Chronology and Sequences of Burials in the late Neolithic Dolmen of Oberbipp (Switzerland). In: *Radiocarbon* 61 (6), S. 1697–1709. DOI: 10.1017/RDC.2019.96.

Suter, Peter; Affolter, Jehanne; Bode, Michael; Brombacher, Christoph; Francuz, John; Löffler, Ingolf et al. (2017): Um 2700 v. Chr. - Wandel und Kontinuität in den Ufersiedlungen am Bielersee Bern. In: *Archäologischer Dienst des Kantons Bern 2 Bände*.

Underhill, Peter A.; Poznik, G. David; Rootsi, Siiri; Järve, Mari; Lin, Alice A.; Wang, Jianbin et al. (2015): The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. In: *European journal of human genetics : EJHG* 23 (1), S. 124–131. DOI: 10.1038/ejhg.2014.50.

Vågene, Åshild J.; Herbig, Alexander; Campana, Michael G.; Robles García, Nelly M.; Warinner, Christina; Sabin, Susanna et al. (2018): Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. In: *Nature ecology & evolution* 2 (3), S. 520–528. DOI: 10.1038/s41559-017-0446-6.

Wang, Chuan-Chao; Reinhold, Sabine; Kalmykov, Alexey; Wissgott, Antje; Brandt, Guido; Jeong, Choongwon et al. (2019): Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. In: *Nature communications* 10. DOI: 10.1038/s41467-018-08220-8.

Whittle Alasdair, Cummings Vicki (2007): Going Over: The Mesolithic-Neolithic Transition in North-West Europe. In: *Proceedings of the British Academy* 144.

Woodward, S. R.; Weyand, N. J.; Bunnell, M. (1994): DNA sequence from Cretaceous period bone fragments. In: *Science (New York, N.Y.)* 266 (5188), S. 1229–1232. DOI: 10.1126/science.7973705.

Appendices

Papers I, II and III

Paper I

Furtwängler, A., Reiter E, Neumann GU, Siebke I, Steuri N, Hafner A, Lösch S, Anthes N, Schuenemann VJ, Krause J. "Ratio of mitochondrial to nuclear DNA affects contamination estimates in ancient DNA analysis." *Scientific reports* 8.1 (2018): 14075.

Ratio of mitochondrial to nuclear DNA affects contamination estimates in ancient DNA analysis.

Authors: Furtwängler A*[1], Reiter E[2], Neumann GU[2], Siebke I[3], Steuri N[4], Hafner A[4], Lösch S[3], Anthes N[5], Schuenemann VJ[2,6,7], Krause J*[8,9,10].

Affiliations:

[1]Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen, Germany. anja.furtwaengler@uni-tuebingen.de.
[2]Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen, Germany.
[3]Department of Physical Anthropology, Institute of Forensic Medicine, University of Bern, Bern, Switzerland.
[4]Institute of Archaeological Sciences and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland.
[5]Institute of Ecology and Evolution, Animal Evolutionary Ecology group University of Tübingen, Tübingen, Germany.
[6]Institute of Evolutionary Medicine, University of Zurich, Zurich, Switzerland.
[7]Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Tübingen, Germany.
[8]Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen, Germany. krause@shh.mpg.de.
[9]Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Tübingen, Germany. krause@shh.mpg.de.
[10]Max Planck Institute for the Science of Human History, Jena, Germany. krause@shh.mpg.de.

* Corresponding author
E-mail: krause@shh.mpg.de, anja.furtwaengler@uni-tuebingen.de

Abstract

In the last decade, ancient DNA research has grown rapidly and started to overcome several of its earlier limitations through Next-Generation-Sequencing (NGS). Among other advances, NGS allows direct estimation of sample contamination from modern DNA sources. First NGS-based approaches of estimating contamination measured

heterozygosity. These measurements, however, could only be performed on haploid genomic regions, i.e. the mitochondrial genome or male X chromosomes, but provided no measures of contamination in the nuclear genome of females with their two X chromosomes. Instead, female nuclear contamination is routinely extrapolated from mitochondrial contamination estimates, but it remains unclear if this extrapolation is reliable and to what degree variation in mitochondrial to nuclear DNA ratios affects this extrapolation. We therefore analyzed ancient DNA from 317 samples of different skeletal elements from multiple sites, spanning a temporal range from 7,000 BP to 386 AD. We found that the mitochondrial to nuclear DNA (mt/nc) ratio negatively correlates with an increase in endogenous DNA content and strongly influenced mitochondrial and nuclear contamination estimates in males. The ratio of mt to nc contamination estimates remained stable for overall mt/nc ratios below 200, as found particularly often in petrous bones but less in other skeletal elements and became more variable above that ratio.

Introduction

The emergence of Next-Generation-Sequencing (NGS) technologies has substantially advanced the field of ancient DNA (aDNA) research[1]. Besides its high throughput, NGS gave rise to analyses of ancient DNA-specific DNA damage to reveal patterns that authenticate ancient DNA. *Post-mortem* DNA continuously accumulates characteristic modifications, in particular deamination at the DNA fragment ends[2,3,4]. Deamination frequency thus increases over time and can therefore reveal the ancient origin of an aDNA sample[3]. As a second step required for authenticating ancient human DNA, NGS data allow estimating contamination levels of human DNA directly[1]. Earlier estimates of contamination levels in NGS data from early modern humans were based on so-called diagnostic positions on the mitochondrial DNA (mtDNA), i.e. nucleotide positions that differ between the sample and a comparative dataset of complete modern human mtDNA sequences from world-wide populations[5]. Reads exhibiting a different allele at these positions than the majority of reads likely constitute contamination. Today, Bayesian approaches allow even more precise estimates by identifying the proportion of sequencing reads that were considered to be authentic aDNA from the studied individual[6].

While quantifications of contamination levels in the haploid and maternally inherited mtDNA are achieved by measuring levels of heterozygosity[1], contamination estimates for the diploid nuclear genome can only be obtained for the haploid sex chromosomes in male individuals, especially for the larger X chromosome[7]. For female individuals, most current studies therefore restrict analyses to nuclear DNA sequences with typical aDNA damage patterns and thus sequences of trusted ancient origin, often reducing datasets by an order of magnitude[8]. Alternatively, female mtDNA contamination estimates are extrapolated to the nuclear level and only female samples with low

mtDNA contamination estimates are used for population genetic analysis on the nuclear genome. Whether this extrapolation is reliable, however, remains untested, in particular given that mitochondrial to nuclear DNA ratios can substantially vary between and even within bone samples[9], potentially affecting the extrapolation of mtDNA to nuclear contamination levels.

The reliability of this extrapolation may also depend on the chosen skeletal element given differences in endogenous DNA content and relative amounts of mitochondrial and nuclear DNA. Recent work has identified the petrous part of the temporal bone as a particularly rich source of endogenous DNA and thus an ideal candidate for aDNA studies[10,11]. Endogenous DNA portions in this skeletal element exceed that in other parts of the skeleton by up to a factor of 400[12]. This reduces the cost of shotgun sequencing for ancient human genomes to sufficient coverage[10] and allows obtaining sequences from geographic regions with typically poor DNA preservation, such as the Near East, Remote Oceania and Africa[13,14,15,16,17]. The exceptionally good DNA preservation in petrous bones has been linked to the high bone density of the bony labyrinth[10], which reaches adult size during gestation[18] and shows reduced bone remodeling compared to the surrounding tissue[19]. These conditions seem to be ideal for DNA preservation. Initial studies suggest low mt/nc ratios for petrous bones[20]. We hypothesize that these low ratios allow reliable estimates of contamination with human DNA using mtDNA, especially for female individuals.

This is the first study to explore systematic differences in mt/nc ratios between skeletal elements and classically used negative controls, and the relationship between mt/nc ratios and contamination estimates. We compare mt/nc ratios and overall DNA preservation in petrous bones, teeth and other skeletal elements in newly produced and publically available aDNA datasets. Our data show that DNA contamination estimates in males strikingly vary with mt/nc ratios, and argue that the typically low mt/nc ratios in petrous bones make them ideal for making reliable contamination estimates using mtDNA.

Results

Differences in the mt/nc ratio between sites

For each of the three tested skeletal elements we investigated samples from different sites (Table 1), labelled P1 through P7 for petrous bones, T1 through T4 for teeth, and B1 through B4 for other bones. The subsample pairs T3/B1, T4/B2 and P1/T1 originate from the same studies, respectively. Given that any systematic differences between sample origins in, for example, general preservation, laboratory protocols, soil type, climatic conditions and taphonomic alterations can introduce between-sample variation, we first assessed the degree of origin-related variation in mt/nc ratios within each skeletal element (Fig. 1). Indeed, we detected significant variation between sample origins in all

skeletal elements (Petrous bones ANOVA $F_{6,50} = 20.15$, $p < 0.001$; Teeth ANOVA $F_{3,181} = 25.25$, $p < 0.001$; Bones ANOVA $F_{3,71} = 9.26$, $p < 0.001$), with at least one sample differing from all remaining ones in each skeletal element (Tukey-HSD tests, Fig. 1). In all the following statistical models comparing differences in mt/nc ratios between tissues, we therefore routinely – and conservatively – accounted for between-origin variation while maintaining the paired nature of samples from the same source by integrating study ID and site ID as random factors.

Differences in the mt/nc ratio between skeletal elements

To assess the mt/nc ratio of potential sources for contamination of archaeological samples, we also measured the mt/nc ratio of human contamination originating from plant extracts[21] and laboratory negative controls. If mt/nc ratios of modern human contamination exceeded those of the ancient human DNA, the relative contamination impact would be stronger on the mtDNA than on the nDNA, and *vice versa*. We found no significant difference in the $\log_e$[mt/nc ratio] between both controls (ANOVA $F_{1,21} = 2.33$, $p = 0.14$), therefore combining them for all further analyses.
The mt/nc ratios showed significant variation (Fig. 2) between the three skeletal elements (petrous bones, teeth and bones) and controls ($Chi^2 = 9.54$, $df = 3$, $p = 0.023$). Petrous bones had significantly lower mt/nc ratios than teeth and showed a trend towards lower mt/nc ratios than other bone samples, while teeth and other bones showed similarly high ratios (Tukey HSD post hoc tests as indicated in Fig. 2). The mt/nc ratios of the combined control group were intermediate between the ratios derived from teeth and bones and those derived from petrous bones. In addition to these differences in mean mt/nc ratios, skeletal elements also showed striking differences in mt/nc ratio variation (Levene's test $F_{3,336} = 15.01$, $p < 0.001$). Variation was again similar between teeth and other bones (Tukey HSD post hoc $P = 0.75$) but significantly larger in these two than in petrous bones (both $P < 0.0001$).

Factors influencing the mt/nc ratio

To assess the degree to which mt/nc ratios vary with sample characteristics within sample origins, we tested relationships with endogenous DNA content, age, and mapping stringency. Despite much between-sample variation, mt/nc ratios almost consistently decreased with increasing endogenous DNA content across skeletal elements and sampling origins (Fig. 3, mixed model endogenous DNA-effect $Chi^2 = 6.57$, $df = 1$, $P = 0.010$, overall regression slope = -0.224 ± 0.082, mean ± SE).
In contrast, we detected no significant relationship between radio carbon dates of the P2, P3, T3 and B2 samples and their $\log_e$[mt/nc ratio] (mixed model sample age-effect $Chi^2 = 0.63$, $df = 1$, $P = 0.426$). Changes in the mapping stringency in terms of mismatches allowed per read (BWA parameter *–n*, with tested values 0.001, 0.01, 0.1, 0.2 and 0.8 corresponding to 5, 3.7, 2, 1.7 and 0 mismatches for the average fragment size of T1 and

T2 of 55 bp) showed some fluctuation in the mt/nc ratio in T1 and T2 teeth, but no directional effect on mean mt/nc ratios (linear regression $F_{1,125} = 2.049$, $P = 0.155$, $R^2 = 0.016$).

Effect of changes in mt/nc ratio on contamination estimates

To evaluate whether mitochondrial and nuclear contamination estimates vary with sample mt/nc ratios, we analysed male individuals where sufficient NGS coverage allowed reliable estimates of mtDNA and nuclear contamination levels. Mitochondrial and X-chromosomal contamination rates were by far most strongly biased towards mt-contamination in samples with high mt/nc ratios (Fig. 4). In contrast, where mt/nc ratios were small, nuclear contamination estimates were close to mtDNA contamination rates with an average ratio of 1.47 ± 0.44 SE. In addition, contamination ratios remained stable across $\log_e$[mt/nc ratio] between 0 and $\log_e(200) = 5.3$ (mixed model $\log_e$[mt/nc ratio]-effect $Chi^2 = 0.80$, df = 1, p = 0.37), but strikingly increased with $\log_e$[mt/nc ratio] at values exceeding $\log_e(200) = 5.3$ ($Chi^2 = 5.95$, df = 1, p = 0.015). This stability is reflected in the fact that samples with a ratio below $\log_e(200) = 5.3$ are located on the bisecting line when plotting MT contamination against X contamination (Fig. 4).

Discussion

In this work, we show that variation in the mt/nc ratio of ancient DNA samples strongly influences the estimates of contamination by human DNA, which is one of the most critical steps when analysinig ancient human DNA. Besides the characteristic age related misincorporation[2], estimates of contamination rates are used for authentication of ancient human DNA. Currently available methods for contamination estimates of the mtDNA and nDNA levels measure heterozygosity in haploid regions of the mtDNA, or the X chromosome in males[6,22,23]. However, at present it is not possible to estimate contamination levels in the nDNA of female individuals. Recent studies of ancient human nDNA either restrict their analysis to male individuals with known nDNA contamination levels or extrapolate contamination levels of mtDNA to nDNA in the case of females.

Our results show that contamination levels extrapolated from the mtDNA to the nDNA might systematically underestimate the actual level of nDNA contamination. This appears particularly likely for teeth and other bones with a high mt/nc ratio, where such extrapolations (as often done for female specimens) may be highly misleading. For example, three "other bone" samples in the current study exhibited less than 3% contamination on the mtDNA, contrasting to nDNA-contamination between 16 and 30% measured on the X-chromosomes. All three samples were characterized by high mt/nc ratios spanning 870 to >56.000. For these samples, it is not possible to determine from the distribution of reads on the X and Y chromosome with certainty if they are males with a contaminated X chromosome or if they are females with male contamination. Either way,

the contamination on the nuclear level is too high for further population genetic analysis and even in absence of Y chromosomal reads as an indicator of contamination, the extrapolation from the mtDNA contamination level could still underestimate the actual nuclear contamination level in case of high mt/nc ratios. Therefore, we advise to take the mt/nc ratio into consideration if only the mtDNA contamination estimate is used in order to choose female samples to be included in population genetic analysis. An effect of mt/nc ratio on the contamination estimates of mtDNA and nDNA as described above was already suggested in the context of analysing Neandertal nuclear and mtDNA[9,24] and receives further support by our statistical analysis.

In contrast to samples with a high mt/nc ratio, samples with a ratio below 200 ($\log_e[200] = 5.3$) seem to provide a rather reliable estimate of contamination on the mitochondrial level (Fig. 4). We included two types of possible contamination in our study: modern human DNA contamination in archaeological samples measured from aDNA plant extracts from 6,000 year old barley[21] and contamination introduced during lab work measured from the negative controls carried along in the entire laboratory workflow. The mt/nc ratios of these potential contamination sources were intermediate to those detected in the three tested skeletal elements (Fig. 2). Therefore, these types of contamination contribute more to the contamination level of mtDNA than nDNA in samples showing lower ratios than these controls. Our systematic comparison of a large number of teeth, diverse bones and petrous bones indicates that the petrous bones display relatively low mt/nc ratios and hence have a high chance to show even lower contamination rates on the nDNA than extrapolated from their mtDNA. In addition, we found that the mt/nc ratio in petrous bones not only tended to be smaller compared to teeth and other bones but also exhibited significantly less variation (Fig. 2). The lower mt/nc ratios in petrous bones increase the chance to obtain reliable extrapolations of nDNA contamination rates from estimated mtDNA contamination. Furthermore, the extraordinarily good preservation of endogenous DNA in petrous bones makes it possible to overcome the limited preservation of aDNA in most environments[12]. Hansen and colleagues[20] show that teeth can have even higher contents of endogenous DNA than petrous bones. In our dataset, however, the majority of petrous bones displayed clearly higher endogenous DNA contents than teeth, leading us to maintain petrous bones as the prime source for highly concentrated and little contaminated endogenous DNA.

From these influences of the mt/nc ratio on the contamination estimates, we conclude that the mt/nc ratio should be considered if mtDNA contamination estimates are used to select female samples for nDNA studies. Reasons for the observed variation in this ratio have already been subject of different studies and no clear reason for the elevation of the mt/nc ratio in some aDNA samples compared to modern DNA has been identified. One possibility could be a bias of the mt/nc ratio originating from bacteria, since mtDNA contains features of a bacterial chromosome and bacterial reads from the environmental

background might map better to the mtDNA elevating the mt/nc ratio. However, mapping with different stringencies (different numbers of allowed mismatches per read, BWA parameter *-n*) did not reveal a particular trend. Allowing more mismatches (*-n* 0.001, *-n* 0.01 and *-n* 0.1) resulted in similar values in the ratio than more stringent mapping (*-n* 0.2 and *-n* 0.8). We could therefore not detect any effect of the bacterial background in aDNA extracts on the correlation between high mt/nc ratios and lower percentages of endogenous DNA.

It has been suggested that the higher ratios in aDNA might originate from better protection of DNA in the mitochondrion due to its double membrane[25]. In addition, Schwarz and colleagues[25] and Allentoft and colleagues[26] conclude from their observation of longer mtDNA fragments compared to nDNA that mtDNA decays more slowly than nDNA. This would result in increasing ratios through time. No such trend is detectable in our data for petrous bones (P2, P3) and teeth (T3), and only a weak trend in other bones (B2). Therefore, we found no indication that different decay rates cause the observed variation in the mt/nc ratio.

Another reason for differences in mt/nc ratios could be higher metabolic activity in particular skeletal elements, as suggested by Higgins and colleagues[27] in their study of different parts of teeth. Their observation indicates that higher mt/nc ratios in dentine may originate from its higher metabolic activity as compared to enamel and cementum. This assumption, of different metabolic activity resulting in different mt/nc ratios could also be applied to other skeletal elements, the petrous bone in particular. In the cortical part of the petrous bone, bone remodelling is suppressed compared to surrounding tissue by a specific signalling-pathway, and the number of metabolic active osteoclasts and osteoblasts is reduced[28]. Furthermore, the more compact bone regions contain a higher number of osteocytes, which have lower numbers of mitochondria[29]. The number of mitochondria in osteocytes decreases from periosteal and endosteal surfaces towards the inner and denser bone parts[30]. Consequently, sampling the hardest part of the petrous bone results in low mt/nc ratios due low mt/nc ratio in this area *ante-mortem*. Within the densest parts of the cortical bone displaying low mt/nc ratios, the endogenous DNA is also better protected against the environmental background of bacteria, fungi and other microbes resulting in high percentages of endogenous DNA.

In conclusion, we showed that the mt/nc ratio is an important value in aDNA authentication. It strongly influences the accuracy of extrapolating nDNA contamination levels from mtDNA contamination estimates. This approach should be used with particular caution if mt/nc ratios exceed 200. Lower mt/nc ratios are often associated with high percentages of endogenous DNA, as typically found in the densest parts of the petrous bones. In this case, low mt/nc ratios are most likely derived from a low mtDNA concentration in this bone region *ante-mortem*, which is caused by a low metabolic activity.

Temporal bones, including the petrous bones have already proven to be valuable in physical anthropology. The compact structure of the petrous bone is ideal for preservation in archaeological contexts, therefore, it can be used to estimate the minimal number of individuals[31]. While age can only be roughly estimated depending on developmental stages[32]. In addition, the discovery of the higher percentage of endogenous DNA in petrous bone compared to other skeletal elements[10,11] constitutes a substantial improvement for the field of aDNA. Our study confirms those findings and shows that mitochondrial to nuclear ratios provide a further argument to extract ancient DNA from petrous bones providing reliable human contamination assessment from both sexes.

Materials and Methods

Dataset

We combined shotgun data from 76 new ancient DNA samples from the late Neolithic dolmen burial site of Oberbipp in Switzerland[33,34], the late Neolithic multiple burial of Muttenz[35], the final Neolithic sites in Switzerland Spreitenbach[36], Seengen[35], Zurzach[35] and cave finds from Wartau[37] (detailed description of archaeological sites and laboratory workflow SI section 1 and 2) with published datasets (Table 1). In total, DNA from 317 ancient individuals and 23 controls was analysed for its mtDNA to nDNA (mt/nc) ratio. For each of the three investigated skeletal elements, samples originated from three sources. Different skeletal elements do not originate from the same individuals. Samples for petrous bones (n = 57) included 40 newly processed samples from the previous mentioned Swiss burials (P1, P4, P5, P6, P7), 13 from Gamba and colleagues[10] (P2) and four from Broushaki and colleagues[13] and Gallego-Llorente and colleagues[14] (P3). Teeth samples (n = 185) included 36 from the Swiss burials (T1, T2), 88 from Allentoft and colleagues[38] (T3) and 61 from Schuenemann and colleagues[39] (T4). P1 and T1 from this study originate from the same multiple burial in Oberbipp, Switzerland but due to the comingled nature of the remains an assignment of the teeth to the petrous bones is not possible. Furthermore, the endogenous DNA in the teeth is too low to allow kinship analysis to identify identical individuals. Therefore, it is possible that teeth and petrous bones are from the same individuals. Finally, diverse bone samples from compact bone parts (n = 75) included 13 from Allentoft and colleagues[38] (B1), 52 from Schuenemann and colleagues[39] (B2) and eight from Günther and colleagues[40] (B3) and two from Broushaki and colleagues[13] (B3). Fastq files were downloaded from the European Nucleotide Archive. Radio carbon dates were used from Gamba and colleagues[10], Broushaki and colleagues[13], Gallego-Llorente and colleagues[14] and Schuenemann and colleagues[39].

Bioinformatic processing

All data were processed with the EAGER pipeline[41]. If necessary, adapters were removed and paired-end data was merged using Clip&Merge[41]. Mapping was performed with BWA with the mismatch parameter set to 0.01 and a seed length of 1,000. If necessary, PCR duplicates were removed using DeDup[41]. Mitochondrial to nuclear ratios were calculated by dividing the mean coverage of the mitochondrial genome by the mean coverage of the nuclear genome to take into account the different length of nuclear and mitochondrial genome. Samples with no reads on the mitochondrial chromosome resulting in a ratio of zero as found in some negative controls, plant extracts, and extremely bad preserved teeth samples were excluded.
Reads mapping to the mitochondrial genome were extracted from the BAM files and the mitochondrial genome was reconstructed using the software *schmutzi*[22]. Mitochondrial contamination was then estimated with a Bayesian approach as described in Fu and colleagues[6]. Sex determination was performed after Skoglund and colleagues[42]. X chromosomal contamination in males with more than 0.5-fold coverage on the X chromosome was estimated using ANGSD[23].

Statistical analysis

Statistical analysis was performed in R version 3.4.3 (R Core Team 2017). We first used one-way ANOVA to assess differences in $\log_e$[mt/nc ratio] (i.e., the raw mt/nc-ratios $\log_e$-transformed to approach normality) between the sub-samples within each of the three different skeletal elements and controls, followed by Tukey HSD *post hoc* tests. Second, to assess overall differences in $\log_e$[mt/nc ratio] between skeletal elements and controls, we performed linear mixed models as implemented in the lme4-package for R[43]. Given substantial variation in $\log_e$[mt/nc ratio] between different sample origins and studies detected above, this model contained the source study as a random slope factor, allowing between-study variation in the main effect while maintaining the paired nature of samples from the same source. Pairwise Tukey HSD *post hoc* tests were calculated from this model using the lsmeans package for R[44]. Differences in the variance of the $\log_e$[mt/nc ratio] between skeletal elements and control were tested using Levene's test, followed by Tukey HSD *post hoc* tests.
To assess how $\log_e$[mt/nc ratio] varies with DNA-content (log-transformed to approach normality) we constructed a linear mixed model as above with $\log_e$[mt/nc ratio] as the response variable, $\log_e$[endogenous DNA] as the predictor, and subsample as random slope factor allowing differential regression slopes between subsamples. In addition, we added skeletal element and its interaction with $\log_e$[endogenous DNA] as fixed factor to assess slope consistency between skeletal elements. Given that this interaction was clearly insignificant ($Chi^2 = 0.18$, $df = 2$, $P = 0.91$) we removed it from the final reported model. Information about endogenous DNA content was only available for all Swiss and

Egyptian samples. An identical model structure was used to assess the relationship between sample age and $\log_e$[mt/nc ratio], now with sample age as the main predictor. Differences in $\log_e$[mt/nc ratio] between different mapping stringencies were assessed using linear regression. Information about $\log_e$[mt/nc ratio] at different mapping stringencies was obtained for the teeth of the Swiss burials sites Oberbipp and Spreitenbach (T1) since they have low endogenous content and a high portion of bacterial background.

Finally, to assess how the ratio of mitochondrial over X-chromosomal contamination varies with $\log_e$[mt/nc ratio] we constructed linear mixed models as above, but separate for samples with low (<5.3) and high (>5.3) $\log_e$[mt/nc ratio]. Contamination ratio served as the response variable, $\log_e$[mt/nc ratio] as the predictor, and skeletal element as a random slope factor allowing differential responses to $\log_e$[mt/nc ratio].

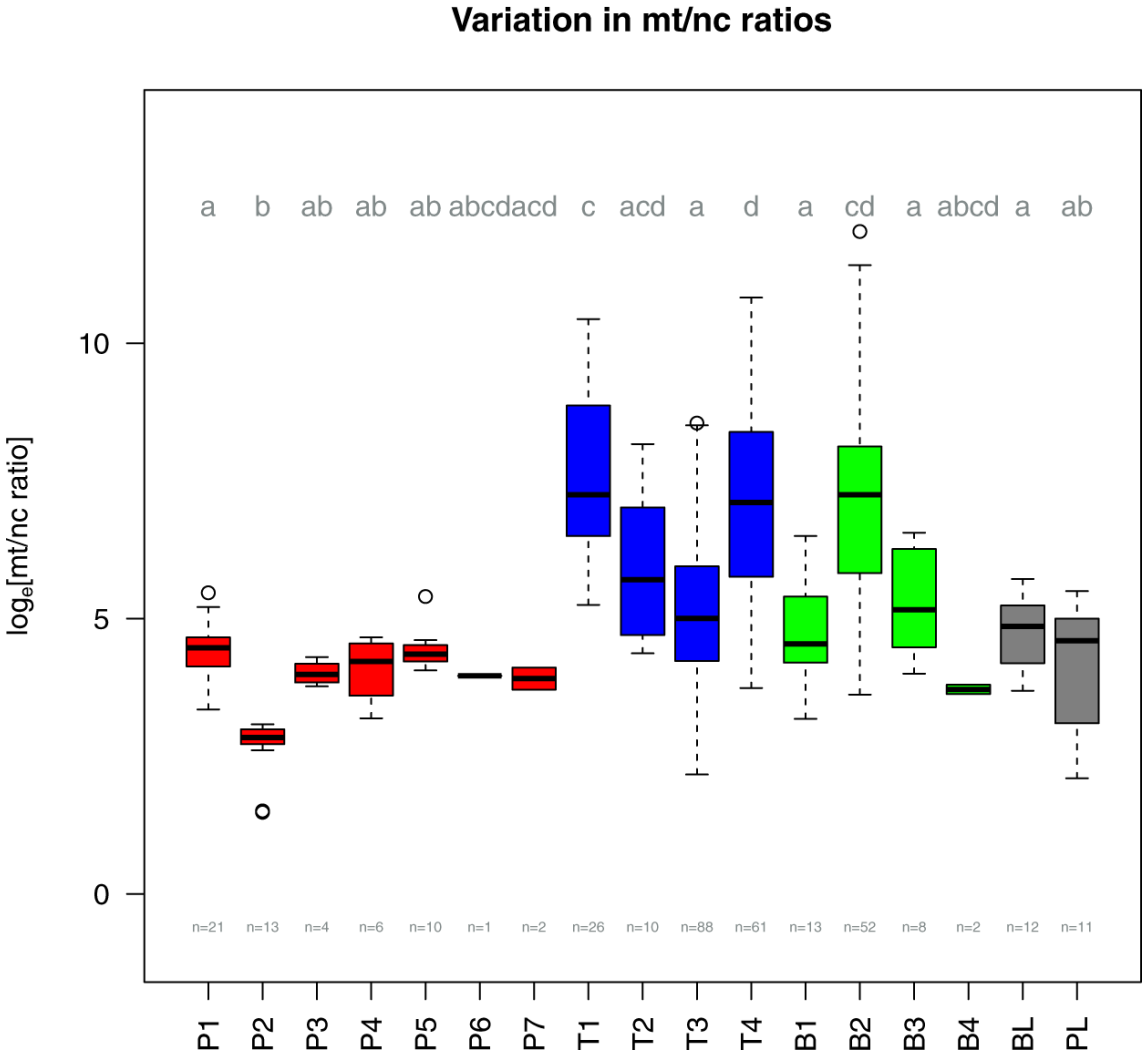


Figure 1: Variation in mt/nc ratios between sample sources. Mitochondrial to nuclear DNA ratios (log-transformed) are shown for petrous bones (red), teeth (blue), other bones (green), and controls (grey). Box plots show the raw data median

(thick line), interquartile range IQR (box), data within 1.5*IQR (flags), and extreme values (dots). Different lower case letters indicate pairwise differences revealed by Tukey-HSD post-hoc tests.

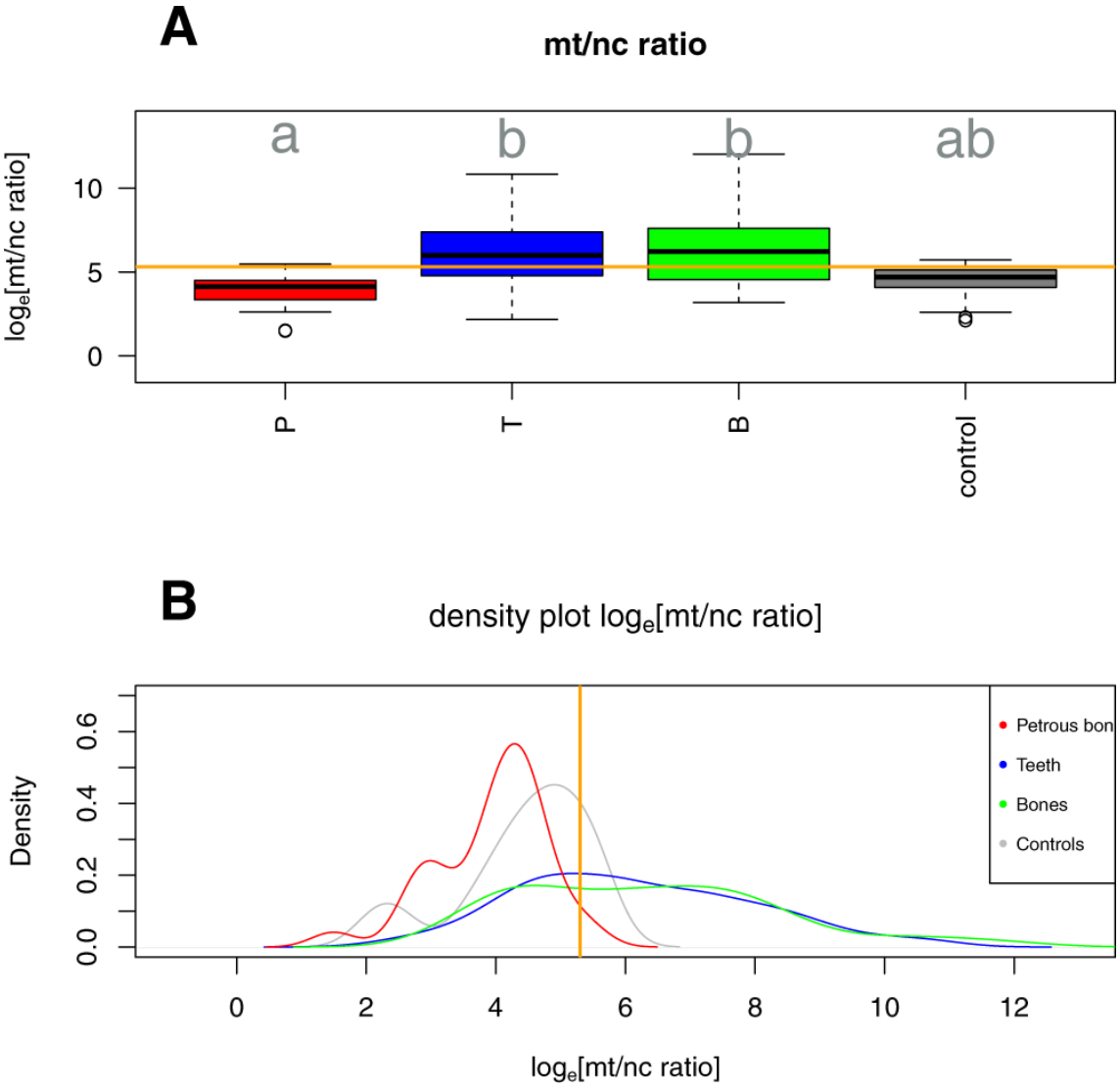


Figure 2: Variation in the mt/nc ratios between skeletal elements. (**A**) Box plots of the mitochondrial to nuclear DNA ratio (log-transformed) of human DNA in petrous bones (P), teeth (T), other bones (B) and controls. Panel (**B**) shows density plots of the same data across elements. Orange lines indicate the suggested threshold of the mitochondrial to nuclear ratio of 200 ($\log_e(200) = 5.3$).

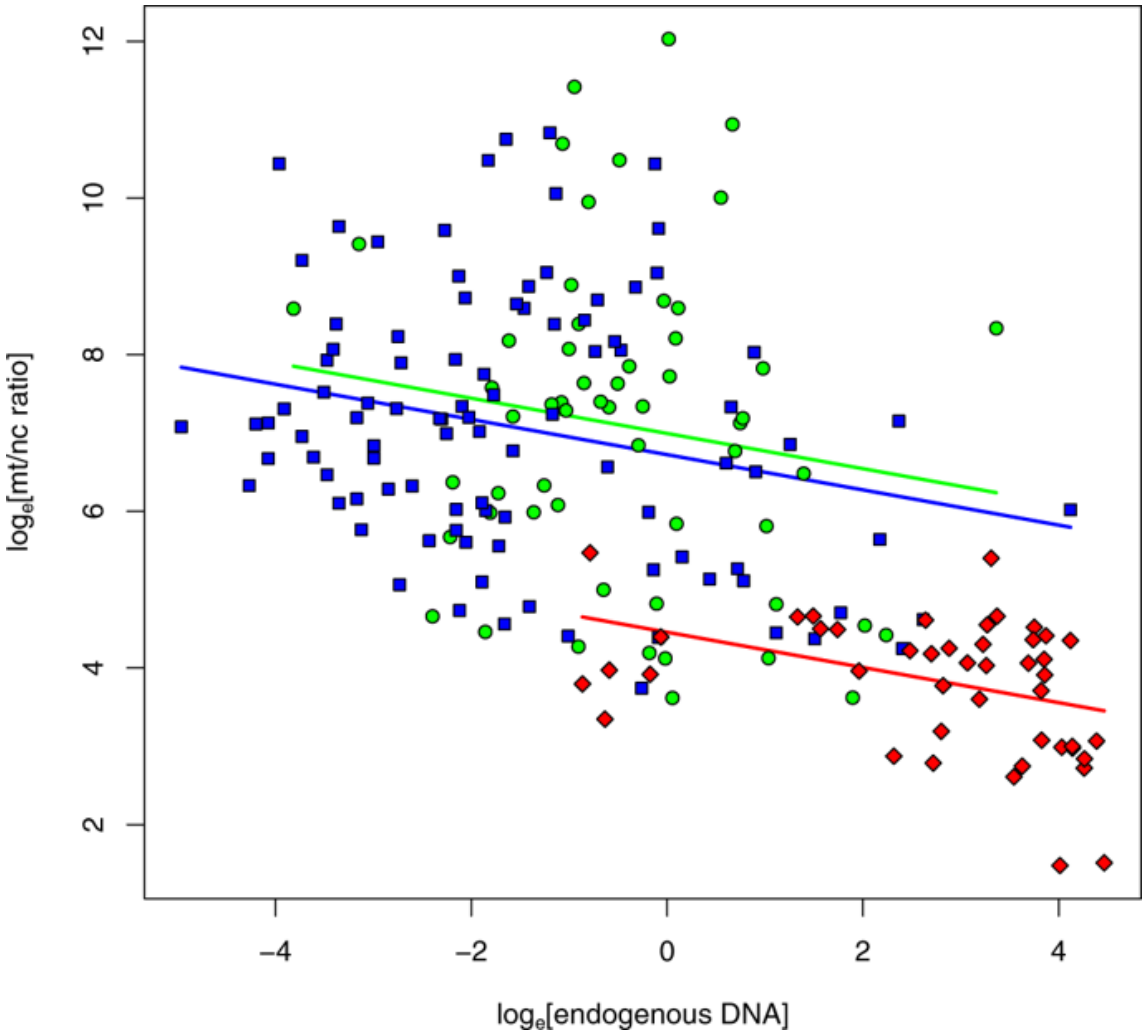


Figure 3: Relationship between the mt/nc ratio and the percentage of endogenous DNA in petrous bones (red), teeth (blue) and other bones (green). Despite overall differences in average mt/nc ratios between skeletal elements and subsamples, these ratios declined consistently when samples contained more endogens DNA.

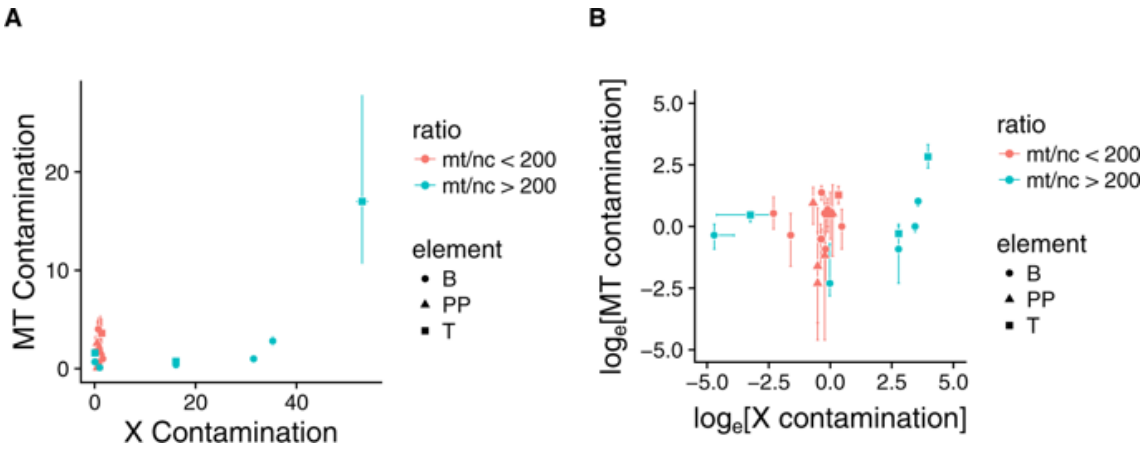


Figure 4: Relationship between male mitochondrial and X-chromosome contamination rates in un-transformed (A) and $\log_e$ transformed (B) data. Different colours indicate samples with low ($<\log_e(200) = 5.3$ red) and high ($>\log_e(200) = 5.3$ blue) mt/nc ratios.

Table 1 Samples from different studies used in this analysis.

Site	Dating	Samples			Reference
		Petrous bones	Teeth	Diverse bones	
Oberbipp, Switzerland	approx. 5500 BP	P1	T1		This study
Spreitenbach, Switzerland	approx. 4500 BP		T2		This study
Europe and Central Asia	3400 BP–600 AD		T3	B1	[38]
Abusir el-Meleq, Egypt	1300 BP–386 AD		T4	B2	[39]
Hungary	5060–1830 BP	P2			[10]
Atapuerca, Spain	5500–3500 BP			B3	[40]
Zagros, Iran	10000–7000 BP	P3		B4	[13]
Zagros, Iran	10000–9700 BP				[14]
Muttenz, Switzerland	Neolithic	P4			This study
Wartau, Switzerland	Neolithic	P5			This study
Seengen, Switzerland	Neolithic	P6			This study
Bad Zurzach, Switzerland	Neolithic/Bronze Age	P7			This study

References

1. Stoneking, M. & Krause, J. Learning about human population history from ancient and modern genomes. NatRev Genet. 12, 603–14 (2011).
2. Briggs, A. W. et al. Patterns of damage in genomic DNA sequences from a Neandertal. Proc Natl Acad Sci USA 104.37, 14616–14621 (2007).
3. Sawyer, S. et al. Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. PLoS ONE. 7, e34131 (2012).
4. Dabney, J., Meyer, M. & Paabo, S. Ancient DNA Damage. Cold Spring Harbor Perspectives in Biology. 5, a012567 (2013).
5. Krause, J. et al. A Complete mtDNA Genome of an Early Modern Human from Kostenki, Russia. Current Biology. 3, 231–236 (2010).
6. Fu, Q. et al. A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes. Current Biology. 7, 553–559 (2013).
7. Rasmussen, M. et al. Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature. 463, 757–62 (2010).
8. Fu, Q. et al. The genetic history of Ice Age Europe. Nature, 243 (2016).
9. Green, R. E. et al. The Neandertal genome and ancient DNA authenticity. EMBO J. 28, 2494–502 (2009).
10. Gamba, C. et al. Genome flux and stasis in a five millennium transect of European prehistory. Nat Comms, 5257 (2014).
11. Rasmussen, M. et al. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature. 506, 225–29 (2014).
12. Pinhasi, R. et al. Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. PLoS ONE. l0, e0129102 (2015).
13. Broushaki, F. et al. Early Neolithic genomes from the eastern Fertile Crescent. Science. 6298, 499–503 (2016).
14. Gallego-Llorente, M. et al. The genetics of an early Neolithic pastoralist from the Zagros, Iran. Sci. Rep., 31326 (2016).
15. Llorente, M. G. et al. Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. Science. 350, 820–22 (2015).
16. Skoglund, P. et al. Genomic insights into the peopling of the Southwest Pacific. Nature. 538, 510–13 (2016).
17. van de Loosdrecht, M. et al. Pleistocene North African genomes link Near Eastern and sub-Daharan African human populations Science, eaar8380 (2018).
18. Richard, C. et al. New insight into the bony labyrinth: A microcomputed tomography study. AurisNasus Larynx. 37, 155–61 (2010).
19. Sorensen, M. S. Temporal bone dynamics, the hard way. Formation, growth, modeling, repair and quantum type bone remodeling in the otic capsule. Acta Otolaryngol Suppl. 512, 1–22 (1994).
20. Hansen, H. B. et al. Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum. PloS ONE 12.1, e0170940 (2017).
21. Mascher, M. et al. Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. Nat Genet. 48, 1089–93 (2016).
22. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome biology 16.1, 1 (2015).
23. Rasmussen, M. et al. An Aboriginal Australian genome reveals separate human dispersals into Asia. Science 334, 6052:94–8 (2011).
24. Green, R. E. et al. Analysis of one million base pairs of Neanderthal DNA. Nature. 7117, 330–336 (2006).
25. Schwarz, C. et al. New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. Nucleic Acids Research. 37, 3215–29 (2009).

26. Allentoft, M. E. et al. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proceedings of the Royal Society B: Biological Sciences. 1748, 4724–4733 (2012).
27. Higgins, D. et al. Differential Nuclear and Mitochondrial DNA Preservation in Post-Mortem Teeth with Implications for Forensic and Ancient DNA Studies. PLoS ONE. 5, e0126935 (2015).
28. Bloch, S. L., Kristensen, S. L. & Sørensen, M. S. The Viability of Perilabyrinthine Osteocytes: A Quantitative Study Using Bulk-Stained Undecalcified Human Temporal Bones. Anat Rec. 7, 1101–1108 (2012).
29. Dudley, H. R. The Fine Structure Of Bone Cells. The Journal of Cell Biology. 3, 627–649 (1961).
30. Frikha-Benayed, D., Basta-Pljakic, J., Majeska, R. J. & Schaffler, M. B. Regional differences in oxidative metabolism and mitochondrial activity among cortical bone osteocytes. Bone, 15–22 (2016).
31. Osterholtz, A. J., Baustian, K. M., Martin, D. L. & Potts, D. T. Commingled Human Skeletal Assemblages: Integrative Techniques in Determination of the MNI/MNE In Commingled and disarticulated human remains. Working toward improved theory, method, and data, edited by A. J. Osterholtz (Springer, New York, NY, 2014), pp. 35–50.
32. Cunningham, C., Scheuer, L. & Black, S. M. Developmental juvenile osteology (Academic Press, Amsterdam, 2016).
33. Ramstein, M., Schimmelpfennig, D. & Lösch, S. Ein neolithischer Dolmen an der Steingasse in Oberbipp. Archäologie Schweiz 37(3), 4–15 (2014).
34. Siebke, I., Furtwängler, A., Hafner, A., Krause, J. & Lösch, S. An Interdisciplinary Project on the Neolithic Population of Modern Switzerland. 86th Annual Meeting of the American Association of Physical Anthropologists. American Journal of Physical Anthropology. p.358 (2017)
35. Bleuer, E., Doppler, T & Fetz, H. Gräber im näheren und weiteren Umfeld von Spreitenbach in Spreitenbach-Moosweg (Aargau, Schweiz): Ein Kollektivgrab um 2500 v Chr (ed. Doppler, T.) 233–266 (Basel 2012).
36. Doppler, T. Spreitenbach-Moosweg: ein Kollektivgrab um 2500 v. Chr. Archäologie Schweiz, Vol. 51 (2012).
37. Stehrenberger T. Obere Höhle bei der "Procha Burg" Höhlenpost 148 (2016).
38. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. Nature. 7555, 167–172 (2015).
39. Schuenemann, V. J. et al. Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. Nat Comms. 8, 15694 (2017).
40. Günther, T. et al. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. Proc Natl Acad Sci USA 38, 11917–11922 (2015).
41. Peltzer, A. et al. EAGER: efficient ancient genome reconstruction. Genome biology 17.1, 1 (2016).
42. Skoglund, P., Stora, J., Götherström, A. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. Journal of Archaeological Science 40.12, 4477–4482 (2013).
43. Bates, D., Maechler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 67(1), 1–48 (2015).
44. Lenth, R. V. Least-Squares Means: The R Package lsmeans. Journal of Statistical Software 69(1), 1–33 (2016).

Acknowledgements: We would like to thank the archaeological services of the Cantons of Aargau, Baselland, Bern, and Sankt Gallen for providing access to their collections as well as Anne Gibbson for proof reading the manuscript.

Author Contributions: J.K. conceived and designed the general research. A.F., E.R. and G.U.N. performed DNA-extraction and the preparation of sequencing libraries. I.S., N.S., S.L. and A.H. provided samples. I.S. performed morphological analysis of the Swiss skeletons from Oberbipp, Wartau and Zurzach. A.F. and N.A. performed bioinformatic and statistical analyses. A.F., N.A., V.J.S. and J.K. wrote the manuscript and S.L. contributed to it. All authors reviewed the manuscript.

Competing financial interests: The authors declare no competing financial interests.

Ratio of mitochondrial to nuclear DNA affects contamination estimates in ancient DNA analysis

Supplementary Information

Authors: Furtwängler A.[1*], Reiter E.[1], Neumann G.U.[1], Siebke I.[2], Steuri N.[3], Hafner A.[3], Lösch S.[2], Anthes A.[4], Schuenemann V.J.[1,5,6], Krause J.[1,6,7*]

Affiliations:
[1]Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Germany
[2]Department of Physical Anthropology, Institute of Forensic Medicine, University of Bern, Switzerland
[3]Institute of Archaeological Sciences and Oeschger Centre for Climate Change Research, University of Bern, Switzerland
[4]Institute of Ecology and Evolution, Animal Evolutionary Ecology group University of Tübingen, Germany.
[5]Institute of Evolutionary Medicine, University of Zurich, Switzerland
[6]Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Germany
[7]Max Planck Institute for the Science of Human History, Jena, Germany

* Corresponding author
E-mail: krause@shh.mpg.de, anja.furtwaengler@uni-tuebingen.de

Contents:

Refer to CD for Supplementary Tables 1 – 4

Supplementary Section 1 Archeological sites

Oberbipp, Canton Bern, Switzerland

Ramstein et al. 2014, Siebke et al. 2017

The Dolmen burial from Oberbipp was discovered and excavated in 2012. First ^{14}C results date to 5500 BP. The dolmen contains a minimal number of 40 individuals as determined from femora including males and females of all age classes. The burial is assigned to the Horgen Culture.

Spreitenbach, Canton Aargau, Switzerland

Doppler 2012

In 1997 the multiple burial of Spreitenbach, Canton Aargau in Switzerland was discovered and excavated. It contained five man, four women, two subadults and a new born. The burial is radio carbon dated to approximately 4500 BP and assigned to the Corded Ware complex.

Muttenz, Canton Basel, Switzerland

Bleuer et al. 2012

The multiple burial of Muttenz was discovered in 1969. It contained five individuals and was assigned to the Bell Beaker Complex.

Wartau, Canton St. Gallen, Switzerland

Stehrenberger 2016

The archaeological site is located in a natural cave. In the 1970s and 1980s, several inspections by archaeologist of the archaeological service of the Canton St. Gallen took place. The cave was accessible to the public at all time and in the 1970s some human remains were recovered illegally by private persons and transferred to the archaeological service in 2001. Radio carbon dates of bone fragments date to the Middle Neolithic period.

Seengen, Canton Aargau, Switzerland

Bleuer et al. 2012

The grave mound of Seengen was excavated in 1993. The burial contained two individuals, one of them consisted only of burned remains. The grave mound is archaeological dated to the Bronze Age.

Bad Zurzach, Canton Aargau, Switzerland

Bleuer et al. 2012, Gutzwiller 1994

The double burial of Bad Zurzach was excavated in 1984. It was assigned to the Bronze Age. Both individuals were buried in contracted position.

Supplementary Section 2 Laboratory workflow

40 petrous bones (P1, P4 - P7) and 36 teeth (T1) from the Neolithic sites in Switzerland were shotgun sequenced. Between 30 mg and 50 mg of powder from coronal dentin for teeth and between 50 mg and 100 mg of bone powder for the petrous bones was used for extraction as described previously (**Dabney et al. 2013)**. Sequencing libraries were prepared after **Meyer and Kircher 2010**. To enable multiplexed sequencing, sample specific barcodes were added to each library by amplification with tailed primers (**Kircher et al. 2011**). To achieve a high copy number of each library, an additional amplification was performed with Herculase II in a final volume of 100 µl consisting of 1 x Herculase II reaction buffer, 0.25 mMdNTPs, 0.4 µM IS 5, 0.4 µM IS 6, 0.01 % Herculase II Fusion DNA Polymerase. The thermal profile started with 2 min at 95 °C followed by sample specific number of cycles of 10 sec 95 °C, 30 sec 65°C, 30 sec 72 °C. This was then followed by 4 min at 72 °C. All samples were pooled equimolar at 10 nM and sequenced on a HiSeq with paired or single ends and NextSeq with paired ends. The mtDNA enrichment for MA315 and MA318 was performed as follows. For the generation of baits a long range PCR was performed on modern human mtDNA as described in **Maricic et al. 2010**. The purified products were sonicated with COVARIS shearing (Covaris, Woburn USA) to generate fragments of ~ 350 bp. For the preparation of the baits this step was followed by a blunt end reaction with Quick Blunting™ Kit of New England Biolabs and followed by a MinElute purification. To these fragments a double-stranded adapter was ligated generated from modified sequences after **Fu et al. 2013** according to the approach for generation of double-stranded adapters described in **Meyer and Kircher 2010**. The two adapters were generated from APL 5 (CGTGGATGAGGAGCCGCAGTG) and adapter1 rev (CACTGCGGCT) for the first adapter, the second adapter consists of APL6 (ATAGGGATCGCACCAGCGTGT) and adapter2 rev (ACACGCTGGT). Adapter ligation was performed using Quick Ligase (New England Biolabs) followed by a MinElute purification with an elution volume of 20 µl . For the fill in of the 5'-overhangs of the adapters the reaction had a total volume of 40 µl with 1 x Isothermal buffer, 125 nM dNTP, 0,4 U/µl Bst polymerase 2.0 from (BioLabs) and was incubated for 20 min at 37 °C and then 20 min at 80 °C. This was followed by an amplification within 8 reaction with Herculase II in a final volume of 100 µl consisting of 1 x Herculase II reaction buffer, 0,25 mM dNTPs, 0,4 µM APL 5, 0,4 µM APL 6, 0,01 % Herculase II Fusion DNA Polymerase and 30 cycles. Single-stranded probes were generated as in **Fu et al. 2013** using the APL2 primer. Before hybridisation 2 µg of target DNA were combined with 250 µM blocking oligonucleotides and incubated at 5 min at 95°C and 5 min at 65°C followed by at least 10 min at 37°C. This reaction was combined with 250 ng single-stranded bait of each fragment. The final hybridisation contains furthermore 1x HI-RPM hybridization buffer (aCGH Hybridization Kit; Agilent) and 1x Agilent blocking agent (aCGH Hybridization Kit; Agilent). The reaction is incubated over two nights at 65°C. For each hybridised library

pool 20 µl of MyOne M-270 streptavidin beads (Invitrogen) were washed twice with 1xBWT and then resuspended in 20 µl BWT. The hybridisation reaction is added to the beads and 160 µl BWT are added and the reaction is incubated at RT for 30 min. The beads were then washed with 200 µl BWT three times and then washed twice with 200 µl preheated (60°C) HWT and an incubation for 2 min. This was followed by washing with 200 µl BWT and 100 µl TET. Then 15 µl TET were added and the copy number was determined by qPCR (End volume of 20 µl: 1X DyNAmo Mastermix and 0,5 µM IS7 and IS8 and temperature profile of 10 min at 95 °C, then 40 cycles of 30 sec at 95 °C, 30 sec at 60 °C and 30 sec at 72 °C) followed by another amplification using Herculase II both directly on the beads. For MA315 and MA318 two more libraries with a half UDG-treatment (**Rohland et al. 2014**) were prepared and enriched for 1,240k nuclear SNPs as described in **Schuenemann et al. 2017** to achieve sufficient coverage on the X chromosome. The libraries MA307 till MA381 were paired-end sequenced on a HiSeq4000 (2x75 cycles). The libraries MA395 till MA482 and the libraries SA65 till SA74 were single-end sequenced on a HiSeq4000 for 75 cycles.

SI Table 5: mitochondrial and X-chromosome contamination rates and mt/nc ratios.

element	mt/nuc_ratio	ln of mt/nc ratio	MT contamination	lower95 MT contamination	upper95 MT contamination	X contamination	SE X contamination	lower95 X contamination	upper95 X contamination
PP	4	1,386294361	1,6	2,5	0,8	0,9	3,71E-04	0,9	0,9
T	9	2,197224577	3,6	5,1	2,5	1,4	8,18E-04	1,4	1,4
PP	17	2,833213344	1,7	5,4	0,3	1,1	3,30E-03	1,11	1,09
PP	20	2,995732274	1,6	4	0,6	1	3,90E-03	1,01	0,99
B	20	2,995732274	0,6	1,1	0,2	0,68	1,60E-03	0,68	0,68
PP	20	2,995732274	0,31	3,5	0,01	0,8	3,50E-03	0,81	0,79
PP	20	2,995732274	0,1	2,1	0,02	0,6	2,40E-03	0,6	0,6
PP	21	3,044522438	2,6	4,9	1,1	0,5	2,40E-03	0,5	0,5
B	29	3,36729583	0,4	1,3	0,1	0,82	1,70E-03	0,82	0,82
B	33	3,496507562	1,7	2,6	0,8	0,8	1,90E-03	0,8	0,8
B	38	3,63758616	0,6	0,9	0,6	0,7	5,30E-04	0,7	0,7
B	45	3,80666249	2	3,1	1	0,9	1,50E-03	0,9	0,9
B	48	3,87120101	1,7	3,3	0,9	0,1	1,00E-03	0,1	0,1
B	54	3,988984047	4	5,2	2,9	0,7	7,90E-04	0,7	0,7
B	72	4,276666119	0,7	1,7	0,2	0,2	1,10E-03	0,2	0,2
PP	74	4,304065093	0,2	0,8	0,01	0,6	4,00E-03	0,61	0,59
B	89	4,48863637	1	2	0,4	1,6	1,10E-03	1,6	1,6
T	411	6,018593215	1,6	1,9	1,2	0,039	3,17E-02	0,1	-0,02
B	708	6,562444094	0,1	0,5	0	0,98	1,20E-03	0,98	0,98
T	795	6,678342115	17	27,8	10,7	53	5,50E-01	54,08	51,92
B	870	6,768493212	2,8	3,3	2,3	35,27	3,14E-02	35,33	35,21
T	3157	8,057377489	0,75	1,1	0,4	16,1	3,43E-02	16,17	16,03
B	4169	8,335431478	0,7	1,1	0,4	0,009	3,35E-03	0,02	0
B	22126	10,00450867	1	1,1	0,8	31,5	3,28E-02	31,56	31,44
B	56387	10,93999391	0,4	1	0	16,1	2,80E-02	16,15	16,05

References

1. Ramstein, M., Schimmelpfennig, D., Lösch, S. Ein neolithischer Dolmen an der Steingasse in Oberbipp. *Archäologie Schweiz* **37(3)**, 4-15 (2014).
2. Siebke I., Furtwängler A., Hafner A., Krause J., Lösch S. An Interdisciplinary Project on the Neolithic Population of Modern Switzerland. 86th Annual Meeting of the American Association of Physical Anthropologists. *American Journal of Physical Anthropology*. p.358 (2017)
3. Bleuer E, Doppler T, Fetz H.. Gräber im näheren und weiteren Umfeld von Spreitenbach in *Spreitenbach-Moosweg (Aargau, Schweiz): Ein Kollektivgrab um 2500 v Chr* (ed. Doppler, T.) 233-266 (Basel 2012)
4. Doppler, T. *Spreitenbach-Moosweg: ein Kollektivgrab um 2500 v. Chr*. Archäologie Schweiz, **Vol. 51** (2012).
5. Stehrenberger T. Obere Höhle bei der "Procha Burg" *Höhlenpost* **148** (2016).
6. Gutzwiller P. Die vorrömische Besiedlung des Fleckens Zurzach. *Jahrbuch der Schweizerischen Gesellschaft für IJr- und Frühgeschichte* **77**:7-33 (1994).
7. Dabney, J., et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences.* **39**, 15758–15763 (2013).
8. Meyer M., Kircher M. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols*. **5448** (2010).
9. Kircher, M., Sawyer, S.Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research.* **1**, e3 (2011).
10. Maricic, T., Whitten, M., Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PloS ONE* **5(11)**, e14004 (2010).
11. Fu, Q., et al. DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences* **110(6)**, 2223-2227 (2013).
12. Rohland N., Harney E., Mallick S., Nordenfelt S., Reich D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **370**, 20130624 (2014).
13. Schuenemann V.J., Peltzer A., et al. Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. *Nat Comms*. **8**, 15694 (2017)

Paper II

Furtwängler A., Neukamm J., Böhme L., Reiter E., Vollstedt M., Arora N., Singh P., Cole S.T., Knauf S., Calvignac-Spencer S., Krause-Kyora B., Krause J., Schuenemann V.J., Herbig A. "Comparison of target enrichment strategies for Pathogen studies in ancient DNA." *bioRxiv*

https://doi.org/10.1101/2020.07.09.195065

Comparison of target enrichment strategies for ancient pathogen DNA

Authors: Anja Furtwängler[1,2*], Judith Neukamm[1,3,4], Lisa Böhme[5], Ella Reiter[1], Melanie Vollstedt[5], Natasha Arora[6], Pushpendra Singh[7], Stewart T. Cole[8], Sascha Knauf[9,10], Sébastien Calvignac-Spencer[11], Ben Krause-Kyora[5,12], Johannes Krause[1,2,12], Verena J. Schuenemann[1,2,3#,*], Alexander Herbig[1,12#,*]

Affiliations:
[1]Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Germany
[2]Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Germany
[3]Institute of Evolutionary Medicine, University of Zurich, Switzerland.
[4]Institute for Bioinformatics and Medical Informatics, University of Tübingen, Germany
[5]Institute of Clinical Molecular Biology, Kiel University, Germany
[6]Zurich Institute of Forensic Medicine, University of Zurich, Switzerland
[7]Indian Council of Medical Research-National Institute of Research in Tribal Health, Jabalpur, MP, India
[8]Institute Pasteur, Paris, France
[9]Deutsches Primatenzentrum GmbH, Leibniz-Institute for Primate Research, Goettingen, Germany
[10]Department for Animal Sciences, Georg-August-University, Goettingen, Germany
[11]Robert Koch Institut, Berlin, Germany
[12]Max Planck Institute for the Science of Human History, Jena, Germany

#These authors jointly supervised this study.

* Corresponding authors: anja.furtwaengler@uni-tuebingen.de, verena.schuenemann@iem.uzh.ch, herbig@shh.mpg.de

Abstract

In ancient DNA research, the degraded nature of the samples generally results in poor yields of highly fragmented DNA, and targeted DNA enrichment is thus required to maximize research outcomes. The three commonly used methods – (1) array-based hybridization capture and in-solution capture using either (2) RNA or (3) DNA baits – have different characteristics that may influence the capture efficiency, specificity, and reproducibility. Here, we compared their performance in enriching pathogen DNA of *Mycobacterium leprae* and *Treponema pallidum*

of 11 ancient and 19 modern samples. We find that in-solution approaches are the most effective method in ancient and modern samples of both pathogens, and RNA baits usually perform better than DNA baits.

Method summary

We compared three targeted DNA enrichment strategies used in ancient DNA research for the specific enrichment of pathogen DNA regarding their efficiency, specificity, and reproducibility for ancient and modern *Mycobacterium leprae* and *Treponema pallidum* samples. Array-based capture and in-solution capture with RNA and DNA baits were all tested in three independent replicates.

Main Text

The field of ancient DNA (aDNA), which studies DNA retrieved from paleontological and archaeological material, was revolutionized by the invention of high-throughput sequencing (HTS). In combination with HTS, the development of targeted DNA enrichment protocols has made a crucial contribution in advancing aDNA research during the last decade.

As DNA decays over time, aDNA is usually only present in trace amounts of highly fragmented sequences (**1, 2, 3**). Detecting endogenous pathogen aDNA from archaeological material is additionally compounded by the larger amount of background DNA from the environment including soil microorganisms. Furthermore, the background of host DNA in ancient remains is an additional obstacle in order to obtain ancient pathogen DNA. Shotgun sequencing of libraries from aDNA extracts to sufficient genomic coverage is, therefore, cost-intensive (**4**). To circumvent this problem, specific regions of interest such as bacterial chromosomes, mammalian mitochondrial genomes, or regions with single-nucleotide-polymorphisms (SNP) are often target-enriched before sequencing (**4**). Aside from its application in aDNA sequencing, targeted DNA enrichment is also useful to retrieve pathogen DNA from clinical samples, particularly for infectious agents that are found in low quantities in the host organism and which are difficult to culture, as is the case for *Mycobacterium leprae* and *Treponema pallidum*. Removal of background DNA prior to sequencing increases the yield of pathogen DNA, and thus allows valuable information for epidemiologists investigating outbreaks to be obtained.

For the enrichment of entire bacterial and mammalian chromosomes, there are currently three methods available, which are based on hybridization capture (**5**): DNA microarrays (here represented by SureSelect from Agilent Technologies), in-solution capture with DNA baits (represented by SureSelect from Agilent Technologies according to Fu and colleagues **(6)**) and in-solution capture with RNA baits (here represented by myBaits® from Arbor Biosciences).

In the case of the DNA array-based method, up to a million artificial DNA baits are printed on the surface of a glass slide **(7)**. Additionally, there is the possibility to perform in-solution capture with baits cleaved from the glass slides and used right away or immortalized in DNA bait libraries (**6**). The second in-solution approach uses up to 100,000 artificial RNA baits. The three approaches rely on the hybridization of target fragments to the complementary sequence

of the baits (immobilized or in-solution), which can be levered to wash background DNA away.

To date there has been to our knowledge, no statistical comparison of the performance of all three methods: microarrays, in-solution capture with DNA baits, and in-solution capture with RNA baits (**6**). So far only microarrays and the in-solution capture with DNA baits were compared for *Salmonella enterica* and no replicates for statistical assessment were produced (**8**). Here, we present results from the enrichment of modern and ancient samples containing pathogen DNA, using the three aforementioned approaches. All samples had previously tested positive but had also shown low amounts of target DNA for *M. leprae* or *T. pallidum* (Supplementary Table 1).

The different enrichment concepts tested were chosen to represent methods as they are applied in ongoing research and therefore not only differ in the technology used (DNA *vs.* RNA baits, immobilized *vs.* in-solution) but also in the design such as bait length and number of unique baits, which might have an effect on the performance.

We used eight ancient samples positive for *M. leprae* and six modern libraries from leprosy patients that were shown to contain *M. leprae* DNA (Supplementary Note 1). Genetic data from the ancient and modern *M. leprae* samples were previously published in **9** and **10**. Samples with less than 0.6 % endogenous bacterial DNA were selected.

Modern *T. pallidum* samples (n=13) were previously published in **12** and **13**. Three ancient extracts of *T. pallidum* were used from **14**. The portion of endogenous DNA for the selected *T. pallidum* samples was below 0,01 % for ancient and modern samples.

Starting from existing sequencing libraries all three methods were applied with three independent replicates each (see Figure. 1 and Supplementary Note 1 for a detailed description of the methods, the newly generated data is available at the Sequence Read Archive under the BioProject PRJNA645054). Following the manufacturer's suggestion for libraries with low yields of target DNA, we performed two successive rounds of hybridization for all methods. To investigate the effectiveness of this procedure, we compared results from the first and second rounds for the in-solution capture with RNA baits. We then evaluated differences in efficiency, reproducibility, and specificity across the three approaches by calculating mean coverage, standard deviation of the mean coverage, enrichment factor (calculated by dividing the % of target DNA after enrichment by the % of target DNA in the shotgun data), and the % of the genome covered 5-fold or more after normalizing the data of each bacterial species to the same number of raw reads (Supplementary Tables 2, 3 & 5 and Supplementary Figures 1 & 2).

For most ancient samples, the highest mean coverage (Figure 2A) is reached with the RNA bait in-solution capture (eight out of eleven, more details can be found in Supplementary Note 2 & 3, and SSupplementary Tables 1 & 2). On average the RNA bait capture results in a 1.5 and 20.0 times higher mean coverage than the DNA bait or the array capture, respectively. As illustrated in Figure. 2B, the highest enrichment factor is obtained in the RNA bait capture of ancient *T. pallidum* DNA (all three samples) and *M. leprae* (four samples showed best results for the RNA bait, three for the DNA bait, and one for the array), with values between 2-150x

higher, compared to the other two approaches. An in-solution approach seems, therefore, to be advantageous for enriching ancient pathogen DNA.
A similar pattern can be observed in the data of the modern *M. leprae* and *T. pallidum* samples (Figures. 2A and 2B) further highlighting the performance of the in-solution approach in general and RNA baits in particular.
In-solution capture with DNA baits was used with robot-assistance in this study whereas the in-solution capture with RNA baits was performed in two different labs. Unsurprisingly, the DNA bait capture showed the smallest differences (2- to 50-fold lower) between the replicates whereas the RNA bait capture showed the largest and the DNA array capture was intermediate. Consistent conditions are therefore crucial for reproducibility.
Another important feature of targeted enrichment is specificity. We estimated the specificity of the three tested methods by comparing the number of reads specific to either *M. leprae* or *T. pallidum* in comparison to general mycobacterial or treponemal reads, respectively (Figure 2 C). Here, differences between the two pathogens can be observed. In the ancient and modern *T. pallidum* samples, the RNA bait capture consistently shows the highest proportion (up to 1.5 times higher) of specific reads. The same trend was observed for the libraries prepared from recent leprosy patient samples, i.e. modern samples of *M. leprae*. Only for ancient *M. leprae* samples, the DNA bait capture is more specific. The highest percentages of specific reads are not necessarily found in samples with high percentages of endogenous DNA in the shotgun data before enrichment.
For ancient and modern samples, due to high efficiency, reproducibility and specificity in-solution approaches are highly recommendable.
Two rounds of hybridization are routinely performed in aDNA research, which is expected to improve enrichment but may also reduce data complexity in terms of portions of unique reads. To formally investigate the effect of the second round of capture, we also sequenced the libraries only enriched with one round of hybridization with the RNA baits and compared the results to the second round of hybridization. The second round of hybridization resulted in an increase in the enrichment factor for ancient and modern *M. leprae* samples (with an average of 2x increase) as well as for *T. pallidum* samples (with an average of 17x increase), demonstrating the utility of such a second round of hybridization capture (Supplementary Table 5). On the other hand, when comparing the library complexity (Figure. 2 D and Supplementary Note 2 & 3, Supplementary Figure 3), we found a substantial loss of complexity after the second round of hybridization in all modern and ancient samples. This loss was reflected in the higher percentage of unique reads in all the reads mapped after the first round. Therefore, if the portion of endogenous DNA in a sample is high in the beginning it may be worthwhile considering whether a single round of capture combined with deeper sequencing is sufficient or even advantageous.
The three protocols also differ in terms of cost and effort. The most cost-intensive is the array-capture approach (~673 € per sample), which requires additional equipment that is not usually necessary with the other approaches. The in-solution capture with DNA baits is, by contrast, cheaper once the baits are cleaved from the glass slide (~56,23€ per sample), but the version

that can be used for the immortalization of the baits by transforming them into a library is not freely available. The in-solution capture with RNA baits is more comparable to the DNA bait capture than to the array with ~109 € per sample and it also needs the lowest number of additional equipment and reagents (Supplementary Table 7).

After a detailed comparison of the three tested methods it can be concluded that for ancient and modern pathogen samples, the RNA bait capture with two rounds of hybridization seems to be the most suitable. The generally high performance of the in-solution approach (mainly the one with RNA baits) for both bacterial species suggests that the findings are highly representative and comparable performance is also expected for a variety of other bacterial/microbial organisms.

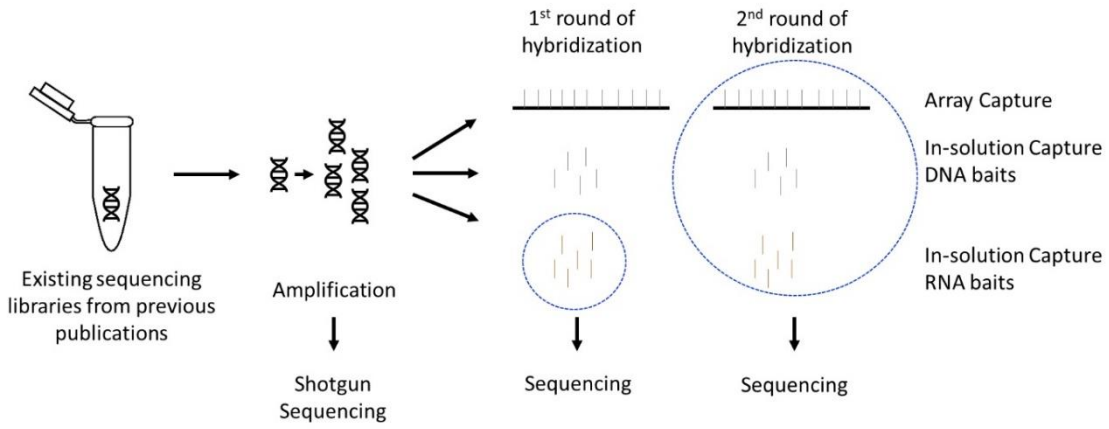


Figure 1. Schematic representation of the workflow. For all samples, the three different enrichment protocols were tested in three independent replicates. Blue circles indicate the libraries that were sequenced at each particular step.

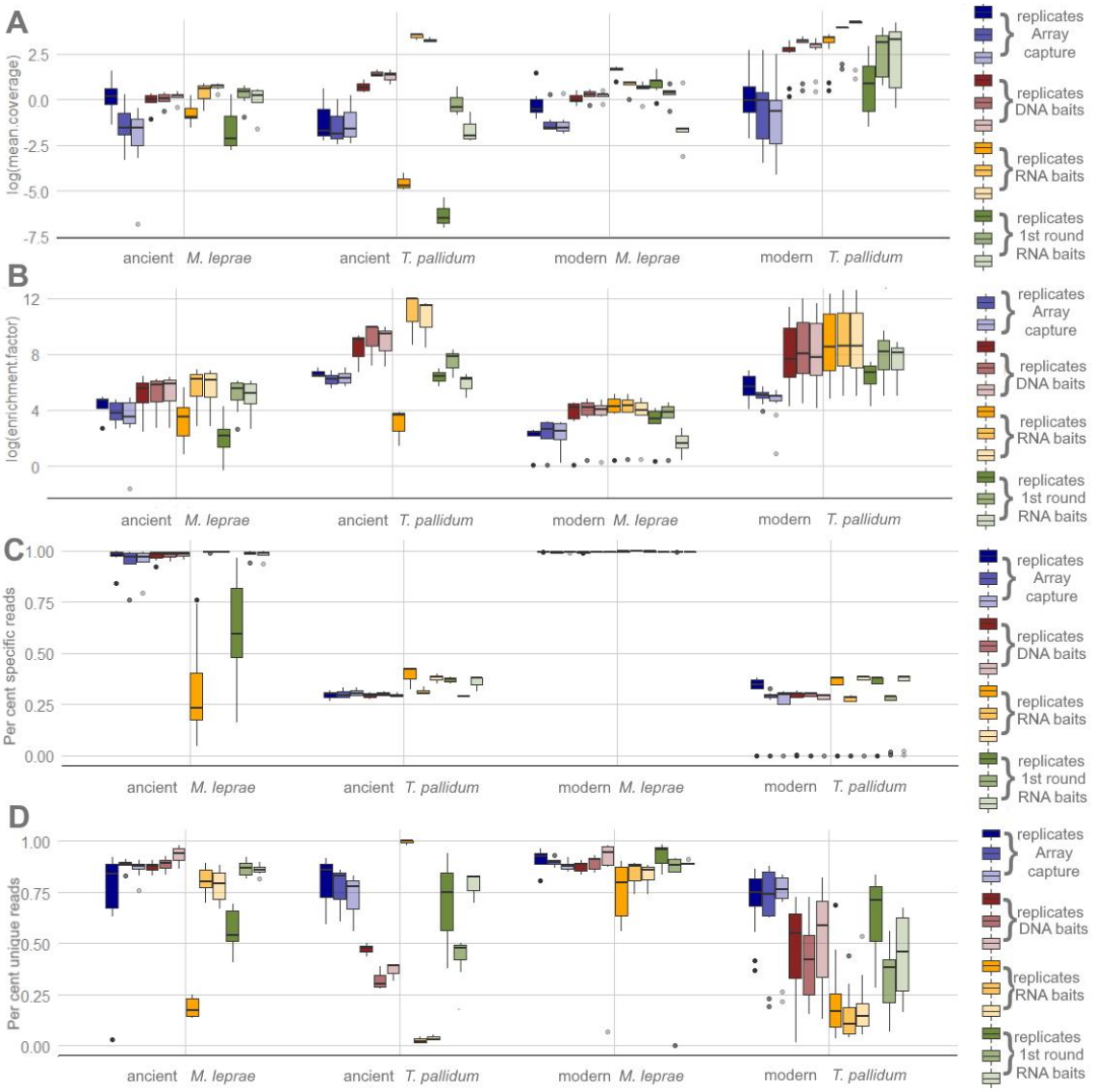


Figure 2. Differences between the three tested protocols in ancient and modern *M. leprae* and *T. pallidum* samples. A) Log-transformed values of the mean coverage. B) log-transformed values of the enrichment factor calculated by dividing the percentage of endogenous DNA by the percentage of endogenous DNA after shotgun sequencing. C) The proportion of specific reads corresponding to *M. leprae* and *T. pallidum* compared to other mycobacterial and treponemal reads, respectively. D) Percentage of unique reads calculated by the number of unique reads divided by the total number of sequences mapped to represent library complexity in *M. leprae* and *T. pallidum* samples.

References

1. Sawyer, Susanna; Krause, Johannes; Guschanski, Katerina; Savolainen, Vincent; Pääbo, Svante (2012): Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. In: *PloS one* 7 (3), e34131. DOI: 10.1371/journal.pone.0034131.
2. Allentoft, Morten E.; Collins, Matthew; Harker, David; Haile, James; Oskam, Charlotte L.; Hale, Marie L. et al. (2012): The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. In: *Proceedings. Biological sciences* 279 (1748), S. 4724–4733. DOI: 10.1098/rspb.2012.1745.
3. Briggs, Adrian W.; Stenzel, Udo; Johnson, Philip L. F.; Green, Richard E.; Kelso, Janet; Prüfer, Kay et al. (2007): Patterns of damage in genomic DNA sequences from a Neandertal. In: *Proceedings of the National Academy of Sciences of the United States of America* 104 (37), S. 14616–14621. DOI: 10.1073/pnas.0704665104.
4. Krause, Johannes (2010): From Genes to Genomes: What is New in Ancient DNA? In: *Mitteilungen der Gesellschaft für Urgeschichte* 19, S. 11–33.
5. Spyrou, Maria A.; Bos, Kirsten I.; Herbig, Alexander; Krause, Johannes (2019): Ancient pathogen genomics as an emerging tool for infectious disease research. In: *Nature reviews. Genetics* 20 (6), S. 323–340. DOI: 10.1038/s41576-019-0119-1.
6. Fu, Qiaomei; Meyer, Matthias; Gao, Xing; Stenzel, Udo; Burbano, Hernán A.; Kelso, Janet; Pääbo, Svante (2013): DNA analysis of an early modern human from Tianyuan Cave, China. In: *Proceedings of the National Academy of Sciences of the United States of America* 110 (6), S. 2223–2227. DOI: 10.1073/pnas.1221359110.
7. Vågene, Åshild J.; Herbig, Alexander; Campana, Michael G.; Robles García, Nelly M.; Warinner, Christina; Sabin, Susanna et al. (2018): Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. In: *Nature ecology & evolution* 2 (3), S. 520–528. DOI: 10.1038/s41559-017-0446-6.
8. Burbano, Hernán A.; Hodges, Emily; Green, Richard E.; Briggs, Adrian W.; Krause, Johannes; Meyer, Matthias et al. (2010): Targeted investigation of the Neandertal genome by array-based sequence capture. In: *Science (New York, N.Y.)* 328 (5979), S. 723–725. DOI: 10.1126/science.1188046.
9. Schuenemann, Verena J.; Singh, Pushpendra; Mendum, Thomas A.; Krause-Kyora, Ben; Jäger, Günter; Bos, Kirsten I. et al. (2013): Genome-wide comparison of medieval and modern Mycobacterium leprae. In: *Science (New York, N.Y.)* 341 (6142), S. 179–183. DOI: 10.1126/science.1238286.
10. Schuenemann, Verena J.; Avanzi, Charlotte; Krause-Kyora, Ben; Seitz, Alexander; Herbig, Alexander; Inskip, Sarah et al. (2018): Ancient genomes reveal a high diversity of Mycobacterium leprae in medieval Europe. In: *PLoS pathogens* 14 (5), e1006997. DOI: 10.1371/journal.ppat.1006997.
11. Knauf, Sascha; Gogarten, Jan F.; Schuenemann, Verena J.; Nys, Hélène M. de; Düx, Ariane; Strouhal, Michal et al. (2018): Nonhuman primates across sub-Saharan Africa are infected with the yaws bacterium Treponema pallidum subsp. pertenue. In: *Emerging microbes & infections* 7 (1), S. 157. DOI: 10.1038/s41426-018-0156-4.
12. Arora, Natasha; Schuenemann, Verena J.; Jäger, Günter; Peltzer, Alexander; Seitz, Alexander; Herbig, Alexander et al.: Origin of modern syphilis and emergence of a pandemic Treponema pallidum cluster. In: *Nat Microbiol* 2 (1), S. 1–6. DOI: 10.1038/nmicrobiol.2016.245.
13. Schuenemann, Verena J.; Kumar Lankapalli, Aditya; Barquera, Rodrigo; Nelson, Elizabeth A.; Iraíz Hernández, Diana; Acuña Alonzo, Víctor et al. (2018): Historic Treponema pallidum genomes from Colonial Mexico retrieved from archaeological remains. In: *PLoS neglected tropical diseases* 12 (6), e0006447. DOI: 10.1371/journal.pntd.0006447.

Author contributions

V.J.S., A.H. and J.K. conceived of the study. B.K. and S.C-S. provided RNA baits and sequencing libraries. N.A., P.S., S.T.C., S.K. provided sequencing libraries. A.F., L.B., E.R., M.V. performed the laboratory work. A.F. and J.N. performed the data analysis. A.F. and A.H. conducted the statistical analysis. A.F. designed the figures. A.F., V.J.S, and A.H. wrote the manuscript with input from all authors. All authors reviewed the manuscript.

Acknowledgements

We thank all our colleagues providing samples for our study: Sarah Inskip (University of Cambridge, UK), Helen Donoghue (University College London, UK), Rodrigo Barquera (Max Planck Institute for the Science of Human History, Germany), Michael Taylor (University of Surrey, UK), Thomas Mendum (University of Surrey, UK), Graham Stewart (University of Surrey, UK), Simon Roffey (The Magdalen Hill Archaeological Research Project (MHARP) Winchester, UK), Phil Marter (The Magdalen Hill Archaeological Research Project (MHARP) Winchester, UK), Katie Tucker (Deutsches Archäologisches Institut, Berlin, Germany), Fabian Leendertz (Robert Koch Insitute, Berlin, Germany), Roman Wittig (Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany), Anna Kjellström (University of Stockholm; Sweden), Christos Economou (University of Stockholm, Sweden), Petr Velemínský (National Museum, Czech Republic) Antónia Marcsik (University of Szeged, Hungary), Erika Molnár (University of Szeged, Hungary), György Pálfi (University of Szeged, Hungary), Valentina Mariotti (University of Bologna, Italy; Aix-Marseille Université, France), Alessandro Riga (University of Florence, Italy), M. Giovanna Belcastro (University of Bologna, Italy; Aix-Marseille Université, France), Jesper L. Boldsen (University of Southern Denmark, Denmark), and Charlotte Avanzi (Colorado State University, USA). The authors would also like to thank the laboratory team of the MPI for the Science of Human History in Jena for extensive support with capture experiments and sequencing and the developer of EAGER 2: Alexander Peltzer and James Fellows Yates (Max Planck Institute for the Science of Human History, Jena, Germany). This work was supported by the University of Zurich's University Research Priority Program "Evolution in Action: From Genomes to Ecosystems" (V.J.S), the Max-Planck Society (J.K., A.H.), the Senckenberg Centre for Human Evolution and Palaeoenvironment (S-HEP) at the University of Tübingen (V.J.S., J.K., A.F.). P.S.'s research work is supported by ICMR, DBT India, R2STOP Canada, and the Leprosy Research Initiative Netherlands. The manuscript has been approved by the Publication Screening Committee of ICMR-NIRTH, Jabalpur, and assigned with the number ICMR-NIRTH/PSC/44/2020.

Comparison of target enrichment strategies for ancient pathogen DNA

Supplementary Information

Authors: Anja Furtwängler[1,2*], Judith Neukamm[1,3,4], Lisa Böhme[5], Ella Reiter[1], Melanie Vollstedt[5], Natasha Arora[6], Pushpendra Singh[7], Stewart T. Cole[8], Sascha Knauf[9,10], Sébastien Calvignac-Spencer[11], Ben Krause-Kyora[5,12], Johannes Krause[1,2,12], Verena J. Schuenemann[1,2,3#,*], Alexander Herbig[1,12#,*]

Affiliations:
[1]Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Germany
[2]Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Germany
[3]Institute of Evolutionary Medicine, University of Zurich, Switzerland.
[4]Institute for Bioinformatics and Medical Informatics, University of Tübingen, Germany
[5]Institute of Clinical Molecular Biology, Kiel University, Germany
[6]Zurich Institute of Forensic Medicine, University of Zurich, Switzerland
[7]Indian Council of Medical Research-National Institute of Research in Tribal Health, Jabalpur, MP, India
[8]Institute Pasteur, Paris, France
[9]Deutsches Primatenzentrum GmbH, Leibniz-Institute for Primate Research, Goettingen, Germany
[10]Department for Animal Sciences, Georg-August-University, Goettingen, Germany
[11]Robert Koch Institut, Berlin, Germany
[12]Max Planck Institute for the Science of Human History, Jena, Germany

#These authors jointly supervised this study.
* Corresponding authors: anja.furtwaengler@uni-tuebingen.de, verena.schuenemann@iem.uzh.ch, herbig@shh.mpg.de

Contents:

Supplementary Note 1 – Laboratory workflow

Sample Selection

The modern and ancient DNA extracts used in this study were previously tested positive for DNA from *Mycobacterium leprae* (**Schuenemann et al. 2013, Mendum et al. 2014, Schuenemann et al. 2018a**) or *Treponema pallidum* (**Arora et al. 2016, Knauf et al. 2018, Schuenemann et al. 2018b**). Existing libraries of the previous studies were used. For a comparison of the three methods under investigation, an additional shotgun sequencing of all samples was performed to determine the percentage of target DNA in the libraries prior to enrichment.

Ethics statement

For all samples used in this study only existing sequencing libraries were used, and no new material was collected. Statements about ethical approval and research permission can be found in the original publications (Supplementary Table 1). In this study only sequencing data of the two bacteria and no human data was generated.

Array capture

The array capture was performed according to the methods described in **Hodges et al. 2007**. The array design was identical to **Schuenemann et al. 2013** (*M. leprae*) and **Arora et al. 2016** (*T. pallidum*). Probe length on both arrays was 60 bp. Modern and ancient samples positive for *M. leprae* were pooled equimolar and captured on two arrays. For the *Treponema* samples, three pools for the capture were prepared: one for modern and ancient syphilis samples respectively and a third of positive extracts originating from different species of nonhuman primates. After the hybridization, the products were quantified by qPCR as described in **Schuenemann et al. 2013**. After determination of the sufficient number of cycles the pools were amplified and after quantification on an Agilent 2100 Bioanalyzer the pools were diluted for sequencing on a HiSeq4000 using a 75bp single-end kit for the first replicates and 75bp paired-end for the following two replicates.

In-solution capture with RNA baits (MYBaits)

The in-solution capture was performed using biotinylated RNA baits from the MYBaits from MYcroarray® according to manufactures instructions. The first replicate was performed in the post-amplification laboratory of the Ancient DNA Laboratory at the Kiel University after the manual of version 1.3.8. The following two replicates were performed in the post-amplification laboratory of the AG Palaeogenetics at the Institute for Archaeological Science in Tübingen after the manual of version 3.02. Samples were pool identical to the array capture and the capture was followed by a similar procedure to prepare for sequencing.

For the baits for *M. leprae* the reference Br4923 (NC_011896) was used as a base for the bait design with a 2x tiling as well and 76,490 baits in the final set. Bait length was 80 bp for *M. leprae*. With regard to *T. pallidum* the baits spanned the simian derived *T. pallidum* ssp. *pertenue* strain Fribourg-Blanc reference genome (NC_021179.1) with a 2x tiling as described in **Knauf et al. 2018**. Bait length for *T. pallidum* is 100bp and the bait set contains 19,925 unique probes. The two rounds from the first repetition were sequenced on a HiSeq4000 with a 75bp single-end kit. The two rounds of the second and third repetition using a 75bp paired-end kit.

In-solution capture with DNA baits (probes derived from arrays)

The in-solution capture with array probes was performed at the Max Planck Institute for the Science of Human History in Jena.
For *M. leprae,* probes were designed based on strains TN (NC_002677.1) and Br4923 (NC_011896.1). For targeted enrichment of *Treponema pallidum* DNA probes were designed on the basis of *T. pallidum* ssp. *pallidum* strains Nichols (NC_000919.1), SS14 (NC_021508.1), Sea 81-4 (NZ_CP003679.1), Mexico A (NC_018722.1), *T. pallidum* ssp. *endemicum* strain Bosnia A (NZ_CP007548.1), and *T. pallidum* ssp. *pertenue* strain Fribourg-Blanc (NC_021179.1). The tiling density is two and one bp for *M. leprae* and *T. pallidum*, respectively. For both target organisms the probe length is 52 bp with an additional 8bp linker sequence (CACTGCGG) as described in **Fu et al. 2013**. Duplicated probes and probes with low sequence complexity were removed. This resulted in 1,125,985 and 1,593,068 unique probe sequences for *T. pallidum* and *M. leprae*, respectively. For each target species the probe set was spread on two Agilent one-million feature SureSelect DNA Capture Arrays. The capacity of the two arrays was filled by randomly duplicating probes from the probe set. The arrays were turned into in-solution DNA capture libraries as described in **Fu et al. 2013**. All three replicates were sequenced on a HiSeq4000 using a 75bp paired-end kit.

Supplementary Note 2 - Bioinformatics and Statistical Analysis

The sequencing data of all samples was processed with the EAGER 2 pipeline (**Peltzer et al. 2016, Fellows Yates et al. 2020,** https://github.com/nf-core/eager). Including mapping with BWA, removal of duplicates, and the generation of damage plots. The enrichment factor for all enriched libraries was calculated by dividing the percentage of endogenous DNA after enrichment by the percentage of endogenous DNA in the shotgun sequencing.
All statistical analysis was performed in R version 426 3.4.3 (R Core Team 2017). The significance between the different features mean coverage, standard deviation of the mean coverage, percentage of the genome covered five-old, fragment length and ancient DNA (aDNA) specific damage in the data of the three tested protocols was assessed in each sample individually with a mixed model as implemented in the lme4-package for R (**Bates et al. 2015**). This model contained the replicate as a random slope factor allowing between-replicate variation in the main effect. Subsequently, p-values were corrected for multiple testing with the Bonferroni correction. Pairwise Tukey HSD post hoc tests were calculated from this model using the *lsmeans* package for R (**Lenth 2016**).
Also, significant differences between the individual replicates (grouped by age and pathogen) were assessed using a linear model as implemented in the stats-package for R (R Core Team (2019)).
The percentage of unique reads was calculated by dividing the number of unique reads by the number of total reads mapped.

Variance within each method

After calculating the absolute value of the pairwise differences between the replicates for each method we used one-way ANOVA to determine the significance of these differences followed by Tukey HSD post hoc tests.

Specificity of the three tested methods

We also used one-way ANOVA to determine the significance of differences in the ratio of specific reads for each sample individually followed by Tukey HSD post hoc tests. Subsequently, p-values were corrected for multiple testing with the Bonferroni correction. General mycobacterial or treponemal reads, as well as specific reads, were determined using the MALT algorithm (**Vågene et al. 2018**). For the ratio, the number of specific reads was divided by either the number of total mycobacterial or treponemal reads.

Data upload

For the samples derived from human patients the reads mapping to the human genome were removed from the fastq files prior data upload with the *--strip_input_fastq* flag of EAGER 2 (**Peltzer et al. 2016, Fellows Yates et al. 2020,** https://github.com/nf-core/eager) while mapping to the hg19 reference genome.

Supplementary Note 3 – Results of the Bioinformatics and Statistical Analysis

Capture efficiency

Detailed results of the tests for significant differences in the mean coverage, standard deviation of the mean coverage, percentage of the genome covered five-fold, enrichment factor, as well as aDNA typical damage and fragment length for each individual sample can be found in Supplementary Table 2 and Supplementary Figures 1 & 2.

Mean coverage and the percentage of the genome covered at least five-fold are highly dependent on the enrichment factor and therefore the results for the two features mirror that of the enrichment factor (see Main Manuscript and Supplementary Table 2).

For the features enrichment factor and mean coverage in the ancient data of both bacteria, the RNA bait capture with two rounds of hybridization shows the best results. However, adjusted p-values do not reach significance. The fragment length is in the data of ancient samples of both bacteria the shortest in the in-solution capture with DNA probes. The differences in fragment length are significant in both cases and increase with the bait length used with the longest fragments in the RNA bait capture with either one or with two rounds of hybridization.

The evenest coverage as represented by lowest values of the standard deviation of the mean coverage is seen in the DNA bait capture in the ancient *T. pallidum* samples and in the data of the RNA bait capture with two rounds of hybridization in the ancient *M. leprae* samples. In both cases, differences are not significant.

The largest portions of the genome covered at least five-fold result from RNA bait capture with two rounds of hybridization in ancient *M. leprae* sample (three to twenty times higher) and from DNA bait capture in the ancient *T. pallidum* samples (in average a hundred times higher, results of statistical significance in Supplementary Table 2). However, also in these cases the adjusted p-values do not reach significance.

An important characteristic of ancient sequencing libraries is the occurrence of the substitution of C by T at the fragment ends (**Briggs et al. 2007**). This is due to the post-mortem decay of the DNA and can be used to authenticate ancient DNA. In the data of both bacteria the array capture results in the highest portion (up to two times higher) of damaged fragment. However, differences are not significant.

Also, in the modern samples enrichment factor and mean coverage are the highest (between three and six hundred times higher) in the data of the RNA bait capture with two rounds of hybridization for both bacteria. All adjusted p-values for the enrichment factor and most of the adjusted p-values for the mean coverage are significant.

The evenest coverage in the modern data is found in the array capture data for *T. pallidum* and in the DNA bait capture data for the *M. leprae*. For *T. pallidum* these differences are significant.

The percentage of the genome covered five-fold is highest for modern *M. leprae* in the RNA bait capture with two rounds of hybridization and for *T. pallidum* in the DNA bait capture. However, only for *M. leprae* differences are significant.

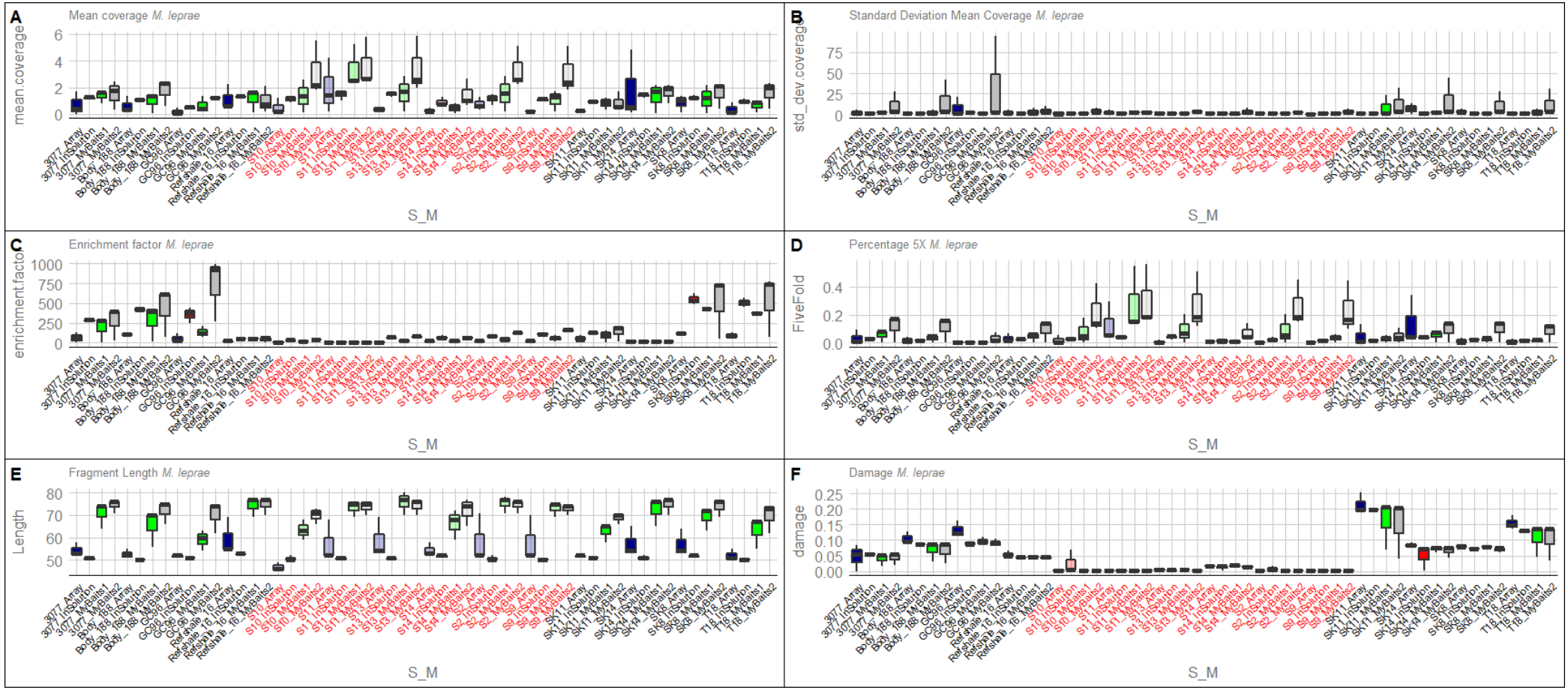
The longest fragment in the modern data are as well found in the methods with the longest baits. For *M. leprae* the RNA bait capture, as for the ancient sample with either one or two rounds of hybridization, results in the longest fragments. For *T. pallidum* the DNA bait capture results in the longest fragments, there the differences also reach significance.
All comparisons between the different methods were also performed between the individual replicates. Detailed results for this comparison can be found in Supplementary Table 3. The general pattern found in sample wise comparison was confirmed with higher statistical significance in the tests performed per replicate.
The number of unique reads in the data of the first and second round of hybridization with the RNA baits does not significantly increase with the second round (Supplementary Figure 3). Showing that the increase in the percentage of endogenous DNA increases while library complexity decreases.

Variance within each method

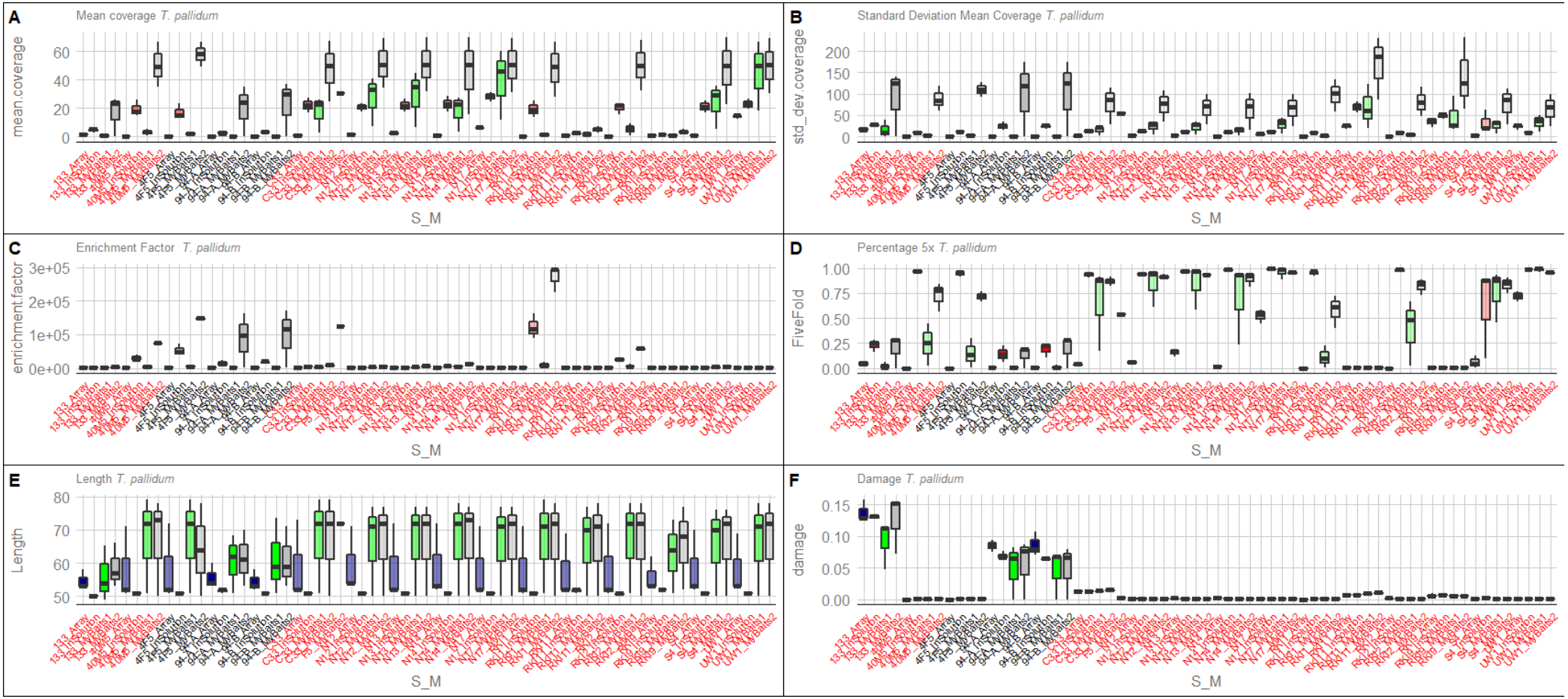
The choice of method significantly affects the differences between the replicates for all tested features in modern and ancient *M. leprae* genomes (Supplementary Table 6). The in-solution capture using DNA probes shows the smallest differences between the replicates besides the enrichment factor. Here the array capture produces the most similar results between the replicates (Supplementary Table 6).
For the data from the ancient and modern syphilis, the variance between the replicates is significantly affected as well. The array capture shows hereby the smallest differences between the single replicates.

Specificity of the different methods

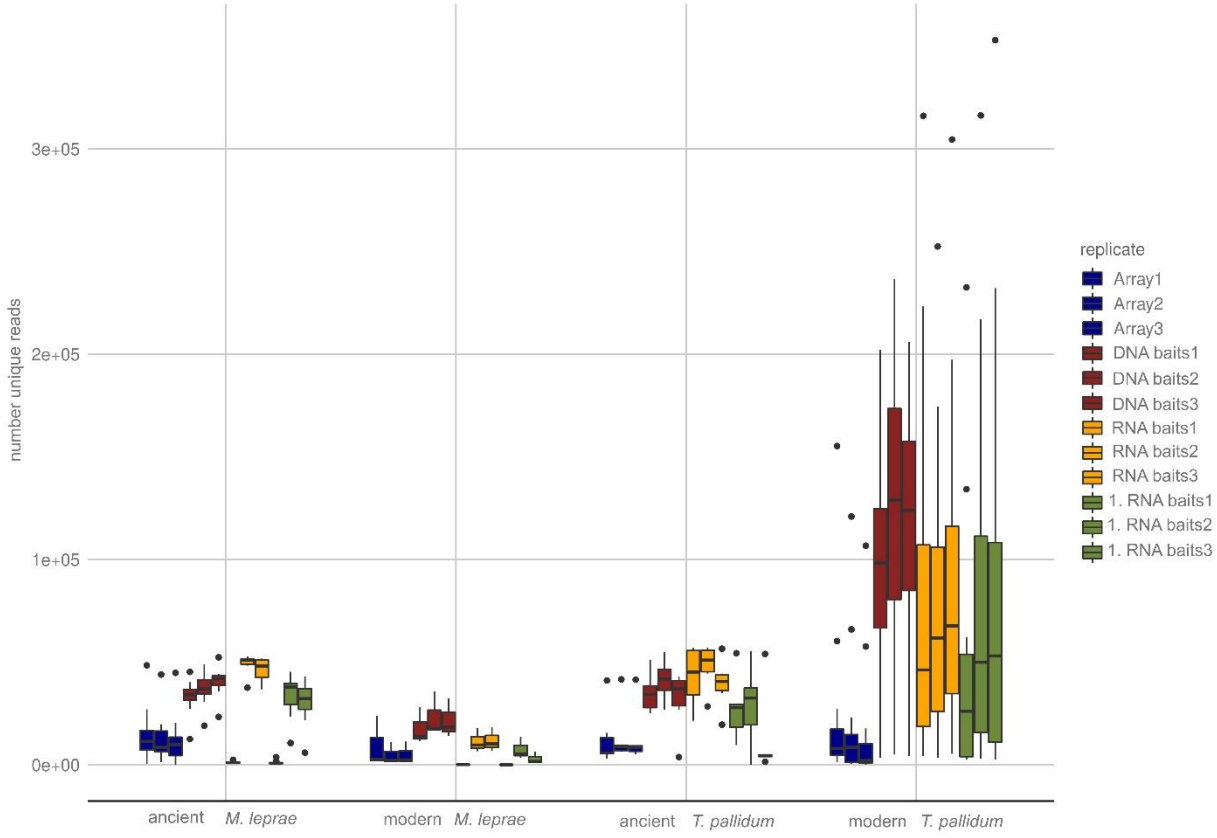
The significance of differences in the ratio of specific reads for either *M. leprae* or *T. pallidum* compared to mycobacterial and treponemal reads, respectively, in each of the ancient and modern samples of both tested bacteria, was assessed (Supplementary Table 4). There is no statistical significance between the methods besides in the data of the ancient *M. leprae* samples. Here the RNA bait capture with two rounds shows significantly the highest ratios of specific reads of *M. leprae* to mycobacterial reads in total. For the ancient *Treponema pallidum* samples, the RNA bait capture with two rounds shows the highest proportions of specific reads.
For the samples of modern *M. leprae*, there is no statistical significance but here the in-solution capture with DNA probes shows the highest proportions of *M. leprae* specific reads. In the modern samples, only one round of capture with RNA baits yields the highest specificity.



Supplementary Figure 1: Comparison of (A) mean coverage, (B) standard deviation of the mean coverage, (C) enrichment factor, (D) and the percentage of the genome covered 5 fold, (E) distribution of the fragment length and (F) frequency of the aDNA damage for the ancient and modern strains of *M. leprae*. Three independent replicates were performed for each method. Labels of the ancient samples are in black and for the modern samples in red. Boxplots of the array are blue, of the DNA bait capture red and the RNA baits capture is green and grey for the first and second round, respectively.



Supplementary Figure 2: Comparison of (A) mean coverage, (B) standard deviation of the mean coverage, (C) enrichment factor, (D) and the percentage of the genome covered 5 fold, (E) distribution of the fragment length and (F) frequency of the aDNA damage for the ancient and modern strains of *T. pallidum*. Three independent replicates were performed for each method. Labels of the ancient samples are in black and for the modern samples in red. Boxplots of the array are blue, of the DNA bait capture red and the RNA baits capture is green and grey for the first and second round, respectively.



Supplementary Figure 3: Number of unique reads for the three replicate batches of the three tested methods. The number of unique reads in the second round of hybridization with the RNA baits does not strongly increase compared to the first round.

Supplementary Table 1: List of all samples used in this study group according to organism and age together with the original publications.

Group	Sample	Age	Organism	Host species	publication
modern *T. pallidum* retrieved from nonhuman primates	40M5160407	modern	*T. pallidum*	nonhuman primate	Knauf et al. 2018
	4F5230307	modern	*T. pallidum*	nonhuman primate	Knauf et al. 2018
	RKI1	modern	*T. pallidum*	nonhuman primate	Knauf et al. 2018
	RKI11	modern	*T. pallidum*	nonhuman primate	Knauf et al. 2018
	RKI2	modern	*T. pallidum*	nonhuman primate	Knauf et al. 2018
	RKI9	modern	*T. pallidum*	nonhuman primate	Knauf et al. 2018
modern *T. pallidum*	N12	modern	*T. pallidum*	human	Arora et al. 2016
	N13	modern	*T. pallidum*	human	Arora et al. 2016
	N14	modern	*T. pallidum*	human	Arora et al. 2016
	N17	modern	*T. pallidum*	human	Arora et al. 2016
	C33	modern	*T. pallidum*	human	Arora et al. 2016
	S4	modern	*T. pallidum*	human	Arora et al. 2016
	UW1	modern	*T. pallidum*	human	Arora et al. 2016
Ancient *M. leprae*	3077	ancient	*M. leprae*	human	Schuenemann et al 2013
	GC96	ancient	*M. leprae*	human	Schuenemann et al 2018a
	Body_188	ancient	*M. leprae*	human	Schuenemann et al 2018a
	SK11	ancient	*M. leprae*	human	Schuenemann et al 2018a
	T18	ancient	*M. leprae*	human	Schuenemann et al 2018a
	Refshale_16	ancient	*M. leprae*	human	Schuenemann et al 2013
	SK14	ancient	*M. leprae*	human	Mendum et al. 2014
	SK8	ancient	*M. leprae*	human	Mendum et al. 2014
modern *M. leprae*	S10	modern	*M. leprae*	human	Schuenemann et al 2013
	S11	modern	*M. leprae*	human	Schuenemann et al 2013
	S13	modern	*M. leprae*	human	Schuenemann et al 2013
	S14	modern	*M. leprae*	human	Schuenemann et al 2013
	S2	modern	*M. leprae*	human	Schuenemann et al 2013
	S9	modern	*M. leprae*	human	Schuenemann et al 2013
ancient *T. pallidum*	94-A	ancient	*T. pallidum*	human	Schuenemann et al 2018b
	133	ancient	*T. pallidum*	human	Schuenemann et al 2018b
	94-B	ancient	*T. pallidum*	human	Schuenemann et al 2018b

Supplementary Table 4: Comparison of the specific reads of the three tested protocols.

Sample	age	Organism	Chisq	Pr(>Chisq)	Array	RNA baits1	RNA baits2	DNA baits	Array - RNA baits1	Array - RNA baits2	Array - DNA baits	RNA baits1 - RNA baits2	RNA baits1 - DNA baits	RNA baits2 - DNA baits	p.adjust
40M5160407	modern	T. pallidum	1.26	0.35	0.28	0.33	0.33	0.28	0.53	0.55	1.00	1.00	0.54	0.56	1.00
4F5230307	modern	T. pallidum	0.73	0.57	0.30	0.33	0.32	0.28	0.85	0.97	0.93	0.99	0.54	0.76	1.00
C33	modern	T. pallidum	1.02	0.43	0.33	0.35	0.36	0.30	0.92	0.91	0.82	1.00	0.49	0.47	1.00
94-A	ancient	T. pallidum	1.48	0.29	0.33	0.35	0.38	0.31	0.92	0.50	0.95	0.84	0.66	0.26	1.00
133	ancient	T. pallidum	7.43	0.01	0.28	0.32	0.34	0.29	0.09	0.01	0.95	0.61	0.18	0.03	0.17
94-B	ancient	T. pallidum	2.38	0.15	0.30	0.35	0.37	0.29	0.53	0.24	1.00	0.91	0.44	0.19	1.00
N12	modern	T. pallidum	0.96	0.46	0.32	0.35	0.35	0.30	0.86	0.86	0.91	1.00	0.51	0.52	1.00
N13	modern	T. pallidum	0.87	0.50	0.33	0.35	0.35	0.30	0.91	0.92	0.87	1.00	0.54	0.54	1.00
N14	modern	T. pallidum	0.92	0.47	0.33	0.35	0.35	0.30	0.88	0.88	0.90	1.00	0.52	0.53	1.00
N17	modern	T. pallidum	0.99	0.45	0.33	0.35	0.35	0.30	0.85	0.85	0.91	1.00	0.51	0.50	1.00
RKI1	modern	T. pallidum	0.97	0.45	0.29	0.33	0.33	0.29	0.69	0.73	1.00	1.00	0.56	0.60	1.00
RKI11	modern	T. pallidum	2.69	0.12	0.00	0.00	0.00	0.00	0.13	1.00	0.66	0.17	0.56	0.76	1.00
RKI2	modern	T. pallidum	1.00	0.44	0.32	0.33	0.33	0.29	0.94	0.95	0.78	1.00	0.47	0.49	1.00
RKI9	modern	T. pallidum	3.50	0.07	0.00	0.01	0.00	0.00	0.10	1.00	1.00	0.10	0.14	1.00	1.00
S4	modern	T. pallidum	0.70	0.58	0.32	0.35	0.35	0.31	0.85	0.87	0.98	1.00	0.65	0.66	1.00
UW1	modern	T. pallidum	1.31	0.34	0.32	0.36	0.36	0.30	0.63	0.65	0.99	1.00	0.45	0.47	1.00
3077	ancient	M. leprae	0.79	0.53	0.99	0.93	0.76	0.99	0.98	0.57	1.00	0.78	0.98	0.57	1.00
GC96	ancient	M. leprae	0.37	0.77	0.80	0.69	0.68	0.94	0.98	0.97	0.95	1.00	0.81	0.79	1.00
Body 188	ancient	M. leprae	0.54	0.67	0.95	0.71	0.68	0.96	0.84	0.80	1.00	1.00	0.82	0.77	1.00
SK11	ancient	M. leprae	0.61	0.63	0.98	0.85	0.74	0.97	0.92	0.67	1.00	0.95	0.92	0.68	1.00
T18	ancient	M. leprae	0.60	0.63	0.96	0.86	0.73	0.98	0.97	0.71	1.00	0.92	0.94	0.64	1.00
Refshale_16	ancient	M. leprae	0.86	0.50	1.00	0.99	0.92	0.99	1.00	0.55	1.00	0.63	1.00	0.57	1.00
SK14	ancient	M. leprae	0.75	0.55	0.99	0.97	0.91	0.99	0.99	0.59	1.00	0.78	0.99	0.59	1.00
SK8	ancient	M. leprae	0.64	0.61	0.98	0.85	0.74	0.99	0.93	0.68	1.00	0.94	0.91	0.64	1.00
S10	modern	M. leprae	6.40	0.02	0.99	1.00	1.00	0.99	0.81	0.17	0.29	0.51	0.08	0.01	0.23
S11	modern	M. leprae	22.37	0.00	1.00	1.00	1.00	0.99	0.21	0.08	0.01	0.89	0.00	0.00	0.00
S13	modern	M. leprae	12.10	0.00	1.00	1.00	1.00	0.99	0.75	1.00	0.00	0.71	0.01	0.00	0.03
S14	modern	M. leprae	11.12	0.00	1.00	1.00	1.00	0.99	0.83	0.30	0.02	0.10	0.07	0.00	0.04
S2	modern	M. leprae	13.49	0.00	1.00	1.00	1.00	0.99	0.04	0.08	0.21	0.94	0.00	0.00	0.02
S9	modern	M. leprae	5.62	0.02	0.99	1.00	1.00	0.99	0.15	0.03	0.93	0.64	0.32	0.06	0.32

Supplementary Table 6: Comparison of the variance within each method tested.

Organism	Feature	Sample Group	Chisq	Pr(>Chisq)	Array	DNA baits	RNA baits1	RNA baits2	Array - RNA baits1	Array - RNA baits2	Array - DNA baits	RNA baits1 - RNA baits2	RNA baits1 - DNA baits	RNA baits2 - DNA baits
M. leprae	Mean Coverage	ancient	19.44	2.21E-04	0.97	0.19	0.89	1.04	0.00	0.99	0.99	0.01	0.00	0.91
		modern	30.28	1.21E-03	0.76	0.34	1.42	2.10	0.59	0.18	0.00	0.01	0.00	0.17
	standard deviation	ancient	39.76	1.20E-05	4.22	0.10	4.01	24.20	0.72	1.00	0.00	0.75	0.00	0.00
	Mean Coverage	modern	28.90	2.35E-03	1.15	0.19	0.95	2.52	0.10	0.96	0.01	0.25	0.00	0.00
	Enrichment Factor	ancient	32.09	5.00E-04	36.60	48.60	152.00	277.00	0.99	0.09	0.00	0.16	0.00	0.06
		modern	30.65	1.01E-06	4.44	14.80	28.90	12.70	0.17	0.00	0.36	0.03	0.97	0.01
	5Xpercentage	ancient	27.64	4.33E-06	0.57	0.01	0.03	0.09	0.01	0.40	0.17	0.43	0.00	0.00
		modern	35.26	1.07E-04	0.04	0.02	0.12	0.21	0.93	0.10	0.00	0.02	0.00	0.06
	damage	ancient	29.31	1.92E-03	4.50	0.42	6.75	5.50	0.00	0.23	0.83	0.00	0.00	0.71
	Length	ancient	24.89	1.63E-02	8.78	0.78	5.78	4.56	0.00	0.21	0.04	0.01	0.07	0.85
		modern	10.30	1.62E-02	0.03	0.01	0.03	0.04	0.35	0.95	0.65	0.13	0.03	0.92
T. pallidum	Mean Coverage	ancient	44.24	1.34E-09	0.20	1.44	0.68	21.60	0.97	1.00	0.00	0.99	0.00	0.00
		modern	101.62	2.20E-16	0.49	5.02	13.80	22.80	0.18	0.00	0.00	0.00	0.00	0.00
	standard deviation	ancient	45.97	5.76E-10	1.79	7.80	10.60	108.00	0.97	0.92	0.00	1.00	0.00	0.00
	Mean Coverage	modern	140.84	2.20E-16	2.47	6.22	20.60	58.40	0.85	0.00	0.00	0.01	0.00	0
	Enrichment Factor	ancient	27.10	5.62E-06	158.00	5.346.00	1.286.00	74.836.00	0.99	1.00	0.00	0.99	0.00	0.00
		modern	14.08	2.80E-03	182.00	7.047.00	2.979.00	5.823.00	0.03	0.68	0.11	0.37	0.96	0.67
	5Xpercentage	ancient	23.05	3.95E-05	0.01	0.09	0.02	0.17	0.09	0.98	0.00	0.18	0.12	0.00
		modern	39.93	1.11E-08	0.03	0.06	0.23	0.08	0.81	0.00	0.48	0.00	0.95	0.00
	damage	ancient	36.99	4.62E-08	4.00	0.22	12.50	10.70	0.20	0.00	0.01	0.00	0.00	0.76
	Length	ancient	101.76	2.20E-16	12.30	0.10	18.00	18.30	0.00	0.01	0.00	0.00	0.00	1.00
		modern	21.26	9.31E-05	0.02	0.00	0.05	0.06	0.63	0.07	0.02	0.00	0.00	0.94

Supplementary Table 7: Comparison of the costs per reaction.

in-solution capture with array derived probes						
Product	**Supplier**	**Art.-Nr.**	**amount**	**Price (€)**	**Price (€)/rxn**	**Machines**
2x Hi-RPM Hybridization Buffer	Agilent	5190-0403	25 ml	480	0,65	Vortexer
bait production			200ul	432	16,20	centrifuge
Herculase II Fusion DNA Polymerases	Agilent	600679	400 rxn	456	2,28	centrifuge for plates
MinElute PCR purification Kit	Qiagen	28004	50 rxn	127	2,54	Magnetic rack (96 well plate)
Acetic acid 100%	Sigma	5438080100	100 ml	61	0,00	Magnetic rack (2 ml tubes)
Dynabeads MyOne T1	ThermoFisher	65601	2 ml	546	10,92	Thermocycler
SeraMag Speedbeads	ThermoFisher	6515210505 0250	15 ml	283,73	0,19	Lightcycler
Oligonucleotide (primers, blocking oligos)	sigma		30 rxn	533,5	17,78	Pipettes
GeneAmp 10x PCR buffer	ThermoFisher	N8080006	1,5 ml	195	3,90	
Cot-1 DNA	ThermoFisher	15279011	500 ug	310	1,55	
Denhardt's Solution 50x	Sigma	D2532-5ML	5 ml	172	0,21	
Additional reagents				**sum/rxn**	**56,23**	
SDS 20 %	Sigma	05030-500ML-F	500 ml	107		
SSC 20%	ThermoFisher	AM9763	1 L	101		
0.5 M EDTA	Sigma	324506-100ML	100 ml	45		
HPLC H2O	Sigma	270733-1L	1 L	24		
5 M NaCl	Sigma	S5150-1L	1 L	68,3		
1 M Tris-HCl, pH 8	ThermoFisher	AM9855G	100 ml	57,25		
Tween-20 100%	Sigma	P1379-25ML	25 ml	13		
1 M NaOH	Applichem	A6579,1000	1 L	27,2		
3 M Sodium acetate pH 5.2	Sigma	S7899-100ML	100 ml	32,4		
PM buffer	Qiagen	19083	500 ml	63,6		
Salmon SpermDNA	ThermoFisher	15632011	5 ml	133		
EtOH	Merck	1009832511	2,5 L	86		

Array						
Product	**Supplier**	**Art.-Nr.**	**amount**	**Price (€)**	**Price (€)/rxn**	**Machines**
Oligo aCGH/ChIP-on-chip Hybridization Kit	Agilent	5188-5220	Blocking Agent + 2x Hi-RPM hybridization buffer 25 rxn	461	18,44	Vortexer
blocking oligos	sigma	BO4, BO6, BO8, BO10		304,64	10,15	Thermoshaker
BD Plastipak	Fisher Scientific	10205194	100 x	61,38	0,61	Thermocycler
Hybridization Gasket	Agilent	G2534-60005	100	203	20,36	Hybridization Oven
SureSelect DNA Capture Array 1 M	Agilent	G3358A	1	613	613,00	1 bigger glass bowl
Microarray Wash Buffer Kit	Agilent	5188-5222	WB 1, WB 2 40 rxn	208	5,20	2 smaller glass bowls
Herculase II Fusion DNA Polymerases	Agilent	600679	400 rxn	456	2,28	Array chamber
MinElute PCR purification Kit	Qiagen	28004	50 rxn	127	2,54	centrifuge
Addtional ragents				**sum /rxn**	**672,59**	Magnetic mixer
HPLC H2O	Sigma	270733-1L	1 L	24		Waterbath
						tweezers
						rack for slides
						pipettes
						Lightcycler

MyBait						
Product	**Supplier**	**Art.-Nr.**	**amount**	**Price (€)**	**Price (€)/rxn**	**Machines**
MyBait Kit			48 rxn	5000	104,17	Vortexer
Herculase II Fusion DNA Polymerases	Agilent	600679	400 rxn	456	2,28	centrifuge
MinElute PCR purification Kit	Qiagen	28004	50 rxn	127	2,54	centrifuge for plates
Additional reagents				**sum /rxn**	**108,99**	Magnetic rack (96 well plate)
1 M Tris-HCl, pH 8	Thermo Fisher	AM9855G	100 ml	57,25		Magnetic rack (2 ml tubes)
Tween-20 100%	Sigma	P1379-25ML	25 ml	13		Thermocycler
HPLC H2O	Sigma	270733-1L	1 L	24		Lightcycler
						Pipettes

References

Arora, Natasha; Schuenemann, Verena J.; Jäger, Günter; Peltzer, Alexander; Seitz, Alexander; Herbig, Alexander et al.: Origin of modern syphilis and emergence of a pandemic Treponema pallidum cluster. In: *Nat Microbiol* 2 (1), S. 1–6. DOI: 10.1038/nmicrobiol.2016.245.

Bates, Douglas; Mächler, Martin; Bolker, Ben; Walker, Steve (2015): Fitting Linear Mixed-Effects Models Using lme4. In: *J. Stat. Soft.* 67 (1). DOI: 10.18637/jss.v067.i01.

Briggs, Adrian W.; Stenzel, Udo; Johnson, Philip L. F.; Green, Richard E.; Kelso, Janet; Prüfer, Kay et al. (2007): Patterns of damage in genomic DNA sequences from a Neandertal. In: *Proceedings of the National Academy of Sciences of the United States of America* 104 (37), S. 14616–14621. DOI: 10.1073/pnas.0704665104.

Burbano, Hernán A.; Hodges, Emily; Green, Richard E.; Briggs, Adrian W.; Krause, Johannes; Meyer, Matthias et al. (2010): Targeted investigation of the Neandertal genome by array-based sequence capture. In: *Science (New York, N.Y.)* 328 (5979), S. 723–725. DOI: 10.1126/science.1188046.

Fu, Qiaomei; Meyer, Matthias; Gao, Xing; Stenzel, Udo; Burbano, Hernán A.; Kelso, Janet; Pääbo, Svante (2013): DNA analysis of an early modern human from Tianyuan Cave, China. In: *Proceedings of the National Academy of Sciences of the United States of America* 110 (6), S. 2223–2227. DOI: 10.1073/pnas.1221359110.

Hodges, Emily; Xuan, Zhenyu; Balija, Vivekanand; Kramer, Melissa; Molla, Michael N.; Smith, Steven W. et al. (2007): Genome-wide in situ exon capture for selective resequencing. In: *Nat Genet* 39 (12), S. 1522–1527. DOI: 10.1038/ng.2007.42.

Mendum, Tom A.; Schuenemann, Verena J.; Roffey, Simon; Taylor, G. Michael; Wu, Huihai; Singh, Pushpendra et al. (2014): Mycobacterium leprae genomes from a British medieval leprosy hospital: towards understanding an ancient epidemic. In: *BMC Genomics* 15 (1), S. 1–8. DOI: 10.1186/1471-2164-15-270.

Mubemba, Benjamin; Gogarten, Jan F.; Schuenemann, Verena J.; Düx, Ariane; Lang, Alexander; Nowak, Kathrin et al. (2019): Geographically structured genomic diversity of non-human primate-infecting Treponema pallidum subsp. pertenue (32).

Peltzer, Alexander; Jäger, Günter; Herbig, Alexander; Seitz, Alexander; Kniep, Christian; Krause, Johannes; Nieselt, Kay (2016): EAGER: efficient ancient genome reconstruction. In: *Genome Biol* 17 (1), S. 1–14. DOI: 10.1186/s13059-016-0918-z.

Fellows Yates, James A.; Lamnidis, Thiseas C.; Borry, Maxime; Valtueña, Aida Andrades; Fagernäs, Zandra; Clayton, Stephen et al. (2020): Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager.

Sawyer, Susanna; Krause, Johannes; Guschanski, Katerina; Savolainen, Vincent; Pääbo, Svante (2012): Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. In: *PloS one* 7 (3), e34131. DOI: 10.1371/journal.pone.0034131.

Schuenemann, Verena J.; Avanzi, Charlotte; Krause-Kyora, Ben; Seitz, Alexander; Herbig, Alexander; Inskip, Sarah et al. (2018): Ancient genomes reveal a high diversity of Mycobacterium leprae in medieval Europe. In: *PLoS pathogens* 14 (5), e1006997. DOI: 10.1371/journal.ppat.1006997.

Schuenemann, Verena J.; Kumar Lankapalli, Aditya; Barquera, Rodrigo; Nelson, Elizabeth A.; Iraíz Hernández, Diana; Acuña Alonzo, Víctor et al. (2018): Historic Treponema pallidum genomes from Colonial Mexico retrieved from archaeological remains. In: *PLoS neglected tropical diseases* 12 (6), e0006447. DOI: 10.1371/journal.pntd.0006447.

Schuenemann, Verena J.; Singh, Pushpendra; Mendum, Thomas A.; Krause-Kyora, Ben; Jäger, Günter; Bos, Kirsten I. et al. (2013): Genome-wide comparison of medieval and modern Mycobacterium leprae. In: *Science (New York, N.Y.)* 341 (6142), S. 179–183. DOI: 10.1126/science.1238286.

Vågene, Åshild J.; Herbig, Alexander; Campana, Michael G.; Robles García, Nelly M.; Warinner, Christina; Sabin, Susanna et al. (2018): Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. In: *Nature ecology & evolution* 2 (3), S. 520–528. DOI: 10.1038/s41559-017-0446-6.

Paper III

Furtwängler A, Rohrlach AB, Lamnidis TC, Papac L, Neumann GU, Siebke I, Reiter E, Steuri N, Hald J, Denaire A, Schnitzler B, Wahl J, Ramstein M, Schuenemann VJ, Stockhammer PW, Hafner A, Lösch S, Haak W, Schiffels S, Krause J ″ Ancient genomes reveal social and genetic structure of Late Neolithic Switzerland." *Nature Communications* **11,** 1915 (2020)

Ancient genomes reveal social and genetic structure of Late Neolithic Switzerland

Authors: Furtwängler A[1], Rohrlach AB[2,3], Lamnidis TC[2], Papac L[2], Neumann GU[1,2], Siebke I[4], Reiter E[1], Steuri N[5], Hald J[6], Denaire A[7], Schnitzler B[8], Wahl J[9,10], Ramstein M[11], Schuenemann VJ[1,12,13], Stockhammer PW[2,14], Hafner A[5,15], Lösch S[4], Haak W[2], Schiffels S[2], Krause J[1,2,11*]

Affiliations:

[1]Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Germany
[2]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany
[3]ARC Centre of Excellence for Mathematical and Statistical Frontiers, School of Mathematical Sciences, The University of Adelaide, Adelaide, South Australia 5005, Australia
[4]Department of Physical Anthropology, Institute of Forensic Medicine, University of Bern, Switzerland
[5]Institute of Archaeological Sciences, Prehistoric Archaeology, University of Bern, Switzerland
[6]Archaeological Office of the District of Constance, Germany
[7]Department of history of arts and Archaeology, University of Burgundy, France
[8]Museum of Archaeology Strasbourg, France
[9]Institute for Archaeological Science, Palaeoanthropology, Eberhard Karls University Tübingen, Germany
[10]State Office for Cultural Heritage Management Baden-Wuerttemberg, Konstanz, Germany
[11]Archaeological Service of the canton of Bern
[12]Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Germany
[13]Institute of Evolutionary Medicine, University of Zurich, Switzerland
[14]Institut für Vor- und Frühgeschichtliche Archäologie und Provinzialrömische Archäologie, Ludwig-Maximilians University Munich, Germany
[15]Oeschger Centre for Climate Change Research, University of Bern, Switzerland

* Corresponding author: krause@shh.mpg.de

Abstract

Genetic studies of Neolithic and Bronze Age skeletons from Europe have provided evidence for strong population genetic changes at the beginning and the end of the Neolithic period. To further understand the implications of these in Southern Central Europe, we analyze 96 ancient genomes from Switzerland, Southern Germany, and the Alsace region in France,

covering the Middle/Late Neolithic to Early Bronze Age. Similar to previously described genetic changes in other parts of Europe from the early 3rd millennium BCE, we detect an arrival of ancestry related to Late Neolithic pastoralists from the Pontic-Caspian steppe in Switzerland as early as 2860–2460 calBCE. Our analyses suggest that this genetic turnover was a complex process lasting almost 1000 years and involved highly genetically structured populations in this region.

Introduction

Genetic studies have revealed that Central Europeans, during the Neolithic, were genetically mixed between indigenous European hunter-gatherers and new incoming people with ancestry related to Western Anatolian early farmers[1,2,3,4,5]. Towards the end of the Neolithic period, just before the transition to the Bronze Age, a second arrival of a new ancestry component in Europe was detected genetically[6,7], coinciding with the emergence of the Central and Eastern European Corded Ware Complex (CWC; encompassing Battle Axe and Single Grave cultural groups, ref. [8]). The new genetic component was most closely related to ancestry from the Pontic–Caspian steppe, found in individuals associated with the Yamnaya complex. While the origin of this new, third European ancestry component has been attested in many European regions[6,7,9], the exact timing of the arrival in other regions, as well as the demographic processes underlying this genetic admixture, are less clear.

Archaeologically, the Neolithic period in Switzerland is dominated by lakeshore and bog settlement sites with organic preservation, inner alpine sites of the Rhône valley, and high mountain pass sites[10]. Apart from settlement remains, stone cist graves from the Chamblandes type and a few megalithic burials towards the end of the Neolithic have been found, such as the dolmen burials of Oberbipp, Sion, Aesch, and others[10,11,12,13]. The rich archaeological record in Switzerland makes the region relevant for studies of population history in Central Europe. This is due to the particularly well-preserved wetland settlements from which the wooden parts provided one of the best dated dendrochronological records in prehistoric Europe[11].

In Switzerland, CWC finds are exclusively from settlements on the banks of the large pre-alpine lakes. They are particularly numerous in the region of Lake Zurich in Eastern Switzerland and the Three Lakes Region in Western Switzerland. The sites on Lake Neuchâtel lie on the South-Western edge of the area influenced by CWC. In Eastern Switzerland, the new, cord-ornamented ceramic style was rapidly adopted, while in Western Switzerland this process had lasted several centuries. High-precision dendrochronological data obtained from the building structures of CWC settlements provide clear approaches to absolute chronology[14]. Although there are numerous Neolithic and Early Bronze Age sites from the lakeshores and moors, there are no burials directly related to them. This is due, among other things, to the fact that the Chamblandes type stone cist tombs were already in use in the fifth millennium BCE. This burial custom, however, most probably ends around 3800 BCE, i.e., exactly at the time when the lakeshore settlements begin to become numerous. The burials of the Early Bronze Age are concentrated in inner alpine regions (Rhone valley, Lake Thun area, and foothills of the Alps) from which no lakeside settlements of this period are known. In periods with many

graves, there are no settlements and vice versa, in periods with many settlements the corresponding burials are missing. The reasons for this are probably of taphonomic nature.
Thus, only four ancient genomes have been published so far for the territory of present-day Switzerland: one Late Pleistocene hunter-gatherer individual from Bichon cave[15] and three individuals associated with the Bell Beaker phenomenon from the dolmen burial of Sion-Petit-Chasseur[9].
The aim of this project is to investigate the transition from the Neolithic to the Bronze Age in Switzerland in detail, with a specific focus on the timing of the arrival, source and mixture process of steppe-related ancestry, and the social and demographic structure before and after this transition. Using relatively dense temporal sampling, we generate genome-wide data from 96 individuals dating to the Neolithic and Early Bronze Age period from Switzerland, Southern Germany, and the Alsace region in France. We also generate data from one individual from the Early Iron Age and the Roman period, respectively. We find the expected large genetic turnover at the beginning of the third millennium BC and a highly genetically structured population in the region of present day Switzerland at that time period. The predominant social structure, furthermore, was probably patrilocal.

Results

Ancient DNA authentication and uniparental markers

The ancient individuals from this study originate from 13 Neolithic and Early Bronze Age sites in Switzerland (Fig. 1b), Southern Germany, and the Alsace region in France. All samples taken from the individuals were radiocarbon dated (Supplementary Note 2 and Supplementary Data 1). In a preliminary screening, 263 samples were enriched for mtDNA. We reconstructed complete mitochondrial genomes, used them to estimate DNA library contamination (Supplementary Data 1), and identified 96 samples that had less than 5% of contamination for further analyses. We determined mtDNA haplogroups using the software *haplogrep* (ref. [16], Supplementary Note 3, Supplementary Fig. 1, and Supplementary Data 1) and found the macrohaplogroups N1a, W, X, H, T2, J, U2, U3, U4, U5a, U5b, K, and U8 in our samples.
For genome-wide analysis, we genotyped all selected individuals on ~1.2 million genomic SNPs[2], also containing 49,704 SNPs on the X chromosome and 32,670 SNPs on the Y chromosome. SNPs on the X chromosome were used to estimate nuclear contamination in male individuals (Supplementary Data 1), and we again used a threshold of 5% to select clean libraries for further analysis (96 individuals). We also determined Y chromosomal haplogroups (Supplementary Note 4, Supplementary Fig. 2, Supplementary Data 1, and Supplementary Table 1).

Population turnover at the transition to the Bronze Age

We combined the genotype data of the new 96 individuals from this study that passed our contamination tests with 399 published ancient genomes from the same time period from

Central and Western Europe as well as Neolithic individuals from Anatolia and the Pontic steppe (individuals annotated as Yamnaya Samara in ref. [6]) and genotype data of modern individuals from the POPRES[17] and the Human origins (HO)[1] datasets for various analyses.

We projected our 96 new ancient genomes from Switzerland and surrounding regions and 52 published ancient genomes selected to reflect the genetic landscape of Europe at different time points onto the first two principal components constructed from 1960 individuals of 38 European populations from the POPRES dataset (Fig. 1c and "Methods"). Two distinct clusters can be identified and were also confirmed by ADMIXTURE analysis (Supplementary Note 5), one consisting of individuals dating to 4770–2500 calBCE, and one comprising individuals dating to 2900–1750 calBCE. The oldest individuals from the sites of Niederried (CH) and Lingolsheim (F) fall close to ancient individuals from Anatolia associated with early agricultural contexts. More recent individuals from the megalithic burials at Oberbipp (CH) and Aesch (CH) are shifted further towards Western Hunter-Gatherers (WHG) and close to modern-day Sardinian individuals, as well as towards Early and Middle Neolithic individuals from Iberia or the Middle Elbe-Saale (MES) region in Central Germany. This shift mirrors an increase of hunter-gatherer-related ancestry during the middle Neolithic that has been described previously for other parts of Europe.

The second distinct cluster is shifted towards the individuals associated with the "Yamnaya" complex, similar to other European groups younger than 2700 BCE, relative to individuals older than 2700 BCE. In this cluster, the oldest individuals are closest to Late Neolithic groups on the steppe, whereas more recent individuals are once again shifted towards the Middle/Late Neolithic cluster. All Final Neolithic and Early Bronze Age individuals fall within the range of modern-day Europeans, but none of the newly sequenced individuals of this study overlap with the present-day Swiss populations in this analysis, suggesting additional population changes in the region after the Middle Bronze Age.

Our individuals sequenced in this study fall in PC space between WHG individuals, Western Anatolian Neolithic Farmers (ANF) and steppe pastoralists from Samara (YAM), similar to other Late Neolithic individuals such as the Tyrolian Iceman[18] and Bronze Age populations such as individuals of the Bell Beaker complex[6] from Europe. Therefore, we modeled them as a three-way mixture between these three populations using *qpAdm* from the ADMIXTOOLS package (ref. [19], "Methods" and Supplementary Tables 2 and Supplementary Data 4). The overall pattern observed from this analysis matched previous analyses of that type[9]. Individuals from older sites (Early Neolithic and Middle Neolithic) are consistent with a two-way mixture between WHG and ANF ancestry, whereas individuals from younger sites after ~2700 BCE exhibit substantial amounts of ancestry related to YAM. Furthermore, the proportions of this component differ strongly between sites and tend to decrease over time (Fig. 2a). This trend is confirmed by further analysis of the ancestry components on an individual level.

Compared with previous studies[2,9] analyzing Neolithic and Bronze Age individuals from present-day Germany and Great Britain, which do not report individuals dating to the transition period directly, in this study we analyze a gapless time-transect covering the

Neolithic to Bronze Age transition. By viewing the YAM-related ancestry component estimated with *qpAdm* over time at an individual level, it becomes apparent that this ancestry was virtually absent before 2700 BCE, followed by a steep increase in parts of the population starting around 2700 BCE (Supplementary Note 7 and Supplementary Data 4). After this rapid increase in individual proportions of YAM-related ancestry from 0% to ~60%, a decrease down to 25–35% can be observed over the next thousand years. We also note four female individuals that can be modeled without any YAM-related ancestry even 1000 years after the appearance of that genetic component in the area. Comparing outgroup-*f3* statistics between the autosomes and the X chromosome of Final Neolithic and Bronze Age individuals we find that autosomes are more closely related to YAM-related ancestry than the X chromosomes are (Supplementary Note 6), consistent with a model in which more males than females brought YAM-related ancestry into the region as already shown by previous studies[20].

We analyzed pairwise genetic differences across all analyzed genomic positions between individuals before and after the genetic turnover and found that the mismatch rates increase, on average, by around 0.009 after 2700 BCE for all populations (Fig. 2c, see also "Methods"). This is more than twice the increase that can be attributed to the between-population variability in rates and indicates a significant increase in genetic diversity after the arrival of the YAM-related ancestry component in Central Europe. Modern populations would be expected to have higher levels on average but are not compared in this analysis since the modern individuals from published datasets usually do not originate from groups with the same background (e.g., being related distantly) as could be expected in multiple burials.

Comparing our newly analyzed individuals from Switzerland with ancient genomes from Great Britain, Iberia, and Germany[2,9,21] we modeled the arrival time of the YAM-related ancestry in the different broadly defined European regions (Fig. 2b, see "Methods"). While our models indicate that the proportions of the YAM-related ancestry peaks earlier in the Swiss dataset (around 2750 BCE) compared with the comparative datasets from refs. [2,9,21] (around 2600 BCE), these differences fall within the uncertainty of the analysis (Supplementary Fig. 6), so may be considered suggestive of an earlier arrival of steppe-related ancestry, but not conclusive. We also caution that differences are likely affected by uneven sampling through time in the three different datasets, and so expect the precision of this analysis to improve with denser temporal sampling in the future.

Timing and duration of the genetic turnover

We used the software *DATES* (ref. [22], https://github.com/priyamoorjani/DATES) to estimate the admixture time between YAM-related and Late Neolithic ancestries in all Final Neolithic and Early Bronze Age individuals from Switzerland with substantial admixture proportion. Our estimates range between 3 and 60 generations ago, with substantial uncertainty. If the mixture occurred as one single event in the history of all individuals, we would expect more recent individuals to have a higher admixture time estimate (i.e., more generations ago) than individuals of older dating. However, we observe only a slight trend towards more generations in individuals with younger C^{14} dates (Fig. 3a), which suggests that the process of

admixture with steppe-related ancestry occurred over several hundred years rather than as a single pulse.

For comparison, we performed a similar analysis for published ancient Final Neolithic and Bell Beaker populations from the MES region in Germany, Great Britain, and Iberia using a mixture between the YAM-related steppe component and the corresponding Middle Neolithic population of the region (Fig. 3c). For the MES region from Germany, Late Neolithic and Early Bronze Age individuals were split in Bell Beaker and Corded Ware groups. Similar to this approach, we also split the Neolithic individuals from Switzerland into a group of individuals associated with the Corded Ware Complex from Spreitenbach and a group of younger individuals from the Bronze Age. All four regions (Iberia, Great Britain, MES and Switzerland) show similar ranges of admixture dates between the steppe-related component and the Neolithic component starting between 3000 and 2500 BCE.

The inferred admixture time describes the time when people of steppe-related ancestry encountered people with Late Neolithic ancestry but does not reveal the location. Admixture could have happened in Switzerland or elsewhere, with already admixed individuals moving to Switzerland.

Neolithic source population for the admixing event

We tested which Neolithic population likely admixed with the incoming people that carried large amounts of YAM-related ancestry, by adding Late Neolithic individuals from Switzerland, Globular Amphora, Iberia Middle Neolithic, and France Neolithic groups separately as additional right populations to the three-way model (WHG, Western Anatolia Neolithic and Yamnaya Samara) used in the above *qpAdm* analyses (ref. [19], Supplementary Data 3 and 4). The model remains fitting for Iberia Middle Neolithic and France Neolithic populations ($p \geq 0.01$ and $p \geq 0.3$, respectively) but it fails when we add Late Neolithic individuals from Switzerland or Globular Amphora as additional right population ($p \leq 2e{-}7$ and $p \leq 2e{-}6$, respectively). This suggests that both the local Swiss Late Neolithic population as well as people associated with the Late Neolithic Globular Amphora culture, located further east, are better proxies of the genetic sources for Final Neolithic and Bronze Age populations from Switzerland than Western ANF and steppe pastoralists.

Kinship before and after the genetic transition

In five burial sites, we identified first-degree relatives using the software *lcmlkin*[23] and *READ*[24], and by calculating pairwise mismatch rates across all analyzed genomic sites between individuals (see "Methods"). Four of these sites contained more than two closely related individuals, which allowed us to reconstruct family trees spanning three generations for Oberbipp, Aesch and Singen (Fig. 4). In these multiple burials, only a few female individuals (four individuals) were buried together with one of their parents or their sons, compared with a higher number (21 individuals) of males buried with their father, brothers or sons, indicating that males likely tended to stay where they were born, while females were likely mobile. This pattern is observed both before and after the arrival of the YAM-related ancestry and is

indicative of patrilocal societies during Late Neolithic times in the studied region, consistent with previous results from Neolithic times throughout Northern and Western Europe[25,26].

Population movement in Switzerland after the Bronze Age

We compared present-day Swiss people to regional Final Neolithic populations (Spreitenbach, Bad Zurzach, Wartau) to test whether there are additional ancestry components in present-day Swiss people. For that analysis, we made use of information available in the present-day dataset (POPRES) about the self-reported language group and split the individuals into three linguistic regions, as they were shown to be distinguishable genetically in previous studies[27]: German-speaking, French-speaking, and Italian-speaking Switzerland. To test for continuity between the ancient and present-day population we used the method *qpWave* from the ADMIXTOOLS package[19] and found that a simple continuity can be rejected ($p = 0.0003$) for all three linguistic regions separately and the entire present-day Swiss population combined, consistent also with the PCA (Fig. 1c).

To assess whether ancient Swiss individuals from the Final Neolithic are symmetrically related to different linguistic present-day groups, or share an excess of alleles with any of them, we calculated D-statistics of the form D(Mbuti, Test, Swiss–French, Swiss–Italian), D(Mbuti, Test, Swiss–German, Swiss–Italian), and D(Mbuti, Test, Swiss–French, Swiss–German) where "Test" are the different Neolithic groups. The first two D-statistics were, with few exceptions, all negative with a $|Z| \geq 1.099$ (maximum values to be found in D(Mbuti, Test, Swiss–French, Swiss–Italian) for Singen with −3.431 and Bad Zurzach −3.068) indicating the least genetic affinity of the Final Neolithic and Early Bronze Age individuals of this study to the Italian-speaking group in our present-day Swiss dataset. For D(Mbuti, Test, Swiss–French, Swiss–German) some variation between the sites can be found (Fig. 5) with the older sites sharing more alleles with the French-speaking group, and the younger sites being more similar with the German-speaking group.

Analysis of functional SNPs

We analyzed the frequencies of several phenotypic SNPs (Table 1, "Methods"). Derived alleles for SLC24A5 associated with light skin pigmentation in Europeans were found in all individuals with this position covered. The frequency of SLC45A2 also causing lighter skin pigmentation tends to increase and the frequency of HERC2 associated with light eye-color tends to decrease towards the Final Neolithic. A mutation associated with lactose tolerance in adulthood (LCT; rs4988235), which is of high frequency in Europe today, is absent in Late and Middle Neolithic samples. The only exception is one Final Neolithic individual from Spreitenbach dating to 2105–2036 calBCE, which is one of the earliest European individuals with this mutation found so far. The near absence of lactose tolerance in these ancient groups is in concordance with previous studies hypothesizing that this mutation arose in the Final Neolithic period and started to increase in frequency after the beginning of the Bronze Age[2].

Discussion

Our study is the first to report a substantial number of ancient genomes from Switzerland, following a trend of population-scale archaeogenetic sequencing studies in Europe[9,21,28], made possible by capture technology. In accordance with previous studies[1,6,7], the Middle and Late Neolithic Swiss individuals are descendants of late European hunter-gatherers and early farmers, whilst the individuals after 2700 BCE also carry steppe-related ancestry[6,7,9]. Genetic similarities between Corded Ware associated individuals from the MES region in Germany and individuals from Spreitenbach, also associated with the Corded Ware Complex, suggest that this complex was associated with a relatively homogenous genetic population throughout/across large parts of Central Europe.

The social and family structures, as reconstructed by biological kinship networks, remain the same before and after the arrival of steppe-related ancestry in the region. The predominant social structure in populations buried at the sites investigated in this study must have been a patrilocal society where males stayed where they were born, and females came from more distant living families, a societal dynamic which has been confirmed by stable isotopes[29] and that has been previously documented for the Middle Neolithic[25]. Also, higher female mobility has been shown during the Early Bronze Age[26,30]. Our study also presents one of the earliest evidence for adult lactose tolerance in Europe, dating to 2105–2036 calBCE.

Unsurprisingly, comparing our ancient individuals from Switzerland with the data of individuals from present-day Switzerland reveals additional changes in the region since the Bronze Age. In the periods following the studied time span, different factors could have influenced the population. In particular, in the so-called migration period from 375 to 538 AD, following the Roman Empire, in which there was widespread migration of peoples within or into Europe[31].

Remarkably, we identified several female individuals without any detectable steppe-related ancestry up to 1000 years after this ancestry arrives in the region, with the most recent woman without such ancestry dating to 2213–2031 calBCE. This suggests a high level of genetic structure in this region at the beginning of the Bronze Age with potential parallel societies living in close proximity to each other. Published stable isotope results for one of these females (MX193 or Individual 3 in the original publication) indicate that she was not of local origin[32]. It can, therefore, be speculated whether admixture between the newly established local population with steppe-related ancestry and mobile females with less or none of it, caused the decline in the relative amount of this ancestry component in the centuries after its arrival in present-day Switzerland. As the parents of those mobile females also could not have carried steppe-related ancestry, it remains to be shown where in Central Europe such populations without this component were present. One possibility could be Alpine valleys, which until today are inhabited by linguistic isolates that exhibit strong genetic differentiation as initial studies on uniparentally inherited markers have suggested[33,34]. But considering the results of Mittnik et al.[35] with similar patterns in the Lech valley, the origin of this steppe-related ancestry component lacking population does not necessarily lie that far south.

Stable isotope analyses[29,32] do not give clear indications if all four females originated and spend their entire life within the region of modern-day Switzerland. Therefore, it cannot be excluded, that these females also originate from regions further south since also some regions e.g., Italy are not genetically described so far for this particular time span. However, individuals without any steppe-related ancestry can be found up until 2479–1945 BCE for example in Iberia or until 2900–1700 BCE in the Minoan population of Crete[21,36] and even later on Sardinia where steppe-related ancestry arrives around 300 CE[37] and where studies of present-day Sardinians found indications of continuity in mountainous regions since Neolithic times[38].
Also noteworthy is the remarkably early arrival of the steppe-related ancestry component in Switzerland, at least as early as or even earlier than in regions of Germany and Great Britain. However, further investigations are needed, especially since datasets from the regions in Great Britain and the MES region in Germany show gaps in the sampling between the Late/Final Neolithic and the Early Bronze Age potentially biasing the results, to draw any conclusions about the exact route in which the steppe-related ancestry spread through Central Europe.

Methods

DNA extraction and library preparation

A total of 263 samples were screened for DNA preservation for this study. A detailed description of the archaeological context of the samples and radiocarbon dates for all individuals included in genome-wide analysis can be found in Supplementary Notes 1 and 2. All pre-PCR steps were performed in the cleanroom facilities at the Institute for Archaeological Sciences in Tübingen. For the reduction of surface contamination, the samples were treated with UV light for 30 min each side. Between 30 and 50 mg of powder from coronal dentin for teeth and between 50 and 100 mg of bone powder for the petrous bones were used for extraction. After the dissolving of the powder in 1 ml extraction buffer (0.45 M EDTA, 0.25 mg/ml proteinase K) at 37 °C overnight, than the supernatant was transferred into 10 ml binding buffer (5 M GuHCl, 40 % Isopropanol, 115 mM NaAc) and the DNA was bound to a silica membrane in the MinElute Columns from Qiagen. Then the membrane was washed two times with 720 µl of the commercial PE buffer form Qiagen, and then the DNA was eluted in 100 µl TET buffer (1 mM EDTA, 10 mM Tris-HCl, 0.05 % Tween-20, ref. [38]). Sequencing libraries were prepared after[39]. For the blunt ending 20 µl extract were combined with 30 µl reaction mix (1× NEB buffer 2, 100 µM dNTP mix, 0.8 mg/ml BSA, 1 mM ATP, 0.4 U/µl T4 Polynucleotide Kinase (BioLabs, Frankfurt), 0.024 U/µl T4 Polymerase also from BioLabs), and then incubated in a thermo cycler for 15 min at 15 °C and then 15 min at 25 °C. The resulting 18 µl was then used for the ligation of the adapters. With a final volume of 40 µl the reaction contained: 1× Quick Ligase buffer, 250 nM Solexa Adapter Mix and 0.125 U/l Quick Ligase (BioLabs, Frankfurt). The incubation was at room temperature for 20 min. This step was followed by a MinElute purification with an elution volume of 20 µl. The 20 µl from the step before were combined with 20 µl of the reaction mix (1× Isothermal buffer, 125 nM dNTP,

0.4 U/µl Bst polymerase 2.0 from BioLabs), and then incubated for 20 min at 37 °C and then 20 min at 80 °C.

To enable multiplexed sequencing, sample-specific barcodes were added to each library by amplification with tailed primers[40]. The PCR reactions had a final volume of 100 µl and the following concentrations: 1× buffer, 0.25 mM dNTP mix, 0.3 mg/ml BSA, 0.2 mM P7, 0.2 mM P5 and 0.025 U/µl Pfu Turbo Polymerase (Agilent Technologies, Santa Clara, USA). The thermal profile started with 2 min at 95 °C This was then followed by ten cycles with 30 s at 95 °C, 30 s at 58 °C, and 1 min at 72 °C followed by 10 min at 72 °C. After MinElute purification the DNA was eluted in 50 µl TET.

To achieve a high copy number of each library, an additional amplification was performed with Herculase II in a final volume of 100 µl consisting of 1× Herculase II reaction buffer, 0.25 mMdNTPs, 0.4 µM IS 5, 0.4 µM IS 6, 0.01 % Herculase II Fusion DNA Polymerase. The thermal profile started with 2 min at 95 °C followed by a sample-specific number of cycles of 10 s at 95 °C, 30 s at 65 °C, and 30 s at 72 °C. This was then followed by 4 min at 72 °C. All samples were pooled equimolar at 10 nM for shotgun sequencing.

Mitochondrial and nuclear capture

In addition to shotgun sequencing, all libraries were enriched for mitochondrial DNA using baits generated from modern DNA via long-range PCR[41]. For extracts with sufficient endogenous DNA after screening two more double-indexed UDG-half treated libraries were prepared similar to the approach described above with the differences that the first master mix consisted of 60 µl with 1× Buffer Tango (Thermo Fisher Scientific), 100 µM dNTPs, 1 mM ATP, and 0.06 U/µl USER enzyme (NEB). After a 30 min incubation at 37 °C, 0.12 U/µl UGI was added and the reaction was incubated for another 30 min at 37 °C. This was followed by the procedure described above[42]. The non-UDG treated library and the two UDG-half libraries were enriched for 1.2 Mio nuclear SNPs using an in-solution hybridization protocol[2,43].

Bioinformatic processing

After sequencing all data were processed using the software package EAGER[44]. Adapters were removed and paired-end data were merged using Clip&Merge[44]. Mapping was performed with BWA with the mismatch parameter set to 0.01 and a seed length of 1000 against the human genome reference GRCh37/hg19. If necessary, PCR duplicates were removed using DeDup[44]. BAM files from different libraries of the same extract were merged after mapping and quality control. Reads mapping to the mitochondrial genome were extracted from the BAM files and the mitochondrial genome was reconstructed and the amount of mitochondrial contamination was estimated using the software *schmutzi*[45]. The web-based tool HaploGrep[46] was used to determine mitochondrial haplogroups. Genetic sex was determined by comparing X chromosomal reads to Y chromosomal reads[47].

For male samples, nuclear contamination was estimated using ANGSD[48]. Samples with mtDNA or nucDNA contamination >5% as well as samples with <10,000 SNPs of the HO dataset were excluded from further analysis.

Population genetic analysis

Pseudo-haploid genotypes for the 1.2 M SNPs were retrieved using *pileupcaller* (https://github.com/stschiff/sequenceTools). A reference dataset of 399 published ancient genomes, the HO dataset or the POPRES dataset were compiled. The overlap between the 1.2 M targeted SNPs and the HO data are 593,054 SNPs, between the POPRES and the target SNPs 133,682.

For the principal component analysis on the newly sequenced individuals, published ancient individuals and the European population from the POPRES *smartpca* from the EIGENSOFT package (version: 16000) was used with the parameters *lsqproject: YES* and *shrinkmode: YES*.

Relative proportions of ancestry components in the newly sequenced individuals, published ancient individuals from Germany, Great Britain, and Iberia were estimated using *qpAdm* (version: 632) from ADMIXTOOLS (ref. [19], https://github.com/DReichLab) using a threshold of 100k SNPs for analysis on an individual level (Supplementary Note 3) and modern reference individuals (Mbuti, Papuan, Onge, Han, and Karitiana) from the HO dataset and published ancient individuals (Ust Ishim, Ethiopia 4500BP, Villabruna, MA1).

Estimation of steppe arrival times

The arrival time of steppe ancestry was modeled from the newly sequenced individuals and published ancient genomes from Germany, Great Britain, and Iberia assuming that proportions of steppe ancestry increase from zero to p_e at some time t_e, and then decrease according to an exponential curve such that they are projected to reach a proportion of p_0 at time zero. We used starting values of $p_0 = 0.1$, $p_e = 0.8$, and t_e equal to the mean date of the samples with nonzero steppe ancestry. We then chose optimal parameter values for the model by finding values for p_0, p_e, and t_e that minimized the residual sum of squares function between fitted and observed steppe ancestry proportions. Analyses were performed using R version 3.4.3 (R Core Team 2017) with the function *optimx*[49]. To avoid downward-biased estimates of p_e, we set individuals with steppe ancestry proportion zero, which are observed after t_e, to have zero residual value.

To incorporate uncertainty in dating estimates, we did not use the mean date for each individual, and instead randomly sampled a date for each individual from a normal distribution truncated such that the upper and lower bounds of the date estimates form a 95% confidence interval around the sampling mean. We repeated this sampling process 100,000 times and found an optimal value for t_e for each sample, for each dataset. Optimal values of t_e were only retained if they produced stable solutions with a positive-definite Hessian matrix.

For the uncertainty in the proportion of steppe ancestry it is assumed it is identically distributed, with mean zero and unbiased for all points, as it is standard in least-squares regression. We report kernel density estimates weighted by the ratio of the exponential of the sum of the squared residuals, compared with the optimal residual sum of squares observed, which is proportional to the ratio of Gaussian likelihood functions. This approach allows parameter values that performed best to be given more weight in the report kernel density

functions and approximates implementing a Monte Carlo Markov chain with a truncated Gaussian prior distribution for sampling times, a Gaussian distribution for the residuals, and a uniform proposal density function for parameter values.

Pairwise mismatch rates

The pairwise mismatch rates were calculated from the genotype file and only pairs with more than 10,000 overlapping SNPs of the 1240k SNP panel were included. Due to the heavy right-skewed distribution of the pairwise mismatch rates, we used a robust linear mixed-effects model to assess the difference between pairwise mismatch rates before and after 2700 BCE. To account for correlations in the data due to the same samples being used to calculate multiple pairwise distances, we included the ID of the samples as random effects. We found that both the time period (before and after 2700 BCE) and the population of origin for samples were significant predictors for the pairwise mismatch rate. Analyses were performed using R version 3.4.3 (R Core Team 2017) via the *robustlmm* package[50].

Admixture date estimates

Admixture dates were estimated using DATES, which was extensively tested in ref. [22], based on the 1240k SNP panel for single individuals and for groups of individuals (ref. [22], https://github.com/priyamoorjani/DATES).

Kinship analysis

Kinship between individuals was assessed using the software READ[23], *lcmlkin*[24], and calculating pairwise mismatch rate on autosomal markers and confirmed by mtDNA haplotype and Y chromosomal haplogroups based on the 1240k SNP panel (Table 1).

qpWave analysis

Population continuity since the Bronze Age was tested with *qpWave* (version: 410) from ADMIXTOOLS (https://github.com/DReichLab) between newly sequenced Swiss Bronze Age individuals and modern Swiss individuals from POPRES using Mbuti, Karitiana, Hakka Taiwan, Papuan, Onge, Han, Hungarian, MA1, EHG, and WHG as outgroups and differences in the genetic affinity between the Bronze Age individuals and the four linguistic regions of Switzerland were assed with D-statistics of the form *f4(X,Y; Test, Outgroup)*.

Functional SNP analysis

Phenotypic SNPs were genotyped using GATK version 3.8[51]. Low coverage positions (e.g., 1×) with reference or alternative alleles carrying A or T were inspected manually to exclude the influence of ancient DNA damage.

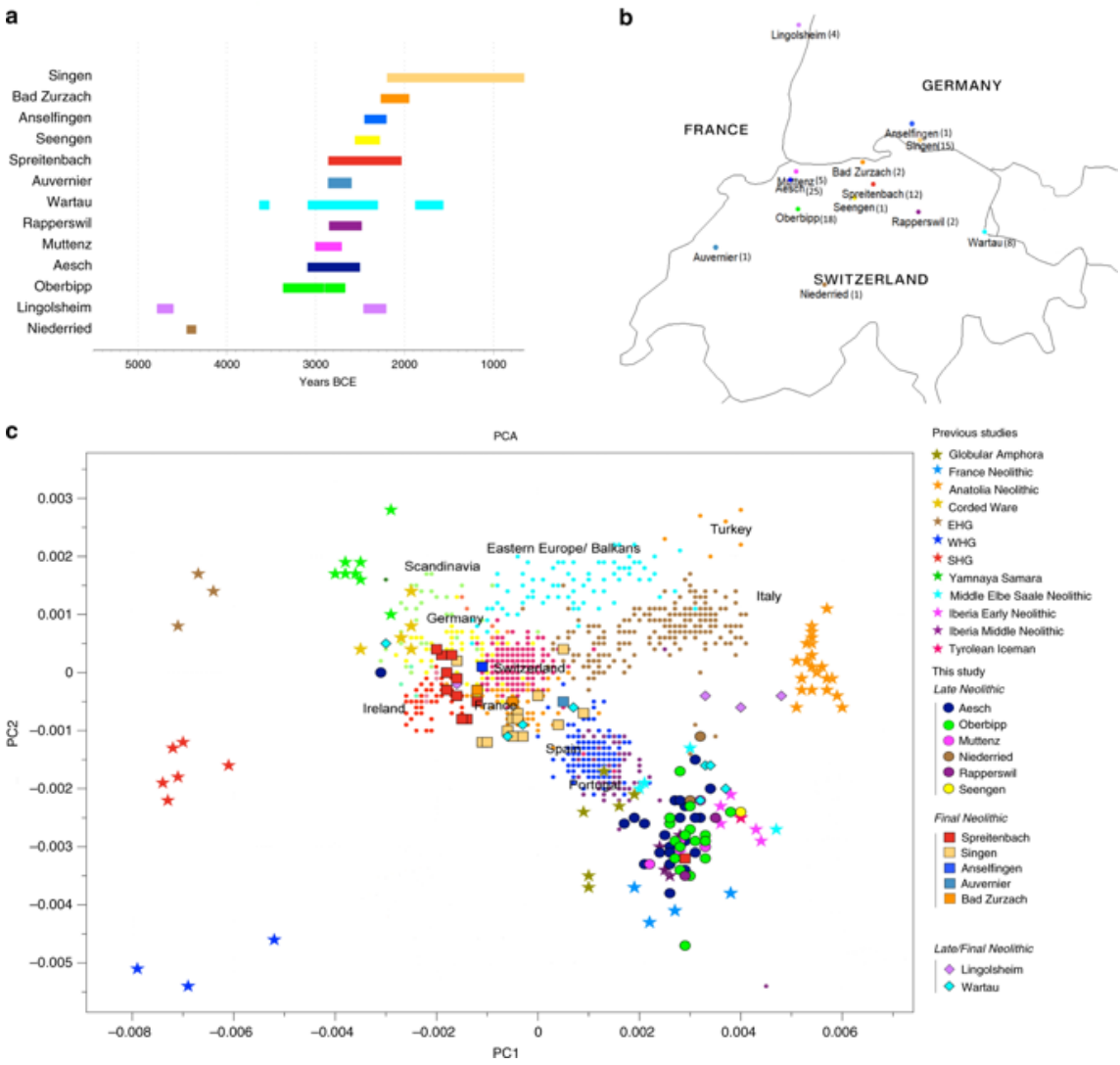


Fig. 1: Genetic, temporal, and spatial structure of individuals in this study. a Time ranges of calibrated radiocarbon dates of the archaeological sites. **b** Geographical distribution of the sites and samples sizes per sites in brackets. Map generated with R version 3.4.3 (R Core Team 2017) using the CIA World Data Bank II currently (mid 2003) available from http://www.evl.uic.edu/pape/data/WDB/. **c** PCA was reconstructed on 1960 modern European individuals of the POPRES dataset and ancient genomes were projected onto it.

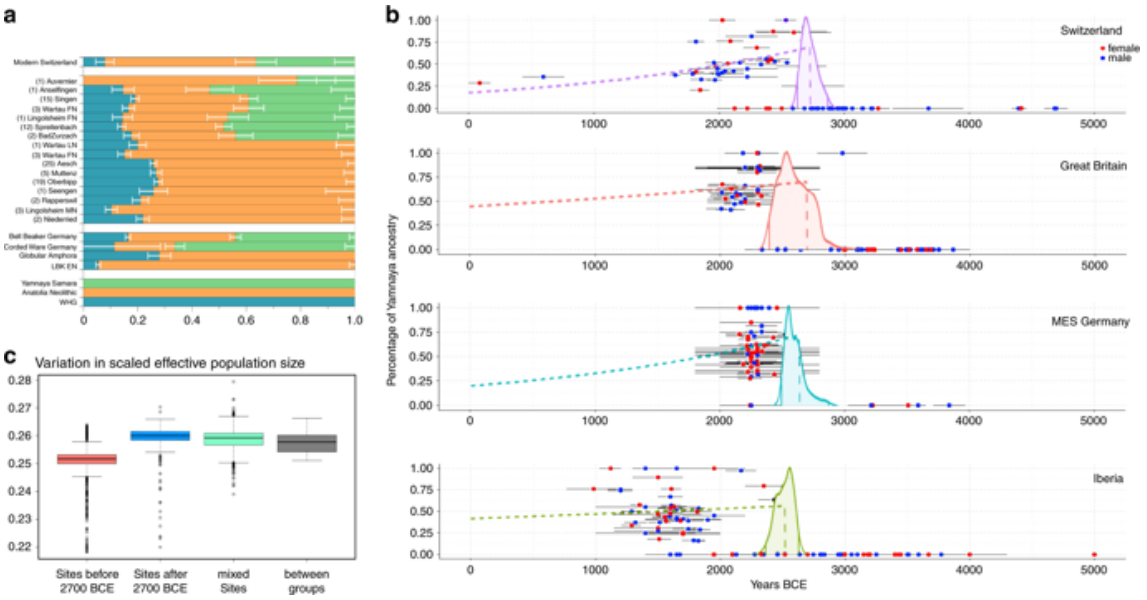


Fig. 2: Genetic turnover at the transition to the Central European Bronze Age. a Three-way qpAdm models of the ancient individuals from Switzerland (number of individuals in brackets) sorted by sites in chronological order (bottom to top) with the source populations WHG, steppe pastoralists (Yamnaya Samara) and Anatolia Neolithic. Error bars represent standard error of the proportion of each component. **b** Relative proportion of the steppe-related ancestry component for each individual in four different regions, calculated with qpAdm and estimates of arrival times (error bars represent the range of C^{14} dating) and following decrease of the component (dashes lines). Red dots represent female individuals and blue dots male individuals. **c** An estimate of genetic diversity between individuals before 2700 BCE and after 2700 BCE and sites with individuals from both periods as well as modern European populations (German and French from the HO dataset).

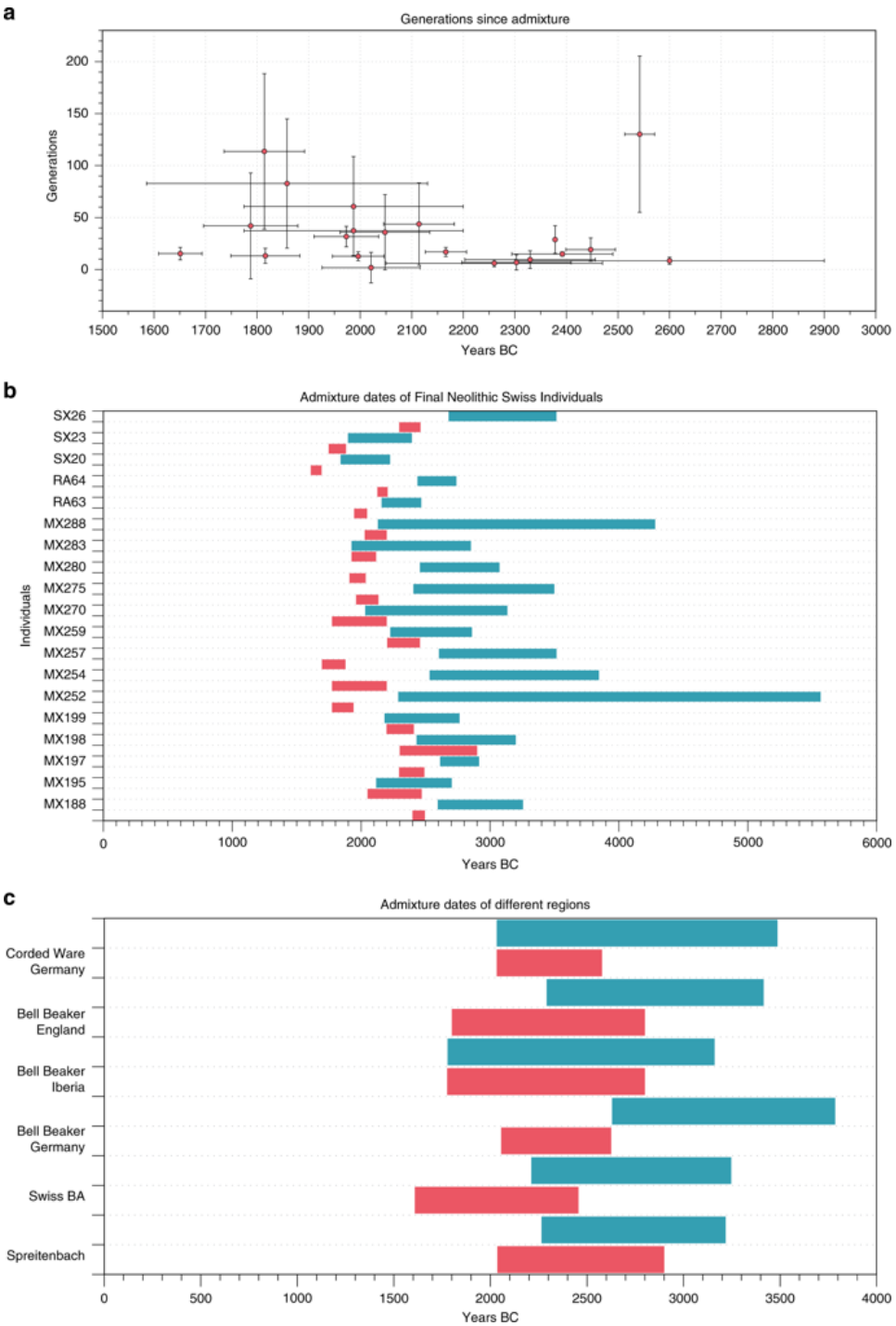
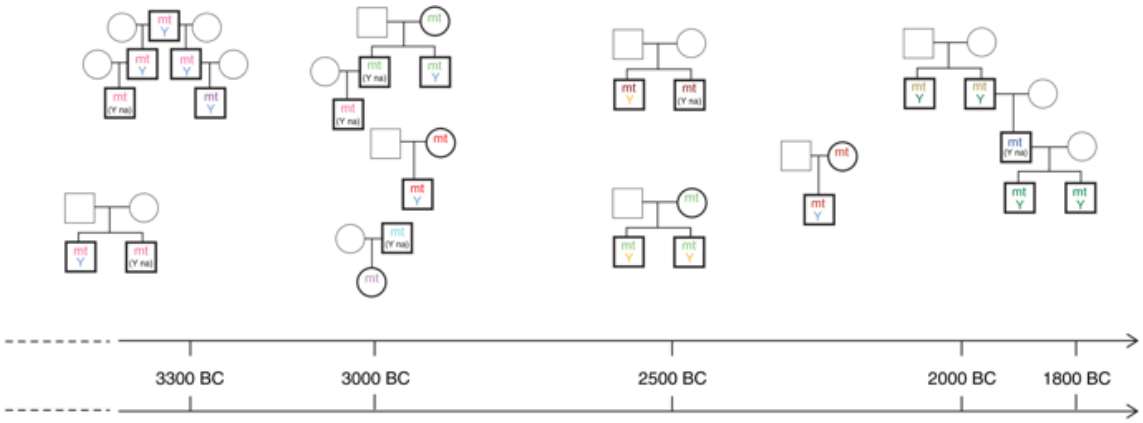


Fig. 3: Estimated admixture times between the Yamnaya-like steppe component and the Middle Neolithic population using DATES software. a Admixture dates of single individuals are plotted against their calC14 dates (horizontal error bars indicate uncertainty in C14 dating and vertical error bars show 95% confidence interval of generation times) and **b** displayed as time range (C14 dates in red and estimated admixture dates in turquoise). **c** Admixture dates of grouped individuals according to their regions of origin were calculated (colors as above).



Site	Oberbipp	Aesch	Spreitenbach	Rapperswill	Singen
Number of individuals withrelatives	7(19)	8(25)	5(12)	2(2)	5(15)
Burial type	Megalithic Dolmen	Megalithic Dolmen	Collective burial	Stone cist	Crouched burials
Region	Central Switzerland	Northwest Switzerland	Central Switzerland	Central Switzerland	South Germany
Range C14 dates related individuals	3337-2908 BC	3078-2877 BC	2860-2050 BC	2463-2274 BC	2135-1892 BC

Fig. 4: Reconstructed family trees from different sites over time. All relationships between the single individuals were reconstructed from autosomal variants and confirmed by uniparentally inherited markers such as mtDNA haplotype and Y chromosomal haplogroup. Individuals with black outline were available for analysis and individuals with gray outline were not found within the burials and are missing. Same colors indicate identical mtDNA haplotypes and matching Y chromosomal haplogroups.

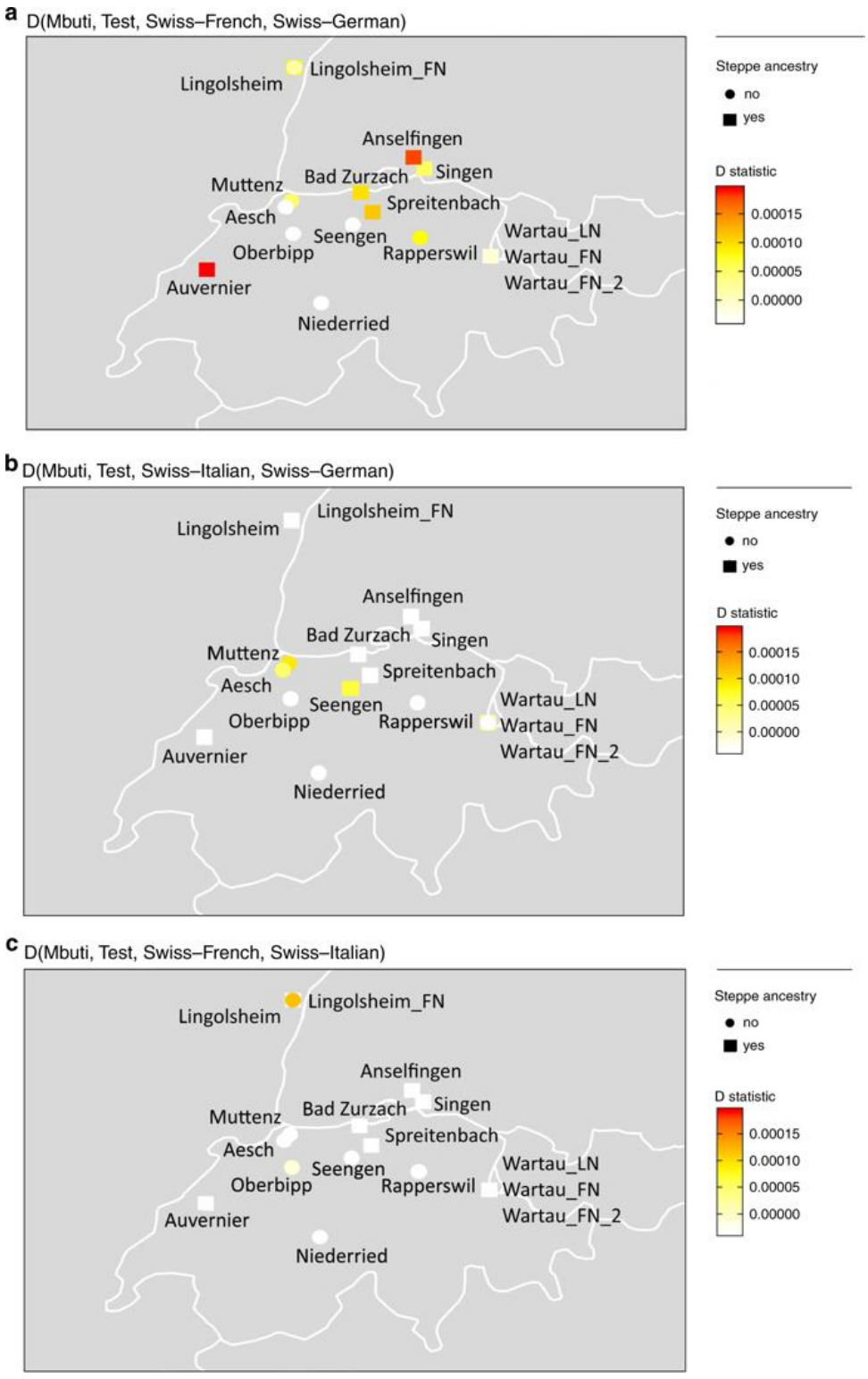


Fig. 5: Genetic affinity between the ancient Individuals and the French-, German-, and Italian-speaking regions of Switzerland. Differences in genetic affinity of the ancient individuals from Switzerland to **a** Swiss–French and Swiss–German, **b** Swiss–Italian and Swiss–German, and **c** Swiss–French and Swiss–Italian. Maps generated with R version 3.4.3 (R Core Team 2017) using the CIA World Data Bank II are currently (mid 2003) available from http://www.evl.uic.edu/pape/data/WDB/.

Table 1: Frequency of the derived allele of four phenotypic SNPs. SLC45A2 and SLC24A5 contribute to lighter skin pigmentation, HERC2 is associated with blue eyes and LCT with lactose tolerance in adults.

	frequency of derived allele SLC45A2 (rs16891982)	number of individuals position covered SLC45A2 (rs16891982)	frequency of derived allele SLC24A5 (rs1426654)	number of individuals position covered SLC24A5 (rs1426654)	frequency of derived allele HERC2 (rs12913832)	number of individuals position covered HERC2 (rs12913832)	frequency of derived allele LCT (rs4988235)	number of individuals position covered LCT (rs4988235)
Late Neolithic								
Niederried	0%	0/2	100%	2/2	50%	2/2	0%	2/2
Oberbipp	31%	13/19	100%	7/19	50%	10/19	0%	11/19
Aesch	57%	23/25	100%	8/25	50%	18/25	0%	23/25
Muttenz	0%	5/5	100%	1/5	75%	4/5	0%	5/5
Seengen	0%	0/2	0%	0/2	0%	0/2	0%	0/2
Final Neolithic								
Spreitenbach	55%	11/12	100%	5/12	36%	10/12	8%	12/12
Bad Zurzach	0%	2/2	100%	1/2	100%	2/2	0%	2/2
Singen	77%	13/15	100%	4/15	13%	8/15	0%	13/15
Anselfingen	0%	1/1	0%	0/1	100%	1/1	0%	1/1
Rapperswil	100%	2/2	0%	0/2	100%	1/2	0%	2/2
Auvernier	0%	0/1	0%	0/1	0%	0/1	0%	0/1
mixed								
Lingolsheim	33%	3/4	100%	1/4	0%	2/4	0%	4/4
Wartau	83%	6/8	0%	0/8	66%	4/8	16%*	6/8

* One individual dating to 789calBCE-2AD.

References

1. Lazaridis, I. et al. Ancient human genomes suggest three ancestral populationsfor present-day Europeans.Nature513, 409–413 (2014).
2. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancientEurasians.Nature528, 499–503,https://doi.org/10.1038/nature16152(2015).
3. Lazaridis, I. et al. Genomic insights into the origin of farming in the ancientnear east.Nature536, 419–424 (2016).
4. Omrak, A. et al. Genomic evidence establishes Anatolia as the source of theEuropean Neolithic gene pool.Curr. Biol.26, 270–275 (2016).
5. Lipson, M. et al. Parallel paleogenomic transects reveal complex genetichistory of early European farmers.Nature551, 368–372 (2017).
6. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe.Nature522, 207–211 (2015).
7. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia.Nature522,167–172 (2015).
8. Furholt, M.Die absolutchronologische Datierung der Schnurkeramik inMitteleuropa und Südskandinavien, Vol 101 (Universitätsforschungen zurprähistorischen Archäologie, 2003).
9. Olalde, I. et al. The Beaker phenomenon and the genomic transformation ofnorthwest Europe.Nature555, 190–196 (2018).
10. Siebke I. et al. Who lived on the Swiss Plateau around 3300 BCE? Analyses ofcommingled human skeletal remains from the dolmen of Oberbipp.Int. J.Osteoarchaeol. https://doi.org/10.1002/oa.2791(2019)
11. Hafner A. & Suter P. J. Das neolithikum in der Schweiz.J. Neolit. Archaeol.5,https://doi.org/10.12766/jna.2003.4(2003).
12. Hafner, A., Heitz, C. & Stapfer, R.Pile-dwellings of the Neolithic and theBronze Age in Switzerland. Long-term Research and Future Tasks, Vol. 10(Schriften des Archäologischen Landesmuseum Schleswig, 2014).
13. Siebke, I. et al. The application of different 3D-scan-systems andphotogrammetry at an excavation—a Neolithic dolmen from Switzerland.Digit. Appl. Archaeol. Cult. Heritage10, e00078 (2018).
14. Suter P. J. et al. Um 2700 v. Chr.—Wandel und Kontinuität in denUfersiedlungen am Bielersee Bern.Archäologischer Dienst des Kantons Bern2Bände (463, 300 Seiten) (2017).
15. Jones, E. R. et al. Upper Palaeolithic genomes reveal deep roots of modernEurasians.Nat. Commun.6, 8912 (2015).
16. Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroupclassification in the era of high-throughput sequencing.Nucleic Acids Res.44,W58–63 (2016).
17. Nelson, M. R. et al. The population reference sample, POPRES: a resource forpopulation, disease, and pharmacological genetics research.Am. J. Hum.Genet.83, 347–358 (2008).
18. Keller, A. et al. New insights into the Tyrolean Iceman's origin andphenotype as inferred by whole-genome sequencing.Nat. Commun.3, 698(2012).

19. Patterson, N. et al. Ancient admixture in human history.Genetics192,1065–1093,https://doi.org/10.1534/genetics.112.145037(2012).
20. Saag, L. et al. Extensive farming in estonia started through a sex-biasedmigration from the Steppe.Curr. Biol.27, 2185–2193.e6 (2017).
21. Olalde, I. et al. The genomic history of the Iberian Peninsula over the past8000 years.Science363, 1230–1234 (2019).
22. Narasimhan, V. M. et al. (2019) The formation of human populations inSouth and Central Asia.Science365,https://doi.org/10.1126/science.aat7487(2019).
23. Lipatov, M., Sanjeev, K., Patro, R. & Veeramah, K. Maximum likelihoodestimation of biological relatedness from low coverage sequencing data.bioRxivhttps://doi.org/10.1101/023374(2015).
24. Monroy Kuhn, J. M., Jakobsson, M. & Günther, T. Estimating genetic kinrelationships in prehistoric populations.PLoS ONE13, e0195491,https://doi.org/10.1371/journal.pone.0195491(2018).
25. Sánchez-Quinto, F. et al. Megalithic tombs in western and northern NeolithicEurope were linked to a kindred society.Proc. Natl Acad. Sci. USA116,9469–9474,https://doi.org/10.1073/pnas.1818037116(2019).
26. Knipper, C. et al. Female exogamy and gene pool diversification at thetransition from the Final Neolithic to the Early Bronze Age in central Europe.Proc. Natl Acad. Sci. USA114, 10083–10088 (2017).
27. Novembre, J. et al. Genes mirror geography within Europe.Nature456(7218),98 (2008).
28. Mathieson, I. et al. The genomic history of southeastern Europe.Nature555,197–203 (2018).
29. Siebke I. et al. Crops vs. animals: regional differences in subsistence strategiesof Swiss Neolithic Farmers revealed by stable isotopes. (2020). (In press).
30. Price, T. D., Bentley, R. A., Lüning, J., Gronenborn, D. & Wahl, J. Prehistorichuman migration in the Linearbandkeramik of Central Europe.Antiquity75,593–603,https://doi.org/10.1017/S0003598X00088827(2001).
31. Stöckli, W. E.Die Schweiz vom Paläolithikum bis zum frühen Mittelalter. VomNeandertaler bis zu Karl dem Grossen=La Suisse du Paléolithique à l'aube duMoyen-Age=La Svizzera dal Paleolitico all'Alto Medioevo, (Verl. Schweizer.Ges. für Ur- und Frühgeschichte, Basel, 1995).
32. Doppler T. (ed)Spreitenbach-Moosweg (Aargau, Schweiz): Ein Kollektivgrabum 2500 v. Chr, (Archäologie Schweiz, Basel, 2012).
33. Berger, B. et al. High resolution mapping of Y haplogroup G in Tyrol(Austria).Forensic Sci. Int. Genet.7, 529–536,https://doi.org/10.1016/j.fsigen.2013.05.013(2013).
34. Pichler, I. et al. Genetic structure in contemporary South Tyrolean isolatedpopulations revealed by analysis of Y-chromosome, mtDNA, and Alupolymorphisms. 2006.Hum. Biol.81, 875–898,https://doi.org/10.3378/027.081.0629(2009).3
5. Mittnik, A. et al. Kinship-based social inequality in Bronze Age Europe.Science366, 731–734 (2019).

36. Lazaridis, I. et al. Genetic origins of the Minoans and Mycenaeans.Nature548, 214–218,https://doi.org/10.1038/nature23310(2017).
37. Fernandes, D. M. et al. The arrival of steppe and Iranian related ancestry inthe islands of the Western Mediterranean.Nature Ecology and Evolution4,334 (2019).
38. Marcus, J. H. et al. Population history from the Neolithic to present on theMediterranean island of Sardinia: an ancient DNA perspective.Nat. Commun.11, 939 (2020).
39. Dabney, J. et al. Complete mitochondrial genome sequence of a MiddlePleistocene cave bear reconstructed from ultrashort DNA fragments.Proc.Natl Acad. Sci.39, 15758–15763 (2013)
40. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highlymultiplexed target capture and sequencing.Cold Spring Harb. Protoc. 5448(2010).
41. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuraciesin multiplex sequencing on the Illumina platform.Nucleic Acids Res.1,e3(2011).
42. Furtwängler, A. et al. Ratio of mitochondrial to nuclear DNA affectscontamination estimates in ancient DNA analysis.Sci Rep.8,1–8.https://doi.org/10.1038/s41598-018-32083-0.
43. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA.Philos. Trans. R.Soc. B370, 20130624 (2014).
44. Fu, Q. et al. DNA analysis of an early modern human from Tianyuan Cave,China.Proc. Natl Acad. Sci.110, 2223–2227 (2013).
45. Peltzer, A. et al. EAGER: efficient ancient genome reconstruction.GenomeBiol.17.1, 1 (2016).
46. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation ofcontamination and endogenous mitochondrial consensus calling for ancientDNA.Genome Biol.16.1, 1 (2015).
47. Mittnik, A., Wang, C.-C., Svoboda, J. & Krause, J. A molecular approach to thesexing of the triple burial at the upper paleolithic site of Dolní Věstonice.PLoSONE11, e0163019 (2016).
48. Rasmussen, M. et al. An Aboriginal Australian genome reveals separatehuman dispersals into Asia.Science334,94–8 (2011).
49. Nash, J. C. & Varadhan, R. Unifying optimization algorithms to aid softwaresystem users: optimx for R.J. Stat. Softw.43,1–14,http://www.jstatsoft.org/v43/i09/(2011).
50. Koller, M. robustlmm: an R package for robust estimation of linear mixed-effects models.J. Stat. Softw.75.6,1–24. (2016).
51. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce frameworkfor analyzing next-generation DNA sequencing data.Genome Res.20,1297–1303 (2010).

Acknowledgments

This study was funded by the German Research Foundation KR 4015/4-1 and the Swiss National Foundation CR31I3L_157024. This project has also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement No 771234 – PALEoRIDER (WH, ABR, LP) and

the Heidelberg Academy of Sciences (WIN project “Times of Upheaval: Changes of Society and Landscape at the Beginning of the Bronze Age”) (PWS, JK). We thank Anna Sapfo Malaspinas, Ivan P. Levkivskyi, Sven Bergman, Michal Feldman, and Choongwon Jeong for helpful discussions and comments. Special thank goes to the Archaeological Service of the Canton of Bern (Switzerland). The authors also would like to thank the Archaeological Services of the Cantons of Aargau (Switzerland), Basel-Landschaft (Switzerland), Solothurn (Switzerland), St. Gallen (Switzerland), and the Archaeological Services of Baden-Württemberg (Germany) and Alsace (France). Furthermore, we thank the Swiss National Museum of Zurich and Sönke Szidat and his team from the LARA laboratory at Bern University for radiocarbon dating and fruitful discussions.

Author Contributions

A.H., S.L., and J.K. conceived of the study. A.F., E.R., and G.U.N. performed laboratory work. A.F., A.B.R., T.C.L., L.P., W.H., and S.S. analyzed data. I.S. and S.L. performed anthropological assessments. I.S., N.S., J.H., A.D., B.S., J.W., P.W.S., S.L. and A.H. assembled and interpreted archaeological material. A.F., A.B.R., and S.S. designed figures. V.J.S. and J.K. supervised laboratory work. A.F., S.S., and J.K. wrote the paper with input from all co-authors.

Competing interests

The authors declare no competing interests.

Ancient genomes reveal social and genetic structure of Late Neolithic Switzerland

Supplementary Information

Authors: Furtwängler A[1], Rohrlach AB[2,3], Lamnidis TC[2], Papac L[2], Neumann GU[1,2], Siebke I[4], Reiter E[1], Steuri N[5], Hald J[6], Denaire A[7], Schnitzler B[8], Wahl J[9,10], Ramstein M[11], Schuenemann VJ[1,12,13], Stockhammer PW[2,14], Hafner A[5,15], Lösch S[4], Haak W[2], Schiffels S[2], Krause J[1,2,11*]

Affiliations:

[1]Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Germany
[2]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany
[3]ARC Centre of Excellence for Mathematical and Statistical Frontiers, School of Mathematical Sciences, The University of Adelaide, Adelaide, South Australia 5005, Australia
[4]Department of Physical Anthropology, Institute of Forensic Medicine, University of Bern, Switzerland
[5]Institute of Archaeological Sciences, Prehistoric Archaeology, University of Bern, Switzerland
[6]Archaeological Office of the District of Constance, Germany
[7]Department of history of arts and Archaeology, University of Burgundy, France
[8]Museum of Archaeology Strasbourg, France
[9]Institute for Archaeological Science, Palaeoanthropology, Eberhard Karls University Tübingen, Germany
[10]State Office for Cultural Heritage Management Baden-Wuerttemberg, Konstanz, Germany
[11]Archaeological Service of the canton of Bern
[12]Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Germany
[13]Institute of Evolutionary Medicine, University of Zurich, Switzerland
[14]Institut für Vor- und Frühgeschichtliche Archäologie und Provinzialrömische Archäologie, Ludwig-Maximilians University Munich, Germany
[15]Oeschger Centre for Climate Change Research, University of Bern, Switzerland
* Corresponding author: krause@shh.mpg.de

Contents:

Refer to CD for Supplementary Data 1 – 4

Supplementary Note 1: Archaeological information of the sites included in this study

Middle Neolithic

CH Niederried Ursisbalm 4458 – 4341 BCE (2sigma cal.)
Tschumi 1920, Hug 1956, Schlaginhaufen 1924
The stone cist burial near Niederried was discovered and excavated in 1913. After rock blasting on a construction site, the graves were found below. The burial had an NW-SE orientation and was constructed with five tiles of chalk. The individual in the grave, an adult, was found in crouched position. Outside the stone cist, the bones of two subadults were found and probably from an earlier burial which was later removed from the cist. Four additional stone cist burials were destroyed by further blasts. Samples from two individuals were used for genetic analyses: A1 and A2 and showed sufficient DNA preservation. The two individuals, one female, and one male are unrelated. Radiocarbon dating hints to an age of 4458 – 4341 calBCE. Samples were provided by the Archaeological Service of the canton of Bern.

F Lingolsheim 4786 – 2208 BC (2sigma cal.)
Lichardus-Itten 1980
The burial ground of Lingolsheim was discovered in spring 1910 in a sandpit and comprises at least 20 graves. The individuals were found in stretched supine position in NW-SE orientation. From these individuals, one petrous bone from tomb 15 was used for genetic analysis. The male individual was dated to 4766-4601 calBCE.
Between 1923 and 1926 a second field session took place in a nearby second sandpit and 17 more graves were discovered. Four of these individuals were buried in crouched position in W-E orientation. Based on found potsherds the graves were assigned to the Rubané culture. From this group, the individual from tomb E was sampled for this study. The male individual dates to 2463-2208 calBCE contradicting the cultural assignment.
The 13 remaining burials were also found in stretched supine position and as the former ones from 1910 assigned to the Grossgartach culture. Petrous bones from the graves 18, 32 and 35 were used for DNA analysis. Only the individuals 18 and 35, a male and a female, contained enough DNA for downstream analysis. Only individual 35 could be dated to 4786-4618 calBCE. Samples were provided by the Museum of Archaeology in Strasbourg.

Late Neolithic

CH Oberbipp 3350-2650 BCE (2sigma cal.)
Ramstein et al. 2014, Siebke et al. 2018, Siebke et al. 2019, Steuri et al. 2019
The Dolmen burial from Oberbipp was excavated in 2012. The site is situated on the Swiss Plateau and on the southern slope of the Jura mountains. The construction of the dolmen consists of large glacial erratics in the form of blocks and slabs. More than 2000 fragmented and commingled skeletal elements were retrieved at the excavation and a minimal number of 42 individuals as determined from femora including males and females of all age classes. An extended supine position could be reconstructed for the buried individuals in the grave

chamber. In total 22 petrous bones from different individuals and 46 teeth could be sampled for aDNA analyses and radiocarbon dating. Samples were provided by the Archaeological Service of the canton of Bern.

CH Muttenz 3010 – 2706 BC (2sigma cal.)
Muttenz, and Hagmann D, editors. 2009; Siebke et al. 2019
The multiple burial of Muttenz was discovered in 1946. Details about the finding situation and the excavation are unknown. It contained at least 11 individuals of which five petrous bones from different individuals could be sampled for aDNA and radiocarbon dating. Samples were provided by the Archaeological Service of the canton of Basel Landschaft.

CH Aesch 3094 – 2500 BC (2sigma cal.)
Löhlein 2011; Stöckli 1995, Schegler 2016, Bay 1936
The dolmen burial of Aesch was excavated in 1907 and 1909. The megalithic burial consisted of one rectangular grave chamber 4m long and 3m wide with a flagged floor. Skeletal remains of 33 adult individuals and 14 children were recovered. 25 of the skeletons comprised petrous bones, which were used for DNA analysis and radiocarbon dating. Samples were provided by the Archaeological Service of the canton of Basel Landschaft.

Final Neolithic/ Early Bronze Age

CH Spreitenbach 2900 – 2031 BC (2sigma cal.)
Doppler 2012, Bleuer et al. 1999
In 1997 the multiple burial of Spreitenbach, Canton Aargau in Switzerland was discovered and excavated. The former wooden construction contained five males, four females, two subadults and a newborn in crouched position. The burials are radiocarbon dated to approximately 2500 BCE. All 12 individuals (pars petrosae) could be sampled for additional aDNA analyses. Samples were provided by the Archaeological Service of the canton of Aargau.

CH Bad Zurzach 2206 – 1946 BC (2sigma cal.)
Bleuer et al. 2012, Doswald et al. 1989
The double burial of Bad Zurzach was excavated in 1984. The grave-pit is in N-S orientation without further stone or wooden construction and contained two adult individuals in contracted position. A contemporaneous burial of the two male individuals is assumed. For both a mature age of death is estimated. The burial was assigned to the Early Bronze Age and both individuals were sampled for aDNA analyses. Samples were provided by the Archaeological Service of the canton of Aargau.

CH Rapperswil 2695 – 2481 BC (2sigma cal.)
Grüninger und Kaufmann 1982
During construction works in 1980, a stone cist burial near Rapperswil was discovered. Within the W-E oriented cist, the skeletal remains of an approximately 35-year-old female buried in extended supine position were found. In the abdominal region, bones of an eight lunar months old and probably unborn foetus were found. From both individuals, mother and child DNA could be retrieved. Samples were provided by the Archaeological Service of the canton of St. Gallen.

CH Wartau 3036 BC – 2 AD (2sigma cal.)
Stehrenberger 2016, Stehrenberger 2019
The human remains are located in a natural cave. In the 1970s and 1980s, several inspections by the archaeological service of the Canton St. Gallen took place. The cave was accessible to the public at all times and in the 1970s some human remains were recovered illegally and were transferred to the archaeological service in 2001. Radiocarbon dates of bone fragments date to the Middle, Late, and Final Neolithic period as well as to the Iron Age/Roman time. Eleven petrous bones were assed for DNA analyses. Samples were provided by the Archaeological Service of the canton of St. Gallen.

D Singen 2199 - 431 BC (2sigma cal.)
Stockhammer et al. 2015, Oelze et al. 2011, Krause 1988
The Early Bronze Age cemetery of Singen (Hohentwiel) was excavated in the 1950s and is one of the largest and most important cemeteries of this period in Southern Germany. In close vicinity, Final Neolithic and later Bronze to Late Iron Age burials were excavated – providing an outstanding continuity of burials at the site. The Early Bronze Age necropolis of Singen is of particular interest, as it was used to define the eponymous "Singen Group" of the Early Bronze Age in Southwestern Germany characterized by the strict positioning of the deceased, stone constructions and particular burial goods. In total, 96 graves were excavated, and it was assumed that they can be related to one kin group. Human remains of approximately 30 individuals were recovered and 15 petrous bones selected for aDNA extraction. Radiocarbon dates of these individuals' range between 2199 - 431 BC. Samples were provided by the Archaeological Office of the District of Constance.

D Anselfingen 2456-2203 BC (2sigma cal.)
Merkl 2016
The site of Anselfingen "Breite", Germany was first discovered in 2008 and excavated during several excavation sessions in the years from 2009 till 2017. Multiple Corded Ware and Bell Beaker associated graves could be recovered. Available for DNA analysis was the petrous bone of an adult individual from a double burial associated with the Bell Beaker phenomenon which was excavated in 2010. This individual was buried in N-S orientation and in crouched position and buried with the remains of a small child. Samples were provided by the Archaeological Office of the District of Constance.

CH Auvernier 2866-2601 BC (2sigma cal.)
Schwegler 2016
The dolmen of Auvernier was discovered in 1876 near the lake Neuchâtel. The complex consisted of 12 vertical stone slabs and two horizontal ceiling tiles. In the burial chamber inhumations of 15 to 20 individuals could be found. Marks on the stone slabs indicate the reuse from an older Neolithic construction which was possibly used around 3000 BCE. The rebuilt dolmen is dated at 2800 BCE at the earliest and was used until the Bronze Age. The only DNA could be retrieved from one individual that dates to 2866-2601 calBCE. Samples were provided by the Archaeological Service of the canton of Bern.

CH Burgäschisee 2862-2581 BC (2sigma cal.)
Ulrich-Bochsler 2012; Schlaginhaufen 1924; Wiedmer-Stern 1904
The burial of Burgäschisee was discovered in 1902 near the shore of the lake Burgäschisee. During 1902 and 1943 some human remains were found. In total inhumation of possibly three individuals could be recovered. After the original description, one individual belongs to the Neolithic period and was buried in crouched position. Not all bones of the skeletons were recovered and preserved. The sample of a femur fragment did not provide enough DNA for subsequent analyses and was excluded from population genetic analysis. Samples were provided by the Archaeological Service of the canton of Bern.

CH Seengen 2463-2274 BC (2sigma cal.)
Bleuer et al. 2012
The remains of a stone cist by Seengen was excavated in 1993. The burial contained two individuals, one of them consisted only of cremated remains. The grave structure is archaeologically dated to the Bronze Age. A petrous bone of the unburned human was used for DNA analyses. . Samples were provided by the Archaeological Service of the canton of Aargau.

Supplementary Note 2: Radiocarbon dating of the newly sequenced individuals

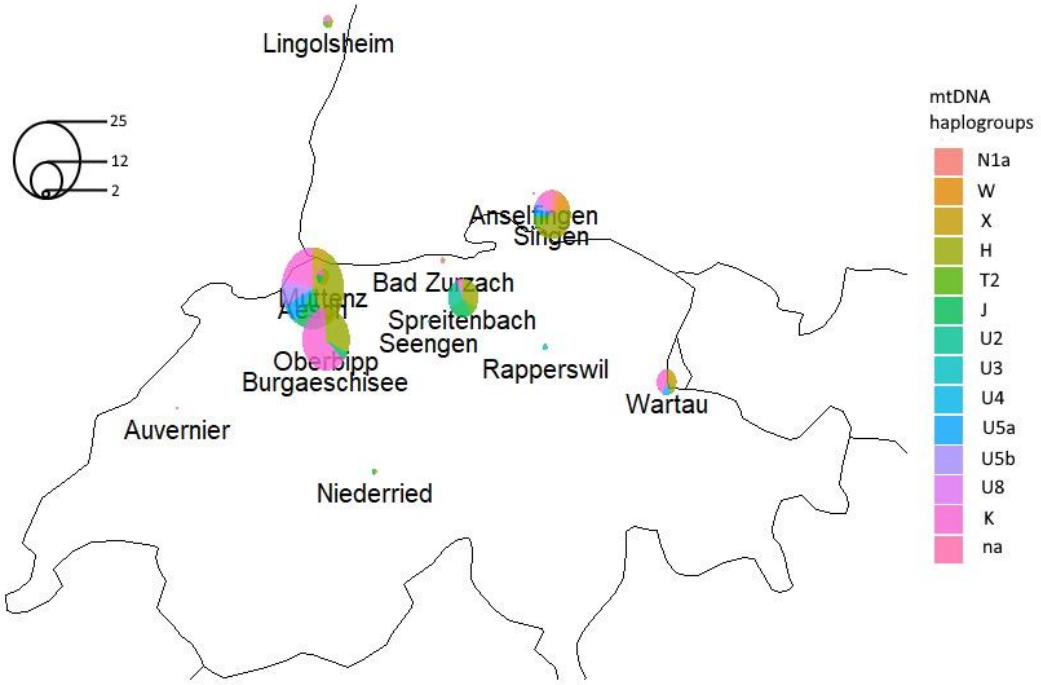
Every individual sequenced in this study was radiocarbon dated (SI table 1). For the individuals 1, 6/14, 7/11, 9, 10/16 and 13/18 the C^{14} dates from the original publication were adopted (**Doppler 2012**). Samples for the new radiocarbon dates were always taken from the same skeletal element (petrous bone, femur or tooth) from which the bone powder for DNA extraction was taken.

The dating was conducted in the LARA (Laboratory for the Analysis of Radiocarbon with AMS) at the University of Bern. The individuals from Oberbipp were also dated at the Curt-Engelhorn-Zentrum in Manheim. In case of large differences between the results of the two measurements, samples were dated a third time at the ORAU (Oxford Radiocarbon Accelerator Unit) in Oxford (**Steuri et al. 2019**).

All burials are far from coastal areas, therefore the reservoir effect due to the consumption of marine fish can be neglected since reservoir effects in freshwater lakes are not detectable so far. Furthermore, the studied populations were retrieved from burials offsite lake shores. Even though many Neolithic settlements in Switzerland were found at the lake shores, the burials are usually not associated with or in close proximity to them. In addition, the data collected for these settlements indicates a high proportion of husbandry and only limited fish consumption. This is also supported by stable isotope analysis of the studied individuals (**Siebke et al 2019**). Therefore, any possible bias based on "hard water effect" can be neglected as well.

Supplementary Note 3: Mitochondrial genome reconstruction

All libraries were enriched for mtDNA following the procedure described in **Futwängler et al. 2018** and sequenced on an Illumina HiSeq at the Max Planck Institute for the Science of Human History in Jena. After processing the raw reads with EAGER (**Peltzer et al 2016**) contamination estimates and consensus sequences were computed with *schmutzi* (**Renaud et al. 2015**). Shared haplotypes were identified using MEGA-X and haplogroups were determined using *haplogrep* (**Weissensteiner et al. 2016**, SI Tab. 1). Percentual portions of the makrohaplogroups N1a, W, X, H, T2, J, U2, U3, U4, U5a, U5b, K, and U8 were calculated for each site (SI Fig. 1).



Supplementary Figure 1 | Composition of the mtDNA haplogroups N1a, W, X, H, T2, J, U2, U,3 U4, U5a, U5b and U8 for each site. The size of the circles indicates the sample size for each site. Map generated with R version 3.4.3 (R Core Team 2017) using the CIA World Data Bank II is currently (mid-2003) available from http://www.evl.uic.edu/pape/data/WDB/

Supplementary Note 4: Y-chromosomal haplogroup assignment

Assignment for Y-chromosomal haplogroups was performed via visual inspection of the captured SNPs, following the nomenclature of the International Society of Genetic Genealogy (ISOGG) version 14.218 (retrieved 19 November 2019, http://www.isogg.org).
In individuals older than 2600 BC originating from the sites of Oberbipp, Aesch and Muttenz only haplogroups belonging to the clades I2a and G2a were found. Besides the youngest Aesch individual Aesch25, which carries a R1b haplogroup.
In the younger site of Spreitenbach associated with the Corded Ware complex only haplogroups of the clade I2a are present. In the contemporaneous sample from Lingolsheim R1b could be determined.
All samples in the dataset younger than 2200 BC all individuals carry R1b, beside MX265 from Singen, which belongs to R1a.

Supplementary Table 1 | Y chromosomal haplogroup assignment for all male individuals

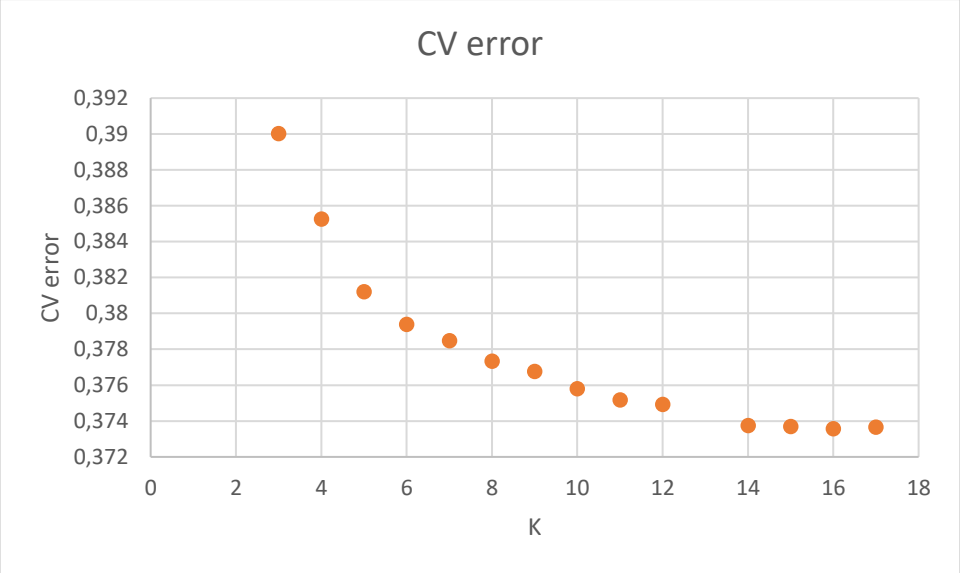
ID	Site	Cal 2 sigma BC	Terminal derived mutation	YHG
Aesch1	Aesch (CH)	3090-2917	PF3239	G2a2a1a2a1
Aesch12	Aesch (CH)	3010-2884	PF3239	G2a2a1a2a1
Aesch13	Aesch (CH)	3016-2901	PF3239	G2a2a1a2a1
Aesch14	Aesch (CH)	3014-2898	PF3239	G2a2a1a2a1
Aesch17	Aesch (CH)	3011-2889	PF3239	G2a2a1a2a1
Aesch19	Aesch (CH)	no collagen	PF3239	G2a2a1a2a1
Aesch20	Aesch (CH)	2913-2878	FGC7739/Z6488	G2a2a1a2a
Aesch21	Aesch (CH)	no collagen	PF3239	G2a2a1a2a1
Aesch22	Aesch (CH)	2892-2694	PF3239	G2a2a1a2a1
Aesch23	Aesch (CH)	2881-2676	PF3239	G2a2a1a2a1
Aesch24	Aesch (CH)	2912-2877	FGC7739/Z6488	G2a2a1a2a
Aesch25	Aesch (CH)	2864-2501	L51/M412/PF6536/S1	R1b1a1b1a1a
Aesch4	Aesch (CH)	3094-2926	PF3239	G2a2a1a2a1
Aesch6	Aesch (CH)	2905-2759	PF3147	G2a2a
Aesch7	Aesch (CH)	no collagen	FGC7739/Z6488	G2a2a1a2a
MX150	Oberbipp Horgen (CH)	3244-3102	L91/PF3246/S285	G2a2a1a2
MX182	Oberbipp Horgen (CH)	3338-3031	PF3239	G2a2a1a2a1
MX183	Oberbipp Horgen (CH)	3344-3037	FGC7739/Z6488	G2a2a1a2a
MX187	Oberbipp Horgen (CH)	3337-2908	PF3239	G2a2a1a2a1
MX188	Spreitenbach CWC (CH)	2495-2399	M423	I2a1a2
MX190	Spreitenbach CWC (CH)	2860 - 2460	M423	I2a1a2
MX191	Spreitenbach CWC (CH)	2570 - 2190	M423	I2a1a2
MX192	Spreitenbach CWC (CH)	2571-2513	M423	I2a1a2
MX195	Spreitenbach CWC (CH)	2470 - 2050	M423	I2a1a2
MX204	Oberbipp Horgen (CH)	no collagen	FGC7739/Z6488	G2a2a1a2a
MX209	Oberbipp Horgen (CH)	no collagen	PF3239	G2a2a1a2a1

ID	Site	Cal 2 sigma BC		YHG
MX210	Oberbipp Horgen (CH)	no collagen	PF3239	G2a2a1a2a1
MX211	Oberbipp Horgen (CH)	3100-2928	PF3239	G2a2a1a2a1
MX212	Oberbipp Horgen (CH)	3323-2581	PF3239	G2a2a1a2a1
MX213	Oberbipp Horgen (CH)	3363-2930	PF3239	G2a2a1a2a1
MX219	Oberbipp Horgen (CH)	3330-3216	PF3147	G2a2a
MX252	Singen (D)	1941-1774	L2/S139	R1b1a1b1a1a2b1
MX254	Singen (D)	no collagen	L2/S139	R1b1a1b1a1a2b1
MX257	Singen (D)	1879-1696	L2/S139	R1b1a1b1a1a2b1
MX258	Singen (D)	2028-1903	P312/PF6547/S116	R1b1a1b1a1a2
MX259	Anselfingen (D)	2456-2203	P312/PF6547/S116	R1b1a1b1a1a2
MX265	Singen (D)	763-431	L146/M420/PF6229	R1a
MX270	Singen (D)	no collagen	L2/S139	R1b1a1b1a1a2b1
MX275	Singen (D)	2135-1961	L2/S139	R1b1a1b1a1a2b1
MX279	Singen (D)	1882-1745	L2/S139	R1b1a1b1a1a2b1
MX283	Singen (D)	2116-1926	L2/S139	R1b1a1b1a1a2b1
MX286	Singen (D)	2029-1892	L2/S139	R1b1a1b1a1a2b1
MX288	Singen (D)	2199-2028	L2/S139	R1b1a1b1a1a2b1
MX298	Wartau (CH)	2620-2448	P303/Page108/PF3340 /S135/Z765	G2a2b2a
MX299	Oberbipp Horgen (CH)	2910-2679	PF3147	G2a2a
MX304	Auvernier (CH)	2866-2601 date from individual of the same burial	CTS5330	R1b1a2a1a
RA58	Muttenz (CH)	2937-2886	FGC7739/Z6488	G2a2a1a2a
RA61	Muttenz (CH)	2905-2865	PF3239	G2a2a1a2a1
RA62	Muttenz (CH)	2921-2886	PF3239	G2a2a1a2a1
RA63	Zuzach (CH)	2046-1946	L2/S139	R1b1a1b1a1a2b1
RA64	Zuzach (CH)	2206-2126	L2/S139	R1b1a1b1a1a2b1
SX10	Rapperswill Zürichstrasse (CH)	new-born or unborn baby of SX8	L166	G2a2a1a2a1a
SX11	Niederried Ursisbalm (CH)	4458 - 4362	PF3239	G2a2b2a1a1
SX20	Wartau (CH)	1693-1609	P312/PF6547/S116	R1b1a1b1a1a2
SX29	Lingolsheim (F)	no collagen	M423	I2a1a2
SX32	Lingolsheim (F)	2463-2208	P310/PF6546/S129	R1b1a1b1a1
SX33	Lingolsheim (F)	4766-4601	L161.1/S185.1	I2a1a2a

Supplementary Figure 2 | Composition of the Y chromosomal haplogroups for each site at different time spans. Map generated with R version 3.4.3 (R Core Team 2017) using the CIA World Data Bank II is currently (mid-2003) available from http://www.evl.uic.edu/pape/data/WDB/.

Supplementary Note 5: ADMIXTURE Analysis

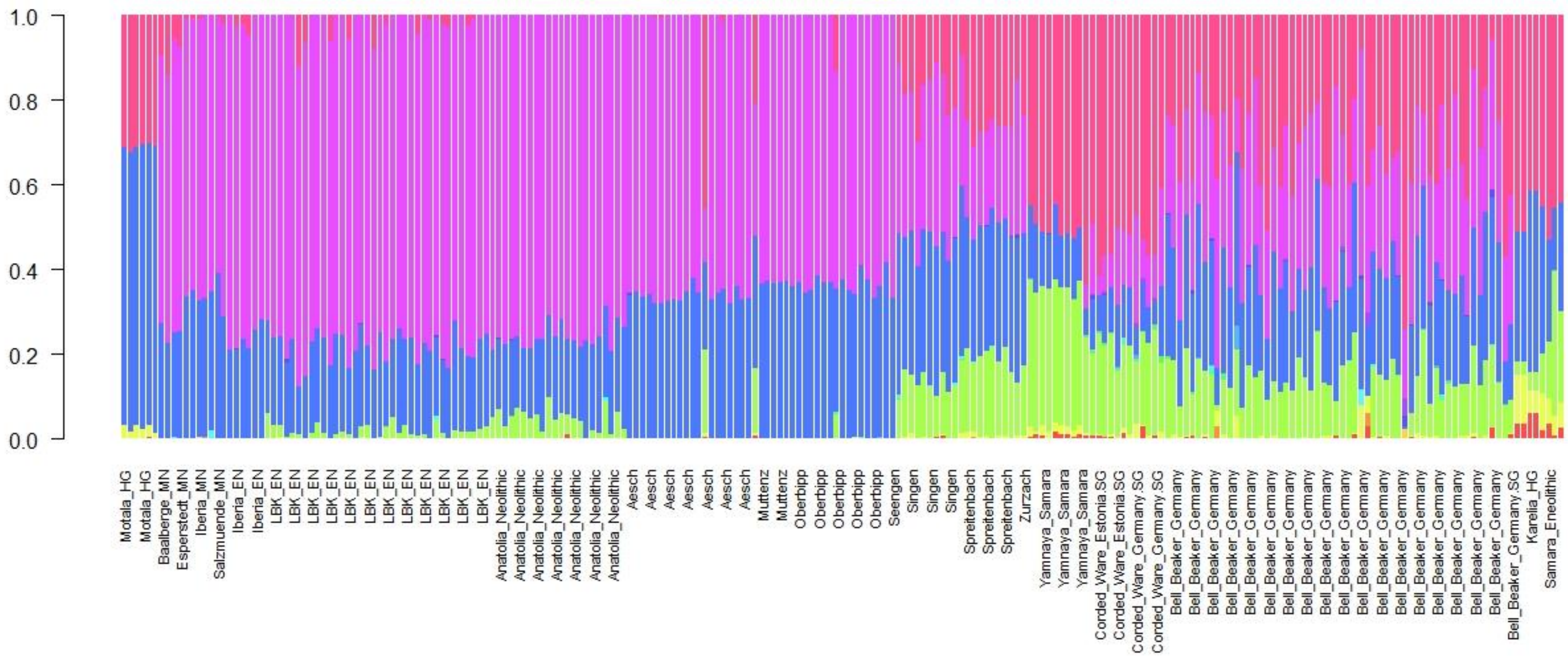
A dataset of all newly sequenced ancient individuals from Switzerland and 275 already published ancient genomes as well as modern populations from the HO dataset were used for ADMIXTURE (version 1.3) analysis. Before starting the analysis, the dataset was pruned for LD using PLINK version v1.07 and the parameters *--indep-pairwise 200 25 0.5*.
The ADMIXTURE software was run in unsupervised mode and the optimal number of k was determined using ADMIXTURE's cross-validation procedure (SI Fig 3).



Supplementary Figure 3 | CV error for k between 3 and 17. The lowest CV value can be found for k=16.

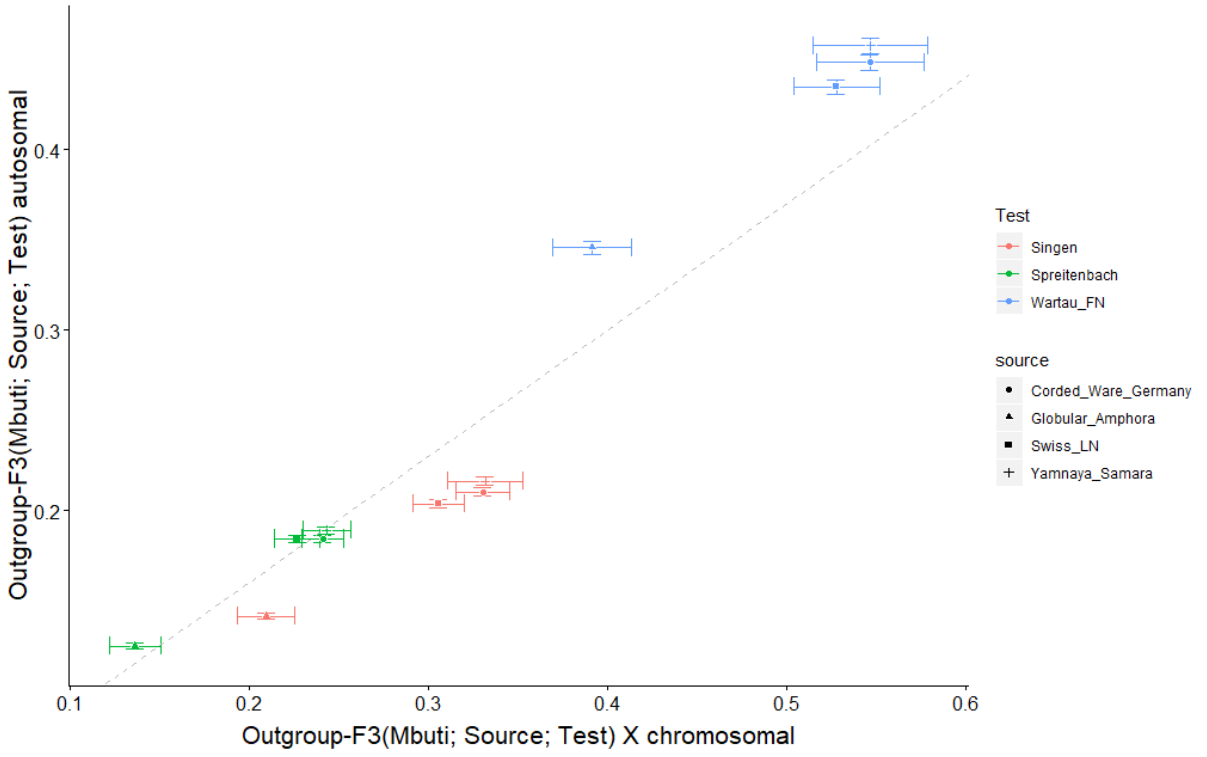
As k=16 produced the lowest CV values, the results of this clustering were plotted (SI Fig 4). These results represent the individuals from sites older than 2700 BCE as a mixture of clusters being similar to the Earls Farmers from Anatolia and WHG. Individuals from sites younger than 2700 BCE also show one additional ancestry cluster which can also be found in Yamnaya individuals from the Pontic Steppe.

Supplementary Figure 4 | Unsupervised ADMIXTURE plot (k=16) of the newly sequenced ancient Swiss individuals and selected published ancient genomes. The clustering supports the results of the PCA plot, displaying the ancient Swiss individuals as a mixture of Anatolian Farmers and WHG for sites older than 2700 BCE and showing additional ancestry in the sites younger than 2700 BCE being similar to Yamnaya-like individuals from the Pontic Steppe.



Supplementary Note 6: Sex-biased Admixture

To assess, if the admixture process between Late Neolithic individuals and immigrants with 'steppe'-related ancestry shows any sex-bias, f3 statistics of the form (Mbuti; Source; Test) using *qp3Pop* from ADMIXTOOLS (https://github.com/DReichLab) were performed on the autosomal SNPs of the HO panel and on 34,207 X chromosomal SNPs that represent the intersect of position on the X chromosome that are covered sufficiently in the individuals from the Final Neolithic sites: Singen, Spreitenbach, and Wartau_FN. As source Corded_Ware_Germany, Yamnaya_Samara, Globular_Amphora, and the Late Neolithic Swiss Sites were tested, similar to the approach described in (**Saag et al. 2017**).
The Late Neolithic individuals from our study (Spreitenbach, Singen, and Wartau_FN) are relatively symmetrically related to the sources Swiss_LN and Globluar_Amphora when comparing X chromosomes to autosomes. Relative to this the Late Neolithic Swiss individuals or our study have a higher affinity on the autosomes than on the X chromosome using the sources CWC and Yamnaya_Samara. The observed effect is stronger using Yamnaya_Samara as a source than when CWC is used. These results suggest that the influence of the steppe was stronger on the male side (SI Fig. 5). These results match the results of previously published data (**Saag et al. 2017)**.



Supplementary Figure 5 | Sex-biased admixture patterns in Final Neolithic Swiss individuals. F 3 values of statistics of the form (Mbuti; Source; Test) for autosomes and the X chromosome. Error bars represent standard deviaition of the *f3* statistics.

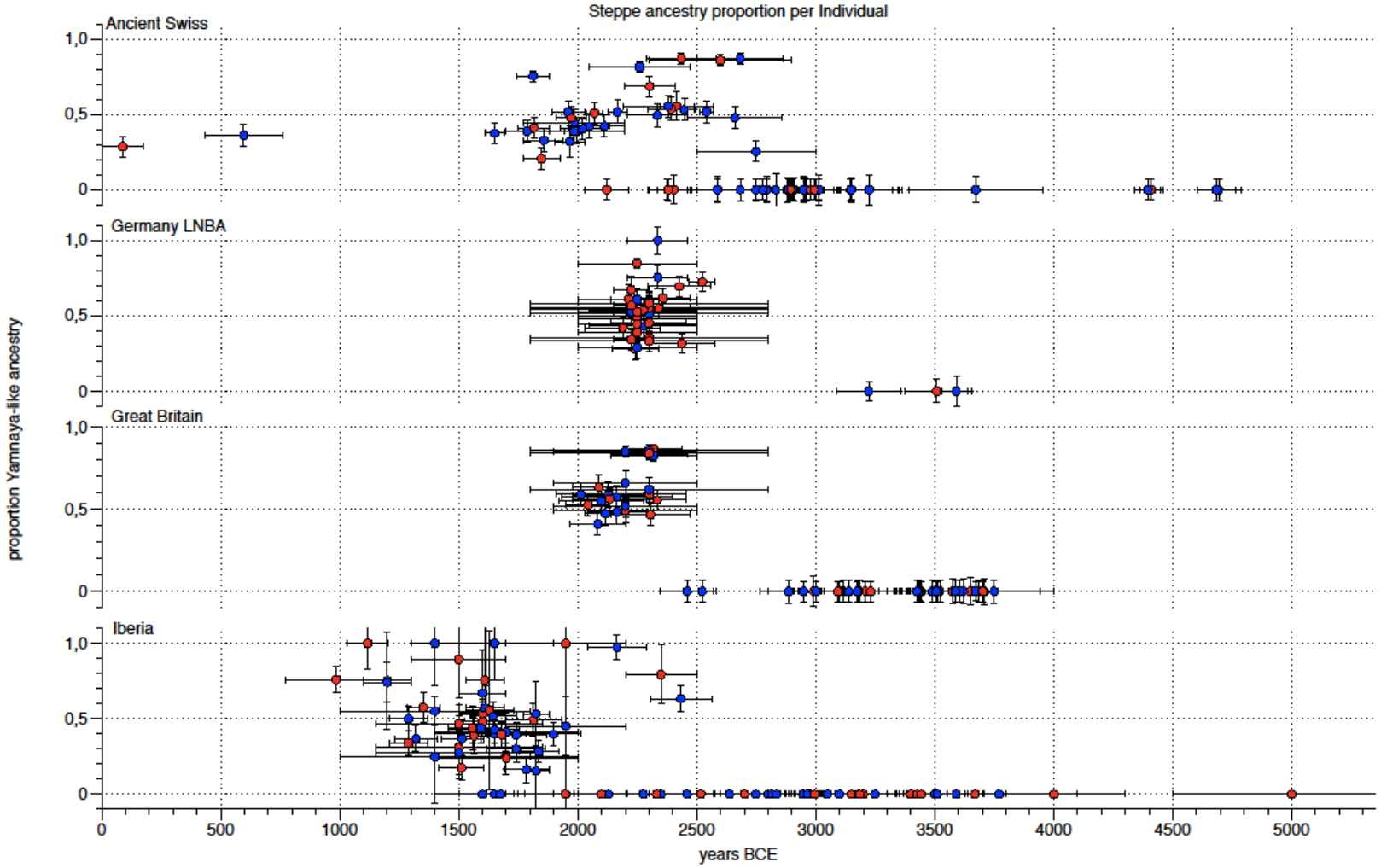
Supplementary Note 7: Proportions of steppe ancestry on individual level

Relative proportions of ancestry components were estimated using *qpAdm* from ADMIXTOOLS (https://github.com/DReichLab).

Each individual's ancestry was modelled using a three-way model with the reference populations WHG, Anatolia Neolithic (ANF) and Yamnaya Samara (YAM). Subsequently, the best supported minimal model was selected and if necessary, thus for individuals for which a two-way model was better supported with either omitting ANF or WHG, the corresponding two-way model was rerun. In cases were also the two-way model was rejected Yamnaya proportions were set either to 0 or 1. Individuals with less than 100k SNPs were filtered out.

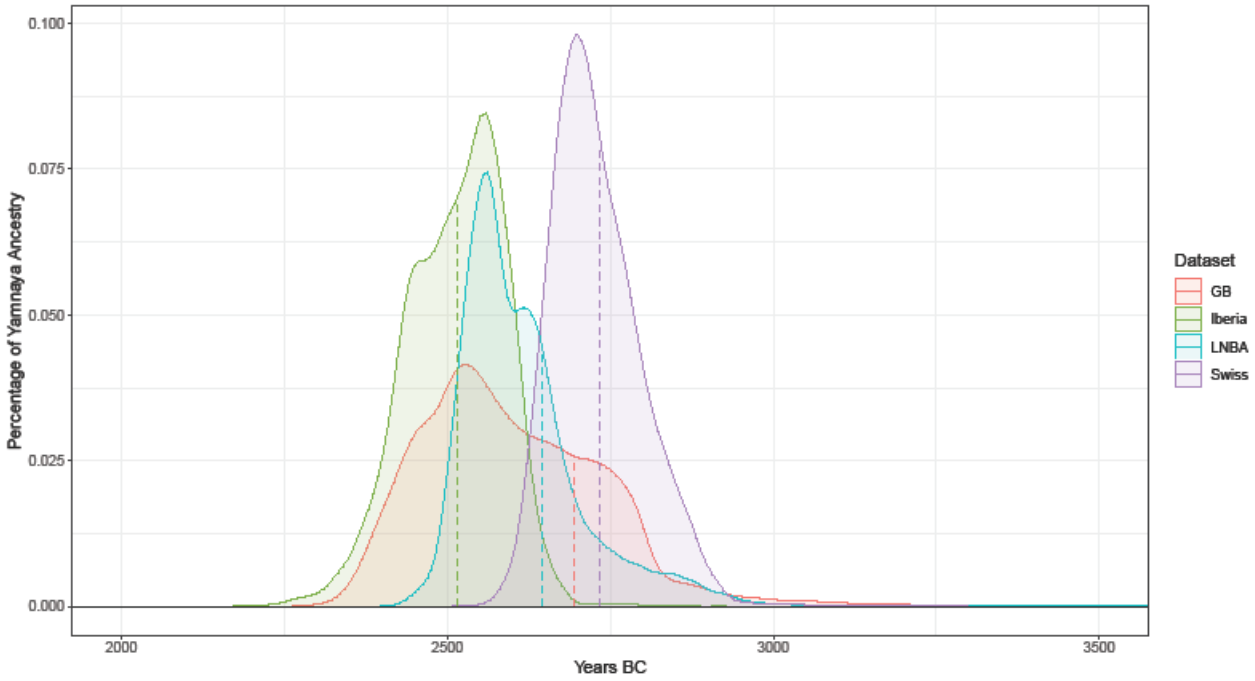
As seen in SI Figure 6 before 2800 BC the steppe component is absent in Central Europe. The earliest individuals with proportions of steppe ancestry greater than zero can be found in the Swiss dataset. However, the ancient Dataset from Great Britain and the MES region in Germany show larger gaps for this specific time span might also biasing the results.

In all regions, an increase of the steppe component around 2500 BC can be seen.



Supplementary Figure 6 | Relative proportions of steppe ancestry in each individual over time. We used *qpAdm* to estimate the proportion of steppe ancestry in each individual. Red dots female individuals and blue dots male individuals. Horizontal error bars represent ranges of C14 dates and vertical error bars represent standard errors of the steppe-related ancestry.

Supplementary Note 8: Estimated arrival time of the steppe component in the different regions



Supplementary Figure 7 | Density plots of the percentage of Yamnaya ancestry in each individual of the regions Great Britain, Iberia, Switzerland, and Germany. The earliest increase can be detected in Switzerland.

Supplementary Table 2 | qpAdm admixture models for the individuals of each site. P-values greater than 0.05 (model is not rejected) marked in green.

Test: WHG Anatolia_Neolithic Yamnaya_Samara
Using the right outgroups: Mbuti Ust_Ishim_HG_published.DG Ethiopia_4500BP.SG MA1_HG.SG Villabruna Papuan Onge Han Karitiana

Population	WHG	Anatolia N	Yamnaya	SD (WHG)	SD (Anatolia N)	SD (Yamnaya)	tail prob (000)	tail prob (001)	tail prob (010)	tail prob (100)
Bell_Beaker_Germany	0,16	0,40	0,44	0,01	0,02	0,02	0,022	0,000	0,000	0,000
Corded_Ware_Germany	0,11	0,22	0,66	0,17	0,04	0,04	0,348	0,000	0,000	0,000
LBK_EN	0,06	0,94	0,00	0,01	0,02	0,00	0,105	0,158	0,000	0,000
Auvernier	0,00	0,79	0,21	0,00	0,14	0,14	0,249	0,212	0,000	0,316
Burgäschisee	0,00	0,70	0,30	0,00	0,15	0,16	0,656	0,355	0,002	0,764
Spreitenbach	0,14	0,38	0,48	0,02	0,03	0,03	0,014	0,000	0,000	0,000
Anselfingen	0,15	0,32	0,54	0,04	0,09	0,09	0,022	0,000	0,000	0,000
Singen	0,19	0,42	0,39	0,02	0,03	0,03	0,126	0,000	0,000	0,000
Oberbipp_F	0,22	0,76	0,02	0,05	0,09	0,09	0,986	0,987	0,000	0,000
Lingolsheim_FN	0,14	0,39	0,47	0,04	0,08	0,08	0,274	0,000	0,000	0,000
Wartau_FN_1	0,17	0,44	0,39	0,02	0,06	0,06	0,621	0,000	0,000	0,000
Wartau_R	0,14	0,59	0,27	0,03	0,07	0,07	0,450	0,005	0,000	0,000
Zurzach	0,18	0,38	0,44	0,03	0,06	0,06	0,719	0,000	0,000	0,000
Lingolsheim_MN	0,20	0,80	0,00	0,02	0,05	0,00	0,209	0,275	0,00	0,000
Muttenz	0,27	0,72	0,00	0,02	0,04	0,00	infeasible	0,227	0,000	infeasible
Niederried	0,22	0,78	0,00	0,02	0,05	0,00	infeasible	0,249	0,000	infeasible
Oberbipp	0,28	0,73	0,00	0,02	0,03	0,00	infeasible	0,123	0,000	infeasible
Seengen	0,26	0,74	0,00	0,05	0,11	0,00	0,304	0,395	0,000	0,000
Rapperswil	0,21	0,79	0,00	0,03	0,06	0,00	infeasible	0,062	0,000	infeasible
Wartau_LN	0,20	0,80	0,00	0,03	0,07	0,00	infeasible	0,579	0,000	0,000
Wartau_FN_2	0,15	0,85	0,00	0,02	0,05	0,00	infeasible	0,200	0,000	infeasible
Aesch	0,26	0,74	0,00	0,01	0,03	0,03	0,135	0,162	0,000	0,000

Supplementary Table 3 | DATES results on individual level with Yamnaya_Samara and Swiss_LN as sources

Sample	Site	sex	maxageBP	minageBP	ageBP	mean generations	generations	gen +/-	maxgen	mingen
MX188	Spreitenbach	m	2495	2399	2447	22,91	19,13	11,29	30,42	7,84
MX190	Spreitenbach	m	2860	2460	2660	495,46	407,65	360,26	767,91	47,39
MX192	Spreitenbach	m	2571	2513	2542	186,36	130,21	75,12	205,32	55,09
MX195	Spreitenbach	m	2470	2050	2260	6,90	6,01	3,33	9,34	2,69
MX197	Spreitenbach	f	2490	2294	2392	14,56	14,89	2,14	17,03	12,75
MX198	Spreitenbach	f	2900	2300	2600	7,81	8,63	3,40	12,03	5,22
MX199	Spreitenbach	f	2409	2197	2303	9,38	6,79	7,41	14,19	-0,62
MX252	Singen	m	1941	1774	1858	111,38	82,75	62,19	144,94	20,56
MX254	Singen	m	2199	1774	1987	50,43	48,08	17,76	65,85	30,32
MX257	Singen	m	1879	1696	1788	54,99	50,92	14,58	65,50	36,34
MX259	Anselfingen	m	2456	2203	2330	10,10	8,54	7,62	16,16	0,92
MX270	Singen	m	2199	1774	1987	30,51	23,85	13,54	37,39	10,32
MX275	Singen	m	2135	1961	2048	39,76	36,18	18,40	54,58	17,78
MX279	Singen	m	1882	1745	1814	35,74	113,66	74,86	188,53	38,80
MX280	Singen	f	2035	1910	1973	35,84	31,70	9,83	41,53	21,87
MX283	Singen	m	2116	1926	2021	18,75	1,77	14,71	16,48	0,00
MX288	Singen	m	2199	2028	2114	68,01	43,70	39,60	83,30	4,10
RA63	Zurzach	m	2046	1946	1996	13,58	12,69	4,15	16,84	8,55
RA64	Zurzach	m	2206	2126	2166	16,54	16,90	4,41	21,31	12,49
SX20	Wartau	m	1693	1609	1651	17,96	15,25	6,00	21,25	9,25
SX23	Wartau	f	1883	1749	1816	16,48	13,18	7,22	20,41	5,96
SX26	Wartau		2461	2295		30,77	28,76	13,385	42,15	15,38
MX189	Spreitenbach	f	2105	2036	2071	243,16	wired number			
MX191	Spreitenbach	m	2570	2190	2380	437,29	wired number			
MX194	Spreitenbach	f	2489	2344	2417	304,99	wired number			
MX196	Spreitenbach	f	2580	2290	2435	5,71	wired number			
MX277	Singen	f	1926	1770	1848	34,16	wired number			
MX256	Singen	f	2132	1922	2027	0,00	wired number			
MX258	Singen	m	2028	1903	1966	28,90	wired number			

MX265	Singen	m	763	431	597	wired number	wired number				
MX298	Wartau	m	2620	2448	2534	wired number	wired number				
MX304	Auvernier	m	2866	2601	2734	208,80	wired number				
MX310	Burgaeschisee	m	2862	2581	2722	135,01	wired number				
MX286	Singen	m	2029	1892	1961	64,10	-271,57	237,27	-34,30	-508,85	
MX251	Singen	f	2197	1981	2089	5,12	-134,78	109,41	-25,38	-244,19	

Supplementary Table 4 | DATES results for different regions

Region	Group	Source1	Source2	C14 range max	C14 range min	generations	gen +/-
Swiss	Spreitenbach_CWC	Yamnaya_Samara	Swiss_LN	2900	2036	10,934	1,78
Swiss	Swiss_EBA	Yamnaya_Samara	Swiss_LN	2456	1609	27,89	3,757
MES	Bell_Beaker_Germany	Yamnaya_Samara	MES_Neolithic	2625	2050	34,817	11,611
Iberia	Bell_Beaker_Iberia	Yamnaya_Samara	Iberia_MN	2800	1776	7,258	7,183
GB	Bell_Beaker_England	Yamnaya_Samara	England_Neolithic	2800	1800	22,114	2,482
MES	Corded_Ware_Germany	Yamnaya_Samara	MES_Neolithic	2578	2033	26,722	9,641

Supplementary Table 5 | Relatedness estimates using PMR, READ and lcMLkin for first and second degree related individuals

		lcMLkin						READ			PMR				
Ind1	Ind2	k0_hat	k1_hat	k2_hat	pi_HAT	nbSNP	r	Relationship	Z_upper	Z_lower	nSNPs	nmismatch	pmismatch	same MT DNA	same Y HG
MX270	MX275	0,315	0,461	0,223	0,454	167647	0,4535	First Degree	7,21	-14,07	196616	38132	0.19394	yes	yes
MX209	MX219	0,231	0,639	0,13	0,45	24866	0,4495	First Degree	8,42	-3,97	29551	5458	0.18470	no	same clade
SX10	SX8	0,13	0,847	0,023	0,446	124180	0,4465	First Degree	9,44	-13,34	145624	28231	0.19386	yes	na
MX195	MX197	0,119	0,873	0,008	0,445	294763	0,4445	First Degree	3,92	-20,07	356136	69688	0.19568	yes	na
MX187	MX212	0,13	0,852	0,018	0,444	175276	0,444	First Degree	8,39	-12,11	208924	39351	0.18835	no	yes
MX192	MX197	0,12	0,873	0,007	0,443	285900	0,4435	First Degree	4,75	-20,44	344641	67734	0.19653	yes	na
MX192	MX195	0,334	0,476	0,19	0,428	268894	0,428	First Degree	1,96	-17,57	323615	63541	0.19635	yes	yes
MX188	MX190	0,382	0,39	0,227	0,422	326168	0,422	First Degree	2,89	-18,45	394714	78180	0.19807	yes	yes
MX254	MX286	0,166	0,828	0,006	0,42	329141	0,42	First Degree	4,23	-23,34	396702	78410	0.19765	no	yes
MX283	MX286	0,185	0,807	0,007	0,411	340960	0,4105	First Degree	5,46	-24,03	411459	81328	0.19766	no	yes
MX275	MX286	0,204	0,784	0,013	0,405	244748	0,405	First Degree	3,41	-23,07	291501	57925	0.19871	no	yes
MX150	MX187	0,248	0,71	0,042	0,397	91284	0,397	First Degree	8,22	-12,18	109655	20743	0.18917	no	same clade
MX254	MX283	0,439	0,422	0,138	0,35	393973	0,349	First Degree	6,13	-16,49	477609	97660	0.20448	yes	yes
MX183	MX211	0,422	0,47	0,108	0,343	258399	0,343	First Degree	0,69	-19,35	310268	62928	0.20282	yes	same clade
MX150	MX209	0,577	0,346	0,077	0,25	72301	0,25	Second Degree	8,62	-0,25	86029	18175	0.21127	no	same clade
MX275	MX283	0,577	0,407	0,016	0,22	286647	0,2195	Second Degree	5,68	-10,40	343024	75468	0.22001	no	yes
MX270	MX286	0,614	0,357	0,028	0,207	189791	0,2065	Second Degree	3,75	-11,33	223865	50077	0.22369	no	yes
MX254	MX275	0,679	0,3	0,021	0,171	277631	0,171	Second Degree	2,92	-13,44	331845	75516	0.22756	no	yes
MX187	MX219	0,754	0,193	0,054	0,15	30547	0,1505	Second Degree	2,92	-3,09	36392	8053	0.22128	no	same clade

MX187	MX209	0,705	0,291	0,003	0,149	181634	0,1485	Second Degree	2,51	-9,68	216996	48874	0.22523	yes	yes
MX150	MX219	0,764	0,223	0,013	0,125	12304	0,1245	Second Degree	1,62	-2,35	14706	3285	0.22338	no	yes
MX150	MX212	0,82	0,121	0,059	0,12	70041	0,1195	Second Degree	1,76	-6,66	83579	18906	0.22621	no	same clade
Aes3	Aes12	0,123	0,863	0,014	0,445	279742	0,4455	First Degree	12,41	-16,25	331487	63428	0.19134	yes	na
Aes3	Aes24	0,583	0,41	0,008	0,213	325204	0,213	Second Degree	7,56	-9,51	386396	84889	0.21969	no	na
Aes12	Aes24	0,143	0,847	0,009	0,433	273774	0,4325	First Degree	7,14	-19,38	324166	62519	0.19286	no	same clade
Aes3	Aes19	0,104	0,891	0,005	0,45	323734	0,4505	First Degree	12,20	-16,34	384537	73727	0.19173	yes	na
Aes19	Aes24	0,577	0,409	0,013	0,218	316956	0,2175	Second Degree	7,21	-9,06	376691	82322	0.21854	no	same clade
Aes11	Aes17	0,106	0,878	0,015	0,454	308832	0,454	First Degree	11,54	-16,95	366976	70202	0.1913	no	na
Aes14	Aes15	0,106	0,878	0,015	0,454	193594	0,454	First Degree	12,84	-14,64	227595	43354	0.19049	yes	na
Aes12	Aes19	0,354	0,437	0,208	0,427	272655	0,4265	First Degree	10,74	-12,73	322847	61942	0.19186	yes	yes
RA63	RA64	0.995	0.005	0	0.002	107385	0.0025	First or Second Degree	5,43	-17.82	68304	15015	0,21983	no	yes

Supplementary Table 6 | D-statistics.

Site	lat	lon	steppe_ancestry	PopA	PopB	F4	Z	ka	ka1	ka2
Spreitenbach	47,255	8,2158	yes	Swiss-French	Swiss-German	0,000114	2,056	6487	6476	98548
Singen	47,767097	8,872239	yes	Swiss-French	Swiss-German	0,000053	0,865	5911	5906	89516
Wartau_FN_1	47,45	9,292	yes	Swiss-French	Swiss-German	-0,000015	-0,196	4775	4776	72386
Wartau_FN_2	47,45	9,292	yes	Swiss-French	Swiss-German	-0,000025	-0,336	4618	4619	70418
Wartau_LN	47,45	9,292	no	Swiss-French	Swiss-German	-0,000009	-0,09	3387	3387	51365
Lingolsheim_FN	48,554457	7,681749	yes	Swiss-French	Swiss-German	0,000082	0,711	2262	2259	34356
Anselfingen	47,854144	8,7737404	yes	Swiss-French	Swiss-German	0,00018	1,371	1754	1750	26337
Zurzach	47,587789	8,293325	yes	Swiss-French	Swiss-German	0,000099	1,207	4199	4193	63804
Burgaeschisee	47,101	7,406	yes	Swiss-French	Swiss-German	-0,000355	-1,425	307	309	4760
Auvernier	46,5835	6,5245	yes	Swiss-French	Swiss-German	0,000193	0,763	330	329	5038
Oberbipp	47,262067	7,659835	no	Swiss-French	Swiss-German	-0,000057	-0,967	6387	6393	96637
Muttenz	47,3122	7,3843	no	Swiss-French	Swiss-German	0,000051	0,767	5401	5397	81812
Seengen	47,1936	8,122	no	Swiss-French	Swiss-German	-0,000435	-2,626	821	826	12666
Niederried	47,0117074	7,2505417	no	Swiss-French	Swiss-German	-0,000198	-2,57	5107	5122	77394
Rapperswil	47,2266239	8,8184374	no	Swiss-French	Swiss-German	0,000079	0,814	2948	2944	44774
Aesch	47,465413	7,601691	no	Swiss-French	Swiss-German	-0,000063	-1,138	6721	6767	102298
Lingolsheim	48,554457	7,681749	no	Swiss-French	Swiss-German	0,000018	0,257	4622	4620	70266
Spreitenbach	47,255	8,2158	yes	Swiss-Italian	Swiss-German	-0,000519	3,573	6516	6465	98548
Singen	47,767097	8,872239	yes	Swiss-Italian	Swiss-German	-0,000534	3,665	5940	5893	89516
Wartau_FN_1	47,45	9,292	yes	Swiss-Italian	Swiss-German	-0,000516	2,736	4798	4760	72386
Wartau_FN_2	47,45	9,292	yes	Swiss-Italian	Swiss-German	-0,000042	0,239	4629	4627	70418
Wartau_LN	47,45	9,292	no	Swiss-Italian	Swiss-German	-0,000233	1,023	3397	3385	51365
Lingolsheim_FN	48,554457	7,681749	yes	Swiss-Italian	Swiss-German	-0,001037	3,866	2280	2244	34356
Anselfingen	47,854144	8,7737404	yes	Swiss-Italian	Swiss-German	-0,000695	2,09	1764	1746	26337
Zurzach	47,587789	8,293325	yes	Swiss-Italian	Swiss-German	-0,000688	3,407	4223	4179	63804
Burgaeschisee	47,101	7,406	yes	Swiss-Italian	Swiss-German	-0,000594	0,93	311	308	4760
Auvernier	46,5835	6,5245	yes	Swiss-Italian	Swiss-German	-0,000898	1,316	334	329	5038
Oberbipp	47,262067	7,659835	no	Swiss-Italian	Swiss-German	0,000046	-0,321	6401	6405	96637

Muttenz	47,3122	7,3843	no	Swiss-Italian	Swiss-German	-0,000092	0,561	5413	5405	81812
Seengen	47,1936	8,122	no	Swiss-Italian	Swiss-German	-0,000069	0,169	826	825	12666
Niederried	47,0117074	7,2505417	no	Swiss-Italian	Swiss-German	0,000068	-0,371	5121	5126	77394
Rapperswil	47,2266239	8,8184374	no	Swiss-Italian	Swiss-German	-0,000389	1,64	2959	2942	44774
Aesch	47,465413	7,601691	no	Swiss-Italian	Swiss-German	-0,00003	0,026	6738	6738	101710
Lingolsheim	48,554457	7,681749	no	Swiss-Italian	Swiss-German	0,000099	-0,534	4627	4633	70266
Spreitenbach	47,255	8,2158	yes	Swiss-French	Swiss-Italian	-0,000405	-2,978	6473	6513	98548
Singen	47,767097	8,872239	yes	Swiss-French	Swiss-Italian	-0,000481	-3,431	5897	5940	89516
Wartau_FN_1	47,45	9,292	yes	Swiss-French	Swiss-Italian	-0,000531	-3,006	4762	4801	72386
Wartau_FN_2	47,45	9,292	yes	Swiss-French	Swiss-Italian	-0,000067	-0,399	4629	4633	70418
Wartau_LN	47,45	9,292	no	Swiss-French	Swiss-Italian	-0,000241	-1,126	3386	3399	51365
Lingolsheim_FN	48,554457	7,681749	yes	Swiss-French	Swiss-Italian	-0,000956	-3,897	2246	2279	34356
Anselfingen	47,854144	8,7737404	yes	Swiss-French	Swiss-Italian	-0,000515	-1,649	1749	1762	26337
Zurzach	47,587789	8,293325	yes	Swiss-French	Swiss-Italian	-0,000589	-3,068	4183	4221	63804
Burgaeschisee	47,101	7,406	yes	Swiss-French	Swiss-Italian	-0,000949	-1,648	307	311	4760
Auvernier	46,5835	6,5245	yes	Swiss-French	Swiss-Italian	-0,000705	-1,099	329	333	5038
Oberbipp	47,262067	7,659835	no	Swiss-French	Swiss-Italian	-0,000011	-0,079	6406	6407	96637
Muttenz	47,3122	7,3843	no	Swiss-French	Swiss-Italian	-0,000041	-0,255	5409	5412	81812
Seengen	47,1936	8,122	no	Swiss-French	Swiss-Italian	-0,000504	-1,273	823	829	12666
Niederried	47,0117074	7,2505417	no	Swiss-French	Swiss-Italian	-0,000131	-0,75	5123	5133	77394
Rapperswil	47,2266239	8,8184374	no	Swiss-French	Swiss-Italian	-0,00031	-1,389	2946	2959	44774
Aesch	47,465413	7,601691	no	Swiss-French	Swiss-Italian	-0,000066	-0,52	6738	6745	101710
Lingolsheim	48,554457	7,681749	no	Swiss-French	Swiss-Italian	0,000117	0,64	4636	4628	70266

References

1. Tschumi, O. Die steinzeitlichen Hockergräber der Schweiz. Anzeiger für schweizerische Altertumskunde : Neue Folge Indicateur d'antiquités suisses : Nouvelle série 22 (1920)
2. Hug E. 1956. Die Anthropologische Sammlung im Naturhistorischen Museum Bern. In: Bern NM, editor. Bern: Naturhistorisches Museum Bern.
3. Schlaginhaufen O. 1924b. Das Hockerskelett von Ursisbalm bei Niederried (Kt. Bern). Jahrbuch des Bernischen Historischen Museums in Bern:96-109.
4. Lichardus-Itten, M. Die Gräberfelder der Grossgartacher Gruppe im Elsaß (Habelt, Bonn, 1980)
5. Ramstein, M., Schimmelpfennig, D., Lösch, S. Ein neolithischer Dolmen an der Steingasse in Oberbipp. Archäologie Schweiz 37(3), 4-15 (2014).
6. Siebke, I. et al. The application of different 3D-scan-systems and photogrammetry at an excavation — A Neolithic dolmen from Switzerland. Digital Applications in Archaeology and Cultural Heritage 10, e00078; 10.1016/j.daach.2018.e00078 (2018)
7. Siebke I, Steuri N, Furtwängler A, Ramstein M, Arenz G, Hafner A, Krause J, and Lösch S. 2019. Who lived on the Swiss Plateau around 3300 BCE? - Analyses of commingled Human Skeletal Remains from the Dolmen of Oberbipp. International Journal of Osteoarchaeology. DOI: 10.1002/oa.2791
8. Bleuer E, Doppler T, Fetz H.. Gräber im näheren und weiteren Umfeld von Spreitenbach in Spreitenbach-Moosweg (Aargau, Schweiz): Ein Kollektivgrab um 2500 v Chr (ed. Doppler, T.) 233-266 (Basel 2012)
9. Löhlein W. Monumentale Grabanlagen der Steinzeit Megalithische Denkmale der Hochrheinregion. Denkmalpflege in Baden-Württemberg 3 2011
10. Stöckli WE, Niffeler U, and Gross-Klee E, editors. 1995. Neolithikum - Néolithique - Neolitico. SPM II. Basel. 385 p.
11. Schwegler U. 2016. Chronologie und Regionalität neolithischer Kollektivgräber in Europa und in der Schweiz Hochwald (Schweiz): LIBRUM Publishers & Editors LLC.
12. Bay R. 1936. Kiefer und Zähne aus dem neolithischen Steinkistengrab bei Aesch. Tätigkeitsbericht der Naturforschenden Gesellschaft Baselland 11 (1936-1938):22-30.
13. Doppler, T. Spreitenbach-Moosweg: ein Kollektivgrab um 2500 v. Chr. Archäologie Schweiz, Vol. 51 (2012).
14. Doswald C. Kaufmann B. Scheidegger S. Ein neolithisches Doppelhockergrab in Zurzach. Archäologie der Schweiz 12 (1989)
15. Grüninger I. and Kaufmann B. Ein Steinkistengrab von Rapperswil SG. Archäologie der Schweiz 5 (1982)
16. Stehrenberger T. Obere Höhle bei der "Procha Burg" Höhlenpost 148 (2016).
17. Stehrenberger T, Siebke I, and Lösch S. 2019. Sekundär bestattet oder «entsorgt»? Die Toten aus der Oberen Höhle Procha Burg (Wartau, Gretschins SG). In: Schweizerische Gesellschaft für Höhlenforschung, editor. Sinterlaken - Nationaler Kongress für Höhlenforschung Interlaken: accepted.

18. Stockhammer, P. W., Massy, K., Knipper, C., Friedrich, R., Kromer, B., Lindauer, S., et al. Rewriting the central European early bronze age chronology: Evidence from large-scale radiocarbon dating. PLoS One, 10(10), e0139705 (2015)
19. OELZE, V. M., NEHLICH, O. & RICHARDS, M. P. 'THERE'S NO PLACE LIKE HOME'-NO ISOTOPIC EVIDENCE FOR MOBILITY AT THE EARLY BRONZE AGE CEMETERY OF SINGEN, GERMANY. Archaeometry 54, 752–778; 10.1111/j.1475-4754.2011.00644.x (2012).
20. R. Krause, Die endneolithischen und frühbronzezeitlichen Grabfunde auf der Nordstadtterrasse von Singen am Hohentwiel. Forschungen und Berichte zur Vor- und Frühgeschichte in Baden-Württemberg 32 (Stuttgart 1988).
21. Matthias B. Merkl CORDED WARE, BELL BEAKERS AND THE EARLIEST BRONZE AGE IN THE HEGAU AND THE WESTERN LAKE CONSTANCE REGION MUSAICA ARCHAEOLOGICA 1 (2016)
22. Ulrich-Bochsler S, Cooper C, Meyer L, Rüttimann D, and Lanz C. 2012. Anthropologische Bearbeitung der Skelettreste vom Burgäschisee. In: Wey O, editor. Die Cortaillod-Kultur am Burgäschisee Materialvorlage und Synthese zu den neolithischen Fundkomplexen von Burgäschisee-Ost, -Südwest, -Süd und -Nord. Bern: Stämpfli Verlag AG. p 111-154.
23. Schlaginhaufen O. 1924a. Die anthropologischen Funde aus den Pfahlbauten der Schweiz. Pfahlbauten, zehnter Bericht. Mitteilungen der Anthropologischen Gesellschaft in Zürich 29(4):220-241.
24. Wiedmer-Stern J. 1904. Archäologie aus dem Oberaargau. Archiv des Historischen Vereins des Kantons Bern 17 (1903-1904)(2).
25. Steuri, N. et al. MULTIPLE RADIOCARBON DATING OF HUMAN REMAINS: CLARIFYING THE CHRONOLOGY AND SEQUENCES OF BURIALS IN THE LATE NEOLITHIC DOLMEN OF OBERBIPP (SWITZERLAND). Radiocarbon II, 1–13; 10.1017/RDC.2019.96 (2019).
26. Siebke I. et al. Crops vs. Animals: Regional Differences in Subsistence Strategies of Swiss Neolithic Farmers revealed by Stable Isotopes. *Submitted* (2019).
27. Furtwängler, Anja, et al. "Ratio of mitochondrial to nuclear DNA affects contamination estimates in ancient DNA analysis." *Scientific reports* **8.1** (2018): 14075.
28. Peltzer, Alexander, et al. "EAGER: efficient ancient genome reconstruction." *Genome biology* **17.1** (2016): 60.
29. Renaud, Gabriel, et al. "Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA." *Genome biology* **16.1** (2015): 224.
30. Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Research* **44**, W58-63 (2016).
31. Saag, L. *et al.* Extensive Farming in Estonia Started through a Sex-Biased Migration from the Steppe. *Current biology : CB* **27,** 2185-2193.e6; 10.1016/j.cub.2017.06.022 (2017).