

Applications of DNA capture in ancient DNA research

Dissertation

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

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

von Alexander Immel

geboren am 22.03.1986 in Bischkek

Gutachter

1: Prof. Dr. Johannes Krause

2: Prof. Dr. Martin S. Fischer

3: Prof. Dr. Oliver Kohlbacher

Tag der Verteidigung: 30.09.2020

Contents

Glossary	II
Summary	IV
Zusammenfassung	VI
List of Figures	VII
List of Tables	VIII
1 ancient DNA research - History and Methods	1
1.1 A Short History of aDNA Research	1
1.2 aDNA Retrieval from Old Bone	2
1.3 DNA Libraries	3
1.4 Next Generation Sequencing	4
2 Applications of aDNA Capture	6
2.1 Current Methods for Target DNA Enrichment	6
2.2 Study 1: Capture of Giant Deer Mitochondrial DNA	7
2.2.1 Archaeological Material	8
2.2.2 Extraction of Giant Deer DNA and Library Preparation	8
2.2.3 Bait Preparation and Capture	8
2.2.4 Sequence Processing and Assembly	9
2.2.5 Multiple Sequence Alignment	11
2.2.6 Phylogenetic Analyses	11
Publication: <i>Mitochondrial Genomes of Giant Deers Suggest their Late Survival in Central Europe</i>	16
2.3 Study 2: Effect of X-rays on ancient DNA	25
2.3.1 Background and Previous Work	25

2.3.2	Investigating X-ray effects on aDNA after Mitochondrial DNA Capture	26
	Publication: <i>Effect of X-ray irradiation on ancient DNA in sub-fossil bones - Guidelines for safe X-ray imaging</i>	27
2.4	Study 3: Immunity Capture	67
2.4.1	Background: Plague, HLA and other Immunity Genes	67
2.4.2	Initial Array Design for Immunity Capture	70
2.4.3	Improved Array Design	72
2.4.4	In-Solution Capture	73
2.4.5	Evaluation of the In-Solution Capture Efficiency	75
2.4.6	Application of the Immunity Capture to 16 th Cent. Plague Victims	75
2.4.7	HLA Typing and Allele Frequency Comparisons	87
2.4.8	CCR5- Δ 32 Mutation in Plague Victims and Contemporary Ellwangen	89
3	Discussion	93
3.1	Giant Deer and Megafauna	93
3.2	X-rays and aDNA	94
3.3	In-Solution Immunity Capture	95
3.4	Stability of the CCR5- Δ 32 Mutation	96
3.5	Frequency Shifts in particular HLA alleles	97
4	Conclusion	99
	Appendix and Supplementary Material	100
	Literature	113
	Veröffentlichungen und Vorträge	134
	Ehrenwörtliche Erklärung	137
	Prozentuale Anteile der eigens erbrachten Leistung	138
	Danksagung	142
	Tabellarischer Lebenslauf	143

Glossary

Abbreviation. Explanation

AD	Anno Domini
aDNA	ancient desoxyribonucleic acid
AIDS	acquired immunodeficiency syndrome
BI	Bayesian Inference
BCE	before Common Era
bp	base pairs
BP	before present
ca.	circa
cal.	calibrated
capture	targeted enrichment of genomic regions of interest using molecular probes
CT	Computed Tomography
cytb	cytochrome b
°C	degree Celsius
DKMS	Deutsche Knochenmarkspenderdatei
df	degrees of freedom
e.g.	example given
fig.	figure
Gy	Gray (SI-unit)
h	hours
HIV	Human Immunodeficiency Virus
HLA	Human Leukocyte Antigen
HTS	High Throughput Sequencing
ibid.	<i>ibidem</i> (lat.) - at the same place
i.e.	<i>id est</i> (lat.) - that is
IGV	Integrative Genomics Viewer
kb	kilo bases
kg	kilogram

Abbreviation.	Explanation
ky	kilo (thousand) years
kya	kilo years ago
kyo	kilo years old
LGM	Last Glacial Maximum (24.5 ky - 18 ky BP)
LPS	lipopolysaccharide
m	meter
mb	mega bases (of DNA)
MCMC	Markov Chain Monte Carlo
MHC	Major Histocompatibility Locus
min	minutes
mio	million
μ l	micro liter
ML	Maximum Likelihood
MP	Maximum Parsimony
MSA	Multiple Sequence Alignment
mtDNA	mitochondrial DNA
mt capture	mitochondrial DNA capture
NK cells	natural killer cells
NGS	Next Generation Sequencing
ntDNA	nuclear DNA
nt	nucleotides
OTU	operational taxonomic unit
PAMP	Pathogen-associated Molecular Patterns
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PRR	Pattern Recognition Receptor
qPCR	quantitative PCR
read	sequence of DNA bases
resp.	respectively
sec.	section
shotgun sequencing	untargeted sequencing of genome-wide DNA
SNP	Single Nucleotide Polymorphism
tab.	table
UDG	Uracil-DNA-Glycosylase
UDG-half	partial Uracil-DNA-Glycosylase treatment
VCF	Variant Call Format
WGS	Whole Genome Sequencing
WHO	World Health Organization
wt	wild type
ya	years ago

Summary

Recent advances in technology and laboratory methods have revolutionized the research field of ancient DNA (aDNA) and significantly extended the boundaries of genetic investigations into the past. Particularly, Next Generation Sequencing (NGS) and targeted DNA enrichment methods allowed ancient DNA to become an inevitable source of fundamental research. Here, I present three studies which comprise diverse objectives of aDNA research by making use of target DNA enrichment and NGS.

In my first study I bioinformatically reconstruct the complete mitochondrial genome of the extinct giant deer after the capture of its mitochondrial DNA (mt capture). Using the reconstructed mitochondrial genome I then apply different methods of phylogenetic analysis which allow to resolve the giant deer's placement within the cervid family tree and reaffirm the fallow deer being its closest extant relative. Mt capture also provides the basis for another study, in which I investigate the effects of X-rays on aDNA. Here, different radiation settings are explored including those ones commonly used in computed tomography (CT) scans of anthropological and palaeontological specimens in order to study potential effects of CT scanning on aDNA preservation. The results allow to define a threshold of 200 Gray (Gy) of maximum absorbed radiation dose wherein no X-ray caused effect can be observed on aDNA, and to specify simple guidelines to avoid impairment of aDNA studies through CT-scans. In my third study I have developed and evaluated a hybridization-capture approach for the enrichment of around 500 human immune system genes, and apply it to compare the immunogenetic makeup of 16th century plague victims from Ellwangen, South Germany, to their contemporary successors. Applying bioinformatic tools allows me to assess kinship within both populations, determine genetic continuity between past and present, and reconstruct the Human Leukocyte Antigen (HLA) allelic profiles of both populations. Comparing frequency distributions of HLA alleles does not show significant differences between both populations, although single HLA-alleles show clearly distinct frequencies, potentially indicative of selective pressure exerted through plague. In summary, specific target enrichment in combination with NGS and the application of particular bioinformatic methods has provided

me with tools to address research questions related to evolutionary biology and aDNA preservation.

Zusammenfassung

Technologischer Fortschritt sowie die Weiterentwicklung von Labormethoden haben das Forschungsfeld der alten DNA (aDNA) revolutioniert. Die aDNA-Forschung ist Dank der Einführung von Next Generation Sequenzier-Technologien (NGS) und Methoden gezielter DNA-Anreicherung zum bedeutenden Bestandteil der Grundlagenforschung geworden. In dieser Arbeit präsentiere ich drei Studien, die unterschiedliche Fragestellungen mithilfe von DNA-Anreicherungsmethoden und NGS zu beantworten suchen.

In meiner ersten Studie gelingt es mir das mitochondriale Genom des Riesenhirschs bioinformatisch zu rekonstruieren, nachdem seine mitochondriale DNA (mtDNA) angereichert werden konnte. Ich verwende anschließend verschiedene Methoden der phylogenetischen Rekonstruktion, um die Positionierung des Riesenhirschs im Stammbaum der Hirsche eindeutig zu identifizieren. Durch diese Analysen kann der Damhirsch als der letzte lebende Verwandte des Riesenhirschs verifiziert werden. Anreicherung von mtDNA ist zugleich die Grundlage für eine weitere Studie, in der ich den Effekt von Röntgenstrahlung auf aDNA untersuche. In dieser Studie werden verschiedene Röntgen-Einstellungen getestet, darunter auch diejenigen, die für die computertomographische (CT) Erfassung von anthropologischen und paläontologischen Befunden verwendet werden. Die Ergebnisse ermöglichen es eine Obergrenze für absorbierte Röntgenstrahlung festzulegen, im Rahmen derer kein Effekt auf aDNA beobachtet werden kann. Diese Obergrenze beträgt 200 Gray (Gy). Weiterhin lassen sich einfache Grundregeln bestimmen, die eine mögliche Beeinträchtigung von aDNA Studien durch vorangegangene CT-Scans verhindern sollen. In meiner dritten Studie habe ich einen Ansatz zur Anreicherung von knapp 500 menschlichen Immunsystem-Genen entwickelt und ausgewertet, den ich anschließend anwende, um die immunogenetische Veranlagung von Pesttoten aus dem 16. Jahrhundert aus Ellwangen, Süddeutschland, mit ihren heutigen Nachfahren zu vergleichen. Durch Anwendung bioinformatischer Programme gelingt es mir Verwandtschaftsverhältnisse innerhalb beider Populationen, sowie die genetische Kontinuität zwischen damals und heute festzustellen, als auch die Allelfrequenzen des humanen Leukozytenantigen-Systems (HLA) beider Populationen zu rekonstruieren. Der Vergle-

ich der Häufigkeitsverteilungen von HLA-Allelen zeigt keine signifikanten Unterschiede, obwohl einzelne HLA-Allele in deutlich unterschiedlichen Häufigkeiten auftreten. Diese Unterschiede könnten auf selektiven Evolutionsdruck hinweisen, der durch die Pest ausgeübt wurde.

Zusammenfassend war es mir durch gezielte DNA-Anreicherung in Kombination mit NGS und der Anwendung spezieller bioinformatischer Programme möglich Antworten auf diverse Fragen der aDNA Forschung zu finden.

List of Figures

2.1	Skeletal remains of the giant deer	9
2.2	Mitochondrial DNA capture	10
2.3	Bioinformatic analysis of giant deer sequencing data	10
2.4	Dichotomy and polytomy	13
2.5	Nucleotide substitution models	14
2.6	Giant deer phylogeny	15
2.7	Immunity capture array	71
2.8	Tiling	73
2.9	Sequence coverage of 3 captured immunity genes	74
2.10	Ellwangen mass graves	77
2.11	Sampling the petrous bone	78
2.12	Target DNA enrichment after in-solution capture	80
2.13	aDNA damage in the ancient Ellwangen samples	81
2.14	Ellwangen individuals projected on top of a PCA of 29 West Eurasian populations	83
2.15	Ellwangen individuals projected on top of a PCA of 67 West Eurasian populations	84
2.16	Ellwangen individuals highlighted separately and projected on top of a PCA of 67 West Eurasian populations	85
2.17	Admixture proportions of 67 West Eurasian and the Ellwangen populations	86
2.18	Kinship estimates of the Ellwangen plague victims	87
2.19	Estimation of the observable effect size	88
2.20	Coverage of the CCR5- Δ 32 region	92
S1	HLA allele frequencies in both Ellwangen populations	111
S2	Frequencies of 30 most common HLA alleles in both Ellwangen populations and German populations (DKMS)	112

List of Tables

2.1	Capture efficiency for 10 positive control samples	76
2.2	Radiocarbon dates of Ellwangen plague victims	78
2.3	Significant HLA-allele frequency differences between Ellwangen plague victims and German populations	90
2.4	CCR5- Δ 32 genotype and allele frequencies among both Ellwangen populations	91

1 ancient DNA research - History and Methods

1.1 A Short History of aDNA Research

Since the first successful retrieval of aDNA (ancient deoxyribonucleic acid) molecules from the extinct South African quagga (Higuchi et al., 1984) as well as from ancient Egyptian mummies in the 1980's (Pääbo, 1985), ancient DNA research has experienced a significant progress. While these first studies used bacterial cloning to amplify small fragments of DNA mainly consisting of molecules with a high copy number, e.g. mitochondrial DNA (mtDNA) with the later Polymerase Chain Reaction (PCR) it became possible to amplify even single copy DNA molecules to a high extent (Pääbo, 1989, Pääbo et al., 1989). However, results of the PCR studies conducted in the early 1990's claiming to present DNA from organisms of the Mesozoic Era and the Miocene, appeared to be modern contaminants (Golenberg et al., 1990, Cano et al., 1993, Woodward et al., 1994).

Apart from contamination, working with aDNA brings along several other challenges. The extracted aDNA constitutes a mixture of endogenous organismic and exogenous environmental DNA such as from bacteria and fungi (Shapiro and Hofreiter, 2014), meaning that the obtainable endogenous DNA is present only in minute amounts. Since no DNA repair takes place in dead organisms, DNA begins to accumulate different forms of damage. Oxidative stress leads to the cleavage of the phosphodiester backbone or the sugar backbone of DNA resulting in strand breaks (Rizzi et al., 2012). Moreover, oxidative damage can lead to depurination resulting in an abasic (base-less) site. Besides oxygen, factors such as environmental salt and pH, water, radiation and above all temperature play an important role in the preservation of aDNA molecules (Lindahl, 1993). Maillard reaction triggered by higher temperatures and alkylation can cause DNA-strand crosslinks and thus prevent the amplification of endogenous DNA in PCR (Rizzi et al., 2012). Therefore, samples preserved in cold conditions such as permafrost soils (Shapiro

et al., 2004) are more likely to provide amplifiable DNA than samples preserved at warm temperatures (Smith et al., 2001).

Nucleases from environmental microorganisms keep degrading the endogenous DNA through strand breaks leading to short DNA fragments with an average fragment length of 50 base pairs (bp) (Green et al., 2008). And last but not least, damage through hydrolysis results in the deamination of cytosine to uracil, which will be misread as a thymine by DNA polymerases during amplification and result in the incorporation of a wrong base (Hofreiter et al., 2001). PCR-based studies were prone to contamination and false positive results. Therefore it became necessary to implement quality control standards in order to provide reliability of the results from upcoming PCR studies. The criteria that need to be fulfilled in order to provide reliable aDNA work involve a physically isolated work area where no sequencing or amplification takes place, positive and negative control reactions, reproducibility of the results and independent replication in different laboratories, as well as quantitative assessment of the target DNA copy number (Cooper and Poinar, 2000). Meanwhile, PCR-based methods of aDNA analysis have been largely replaced by the Next Generation Sequencing (NGS) methods, which imply new criteria of quality control. These criteria include the presence of taphonomic markers in aDNA such as hydrolytic deamination of cytosines to thymines (Briggs et al., 2007); a high number of unique molecules in an aDNA extract, a high percent of sequencing reads matching to a reference organism's DNA sequence, an even read distribution and a low edit distance towards the reference DNA sequence, and last but not least positive and negative sequencing controls (Key et al., 2017).

1.2 aDNA Retrieval from Old Bone

The preservation of DNA mainly depends on the feasibility of its surrounding material to withstand environmental stress. Since soft organic tissue decays unless being mummified, teeth and bones constitute a natural reservoir to preserve DNA to a much higher extent than soil. While teeth consist of dentin that is surrounded by enamel and protects the dental pulp, bones are mainly composed of a bony matrix that consists of an inorganic and an organic part. Both, teeth and bones contain two major components - hydroxyapatite and collagen.

DNA was shown to have a high binding affinity to hydroxyapatite (Lindahl, 1993) because it forms a complex with it (Brundin et al., 2013), and so, higher DNA fractions could be recovered from hydroxyapatite than from collagen (Schwarz et al., 2009, Cam-

pos et al., 2012). On the other hand DNA can bind to collagen and provide a matrix for the aggregation of collagen molecules into fibrils which induces the fibrillogenesis of collagen (Kitamura et al., 1997). Therefore, aDNA extraction protocols for teeth and bones must address both - the mineral part and the organic part. EDTA (ethylenediaminetetraacetic acid) is usually used for the demineralization of bones and teeth since it sequesters calcium and iron ions (Auld, 1995). Proteinase K can be used to digest proteins and cleave crosslinks between DNA and peptides (Rohland and Hofreiter, 2007). Adding phenol-chloroform followed by centrifugation will provide two distinct phases - a solid phase with proteins, lipids as well as other tissues and an aquatic phase containing nucleic acids and salts (Leonard et al., 2000). However, DNA can be separated even more efficiently from other solved components by binding to a silica membrane. Chaotropic salts such as guanidinium thiocyanate (Rohland and Hofreiter, 2007) or guanidine hydrochloride (Dabney et al., 2013) destabilize hydrogen bonds and Van-der-Waal forces and allow positively charged ions to form a salt bridge between a negatively charged silica membrane (e.g. in *MinElute* spin columns, *Qiagen*) and the negatively charged backbone of DNA. Bound DNA can be precipitated and purified from salts, proteins and other components using alcohol such as ethanol or isopropanol, and finally released from the silica membrane by water or elution buffers such as Tris-EDTA-Tween (TET) (Dabney et al., 2013).

The best outgoing material for aDNA extraction from bones and teeth is fine powder produced by grinding (Rohland and Hofreiter, 2007). Before sampling, it is recommended to decontaminate the surface of the bone (or tooth, resp.) from contaminating environmental and human DNA. This can be achieved by either applying bleach (sodium hypochlorite - NaOCl) to the surface, which will destroy DNA-molecules (Kemp and Smith, 2005) or by using UV-irradiation to crosslink superficial DNA-molecules, which will produce thymine dimers that inhibit polymerases and prevent DNA from amplification and sequencing (Champlot et al., 2010).

1.3 DNA Libraries

After extraction and purification aDNA has to be prepared for sequencing. A proof reading polymerase is used to repair the overhanging 5'- and 3'-ends of the double stranded DNA molecules by either filling in a 5'-overhang or by removing a 3'-overhang (Briggs et al., 2010). After this blunt end repair two sequencing adapter molecules P7 and P5 are ligated to the ends of the DNA molecules and nicks in DNA strands are

repaired in a fill-in reaction using a polymerase with strand-displacement activity (Meyer and Kircher, 2010). Finally, two DNA molecules with known nucleotide sequences are added by amplification to the adapters. The combination of these nucleotides is unique for every individual sample allowing to pool (multiplex) all samples and separate them bioinformatically at a later stage (Kircher et al., 2012). Optionally, before blunt end repair, terminal deamination damage can be removed completely (Briggs et al., 2010) or partially (Rohland et al., 2015) using uracil-DNA-glycosylase (UDG), an enzyme that removes uracil residues leaving abasic DNA sites which then can be cleaved by endonuclease VIII activity. Instead of double stranded sequencing libraries it is also possible to produce single stranded libraries, which have been demonstrated to provide higher DNA complexity (Gansauge and Meyer, 2013) and access to shorter molecules (Gutaker et al., 2017) compared to double-stranded DNA libraries.

1.4 Next Generation Sequencing

DNA-Sequencing is the determination of the sequence of nucleotides in a DNA molecule. The first sequencing methods were based on chemical modification of DNA and cleavage at specific sites (Maxam and Gilbert, 1977) or incorporation of modified nucleotides leading to termination of strand synthesis (Sanger et al., 1977), followed by separation of the resulting DNA molecules through gel electrophoresis. These methods could generate only a few thousands of sequences, so called *reads*, in a week and required a lot of manual work. Next Generation Sequencing (NGS) denominates technologies such as pyrosequencing (Margulies et al., 2005), sequencing by synthesis (Bennett et al., 2005), sequencing by ligation (Pandey et al., 2008), sequencing by hybridization, and ion semiconductor sequencing (Rusk, 2011). These technologies can generate up to billions of reads in a week and therefore are also referred to as High Throughput Sequencing (HTS). In this work sequencing shall hereafter refer to *Illumina* sequencing by synthesis using the sequencing platforms HiSeq 2500[™], NextSeq 500[™], HiSeq 3000 and HiSeq 4000[™]. These methods are based on the detection of base-specific light signals that are emitted after a fluorophore-labeled nucleotide becomes incorporated during the synthesis of a complementary DNA-molecule, and provide large amounts of high quality reads at relatively low sequencing costs (Illumina, 2016).

The rapid development of HTS technologies in the recent years has allowed to reach an unprecedented resolution in aDNA research. A significant role in this development was played by the sequencing costs. In coherence with Moore's law, which is an observation

on computer hardware made by Gordon Moore in 1965 describing the doubling of computational power every two years (Moore, 1965), the sequencing costs per mega base of DNA sequence had been linearly decreasing until the takeover of Next Generation Sequencing (NGS) methods in 2005 (Wetterstrand, 2016). Since then sequencing costs per (human) genome decrease by an order of magnitude every year, and nowadays whole genome sequencing (WGS) costs 1000 US \$ on average (NIH, 2019).

2 Applications of aDNA Capture

2.1 Current Methods for Target DNA Enrichment

While decreasing sequencing costs being a positive trend, another approach has become very popular: targeted DNA capture is a cost-efficient way to obtain a high sequence coverage (i.e. high representation by sequencing reads) of genetic regions of interest, such as specific markers or single nucleotide polymorphisms (SNPs). This is of special interest to aDNA research, since aDNA molecules are often badly preserved and present in minimal amounts, which makes shotgun sequencing (i.e. untargeted sequencing of the whole genome eventually including environmental DNA) very cost-intensive, if a sufficient genomic coverage has to be reached.

Capture by hybridization is based on the principle of the complementarity of DNA. In contrast to Polymerase Chain Reaction (PCR), which uses specific primers to amplify the region of interest, it makes use of nucleic acid sequences (probes) complementary to the genomic regions of interest, which hybridize to these probes and become sequenced, while fragments that did not hybridize are discarded (Peñalba et al., 2014). To date, two approaches are common for the enrichment of regions of interest (target regions): solid capture and in-solution capture. Solid capture is based on microarrays with clusters of nucleic acid probes attached to their surfaces (Hodges et al., 2009), while these probes can also be in solution (Gnirke et al., 2009, Maricic et al., 2010, Mason et al., 2011, Li et al., 2013). Capture by hybridization has proven itself efficient in target enrichment of modern and archaic human mitochondrial DNA (mtDNA) (Maricic et al., 2010, Krause et al., 2010, Briggs et al., 2009) and nuclear DNA (ntDNA) (Bi et al., 2012, Haak et al., 2015, Fu et al., 2016), as well as in enrichment of late Pleistocene megafaunal mtDNA (Dabney et al., 2013, Enk et al., 2016, Soubrier et al., 2016).

2.2 Study 1: Capture of Giant Deer Mitochondrial DNA

As a frequent case of targeted enrichment, mitochondrial DNA (mtDNA) capture has become a widely used approach in the last eight years. Mitochondria are cell organelles that are involved into cellular energy supply, cell cycle, cell growth, signalling and cell death (McBride et al., 2006), and contain their own DNA. The mtDNA is circular and present in 10^2 to 10^4 copies per cell (Furtwaengler et al., 2018), whereas nuclear DNA (ntDNA) is only present in two copies in somatic cells. The higher abundance of mtDNA over ntDNA makes it attractive for genetic analyses especially in aDNA research. Compared to ntDNA capture, mtDNA capture is a relatively economical and simple approach. Following the protocol after Maricic *et al.* 2010, a long-range PCR is used to generate two long-range PCR products that encompass the complete mitochondrial genome. These products are sheared into 150-850 base pair (bp) fragments, that then get biotinylated and purified, before being bound to streptavidin-coated magnetic beads (such as *SPRIselect*) (Maricic et al., 2010). The so generated bait are being pooled together with aDNA-sequencing libraries for hybridization of complementary DNA molecules to the bait. Molecules, that have not hybridized after 48 hours (h) of incubation, are washed away, thereby allowing for an enriched concentration of target molecules for sequencing. Using mitochondrial DNA capture we were able to capture and reconstruct two almost complete mitochondrial genomes of the extinct giant deer *Megaloceros giganteus* (also erroneously misnamed the "Irish Elk", (Gould, 1977)) from 12.500 years old (12.5 kyo) skeletal remains. The giant deer was believed to have gone extinct after the Last Glacial Maximum (LGM, ca. 20.000 - 12.500 years ago (kya)) in Central and Western Europe, although the last members of its populations survived to around 7.7 kya in Western Siberia (Stuart et al., 2004). Furthermore, a debated question concerned the phylogenetic placement of the giant deer - a closer relatedness to the red deer *Cervus elaphus* was proposed based on post-cranial skeletal elements (Pfeiffer, 2002), whereas large palmate antlers suggested a closer genetic affinity to the fallow deer *Dama dama* (Freudenberg, 1914, Lister, 1984). Earlier analyses of cytochrome b and regions of the mitochondrial D-loop provided evidence for a closer genetic relatedness to the fallow deer although the statistical support was rather low (Lister et al., 2005, Hughes et al., 2006). However, complete mitochondrial giant deer genomes were missing, which would have provided enough information to finally resolve the giant deer's positioning in the cervid family tree.

2.2.1 Archaeological Material

When bone fragments from large cervids were discovered in the Swabian Alb during an excavation at the Hohlenstein Stadel site (ST) and the Hohlefels site (HF) in Baden-Wuerttemberg (Southwestern Germany) in 2008, they were assumed to belong to Eurasian elk (*Alces alces*). Among the skeletal remains a piece of an atlas, two fragments from a scapula, two pelvic fragments, two ribs, a molar, a skull fragment, a fragmented tibia and a metatarsus fragment were discovered (fig. 2.1 A). However, some morphological traits appeared not completely consistent with elk, but the limited resolution provided by the morphology of the sparse and fragmented bone material did not allow an unequivocal identification of the species (fig. 2.1 B - E). Therefore, the tibia and the metatarsus fragments were used for a palaeogenetic analysis.

2.2.2 Extraction of Giant Deer DNA and Library Preparation

After the palaeontological material had been obtained, DNA had to be extracted, purified and prepared for sequencing. DNA extraction was performed with 50 mg bone powder per sample as described in section 1.2 using a guanidinium-silica based method (Rohland and Hofreiter, 2007). DNA libraries were prepared and indexed following published protocols (Meyer and Kircher, 2010, Kircher et al., 2012) as described in section 1.3. All extractions and pre-amplification steps of the library preparations were performed in a clean room and negative controls were used for each reaction. Indexed libraries were amplified, purified and quantified outside the clean room.

2.2.3 Bait Preparation and Capture

In parallel, bait had to be prepared for mitochondrial DNA capture (mt capture). For bait preparation DNA from a related species has to be used. DNA was extracted from a roe deer's (*Capreolus capreolus*) blood sample. To amplify a sequence of DNA, DNA polymerases require a 3' hydroxyl group (OH) as a starting point. Oligonucleotides that provide this 3' OH group can be generated and ordered as a set of a forward and a reverse primers that will specifically anneal to the target DNA sequence. Following the protocol after Meyer and Kircher 2010, 20 nt long primers were designed using the Primer3Plus online software (Untergasser et al., 2007) and used to produce two long-range PCR products that encompass the complete 16.6 kilo bases (kb) long roe deer mtDNA. The PCR products were sheared, biotinylated and immortalized on streptavidin-coated magnetic beads followed by capture through hybridization as described above (sec. 2.2) and

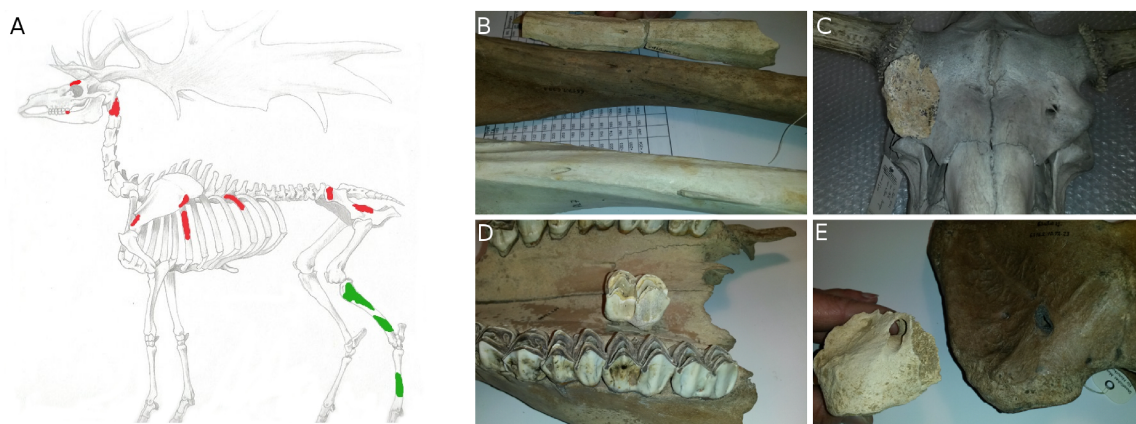


Figure 2.1: Skeletal remains of *Megaloceros giganteus*. **A** *Megaloceros giganteus* skeleton (template obtained from http://img13.deviantart.net/2f6c/i/2010/065/1/9/megaloceros_skeleton_by_cynothic.jpg). Highlighted skeletal fragments were discovered during the excavation and fragments highlighted in green were used for palaeogenetic analyses. **B** Tibia fragments (top) compared to a giant deer tibia (middle) and a tibia of an elk (bottom). **C** *M. giganteus* skull fragment (on the left) compared to an elk skull. **D** *M. giganteus* molar compared to the mandible of a giant deer museum specimen. **E** Fragment of a *M. giganteus* atlas compared to a complete giant deer atlas.

sequencing on the Illumina Hiseq 2500 platform (fig. 2.2).

2.2.4 Sequence Processing and Assembly

Although partial giant deer mtDNA sequences consisting of cytochrome b (cytb) were published previously (Lister et al., 2005, Hughes et al., 2006), so far whole mitochondrial genomes had not been reconstructed. Sequence reads were assigned to each sample based on their combination of index molecules. Forward and reverse reads were merged into single sequences if they overlapped by at least 11 bp (Kircher et al., 2011) and filtered for a length of at least 30 bp. The quality filtered reads were aligned to the complete mitochondrial reference sequence of a roe deer (*Capreolus capreolus*, NC_020684.1) using a custom mapping iterative assembler (MIA) (Green et al., 2008). This program was initially designed to reconstruct the mitochondrial genome sequence of the Neandertal and accounts for aDNA deamination damage and short reads (Green et al., 2008). It aligns reads to the reference sequence and generates a consensus sequence by collapsing all reads covering a certain region and calling a consensus base based on the majority of basecalls in the reads at each position. The consensus sequence is used as a new reference sequence and the process is iterated until the consensus sequences converge.

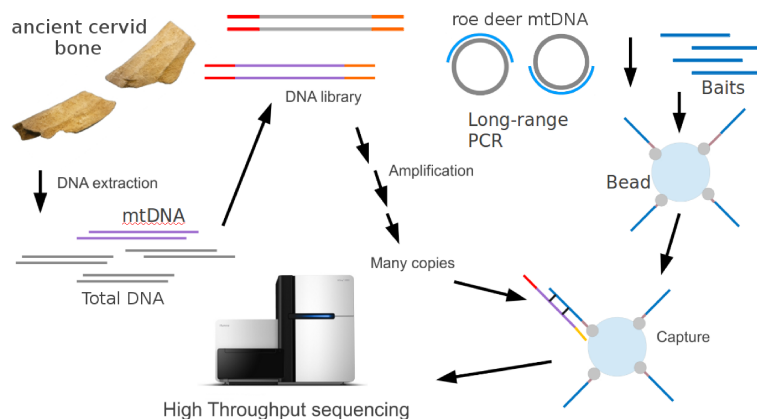


Figure 2.2: Mitochondrial DNA capture of the giant deer’s mtDNA. Total DNA is extracted from a metatarsal fragment of a giant deer (*Megaloceros giganteus*) and turned into sequencing libraries. To generate bait for mtDNA capture, two long-range PCR products are produced from related mtDNA of a roe deer (*Capreolus capreolus*) and fragmented into shorter molecules which are biotinylated and ligated to streptavidin-coated magnetic beads. Sequencing libraries are made single-stranded and pooled together with the bait made from roe deer mtDNA. Giant deer mtDNA molecules bind to the roe deer mtDNA bait because of sequence complementarity, get eluted and can be used for amplification and sequencing, whereas unbound environmental DNA is washed away.

A consensus base was called only if the coverage at this position was at least 5-fold. In addition, aDNA damage patterns were called using a script from the MIA package (fig. 2.3).

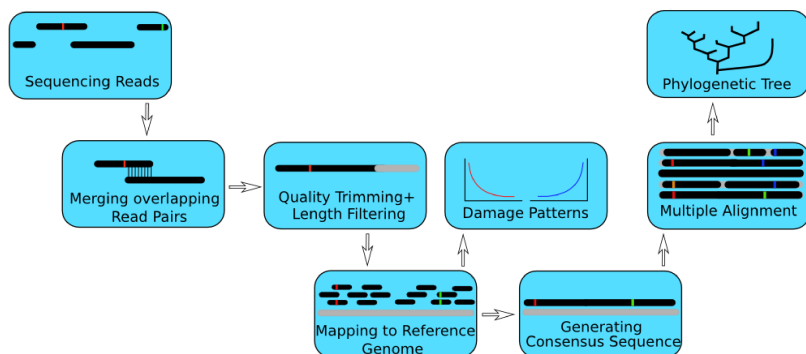


Figure 2.3: Bioinformatic analysis of sequencing data. Reads are merged, quality filtered and aligned to a reference sequence to generate a consensus sequence. A multiple sequence alignment is performed including the generated consensus sequence. Based on the alignment, different methods can be used to construct a phylogenetic tree. In addition, deamination damage patterns can be visualized and used for authentication of aDNA.

2.2.5 Multiple Sequence Alignment

Multiple sequence alignments (MSA) are among the most widely used bioinformatic analyses. Multiple sequence alignment programs such as *ClustalW* and *ClustalX* make use of dynamic programming algorithms such as *Neighbor-Joining*, which cluster sequences after computing all pairwise sequence distances and storing them into a distance matrix (Larkin et al., 2007). Sequences with the lowest distance are put together into a single cluster and the distance matrix is recalculated, followed by joining the sequence with the lowest distance to the cluster. Using the program MEGA6 (Tamura et al., 2013) the reconstructed mtDNA consensus sequences of both giant deer specimens were run together in a multiple sequence alignment with 44 cervid mtDNA consensus sequences available on GenBank (Benson et al., 2014) using the *ClustalW* algorithm, followed by a phylogenetic analysis (results are presented in the publication below: page 5, fig. 3.) Apart from the reconstruction and the phylogenetic analysis of the two ancient mitochondrial genome sequences, it had to be shown that the reconstructed sequences originate from *M. giganteus*. The only comparable data published so far were a cytochrome b (cytb) sequence from a *M. giganteus* specimen from western Siberia (Lister et al., 2005) and another one from Ireland (Hughes et al., 2006). Therefore, for our giant deer specimens, cytb sequences were reconstructed using the cytb reference sequence of a roe deer (*Capreolus capreolus*, Y14951.1) instead of its complete mitochondrial reference sequence (NC_020684.1). Cytb sequences presented in (Lister et al., 2005; AM072744.1) and (Hughes et al., 2006; AM182645.1) were obtained from GenBank together with cytb sequences of roe deer (Y14951.1), fallow deer (AJ000022.1, AM072742.1), hog deer (AY035874.1), Wapiti (AF423199.1), Sika deer (B021095.1), red deer (AF423195.1), Eld's deer (AY157735.1), Axis deer (AM072743.1) and a cytb sequence of Eurasian elk (KC337273). The sequences were supplemented with the reconstructed cytb sequences of both giant deer specimens from Hohlenstein-Stadel and Hohlefelds to perform a MSA using the *ClustalW* algorithm in MEGA6. Small deer *Muntiacus muntjak* (FJ556562) cytb sequence was also included as an outgroup.

2.2.6 Phylogenetic Analyses

Phylogenetic reconstruction aims at a graphical representation of the inferred evolutionary history of operational taxonomic units (OTU) such as DNA or protein sequences or even species. This can be in the form of a tree diagram or a phylogenetic network. Different algorithms have been developed for this type of analysis. In the following, the

application of three popular phylogenetic reconstruction concepts shall be illustrated using the example of the *M. giganteus* cytb: Maximum Parsimony (MP), Maximum Likelihood (ML) and Bayesian Inference (BI).

Following Ockham's razor principle MP assumes that the best evolutionary scenario is the one that explains the evolutionary history of an OTU in the most parsimonious way, i.e. assuming the lowest amount of phylogenetic changes (Baum and Smith, 2013). Given a data set, such as a MSA, the MP method tries to identify the largest subset of common positions among the OTUs, while keeping the amount of mutations to a minimum. The tree with the shortest length, i.e. the least amount of changes, is chosen as the best tree. Algorithmically MP minimizes the number of mutations in a phylogenetic tree by assigning character states to interior tree nodes without taking branch lengths into account. This can cause attraction between long branches which will for example erroneously show the taxa on their terminal nodes (leafs) as closely related despite a shorter genetic distance between those taxa and taxa at the ends of shorter branches (Bergsten, 2005). This algorithmic artifact is known as long branch attraction and is especially problematic in case of convergent evolution: shared, but independently derived traits will be grouped together assuming a common ancestor. Compared to other methods MP gives the most simple explanation for given data.

ML is a parameter-oriented approach for the reconstruction of a phylogeny. Given a MSA and parameters such as the evolutionary substitution model, it calculates all possible phylogenetic trees and computes the likelihood for every tree based on the given parameters. The tree with the highest likelihood for the given parameters is chosen as the ML tree. Initially, a starting tree is generated and the overall tree-topology likelihood is computed from all single site likelihoods based on the parameters. Then, the tree topology gets changed and the likelihood is recalculated. This is repeated until the topology with the ML is found (Schmidt and Haeseler, 2009). ML considers branch lengths and nucleotide substitution rates, and therefore offers a more realistic evolutionary scenario than MP.

A phylogenetic tree is called fully resolved if each of its inner nodes has exactly two descendant lineages. The branching is called dichotomous and the tree is called binary (Baum and Smith, 2013). This does not have to be always the case, since the actual evolutionary scenario might have produced more than two descendant lineages in a radiation event, a so called polytomy (fig. 2.4). In phylogenetic trees, polytomies more often indicate a branching uncertainty. The bootstrap method was developed in order to test the reliability of phylogenetic trees. It re-samples the original data set by replacement of

data points to create a set of pseudo-replicates of the same size as the original data, each of which is analyzed in the same way as the original data set. The bootstrap method then calculates the variation among the resulting phylogenetic estimates in order to determine the error in making phylogenetic estimates from the original data (Felsenstein, 1985). The lower the error, the higher is the statistical bootstrap support. A high bootstrap support for a clade is indicated by a number of > 0.8 at the branching node, meaning that if e.g. 1000 bootstrap replicates were generated, at least 800 replicates would show that particular clade (Baum and Smith, 2013).

BI is based on the posterior probability of a phylogenetic tree. Similar to ML it requires parameters and uses a likelihood function to calculate the probability of a phylogenetic tree. Using Bayes' Theorem (López Puga et al., 2015) BI infers the conditional "posterior" probability of a phylogenetic tree under "prior" parameters depending on the given data. BI became significantly efficient in combination with the Markov Chain Monte Carlo (MCMC) algorithm: After choosing a random starting tree from a set of possible phylogenies and random parameters, a new tree and new parameters are chosen. The chosen tree and the parameters represent a state. The conditional probability for the new state depending on the given data is calculated. If the probability is higher compared to the previous state, the new state is chosen. This process is run millions of times. The algorithm visits different trees and calculates posterior probabilities until reaching an equilibrium distribution (Yang and Rannala, 2012). The tree with the highest posterior probability is chosen to best represent the data assuming that the given priors are correct.

Inferring phylogeny with BI requires prior selection of an appropriate evolutionary model. Different assumptions of nucleotide substitution frequencies and rates have been implemented into various models for the correction of unseen or multiple substitutions. For example, Jukes & Cantor (JC69) is the simplest correction model which predicts that all nucleotide substitutions happen at the same rate (Jukes and Cantor, 1969). Higher

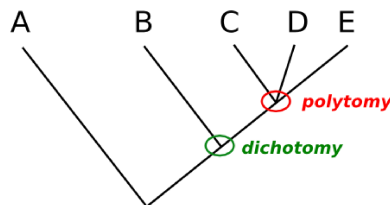


Figure 2.4: Dichotomy and polytomy. An example phylogenetic tree indicating a dichotomous split at the node comprising taxon B and C,D,E, and a polytomous split at the node including taxa C,D,E.

substitution rates are proposed for transitions than for transversions in Kimura’s K80 model (Kimura, 1980) assuming that substitutions happen at the same frequency in both directions (fig. 2.5). On the contrary, the HKY model after Hasegawa, Kishino and Yano weights substitution directions differently assuming that pyrimidines (cytosine and thymine) appear more frequently than purines (adenine and guanine) and assigns the highest substitution rate between thymines and cytosines (Hasegawa et al., 1985). Models that are based on a strict molecular clock assume that the nucleotide substitution rate is constant among different species and at all sites, whereas models with a relaxed molecular clock account for a variable rate of substitutions at different sites and between different species, and therefore represent the evolutionary history in a more realistic way (Drummond et al., 2006). Moreover, programs such as jModelTest (Darriba et al., 2012) are available for testing which substitution model fits best the given data.

Using the program jModelTest 2.1.4 the HKY model was identified as the best-fit nucleotide substitution model for reconstructing a phylogeny from the multiple cervid cytb sequence alignment described above. BI was conducted in MrBayes3 (Ronquist and Huelsenbeck, 2003) using the HKY nucleotide substitution model and a tree search based on a random starting tree. In three independent runs four Markov chains were run for 5 000 000 times, with tree sampling every 100 times. The first 25 % of the generated trees were used as a training set and therefore not regarded for inference of the phylogeny.

In parallel ML and MP phylogenies were generated in MEGA6 based on the cytb MSA. After removing all gaps caused by missing data or deletions (resp. insertions) phylogenies were reconstructed from a common sequence length of 640 informative positions using a bootstrap support of 5000 data set replicates. Reconstructions of phylogenetic relationships under maximum likelihood and maximum parsimony resulted in exactly

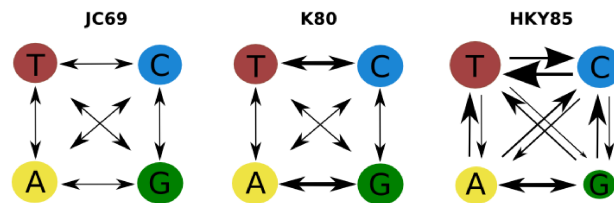


Figure 2.5: Nucleotide substitution models after Yang and Rannala, 2012. Nucleotide frequencies are represented by the size of the circle assuming that the substitution process is in equilibrium. The thicker the arrow, the higher is the substitution rate among the four nucleotides (T, C, A and G). While JC69 and K80 predict equal proportions of the four nucleotides, HKY85 predicts different proportions.

the same tree topology (fig. 2.6 B). Both methods provided a bootstrap support of 100 % for the specimens from the Swabian Alb (ST, HF) forming a common clade with the giant deer specimens from Siberia (Lister et al., 2005) and Ireland (Hughes et al., 2006) which reaffirms the authenticity of the Swabian Alb specimens being of *Megaloceros giganteus* origin. The phylogenetic position of the giant deer specimens was independently confirmed by the Bayesian method. Interestingly, BI showed both specimens from the Swabian Alb and the specimen provided by Lister et al., 2005 as a result of polytomy (fig. 2.6 A). Although the posterior probability at this node being low (51 %), which suggests an unresolved phylogeny, the Irish specimen forms a sister clade to this node supported by a posterior probability of 100 %. Consistent with the results from full mtDNA sequences presented in the publication attached below (page 5, fig. 3) all three different cyt b-based phylogenetic reconstruction approaches independently support giant deer being closer related to fallow deer and clearly isolated from other cervids, such as red deer, Eurasian elk and roe deer.

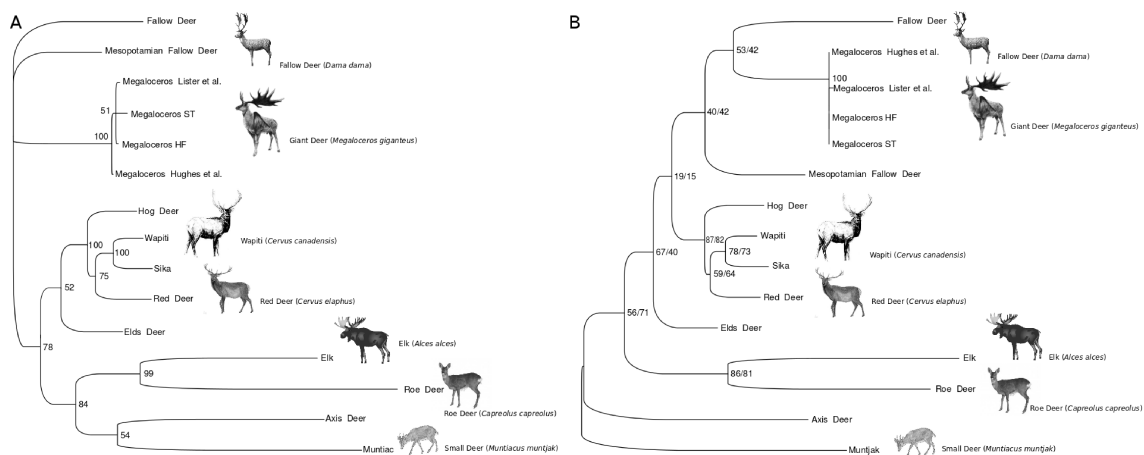


Figure 2.6: Inferred phylogeny of giant deer using 3 different methods. **A** BI tree obtained using the HKY evolution model and 3 independent runs for 5×10^6 generations. Node numbers indicate posterior probability values. **B** ML/MP tree based on 640 informative positions supported by 5000 bootstrap replicates. Node numbers indicate bootstrap values for ML/MP. Swabian Alb specimens fall together with giant deer from Ireland and Siberia confirming their giant deer origin.

SCIENTIFIC REPORTS

OPEN

Mitochondrial Genomes of Giant Deers Suggest their Late Survival in Central Europe

Received: 17 December 2014

Accepted: 20 April 2015

Published: 08 June 2015

Alexander Immel¹, Dorothée G. Drucker², Marion Bonazzi³, Tina K. Jahnke⁴, Susanne C. Münzel⁵, Verena J. Schuenemann¹, Alexander Herbig^{1,6}, Claus-Joachim Kind⁴ & Johannes Krause^{1,6,7}

The giant deer *Megaloceros giganteus* is among the most fascinating Late Pleistocene Eurasian megafauna that became extinct at the end of the last ice age. Important questions persist regarding its phylogenetic relationship to contemporary taxa and the reasons for its extinction. We analyzed two large ancient cervid bone fragments recovered from cave sites in the Swabian Jura (Baden-Württemberg, Germany) dated to 12,000 years ago. Using hybridization capture in combination with next generation sequencing, we were able to reconstruct nearly complete mitochondrial genomes from both specimens. Both mtDNAs cluster phylogenetically with fallow deer and show high similarity to previously studied partial *Megaloceros giganteus* DNA from Kamyshlov in western Siberia and Killavullen in Ireland. The unexpected presence of *Megaloceros giganteus* in Southern Germany after the Ice Age suggests a later survival in Central Europe than previously proposed. The complete mtDNAs provide strong phylogenetic support for a *Dama-Megaloceros* clade. Furthermore, isotope analyses support an increasing competition between giant deer, red deer, and reindeer after the Last Glacial Maximum, which might have contributed to the extinction of *Megaloceros* in Central Europe.

The extinct giant deer *Megaloceros giganteus* (also Irish Elk), first described by Blumenbach in 1799¹, stands out amongst the Pleistocene megafauna not only due to its sheer body size, but also because of its immense antlers, which spanned up to 4 m in diameter and weighted up to 45 kg². Appearing in the fossil record around 400,000 years ago (ya)³, its populations are thought to have ranged from Ireland to Lake Baikal⁴. Many theories have been proposed to account for its pattern of distribution across Eurasia in the Late Pleistocene and its extinction in the early Holocene. One unresolved question concerns the reason for the absence of giant deer during the Last Glacial Maximum (LGM, 20,000 – 12,500 ya) in Western and Central Europe, implying that these species had completely withdrawn from the region⁵. Before its purported extinction ca. 6,900 ya in western Siberia, *Megaloceros* recolonized northwestern Europe in the Late Glacial Interstadial⁵; however, no evidence has thus far been found for the presence of giant deer in Southern- and Central Europe.

The phylogenetic position of *Megaloceros* within the family *Cervidae* is still debated. The presence of large palmate antlers in extant fallow deer (*Dama dama*) and *Megaloceros giganteus* suggests a close relationship between the two species^{6,7}, whereas postcranial skeletal characters place *Megaloceros* into a group comprising red deer (*Cervus elaphus*), cheetal (*Axis*), and bush antlered deer (*Eucladoceros*)⁸.

¹Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen, Germany.

²Department of Geosciences, Palaeobiology, University of Tübingen, Tübingen, Germany. ³Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany. ⁴State Office for Cultural Heritage Baden-Wuerttemberg, Berliner Straße 12, D-73728 Esslingen, Germany. ⁵Institute for Archaeological Sciences, Archaeozoology, University of Tübingen, Tübingen, Germany. ⁶Max Planck Institute for the Science of Human History, Khalaische Straße 10, 07745 Jena, Germany. ⁷Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Tübingen, Germany. Correspondence and requests for materials should be addressed to J.k. (email: johannes.krause@uni-tuebingen.de)



Figure 1. Ancient large cervid remains. (a) A tibia fragment (ST/213/203/144) from the Hohlenstein Stadel cave and (b) a metatarsus fragment (HF/65/100) from the Hohle Fels cave site.

Primer	Sequence
Roe_Deer_mt1_forward	AAGCAAGGCACTGAAAATGC
Roe_Deer_mt1_reverse	TTGGTACAGGATAGGGTCTCC
Roe_Deer_mt2_forward	AACCGCACATGCATTGTAA
Roe_Deer_mt2_reverse	GGTTGTTTGCAGTGACGAGA
Roe_Deer_mt3_forward	CATCATGACCACAAGCTCCG
Roe_Deer_mt3_reverse	CGTGTGCTTGATACCAGCTC

Table 1. Forward and reverse primer pairs used to generate bait from roe deer mtDNA for targeted mtDNA enrichment.

Previous genetic studies on short regions of mitochondrial DNA (mtDNA) provide evidence for a closer relationship to fallow deer^{9,10}; however, the statistical support for the *Dama* - *Megaloceros* clade is low, likely due to the partial mtDNA regions studied and the limited availability of modern cervid mtDNA for comparison^{11,12}. Here we present two reconstructed nearly complete mitochondrial genomes of large cervid bone fragments found in two cave sites in Southern Germany dating to the late glacial period (ca. 12,000 ya). Phylogenetic analyses reveal that both specimens derive from *Megaloceros giganteus*. Comparison against mtDNAs from 44 extant deer species provides furthermore strong support for a *Dama* - *Megaloceros* mtDNA clade.

Results and Discussion

Morphological Analyses and Dating. During an excavation at the Hohlenstein Stadel cave at the Lone Valley (Baden-Württemberg, Southwestern Germany), an accumulation of large cervid bones was discovered including an almost complete atlas, two scapulae and two pelvic fragments, two ribs, a tooth (M3) and six fragments of a tibia. Apart from these finds, a metatarsus shaft fragment from a large cervid was obtained from the Hohle Fels cave at Schelklingen (Baden-Württemberg, Southwestern Germany). We genetically analyzed one of the six tibia fragments (ST/213/203/144, Fig. 1a) recovered from Hohlenstein Stadel cave, dated to $12,175 \pm 50$ uncal ya (ETH-41223), and the metatarsus fragment (HF/65/100, Fig. 1b) from Hohle Fels cave, which was dated to $12,370 \pm 30$ uncal ya (MAMS-16557). Radiocarbon ages were calibrated based on IntCal13 curve and calculated using the Calib 7.0 program.

mtDNA assembly. In order to reconstruct the mitochondrial genomes of our ancient cervid specimens, we isolated the total DNA, turned it into DNA sequencing libraries and enriched for mtDNA using bait generated from modern roe deer with specific primers (table 1). Illumina sequencing on a HiSeq2500 produced 944,648 and 6,123,389 merged reads, for Hohlenstein Stadel and Hohle Fels, respectively. These reads were mapped to the mitochondrial reference sequence of a roe deer (NC_020684.1) and a fallow deer (JN632629.1). Consensus sequences were generated for each sample based on at least 3-fold coverage. Both ancient samples produced identical consensus sequences for overlapping regions, demonstrating mitochondrial similarity. However, using fallow deer¹³ as a mapping reference, created an almost complete mtDNA sequence (91.52% for Hohlenstein Stadel and 99.99% for Hohle Fels), with 7634 unique mapping fragments for the Hohlenstein Stadel sample and 1,009,775 unique mapping fragments for Hohle Fels (table 2). The consensus sequence generated from positions with at least 3-fold coverage

Sample	¹⁴ C cal BP 2 sigma	Total Merged Reads	Unique mapped Reads	Unique Average Coverage	Average Read Length	%C → T substitutions
Hohlenstein Stadel	13904 – 14215	944,648	7,634	28x	60	27
Hohle Fels	14153 – 14681	6,123,389	1,009,775	5296x	85	49
Hohlenstein Stadel EB	13904 – 14215	422,344	193	1.3x	109	0
Hohlenstein Stadel LB	13904 – 14215	320,132	264	1.8x	109	0
Hohle Fels EB	14153 – 14681	50,832	48	0.2x	76	0
Hohle Fels LB	14153 – 14681	337,962	25	0.17x	108	0

Table 2. Mapping results for the Hohlenstein Stadel sample (ST/213/203/144) and the Hohle Fels sample (HF/65/100). EB: Extraction blank, LB: library blank. Columns from left to right: Sample, calibrated radiocarbon date, number of merged reads, number of unique mapped reads to the fallow deer mtDNA sequence, average mitochondrial genome coverage, average read length, and frequency of C to T substitutions at 5' end.

from the high-coverage Hohle Fels sample was subsequently used as reference sequence to re-align the Hohlenstein Stadel fragments. Remapping to this consensus sequence provided 7782 mapped fragments after duplicate removal for the Hohlenstein Stadel sample.

DNA damage patterns. To authenticate the sequenced fragments as ancient, the frequency of terminal substitutions was analyzed. It has been suggested that C to T substitutions at the 5' end and G to A substitutions at the 3' end are likely caused by deamination of cytosine causing miscoding lesions; these accumulate over time, and hence are characteristic of ancient DNA¹⁴. DNA fragments with frequencies of at least 20% for both types of substitutions at the 5' and 3' end can be regarded as authentic ancient DNA¹⁵. We observed a substitution frequency of 27% for C to T changes at the 5' ends and a substitution frequency of 25% of G to A at the 3' ends of sequence reads in the Hohlenstein Stadel sample (Fig. 2a). In the Hohle Fels sample 49% of the sequence reads showed a C to T substitution at the 5' ends and 48% a G to A substitution at the 3' ends (Fig. 2b). These values almost reach the theoretical maximum of 50% deamination at single stranded overhangs with a double strand library preparation protocol and indicate authentic ancient mtDNA in both faunal remains.

Phylogenetic Analysis. The reconstructed mtDNAs from both ancient cervid bones were aligned with 44 publicly available full mitochondrial genomes of extant cervids. The hypervariable D-loop region was excluded from the cervid mtDNA alignment due to its fast evolutionary rate that may decrease the phylogenetic resolution. We used MEGA 6.0.6 to construct both a maximum-likelihood (ML) tree (Fig. 3a) and a maximum-parsimony (MP) tree (Fig. 3b) and robustness of both methods was tested with 1000 bootstrap replicates. The best-fit substitution model for ML was identified with MEGA 6.0.6 to be the General Time Reversible (GTR+G+I) model (BIC score = 215112.733). A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.3366)). The rate variation model allowed for sites to be evolutionarily invariable ([+I]). Both tree reconstructions were based on a total of 14,147 positions. Alpine musk deer (*Moschus chrysogaster*, KC425457.1) was chosen as an outgroup. Both ML and MP topologies place the mtDNA sequences reconstructed from our two ancient cervid bones (Hohlenstein Stadel and Hohle Fels) in a completely resolved clade with both extant fallow deer subspecies (*Dama dama* and *Dama mesopotamica*), and exclude them from the group comprising Pere David's Deer (*Elaphurus davidianus*), *Rusa* sp. and red deer (*Cervus elaphus* sp.). The other large cervid present in Europe during the Pleistocene, the European elk, *Alces alces*, can also be excluded as the source of our ancient cervid bones.

The relationship of our ancient cervids to both *Dama* species within the *Dama* clade is, however, not completely resolved. As *Dama* was only introduced to Europe in the Medieval period¹⁶ and is absent in the palaeontological record of Western Eurasia, and since both ancient cervid bones derive morphologically from a large cervid, which is not elk (*Alces alces*) based on phylogenetic evidence, we conclude that both specimens originate from *Megaloceros giganteus*. To test this hypothesis we further compared our reconstructed ancient cervid mtDNAs to cytochrome b (*cytb*) regions of *Megaloceros* mtDNA previously published (table 3). We observe 1 nucleotide difference (mismatch) for the Hohlenstein Stadel sample compared to the previously published *cytb* sequence from a complete *M. giganteus* skeleton from the Kamyshlov site in western Siberia⁹ and no nucleotide differences compared to the *cytb* sequence obtained from a *M. giganteus* astragalus from Killavullen in Ireland¹⁰. However, 7 bases in the Hohlenstein Stadel *cytb* sequence could not be resolved because the coverage was too low at those positions. We find only 1 mismatch between the Hohle Fels sample and each of the previously published *M. giganteus cytb* sequences. For the same region we find 89 mismatches between the Hohlenstein Stadel sample and fallow deer⁹ and 87 mismatches compared to red deer¹⁷, whereas for the Hohle Fels sample we observe 104 mismatches compared to fallow deer and 97 mismatches to red deer, respectively.

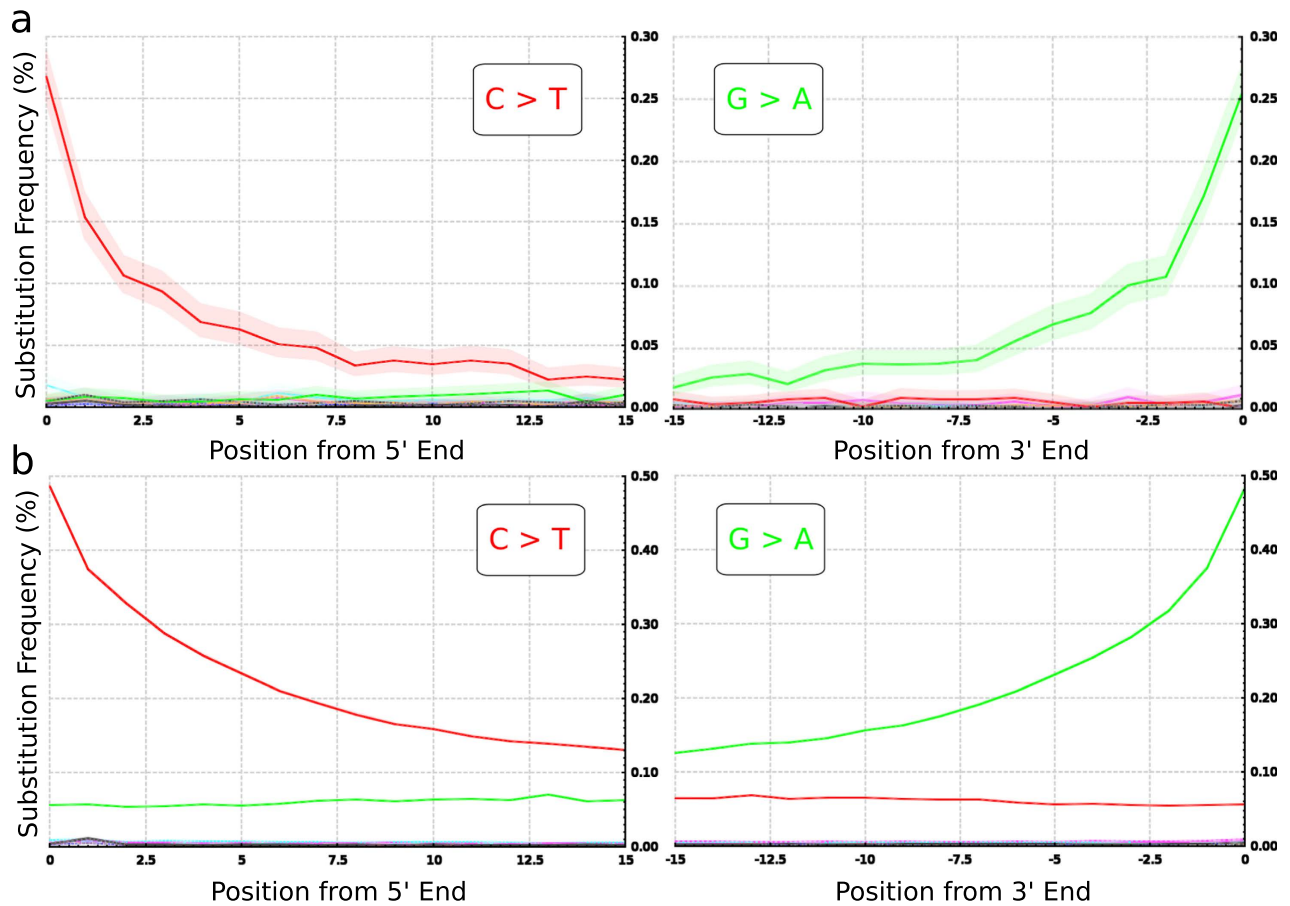


Figure 2. Substitution pattern at the 5' and 3' ends of the aligned sequence reads from the Hohlenstein Stadel sample (a) and the Hohle Fels sample (b). The misincorporation plots were generated using a custom software extension package (Krause J. *et al.* A complete mtDNA Genome of an Early Modern Human from Kostenki, Russia. *Curr. Biol.* 20, 231–236 (2010)).

The small nucleotide distance of our ancient large cervids and both previously determined giant deer *cytb* sequences confirm the attribution of our samples to *Megaloceros giganteus*. The close genetic relationship between the large cervid bone from Hohle Fels and the giant deer skeleton from Kamyshlov in western Siberia (AM072744.1)⁹ and the giant deer astragalus from Killavullen in Ireland (AM182645.1)¹⁰, respectively, suggests furthermore a close maternal relationship and low genetic diversity between Northwestern European, Eastern and Central European giant deer populations.

Stable Isotopes. To evaluate the stable isotope signature of our ancient cervids from the Hohle Fels and Hohlenstein Stadel cave sites, stable isotopes from collagen carbon (¹³C) and nitrogen (¹⁵N) were measured and compared to large cervids present in Central Europe during the Pleistocene such as red deer, reindeer, and giant deer^{18–20}. Pre-LGM (ca. 35,000 uncal ya) *Megaloceros* samples from Southern France and Belgium typically show isotopic signatures of collagen comparable to those of red deer (*Cervus elaphus*), while reindeer (*Rangifer tarandus*) provides systematically higher $\delta^{13}\text{C}_{\text{coll}}$ values likely due to the consumption of lichen¹⁸ (Fig. 4a). During the Late Glacial period (13,000 – 12,000 uncal ya), the ¹³C-based distinction among larger cervids from the Swabian, French, and Swiss Jura decreases. Stable isotope signatures of our both ancient cervid bones from the Hohle Fels cave and the Hohlenstein Stadel cave fall inside the red deer-reindeer cluster reflecting a potential overlap in diet and habitat (Fig. 4b).

Discussion

We obtained nearly complete mtDNA sequences from two ancient cervid bones from the Swabian Alb dated to $12,175 \pm 50$ uncal ya (13,904 – 14,215 cal ya) and $12,370 \pm 30$ uncal ya (14,153 – 14,681 cal ya), respectively. Both sequences are distinct from 44 mtDNAs of extant cervids. The phylogenetic analyses suggest that the reconstructed mtDNAs are maternally closely related to fallow deer (*Dama*). Based on the phylogenetic position of our reconstructed ancient mtDNAs, their close genetic relationship to the previously determined partial *cytb* sequence from a complete giant deer skeleton from western Siberia⁹

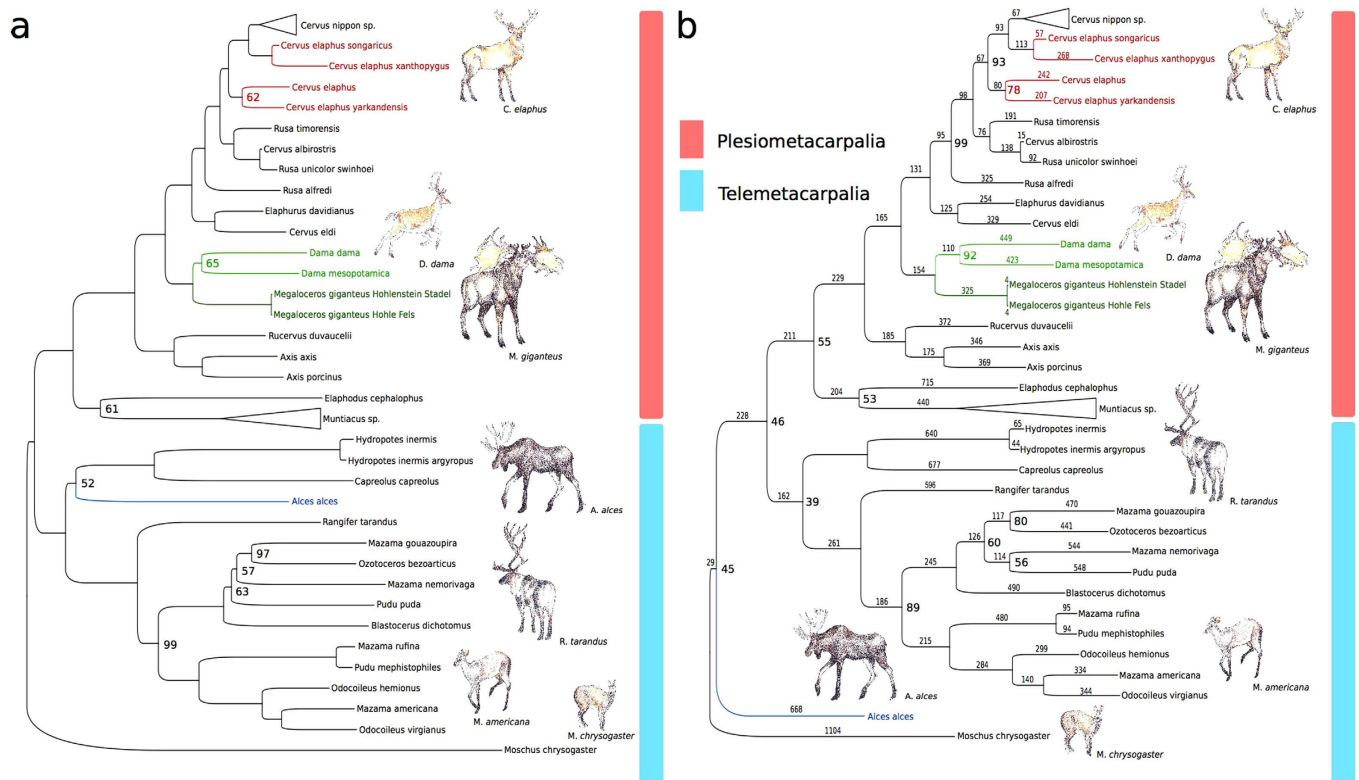


Figure 3. Phylogenetic trees of full mtDNA sequences from 44 extant cervid species and two ancient mtDNA sequences from two ancient cervid bones likely representing *Megaloceros giganteus*. Each tree is based on 14,147 positions. Bootstrapping was performed with 1000 bootstrap replicates. Only bootstrap values different from 100 are indicated at inner nodes. (a) Maximum-likelihood tree based on the General Time Reversible (GTR+G+I) model³³. (b) Maximum-parsimony tree. Branch-numbers in the Parsimony tree indicate the accumulated steps of genetic change (base substitutions) for each species after the divergence from its most recent common ancestor. Both topologies place *Megaloceros giganteus* together with fallow deer (*Dama* sp.) into a distinct clade from red deer (*Cervus elaphus*). Both trees were rooted with musk deer (*Moschus chrysogaster*) as outgroup. The deer drawings were kindly prepared and provided by Kerttu Majander.

Sequence	<i>M. giganteus</i> cytb Kamyshlov (AM072744.1) ⁹	<i>M. giganteus</i> cytb Killavullen (AM182645.1) ¹⁰	Fallow Deer cytb (AJ000022.1) ⁹	Red Deer cytb (AB924664.1) ¹⁷
Hohlenstein Stadel cytb	1	0	89	87
Hohle Fels cytb	1	1	104	97

Table 3. Number of nucleotide differences between the reconstructed and previously published cervid *cytb* sequences.

and to the complete *cytb* sequence from a giant deer astralagus from Ireland¹⁰ and due to the absence of fallow deer in Europe in the Pleistocene as well as due to the size of the bones, we conclude that both specimens derive from *Megaloceros giganteus*. We find strong support for a close maternal relationship to both *Dama* species. The maternal relationship within the *Megaloceros-Dama* clade however could not be resolved in our phylogeny suggesting an almost equal genetic distance of the two fallow deer species and giant deer. Our results disagree with the morphological conclusions that *Megaloceros* is closer related with a group comprising *Cervus*, *Axis*, and *Eucladoceros*^{8,21}, and that the occurrence of palmate antlers in *Megaloceros* and *Dama* must be the result of homoplasy. Our results also disagree with the conclusions derived from short mtDNA sequences such as partial *cytb* reported by Kuehn and colleagues²² which suggested a *Cervus-Megaloceros* clade and attributed the palmate antlers in *Megaloceros* and *Dama* to

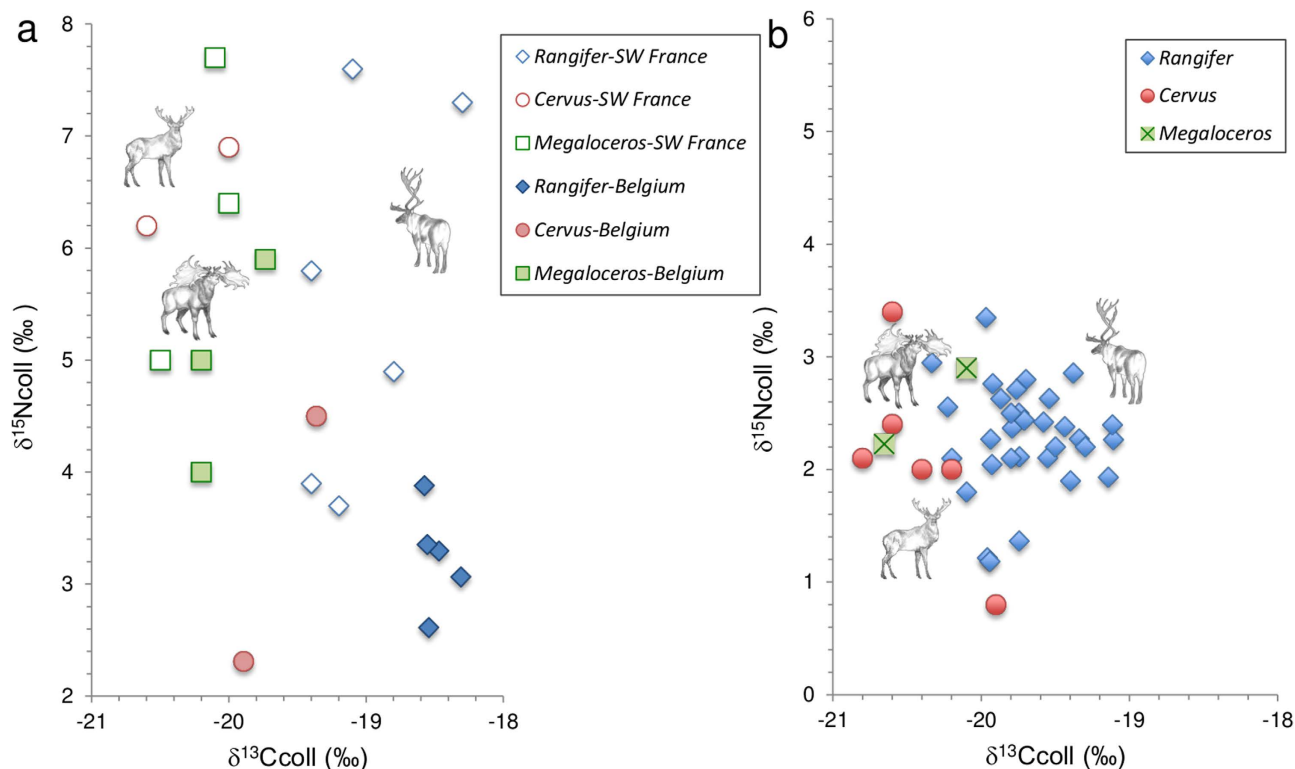


Figure 4. (a) Stable isotope values of reindeer (*Rangifer*), red deer (*Cervus*) and *Megaloceros* before the LGM in SW France and Belgium and (b) after the LGM in the Swabian, Swiss and French Jura indicate a decrease in the distinction between the isotope signatures of the three cervid species after the LGM, which might be due to overlapping diet and habitat. The deer drawings were kindly prepared and provided by Kerttu Majander.

homoplasmy. Instead, our results support the hypothesis of a *Megaloceros-Dama* clade suggested by previous studies based on morphological features and phylogenetic analyses from *cytb*^{2,3,9,10}.

Megaloceros is traditionally considered a species adapted to open-areas that could have suffered from the development of forest in the early Holocene, most likely because the huge antlers would have restricted the movement of males in dense woodland⁵. However, anatomy and distribution suggest that *Megaloceros* was a mixed feeder⁵, and carbon-13 results on enamel of *Megaloceros* of previous interglacial periods support the possibility of a boreal habitat²³. Isotope signatures of our two *Megaloceros* samples reveal that the diversity in habitat and diet decreased after the LGM between *Megaloceros*, red deer, and reindeer, probably resulting in an increased competition among the deer species in Central Europe. The overlapping niches and thus the increased competition with other deer species could explain, at least in part, the local extinction of *Megaloceros* in Southern Germany.

In conclusion, we generated two almost complete mitochondrial genomes from large cervid bones from the Hohle Fels and Hohlenstein Stadel caves in Southwestern Germany that date back to 12,000 uncal ya. Phylogenetic comparisons to contemporary deer mtDNA and previously determined ancient *cytb* DNA suggest that both mtDNA genomes derive from *Megaloceros giganteus*, which demonstrates its presence in Central Europe after the LGM. The close maternal relationship with the two fallow deer species resulting in a near polytomy in the *Dama-Megaloceros* clade questions the morphology-based grouping of giant deer and fallow deer in two separate genera. To date there has been no evidence that *Megaloceros* recolonized Central Europe after the LGM⁵; our findings provide support that *Megaloceros* returned to central parts of Europe even if the presence of humans might have hindered the re-colonization. In addition stable isotopes from our ancient cervid bones suggest a direct competition with other cervid species at the onset of the Holocene, potentially due to the lack of niche partitioning. Thus environmental factors may have played an important role in the final extinction of the giant deer.

Material and Methods

Extraction of ancient DNA. Bone samples were exposed to UV-light overnight to remove surface contamination. A sample of 50mg was removed from the inside of the longbone of each bone using a dentistry drill. DNA extraction was carried out using a guanidinium-silica based method²⁴. For each sample a DNA library was prepared according to published protocols²⁵. Sample-specific indexes were added to both library adapters to allow differentiation between individual samples after pooling and

multiplex sequencing²⁶. Indexed libraries were amplified in 100 µl reactions followed by purification over Qiagen MinElute spin columns (Qiagen, Hilden, Germany) and quantification using Agilent 2100 Bioanalyzer DNA 1000 chip. Target enrichment of mitochondrial DNA was performed by capture of the pooled libraries using bait generated from modern roe deer (*Capreolus capreolus*) mitochondrial DNA²⁷. The bait was generated by use of three primer sets (table 1) designed with the Primer3Plus software package. All extractions and pre-amplification steps of the library preparation were performed in clean room facilities and one negative control was included for each reaction.

Sequence Processing, Assembly, Duplicate Removal. Library pools were sequenced on the Illumina HiSeq 2500 platform using two index reads (2*100 + 7 + 7 cycles) following the manufacturer's protocol. De-indexing was performed by sorting all sequences corresponding to their p7 and p5 combinations using the CASAVA software version 1.8. Forward and reverse reads were merged into single sequences if they overlapped by at least 11 bp²⁸. Unmerged reads were discarded and merged reads were filtered for a length of at least 30 bp. Mapping of the length-filtered reads and removal of duplicate reads was performed using a custom mapping iterative assembler (MIA) which was developed to take into account sequence errors which commonly occur from ancient DNA damage²⁹. Reads were mapped to a full mitochondrial genome reference sequence of the roe deer, *Capreolus capreolus* (NC_020684.1). To achieve a higher resolution in the topology, in a second round sequence reads were mapped to a full mitochondrial genome of the fallow deer *Dama dama* (JN632629.1), and in a third round to the consensus sequence of our ancient putative *Megaloceros giganteus* sample HF/65/100, which was generated by mapping to the *Dama dama* mitochondrial reference sequence as described.

Analysis of ancient DNA damage patterns. C to T and G to A substitution patterns were obtained from the sequences using a custom software developed as an extension package to handle the output format of the mapping iterative assembler²⁹.

Multiple Sequence Alignment and Molecular Phylogenetic Analyses. A multiple sequence alignment was generated from 44 full mitochondrial genome Genbank³⁰ sequences of extant cervid taxa together with the assembled mitochondrial genome sequences of our putative *Megaloceros* samples using ClustalW (Larkin, M.A. *et al.* ClustalW and ClustalX version 2. *Bioinformatics* 23: 2947–2948 (2007)). Alignments and phylogeny constructions were conducted in MEGA 6.0.6. The mitochondrial D-Loop region was excluded using the BioEdit sequence alignment editor (Hall, T.A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41:95–98 (1999)) and phylogenies were constructed from a total of 14,147 positions. Both maximum-likelihood and maximum-parsimony topologies were generated for all positions for which coverage was at least three-fold in each of the ancient sequences. Alignment columns with gaps or missing data were eliminated. Bootstrap support values were obtained over 1000 replicate data sets, using alpine musk deer as an outgroup (*Moschus chrysogaster*, KC425457.1). The phylogenetic trees were edited in FigTree version 1.4.0 (<http://tree.bio.ed.ac.uk/software/figtree>).

Pairwise Comparison of Cytochrome b Sequence Differences. To identify cytochrome b (*cytb*) coordinates within our reconstructed mitochondrial sequences, these were aligned to previously published *cytb* sequences of *Megaloceros giganteus* (AM072744.1, AM182645.1), fallow deer (AJ000022.1), and red deer (AB924664.1) using MEGA 6.0.6 and sequences outside of the aligned regions were discarded. Nucleotide sequence differences were then calculated by pairwise alignment between each of our ancient *cytb* sequences and each of the previously published *cytb* sequences using BLAST.

Stable Isotope Analyses. To study the habitat pattern revealed by stable isotopes, collagen was extracted from both bone fragments and carbon (¹³C) and nitrogen (¹⁵N) were measured. The results were combined with stable isotope data obtained from ancient reindeer (*Rangifer tarandus*) and red deer (*Cervus elaphus*) remains of the Swabian, French, and Swiss Jura dating 13,000 to 12,000 uncal year ago, which corresponds roughly to the GI-1e interstadial^{18–20}. The results were compared to stable isotopes from morphologically defined deer specimens including giant deer (*Megaloceros giganteus*) dating to the pre-LGM from southwestern France (SW France) and Belgium^{18,31}.

Isotopic analysis ($\delta^{13}\text{C}_{\text{coll}}$, $\delta^{15}\text{N}_{\text{coll}}$) was conducted at the Department of Geosciences of Tübingen University using a Thermo Quest Delta+XL mass spectrometer coupled to a NC2500 CHN-elemental analyzer, which provides elemental analysis (C_{coll} , N_{coll}). The international standards used include marine carbonate (V-PDB) for $\delta^{13}\text{C}$ and atmospheric nitrogen (AIR) for $\delta^{15}\text{N}$. Analytical error, based on within-run replicate measurement of laboratory standards (albumen, modern collagen, USGS 24, IAEA 305A), was $\pm 0.1\text{‰}$ for $\delta^{13}\text{C}$ values and $\pm 0.2\text{‰}$ for $\delta^{15}\text{N}$ values. Reliability of carbon and nitrogen isotopic values was established by measuring the chemical composition, with $\text{C}/\text{N}_{\text{coll}}$ atomic ratio within the range of 2.9 to 3.6³².

References

1. Blumenbach, J. F. Handbuch der Naturgeschichte. Sechste Auflage. *Nebst zwey Kupfertafeln*. (Dieterich, Göttingen, 1799).
2. Gould, S. J. The Origin and Function of 'Bizarre' Structures: Antler Size and Skull Size in the 'Irish Elk', *Megaloceros giganteus*. *Evolution* **28**, 191–200 (1974).
3. Lister, A. M. The evolution of the giant deer, *Megaloceros giganteus* (Blumenbach). *Zool. J. Linnean. Soc.* **112**, 65–100 (1994).
4. Kahlke, R. D. Die Entstehungs-, Entwicklungs- und Verbreitungsgeschichte des Oberpleistozänen *Mammuthus-Coelodonta-Faunencomplexes* in Eurasien (Grobsauger). *Abh. Senckenb. Natf. Ges.* **546**, 1–64 (1994).
5. Stuart, A. J., Kosintsev, P. A., Higham, T. F. & Lister, A. M. Pleistocene to Holocene extinction dynamics in giant deer and woolly mammoth. *Nature* **431**, 684–689 (2004).
6. Freudenberg, W. Die Säugetiere des älteren Quartärs von Mitteleuropa. *Geologische und Paläontologische Abhandlungen* **12**, 455–670 (1914).
7. Lister, A. M. Evolutionary and ecological origins of British deer. *P. Roy. Soc. Edinb.* **82B**, 205–229 (1984).
8. Pfeiffer, T. The first complete skeleton of *Megaloceros verticornis* (Dawkins, 1868) Cervidae, Mammalia, from Bilshausen (Lower Saxony, Germany): description and phylogenetic implications. *Fossil Record* **5**, 289–308 (2002).
9. Lister, A. M. *et al.* The phylogenetic position of the 'giant deer' *Megaloceros giganteus*. *Nature* **438**, 850–853 (2005).
10. Hughes, S. *et al.* Molecular phylogeny of the extinct giant deer, *Megaloceros giganteus*. *Mol. Phylogenet. Evol.* **40**, 285–291 (2006).
11. Irwin, D. M., Kocher T. D. & Wilson A. C. Evolution of the cytochrome b gene of mammals. *J. Mol. Evol.* **32**, 128–144 (1991).
12. Honeycutt, R. L., DeWoody J. A. & Skow L. C. Microsatellite markers in white-tailed deer. *J. Hered.* **86**, 317–319 (1995).
13. Hassani, A. *et al.* Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *C. R. Biol.* **335**, 32–50 (2012).
14. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. USA* **104**, 14616–14621 (2007).
15. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Pääbo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* **7**, e34131 (2012).
16. Ueckermann, E. & Hansen, P. *Das Damwild. Naturgeschichte - Hege - Jagd*. (Parey, Hamburg 1994).
17. Luciani, S. & Olivieri, C. Positioning the red deer (*Cervus elaphus*) hunted by the Tyrolean Iceman into a mitochondrial DNA phylogeny. *PLoS ONE* **9**, e100136 (2014).
18. Bocherens, H., Drucker, D. G., Billiou, D., Patou-Mathis, M. & Vandermeersch B. Isotopic evidence for diet and subsistence pattern of Saint-Césaire I Neanderthal. *J. Hum. Evol.* **49**, 71–87 (2005).
19. Drucker, D. G., Bridault, A., Cupillard, C., Hujic, A. & Bocherens, H. Evolution of habitat and environment of red deer (*Cervus elaphus*) during the Late-glacial and early Holocene in eastern France (French Jura and the western Alps) using multi-isotope analysis ($\delta^{13}\text{C}$, $\delta^{15}\text{N}$, $\delta^{18}\text{O}$, $\delta^{34}\text{S}$) of archaeological remains. *Quatern. Int.* **245**, 268–278 (2011a).
20. Drucker, D. G., Kind, C. J. & Stephan, E. Chronological and ecological information on Late-glacial and early Holocene reindeer from northwest Europe using radiocarbon (^{14}C) and stable isotope (^{13}C , ^{15}N) analysis of bone collagen: case study in southwestern Germany. *Quatern. Int.* **245**, 218–224 (2011b).
21. Pfeiffer, T. Die Stellung von *Dama* (Cervidae, Mammalia) im System plesiometa-carpaler Hirsche des Pleistozäns. - Phylogenetische Rekonstruktion – Metrische Analyse. *Cour. Forsch. Inst. Senck.* **211**, 1–218 (1999).
22. Kuehn, R., Ludt, C. J., Schroeder, W. & Rottmann O. Molecular phylogeny of *Megaloceros giganteus*-the giant deer or just a giant red deer? *Zoolog. Sci.* **22**, 1031–1044 (2005).
23. Pushkina, D. & Raia, P. Human influence on distribution and extinctions of the late Pleistocene Eurasian megafauna. *J. Hum. Evol.* **54**, 769–782 (2008).
24. Rohland, N. & Hofreiter, M. Ancient DNA extraction from bones and teeth. *Nat. Protoc.* **2**, 1756–1762 (2007).
25. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* doi:10.1101/pdb.prot5448 (2010).
26. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, 1–8 (2012).
27. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* **5**, e14004 (2010).
28. Kircher, M., Heyn, P. & Kelso, J. Addressing challenges in the production and analysis of illumina sequencing data. *BMC Genomics* **12**, 382–382 (2011).
29. Green, R. E. *et al.* A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416–426 (2008).
30. Benson, A. D., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. GenBank. *Nucleic Acids Res.* **33**, 34–38 (2005), <http://www.ncbi.nlm.nih.gov/genbank>, Date of access: 28/09/2014.
31. Bocherens, H. *et al.* Isotopic evidence for dietary ecology of cave lion (*Panthera spelaea*) in North-Western Europe: prey choice, competition and implications for extinction. *Quatern. Int.* **245**, 249–261 (2011).
32. DeNiro, M. J. Postmortem preservation and alteration of *in vivo* bone collagen isotope ratios in relation to palaeodietary reconstruction. *Nature* **317**, 806–809 (1985).
33. Nei, M. & Kumar, S. *Molecular Evolution and Phylogenetics*. (Oxford University Press, New York, 2000).

Acknowledgements

We thank Kerttu Majander for the elaborate deer drawings and Adrian Lister as well as Kirsten Bos for valuable comments that greatly improved the manuscript. This research was funded by European Research Council Starting Grant APGREID (to A.I. and J.K.). The European Social Fund and Ministry of Science, Research and Arts of Baden-Württemberg funded the current position of Dorothée G. Drucker.

Author Contributions

A.I. and J.K. conceived and designed the research. T.K.J., S.C.M. and C.J.K. provided the bone samples and conducted initial morphological analyses. M.B. and V.J.S. performed the extraction of mtDNA and the preparation of sequencing libraries. A.I. and A.H. performed the bioinformatic analyses and D.G.D. conducted the stable isotope analyses. A.I. wrote the manuscript and J.K. and D.G.D. mostly contributed to the discussion. All authors reviewed the manuscript.

Additional Information

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

How to cite this article: Immel, A. *et al.* Mitochondrial Genomes of Giant Deers Suggest their Late Survival in Central Europe. *Sci. Rep.* **5**, 10853; doi: 10.1038/srep10853 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

2.3 Study 2: Effect of X-rays on ancient DNA

2.3.1 Background and Previous Work

Computed tomography (CT) allows high resolution insights into archaeological remains without destructive sampling. On the other hand CT makes use of X-rays, a form of ionizing radiation, that has been shown to cause mutations, DNA strand breaks and other molecular structural modifications of DNA (Muller, 1927, Wolff, 1967, Grosovsky et al., 1988, Liber et al., 1986). This happens either by directly affecting the DNA molecules through X-ray photons or by producing radicals which damage the DNA (Roots and Okada, 1975, Lindahl, 1993). These mechanisms were observed in contemporaneous DNA and questioned the non-destructiveness of CT applied to precious archaeological material, such as Neandertal bones, often leading to a conflict of interests between (archaeo-)geneticists and anthropologists. Since aDNA is already fragmented, chemically modified and usually present in minute amounts, it is likely that the X-ray effect on aDNA can exacerbate its condition. Temporary darkening of CT-scanned teeth from palaeontological remains was observed suggesting potential molecular damage under the surface (Richards et al., 2012).

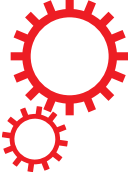
Previous studies have tried to simulate the effect of X-rays on aDNA by using modern pig bones and preserved bird skins from museum specimens as a proxy for archaeological material (Goetherstrom et al., 1995, Grieshaber et al., 2008, Paredes et al., 2012). However, their results were contradictory: while Goetherstrom stated that DNA gets degraded when the pig bones are scanned, Grieshaber could not find any observable effect after exposing the sample to X-radiation, and Paredes observed no significant difference between exposed and unexposed tissue. These studies used PCR or capillary gel-electrophoresis to assess the DNA fragmentation caused by CT-emitted X-rays, thus no sequencing has been done on aDNA exposed to X-radiation so far. The goal of my second study was therefore to find out if effects of X-radiation can be detected on aDNA after CT-scans and to determine a threshold of a maximum absorbed radiation dose hazard-free to aDNA integrity.

2.3.2 Investigating X-ray effects on aDNA after Mitochondrial DNA Capture

We performed mitochondrial capture (Maricic et al., 2010) on DNA extracted from CT-scanned teeth and bones of Late Pleistocene megafaunal specimens including cave bear, steppe bison, giant deer and roe deer, and produced NGS data that allowed to assess the effect of X-rays on aDNA at higher resolution than it was previously possible. The analysis focused on aDNA quantity, fragment length and damage patterns. The aDNA quantity was measured in a quantitative PCR (qPCR) after extraction and before mt capture, while fragment length and damage patterns were assessed after mt capture and sequencing. The dose of absorbed ionizing radiation is defined as Gray (Gy), where $1 \text{ Gy} = 1 \frac{\text{J}}{\text{kg}}$ (Joule per kilogram) describes the energy load absorbed per mass unit. We found that a high X-radiation dose of 170 kGy detrimentally affects the amount of aDNA. For comparison, the absorbed energy after a CT chest scan is 7 mGy, which is 24×10^6 times lower than the energy dose applied in the first experiment of our study. In a serial scan the absorbed radiation dose was increased from 93 Gy to 93 kGy. Since X-radiation induces strand breaks, the fragment length and the total amount of aDNA molecules decreased with accumulating radiation dose. Interestingly, X-radiation decreased the aDNA-characteristic cytosine to thymine (C to T) substitution patterns. This can be explained by the artificial introduction of new molecules without terminal substitution patterns through X-ray-induced strand breaks: while the original substitution patterns stay unchanged, X-rays split the molecules by inducing double strand breaks (Bradley and Kohn, 1979) which results in several shorter molecules whereof only one contains the original terminal C to T substitution (publication below, page 5, fig. 4 a).

To simulate the effect of a conventional CT-scan, crushed cave bear bone material was scanned together with cave bear mandibles. The total absorbed energy was 720 Gy. Also here, changes in aDNA amounts, fragment lengths and substitution frequencies were observed. Based on the observations from the performed experiments we defined 200 Gy as an upper bound for the absorbed energy dose to not cause any detectable effects on aDNA molecules. This dose is ca. 8000 times higher than the highest absorbed dose for a medical scan, and is therefore far beyond any radiation dose that would have been produced by current CT-settings. Moreover the dose of 200 Gy exceeds by 20×10^6 times the dose which is absorbed from airport luggage scans. We therefore state that luggage scans of archaeological samples do not harm aDNA.

SCIENTIFIC REPORTS



OPEN

Effect of X-ray irradiation on ancient DNA in sub-fossil bones – Guidelines for safe X-ray imaging

Received: 06 June 2016
Accepted: 15 August 2016
Published: 12 September 2016

Alexander Immel^{1,2,*}, Adeline Le Cabec^{3,4,*}, Marion Bonazzi^{5,*}, Alexander Herbig¹, Heiko Temming³, Verena J. Schuenemann^{2,12}, Kirsten I. Bos¹, Frauke Langbein², Katerina Harvati⁶, Anne Bridault⁷, Gilbert Pion⁸, Marie-Anne Julien^{9,10}, Oleksandra Krotova¹¹, Nicholas J. Conard¹², Susanne C. Münzel¹³, Dorothée G. Drucker¹⁴, Bence Viola^{3,15}, Jean-Jacques Hublin³, Paul Tafforeau⁴ & Johannes Krause^{1,2,12}

Sub-fossilised remains may still contain highly degraded ancient DNA (aDNA) useful for palaeogenetic investigations. Whether X-ray computed [micro-] tomography ([μ]CT) imaging of these fossils may further damage aDNA remains debated. Although the effect of X-ray on DNA in living organisms is well documented, its impact on aDNA molecules is unexplored. Here we investigate the effects of synchrotron X-ray irradiation on aDNA from Pleistocene bones. A clear correlation appears between decreasing aDNA quantities and accumulating X-ray dose-levels above 2000 Gray (Gy). We further find that strong X-ray irradiation reduces the amount of nucleotide misincorporations at the aDNA molecule ends. No representative effect can be detected for doses below 200 Gy. Dosimetry shows that conventional μ CT usually does not reach the risky dose level, while classical synchrotron imaging can degrade aDNA significantly. Optimised synchrotron protocols and simple rules introduced here are sufficient to ensure that fossils can be scanned without impairing future aDNA studies.

Since its discovery, X-ray imaging has found a broad range of applications in medical, anthropological and palaeontological studies. X-ray computed [micro-] tomography ([μ]CT) scans are routinely used to generate three-dimensional (3D) models of fossil remains, to explore internal structures, which can essentially help to distinguish between specimens, as well as to provide virtual replicas of the fossils that can be shared for analysis with other institutions. Furthermore, CT scanning of fossil and mummified remains prior to analysis requiring invasive/destructive sampling has been recommended and is routinely performed, in order to preserve valuable internal morphological information¹. Until now, about half of the sub-fossil Pleistocene human remains used for aDNA analysis, including Neandertal and Denisovan, were scanned before sampling (Supplementary Table 1). However, recent concerns have been raised regarding the potential deleterious effect of X-rays on the retrieval of aDNA², especially in the case of μ CT performed using synchrotron sources. These concerns arose originally after the observation of transitory darkening of translucent or white enamel when submitted to high

¹Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany. ²Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen, Germany. ³Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. ⁴European Synchrotron Radiation Facility, Grenoble, France. ⁵Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany. ⁶Senckenberg Center for Human Evolution and Palaeoecology, Palaeoanthropology, University of Tübingen, Tübingen, Germany. ⁷CNRS UMR 7041 ArScAn, Equipe Archéologies environnementales, F-92023 Nanterre Cedex, France. ⁸Association départementale pour la recherche archéologique en Savoie, F-73230 Saint-Alban-Leyse, France. ⁹Centre for the Archaeology of Human Origins, Archaeology Department, University of Southampton, Southampton, UK. ¹⁰Unité Histoire naturelle de l'Homme préhistorique (UMR 7194), Sorbonne Universités, Muséum national d'Histoire naturelle, CNRS, Paris, France. ¹¹Department of Stone Age, Institute of Archaeology, National Ukrainian Academy of Science, Kiev, Ukraine. ¹²Senckenberg Center for Human Evolution and Palaeoenvironment, University of Tübingen, Tübingen, Germany. ¹³Institute for Archaeological Sciences, Archaeozoology, University of Tübingen, Tübingen, Germany. ¹⁴Department of Geosciences, Palaeobiology, University of Tübingen, Tübingen, Germany. ¹⁵Department of Anthropology, University of Toronto, Toronto, Canada. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to P.T. (email: paul.tafforeau@esrf.fr) or J.K. (email: krause@shh.mpg.de)

Conditions	Experiment 1: Extreme Irradiation	Experiment 2: Exposure Time Series	Experiment 3: classical synchrotron scan
Average Energy	75.85 keV	74.69 keV	126.85 keV
Filters	Al (3 mm), Cu (0.1 mm), W (0.1 mm)	Al (1.5 mm), Cu (0.1 mm), W (0.1 mm)	Al (23 mm), Cu (6 mm)
Dose Rate	83.41 Gy/s	93.72 Gy/s	1.61 Gy/s
Total exposure time	2040 s	From 1s to 1000 s	447 s
Total Delivered Dose	170.15 kGy	93.72 Gy – 93.72 kGy	720 Gy

Table 1. Overview of the used conditions during the three conducted experiments.

resolution synchrotron scans (typically sub- μm voxel size), that are used to investigate dental microstructures^{2,3}. This darkening can be easily removed by using low power, low energy UV (375 nm, ideally from LED), as already demonstrated on hundreds of teeth imaged at the European Synchrotron Radiation Facility (ESRF). Nevertheless it indeed indicates a significant effect of the X-rays on the specimen, mostly due to electronic excitation, but also partially to ionisation^{4,5}. One of the most relevant applications of synchrotron μCT , but also the most controversial in regards to the risks for aDNA, is for non-destructive investigations of dental structures and microstructures in palaeoanthropology, using virtual palaeohistology approaches^{3,6–8}.

Numerous studies describe the effects of X-ray irradiation on DNA molecules in biological conditions, particularly in medical contexts. Mutations, DNA strand breaks, chemical modifications of bases, and structural changes such as gene order rearrangements have been identified^{4,9–11}. Creation and diffusion of radicals during X-ray exposure have been linked to molecular structural modifications^{12,13}, since X-ray photons are known to contribute to strand breaks¹⁴. In living organisms, repair mechanisms exist to withstand moderate levels of radiation-induced damage¹⁵. In the absence of repair mechanisms, the damages would accumulate with increasing X-ray dose (each new scan adding to the effects of the previous ones). When compounded with additional DNA modifications such as oxidative and hydrolytic damage as a result of natural decay and taphonomic processes^{16,17}, downstream genetic analyses may be affected in cases of significant levels of X-ray dose accumulation. Particularly the hydrolytic conditions seem to play a critical role: as demonstrated in a recent study on simulated effects of X-radiation on fragmented DNA in dry, wet and frozen states, the highest probability of radiation-induced DNA damage occurs in a wet state¹⁸.

Earlier studies have used modern pig (*Sus scrofa*) bones as proxies for archaeological samples, where molecular damage was determined based on the success of Polymerase Chain Reaction (PCR) amplification^{19,20}. Preserved bird skins from museum specimens were evaluated via capillary electrophoresis-based DNA quantification to infer the level of fragmentation induced by X-ray exposure²¹. The results of these investigations, however, were not consistent, potentially owing to inappropriate or non-homogenous samples, the use of low sensitivity measures to evaluate DNA concentrations, or insufficient control of X-ray dose. To date, no single investigation has reliably assessed the effects of X-ray irradiation on authentic aDNA molecules. A robust investigation of the effects of accumulation of X-ray dose on ancient DNA integrity was, therefore, urgently needed.

In this study we evaluate the effects of X-ray radiation on aDNA from Late Pleistocene megafaunal teeth and bones. In order to assess the effects of a large range of X-ray dose deposition, we used powerful polychromatic X-rays instead of a conventional X-ray source. Homogenised aliquots of crushed bone and dental tissues were exposed to increasing synchrotron X-ray doses using different exposure times at the ESRF. The dose rate and integrated dose were quantified as water equivalent surface dose for each experimental setup. After exposure, we used both quantitative PCR and next generation sequencing to evaluate the effects of synchrotron irradiation on: (1) aDNA quantity, (2) aDNA fragment length and (3) aDNA-characteristic nucleotide misincorporation patterns.

Our results reliably demonstrate a clear relationship between increasing X-ray dose deposition and level of aDNA damage, likely through increased strand breakage. We observe that strong X-ray exposure can significantly degrade aDNA (surface dose above 10 000 Gray (10 kGy)); however, no effect can be demonstrated for a dose below 200 Gy.

Dosimetry on classical synchrotron configurations, as well as on new configurations optimised for low dose imaging used at the ESRF on the beamline ID19 for palaeoanthropology, and finally on two conventional μCT scanners (a BIR ACTIS 225/300 and a Skyscan 1173), demonstrate that in the vast majority of cases conventional microtomographs are well below the detection limit of any defect on aDNA.

Results

Crushed material from 38 Late Pleistocene animal bones and teeth consisting of bison from the Ukraine, cave bear, and giant deer from the Swabian Jura, and roe deer from the French Jura were exposed to various configurations of polychromatic synchrotron beam on the beamline BM05 at the ESRF (Table 1). In the first experiment, aliquots were exposed to an extremely high level of X-ray radiation. For the second experiment, aliquots were irradiated with increasing exposure time. In the third experiment, we exposed aliquots to irradiation applicable in conventional high quality imaging μCT .

Effect of extreme irradiation on aDNA quantity. In the first experiment, aliquots from 11 well preserved Late Pleistocene bones were exposed for 34 minutes to an extreme dose of X-ray radiation of 170 kGy (water equivalent surface dose). Subsequent processing of each sample was performed alongside a non-scanned control, a negative extraction control, and a negative library-preparation control. The amount of aDNA that could

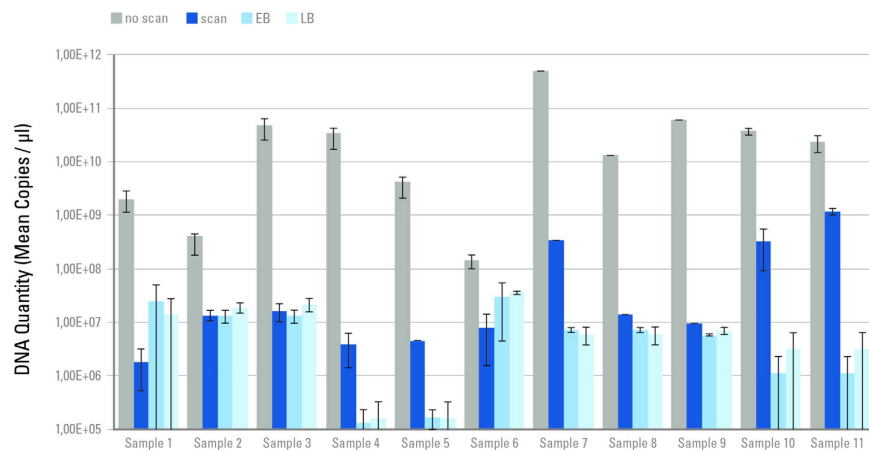


Figure 1. aDNA quantitation of 11 ancient samples after exposure to a radiation dose of 170 kGy. Each column represents a mean copy number obtained from aliquots of two independent DNA-libraries made for each sample. “scan”: scanned aliquot, “no scan”: non-scanned control, “EB”: extraction blank, “LB”: library blank. Copy numbers were normalised by the amount of extract included in each library and mean copy numbers were calculated from both libraries. Values and standard deviation are shown on a logarithmic scale.

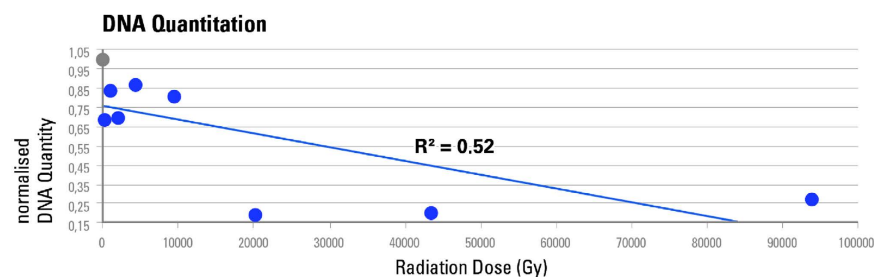


Figure 2. qPCR-based DNA quantitation after irradiation from 0 to 93.72 kGy. Normalised mean aDNA amounts are plotted against cumulative radiation dose received after 0 s (0 Gy), 2.2 s (~206 Gy), 10 s (~937 Gy), 21.5 s (~2.02 kGy), 46.4 s (~4.35 kGy), 100 s (~9.37 kGy), 215.4 s (~20.2 kGy), 464.2 s (~43.5 kGy) and 1000 s (~93.72 kGy). The non-scanned control is represented as a grey circle, the scans as blue circles. Values were normalised by the amount of extracted material and the corresponding non-scanned controls. Mean copy numbers were then obtained per exposure time group consisting of two aliquots from independent DNA-libraries.

be quantified via qPCR decreased substantially in the scanned aliquots compared to the non-scanned controls, indicating a highly damaging effect of this high X-ray radiation dose (Fig. 1). No sequencing data were produced for these exposed samples, as the amplifiable aDNA quantities after scanning were too low, being comparable to the negative controls.

Effect of increasing X-ray dose on aDNA quantity. To investigate a possible correlation between an increasing radiation dose and its effect on aDNA, and to define what level of irradiation could be considered harmful for future aDNA analyses aliquots from different bones were exposed to increasing dose levels, from 0 to 93.2 kGy, by changing exposure time from 0 to 1000 s. The estimated number of total library molecules from each sample was normalised by the amount of extracted material and the corresponding non-scanned aliquot. Mean values were then calculated for each exposure time group. Our results indicate a negative correlation ($R^2 = 0.52$) between increasing radiation dose and the number of amplifiable aDNA molecules (Fig. 2).

Effect of increasing X-ray dose on aDNA molecule length. In living organisms X-ray radiation is known to induce double strand breaks (DSBs)²². Average DNA molecule length is thus assumed to become shorter after radiation exposure. To test if the same effect applies to ancient molecules, we performed mitochondrial capture²³ and sequenced all enriched libraries from samples that were previously exposed to dose from 0 to 43.5 kGy. No DNA could be sequenced from the 1000 s (93.72 kGy) sample. Sequence data were pre-processed, filtered, and mapped to the mitochondrial reference sequences of the corresponding organisms. Between 4,208,541 and 2,425 fragments mapped to the mitochondrial DNA (mtDNA) reference sequences of the corresponding organism. Mean fragment lengths were determined for mapped (endogenous) and overall (including non-mapped) fragments, and samples were again normalised as described above. The calculated fragment length

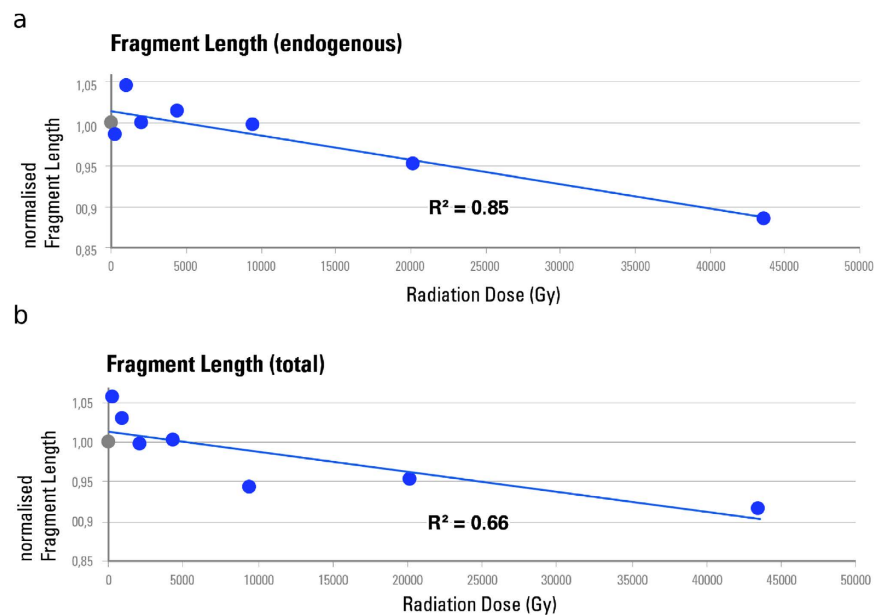


Figure 3. aDNA fragment lengths determined after irradiation from 0 to 43.5 kGy. Normalised mean fragment lengths are plotted against cumulative radiation dose received after 0 s (0 Gy), 2.2 s (~206 Gy), 10 s (~937 Gy), 21.5 s (~2.02 kGy), 46.4 s (~4.35 kGy), 100 s (~9.37 kGy), 215.4 s (~20.2 kGy) and 464.2 s (~43.5 kGy). The non-scanned control is represented as a grey circle, the scans as blue circles. **(a)** Mean fragment lengths are calculated for endogenous aDNA fragments only, as determined by mtDNA mapping. **(b)** Mean fragment lengths are calculated for total DNA including those fragments that did not map to the corresponding mitochondrial reference sequence. Values were normalised using the means of the corresponding non-scanned control aliquots, and averaged per exposure time group consisting of two aliquots from independent DNA-libraries.

values were plotted against radiation dose. We observe a strong correlation indicating an almost linear decrease of endogenous ($R^2 = 0.85$; Fig. 3a) and overall DNA ($R^2 = 0.66$; Fig. 3b) fragment length with increased radiation dose.

Since the observable fragment length can be biased through extraction, library preparation, capture and sequencing of the DNA, the decay constant λ from the exponential fragment length decline model by Allentoft *et al.*²⁴ was utilized to compute the unbiased average fragment length²⁵. Although the computed unbiased average fragment length is shorter than the mean fragment length, in both cases the general trend shows a decreasing fragment length towards a higher accumulated radiation dose (Supplementary Data 4).

Increasing X-ray dose and aDNA misincorporation patterns. Since DNA repair mechanisms are absent in dead organisms, chemical modifications of DNA will accumulate. The most common type of such DNA damage is the deamination of cytosines into uracils that causes a nucleotide misincorporation during amplification evident as a C to T substitution most prominently at the 5' terminus of sequenced aDNA molecules²⁶. It has been suggested that the amount of C to T substitutions at the 5' terminus accumulates over time, and that this can be used to authenticate ancient DNA^{27,28}. When DNA molecules get exposed to radiation strong enough to induce strand breaks, the proportion of molecules with terminal substitutions should theoretically decrease, contributing to an associated decrease in the overall C to T substitution frequency (Fig. 4a). A destructive effect of X-ray radiation can thus be inferred by reduced terminal substitution frequencies.

To measure the effect of different X-ray doses on nucleotide misincorporation patterns, C to T substitution frequencies were measured for the first position from the 5' end from the DNA sequence of the exposed aliquots. Data were normalised by the corresponding non-scanned control aliquots and plotted against radiation dose (Fig. 4b). Predictably, substitution frequencies correlate negatively with exposure time, suggesting that a higher X-ray dose lowers the amount of C to T nucleotide misincorporations. This is best explained by DNA strand breaks induced by radiation that introduce new 5' ends, thus lowering the overall proportion of molecules with terminal C to T substitutions.

Effect of synchrotron high quality imaging on aDNA. The results of the first two experiments show no representative effect below 200 Gy, and nearly no effect up to 2 kGy. In order to better assess effects in this dose range, we included three aliquots from cave bear specimens in a series of real μ CT scans performed with a classical high quality synchrotron setup. The total water equivalent surface dose delivered during this scan was 720 Gy, but with an average energy higher than for the previous experiments (127 keV instead of 75 keV). Compared to

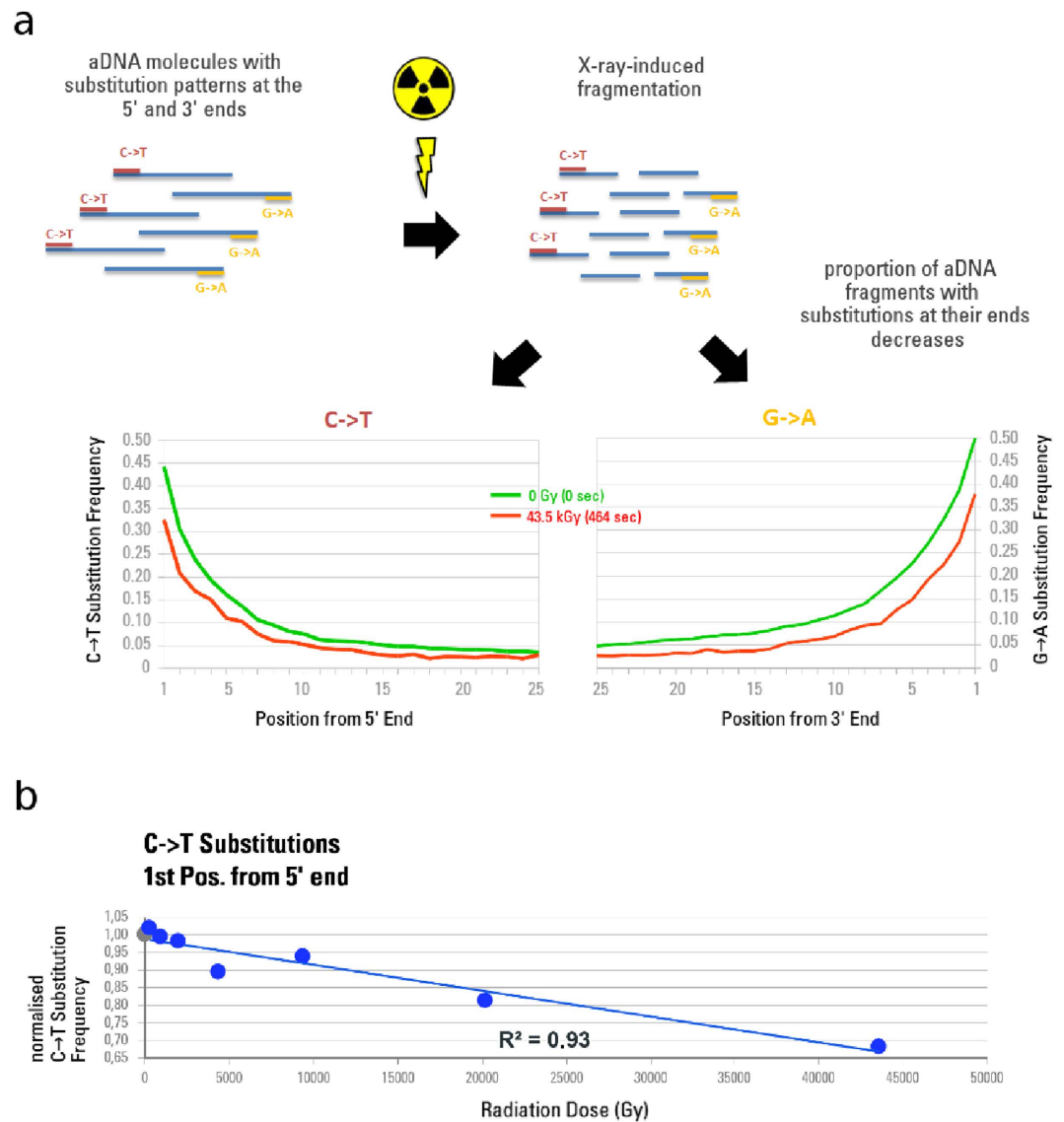


Figure 4. X-ray-induced effects on aDNA nucleotide misincorporation patterns for the first 25 positions from both ends of the molecule. (a) The misincorporation frequencies of the non-scanned controls are shown as a green line. Lower frequencies (red line) of C to T substitutions at the 5' ends and G to A substitutions at the 3' ends can be observed after irradiation at 43.5 kGy, since the fraction of aDNA fragments without terminal nucleotide misincorporations increases through exposure. **(b)** Normalised C to T substitution frequencies of the 1st position from 5' end of endogenous DNA molecules are plotted against cumulative radiation dose received after 0 s (0 Gy), 2.2 s (~206 Gy), 10 s (~937 Gy), 21.5 s (~2.02 kGy), 46.4 s (~4.35 kGy), 100 s (~9.37 kGy), 215.4 s (~20.2 kGy) and 464.2 s (~43.5 kGy). The non-scanned control is represented as a grey circle, the scans as blue circles. Substitution frequencies were normalised by the corresponding non-scanned controls, and averaged per exposure time group consisting of two aliquots from independent DNA-libraries.

their non-scanned controls, two out of three libraries showed an increase in the number of amplifiable library molecules after X-ray exposure (Fig. 5a).

Libraries from all three samples were captured for mtDNA, sequenced, and mapped to the cave bear mitochondrial reference. Since sample 21 remained with only 127 reads after the mapping step, it was not considered further as a threshold of at least 1000 reads was required for further analysis. No consistent trend could be detected for changes in mean fragment length between scanned and non-scanned fractions of samples 20 and 22 (Fig. 5c). We therefore applied the decay constant λ^{24} to calculate the unbiased average fragment length²⁵, which did not differ significantly between scanned and non-scanned aliquots. In contrast, all samples showed a decrease in mean fragment length as a result of scanning when total reads were considered, applying the same filtering criteria for fragment length and sequence quality (Fig. 5b). Except for sample 21 this behaviour was also represented by the λ -based calculated fragment length. The obtained mean, median and lambda-based fragment length values are shown in Supplementary Data 4. We also find that samples 20 and 22 have lowered C to T misincorporation frequencies for the 5' terminal nucleotide positions after scanning (Fig. 5d).

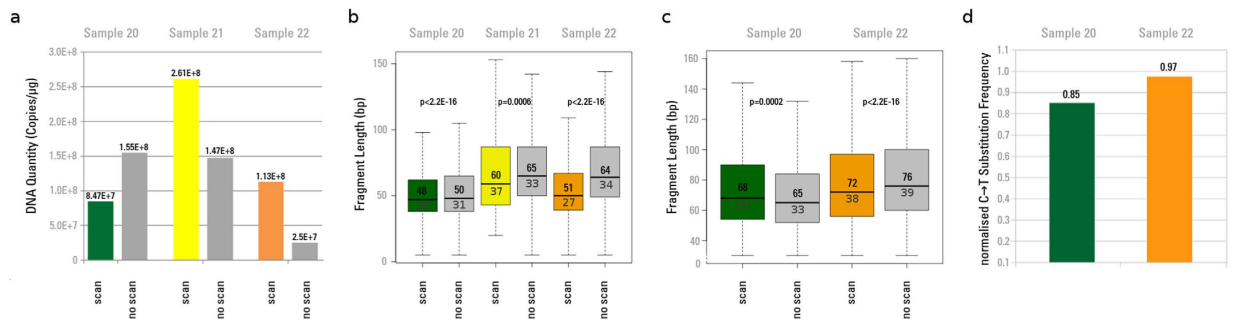


Figure 5. aDNA quantitation, fragment length and misincorporation pattern analyses after high quality imaging synchrotron scan. (a) aDNA quantitation after exposure to 720 Gy during a real synchrotron μ CT scan. “scan”: number of aDNA molecules after scanning the aliquot with 720 Gy, “no scan”: non-scanned control. Values were normalised by the amount of extracted material. (b) aDNA fragment lengths after exposure to 720 Gy (“scan”) and non-scanned controls (“no scan”). For each aliquot boxplots were generated from fragment lengths of the total DNA content including those fragments that did not map to the target organism’s (here: cave bear) mtDNA. P-values were obtained for each pair of scanned and non-scanned aliquots using Student’s t-test to assess significant differences in mean fragment lengths. The median fragment length is shown in the upper quartile and the computed λ -based average fragment length^{24,25} is shown in the lower quartile of the boxplot. (c) Boxplots were generated only from endogenous (mapped) aDNA molecule fragment lengths. The median fragment length is shown in the upper quartile and the computed λ -based average fragment length is shown in the lower quartile. Sample 21 was discarded because of a low number of mapped reads (<1000). (d) C to T substitution frequencies of the 1st position from the 5’ end of endogenous aDNA molecules mapped against cave bear mtDNA after X-ray exposure to 720 Gy. The values were normalised by their corresponding non-scanned controls.

surface dose (Gy)	aDNA quantity ratio	fragment length ratio endogenous DNA	fragment length ratio total DNA	C/T misincorporation ratio	general effect of X-ray irradiation on aDNA
0	~1	~1	~1	~1	no detectable effect
200	~1	~1	~1	~1	negligible damages
2000	0,85	~1	~1	0,97	acceptable damages
100000	0,8	~1	0,98	0,93	significant damages
200000	0,35	0,95	0,95	0,85	serious damages
45000	0,3	0,88	0,92	0,67	strong damages
170000	0,1	N.A.	N.A.	N.A.	no usable DNA left

Figure 6. Summary of effects of X-ray dose on aDNA quantity, molecule length and C to T misincorporation frequencies. Coloured lines represent the evaluated risk for the interval between two dose levels. Shown are normalised values corresponding to each applied X-ray surface dose. Normalisation was done by the corresponding non-scanned aliquot. A lower value indicates a more deleterious effect. No effect for dose below 200 Gy could be detected.

Evaluation of X-ray dose dependent aDNA degradation. Among those samples tested for different exposure times, we cannot observe any effect for doses below 200 Gy. The observed effects for doses between 200 and 2000 Gy are either not detectable or negligible. Above this dose level, we observe significant effects. Above that dose level, we observe more severe effects with a rapid decrease in amplifiable DNA, and dramatic effects for doses above 100 kGy that would make aDNA analysis likely impossible (Fig. 6).

Risk assessment for X-ray imaging and future aDNA analyses. We performed dosimetry of the most relevant configurations used for μ CT on sub-fossils with both conventional and synchrotron sources. Dosimetry on imaging systems was performed in three different steps: classical synchrotron experiments, low dose synchrotron experiments and conventional microtomographs.

All measurements and synthesis of results can be found in Supplementary Data 1 and Supplementary Data 2. By compiling all the measurements obtained for the conventional microtomographs, we propose a water equivalent surface dose estimator relevant for most of the conventional X-ray imaging systems used to image sub-fossils (Supplementary Data 3).

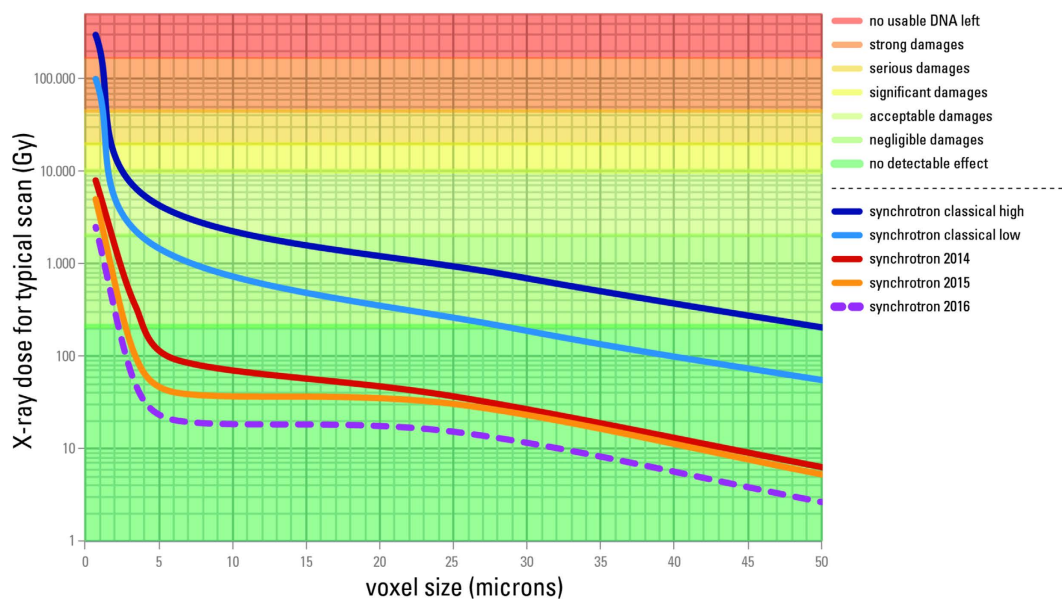


Figure 7. Synchrotron X-ray dose and associated aDNA damages. Typical water surface equivalent X-ray dose per scan and associated damages to aDNA depending on voxel size for synchrotron tomography configurations at the ESRF. Thanks to strong efforts to reduce dose since 2013, the average dose level was reduced by factor 30 in 2014, a factor 62 in 2015, and has reached a factor 125 in early 2016.

By combining the results from aDNA analysis and X-ray imaging devices dosimetry, it is possible to evaluate the risk level for future aDNA analyses for both synchrotron scanning and conventional μ CT scanning (Supplementary Data 2). From our results it is clear that classical synchrotron configurations can indeed reach dangerous levels for aDNA, especially when working at voxel sizes below $10\ \mu\text{m}$ (Fig. 7). Nevertheless, the new systems developed on the ID19-beamline at the ESRF since 2013 put all the configurations used for palaeoanthropology below the detection limit of any defect on aDNA, except the sub- μm resolution setup used for enamel microstructure (Supplementary Information, Note 1, Supplementary Figure 2).

Dosimetry experiments on the two conventional microtomographs at the Max Planck Institute for Evolutionary Anthropology (MPI-EVA) allowed for the assessment of corresponding dose levels for typical scans and to infer the potential risks of aDNA degradation (Fig. 8).

A bibliographic survey (Supplementary Table 1) as well as estimation of delivered doses for several specimens scanned at the MPI-EVA before aDNA sampling (Supplementary Information, Supplementary Table 2) show that in most cases the delivered doses are well below the detection limit of any defect on aDNA, and that samples that were scanned before aDNA sampling do not seem to be different in aDNA-retrieval success rate compared to non-scanned samples.

Discussion

X-ray-induced damage to aDNA is not yet well understood. Two mechanisms have been suggested. Oxidative damage, either directly caused by ionising radiation or mediated through water radicals, may lead to strand fragmentation or nucleotide modifications including hydantoin, which block DNA polymerases and hence prevent molecules from being amplified^{4,5,16,29,30}. Alternatively, X-ray-induced creation of free radicals such as hydroxide may lead to single and/or double strand breaks that may reduce DNA into smaller fragments^{14,16,22,29}. In the event of strand fragmentation, an apparent loss of DNA is expected as smaller fragments are assumed to be lost during the purification steps in the DNA extraction. Previous studies have attempted to assess the mechanisms of radiation-induced damage through PCR and electrophoresis-based analyses, and have generated contradictory results in terms of effects on DNA fragmentation^{19–21}. The samples used for these investigations, modern bones and preserved museum soft tissue such as skin specimens, may not be ideal proxies for evaluating the effect of radiation to aDNA in ancient bones, and although effects of X-radiation on aDNA were recently calculated *in-silico*¹⁸, the simulations were lacking support from real data obtained from authentic aDNA.

Here we used skeletal remains of late Pleistocene megafauna such as cave bear, giant deer and bison, as well as roe deer to further investigate X-ray-induced effects on authentic aDNA. In order to better control all the parameters of the beam and have results relevant for both synchrotron and conventional X-ray sources, we used a polychromatic synchrotron beam covering a large range of dose level and energy relevant for both kinds of source (Supplementary Information, Note 2, Supplementary Figures 3–5). We observed an almost complete loss of amplifiable aDNA in sample aliquots exposed to a high X-ray dose of 170 kGy compared to non-scanned controls. Furthermore, from our second experiment, we observed a decline in aDNA quantity as inferred from qPCR data that correlated with an accumulation in X-ray dose. An almost linear trend was observed in a decrease of mean aDNA fragment length with an increasing radiation dose for endogenous as well as for total DNA molecules. We also found a linear decrease in C to T substitution frequencies at the 5' end of the DNA fragments with

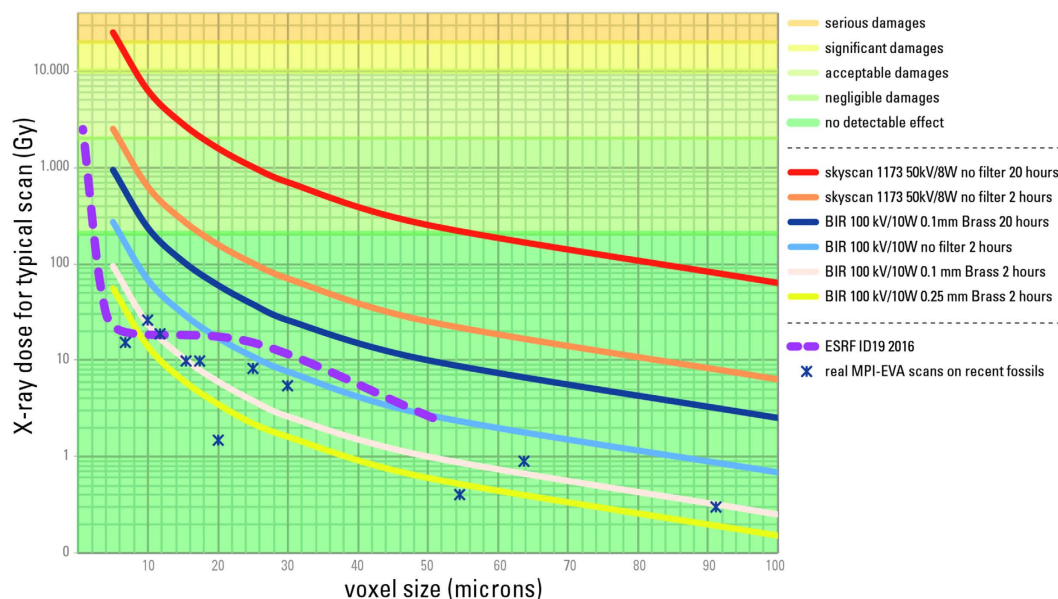


Figure 8. Conventional μ CT X-ray dose and associated aDNA damages. Typical water surface equivalent X-ray dose and associated degradation level of aDNA for conventional μ CT experiments depending on source parameters, filters, voxel size and scan duration. Plain curves represent typical scanning settings at the MPI-EVA and the ESRF. The dashed curve for ESRF ID19-beamline is calculated based on dose measurements to show the effect of using no filter or thin filter. Only scans without filters, especially repetitions of such scans, can lead to substantial degradation of aDNA. The dose profile of ESRF current configurations is given for comparative purpose.

increasing radiation dose. This is best explained by strand fragmentation, where new 5' ends created by radiation will have a lower frequency of the characteristic damage observed at the ends of ancient DNA fragments. Since this type of nucleotide misincorporation pattern is frequently used to authenticate aDNA molecules³¹, this trend is deserving of special attention, especially with regard to fossils that underwent a high level of irradiation during their scanning history. The observed damage seems to be proportional to a cumulative X-ray dose, an effect that was previously suggested for modern DNA³².

A large amount of DNA defects induced by X-rays in biological conditions are due to the production of extremely reactive free radicals mostly coming from water molecules. Without water, secondary chemical reactions due to free radicals should not take place when scanning dry fossils. Most likely, only direct interactions between X-rays and aDNA molecules (e.g. from direct beam and scattering effects) would have a significant impact. In their analysis of desiccated ancient soft tissues, Paredes *et al.* did not observe any damaging effects of μ CT scanning on DNA²¹, whereas Goetherstrom *et al.* and Grieshaber *et al.* purported to have seen X-ray-induced DNA-damage in fresh modern pig bones^{19,20}. Based on these data, the amount of water molecules most probably plays a significant role where a higher number of OH-radicals generated in non-desiccated material would ultimately contribute to an increased level of X-ray-induced DNA fragmentation³³. Strong support for this assumption is given through simulations provided by Wanek and Rühli who calculated a higher probability for radiation-induced aDNA damage to wet objects based on a higher radical yield, whereas radiation effects on dry objects were non-significant¹⁸.

In the case of three samples irradiated during a real tomography experiment (Experiment 3), we observed an apparent increase in aDNA molecules after X-ray exposure with a dose of 720 Gy at high energy in two out of three samples. This may be best explained by double strand breaks introduced by radiation that increase the number of molecules in our library that can be measured via qPCR post ligation of adapters. Even though the same amount of fluorescence signal should theoretically be detected if a molecule was split into multiple fragments, addition of the 132 bp of adapter to each molecule during library preparation will almost double the number of base pairs for a typical aDNA fragment that would be detected by SYBR Green dye intercalation, thus leading to a higher fluorescence signal. The increase in aDNA molecule number is further supported by an observed consistent decrease of mean fragment length for total DNA molecules. One of our samples, Sample 20, however, showed a decrease in aDNA quantity by about 60%. Since this sample derived from a different environmental context (different physicochemical factors) than the other two samples subjected to this test, preservation state of the sample, especially with regard to the level of mineralisation and the amount of water exposure, may influence the proportion of aDNA molecules that could be fragmented via the exposure to radiation. An initially higher level of fragmentation for this sample is suggested by its shorter mean read length in the non-scanned aliquot compared to Samples 21 and 22 (Fig. 5b). Further fragmentation due to radiation exposure could have shortened fragments to the point where they would be lost during the purification steps performed during aDNA extraction.

Despite this decrease in mean fragment length for total DNA, no consistent decline in mean length for endogenous captured cave bear mtDNA molecules could be detected. Furthermore, the λ -based average fragment length did not show any significant differences between scanned and non-scanned aliquots, and the aDNA specific C to T substitutions were only slightly lower in the scanned aliquots. Our lab protocols, implying purification steps and the application of mitochondrial capture prior to sequencing, as well as additional stringent filtering criteria applied to the sequences to retain only endogenous molecules, such as mapping to a reference and applying a minimum mapping-quality-filter, may have ultimately reduced the total number of DNA molecules and reads subject to analysis, with the shorter or heavily damaged fragments having been completely removed. Another possibility is that DNA stemming from ancient bone is better protected from ionising radiation than surface contaminants since endogenous DNA will have a higher binding affinity to hydroxyapatite and collagen^{16,34}. Hydroxyapatite forms a complex with DNA molecules³⁵, and components of the bony matrix might potentially absorb penetrating radiation, leading to a partial shielding. Higher fractions of DNA have been recovered from hydroxyapatite than from collagen emphasizing the importance of the mineral fraction of bone for the preservation of aDNA^{36,37}. Finally, endogenous aDNA is more protected from water than contaminating DNA, and therefore less affected by secondary damages due to free water-radicals.

The results for this configuration obtained during a high quality imaging synchrotron scan indicate that for surface doses below 2000 Gy, the effects on aDNA would not lead to substantial degradation and would not compromise palaeogenetic investigations.

Scans that are performed on sub-fossils with conventional machines are well below the detection limit of any defects on aDNA. The detection limit of 200 Gy for aDNA is 8 000 times higher than the highest dose for a medical scan and 20 million times higher than the dose exerted during luggage scans in airports. A comparison of X-ray dose levels achieved by various X-ray technologies is presented in Supplementary Table 3. Indeed, and for the sake of comparison, with the current X-ray devices in most airports, the total dose delivered during luggage scan is around 10 μ Gy (Supplementary Information, Note 3). Thus, only very long or repeated scans without any metallic filter with voxel sizes lower than 20 μ m can endanger aDNA integrity. High surface doses obtained for scans without any filters are due mostly to low energy X-rays. Since the lowest energies of the X-ray spectrum would enhance the beam hardening effect (stronger absorption of the lowest energies in the X-ray spectrum by the sample, leading to higher average energy of the X-rays behind the sample than before it, resulting into artefacts in the reconstructed data, see Supplementary Information, Note 1, Supplementary Figure 1), a filter of at least 0.1 mm of copper or brass should be used by default for any scan of sub-fossils, thereby providing a higher data quality, while avoiding aDNA-degradation even in case of multiple scans.

It should be noted that dose values used for dosimetry and risk assessment presented in this study were measured for the ESRF configurations on the ID19-beamline, where much optimisation has been performed during the last 15 years to image fossils. Furthermore, additional optimisation has since been performed to reduce X-ray dose for sub-fossil imaging. These optimisations do not reflect what is available at other synchrotron light sources. Careful dosimetry, setup and spectrum optimisation should therefore be performed on other beamlines before scanning sub-fossil remains that could be potentially subject to aDNA studies.

Conclusion

In summary, our results confirm that X-ray irradiation of sub-fossils can have a detrimental effect on aDNA integrity when the total water equivalent surface dose exceeds 200 Gy; the degradation increase being roughly linear with the dose accumulation. While the effect is very limited up to 2 kGy, the degradation can reach dramatic levels for doses exceeding 100 kGy. Based on the results presented above, we have defined the detection limit at 200 Gy due to the limited number of samples irradiated in the 200–720 Gy dose range. Nevertheless additional experiments would be necessary to define this limit more precisely. In order to ensure safe scanning, we therefore suggest the following guidelines for the scanning of sub-fossil remains:

Recommendation for conventional μ CT of sub-fossils.

- Never perform scans without a metallic filter. Always use filters to remove lowest energies of the spectrum. We recommend a systematic use of at least 0.1 mm of copper or brass filter for scanning of sub-fossils. Such a filter will ensure safe scanning while increasing data quality by reducing the beam hardening.

Recommendations for synchrotron scanning of sub-fossils.

- As synchrotron scanning can clearly lead to aDNA degradation, careful dosimetry has to be performed before any real experiment on a synchrotron beamline involving scanning of sub-fossils. Only high optimisation effort can put synchrotron setup in the safe dose region.
- Perform dose estimation/measurements before doing the actual scan. Take into account overlap of scans.
- Phase contrast being up to 1000 times more sensitive than absorption one, always use it to optimise results while keeping the dose as low as possible.
- Always perform a precise collimation of the beam (beam size fitting to the field of view) to ensure that a dose as low as possible is delivered to non-imaged areas.
- For sub- μ m resolution scans, orientate the samples carefully to minimise the amount of material crossing the beam path.

Recommendations for sub-fossil scanning in general.

- Never perform μ CT imaging on wet specimens, such as frozen samples, and samples that were recently cleaned with water.
- Do not put sub-fossil samples into water (e.g. to reduce beam hardening) when performing a μ CT scan.
- Do not perform scans at higher resolution than necessary, as the dose increases roughly at the square power of the increase of resolution.
- Dose is cumulative: do not perform multiple scans when similar data already exist. Data sharing (under responsibility of the curators) and public repository databases play an important role on this.
- Record precise information about all scanning parameters and scanning geometry and provide them to curators in order to track the whole scanning history for a given specimen.
- Before any new scanning of a specimen, take into account its scanning history to estimate the total dose accumulation (especially in the case of synchrotron experiments).

These simple recommendations can ensure that both conventional and synchrotron X-ray scanning of sub-fossils will not hinder future aDNA investigations and still allow studying the morphological or microstructural information by X-ray imaging before destructive sampling.

Material and Methods

Samples. Since sub-fossilised human remains were too rare to be used for this study, we collected animal samples considered old enough to be representative of ancient human bones, such as Neandertals. In total, a maximum of 38 samples from ancient faunal remains could be obtained. The selected species represent cave bear (*Ursus spelaeus*), giant deer (*Megaloceros giganteus*), roe deer (*Capreolus capreolus*) and steppe bison (*Bison priscus*) which were obtained from the Swabian Jura, the French Jura and from the Ukraine. The samples were UV-irradiated overnight to remove surface contamination, crushed into 0.5–2.0 mm² fragments and subdivided into 200 mg aliquots. To ensure non-biased results, a double-blind approach was used: no information about aliquot origin and preservation state was given to the persons in charge of the synchrotron scanning. As well, no information concerning the scanning settings and the organisation of scanned aliquots and control aliquots was available for the persons in charge of the aDNA extraction and quantitation.

X-ray exposure experiments and dosimetry. *Synchrotron (ESRF).* All X-ray exposure experiments were conducted on the BM05 beamline at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. The dose rate for the various setups was measured using a PTW Unidos E (T10021) dosimeter equipped with a TM31010 semiflex ionisation chamber. Measurements were conducted as surface water-equivalent dose, the ionisation chamber sensitive part being completely covered by the X-ray beam. The total integrated dose for a complete scan was then calculated as the total exposure time multiplied by the dose rate (see Table 1).

In a first experiment, aliquots from 11 different samples were irradiated every 45 degrees over 360 degrees using a polychromatic beam (50 × 5 mm²) with the following settings: BM05 bending magnet white beam filtered with 3 mm of aluminium, 0.1 mm of copper and 0.1 mm of tungsten leading to an average energy of 75.85 keV (Supplementary Information, Note 2, Supplementary Figure 3); 200 mA in the storage ring, sample at 50 m from the source; 17 min/scan; each part of sample scanned twice. The measured dose rate was 83.41 Gy/s, yielding a total delivered dose at every position in the samples of 170.15 kGy.

In a second experiment, aliquots were scanned using a polychromatic beam (45 × 3.1 mm²) with the same settings as described above (except 1.5 mm of aluminium instead of 3 mm, Supplementary Information, Note 2, Supplementary Figure 4), but at different exposure times (0 s, 1 s, 2.2 s, 4.6 s, 10 s, 21.5 s, 46.4 s, 100 s, 215.4 s, 464.2 s, and 1000 s) and a measured dose rate of 93.72 Gy/s which yielded a maximum delivered X-ray dose of 93.72 kGy. For each different time of exposure, the delivered dose is as follows: 0 Gy, 93.72 Gy, 206 Gy, 431.11 Gy, 937 Gy, 2.02 kGy, 4.35 kGy, 9.37 kGy, 20.19 kGy, 43.5 kGy and 93.7 kGy. A fast shutter was used to ensure that the total exposure time was equally distributed over 360 degrees every 45 degrees.

In a third experiment, aliquots were placed in a plastic tube together with cave bear (*Ursus spelaeus*) mandibles that underwent high quality scanning for other research purposes. Each sample was scanned twice (50% overlap between consecutive vertical scans). Scanning parameters involved a polychromatic beam at an average energy of 127 keV, obtained by filtering the bending magnet beam by 23 mm of aluminium and 6 mm of copper (Supplementary Information, Note 2, Supplementary Figure 5). We used a propagation distance for phase contrast of 2.4 m, and a pixel size of 29.88 μ m. The total exposure time for each part of the samples was 447 s, with a measured dose rate of 1.61 Gy/s (dosimeter in the plastic tube), leading to a total delivered X-ray dose of 720 Gy.

Conventional μ CT scanners (MPI-EVA). The dosimetry characterisation for conventional μ CT-scanners was performed at the Max Planck Institute for Evolutionary Anthropology (MPI-EVA, Leipzig, Germany) on the two portable industrial μ CT-scanners: a BIR ACTIS 225/300 and a Skyscan 1173. The dose rate was measured using the same model of dosimeter as for the synchrotron measurements (PTW: UNIDOS Webline T10021, Ionisation chamber TM31010), with the same calibration protocol using a ⁶⁰Co radioactive source. For both scanners, the dose rate was measured in different conditions of irradiation, involving variation in power, voltage, absence or presence of metallic filter of various thickness and material. Ranges of parameters explained below include conditions where the dose rate was not measurable in some cases, but this clearly appears in Supplementary Data 1.

For the BIR, three sessions of measuring were run. Firstly, the dosimeter was fixed on the sample stage at 250 mm from the X-ray source, initially with no metallic filter and then with various filters (0.25 mm brass, 0.5 mm brass, 1 mm brass, 1 mm aluminium), the operators were then able to measure the dose rate for power

Primer	Sequence
Roe_Deer_mt1_forward	AAGCAAGGCACTGAAAATGC
Roe_Deer_mt1_reverse	TTGGTACAGGATAGGGTCTCC
Roe_Deer_mt2_forward	AACCGCACATGCATTGTAA
Roe_Deer_mt2_reverse	GGTTGTTTGCAGTGACGAGA
Roe_Deer_mt3_forward	CATCATGACCACAAGCTCCG
Roe_Deer_mt3_reverse	CGTGTGCTTGATACCAGCTC
Brown_Bear_mt1_forward	AGGTCTCGTCGACAGTCAAAT
Brown_Bear_mt1_reverse	AGTCCCTGCACCTGCTTCTA
Brown_Bear_mt2_forward	TAGAAGCAGGTGCAGGGACT
Brown_Bear_mt2_reverse	TGATGTTGAGGTCGGTGTGT
Brown_Bear_mt3_forward	CAGTAGCCCTGTTTCGTCACA
Brown_Bear_mt3_reverse	CAGATCGCTTAGGGTCCAA
Bison_mt1_forward	ACCGCGGTCATACGATTAAC
Bison_mt1_reverse	AATTGCGAAGTGGATTTGG
Bison_mt2_forward	ATGAGCCAAAATCCACTTCG
Bison_mt2_reverse	TGTATTGCGTCTGCTCGTC
Bison_mt3_forward	CGAATCCACAGCCGAACATAT
Bison_mt3_reverse	TATAAAGCACCGCCAAGTCC

Table 2. Forward and reverse primer pairs used to generate bait for targeted mtDNA enrichment.

ranging from 5 to 150 W and voltage from 50 to 200 kV. Secondly, for each filtering condition mentioned above, the dose rate was measured at 50 kV, 130 kV and 200 kV and placing the dosimeter at an increasing distance from the dosimeter from 20 mm to 500 mm. Last, the influence of distance to beam axis was tested at 130 kV, 50 W, 250 mm from the X-ray source, and with 0.5 mm of brass. This involved combinations of variations in altitude (z ranging from 0 to 100 mm) and in translation (from 0 to 100 mm). It has to be noted that 130 kV is the most commonly used voltage when scanning at MPI-EVA.

For the Skyscan 1173, the dosimeter was set at the centre of the rotation stage, first at 125 mm from the X-ray source (\Leftrightarrow 17.2 μ m, commonly used distances for scans at MPI-EVA), and measures were taken twice during irradiations without filter, with 1 mm aluminium, 0.25 mm brass, with a power ranging from 2 W to 8 W, and a voltage from 50 kV to 130 kV. Second, with the same range of power, the sample stage was moved from 50 mm to 250 mm to the X-ray source without filter (at 50 kV and 80 kV), 1 mm aluminium and 100 kV, and with 0.25 mm brass and 130 kV. Last, the dose rate was measured at 130 kV, with 0.25 mm brass, at 125 mm from the source, with a power 2–8 W, and with a translation in “z” from 0 to 60 mm.

aDNA Extraction and Preparation of Sequencing Libraries (University of Tübingen). Prior to extraction samples were UV-irradiated overnight. DNA extraction was carried out using 50 mg as starting material based on a guanidinium-silica based extraction method³⁸. For each sample a DNA library was prepared according to published protocols^{39,40} using 20 μ l of extract. Sample-specific indexes were added to both library adapters to allow differentiation between individual samples after pooling and multiplex sequencing⁴⁰. Indexed libraries were amplified in 100 μ l reactions in a variable number of 9 to 14 cycles to reach the amplification plateau, followed by purification over Qiagen MinElute spin columns (Quiagen, Hilden, Germany).

Quantitation of aDNA Amount. qPCR quantification of the libraries was carried out on a Roche LightCycler 480 using the DyNAmo HS SYBR Green qPCR kit by Thermo Scientific. qPCR primers and standards are described in Meyer and Kircher 2010³⁹. All DNA quantities were numbered as a number of copies - i.e. a number of fragments with an adapter on each end - per μ l of library. As there is no amplification during library preparation, variation in the number of copies between libraries is assumed to reflect the variation in the DNA quantity between extracts. The number of copies for each library was normalised by the amount of extracted material in order to prevent a bias due to sampling. Results were analysed using the Roche Lightcycler's 480 integrated software, Microsoft Excel and the R statistical software⁴¹. The obtained values were normalised by division through the corresponding value of the non-scanned control aliquot. Except for the last experiment, mean values were calculated per scan group consisting of two aliquots from independent sequencing libraries. See Supplementary Data 4 for normalised DNA quantitations in figures 1, 2 and 5a.

Enrichment of aDNA. Target enrichment of mtDNA was performed by capture of the pooled libraries using bait generated from roe deer (*Capreolus capreolus*), brown bear (*Ursus arctos*) and bison (*Bison bison*) mtDNA²⁵. The bait was generated by use of three primer sets (Table 2) designed with the Primer3Plus software package⁴². All extractions and pre-amplification steps of the library preparation were performed in clean room facilities and negative controls were included for each reaction. Globally, all criteria established for ancient DNA studies authenticity were respected.

Sequencing and Preprocessing of Raw Sequences. Indexed library pools were sequenced on the Illumina HiSeq 2500 platform with 2*100 + 7 + 7 cycles⁴⁰. Demultiplexing was performed by sorting all the

sequences corresponding to their p7 and p5 index combinations. Forward and reverse reads were merged into single sequences if they overlapped by at least 11 bp⁴³. Unmerged reads were discarded and merged reads were filtered for a length of at least 30 bp.

Mapping. The *preprocessed* sequences were mapped to one of the corresponding complete mitochondrial genome reference sequences: *Capreolus capreolus* (NC_020684.1), *Bison bison* (NC_012346.1) or *Ursus spelaeus* (FM177760.1). Mapping was performed using BWA⁴⁴ with seeding turned off and a reduced mapping stringency parameter “-n 5” to account for up to five mismatches in ancient DNA reads⁴⁵. Duplicate removal was performed on those reads that showed identical start and end coordinates only by using a custom software now integrated into the EAGER-pipeline⁴⁶. The produced BAM files were filtered for sequences with a mapping quality of at least 20. We set a threshold of at least 1000 mapped reads for an aliquot to be considered in further analyses.

Fragment Length Distribution Analysis. Read lengths were obtained from quality filtered BAM files using SAMtools⁴⁷. Read length means and boxplots were obtained using the R statistical software⁴¹. Significance testing was assessed using the unpaired Student’s t-test in R. Normalisation of the data was achieved by dividing the obtained mean values by the mean fragment length of the corresponding non-scanned control aliquots. Scatter plots for regression analysis were generated in Microsoft Excel. See Supplementary Data 4 for normalised mean fragment length values in figures 3a and 3b, as well as Supplementary Data 4 and Supplementary Data 5 for fragment length distributions in figures 5b and 5c.

Recalculating Fragment Length based on λ . Read length distributions obtained by SAMtools⁴⁷ were plotted and the decay constant λ was determined from fitting the declining part of the plot into power regression. The unbiased average fragment length was calculated by dividing $1/\lambda^{24,25}$. See Supplementary Data 4 for plotted normalised lambda fragment length values corresponding to figures 3a and 3b, and lambda fragment length values utilised in figures 5b and 5c.

Analysis of aDNA C to T Substitution Frequencies. C to T misincorporation frequencies typical of aDNA were obtained using mapDamage 2.0⁴⁸. Data was normalised by division through the corresponding value of the non-scanned control for each position from 5’ end. Scatter plots and bar plots were generated in Microsoft Excel. For the exposure time experiment, mean values were calculated per scan group consisting of two aliquots from independent sequencing libraries. See Supplementary Data 4 for the substitution frequencies in figures 4a, 4b and 5d.

References

- Hublin, J. J. *et al.* Suggested guidelines for invasive sampling of hominid remains. *J Hum Evol* **55**, 756–757 (2008).
- Richards, G. D., Jabbar, R. S., Horton, C. F., Ibarra, C. L. & MacDowell, A. A. Color changes in modern and fossil teeth induced by synchrotron microtomography. *Am J Phys Anthropol* **149**, 172–180 (2012).
- Tafforeau, P. & Smith, T. M. Nondestructive imaging of hominoid dental microstructure using phase contrast X-ray synchrotron microtomography. *J Hum Evol* **54**, 272–278 (2008).
- Grososky, A. J., de Boer, J. G., de Jong, P. J., Drobetsky, E. A. & Glickman, B. W. Base substitutions, frameshifts, and small deletions constitute ionizing radiation-induced point mutations in mammalian cells. *Proc Natl Acad Sci USA* **85**, 185–188 (1988).
- Teoule, R. Radiation-induced DNA damage and its repair. *Int J Radiat Biol Relat Stud Phys Chem Med* **51**, 573–589 (1987).
- Le Cabec, A., Tang, N. & Tafforeau, P. Accessing developmental information of fossil hominin teeth using new synchrotron microtomography-based visualization techniques of dental surfaces and interfaces. *Plos One* **10**, e0123019 (2015).
- Smith, T. M. *et al.* Dental ontogeny in pliocene and early pleistocene hominins. *Plos One* **10**, e0118118 (2015).
- Smith, T. M. *et al.* Dental evidence for ontogenetic differences between modern humans and Neanderthals. *Proc Natl Acad Sci USA* **107**, 20923–20928 (2010).
- Liber, H. L., Leong, P. M., Terry, V. H. & Little, J. B. X-rays mutate human lymphoblast cells at genetic loci that should respond only to point mutagens. *Mutat Res.* **163**, 91–97 (1986).
- Muller, H. J. Artificial Transmutation of the Gene. *Science* **66**, 84–87 (1927).
- Wolff, S. Radiation Genetics. *Annu. Rev. Genet.* **1**, 221–224 (1967).
- Bushong, S. C. *Radiologic Science for Technologists: Physics, Biology and Protection.* (Mosby, 2001).
- Jefferies, K. In *Introduction to Radiography and Patient Care* (eds A. M. Adler & R. R. Carlton) Ch. 9 (WB Saunders Company, 1999).
- Roots, R. & Okada, S. Estimation of life times and diffusion distances of radicals involved in x-ray-induced DNA strand breaks of killing of mammalian cells. *Radiat Res* **64**, 306–320 (1975).
- Clancy, S. DNA damage & repair: mechanisms for maintaining DNA integrity. *Nature Education* **1** (2008).
- Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).
- Mitchell, D., Willerslev, E. & Hansen, A. Damage and repair of ancient DNA. *Mutat Res.* **571**, 265–276 (2005).
- Wanek, J. & Ruhl, F. J. Risk to fragmented DNA in dry, wet, and frozen states from computed tomography: a comparative theoretical study. *Radiat Environ Biophys* (2016).
- Götherstrom, A., Fischer, C. & Linden, K. X-raying ancient bone: a destructive method in connection with DNA analysis. *Laborativ Arkeologi* **8**, 26–28 (1995).
- Grieshaber, B. M., Osborne, D. L., Doubleday, A. F. & Kaestle, F. A. A pilot study into the effects of X-ray and computed tomography exposure on the amplification of DNA from bone. *J. Archaeol. Sci.* **35**, 681–687 (2008).
- Paredes, U. M. *et al.* Micro-CT X-rays do not fragment DNA in preserved bird skins. *J. Zool. Sys. Evol. Res.* **50**, 247–250 (2012).
- Bradley, M. O. & Kohn, K. W. X-ray induced DNA double strand break production and repair in mammalian cells as measured by neutral filter elution. *Nucleic Acids Res.* **7**, 793–804 (1979).
- Maricic, T., Whitten, M. & Paabo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *Plos One* **5**, e14004 (2010).
- Allentoft, M. E. *et al.* The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings. Biological sciences/the royal society* **279**, 4724–4733, doi: 10.1098/rspb.2012.1745 (2012).

25. Deagle, B. E., Eveson, J. P. & Jarman, S. N. Quantification of damage in DNA recovered from highly degraded samples—a case study on DNA in faeces. *Frontiers in zoology* **3**, 11, doi: 10.1186/1742-9994-3-11 (2006).
26. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci USA* **104**, 14616–14621 (2007).
27. Krause, J. *et al.* The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**, 894–897, doi: 10.1038/nature08976 (2010).
28. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Paabo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *Plos One* **7**, e34131 (2012).
29. Hoss, M., Jaruga, P., Zastawny, T. H., Dizdaroglu, M. & Paabo, S. DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Res.* **24**, 1304–1307 (1996).
30. Pääbo, S. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc Natl Acad Sci USA* **86**, 1939–1943 (1989).
31. Stoneking, M. & Krause, J. Learning about human population history from ancient and modern genomes. *Nat Rev Genet* **12**, 603–614 (2011).
32. Egerton, R. F., Li, P. & Malac, M. Radiation damage in the TEM and SEM. *Micron* **35**, 399–409 (2004).
33. Folkard, M., Prise, K. M., Brocklehurst, B. & Michael, B. D. DNA damage induction in dry and hydrated DNA by synchrotron radiation. *J. Phys. B: At. Mol. Opt. Phys.* **32**, 2753–2761 (1999).
34. Mrevlishvili, G. M. & Svintradze, D. V. DNA as a matrix of collagen fibrils. *Int J Biol Macromol* **36**, 324–326 (2005).
35. Brundin, M., Figdor, D., Sundqvist, G. & Sjogren, U. DNA binding to hydroxyapatite: a potential mechanism for preservation of microbial DNA. *J Endod* **39**, 211–216 (2013).
36. Campos, P. F. *et al.* DNA in ancient bone - where is it located and how should we extract it? *Ann Anat* **194**, 7–16 (2012).
37. Schwarz, C. *et al.* New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. *Nucleic Acids Res.* **37**, 3215–3229 (2009).
38. Rohland, N. & Hofreiter, M. Ancient DNA extraction from bones and teeth. *Nat Protoc.* **2**, 1756–1762 (2007).
39. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc.* **2010**, pdb.prot5448 (2010).
40. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
41. R Development Core Team. R: A Language and Environment for Statistical Computing. *The R Foundation for Statistical Computing, Vienna, Austria* (2011).
42. Untergasser, A. *et al.* Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* **35**, W71–W74 (2007).
43. Kircher, M., Heyn, P. & Kelso, J. Addressing challenges in the production and analysis of illumina sequencing data. *BMC Genomics* **12**, 382 (2011).
44. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
45. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci USA* **110**, 15758–15763 (2013).
46. Peltzer, A. *et al.* EAGER: efficient ancient genome reconstruction. *Genome Biology* **17**, doi: 10.1186/s13059-016-0918-z (2016).
47. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
48. Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).

Acknowledgements

We thank the ESRF BM05 beamline to have provided access to beamtime, as well as Thierry Brochard for his help with the dosimetry experiments. The computational work was performed on the computational resource bwGRiD Cluster Tübingen funded by the Ministry of Science, Research and the Arts Baden-Württemberg and the Universities of the State of Baden-Württemberg, Germany, within the framework program bwHPC. We want to thank Marek Dynowski for providing access to the computational resource bwGRiD Cluster Tübingen and his technical support. We are grateful to Smiths Heimann GmbH, Andreas Frank, Arno Folkerts and Christian Rauth (Wiesbaden, Germany) for sharing information about inspection devices in airports, as well as to Anne Bonnin (Paul Scherrer Institut, Switzerland). We want to acknowledge Morten Allentoft for his help and advice on applying the λ -exponential fragment length decline model to our data. This research was partially funded by the European Research Council Starting Grant APGREID (to A.I., A.H., K.B. and J.K.). We want to thank James Yates for proofreading the manuscript and improving the language. We thank Cosimo Posth and Maria Spyrou for their helpful comments. We also wish to acknowledge Michel Toussaint for providing access to the Engis 2 Neandertal material. We are grateful to Svante Pääbo for his help and support at an earlier stage of this project.

Author Contributions

J.K., P.T., M.B., A.I. and A.L.C. conceived and designed the general research. A.I., M.B. and J.K. further designed the genetic analyses while P.T. and A.L.C. designed and conducted the irradiation-experiments at the ESRF. P.T., A.L.C. and H.T. designed the conventional scanner dosimetry measurements that were performed by A.L.C. and H.T. at the MPI-EVA. M.B., F.L. and V.J.S. performed mtDNA-extraction and the preparation of sequencing libraries at the University of Tübingen. N.J.C., S.C.M., A.B., G.P., D.G.D., M.-A.J. and O.K. provided samples. B.V., K.H., P.T. and J.-J.H. facilitated the research and provided access to infrastructure. A.I. and A.H. performed bioinformatic analyses at the MPI-SHH. A.I., P.T. and J.K. wrote the manuscript and A.L.C. and K.B. contributed to it. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

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

How to cite this article: Immel, A. *et al.* Effect of X-ray irradiation on ancient DNA in sub-fossil bones - Guidelines for safe X-ray imaging. *Sci. Rep.* **6**, 32969; doi: 10.1038/srep32969 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016

EFFECT OF X-RAY IRRADIATION ON ANCIENT DNA IN SUB-FOSSIL BONES – GUIDELINES FOR SAFE X-RAY IMAGING.

SUPPLEMENTARY INFORMATION

Authors

Alexander Immel^{1,2*}, Adeline Le Cabec^{3,4*}, Marion Bonazzi^{5*}, Alexander Herbig¹, Heiko Temming³, Verena J. Schuenemann², Kirsten I. Bos¹, Frauke Langbein², Katerina Harvati⁶, Anne Bridault⁷, Gilbert Pion⁸, Marie-Anne Julien⁹, Oleksandra Krotova¹⁰, Nicholas J. Conard¹¹, Susanne C. Münzel¹², Dorothée G. Drucker¹³, Bence Viola^{14,3}, Jean-Jacques Hublin³, Paul Tafforeau^{4#} and Johannes Krause^{1,2,11#}

* These authors contributed equally, # corresponding author for X-ray imaging and dosimetry: paul.tafforeau@esrf.fr, # corresponding author for aDNA analysis: krause@shh.mpg.de

Affiliations

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

²Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen, Germany;

³Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany;

⁴European Synchrotron Radiation Facility, Grenoble, France;

⁵Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany;

⁶Senckenberg Center for Human Evolution and Palaeoecology, Palaeoanthropology, University of Tübingen, Tübingen, Germany

⁷CNRS UMR 7041 ArScAn, Equipe Archéologies environnementales, F-92023 Nanterre Cedex, France;

⁸Association départementale pour la recherche archéologique en Savoie, F-73230 Saint-Alban-Laysse, France;

⁹Centre for the Archaeology of Human Origins, Archaeology Department, University of Southampton;

¹⁰Department of Stone Age, Institute of Archaeology, National Ukrainian Academy of Science, Kiev, Ukraine;

¹¹Senckenberg Center for Human Evolution and Palaeoenvironment, University of Tübingen, Tübingen, Germany;

¹²Institute for Archaeological Sciences, Archaeozoology, University of Tübingen, Tübingen, Germany;

¹³Department of Geosciences, Palaeobiology, University of Tübingen, Tübingen, Germany;

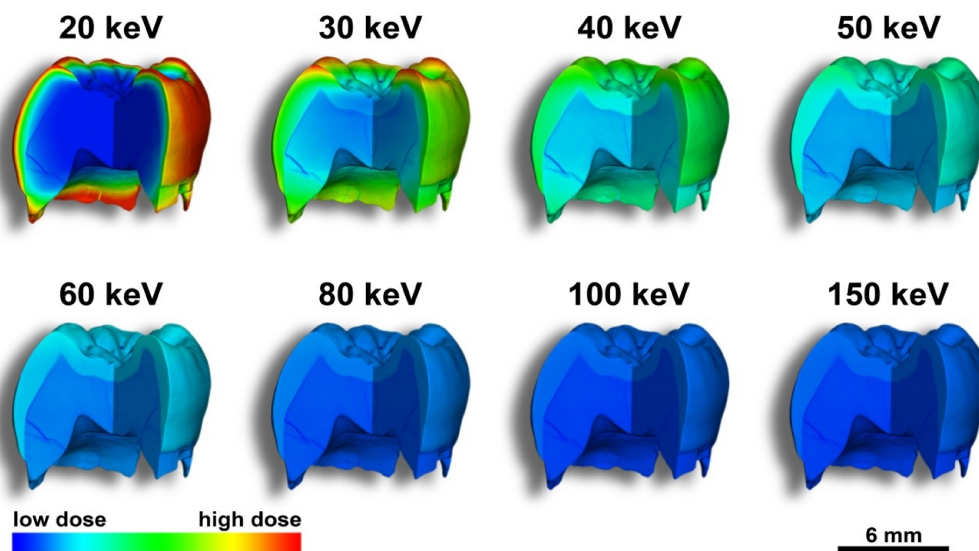
¹⁴Department of Anthropology, University of Toronto, Toronto, Canada;

Supplementary Note 1 – X-ray imaging and dose deposition

The first critical aspect to assess X-ray dose is linked to the voxel size of the scan (often presented as resolution, even if this term is not really adapted). Synchrotron μ CT is using parallel beam geometry; hence the resolution and magnification are obtained thanks to the X-ray detector. In most of the cases, it is an indirect detector based on a scintillator screen (mostly single crystals for high quality phase contrast imaging) coupled to a CCD or CMOS camera through an optical device (e.g. microscope, photographic objectives, optical taper). For a constant X-ray spectrum, the X-ray dose necessary to perform a scan with a given voxel size depends mostly on the detector properties. By opposition, in conventional sources, the magnification effect is obtained only thanks to the conical geometry of the X-ray source, by displacing the sample along the X-ray cone. In this case, for constant X-ray spectrum, the dose necessary

for a scan of a given voxel size depends on the distance between the source and the detector. This major difference between synchrotron and conventional X-ray source makes possible to derive X-ray dose relatively easily for conventional sources from calibration points by geometric calculation, whereas it is necessary to test all the different detectors combinations (scintillator/optic/sensor) for the synchrotron configurations.

The second critical aspect for X-ray dose is linked to the X-ray spectrum. Low energy X-rays are more easily absorbed by the samples than high energy X-rays. Nevertheless, low energy X-rays can also bring higher contrast level than high energy ones. Tomography requires that a sufficient amount of X-rays goes through the sample (typically the lowest transmission has to be above 10%, but often results are better for minimum transmission above 20%). In case of broad X-ray spectrum, the low energies can be completely absorbed by the sample when higher energies can go through (Supplementary Fig. 1). This very well-known effect is called beam hardening, which is leading to typical artifacts when scanning with such broad spectrum. Nowadays, very efficient algorithms can correct most of these artifacts¹, but it remains that the low energy photons are depositing a large part of the dose as they can be totally absorbed. Adapted filtering of the source spectrum using adapted metallic filter allows removing the lowest energies in order that the X-rays used to scan the specimen are really useful to obtain the data.



Supplementary Figure 1. 3D simulation of dose deposition pattern in a fossil molar depending on the photons energy (lower first molar of the Engis 2 Neandertal child from Belgium, scan was originally published in Smith *et al.* 2010², data deposited on the ESRF public database <http://paleo.esrf.eu>). Calculations are based on a constant amount of photons for the different energies, i.e. not for constant surface dose. Low energies are completely stopped in the superficial layers of the sample and will then deposit high dose level in surface and subsurface, when high energies go through the sample in a more uniform way.

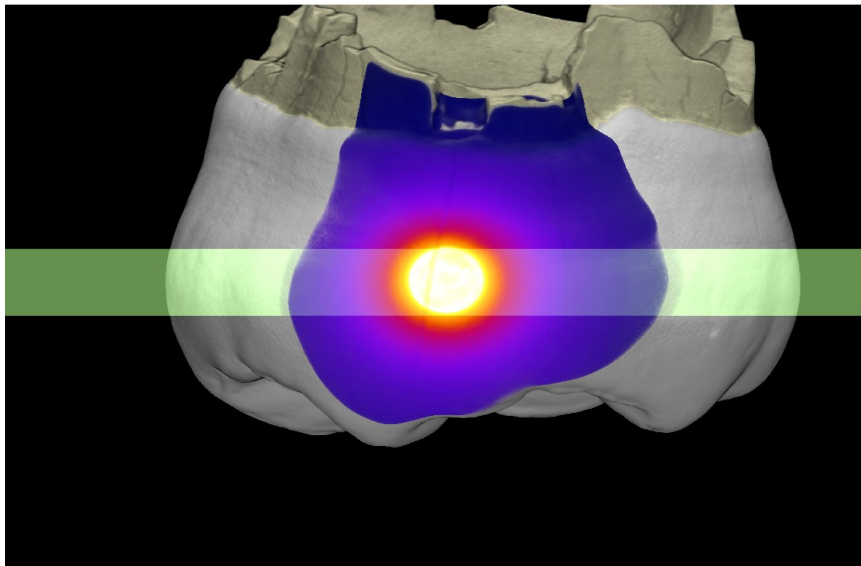
When performing X-ray tomography, it is then important to balance all these effects to find the optimal configurations allowing the narrower X-ray spectrum, with energy high enough to reach sufficient transmission through the sample without going to too high energy that would tend to reduce contrast.

Synchrotrons are well known to allow imaging using monochromatic beam, that remove *de facto* the beam hardening effect^{1,3}. Nevertheless, narrow polychromatic beam can bring results very close, without visible beam hardening, while allowing less ring artifacts (better beam profile and stability), and faster scans.

Since 2011, all the fossil specimens scanned on the beamline ID19 at the ESRF are imaged using these high quality direct “pink” (meaning narrow spectrum polychromatic) beams^{4,6}. It has to be noted that the use of high quality pink beam or of monochromatic beam does not really have an impact on the delivered dose as in both cases the average energy is similar as well as the dynamic level on the detector.

Imaging of fossils with synchrotrons also implies in nearly all the cases the use of propagation phase contrast. This technique can be up to 1000 times more sensitive to small density differences than X-ray absorption used in conventional systems, and is often giving better results with energies higher than those used for absorption. Phase contrast is nowadays the most important reason why using synchrotron sources to image fossils, especially for observation of small structures such as incremental lines in teeth or bones microstructures. It is also the key to reduce the X-ray dose for recent fossils by using as much as possible the high sensitivity given by this approach. All the results presented in the present paper for low dose synchrotron imaging are then based on propagation phase contrast, and would not be relevant for pure absorption imaging.

Classical sub- μm resolution configurations could even reach the level of total destruction of aDNA, but these configurations are restricted to small irradiated volumes (Supplementary Fig. 2) thanks to the precise beam collimation of synchrotron sources, i.e. the possibility to adapt the beam size to the field of view using absorbing slits systems.



Supplementary Figure 2. Typical 3D dose deposition pattern of a sub- μm resolution scan for enamel microstructure as performed at the ESRF (3D simulation performed on the same lower first molar of the Engis 2 Neandertal child than in Supplementary Fig. 1). Only the central yellow part can reach the high dose level reported in the present paper, the dose rapidly decreases for all the other parts crossed by the beam during the scan depending on the distance to the imaged part. Beam scattering (not simulated here) will also contribute to general dose level, but high resolution dosimetry experiments shows that it remains far less important than the dose deposition due to the direct beam, and its contribution decreases very rapidly with the distance to the direct beam (negligible after typically 200 μm in the geometry presented here).

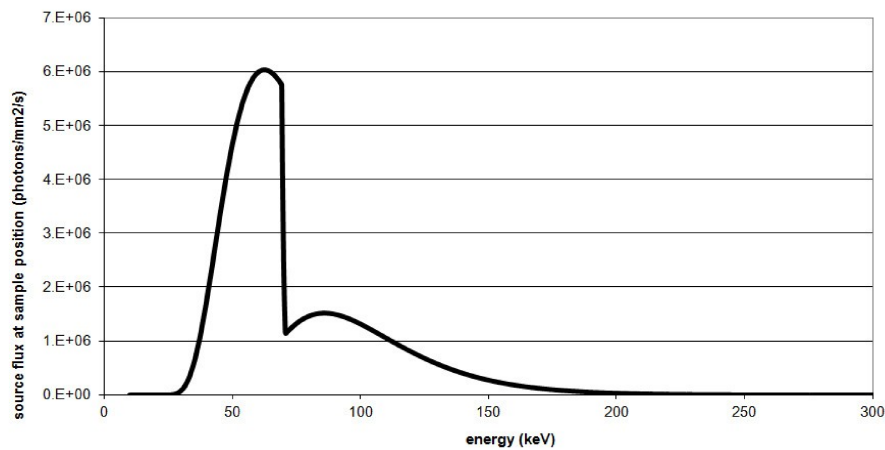
Low resolution scans (voxel size larger than 20 μm , implying full irradiation of the specimens) were typically in the safe zone even with classical configurations, but the new configurations are well below the detection limit. The most detrimental configurations with the classical synchrotron scans were in the 5 μm range, where large areas were scanned while having substantial level of dose (typically complete teeth for dental development), but these scans would not have really endangered aDNA studies, except in case of multiple scans. The highest efforts for dose reduction were then

applied in this resolution range, as it was potentially the most dangerous one, and as it is a critical one for virtual dental and bones palaeohistology. The new configurations implemented on ID19 are now well below the detection limit. Nowadays, only the sub- μm resolution scans typically used to observe enamel microstructures can still reach dose level that could have limited consequences on aDNA, but due to the small beam size, the concerned areas are very limited (typically small cylinders of $4 \times 2 \text{ mm}$), and concern mostly enamel, where no sampling for aDNA would be done anyway. All in all, the complete set of configurations available at the ESRF for scanning of recent fossils can be considered as safe for future aDNA studies, as long as good care is taken to perform the experiments (especially avoiding multiple scans whenever they are not necessary). Further efforts are ongoing to decrease dose for the sub- μm setup by further factor 2 to 3, and setups for voxel sizes larger than $10 \mu\text{m}$ by factor 2 to 10.

In the case of conventional scanner imaging, it has to be noted that the high surface dose obtained for scans without any filters are due mostly to low energy X-rays. This effect was especially visible with the skyscan1273 scanner for which dose rates without filters were really higher than expected after the experiment on the BIR scanner. Even if the cause of this higher dose is not really clear, it appears to be due to low energy part of the spectrum as even a thin aluminium filter can completely remove it. Hence, it could lead to substantial aDNA degradation in sub-surface of a specimen, but not in depth as the specimen itself would act as a filter and stop these low energy X-rays.

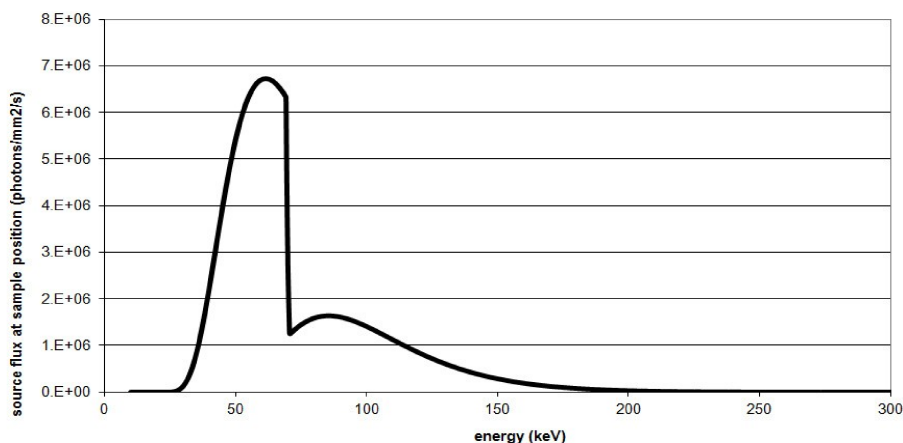
Supplementary Note 2 – X-ray spectrums used for irradiation and imaging experiments

Experiment 1, extreme irradiation



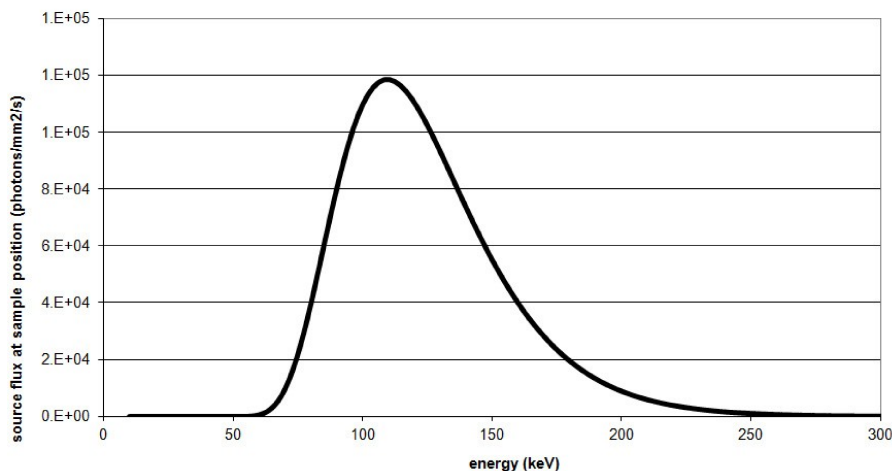
Supplementary Figure 3. X-ray spectrum used for the extreme irradiation experiment on the beamline BM05 using the white beam produced by the 0.85 tesla bending magnet filtered with 3mm of aluminium, 0.1 mm of copper and 0.1 mm of tungsten. Sample at 50 m from the source. Dose rate of 83.41 Gy/s.

Experiment 2, exposure time series



Supplementary Figure 4. X-ray spectrum used for the time series irradiation experiments on the beamline BM05 using the white beam produced by the 0.85 tesla bending magnet filtered with 1.5 mm of aluminium, 0.1 mm of copper and 0.1 mm of tungsten. Sample at 50 m from the source. Dose rate of 93.72 Gy/s.

Experiment 3, high quality classical synchrotron tomography



Supplementary Figure 5. X-ray spectrum used for the high quality classical synchrotron tomography experiment on the beamline BM05 using the white beam produced by the 0.85 tesla bending magnet filtered with 23 mm of aluminium and 6 mm of copper. Sample at 50 m from the source. Dose rate of 1.61 Gy/s.

Supplementary References

- 1 Olejniczak, A. J., Tafforeau, P., Smith, T. M., Temming, H. & Hublin, J. J. Technical note: compatibility of microtomographic imaging systems for dental measurements. *American journal of physical anthropology* **134**, 130-134, doi:10.1002/ajpa.20615 (2007).
- 2 Smith, T. M. *et al.* Dental evidence for ontogenetic differences between modern humans and Neanderthals. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 20923-20928, doi:10.1073/pnas.1010906107 (2010).
- 3 Tafforeau, P. *et al.* Applications of X-ray synchrotron microtomography for non-destructive 3D studies of paleontological specimens. *Applied Physics A* **83**, 195-202 (2006).
- 4 Le Cabec, A., Tang, N. & Tafforeau, P. Accessing developmental information of fossil hominin teeth using new synchrotron microtomography-based visualization techniques of dental surfaces and interfaces. *PloS one* **10**, e0123019, doi:10.1371/journal.pone.0123019 (2015).
- 5 Sanchez, S., Ahlberg, P. E., Trinajstic, K. M., Mirone, A. & Tafforeau, P. Three-dimensional synchrotron virtual paleohistology: a new insight into the world of fossil bone microstructures. *Microscopy and microanalysis: the official journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada* **18**, 1095-1105, doi:10.1017/s1431927612001079 (2012).
- 6 Smith, T. M. *et al.* Dental ontogeny in pliocene and early pleistocene hominins. *PloS one* **10**, e0118118, doi:10.1371/journal.pone.0118118 (2015).

Supplementary Table 1

Specimen	Taxon	Scanning date prior to aDNA sampling	Anatomical element sampled for aDNA	aDNA sampling date	aDNA publications	Scanned / irradiated before aDNA sampling	No scan/ irradiation before aDNA sampling	Unlikely scanned/ irradiated before aDNA sampling	No record
1	Abri Pataud (France, 28-24.9 ky cal. BP)	<i>H. sapiens</i>	Was unlikely X-rayed or CT-scanned before aDNA sampling (Musée de l'Homme, Paris, 05/05/2015)	Inventory number (whole skeleton ID): 26233 et 26234-A? but no further record (Musée de l'Homme, Paris, 29/04/2015)	2002 (Musée de l'Homme, Paris, 04/05/2015)	1		1	
2	Aboriginal Australian (Australia, ~100ky)	<i>H. sapiens</i>	no scan (M. Rasmussen, 2009/06/2015)	Hair (donation by an Aboriginal man from S.W. Australia in the early 20th century).	Early 2010 (M. Rasmussen, 29/06/2015)	2	1		
3	Anzick-1 (Western Montana, USA, 11.1-10.7 BPky)	<i>H. sapiens</i>	none (S. Anzick, 23/12/2015)	Petrous portion of the temporal bone - left side (S. Anzick, 23/12/2015)	-	3	1		
4	Baigara 1 (Western Siberia, 8 ka)	<i>H. sapiens</i>	28/02/2011 at MPI-EVA.	Left talus	03/03/2011 (V. Slon, 26/05/2015)	(Anja Heinze et al., in prep.)	1		
5	Borgo Nuovo (Italy, 5.9-6.4 ky)	<i>H. sapiens</i>	none (G. Barbujani, 30/06/2015)	tooth/bone	no record.	4			1
6	Chagyrskaya 2 (Altai Mountains, end of MIS 4; ~60ky)	Neand.	08/05/2010 at MPI-EVA.	C1 (atlas)	08/03/2010 (V. Slon, 26/05/2015)	(MPI-EVA, in prep)	1		
7	Chagyrskaya 6	Neand.	August 2011 at the veterinary medicine school of Leipzig.	Mandible	06/09/2011 and 01/12/2011 at MPI-EVA.	Sawyer et al., in prep.	1		
8	Chagyrskaya – SP3393	Neand.	X-rays in Moscow (B. Viola, 08/01/2016) and µCT scan on 24/09/2014 at MPI-EVA.	Distal hand phalanges	19/11/2014 at MPI-EVA.	(Slon et al., in prep)	1		
9	Chagyrskaya – SP3394	Neand.	X-rays in Moscow (B. Viola, 08/01/2016) and µCT scan on 24/09/2014 at MPI-EVA.	Ulna	19/11/2014 at MPI-EVA.	5	1		
10	Cova del Gegant (Spain, first half of the Würm: ~100-40 kya)	Neand.	2 medical CT-scans on 06/04/2004 (R. Quam, 03/06/2015)	Mandible	1 or 2 years after the scans (E. Subiro, 12/01/2016)	6	1		
11	Cro-Magnon 3 (France, 28ky BP)	<i>H. sapiens</i>	Was unlikely X-rayed or CT-scanned before aDNA sampling (Musée de l'Homme, Paris, 05/05/2015)	Right femur 4328? (Musée de l'Homme, Paris, 05/05/2015)	2002 (Musée de l'Homme, Paris, 04/05/2015)	1		1	
12	Denisova 2 - SP3276 (Altai Mountains in southern Siberia, likely >50 kyBP)	Denisovan	X-rayed in 2004 (B. Viola, 08/01/2016), µCT scan on 10/02/2014 at MPI-EVA.	LLDM2	14/04/2014 at MPI-EVA.	(Slon et al., in prep)	1		
13	Denisova 3	Denisovan	Has been sampled before scanning (scan: 11/12/2009), then scanned (date not recovered) and sampled again at MPI-EVA.	HP5D	extractions on 10/08/2009 and 15/12/2009 at MPI (01/07/2015, V. Slon)	7,8,9,10	1	1	
14	Denisova 4 / DENISOVA_11.1G_2	Denisovan	25/01/2010 at MPI-EVA (before aDNA sampling).	ULM3	End of 01/2010 at MPI-EVA.	8,11	1		
15	Denisova 5	Neand.	26/01/2011 at MPI-EVA (before aDNA sampling).	Proximal foot phalanx (4 or 5)	03/2011 at MPI-EVA.	10,12	1		
16	Denisova 8/ DENISOVA_Gensample3	<i>Ursus sp./</i> Neand./ Denisovan?	07/02/2011 at MPI-EVA.	Molar	10/02/2011 (V. Slon, 26/05/2015)	11	1		
17	El Sidron SD 1253 (Spain, ~49kya)	Neand.	19/10/2006 (A. Rosas, 26/05/2015)	Long bone fragment	11/2006 (A. Rosas, 26/05/2015)	10,13,14	1		
18	El Sidron SD 441 = S11	Neand.	None (A. Rosas 26/05/2015; first scan on 22/12/2005)	URI2	25/02/2005 (A. Rosas 26/05/2015)	15,16,17	1		
19	El Sidron SD 1252	Neand.	30/09/2005 (A. Rosas, 26/05/2015)	Femur fragment	November 2005 (C. Lalueza-Fox, 28/06/2015)	13	1		
20	El Sidron SD-011	Neand.	None (A. Rosas, 26/05/2015)	Broken mandible fragment of Adult 3	27/10/2009 (A. Rosas, 26/05/2015)	18	1		
21	El Sidron SD-331c	Neand.	17/12/2005 (A. Rosas, 26/05/2015)	LLM1 of Adult 4	08/05/2009 & 16/06/2009 (A. Rosas, 26/05/2015)	18	1		
22	El Sidron SD-500	Neand.	20/12/2005 (A. Rosas, 26/05/2015)	LLP4 of Adolescent 3	27/10/2009 (A. Rosas, 26/05/2015)	18	1		
23	El Sidron SD-566	Neand.	None (A. Rosas, 26/05/2015)	URP3 of Adolescent 2	26/02/2004 (A. Rosas, 26/05/2015)	18	1		
24	El Sidron SD-634	Neand.	None (A. Rosas, 26/05/2015)	Right 5th metatarsal of the Infant	25/03/2010 (A. Rosas, 26/05/2015)	18	1		
25	El Sidron SD-753	Neand.	22/12/2005 (A. Rosas, 26/05/2015)	LLC of Adult 6	21/10/2008 (A. Rosas, 26/05/2015)	18	1		
26	El Sidron SD-763a	Neand.	19/12/2005 (A. Rosas, 26/05/2015)	LRP4 of Adolescent 1	27/10/2009 (A. Rosas, 26/05/2015)	18	1		
27	El Sidron SD-763b	Neand.	16/06/2005 (A. Rosas, 26/05/2015)	Fragment of right ulna of Juvenile 2	21/10/2008 (A. Rosas, 26/05/2015)	18	1		
28	El Sidron SD-1161	Neand.	None (A. Rosas, 26/05/2015; first scan on 22/12/2005)	URC of Adult 4	14/06/2005 (A. Rosas, 26/05/2015)	18	1		
29	El Sidron SD-1240	Neand.	None (A. Rosas, 26/05/2015)	URI2 of Adult 2	16/06/2009 (A. Rosas, 26/05/2015)	18	1		
30	El Sidron SD-1327h	Neand.	None (A. Rosas, 26/05/2015)	LLM1 of Adult 5	16/06/2009 (A. Rosas, 26/05/2015)	18	1		
31	El Sidron SD-1634	Neand.	None (A. Rosas, 26/05/2015; first scan on 15/03/2011)	Fragment of right femur of Juvenile 1	25/03/2010 (A. Rosas, 26/05/2015)	18	1		
32	Engis 2 (Belgium, Mousterian)	Neand.	Medical scanner in 1983 or 1984 and 1999 (M. Toussaint, 04/05/2015)	Left parietal	1994, Munich (M. Toussaint, 04/05/2015)	1	1		
33	Feldhofer 1 ("Neandertal 1") [Type specimen] (Germany, ~40 ky)	Neand.	several conventional X-rays, several clinical CT scans (1975 in Tübingen, 1993 in Göttingen) (R. Schmitz, 09/09/2015)	Right humerus	1996	19,20	1		
34	Feldhofer 1	Neand.	micro-CT scan in 2004 (R. Schmitz, 09/09/2015)	Right tibial shaft (NN4)	2009 (R. Schmitz, 09/09/2015)	21,22	1		
35	Feldhofer 2	Neand.	micro-CT scan in 2004 (R. Schmitz, 09/09/2015)	Right humerus (NN1)	Sampling for Schmitz et al 2002 before the 2004 scanning; sampling for Briggs et al. 2009 after the 2004 CT (R. Schmitz, 09/09/2015).	21,22	1	1	
36	Kent's Cavern (KC4) (United Kingdom, 43-42 ky cal H. sapiens BP)	<i>H. sapiens</i>	30/11/2006, NHM London (T. Higham, 11/05/2015; M. Fagan, 18/05/2015)	Dentine from URP3	2007 (T. Higham, 11/05/2015; M. Fagan, 18/05/2015)	23	1		
37	Kostenki 14 (K14) (Kostenki-Borshchevo, Russia, 38.7 to 36.2 ky)	<i>H. sapiens</i>	none (V. Moiseyev, 30/06/2015; 24/07/2015)	Dorsal side of the shaft of the left tibia.	End of 2013 (V. Moiseyev, 06/08/2015)	24	1		
38	La Braña 1 (La Braña-Arintero, Spain, 7 ky)	<i>H. sapiens</i>	Femur and skull + jaw radiographed at Hospital de León in March 2007, skull radiographed in Feb. 2009 at University Hospital of Granada (J. M. Vidal Encinas, 08/07/2015)	roots of ULM3 and URM3, one fragment of cortical bone from femur	18/04/2013 (C. Lalueza-Fox, 28/06/2015)	25	1		
39	La Chapelle-aux-Saints (France, ~60 ky)	Neand.	Was unlikely X-rayed or CT-scanned before aDNA sampling (Musée de l'Homme, Paris, 05/05/2015)	Vertebra? (A. Langaney, 06/05/2015), no further record (Musée de l'Homme, Paris, 29/04/2015)	2002 (Musée de l'Homme, Paris, 04/05/2015)	1		1	
40	La Ferrassie 1 (France, ~43-45 ky)	Neand.	Was unlikely X-rayed or CT-scanned before aDNA sampling (Musée de l'Homme, Paris, 05/05/2015)	post-cranial element? no further record (Musée de l'Homme, Paris, 29/04/2015)	2002 (Musée de l'Homme, Paris, 04/05/2015)	1		1	

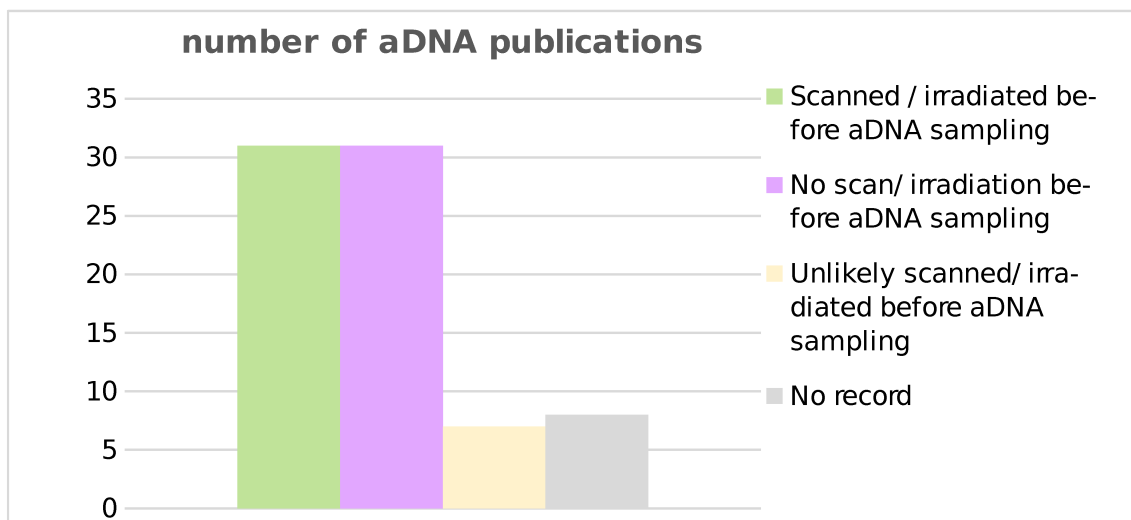
41	La Ferrassie 2 (France, ~43–45 ky)	Neand.	Was unlikely X-rayed or CT-scanned before aDNA sampling (Musée de l'Homme, Paris, 05/05/2015)	Left distal hand phalanx ? (Musée de l'Homme, Paris, 05/05/2015)	2002 (Musée de l'Homme, Paris, 04/05/2015)	1		1			
42	La Madeleine (adult) (France, ~13 ky)	<i>H. sapiens</i>	Was unlikely X-rayed or CT-scanned before aDNA sampling (Musée de l'Homme, Paris, 05/05/2015)	post-cranial element ? but no further record (Musée de l'Homme, Paris, 29/04/2015)	2002 (Musée de l'Homme, Paris, 04/05/2015)	1		1			
43	La Quina 5 (France, ~65 ky)	Neand.	Was unlikely X-rayed or CT-scanned before aDNA sampling (Musée de l'Homme, Paris, 05/05/2015)	Left and right humeri? but no further record (Musée de l'Homme, Paris, 29/04/2015)	2002 (Musée de l'Homme, Paris, 04/05/2015)	1		1			
44	Les-Rochers-de-Villeneuve (France, 40.7 ky BP)	Neand.	none (C. Beauval, 14/04/2015)	RdV 1: fragment of left femur diaphysis.	2004 or 2005	26	1				
45	Loschbour (Luxembourg, ~8 ky)	<i>H. sapiens</i>	None, (CT scan of the skull on 16/06/2009 whereas URM1 had already been taken out; D. Delsate & F. Lebrun, 02/07/2015)	'Tooth 16': URM1 from a male skeleton	10/06/2009 (D. Delsate, 02/07/2015)	27	1				
46	Mal'ta - MA 1 (South-central Siberia, 24 ky)	<i>H. sapiens</i>	likely X-rayed	Right humerus	?	28	1				
47	Mezmaiskaya 1 (Northern Caucasus, 60-70 ky)	Neand.	none (L. Golovanova, 16/05/2015)	Two small fragments of a rib (L. Golovanova, 16/05/2015)	1998-1999 (L. Golovanova, 16/05/2015)	29	1				
48	Mezmaiskaya 1	Neand.	May-June 2003: Medical CT-scan of the Mezmaiskaya 1 skeleton, St. Petersburg (L. Golovanova, 16/05/2015)	A small rib fragment of Mezmaiskaya 1 (L. Golovanova, 16/05/2015)	Brought in January 2006 to MPI-EVA (L. Golovanova, 16/05/2015); Extractions on 23/11/2006 and June 2008 (S. Sawyer, 13/04/2015)	14,22	1				
49	Mezmaiskaya 1	Neand.	May-June 2006: Second CT-scanning of the Mezmaiskaya 1 skeleton MPI-EVA portable micro-CT scanner, St. Petersburg (L. Golovanova, 16/05/2015)	A small rib fragment of Mezmaiskaya 1 (L. Golovanova, 16/05/2015)	Brought in 2009 to MPI-EVA (L. Golovanova, 26/05/2015); extraction on 30/03/2011 (S. Sawyer, 13/04/2015).	12,30	1				
50	Mezmaiskaya 2 (Northern Caucasus, ~41 ky)	Neand.	May-June 2006: First CT-scanning of the Mezmaiskaya 2 skull; MPI-EVA portable micro-CT scanner, St. Petersburg (L. Golovanova, 16/05/2015)	An unidentified skull fragment of Mez. 2 (L. Golovanova, 26/05/2015)	Brought in January 2006 to MPI-EVA (L. Golovanova, 26/05/2015)	14,30	1				
51	Mezzocorona (Italy, 6.3–6.4ky)	<i>H. sapiens</i>	None (G. Barbujani, 30/06/2015)	tooth/bone	no record.	4		1			
52	Monte Lessini (Italy, 50 ky)	Neand.	None (D. Caramelli, 06/05/2015)	Skull fragment	2006 (D. Caramelli, 06/05/2015)	31	1				
53	Monte Lessini	Neand.	None (D. Caramelli, 06/05/2015)	Mezzena Jaw	2013 (D. Caramelli, 06/05/2015)	32	1				
54	Motala 12, site of Kanaljorden (Sweden, ~8 ky)	<i>H. sapiens</i>	None (F. Hallgren, 02/07/2015)	Mandibular bone and M2 (F. Hallgren, 02/07/2015)	November 2010 (F. Hallgren, 02/07/2015) 14 more samples from 13 specimens sampled in the same conditions.	27	1				
55	Paglicci 12 (25 kya, Italy)	<i>H. sapiens</i>	No record? (D. Caramelli, 26/05/2015)	Rib	2002 (D. Caramelli, 26/05/2015)	33		1			
56	Paglicci 25 (23kya, Italy)	<i>H. sapiens</i>	No record? (D. Caramelli, 26/05/2015)	Femur	2002 (D. Caramelli, 26/05/2015)	33		1			
57	Oase 1, Peștera cu Oase (Romania, ~37-42 ky)	<i>H. sapiens</i>	Medical CT scan in 2003 (S. Constantin, 07/07/2015)	Mandible	29/09/2009 at MPI-EVA (M. Hajdinjak, 24/06/2015).	34	1				
58	Okladnikov 7 (Altai Mountains, ~45 ky)	Neand.	none (B. Viola, 14/04/2015)	Subadult humerus shaft	late 2005 or early 2006 at MPI-EVA.	30,35	1				
59	Okladnikov 8	Neand.	none (B. Viola, 14/04/2015)	Subadult femur	late 2005 or early 2006 at MPI-EVA.	35	1				
60	Okladnikov 9 (~40 ky BP)	Neand.	09/12/2005 (B. Viola, 14/04/2015)	Adult humerus (distal half).	2010	- no aDNA recovered (taphonomy: shallow deposit)	1				
61	Saqqaq culture, site Qeqertarsussuk (North-western Greenland, ~4 ky)	<i>H. sapiens</i>	none (M. Rasmussen, 29/06/2015)	Four permafrost-preserved human hair tufts from individual Qt 86 85/261: 12	Early 2009 (M. Rasmussen, 29/06/2015)	36	1				
62	Scladina 1 (Belgium, ~127 ky)	Neand.	none (M. Toussaint, 13/04/2015)	Scla 4A-4: URM1 buccal root	1993 (Münich; (M. Toussaint, 13/04/2015)	- no aDNA (taphonomy?)					
63			none (M. Toussaint, 13/04/2015)	Scla-4A-13: LRdm2	Feb. 2001 (M. Toussaint, 14/04/2015)	37	1				
64			micro-CT scanned at MPI on 29/11/2006 (MPI scanning records), and at Antwerp University (date not recorded)	Scla 4A-4: URM1	2007 (MPI-EVA, M. Toussaint, 14/04/2015)	- no aDNA (taphonomy?)					
65	Sima de los Huesos – Femur XIII (AT-2944) (Spain, > 300ky)	<i>H. heidelbergensis</i>	CT-scanned prior to aDNA sampling (M. Meyer and J.-L. Arsuaga, 04/07/2015)	Right femur (sampling in the breaks of the 3 pieces)	30/11/2012 (M. Meyer, 04/07/2015)	38	1				
66	'Stuttgart' from Viesenhäuser Hof (Germany, ~7 ky)	<i>H. sapiens</i>	none (J. Krause, 01/07/2015)	'Tooth 47': LRM2 of a female skeleton (LBK380)	18/10/2012 (A. Mittnik, 07/08/2015)	27	1				
67	Tyrolean Iceman – Ötzi (Italy, 5.3 ky)	<i>H. sapiens</i>	≥ 2 classical radiographs, 2 CT scans in Innsbruck, one in Bolzano in 2005 (A. Zink, 16/12/2015)	Left ilium	Oct. 2009 (A. Zink, 16/12/2015)	39,40	1				
68	Teshik-Tash (Uzbekistan, 57-24 ky)	Neand.	Was X-rayed many times (as early as 1949). (B. Viola 14/04/2015)	Left femur	2003 or 2004 (J. Krause, 04/05/2015) at MPI-EVA.	35	1				
69	Ust'-Ishim 1 (Siberia, ~45 ky)	Neand.	15/05/2012 (MPI), was medical CT-scanned before coming to MPI as well. (B. Viola)	Left femur	after 15/05/2012 (B. Viola) at MPI-EVA.	41	1				
70	Valdegoba 1 (VB1) (Spain, 48.5 ky BP)	Neand.	24/04/1998, Madrid, medical CT scan of the mandible (R. Quam, 08/05/2015)	LLM1 (was <i>in situ</i> in the mandible).	22/05/2006 (L. Dalén, 07/05/2015; R. Quam, 08/05/2015)	42	1				
71	Villabruna (Italy, ~13.8–14.3)	<i>H. sapiens</i>	none (G. Barbujani, 30/06/2015)	tooth/bone	no record.	4		1			
72	Vindija 75 (Croatia, 28-29 ky BP)	Neand.	No record (J. Mauch Lenardić, 07/05/2015)	No record (J. Mauch Lenardić, 07/05/2015)	no record.	43		1			
73	Vindija77 (=Vi.77/255 G3, Vin #H51)	Neand.	No record (J. Mauch Lenardić, 07/05/2015)	No record (J. Mauch Lenardić, 07/05/2015)	no record.	1		1			
74	Vi 33.15	Neand.	25/07/2007 at MPI-EVA	Shaft splinter (femur?)	1st extraction on 25/07/2007 at MPI-EVA.	10	1				
75	Vi 33.16 (=Vindija 80=Vi.80)	Neand.	22/06/2006 and on 26/02/2007 at MPI-EVA (before aDNA sampling).	Fragment of tibia	no record.	1,14,22,44,45		1			
76	Vi. 33.25 (=Vi75 I)	Neand.	Highly likely that no scan before aDNA sampling (T. Maričić, 17/06/2015)	Elongated shaft splinter, morphology not diagnostic. (C. Verna, Feb. 2007)	09 and 17/01/2007, 26/03/2008, 17 and 25/06/2009 at MPI-EVA (T. Maričić, 17/06/2015).	14,22	1				
77	Vi. 33.26 (=Vi1977 G)	Neand.	Highly likely that no scan before aDNA sampling (T. Maričić, 17/06/2015)	Long shaft splinter compatible with the latero-posterior face of a human femur (C. Verna, Feb. 2007)	10/01/2007, 18/02/2008, 17/06/2009 at MPI-EVA (T. Maričić, 17/06/2015)	14	1				
TOTAL							77	31	31	7	8

Supplementary Table 1:

A bibliographic survey showing publications with record of CT-scanned and non scanned sub-fossil remains from 1949 to 2016. Shown is the specimen, the taxon, the scanning date prior to aDNA sampling (if applies), the scanned anatomical element, sampling date for aDNA extraction, and the corresponding publication. Entries showing specimens that were scanned before aDNA sampling (31 in total) are shown in green, the unscanned Sub-fossils (also 31 in total) are shown in purple. In 7 cases it was unlikely that a scan was performed before aDNA sampling (indicated in beige), and in 8 cases no record exists (shown in grey). 50 % of all sub-fossil remains listed here were scanned prior to aDNA sampling.

Acknowledgements of curators and researchers that helped gathering this information:

Michel Toussaint, Viviane Slon, Aurélie Fort, Alain Froment, Antoine Balzeau, Mario Cech, Philippe Menecier, André Langaney, Antonio Rosas, Almudena Estalrich Albo, Antonio García Tabernero, Liubov Golovanova, Vladimir Doronichev, Tomislav Maričić, Christine Verna, Philip Nigst, Marta Mirazon Lahr, Vyacheslav Moiseyev, Eske Willerslev, Morten Rasmussen, Kelly Graf, Jean-Jacques Cleyet-Merle, Love Dalén, David Caramelli, Mateja Hajdinjak, Ralf Schmitz, Jadranka Mauch Lenardić, Susanna Sawyer, Cédric Beauval, Annamaria Ronchitelli, Rolf Quam, Maria Eulàlia Subirà i de Galdàcano, Thomas Higham, Beth Shapiro, Chris Stringer, Michael Fagan, Stefano Benazzi, Francesco Mallegni, Carles Lalueza-Fox, Juan Luis Arsuaga, Manuel Antonio Garcia Garrido, Matthias Meyer, Albert Zink, Iosif Lazaridis, David Reich, Guido Barbujani, Fredrik Hallgren, Dominique Delsate, Silviu Constantin, Matthias Meyer, Julio Manuel Vidal Encinas, María Encina Prada Marcos, Vyacheslav Moiseyev, Sarah Anzick.



1. Serre, D. *et al.* No evidence of Neandertal mtDNA contribution to early modern humans. *PLoS Biol* **2**, 313–317 (2004).
2. Rasmussen, M. *et al.* An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science* **334**, 94–98 (2011).
3. Rasmussen, M. *et al.* The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225–229 (2014).
4. Di Benedetto, G. *et al.* Mitochondrial DNA sequences in prehistoric human remains from the Alps. *Eur. J. Hum. Genet.* **8**, 669 (2000).
5. Korlevic, P. *et al.* Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015).
6. Arsuaga, J. *et al.* in *Continuity and Discontinuity in the Peopling of Europe* (eds. Condeani, S. & Weniger, G.-C.) 213–217 (Springer Netherlands, 2011).
7. Krause, J. *et al.* The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**, 894–897 (2010).
8. Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010).
9. Meyer, M. *et al.* A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* **338**, 222–226 (2012).
10. Castellano, S. *et al.* Patterns of coding variation in the complete exomes of three Neandertals. *Proc. Natl. Acad. Sci.* **111**, 6666–6671 (2014).
11. Sawyer, S. *et al.* Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *Proc. Natl. Acad. Sci.* **112**, 15696–15700 (2015).
12. Prufer, K. *et al.* The complete genome sequence of a Neandertal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
13. Lalueza-Fox, C. *et al.* Mitochondrial DNA of an Iberian Neandertal suggests a population affinity with other European Neandertals. *Curr. Biol.* **16**, R629–R630 (2006).
14. Green, R. E. *et al.* A Draft Sequence of the Neandertal Genome. *Science* **328**, 710–722 (2010).
15. Fortea, J. *et al.* La cueva de el Sidron (Borines, Piloña, Asturias): primeros resultados. (2003).
16. Lalueza-Fox, C. *et al.* Neandertal Evolutionary Genetics: Mitochondrial DNA Data from the Iberian Peninsula. *Mol. Biol. Evol.* **22**, 1077–1081 (2005).
17. Rosas, A. *et al.* Paleobiology and comparative morphology of a late Neandertal sample from El Sidrón, Asturias, Spain. *Proc. Natl. Acad. Sci.* **103**, 19266–19271 (2006).

18. Lalueza-Fox, C. *et al.* Genetic evidence for patrilocal mating behavior among Neandertal groups. *Proc. Natl. Acad. Sci.* **108**, 250–253 (2011).
19. Krings, M. *et al.* Neandertal DNA Sequences and the Origin of Modern Humans. *Cell* **90**, 19–30 (1997).
20. Krings, M., Geisert, H., Schmitz, R. W., Krainitzki, H. & Pääbo, S. DNA sequence of the mitochondrial hypervariable region II from the Neandertal type specimen. *Proc. Natl. Acad. Sci.* **96**, 5581–5585 (1999).
21. Schmitz, R. W. *et al.* The Neandertal type site revisited: Interdisciplinary investigations of skeletal remains from the Neander Valley, Germany. *Proc. Natl. Acad. Sci.* **99**, 13342–13347 (2002).
22. Briggs, A. W. *et al.* Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes. *Science* **325**, 318–321 (2009).
23. Higham, T. *et al.* The earliest evidence for anatomically modern humans in northwestern Europe. *Nature* **479**, 521–524 (2011).
24. Seguin-Orlando, A. *et al.* Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113–1118 (2014).
25. Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225–228 (2014).
26. Beauval, C. *et al.* A late Neandertal femur from Les Rochers-de-Villeneuve, France. *Proc. Natl. Acad. Sci.* **102**, 7085–7090 (2005).
27. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
28. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
29. Ovchinnikov, I. V. *et al.* Molecular analysis of Neandertal DNA from the northern Caucasus. *Nature* **404**, 490–493 (2000).
30. Skoglund, P. *et al.* Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl. Acad. Sci.* **111**, 2229–2234 (2014).
31. Caramelli, D. *et al.* A highly divergent mtDNA sequence in a Neandertal individual from Italy. *Curr. Biol.* **16**, R630 (2006).
32. Condemni, S. *et al.* Possible Interbreeding in Late Italian Neandertals? New Data from the Mezzena Jaw (Monti Lessini, Verona, Italy). *PLoS One* **8**, e59781 (2013).
33. Caramelli, D. *et al.* Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. *Proc. Natl. Acad. Sci.* **100**, 6593–6597 (2003).
34. Fu, Q. *et al.* An early modern human from Romania with a recent Neandertal ancestor. *Nature* **524**, 216–219 (2015).
35. Krause, J. *et al.* Neandertals in central Asia and Siberia. *Nature* **449**, 902–904 (2007).
36. Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).
37. Orlando, L. *et al.* Revisiting Neandertal diversity with a 100,000 year old mtDNA sequence. *Curr. Biol.* **16**, R400–R402 (2006).
38. Meyer, M. *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403–406 (2014).
39. Keller, A. *et al.* New insights into the Tyrolean Iceman’s origin and phenotype as inferred by whole-genome sequencing. *Nat Commun* **3**, 698 (2012).
40. Sikora, M. *et al.* Population genomic analysis of ancient and modern genomes yields new insights into the genetic ancestry of the Tyrolean Iceman and the genetic structure of Europe. *PLOS Genet.* **10**, e1004353 (2014).
41. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014).
42. Dalén, L. *et al.* Partial Genetic Turnover in Neandertals: Continuity in the East and Population Replacement in the West. *Mol. Biol. Evol.* **29**, 1893–1897 (2012).
43. Krings, M. *et al.* A view of Neandertal genetic diversity. *Nat. Genet.* **26**, 144–146 (2000).
44. Green, R. E. *et al.* Analysis of one million base pairs of Neandertal DNA. *Nature* **444**, 330–336 (2006).
45. Green, R. E. *et al.* A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416–426 (2008).

Supplementary Table 2

Site	Taxon	Specimen ID	bone/tooth	ESTIMATED DOSE	Ontoitem	scan ID	Scanner	kV	uA
Baigara	H. sapiens	Baigara 1	bone	0.9 Gy	TALL	10001241/000/001	BIR SN001 ACTIS5	130	110
Chagyrskaya	Neandertal	Chagyrskaya 2	bone	9.7 Gy	C1	20003197/000/001	Skyscan 1172	100	100
Chagyrskaya	Neandertal	Chagyrskaya - SP3393	bone	18.7 Gy	HP	20003919/000/001	SkyScan1173	100	62
Chagyrskaya	Neandertal	Chagyrskaya - SP3394	bone	8.3 Gy	ULNL ULNR	20003920/000/001	SkyScan1173	100	62
Denisova	Denisovan	Denisova 2 - SP3276	tooth	25.8 Gy	LLDM2 LRDM2	20003694/000/001	SkyScan1173	100	62
Denisova	?	Denisova 3	?	15.2 Gy	?	?	skyscan 1172	100	100
Denisova	Denisovan	Denisova 4	tooth	1.5 Gy	ULM3	10000796/002/001	BIR SN001 ACTIS5	130	100
Denisova	Neandertal	Denisova 5	bone	9.9 Gy	FP4or5P	20003531/000/001	Skyscan 1172	100	100
Denisova	?	Denisova 8	tooth	5.4 Gy	molar	10001219/000/004	BIR SN001 ACTIS5	130	110
Ust'-Ishim	H. sapiens	Ust'-Ishim 1	bone	0.3 Gy	FEML	10001598/000/001	BIR SN001 ACTIS5	130	100
Vindija	Neandertal	Vi 33.16 (Vi 80)	bone	0.4 Gy	TIB?	10000096/001/001	BIR SN001 ACTIS5	130	100
Site	Taxon	Specimen ID	bone/tooth	filters	total scanning time	nb of scans	nb of projections	frame averaging	exposure time [ms]
Baigara	H. sapiens	Baigara 1	bone	0.25 mm Brass	02:13:20	3	2500	3	333.3
Chagyrskaya	Neandertal	Chagyrskaya 2	bone	0.5 mm Al + 0.04 mm Cu	4:26:10	3	3000	2	2655
Chagyrskaya	Neandertal	Chagyrskaya - SP3393	bone	Al 1.0 mm	3:22:40	1	2400	4	991
Chagyrskaya	Neandertal	Chagyrskaya - SP3394	bone	Al 1.0 mm	3:22:26	2	2400	4	991
Denisova	Denisovan	Denisova 2 - SP3276	tooth	Al 1.0 mm	3:23:26	1	2400	4	991
Denisova	?	Denisova 3	?	Al 0.5 mm	2:57:00	1	3000	2	885
Denisova	Denisovan	Denisova 4	tooth	0.25mm Brass	?	?	2500	2	333.3
Denisova	Neandertal	Denisova 5	bone	0.5 mm Al + 0.04 mm Cu	3:31:50	2	2400	2	1770
Denisova	?	Denisova 8	tooth	0.25 mm Brass	02:46:40	4	2500	3	333.3
Ust'-Ishim	H. sapiens	Ust'-Ishim 1	bone	0.50 mm Brass	00:56:40	9	2500	2	333.3
Vindija	Neandertal	Vi 33.16 (Vi 80)	bone	0.25 mm Brass	00:55:33	4	1250	2	333.3
Site	Taxon	Specimen ID	bone/tooth	rotation step (deg)	Distance [object] to source [mm]	geometry	Pixel size [um]		
Baigara	H. sapiens	Baigara 1	bone	0.144	330.0	360	63.728		
Chagyrskaya	Neandertal	Chagyrskaya 2	bone	0.120	260.1	360	17.3974		
Chagyrskaya	Neandertal	Chagyrskaya - SP3393	bone	0.150	85.9	360	11.8304		
Chagyrskaya	Neandertal	Chagyrskaya - SP3394	bone	0.150	182.2	360	25.09851		
Denisova	Denisovan	Denisova 2 - SP3276	tooth	0.150	72.9	360	10.03552		
Denisova	?	Denisova 3	?	0.120	127.3	360	6.8		
Denisova	Denisovan	Denisova 4	tooth	0.144	110.0	?	20.014		
Denisova	Neandertal	Denisova 5	bone	0.150	187.6	360	15.4595		
Denisova	?	Denisova 8	tooth	0.144	165.0	360	30.065		
Ust'-Ishim	H. sapiens	Ust'-Ishim 1	bone	0.144	500.0	360	91.236		
Vindija	Neandertal	Vi 33.16 (Vi 80)	bone	0.288	300.0	360	54.6		

Supplementary Data 1

Measurements with BIR (Part 2)

Variation in SOD 24/06/2015 : we did a zeroing before starting the variation in SOD.
µGy/s

Set #1

No metallic filter, 50 kV												
10 W	Ncur	20 W	40 W	50 W	Ncur	75 W	Ncur	100 W	Ncur	150 W	Ncur	
20 mm	10720	230	88880	450	143500	670	190000	895	296100	1505	337300	1790
50 mm	8339	230	16170	450	23540	670	31120	895	48730	1505	55390	1790
100 mm	2086	230	4049	450	5884	670	779.6	895	12240	1505	13890	1790
150 mm	926.1	230	1795	450	2619	670	3456	895	5444	1505	6165	1790
165 mm	762.6	230	1481	450	2162	670	2849	895	4481	1505	5084	1790
200 mm	519	230	1004	450	1467	670	1942	895	3043	1505	3438	1790
300 mm	330.6	230	640.2	450	936.3	670	1234	895	1939	1505	2205	1790
400 mm	238.3	230	440.4	450	646.8	670	858.9	895	1346	1505	1528	1790
500 mm	168.8	230	324.6	450	474.6	670	624.3	895	938.7	1505	1118	1790
100.5	230	248.4	450	361.8	670	479.4	895	751.8	1505	852.3	1790	
450 mm	100.5	230	195.3	450	285	670	377.1	895	591.6	1505	719.9	1790
500 mm	82.2	230	159.3	450	231.6	670	306.3	895	479.1	1505	546	1790

Set #2

0.25 mm Brass filter, 50 kV												
10 W	Ncur	20 W	40 W	50 W	Ncur	75 W	Ncur	100 W	Ncur	150 W	Ncur	
20 mm	2696	230	5208	450	7244	665	9495	890	11480	1120	15170	1795
50 mm	434.4	230	834.3	450	1166	665	1516	890	1845	1120	2542	1795
100 mm	108.6	230	207.6	450	209.4	665	377.7	890	461.7	1120	634.2	1795
150 mm	47.7	230	92.4	450	129	665	167.7	890	204.3	1120	281.1	1795
165 mm	39.6	230	76.2	450	106.5	665	138.6	890	168.9	1120	232.8	1795
200 mm	26.4	230	51	450	72.3	665	93.9	890	115.2	1120	158.1	1795
250 mm	17.1	230	33	450	45.9	665	60.3	890	73.2	1120	101.4	1795
300 mm	11.4	230	22.8	450	32.4	665	41.7	890	51	1120	69.9	1795
350 mm	8.4	230	16.8	450	23.1	665	30	890	37.2	1120	51	1795
400 mm	6.2	230	12.4	450	17.7	665	23	890	28.1	1120	38.6	1795
450 mm	5.1	230	10.2	450	14.1	665	18.6	890	22.3	1120	30.9	1795
500 mm	4.2	230	8.1	450	12	665	15.3	890	18.3	1120	24.9	1795

Set #3

0.5 mm Brass filter, 50 kV												
10 W	Ncur	20 W	40 W	50 W	Ncur	75 W	Ncur	100 W	Ncur	150 W	Ncur	
20 mm	658.2	230	1204	450	1680	660	2187	890	2649	1150	3523	1800
50 mm	105.6	230	193.5	450	271.2	660	352.2	890	425.1	1150	569.1	1800
100 mm	26.4	230	48.3	450	67.5	660	87.6	890	106.5	1150	142.2	1800
150 mm	11.7	230	21.3	450	30	660	39	890	47.7	1150	62.7	1800
165 mm	9.6	230	17.7	450	24.9	660	32.4	890	38.4	1150	52.4	1800
200 mm	6.6	230	11.7	450	16.8	660	21.9	890	26.7	1150	35.2	1800
300 mm	3.6	230	6.3	450	9.3	660	11.8	890	14.5	1150	18.6	1800
400 mm	2.6	230	4.5	450	6.6	660	8.3	890	10.1	1150	13.2	1800
500 mm	1.8	230	3.1	450	5.1	660	7.2	890	8.4	1150	11.1	1800
100.5	230	3.3	450	4.2	660	5.4	890	6.3	1150	9	1800	
450 mm	0.9	230	2.4	450	3.3	660	4.2	890	5.1	1150	7.2	1800
500 mm	0.9	230	1.8	450	2.4	660	3.3	890	3.9	1150	5.7	1800

Set #4

1 mm Brass filter, 50 kV												
10 W	Ncur	20 W	40 W	50 W	Ncur	75 W	Ncur	100 W	Ncur	150 W	Ncur	
20 mm	66.3	230	120	450	159.6	600	204.9	890	299.7	1150	290.4	1800
50 mm	10.8	230	19.2	450	26.1	600	33	890	38.7	1150	47.1	1800
100 mm	2.6	230	4.8	450	9.3	600	7.8	890	17.90	1150	24.970	1800
150 mm	0.9	230	1.5	450	2.1	600	2.9	890	3.3	1150	4.2	1800
200 mm			0.9	450	1.2	600	1.8	890	2.4	1150	2.7	1800
250 mm					0.6	600	1.5	890	1.5	1150	1.5	1800
300 mm						600	0.9	890	0.6	1150	1.2	1800
350 mm										0.9	1800	
400 mm												
450 mm												
500 mm												

Set #13

1 mm Aluminum filter, 50 kV												
10 W	Ncur	20 W	40 W	50 W	Ncur	75 W	Ncur	100 W	Ncur	150 W	Ncur	
20 mm	384.0	230	719.9	450	965.0	600	1180	890	1590.0	1125	1800.0	1800
50 mm	64.5	230	129.0	450	169.5	600	219.0	890	279.0	1125	360.0	1800
100 mm	990.9	230	1902	450	2744	665	3627	890	4477	1125	6459	1800
150 mm	442.8	230	853.2	450	1231	665	1625	890	2007	1125	2888	1800
165 mm	365.1	230	705.3	450	1017	665	1347	890	1657	1125	2397	1800
200 mm	249.9	230	480.3	450	693.3	665	917.1	890	1131	1125	1548	1800
300 mm	159.6	230	307.2	450	444.9	665	585.9	890	726	1125	1048	1800
400 mm	110.7	230	213.9	450	309.6	665	410.7	890	503.4	1125	736	1800
500 mm	81.3	230	156.9	450	227.7	665	300.6	890	370.2	1125	514	1800
100.5	230	62.1	230	119.4	450	17.4	665	229.2	890	283.5	1125	409.8
450 mm	49.8	230	94.8	450	137.1	665	181.2	890	223.8	1125	323.1	1800
500 mm	39.9	230	76.8	450	111.3	665	147.6	890	182.7	1125	262.2	1800

Set #6

No metallic filter, 130 kV												
10 W	Ncur	20 W	40 W	50 W	Ncur	75 W	Ncur	100 W	Ncur	150 W	Ncur	
20 mm	77960	90	152500	175	222000	258	296500	342	463500	410	528500	770
50 mm	12760	90	24990	175	36770	258	48730	342	59840	410	88300	770
100 mm	3208	90	6279	175	9228	258	12260	342	15090	410	22220	770
150 mm	1500	90	2785	175	4100	258	5444	342	6707	410	9880	770
165 mm	1174	90	2301	175	3381	258	4497	342	5643	410	8140	770
200 mm	800.1	90	1565	175	2306	258	3058	342	3769	410	5575	770
250 mm	511.2	90	1001	175	1474	258	1960	342	2411	410	3563	770
300 mm	356.1	90	695.1	175	1022	258	1356	342	1679	410	2477	770
350 mm	260.7	90	509.7	175	751.5	258	997.2	342	1233	410	1821	770
400 mm	200.7	90	392.4	175	575.4	258	765.9	342	947.1	410	1394	770
450 mm	157.8	90	310.2	175	457.2	258	607.8	342	751.5	410	1105	770
500 mm	130.2	90	253.2	175	377.8	258	495.6	342	613.5	410	900.3	770

Set #7

0.25 mm Brass filter, 130 kV												
20 mm	Ncur	20 W	40 W	50 W	Ncur	75 W	Ncur	100 W	Ncur	150 W	Ncur	
20 mm	21700	90	42440	175	62650	258	81300	342	100100	460	132000	795
50 mm	3525	90	6874	175	10170	258	13200	342	16410	460	23870	795
100 mm	885	90	1721	175	2546	258	3313	342	4124	460	5995	795
150 mm	391.8	90	765.6	175	1137	258	1476	342	1836	460	2683	795
165 mm	323.4	90	631.8	175	938.7	258	1222	342	1523	460	2205	795
200 mm	220.8	90	430.2	175	639.6	258	830.7	342	1036	460	1506	795
250 mm	140.7	90	275.4	175	408.6	258	532.8	342	662.7	460	962.4	795
300 mm	97.2	90	191.1	175	283.5	258	369.3	342	461.1	460	672.9	795
350 mm	72	90	141.3	175	209.1	258	272.1	342	339.3	460	495	795
400 mm	53.9	90	106.6	175	156.6	258	201.6	342	251.6	460	361.3	795
450 mm	43.9	90	86	175	128.1	258	168	342	209.4	460</		

Supplementary Data 1

Measurements with BIR (Part 3)

Set #12 No metallic filter, 200 kV

	10 W	20 W	30W	Ncur	40 W	50 W	75 W	100 W	160 W	320 W	Ncur	
20 mm	77040	63	141000	118	200500	173	259300	380	466600	720	592200	1420
50 mm	14130	63	26140	118	37540	173	49140	380	89280	720	114500	1420
100 mm	3677	63	6851	118	9860	173	13060	258	16270	380	23560	720
150 mm	1658	63	3091	118	4474	173	5901	380	10730	720	13790	1420
165 mm	1370	63	2560	118	3709	173	4902	258	6087	380	8854	720
200 mm	936.9	63	1749	118	2539	173	3358	258	4131	380	6059	720
250 mm	601.5	63	1129	118	1635	173	2164	258	2694	380	3894	720
300 mm	368	63	676	118	1137	173	1494	258	1933	380	2733	720
350 mm	237.3	63	436.6	118	745.5	173	986.6	258	1268	380	1820	720
400 mm	237.3	63	444.3	118	646.5	173	858.6	258	1056	380	1546	720
450 mm	188.4	63	354.3	118	517.5	173	685.8	258	843.6	380	1233	720
500 mm	156.3	63	293.1	118	423.9	173	563.4	258	692.4	380	1015	720

Set #11 0.25 mm Brass filter, 200 kV

	10 W	20 W	30W	Ncur	40 W	50 W	75 W	100 W	160 W	320 W	Ncur	
20 mm	33080	63	60400	116	87700	173	115700	258	140900	358	206300	690
50 mm	5974	63	11000	116	16080	173	21400	258	26240	358	38410	690
100 mm	1551	63	2864	116	4214	173	5627	258	6914	358	10070	690
150 mm	700.5	63	1292	116	1904	173	2548	258	3124	358	4535	690
165 mm	578.1	63	1068	116	1573	173	2106	258	2583	358	3712	690
200 mm	394.5	63	731.1	116	1080	173	1445	258	1770	358	2570	690
250 mm	277.3	63	517.6	116	766.3	173	1033	258	1296	358	1924	690
300 mm	192.9	63	377.6	116	548.4	173	735.6	258	948.6	358	1374	690
350 mm	129.6	63	242.1	116	359.4	173	481.2	258	590.7	358	845.4	690
400 mm	100.5	63	187.2	116	276.3	173	372.6	258	456.9	358	652.5	690
450 mm	80.1	63	150.6	116	222.6	173	298.2	258	369.3	358	525.6	690
500 mm	67.5	63	125.7	116	185.7	173	249	258	304.2	358	432.6	690

Set #9 0.5 mm Brass filter, 200 kV

	10 W	20 W	30W	Ncur	40 W	50 W	75 W	100 W	160 W	320 W	Ncur	
20 mm	25190	63	46640	118	67390	173	88620	240	127400	350	184000	728
50 mm	4526	63	8442	118	12320	173	16200	240	21260	350	31050	728
100 mm	1175	63	2201	118	3218	173	4250	240	5429	350	7843	728
150 mm	528.9	63	991.2	118	1443	173	1934	240	2418	350	3514	728
165 mm	426.4	63	820.5	118	1139	173	1536	240	1996	350	2863	728
200 mm	287.2	63	549.2	118	817.8	173	1086	240	1391	350	2049	728
250 mm	192.9	63	367.2	118	526.5	173	698.7	240	877.2	350	1281	728
300 mm	134.4	63	252	118	367.8	173	488.1	240	610.5	350	887.4	728
350 mm	99.3	63	186.3	118	272.7	173	360.9	240	451.5	350	657.9	728
400 mm	76.2	63	144.3	118	209.4	173	277.8	240	348.6	350	508.8	728
450 mm	61.8	63	117.3	118	169.2	173	226.5	240	282.9	350	412.8	728
500 mm	51.6	63	98.1	118	144	173	186.9	240	235.8	350	342	728

Set #10 1 mm Brass filter, 200 kV

	10 W	20 W	30W	Ncur	40 W	50 W	75 W	100 W	160 W	320 W	Ncur	
20 mm	17340	63	32380	118	46380	173	60430	238	73700	335	106100	715
50 mm	3093	63	5816	118	8365	173	11030	238	13590	335	19750	715
100 mm	801.6	63	1509	118	2287	173	2882	238	3582	335	5178	715
150 mm	467.2	63	876.7	118	1317	173	1696	238	2490	335	3416	715
165 mm	307.2	63	562.8	118	815.7	173	1086	238	1391	335	1940	715
200 mm	204.3	63	385.2	118	559.2	173	741.9	238	913.5	335	1330	715
250 mm	131.4	63	248.4	118	361.2	173	479.7	238	590.1	335	857.1	715
300 mm	91.2	63	173.1	118	252.6	173	336	238	413.1	335	599.4	715
350 mm	68.4	63	128.7	118	186.6	173	248.4	238	305.7	335	443.4	715
400 mm	51.9	63	98.7	118	143.7	173	191.7	238	236.1	335	342.3	715
450 mm	42	63	79.8	118	117.6	173	156.3	238	193.2	335	278.7	715
500 mm	35.7	63	68.4	118	99.3	173	130.5	238	160.2	335	232.5	715

Set #15 1 mm Aluminum filter, 200 kV

	10 W	20 W	30W	Ncur	40 W	50 W	75 W	100 W	160 W	320 W	Ncur	
20 mm	60620	63	113500	118	163900	173	211600	232	259200	328	373200	675
50 mm	10930	63	20380	118	29680	173	38620	232	47640	328	69320	675
100 mm	2819	63	5303	118	7482	173	9820	232	12490	328	18160	675
150 mm	1650	63	3179	118	4688	173	6188	232	7860	328	11290	675
165 mm	1050	63	1979	118	2888	173	3788	232	4680	328	6789	675
200 mm	717	63	1353	118	1977	173	2590	232	3212	328	4637	675
250 mm	461.4	63	870.9	118	1271	173	1664	232	2069	328	2980	675
300 mm	320.4	63	605.1	118	884.1	173	1162	232	1445	328	2081	675
350 mm	238.2	63	447.3	118	653.1	173	858.9	232	1064	328	1537	675
400 mm	182.4	63	344.4	118	501.6	173	660.3	232	820.8	328	1184	675
450 mm	145.8	63	276.6	118	402.6	173	530.4	232	657.6	328	948	675
500 mm	120.9	63	228.6	118	331.8	173	437.1	232	540.6	328	777.6	675

Supplementary Data 1

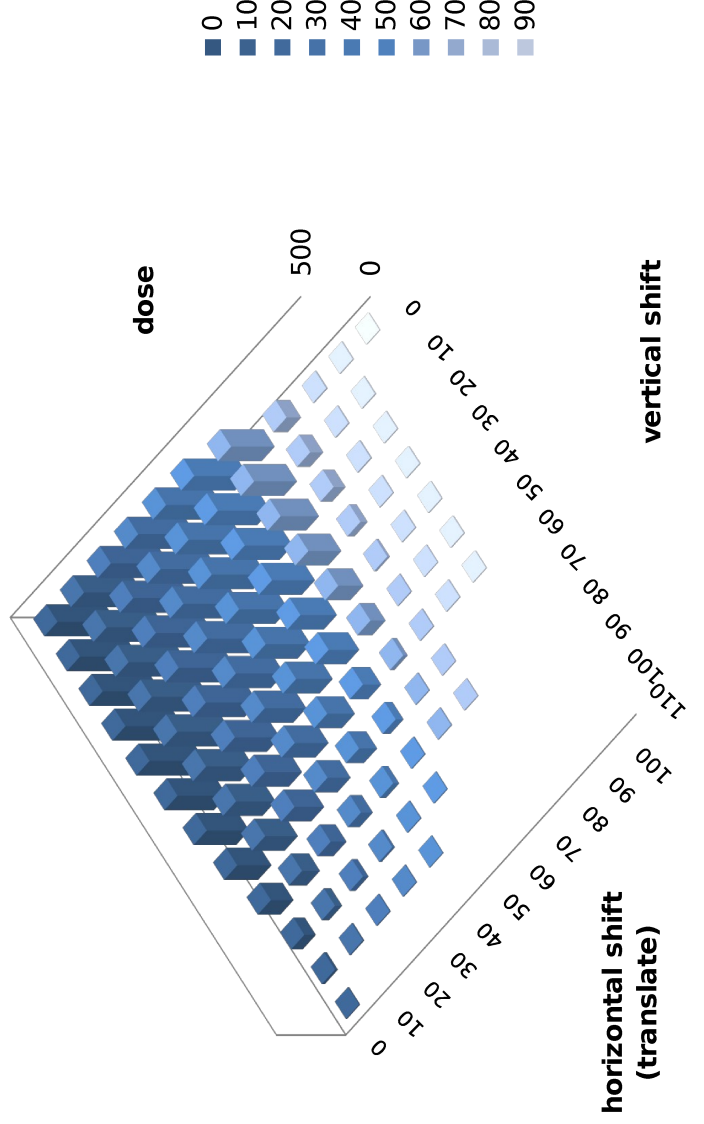
Measurements with BIR (Part 4)

0.5mm Brass, 35 measurements of dose rate, source/dosimeter distance 250 mm

W	uA													
	5	10	15	20	25	30	40	50	75	100	125	150	175	200
100	TheoCur 100.00	TheoCur 100.00	TheoCur 115	TheoCur 150.00	TheoCur 200.00	TheoCur 250.00	TheoCur 300.00	TheoCur 400.00	TheoCur 500.00	TheoCur 750.00	TheoCur 1000.00	TheoCur 1250.00	TheoCur 1500.00	TheoCur 1750.00
130	Ncur 38.46	Ncur 76.92	Ncur 90	Ncur 115.38	Ncur 153.85	Ncur 192.31	Ncur 230.77	Ncur 307.69	Ncur 384.62	Ncur 576.92	Ncur 769.23	Ncur 961.54	Ncur 1153.85	Ncur 1346.15
200	Ncur 25.00	Ncur 50.00	Ncur 65	Ncur 75.00	Ncur 100.00	Ncur 125.00	Ncur 150.00	Ncur 200.00	Ncur 250.00	Ncur 375.00	Ncur 500.00	Ncur 625.00	Ncur 750.00	Ncur 875.00

"Vertical variation"
0.5 mm Brass filter, 130 kV, @250mm SOD, 50W, NCur 450

h : v	0	10	20	30	40	50	60	70	80	90	100	110
0	423	421.5	417.6	412.5	399	356.7	306.6	246.3	163.5	82.2	21	7.8
10	422.4	420.3	420	414.3	390.3	341.7	284.4	234.9	129.6	52.2	11.4	
20	419.4	417.6	414.3	410.1	373.2	324	269.1	214.2	113.4	41.1	9.6	
30	412.5	410.7	406.8	394.2	348.3	299.4	243.9	167.1	85.8	19.8	9	
40	402.3	399.6	391.5	355.2	323.1	264.6	199.8	129.9	43.5	9.9	8.7	
50	388.8	378.3	345.6	309	264.3	220.8	133.5	56.7	10.5	9		
60	314.4	308.1	277.5	237.6	182.4	110.1	33.3	10.2	9.3			
70	105	95.1	85.2	52.5	19.5	10.5	9.6	9.3	8.7			
80	11.1	11.1	11.1	10.5	10.2	9.6	9.6					
90	9.9	9.9	9.9	9.3	9.3	9	8.7					
100	8.4											



Supplementary Data 1

Measurements with SkyScan 1173

SkyScan 1173

maximum tension 130 kV
 maximum power 8 W
 corresponding intensity 62 uA
 minimum source/sample distance 50 mm
 maximum source/sample distance 250 mm
 commonly used filters 1mm Al, 0.25mm brass

Configurations to test with the dosimeter chamber fixed on the center of the rotation stage, full illumination of the chamber in dose rate mode (adapt the measurement scale), in surface dose wat the measurements can then be adapted if necessary, as long as exact parameters are noted. Especially for the distance scans. Total number of measurements is 126

detector pixel size = 50 μm

SID = 364 mm

SOD = 50mm

SOD=250mm

6.96 μm

34:3:4μm

Zeroing of the dosimeter before starting the measurements
dose in μGy

at 125mm = 17.2μm

no metallic filter, 20 measurements of dose rate, source/dosimeter distance 125 mm

set#1	2W	4W	6W	8W
50 kV	7151	40	15300	80
70 kV	7400	29	21480	57
90 kV	6446	22	12870	44
110 kV	5672	18	11310	36
130 kV	4982	15	10240	31

Second run, but same effect

set#4	2W	4W	6W	8W
50 kV	7309	40	14910	80
70 kV	7047	29	14150	57
90 kV	6422	22	12760	44
110 kV	5693	18	10910	36
130 kV	5046	15	10170	31

No metallic filter, 80 kV, 33 measurements

set#5	2W	4W	6W	8W
50 mm = 6.86μ	56970	106900	154900	298000
100 mm = 13.74μm	11750	23040	31740	43650
125 mm = 17.17μm	6332	13620	18580	25920
150 mm = 20.60μm	4532	8997	12040	16990
200 mm = 27.47μm	2392	4704	6302	8993
250 mm = 34.34μm	1451	2864	3834	5475

No metallic filter, 50 kV, 33 measurements

set#8	2W	4W	6W	8W
50 mm = 6.86μ	61740	121500	179900	237100
100 mm = 13.74μm	12980	25510	37100	49550
125 mm = 17.17μm	7673	15070	21620	28220
150 mm = 20.60μm	5073	9860	13940	18010
200 mm = 27.47μm	2637	5189	7272	9977
250 mm = 34.34μm	1596	3153	4379	5973

"Vertical variation"

0.25 mm Brass filter, 130 kV, @125mm SOD, 33 measurements

vertical shift	2W	4W	6W	8W
0mm	107.7	223.2	330.6	440.7
10mm	107.7	223.2	329.4	438.3
20mm	104.7	214.5	315.3	422.7
30mm	95.4	195	283.5	383.1
35mm	33.3	71.4	104.4	140.4
40mm	8.1	20.1	30.3	41.7
50mm	5.7	14.4	23.1	31.8
60mm	8.4	13.8	19.5	27.1

only vertical but no horizontal shift possible

sod	pixel
200	364
p	U
8	130
8	61.5

1mm Al, 20 measurements of dose rate, source/dosimeter distance 125 mm

set#2	2W	4W	6W	8W
50 kV	309.6	613.5	918.3	1220
70 kV	380.7	743.1	1117	1467
90 kV	409.5	801.6	1215	1587
110 kV	431.7	856.2	1301	1686
130 kV	445.8	895.1	1333	1763

0.25mm brass filter, 20 measurements of dose rate, source/dosimeter distance 125 mm

set#3	2W	4W	6W	8W
50 kV	11.4	26.4	40.2	55.2
70 kV	36	73.5	112.2	149.4
90 kV	62.7	129	197.4	259.2
110 kV	90.6	182.4	280.5	368.1
130 kV	117.6	243.6	360.3	478.8

0.25mm brass filter, 130 kV, 33 measurements

set#7	2W	4W	6W	8W
50 mm	852.3	1746	2591	3407
100 mm	191.7	396.1	591.3	784.5
125 mm	118.2	244.2	362.4	480.6
150 mm	78.3	165	244.2	324.6
200 mm	42.6	91.2	137.1	181.5
250 mm	26.7	57.6	87.9	115.8

1mm aluminum filter, 100 kV, 33 measurements

set#6	2W	4W	6W	8W
50 mm	3070	5897	8752	11920
100 mm	698.4	1366	1997	2715
125 mm	427.8	837	1225	1642
150 mm	287.7	566.4	827.7	1107
200 mm	159.3	314.1	460.5	617.1
250 mm	101.1	201	293.4	393.9



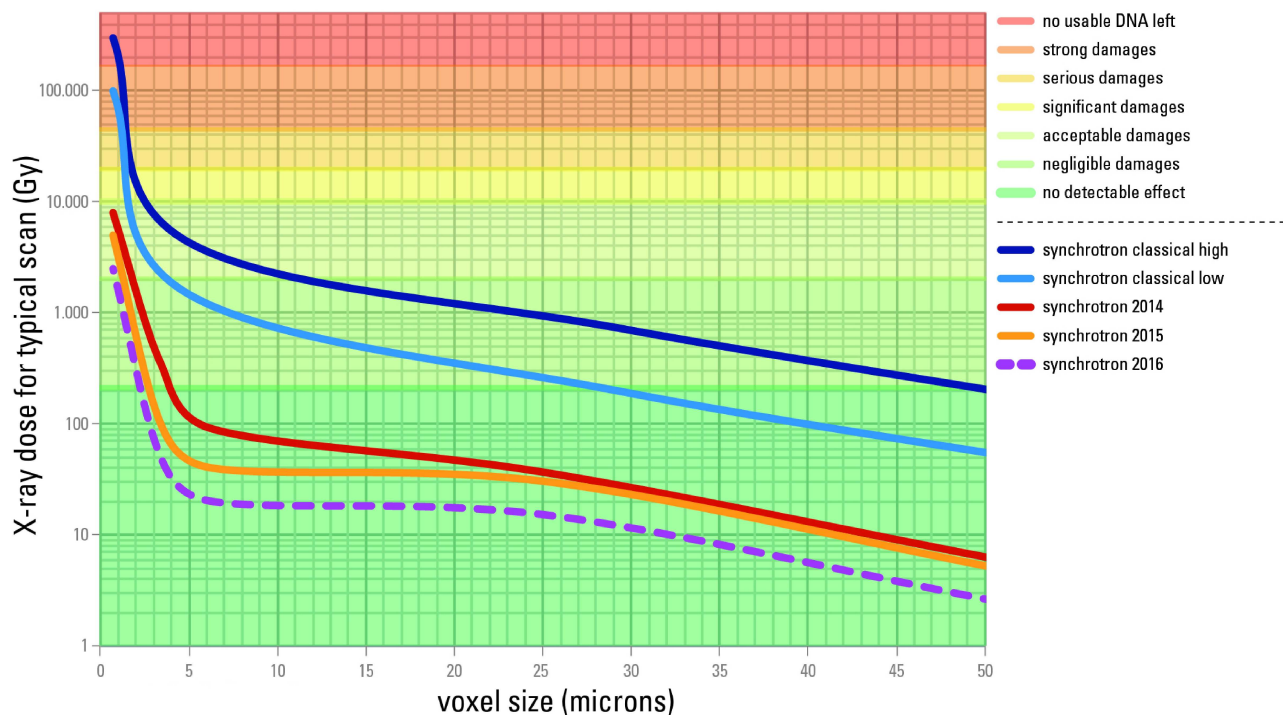
Supplementary Data 2

Synchrotron Dosimetry

synchrotron configurations									
classical high		classical low		2014		2015		2016	
voxel size (um)	surface dose (Gy)	voxel size (um)	surface dose (Gy)	voxel size (um)	surface dose (Gy)	voxel size (um)	surface dose (Gy)	voxel size (um)	surface dose (Gy)
59	120.0	59.0	32.5	50.74	6	50.74	5	50.74	2.5
29.5	720	29.5	195	25.37	36	25.37	30	25.37	15
4.9	4400	4.9	1500	6.34	90	6.34	40	6.34	20
0.7	300000	0.7	100000	3.4	350	3.4	100	3.4	50
				0.7	8000	0.7	5000	0.7	2500

surface dose (Gy)	aDNA quantity ratio	fragment length ratio endogenous DNA	fragment length ratio total DNA	C/T misincorporation ratio	general effect of X-ray irradiation on aDNA
0	~1	~1	~1	~1	no detectable effect
200	~1	~1	~1	~1	negligible damages
2000	0.85	~1	~1	0.97	acceptable damages
10000	0.8	~1	0.98	0.93	significant damages
20000	0.35	0.95	0.95	0.85	serious damages
45000	0.3	0.88	0.92	0.67	strong damages
170000	0.1	N.A.	N.A.	N.A.	no usable DNA left

X-ray dose for a typical scan (Gray)



Risk color coding (used to define the color regions on the graphs)

no detectable effect	step	negligible damages		step	acceptable damages		step	significant damages		step	serious damages		step	strong damages		step	no usable DNA left		step
-50	180.0	-50	1980.0	1.025	-50	9980.0	1.017	-50	19500.0	1.008	-50	44500.0	1.009	-50	165000.0	1.014	-50	1000000.0	1.02
200	180.0	200	1980.0	min	200	9980.0	min	200	19500.0	min	200	44500.0	min	200	165000.0	min	200	1000000.0	min
200	180.0	200	1980.0	0.1	200	9879.1	2060	200	19500.0	10020	200	44500.0	20020	200	165000.0	45500	200	1000000.0	175000
-50	180.0	-50	1980.0	max	-50	9713.9	max	-50	19500.0	max	-50	44500.0	max	-50	163494.5	max	-50	1000000.0	max
-50	180.0	-50	1980.0	180	-50	9551.6	9980	-50	19500.0	19500	-50	44500.0	44500	-50	161237.2	165000	-50	1000000.0	1000000
200	180.0	200	1938.1		200	9391.9		200	19500.0		200	44500.0		200	159011.1		200	1000000.0	
200	180.0	200	1890.8		200	9234.9		200	19500.0		200	44440.4		200	156815.6		200	1000000.0	
-50	180.0	-50	1844.7		-50	9080.5		-50	19500.0		-50	44044.0		-50	154650.5		-50	999662.0	
-50	180.0	-50	1799.7		-50	8928.7		-50	19500.0		-50	43651.1		-50	152515.3		-50	980060.7	
200	165.5	200	1755.8		200	8779.5		200	19500.0		200	43261.8		200	150409.6		200	960843.9	
200	151.8	200	1713.0		200	8632.7		200	19500.0		200	42875.9		200	148332.9		200	942003.8	
-50	139.3	-50	1671.2		-50	8488.4		-50	19500.0		-50	42493.5		-50	146284.9		-50	923533.1	
-50	127.8	-50	1630.4		-50	8346.5		-50	19413.0		-50	42114.4		-50	144265.2		-50	905424.6	
200	117.2	200	1590.7		200	8207.0		200	19258.9		200	41738.8		200	142273.4		200	887671.2	
200	107.5	200	1551.9		200	8069.8		200	19106.0		200	41366.5		200	140309.1		200	870265.9	
-50	98.7	-50	1514.0		-50	7934.9		-50	18954.4		-50	40997.5		-50	138371.9		-50	853201.9	
-50	90.5	-50	1477.1		-50	7802.3		-50	18804.0		-50	40631.8		-50	136461.4		-50	836472.4	
200	83.0	200	1441.1		200	7671.9		200	18654.7		200	40269.4		200	134577.3		200	820071.0	
200	76.2	200	1405.9		200	7543.6		200	18506.7		200	39910.2		200	132719.3		200	803991.2	
-50	69.9	-50	1371.6		-50	7417.5		-50	18359.8		-50	39554.2		-50	130886.8		-50	788226.6	
-50	64.1	-50	1338.2		-50	7293.5		-50	18214.1		-50	39201.4		-50	129079.7		-50	772771.2	
200	58.8	200	1305.5		200	7171.6		200	18069.5		200	38851.7		200	127297.6		200	757618.8	
200	54.0	200	1273.7		200	7051.7		200	17926.1		200	38505.2		200	125540.0		200	742763.6	
-50	49.5	-50	1242.6		-50	6933.9		-50	17783.9		-50	38161.7		-50	123806.7		-50	728199.6	
-50	45.4	-50	1212.3		-50	6818.0		-50	17642.7		-50	37821.3		-50	122097.3		-50	713921.1	
200	41.7	200	1182.7		200	6704.0		200	17502.7		200	37484.0		200	120411.6		200	699922.7	
200	38.2	200	1153.9		200	6591.9		200	17363.8		200	37149.6		200	118749.1		200	686198.7	
-50	35.1	-50	1125.8		-50	6481.7		-50	17226.0		-50	36818.3		-50	117109.6		-50	672743.8	
-50	32.2	-50	1098.3		-50	6373.4		-50	17089.3		-50	36489.9		-50	115492.7		-50	659552.8	
200	29.5	200	1071.5		200	6266.9		200	16953.6		200	36164.4		200	113898.1		200	646620.4	
200	27.1	200	1045.4		200	6162.1		200	16819.1		200	35841.8		200	112325.5		200	633941.5	
-50	24.8	-50	1019.9		-50	6059.1		-50	16685.6		-50	35522.1		-50	110774.7		-50	621511.3	
-50	22.8	-50	995.0		-50	5957.8		-50	16553.2		-50	35205.3		-50	109245.2		-50	609324.8	
200	20.9	200	970.7		200	5858.2		200	16421.8		200	34891.2		200	107736.9		200	597377.3	
200	19.2	200	947.1		200	5760.3		200	16291.5		200	34580.0		200	106249.4		200	585664.0	
-50	17.6	-50	924.0		-50	5664.0		-50	16162.2		-50	34271.6		-50	104782.5		-50	574180.4	
-50	16.1	-50	901.4		-50	5569.3		-50	16033.9		-50	33965.9		-50	103335.8		-50	562921.9	
200	14.8	200	879.4		200	5476.2		200	15906.6		200	33662.9		200	101909.1		200	551884.3	
200	13.6	200	858.0		200	5384.7		200	15780.4		200	33362.7		200	100502.0		200	541063.0	
-50	12.5	-50	837.1		-50	5294.7		-50	15655.2		-50	33065.1		-50	99114.4		-50	530453.9	
-50	11.4	-50	816.6		-50	5206.2		-50	15530.9		-50	32770.1		-50	97746.0		-50	520052.9	
200	10.5	200	796.7		200	5119.2		200	15407.7		200	32477.8		200	96396.4		200	509855.8	
200	9.6	200	777.3		200	5033.6		200	15285.4		200	32188.1		200	95065.5		200	499858.6	
-50	8.8	-50	758.3		-50	4949.4		-50	15164.1		-50	31901.0		-50	93753.0		-50	490057.4	
-50	8.1	-50	739.8		-50	4866.7		-50	15043.7		-50	31616.5		-50	92458.6		-50	480448.5	
200	7.4	200	721.8		200	4785.4		200	14924.3		200	31334.5		200	91182.0		200	471027.9	
200	6.8	200	704.2		200	4705.4		200	14805.9		200	31055.0		200	89923.1		200	461792.1	
-50	6.3	-50	687.0		-50	4626.7		-50	14688.4		-50	30778.0		-50	88681.5		-50	452737.3	
-50	5.7	-50	670.3		-50	4549.4		-50	14571.8		-50	30503.4		-50	87457.1		-50	443860.1	
200	5.3	200	653.9		200	4473.3		200	14456.1		200	30231.4		200	86249.6		200	435157.0	
200	4.8	200	638.0		200	4398.6		200	14341.4		200	29961.7		200	85058.8		200	426624.5	
-50	4.4	-50	622.4		-50	4325.0		-50	14227.6		-50	29694.5		-50	83884.4		-50	418259.3	
-50	4.1	-50	607.2		-50	4252.7		-50	14114.7		-50	29429.6		-50	82726.3		-50	410058.1	
200	3.7	200	592.4		200	4181.6		200	14002.6		200	29167.1		200	81584.1		200	402017.8	
200	3.4	200	578.0		200	4111.7		200	13891.5		200	28906.9		200	80457.7		200	394135.1	
-50	3.1	-50	563.9		-50	4043.0		-50	13781.3		-50	28649.1		-50	79346.8		-50	386406.9	
-50	2.9	-50	550.1		-50	3975.4		-50	13671.9		-50	28393.5		-50	78251.3		-50	378830.3	
200	2.6	200	536.7		200	3909.0		200	13563.4		200	28140.3		200	77170.9		200	371402.3	
200	2.4	200	523.6		200	3843.6		200	13455.7		200	27889.3		200	76105.4		200	364119.9	
-50	2.2	-50	510.8		-50	3779.4		-50	13348.9		-50	27640.5		-50	75054.7		-50	356980.3	
-50	2.0	-50	498.4		-50	3716.2		-50	13243.0		-50	27394.0		-50	74018.4		-50	349980.7	
200	1.9	200	486.2		200	3654.1		200	13137.9		200	27149.6		200	72996.5		200	343118.3	
200	1.7	200	474.4		200	3593.0		200	13033.6		200	26907.5		200	71988.6		200	336390.5	
-50	1.6	-50	462.8		-50	3533.0		-50	12930.2		-50	26667.4		-50	70994.7		-50	329794.6	
-50	1.4	-50	451.5		-50	3473.9		-50	12827.6		-50	26429.6		-50	70014.5		-50	323328.0	
200	1.3	200	440.5		200	3415.8		200	12725.8		200	26193.8		200	69047.8		200	316988.3	
200	1.2	200	429.7		200	3358.7		200	12624.8		200	25960.2		200	68094.5		200	310772.8	
-50	1.1	-50	419.3		-50	3302.6		-50	12524.6		-50	25728.6		-50	67154.3		-50	304679.2	
-50	1.0	-50	409.0		-50	3247.4		-50	12425.2		-50	25499.1		-50	66227.2		-50	298705.1	
200	0.9	200	399.1		200	3193.1		200	12326.6		200	25271.7		200	65312.8		200	292848.2	
200	0.9	200	389.3		200	3139.7		200	12228.7		200	25046.3		200	64411.0		200	287106.0	
-50	0.8	-50	379.8		-50	3087.2		-50	12131.7		-50	24822.9		-50	63521.7		-50	281476.5	
-50	0.7	-50	370.6		-50	3035.6		-50	12035.4		-50	24601.5		-50	62644.7		-50	275957.4	
200	0.7	200	361.5		200	2984.9		200	11939.9		200	24382.0		200	61779.8		200	270546.4	
200	0.6	200	352.7		200	2935.0		200	11845										

Supplementary Data 2

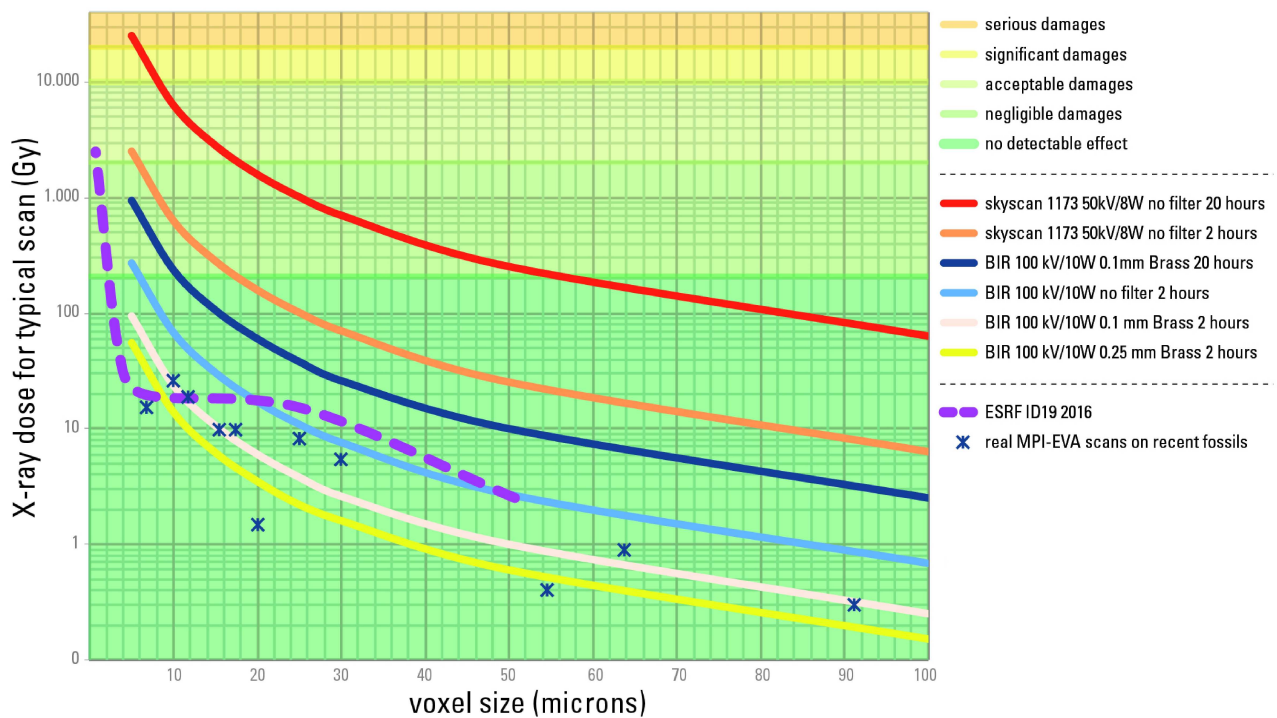
Conventional Scanner Dosimetry

surface dose (Gy) for MPI-EVA BIR scanner at 100 kV / 10W				
voxel size (um)	no filter 2h	0.1 mm brass 2h	0.25 mm brass 2h	0.1mm brass 20 h
100	0.7	0.3	0.2	2.5
50	2.7	1	0.6	10
30	7.6	2.6	1.6	26
25	10.9	3.8	2.2	38
20	17.1	6	3.5	60
15	30.5	10.6	6.2	106
10	67.9	23.6	13.9	236
5	271.4	94.4	55.5	944

surface dose (Gy) for MPI-EVA skyscan scanner at 50 kV / 8W		
voxel size (um)	no filter 2h	no filter 20 h
100	6.3	63.3
50	25.3	253
30	70.2	702
25	101.2	1012
20	158.4	1584
15	282.3	2823
10	629.3	6293
5	2517	25170

MPI typical scans dose estimations	
voxel size	estimated dose
91.2	0.3
63.7	0.9
54.6	0.4
30.1	5.4
25.1	8.3
20.0	1.5
17.4	9.7
15.5	9.9
11.8	18.7
10.0	25.8
6.8	15.2

X-ray surface dose for conventional microtomographs



Supplementary Data 3

Dose Estimator for μ CT

Source parameters

tension (kV) **160** max kV
 intensity (μ A) **100.0** **OK** 200
 power (W) **16.0** **OK** 100
 max W

filters

Aluminum (mm) **0** 0-1mm **OK**
 copper/Brass (mm) **0.1** 0-1 mm **OK**

reference dose rate at 250 mm from the source (uGray/s) 408.1

manual input of the reference dose rate in uGray/s (0 to desactivate) => **0**
 manual entry of the distance for reference dose rate (mm) => **0**

setup geometry

source sample distance (mm) **250.00**
 detector pixel size (μ m) 127
 source detector distance (mm) 696
 object voxel size (μ m) 45.6

scan parameters

number of projections 3600
 exposure time per frame (s) 1
 readout dead time with beam on 0
 number of frame repetition 2
 number of spiral turn / stages 1
 total exposure time (s) 7200
 total exposure time (min) 120.0
 total exposure time (hours) 2.0

estimated dose rate at sample level (uGy/s) 408.1 **0** <= manual input of measured dose rate (0 to desactivate)

total delivered surface dose (mGy) **2938**
 total delivered surface dose (Gy) **2.9**

integrated dose for multiple / spiral scan (mGy) **2938.0**
 integrated dose for multiple / spiral scan (Gy) **2.9**

IMPORTANT: ALL THE DOSE ESTIMATIONS ARE MADE FOR WATER EQUIVALENT SURFACE DOSE AT THE CENTER OF THE SOURCE CONE

all the inputs for the surface dose estimations are in the green boxes, only these cells can be selected

scanner parameters	detector pix size (μ m)	source/det distance (mm)	max current (kV)	max power (W)	vertical cone aperture (deg)
BIR MPI-EVA	127	696	200	100	20
skyscan 1173	50	364	130	8	10
skyscan 1172	11.66	218.8	160	10	10
user defined	50	915	200	100	37

selected scanner **1** 127 696 200 100 20

BIR MPI-EVA

0 <= manual input of scan duration in minutes (0 to desactivate)

Parameters for spiral / multi-stage scans

vertical cone aperture (deg) **20**
 vertical beam size at the sample position (mm) 91.0
 vertical displacement per spiral / stage turn (mm) **77.76**

proportion of the total scanning time in direct beam 100.0%
 dose ponderation factor with distance from beam center 100.0%
 proportion of maximum dose 100.0%

Supplementary Data 3

BIR references dose rates (Part 1)

power(W)	dose rate
5	142.9
10	263.5
15	385.2
20	501.7
25	619.8
30	738.6
40	976.3
50	1205.0
75	1765.1
100	2282.2

polynomial regression for all power values

a	b	c
-0.021592287	24.854917332	15.903360609

16.0 **408.1** dose rate value used as reference for this source configuration

power(W)	dose rate
5	309.5
10	561.6
15	822.5
20	1078.5
25	1328.6
30	1586.7
40	2094.2
50	2590.3
75	3831.4
100	4952.2

polynomial regression for all power values

a	b	c
-0.038603576	53.173917621	33.148216262

16.0 **874.0** dose rate value used as reference for this source configuration

reference configure 5 W 250 mm source-object distance

reference measurement s kV	brass thickness		
	0	0.25	0.5
50	166.2	8.4	1.8
100	249.9	48.9	26.1
130	286.8	77.1	48.3
150	288.3	90.6	61.8
200	340.2	139.2	103.5

polynomial regression for all brass thickness extrapolation $y=ax^2+bx+x$

brass thickness	a	b	c
0	-0.003952007	2.1112920557	72.8950867758
0.25	0.0005451626	0.7327636764	-29.5083109264
0.5	0.0015895087	0.2947082865	-16.7884135508
1	0.0024336006	-0.114727789	-0.6938645808

extrapolated results for given energy

brass filter	thick(L/3)	energy (kV)
0	0	160
0.25	0.6299605249	309.5
0.5	0.793700526	101.7
1	1	43.2

extrapolated dose rate value for given brass thickness

brass thickness (mm)	dose rate (uGray/s)
0	142.9

reference configure 5 W 250 mm source-object distance

reference measurement s kV	Aluminium filter		
	0	1	1
50	166.2	82.5	15.7
100	249.9	155.7	190.2
130	286.8	190.2	204.9
150	288.3	204.9	252.3
200	340.2	252.3	

polynomial regression for all energies extrapolation $y=ax^2+bx+x$

Al thickness	a	b	c
0	-0.003952007	2.1112920557	72.8950867758
1	-0.002789597	1.8152752872	-0.3318992911

extrapolated results for given energy

Al thick	energy (kV)
0	160
1	309.5
1	218.7

linear regression for all Al thickness

a	b
-90.83196871	309.53044048

NOT ACCURATE, MORE CALIBRATION POINTS ARE NECESSARY

extrapolated dose rate value for given brass thickness

Al thickness (mm)	dose rate (uGray/s)
0	309.5

reference configure 10 W 250 mm source-object distance

reference measurement s kV	brass thickness		
	0	0.25	0.5
50	326.1	16.5	3.3
100	464.4	90	48.9
130	511.8	139.2	87.9
150	542.4	171.6	116.1
200	606	258	192.3

polynomial regression for all brass thickness extrapolation $y=ax^2+bx+x$

brass thickness	a	b	c
0	-0.007071494	3.6019878758	166.263578871
0.25	0.001205662	1.3110567587	-52.2932045855
0.5	0.002927262	0.5205544855	-30.8875336104
1	0.0034545246	-0.004155414	-9.3305793205

extrapolated results for given energy

brass filter	thick(L/3)	energy (kV)
0	0	160
0.25	0.6299605249	561.6
0.5	0.793700526	188.3
1	1	78.4

polynomial regression for all brass thickness extrapolations

a	b	c
296.40435979	-779.6972329	561.575503782

extrapolated dose rate value for given brass thickness

brass thickness (mm)	dose rate (uGray/s)
0	263.5

reference configure 10 W 250 mm source-object distance

reference measurement s kV	Aluminium filter		
	0	1	1
50	326.1	159.6	289.9
100	464.4	289.9	345.6
130	511.8	345.6	381.6
150	542.4	381.6	460.5
200	606	460.5	

polynomial regression for all energies extrapolation $y=ax^2+bx+x$

Al thickness	a	b	c
0	-0.007071494	3.6019878758	166.263578871
1	-0.004899184	3.2128190662	-12.671791738

extrapolated results for given energy

Al thick	energy (kV)
0	160
1	561.6
1	401.3

linear regression for all Al thickness

a	b
-160.2476617	561.55140455

NOT ACCURATE, MORE CALIBRATION POINTS ARE NECESSARY

extrapolated dose rate value for given brass thickness

Al thickness (mm)	dose rate (uGray/s)
0	561.6

Supplementary Data 3

BIR references dose rates (Part 2)

reference configuration 15 W 250 mm source-object distance

brass thickness (mm)	brass thickness		
	0.25	0.5	1
50	481.5	24.3	0
100	686.4	133.8	32.7
130	757.2	206.4	73.5
150	795.9	252.9	103.8
200	874.8	371.7	186.3

brass filter thick²(1/3) energy (kV)

0	822.5
0.25	0.62996062 275.8
0.5	0.79370053 186.1
1	115.9

brass thickness (mm) dose rate (uG/yr/s)

0.1	385.2
-----	-------

polynomial regression for all energies extrapolation $y=ax^2+bx+x$

brass thickness	a	b	c
0	-0.01226628 5.64429401	233.4332192618	
0.25	0.00077883 2.12838098	-84.6708394258	
0.5	0.00353358 0.93060088	-51.2828403813	
1	0.00483315 0.05218724	-16.1488144708	

polynomial regression for all brass thickness extrapolations

a	438.779571 -1145.92076
b	822.5785963594
c	

reference configuration 20 W 250 mm source-object distance

brass thickness (mm)	brass thickness		
	0.25	0.5	1
50	632.4	31.5	6.6
100	906.9	176.7	95.7
130	996.3	271.5	171.6
150	1036	328.8	222.9
200	1148	477.9	355.1

brass filter thick²(1/3) energy (kV)

0	1078.5
0.25	0.62996052 358.6
0.5	0.79370053 244.4
1	152.5

brass thickness (mm) dose rate (uG/yr/s)

0.1	501.7
-----	-------

reference configuration 15 W 250 mm source-object distance

Al thickness	Aluminum filter		
	0	1	1
50	481.5	234.3	
100	686.4	428.4	
130	757.2	508.8	
150	795.9	559.5	
200	874.8	661.8	

extrapolated results for given energy

Al thick	energy (kV)
0	160
1	822.5
1	585.9

extrapolated dose rate value for given brass thickness

Al thickness (mm) dose rate (uG/yr/s)

0	822.5
---	-------

reference configuration 20 W 250 mm source-object distance

Al thickness	Aluminum filter		
	0	1	1
50	632.4	310.2	
100	906.9	568.5	
130	996.3	672.9	
150	1036	725.4	
200	1148	870.6	

extrapolated results for given energy

Al thick	energy (kV)
0	1078.5
1	769.1

extrapolated dose rate value for given brass thickness

Al thickness (mm) dose rate (uG/yr/s)

0	1078.5
---	--------

polynomial regression for all energies extrapolation $y=ax^2+bx+x$

Al thickness	a	b	c
0	-0.01226628 5.64429401	233.4332192618	
1	-0.00881673 5.03021051	6.804962112	

linear regression for all Al thickness

a	
b	
c	-236.573069 822.50344

NOT ACCURATE, MORE CALIBRATION POINTS ARE NECESSARY

polynomial regression for all energies extrapolation $y=ax^2+bx+x$

Al thickness	a	b	c
0	-0.01612429 7.38246634	310.0836445857	
1	-0.01136633 6.51672457	17.12903939547	

linear regression for all Al thickness

a	
b	
c	-309.436447 1078.5065

NOT ACCURATE, MORE CALIBRATION POINTS ARE NECESSARY

Supplementary Data 3

BIR references dose rates (Part 3)

25 W 250 mm source-object distance

brass filler thickness (mm)	brass thickness		
	0	0.5	1
50	776.1	81.1	0
100	1108	216	117.3 53.7
130	1236	336	213.6 120.9
150	1280	410.1	274.8 168
200	1409	568	434.4 298.5

energy (kV) **160**
 thick*(1/3)
 0 1236.6
 0.25 0.6299605249 443.5
 0.5 0.793700626 301.5
 1 187.5

brass thickness (mm) dose rate (uGray/s)
0.1 619.8

polynomial regression for all energies extrapolation $y=ax^2+bx+x$

brass thickness	a	b	c
0	-0.0268897091	9.3886795893	364.451992778
0.25	-0.0008419017	3.7741783427	-151.6223172818
0.5	0.0046694305	-1.731960352	-92.5731361525
1	0.007334461	0.1870993989	-30.1619164019

polynomial regression for all brass thickness extrapolations
 720.9020748247 -1864.0670764054 1329.7103494944

reference configuration 25 W 250 mm source-object distance

reference measurements kV	Aluminum filter		
	0	378	1
50	776.1	700.2	
100	1108	835.8	
130	1236	900.9	
150	1280	1067	
200	1409		

extrapolated results for given energy
 energy (kV) **160**
 Al thick 0 1236.6
 1 950.9

extrapolated dose rate value for given brass thickness
 Al thick
 ness
 (mm dose rate (uGray/s))
0 1329.6

polynomial regression for all energies extrapolation $y=ax^2+bx+x$

Al thickness	a	b	c
0	-0.0268897091	9.3886795893	364.451992778
1	-0.0159530139	8.4994373992	-3.1707162063

linear regression for all Al thickness
 a b
 -378.6565991787 1329.559070154

NOT ACCURATE, MORE CALIBRATION POINTS ARE NECESSARY

30 W 250 mm source-object distance

brass filler thickness (mm)	brass thickness		
	0	0.5	1
50	919.5	44.4	9.6 0
100	1328	259.5	141 65.1
130	1469	399	263.2 148.4
150	1535	469	329.1 201.3
200	1672	696.9	524.1 360.9

energy (kV) **160**
 thick*(1/3)
 0 1586.7
 0.25 0.6299605249 527.7
 0.5 0.793700626 361.5
 1 225.3

brass thickness (mm) dose rate (uGray/s)
0.1 738.6

reference configuration 30 W 250 mm source-object distance

reference measurements kV	Aluminum filter		
	0	442.2	1
50	919.5	833.1	
100	1328	891.5	
130	1469	1075	
150	1535		
200	1672		

extrapolated results for given energy
 energy (kV) **160**
 Al thick 0 1586.7
 1 1127.0

extrapolated dose rate value for given brass thickness
 Al thick
 ness
 (mm dose rate (uGray/s))
0 1586.7

polynomial regression for all energies extrapolation $y=ax^2+bx+x$

brass thickness	a	b	c
0	-0.0268897091	11.6565976314	410.9588120264
0.25	-0.0008054969	-4.6909735661	-185.919555121
0.5	0.0060076432	-1.9624263595	-106.2861647519
1	0.0091710682	0.1459327394	-32.8117515165

polynomial regression for all brass thickness extrapolations
 866.277159046 -2230.453469514 1586.78182299

NOT ACCURATE, MORE CALIBRATION POINTS ARE NECESSARY

polynomial regression for all energies extrapolation $y=ax^2+bx+x$

Al thickness	a	b	c
0	-0.0268897091	11.6565976314	410.9588120264
1	-0.021426551	10.6867104864	-34.2455145441

linear regression for all Al thickness
 a b
 -459.6678210706 1586.6762796382

Supplementary Data 3

BIR references dose rates (Part 4)

W 250 mm source-object distance

0	0.25	0.5	1
1220	59.1	13.3	0
1775	344.7	187.8	86.7
1925	527.4	334.2	187.8
2034	648	431.4	264.9
2199	922.5	687.3	473.4

thick^k (L/3) **160**
 energy (kV) **160**
 0 2094.2
 0.62996052 698.7
 0.79370053 475.3
 1 296.0

dose rate (μGray/s) **976.3**

polynomial regression for all brass thickness extrapolation $y=ax^2+bx+x$

brass thickness	a	b	c
0	-0.036355945	15.4730902	549.2586609142
0.25	-0.00121198	16.09230408	-244.8703994356
0.5	0.00752942	2.65669186	-142.5087264727
1	0.01192812	0.21558245	-43.8422390614

polynomial regression for all brass thickness extrapolations
 a 11.36.54874 -2936.50011 2094.4135384096
 b
 c

W 250 mm source-object distance

0	0.25	0.5	1
1505	72.3	15.6	0
2211	424.2	233.1	108.3
2412	651.6	417	234.9
2482	797.1	531.9	328.8
2714	1143	839.4	585

thick^k (L/3) **160**
 energy (kV) **160**
 0 2590.3
 0.62996052 862.7
 0.79370053 585.5
 1 367.4

dose rate (μGray/s) **1205.0**

reference configuration 40 W 250 mm source-object distance

reference measurements kV	Aluminium filter	
	0	1
50	1270	587.7
100	1775	1119
130	1925	1310
150	2034	1436
200	2199	1655

extrapolated results for given energy

energy (kV) **160**
 Al thick 0 2094.2
 1 1498.5

extrapolated dose rate value for given brass thickness

Al thickness (mm) **0**
 dose rate (μGray/s) **2094.2**

polynomial regression for all energies extrapolation $y=ax^2+bx+x$

Al thickness	a	b	c
0	-0.036355945	15.4730902	549.2586609142
1	-0.02960635	14.4341488	-52.9981178196

linear regression for all Al thickness
 a -595.633847 2094.17709
 b

NOT ACCURATE, MORE CALIBRATION POINTS ARE NECESSARY

reference configuration 50 W 250 mm source-object distance

reference measurements kV	Aluminium filter	
	0	1
50	1505	714
100	2211	1376
130	2412	1628
150	2482	1751
200	2714	2036

extrapolated results for given energy

energy (kV) **160**
 Al thick 0 2590.3
 1 1844.3

extrapolated dose rate value for given brass thickness

Al thickness (mm) **0**
 dose rate (μGray/s) **2590.3**

polynomial regression for all energies extrapolation $y=ax^2+bx+x$

Al thickness	a	b	c
0	-0.04725451	19.6091582	662.5839648008
1	-0.03758406	18.078417	-85.9877780484

linear regression for all Al thickness
 a -746.050931 2590.33382
 b

NOT ACCURATE, MORE CALIBRATION POINTS ARE NECESSARY

Supplementary Data 3

BIR references dose rates (Part 5)

75 W 250 mm source-object distance

brass filter thickness (mm)	brass thickness	
	0.25	0.5
50	2185	3916
100	3265	6177
130	3538	952.2
150	3706	1157
200	3998	1660

energy (kV) **160**
 Al thick 0 3831.4
 1 2721.9
 linear regression for all Al thickness a b
 -1109.54622 3831.44097

brass thickness (mm) **0.1**
 dose rate (uGray/s) **1765.1**

polynomial regression for all energies extrapolation y

brass thickness	a	b	c
0	-0.07415669	30.3074187	880.6651185579
0.25	-0.00239954	11.0437534	-450.10946
0.5	0.01319303	4.89736354	-204.379398
1	0.02126438	0.46535009	-82.5733317

polynomial regression for all brass thickness extrapc
 a b c
 2155.84911 -5452.98862 3831.69915

100 W 250 mm source-object distance

brass filter thickness (mm)	brass thickness	
	0.25	0.5
50	2913.33333	132.8
100	4236	796.8
130	4602	1229
150	4781	1494
200	5165	2152

energy (kV) **160**
 Al thick 0 4952.2
 1 3497.9
 linear regression for all Al thickness a b
 -1454.3165 4952.244

brass thickness (mm) **0.1**
 dose rate (uGray/s) **2282.2**

polynomial regression for all energies extrapolation y

brass thickness	a	b	c
0	-0.09115237	37.3793359	1305.05084
0.25	-0.00176908	13.9607054	-565.245612
0.5	0.01619717	6.43047162	-343.903398
1	0.02763889	0.46026162	-59.83395013

polynomial regression for all brass thickness extrapc
 a b c
 2760.37071 -7034.63804 4952.70595

reference configuration 75 W 250 mm source-object distance

reference measurements kV	Aluminum filter	
	0	1
50	2185	1045
100	3265	2022
130	3538	2389
150	3706	2662
200	3998	3025

extrapolated results for given energy

energy (kV) **160**
 Al thick 0 3831.4
 1 2721.9
 linear regression for all Al thickness a b
 -1109.54622 3831.44097

extrapolated dose rate value for given brass thickness

Al thickness (mm) **0**
 dose rate (uGray/s) **3831.4**

polynomial regression for all energies extrapolation y=ax²+bx+x

Al thickness	a	b	c
0	-0.07415669	30.3074187	880.6651185579
1	-0.026263921	26.1573478	-115.7171840628

reference configuration 100 W 250 mm source-object distance

reference measurements kV	Aluminum filter	
	0	1
50	2913.33333	1393
100	4236	2628
130	4602	3086
150	4781	3323
200	5165	3865

extrapolated results for given energy

energy (kV) **160**
 Al thick 0 4952.2
 1 3497.9
 linear regression for all Al thickness a b
 -1454.3165 4952.244

extrapolated dose rate value for given brass thickness

Al thickness (mm) **0**
 dose rate (uGray/s) **4952.2**

NOT ACCURATE, MORE CALIBRATION POINTS ARE NECESSARY

polynomial regression for all energies extrapolation y=ax²+bx+x

Al thickness	a	b	c
0	-0.09115237	37.3793359	1305.0508433146
1	-0.06834523	33.303335	-80.9682229284

2.4 Study 3: Immunity Capture

2.4.1 Background: Plague, HLA and other Immunity Genes

The plague pathogen *Yersinia pestis*

The goal of my third study was to capture human immune system genes from medieval victims of plague caused by the gram-negative bacterium *Yersinia pestis*, the etiological agent of the Black Death (Benedictow, 2004, Bos et al., 2011). *Y. pestis* DNA was previously detected in those plague victims allowing to reconstruct a complete *Y. pestis* genome (Spyrou et al., 2016).

Y. pestis is endemic in rodent populations and invades the human host through the bite of an infected rodent flea *Xenopsylla cheopis* (Bitam et al., 2010). In flea guts it is able to form biofilm (Hinnebusch and Erickson, 2008), whereas in the human host the pathogen is transported via the bloodstream into the lymph nodes, where it proliferates and produces toxins causing severe inflammation which results in swollen and painful lymph nodes, the characteristic "buboes" (Comer et al., 2010). This, the so called bubonic plague, is the most common form of plague. It has an incubation period of up to seven days and a mortality rate ranging from 30 - 60 % (WHO, 2019). Advanced bubonic plague can spread into the lungs and be transmitted via droplet infections from human to human, causing pneumonic plague (Butler, 1983). The pneumonic plague is the most virulent form of *Y. pestis* infection, which is almost always fatal if untreated, and characterized by symptoms such as severe pneumonia, high fever, cough, chills and other symptoms after a short incubation time of as short as 24 h (WHO, 2019). The third and rarest known form of plague is the septicemic plague. It is characterized by febrility, gastro-intestinal disturbances, chills and headache after a blood infection by *Y. pestis* (Perry and Fetherston, 1997).

In the recent years many studies have investigated the evolution of the plague pathogen (Achtman et al., 1999, Bos et al., 2011, Cui et al., 2013, Wagner et al., 2014, Rasmussen et al., 2015, Bos et al., 2016, Feldman et al., 2016, Andrades Valtuena et al., 2017, Spyrou et al., 2016, Spyrou et al., 2018, Keller et al., 2019). *Y. pestis* was shown to have been infecting humans since at least the Late Neolithic (Andrades Valtuena et al., 2017, Rascovan et al., 2019). The genetic changes that contributed to the increased virulence of *Y. pestis* include the upraise of the *ymt* gene, allowing the pathogen to survive in the flea gut and use the flea as a vector for transmission (Hinnebusch et al., 2002), a mutation in the *pla* gene allowing deep penetration into lung tissue and therefore essential only

for the pneumonic form of plague (Zimbler et al., 2015), as well as the loss of function of the *flhD* gene regulating the expression of flagellin (Minnich and Rohde, 2007), which triggers the mammalian innate immune system (Hayashi et al., 2001). Historical records indicate that the Black Death killed up to half of the European population from 1346 to 1353 (Benedictow, 2004) and plague still emerges nowadays, especially in developing countries.

Most of the genetic changes that modern day *Y. pestis* strains acquired after diverging from the Black Death strains do not contribute to an increased virulence of the plague bacillus (Bos et al., 2011). Thus, apart from the introduction of antibiotics, improved hygiene conditions or climatic changes and different vector dynamics, the host immunity may have adapted to the effect of *Y. pestis*. This scenario implies a plague driven positive selection on particular immunity related genes owned by plague victims, and further suggests that advantageous genes or alleles responsible for plague resistance should be present at increased frequencies in places previously affected by plague.

HLA and other immunity genes

With a current number of 25,756 alleles (hla.alleles.org, 2020) the Human Leukocyte Antigen (HLA) complex is the most polymorphic gene complex in the human genome (Albert and Baur, 1984). The HLA is the human version of the vertebrate Major Histocompatibility Locus (MHC) located on the short arm of the human chromosome 6 (Erlich et al., 2001). Balancing selection exerted through pathogens was proposed as a driving mechanism for maintaining the high allelic diversity of the HLA locus (Trowsdale, 2011). HLA genes encode cell-surface glycoproteins which present intracellular and extracellular peptides to the innate immune system and therefore play a major role for the immune system as well as for tissue compatibility after organ and bone marrow transplantations (Mahdi, 2013).

The HLA class I consists of the major genes A, B, C and the minor genes E, F, G, which encode cell surface proteins that present cell-own and viral peptides to CD8+ cytotoxic T-cells (Pamer and Cresswell, 1998). The HLA class II consists of the classical genes DP, DQ, DR and the non-classical genes DM and DO. HLA-DP, DQ and DR encode cell surface proteins that present extracellular antigens, e.g. bacterial peptides, to CD4+ helper T-cells (Watts, 1997). The non-classical HLA-II genes DM and DO encode non-binding proteins that regulate the loading of peptides from self and foreign antigens (Mellins and Stern, 2014).

The leading question here was whether plague victims carried disadvantageous alleles

for the presentation of *Y. pestis*-derived peptides while plague survivors had advantageous alleles that they would have inherited to following generations. In short, whether a contemporary population has significantly different HLA-allele frequencies compared to plague victims who existed in the same geographic region.

Other immunity-related genes important for the immune response to pathogens such as *Y. pestis* comprise pattern recognition receptors (PRRs) involved into the direct recognition of bacterial Pathogen-Associated Molecular Patterns (PAMPs) and signalling molecules. Main PRR gene families include Toll-like Receptors (TLRs) (Hayashi et al., 2001, Lu et al., 2008), Interleukins (IL) and Interleukin Receptors (ILRs) (Goritzka et al., 2015), Killer-Cell-Immunoglobulin-like Receptors (KIRs) (Augusto and Petzl-Erler, 2015), inflammasome-involved Nucleotide-binding Domain and Leucine-rich Repeat containing Receptors (NLRs) (Man and Kanneganti, 2015), signal transducing Leukocyte Immunoglobulin-like Receptors (LILRBs) (Kang et al., 2016), as well as other cell-surface glycoproteins from the large Cluster of Differentiation (CD) gene family (Nymo et al., 2016). More than that, proteasome (PSM) genes are involved in the degradation of infected cells and the regulation of the NF- κ B pathway (Sun et al., 2016) which plays an important role in immune responses. Cytokines such as Chemokine Ligands (CCLs) (Bardina et al., 2015) and their Chemokine Receptors (CCRs) (Fricker, 2015), Tumor Necrosis Factors (TNFs) and Interleukins (ILs) (Nymo et al., 2016), Interferons (IFNs) (Wang et al., 2016), as well as transcription factors such as NF- κ B (Kopitar-Jerala, 2015) are involved into signal-transduction and induce inflammatory responses to pathogens.

Interaction between *Y. pestis* and the Human Immune System

After having penetrated through the human skin barrier *Y. pestis* is recognized by the Toll-like receptor 4 (TLR4) which senses lipopolysaccharides (LPS) that constitute a major part of the cell wall in gram-negative bacteria (Takahashi and Kawai, 2007). TLR4 recruits downstream signal transduction adaptor proteins which then induce the expression of proinflammatory cytokine genes and transcription factors such as AP-1 and NF- κ B (Lu et al., 2008). However, *Y. pestis* can evade recognition by altering the fatty acid chains of the LPS Lipid A from hexa-acylated to tetra-acylated after the temperature transition from flea (26 °C) to human (37 °C) (Li and Yang, 2008). Moreover, when *Y. pestis* is phagocytosed by macrophages at an early stage of infection, it can proliferate in the macrophages, where it is protected from other components of the human immune system (Li and Yang, 2008), and produces virulence factors such as *Yersinia*

outer proteins (Yops), F1 antigen and LcrV antigen (Perry and Fetherston, 1997). The F1 antigen forms a capsule around the bacterium protecting it from engulfment through macrophages. In macrophages *Y. pestis* produces type III secretion system (T3SS) surface proteins (Cornelis, 2002) that allow the bacterium to inject Yops virulence factors into host cells such as macrophages, neutrophils and dendritic cells after the release from macrophages and disable host immune responses. E.g. YopM can deplete Natural Killer cells (NK), whereas YopH, YopE, YopT and YopO inhibit phagocytosis by macrophages and neutrophils (Li and Yang, 2008). YopP inhibits TNF- α and IL-8 release from infected cells, which suppresses the activation of NK cells (Boland and Cornelis, 1998). Moreover, the *Y. pestis* LcrV antigen that is involved into the T3SS formation has been shown to also suppress proinflammatory TNF- α and IFN- γ cytokines (Sing et al., 2002) and to inhibit the chemotaxis of neutrophils to the inflammation site (Welkos et al., 1998).

Host dendritic cells (DCs) sense invading pathogens and trigger the release of PRRs to activate T-cells, while disintegrating the pathogen within their phagosome and presenting microbial peptides with their HLA class II molecules to CD4+ T-cells (Anderson et al., 2017). *Y. pestis* paralyzes DCs by diminishing their cytoskeleton rearrangement (Velan et al., 2006). More than that, the YopH effector protein depletes T-cells by inducing their apoptosis (Bruckner et al., 2005).

2.4.2 Initial Array Design for Immunity Capture

Microarrays for target enrichment of nucleic acids can be purchased commercially. They are available for various purposes (e.g. gene expression analysis, (micro-) RNA and DNA capture) in various layouts, depending on the amount of desired probes (ranging from 244 000 to 1 million (mio) features, e.g. *Roche NimbleGen Sequence Capture*, *Agilent eArray*). Using an online user interface (e.g. *Agilent SureDesign*) customers can upload their target sequences which will be printed as nucleotide sequences on the surface of the microarray. After DNA extraction and purification, and the construction of DNA sequencing libraries, the indexed double stranded libraries are being denatured into single strands and applied to the surface of the microarray. The hybridization between target molecules and bait on the solid surface of the microarray happens at a temperature of 65 °C during an incubation period of 48-65 h (Hodges et al., 2009). After the hybridization, non-binding (non-target) DNA molecules are washed away, and the target DNA is eluted and purified. The target molecules are amplified using PCR, quantified and become ready for sequencing. Optionally, a second round of capture can

be applied to the captured target DNA using a new array.

In a first approach genes involved in the innate immune system were selected including a set of different Pattern Recognition Receptors (PRR) and a panel of 11179 worldwide alleles of the Human Leukocyte Antigen (HLA) gene complex (Supplementary File 1). To encompass the whole present allelic diversity of the HLA and KIR genes a dataset was compound from the Immuno Polymorphism Database (Robinson et al., 2013). The locations of the genes in Supplementary File 1 and their exonic and intronic regions were obtained from the Ensembl Genome browser (Aken et al., 2017) with reference to the human genome assembly hg19 (Lander et al., 2001). Based on the chromosomal coordinates, exonic nucleotide sequences were extracted from the hg19 nucleotide sequence for all the target genes. For the HLA genes, the KIR genes, the MIC (MHC class I chain related, (Collins, 2004)), and the TAP genes (transporter associated with antigen processing, (McCluskey et al., 2004)) also the intronic regions were obtained. According to *Agilent* a probe length of 60 nucleotides provides the optimal balance between sensitivity and specificity. 60 bp long probes were designed from the previously extracted target sequences based on a 3 bp overlap (tiling). The target regions were extended to include 60 bp into the flanking regions. All duplicate probe sequences were filtered out yielding 762833 unique probes with the fraction of HLA probes constituting 37 % of all probes. The probeset was uploaded onto the *Agilent SureDesign* server and a 1 mio. feature microarray was ordered (fig. 2.7).

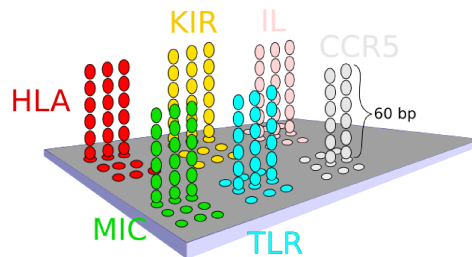


Figure 2.7: Simplified illustration of an *Agilent* capture array with up to one million clusters of molecular 60 bp long probes on its surface, designed to target human immunity-related genes.

To evaluate the target enrichment performance of the immunity capture array, human DNA samples with known HLA-genotypes were used as positive controls. These samples were previously HLA-genotyped and provided by Paul J. Norman (Stanford University). The modern human DNA was sheared to match the average fragment length expected from ancient DNA and turned into sequencing libraries as described in section 1.3. Cap-

ture was performed as described above before sequencing.

HLA genotypes were assessed in *OptiType*. *OptiType* is an in-silico prediction software for 4-digit HLA typing. It identifies a combination of HLA class I alleles as the correct genotype, if this combination maximizes the number of aligned reads (Szolek et al., 2014). *OptiType* uses reads with a length of 37 bp to 101 bp and can work with as little as a 12 fold coverage on the HLA loci, which makes it suitable for aDNA analyses. *OptiType* was able to identify the correct genotypes for the Stanford reference samples, which verified the reliability of the immunity capture array.

After mapping to hg19 the sequencing data was visually inspected in the *Integrative Genomics Viewer* (IGV) (Robinson et al., 2011): an over-representation of sequences mapping to HLA genes could be observed yielding high coverage of the captured HLA regions, but only shallow coverage of other immunity-related genetic regions of interest. This was due to the high proportion of HLA probe sequences determined by the exceptionally high polymorphism of the HLA region that was completely encompassed during the initial array design. This not only left little room for probes of other target genes, but also caused an uneven distribution of probes: the most polymorphic regions had significantly more probes than less polymorphic regions. Moreover, repetitive intronic regions had not been masked thus their probes would hybridize to sequences from multiple regions in the human genome causing peaks of high coverage.

2.4.3 Improved Array Design

In order to solve the problem of the capture bias towards the HLA-genes, in the next approach the number of probes for the HLA locus was reduced to a selection of "sensitive" probes: using a greedy algorithm a subset of full-length alleles was selected whose tiling can encompass most of the previously designed 60 bp probes with up to six mismatches. This set included probes that mainly originated from large HLA allele families (such as A*24:02, B*39:01 or C*16:01) allowing a probe to target multiple HLA alleles within the same allele family at the same time. Since a probe of 60 bp length allows to hybridize a complementary molecule with a tolerance of up to 6 nucleotide mismatches, with the new probeset 95 % of all possible HLA polymorphisms could be covered. On top, a set of unique probes was selected to target the remaining 5 %, but using 160-mers instead of full alleles. Moreover, repetitive intronic regions were excluded, the tiling was changed to 5 bp overlap (fig. 2.8) and duplicate probe sequences were removed, thus, reducing the total proportion of HLA probes to 15 %.

In addition, a more stringent filtering was applied to the new probeset: the probe se-

```

CACGTTTCTGGAGCAGGTTAAACATGAGTGTCAATTCCTCAACGGGACGGAGCGGGTGCGGTTCTGGA
CACGTTTCTGGAGCAGGTT
TTCTTGGAGCAGGTTAAACA
GGAGCAGGTTAAACATGAGT
AGGTTAAACATGAGTGTCAAT
AAACATGAGTGTCAATTCCT
TGAGTGTCAATTCCTCAACG
GTCATTCCTCAACGGGACG
TTCTTCAACGGGACGGAGCG
CAACGGGACGGAGCGGGTGC
GGACGGAGCGGGTGCGGTTCT
GAGCGGGTGCGGTTCTGGA

```

Figure 2.8: Tiling. An example target sequence covered by 20 bp long probes with a 5 bp tiling.

quences that remained after duplicate removal were aligned to the human reference sequence hg19 with the sensitive mapper *RazerS3* (Weese et al., 2012) using an identity threshold of at least 95 % between the probe and the reference sequence. Using the program *Samtools* (Li et al., 2009) the individual number of alignments to hg19 was counted for each probe. Probe sequences that aligned more often than 20 times were regarded as unspecific and were removed from the probeset. The remaining probes were re-aligned to hg19 to ascertain that no probe mapped more frequently than 20 times. The final probeset contained 229604 unique probes and was ordered using the *Agilent SureDesign* service. After capture and sequencing of the Stanford reference samples, their HLA alleles were correctly called in *OptiType* and reaffirmed the accuracy of the improved array design. Moreover, evaluation of the re-designed immunity array demonstrated that the coverage of the non-HLA genes now was comparable to the coverage of the HLA genes (fig. 2.9).

2.4.4 In-Solution Capture

In the third approach probes were re-designed to meet the specifications of in-solution capture in order to standardize the immunity capture approach and save costs (Gnirke et al., 2009). In-solution capture has been widely used in aDNA studies e.g. to investigate population genetics, human migrations and selection by targeting different panels of polymorphic sites (Fu et al., 2013, Haak et al., 2015, Mathieson et al., 2015).

In contrast to solid capture, the bait is not bound to a solid surface of a microarray but free in buffer solution. As a first step, however, the bait is ordered on an *Agilent* microarray and cleaved from its surface. Since it is modified to contain a primer sequence, bait can be re-amplified using PCR, followed by biotinylation in a second round of PCR (Fu et al., 2013). Biotinylated bait is then pooled together with the sequencing libraries so that complementary target DNA molecules can hybridize to the bait. Because of biotin-streptavidin interaction, the addition of streptavidin-coated magnetic beads (e.g.

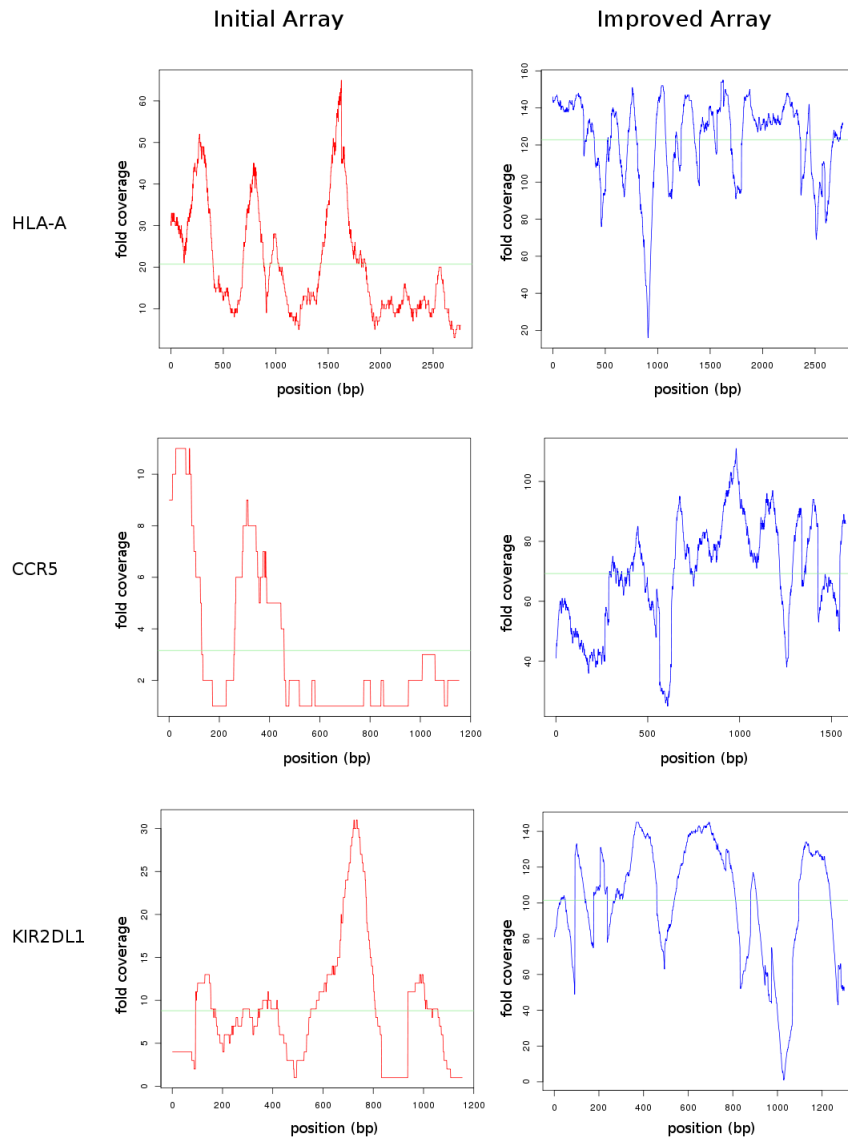


Figure 2.9: An example of coverage comparison based on the coverage of the captured genes HLA-A, CCR5 and KIR2DL1 using the same sample for the initial and the improved capture array. The average sequence coverage for the respective genes (indicated by the green line) is generally higher after capture with the improved array design compared to the first array design. Plots were generated in R statistical language (R Development Core Team, 2008).

AMPure XP SPRIselect) allows to select the target fraction of the DNA. Unbound DNA is washed away while bound target DNA is eluted from the bait, purified, amplified and sequenced.

Probes from the improved array design were modified by replacement of the last 8 bp from the 3' end of each probe by a specific primer sequence, that allows to re-amplify

the probe molecules in a PCR after cleaving them off from the surface of the micro-array (Fu et al., 2013). Trimming off the last 8 terminal bases was performed by use of the *FASTX* toolkit (HannonLab, 2009) followed by replacing them with the actual primer sequence, removal of duplicate sequences, alignment to the human reference sequence hg19 and filtering for probe sequences that aligned more than 20 times as described above (sec. 2.4.3).

2.4.5 Evaluation of the In-Solution Capture Efficiency

Evaluation of the target enrichment was conducted using the HLA-genotyped positive controls as described in section 2.4.2 and *OptiType* was able to identify the correct genotypes for the human control samples consolidating the accuracy of the in-solution immunity capture. In addition, the sequencing libraries were shotgun sequenced without prior enrichment. After sequencing, the total number of reads after adapter removal, the number of reads that aligned (mapped) to the human genome reference sequence hg19 and the percentage of reads mapping to the hg19 human reference sequence (endogenous DNA) were assessed within the EAGER-pipeline (Peltzer et al., 2016), while the average read length and the number of reads in target were obtained using a custom program. The enrichment factor was then calculated as follows: for both - the captured and the shotgun libraries - the number of reads in target regions was normalized by the number of reads that mapped to hg19. The normalized number of reads in target after capture was divided by the normalized number of reads in target after shotgun sequencing (tab. 2.1). On average, this number was 640 times higher after in-solution capture indicating a high enrichment efficiency.

2.4.6 Application of the Immunity Capture to 16th Cent. Plague Victims

Archaeological Material

The novel in-solution immunity capture was applied to DNA obtained from human skeletal material of plague victims of a historically documented plague outbreak in Ellwangen (Baden-Wuerttemberg, South Germany). 44 individuals were obtained from three mass graves (archaeological records 549, 559, 706) during an excavation led by the ministry for the preservation of monuments (Landesamt für Denkmalpflege), Baden-Wuerttemberg in 2014 and 2015 (fig. 2.10). Radiocarbon dating was performed on 12 of those indi-

Table 2.1: EAGER results and capture efficiency for 10 human DNA positive control samples with previously established HLA-genotypes. *Capture* indicates libraries after the in-solution immunity capture, *Shotgun* indicates the libraries without previous enrichment. *Endogenous DNA (%)*: percentage of total reads mapping to hg19. *Reads in Target*: Number of reads falling into the target regions of the in-solution immuno capture. *Enrichment Factor*: $\frac{\text{normalized reads in target after capture}}{\text{normalized reads in target after shotgun}}$

Sample	Total Reads after Adapter Removal	Mapped Reads (hg19)	Endogenous DNA (%)	Average Read Length	Reads in Target	Enrichment Factor
COX_Capture	5 329 525	5 028 640	94.4	75.4	1 864 643	668.9
COX_Shotgun	10 617 305	9 748 356	91.8	72.1	5 404	
GBL_Capture	7 346 822	7 046 816	95.9	75.6	2 584 206	639.6
GBL_Shotgun	9 844 279	9 268 587	94.2	72.3	5 314	
MOU_Capture	5 899 988	5 668 114	96.1	75.5	2 024 780	632.3
MOU_Shotgun	9 275 108	8 756 612	94.4	72.5	4 947	
PBB_Capture	6 074 598	5 803 782	95.5	75.5	2 011 848	631.9
PBB_Shotgun	10 141 481	9 489 475	93.6	72.5	5 206	
PGF_Capture	4 379 895	4 197 934	95.8	75.6	1 553 686	628.3
PGF_Shotgun	8 361 144	7 845 175	93.8	72.4	4 621	
RSH_Capture	5 408 967	5 170 051	95.6	75.6	1 951 595	659.7
RSH_Shotgun	7 600 117	7 112 366	93.6	72.8	4 070	
SPL_Capture	4 053 284	3 864 052	95.3	75.6	1 419 911	640.9
SPL_Shotgun	9 058 986	8 494 267	93.8	72.6	4 870	
SSTO_Capture	8 164 540	7 718 107	94.5	75.6	2 396 694	596.2
SSTO_Shotgun	8 017 032	7 361 621	91.8	71.9	3 834	
T7526_Capture	3 215 111	3 064 910	95.3	75.6	1 106 450	623.9
T7256_Shotgun	8 336 150	7 787 939	93.4	72.7	4 506	
WT49_Capture	6 538 212	6 253 695	95.6	75.6	2 325 988	673.4
WT49_Shotgun	10 635 275	9 955 516	93.6	72.4	5 499	

viduals. One individual (Ellwangen 2) had too little collagen and could not be dated. One individual (Ellwangen 7) was dated to the 13th - 14th century AD, while the remaining 10 samples showed ages falling into the 15th - 17th century AD (tab. 2.2).

Modern DNA for Reference

Saliva samples for DNA extraction were obtained from 51 modern day inhabitants of Ellwangen in September 2014 and April 2015 using *Whatman Omniswab* cheek swabs. Only individuals whose families have been resident since at least four generations in Ellwangen were asked to contribute their saliva samples, in order to avoid a potential bias that might have been caused through admixture because of migration. Consent was obtained from the contributing persons and their samples were made anonymous disallowing a re-identification of the donor.



Figure 2.10: Mass graves discovered at the market place in Ellwangen, Baden-Wuerttemberg during an excavation in 2014/2015 led by the ministry for the preservation of monuments (Landesamt für Denkmalpflege), Baden-Wuerttemberg. Images were kindly provided by Prof. Dr. Joachim Wahl.

Sampling and DNA Extraction

Former studies have shown that the petrous portion of the temporal bone in the (human) skull provides five times more endogenous human DNA compared to teeth (Gamba et al., 2014, Pinhasi et al., 2015). Therefore petrous bones from the plague victims were obtained for sampling. Following Pinhasi *et al.* 2015, petrous pyramids were cut longitudinally in order to enable access to the bony labyrinth (fig. 2.11), which is the densest part of the mammalian body (Frisch et al., 1998). After cleaning the surface on one side of the bony labyrinth with a drill bit, bone powdering was performed along the semi-circular canal, which yielded 80-120 mg bone powder. DNA extraction was carried out using up all the bone powder following a guanidinium-silica based extraction method (Rohland and Hofreiter, 2007).

Isolation of genomic DNA from saliva was performed using the QIAamp DNA Blood Mini Kit following the Quiagen protocol. Isolated modern DNA was sheared to an average fragment length of 300 bp using the Covaris M220 Focused-ultrasonicator.

Preparation of DNA Libraries and In-Solution Capture

DNA libraries from modern and ancient samples were prepared using 20 μ l of extract following published protocols (Kircher et al., 2012, Meyer and Kircher, 2010). For the ancient samples partial uracil-DNA-glycosylase (UDG-half) treatment was applied

Table 2.2: Radiocarbon dates of 11 out of 44 Ellwangen mass grave samples. Cal 1 sigma: sigma 1 statistical estimation of the calibrated date falling between the intercepts at a 68 % probability. Cal 2 sigma: sigma 2 statistical estimation of the calibrated date falling between the intercepts at a 95 % probability.

Sample	Cal 1 sigma	Cal 2 sigma
Ellwangen 1	cal AD 1468-1619	cal AD 1454-1631
Ellwangen 3	cal AD 1493-1630	cal AD 1473-1634
Ellwangen 4	cal AD 1473-1622	cal AD 1455-1633
Ellwangen 5	cal AD 1484-1626	cal AD 1460-1633
Ellwangen 6	cal AD 1459-1616	cal AD 1450-1630
Ellwangen 7	cal AD 1290-1385	cal AD 1284-1390
Ellwangen 8	cal AD 1449-1607	cal AD 1443-1618
Ellwangen 9	cal AD 1471-1620	cal AD 1455-1632
Ellwangen 10	cal AD 1485-1627	cal AD 1462-1633
Ellwangen 11	cal AD 1495-1632	cal AD 1481-1638
Ellwangen 12	cal AD 1425-1441	cal AD 1415-1447

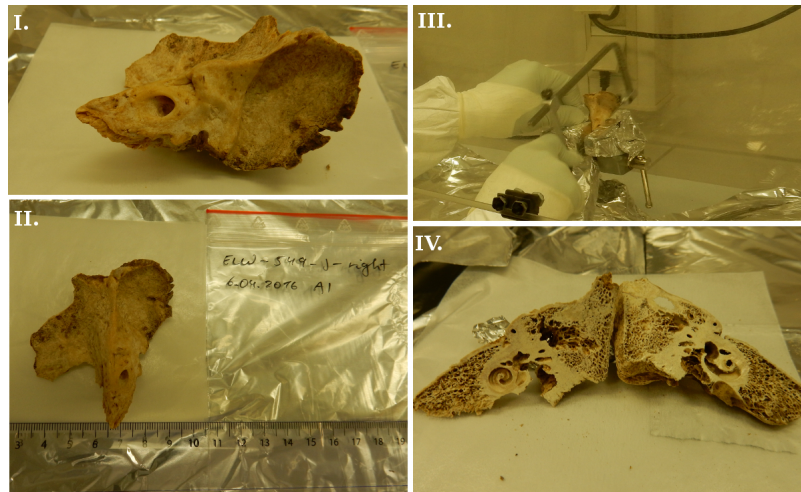


Figure 2.11: Sampling the petrous bone. **I.** and **II.** Right petrous bone from an individual obtained from the mass grave 549. **III.** Longitudinal cut through the petrous bone; **IV.** Medial view of a longitudinal cut through the petrous bone exposing the densest part of the bone and the semi-circular canal, where drilling was performed to produce bone powder.

(Rohland et al., 2015). Sample-specific index combinations were added to the sequencing libraries in order to allow differentiation between the individual samples after pooling and multiplex sequencing (Kircher et al., 2012). Indexed libraries were amplified in 100 μ l reactions in a variable number of 1 to 7 cycles to reach the required concentration for enrichment, followed by purification over Quiagen MinElute columns. Indexed library

pools were single-end shotgun-sequenced on the Illumina Hiseq 4000 with 75+8 cycles providing around 10-11 million sequencing reads per sample. Finally, using the latest design of probes for in-solution capture, target enrichment was performed on the same libraries. The indexed, captured library pools were then single-end sequenced on the Illumina Hiseq 4000 with 75+8 cycles providing around 11 million reads per sample.

Processing the Raw Sequences

Sequences were sorted by their index combinations and Illumina sequencing adapters were removed. An initial analysis of the sequence data was performed using the EAGER pipeline (Peltzer et al., 2016). Pre-processed sequences were mapped to the human reference sequence hg19 (Lander et al., 2001) using BWA 0.7.12 (Li and Durbin, 2010) with a default seed length of 32, mapping quality of 30 and a reduced mapping stringency "-n 0.01" to account for mismatches in ancient DNA from the medieval plague material. C to T misincorporation frequencies typical of aDNA were obtained using mapDamage 2.0 (Jonsson et al., 2013) in order to assess the authenticity of the ancient DNA fragments (Briggs et al., 2007). After the validation of terminal damage, the positions with substitutions at both ends of the reads were trimmed off in order to remove terminal damage. Mapping the saliva DNA sequence reads to hg19 was performed with the same parameters as for aDNA reads to assure comparability. Mapping results are shown in Supplementary File 2, Part 1, and Supplementary File 3, Part 1, respectively.

Evaluation of the Enrichment Efficiency on aDNA

The enrichment efficiency of the in-solution immuno capture was evaluated as described in sec. 2.4.5, but using the ancient DNA samples instead of the modern DNA control samples. The normalized amount of reads in target was on average 370 times higher after capture (fig. 2.12).

Combination of Capture and Shotgun Sequences

Both sequence data sets - the sequenced capture and shotgun libraries - were combined into a single dataset in order to maximally increase the number of various SNPs suitable for sex determination, contamination estimation, kinship inference and population genetic analyses, such as PCA and admixture.

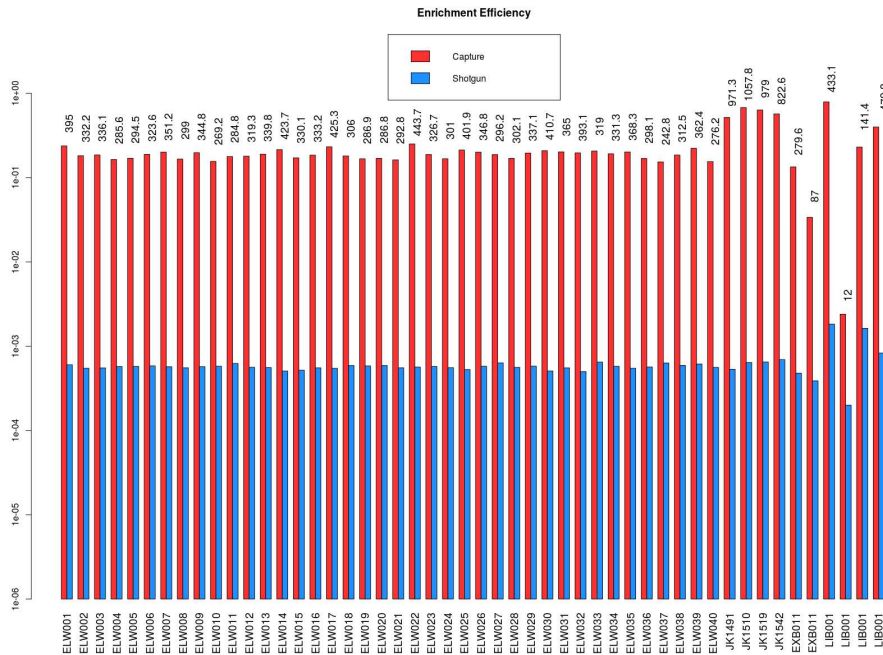


Figure 2.12: Target DNA enrichment after in-solution capture ("capture", red) and uncaptured libraries ("shotgun", blue). The number of reads in target is normalized by the total number of reads after adapter clipping and shown on a logarithmic scale. "EXB" = extraction blanks, "LIB" = library blanks. The enrichment factor ($\frac{\text{normalized reads in target after capture}}{\text{normalized reads in target after shotgun}}$) is shown on top of each pair of samples. On average, 370 times more reads end up in target after in-solution capture compared to sheer shotgun sequencing.

Sex Determination

Sex determination was assessed on shotgun-sequencing data of ancient and modern DNA samples based on the ratio of sequences aligning to their X and Y chromosomes (Skoglund et al., 2013). In males the ratio between the coverage on the X chromosome and the average coverage of the autosomes should be around 0.5, since males have only one copy of the X chromosome, whereas in females it should be around 1. In analogy, the coverage on the Y chromosome should be around half the average coverage of autosomes in males, and it should be zero in females. 26 out of 44 individuals in the ancient Ellwangen population and 30 out of 53 individuals in the modern population were found to be males (Supplementary File 2, Part 1 and Supplementary File 3, Part 1), while the number of female individuals in each population was 18. In five cases of the modern Ellwangen population the genetic sex could not be determined due to low sequence coverage of the sex chromosomes.

Contamination Estimation

Since aDNA is present in only minute amounts, contamination from contemporary human DNA constitutes a major problem especially for the work with ancient human genetic material. Contaminated samples must be excluded from the analysis. A way to investigate whether a sample is contaminated is by detection of 5' terminal cytosine to thymine (C→T) substitutions caused through deamination of cytosine (Hofreiter et al., 2001) and, accordingly, 3' terminal substitutions of guanine by adenine. These substitutions accumulate over time and can be used to authenticate ancient DNA (Briggs et al., 2007, Krause et al., 2010, Sawyer et al., 2012). A low amount of terminal substitutions in an ancient sample can indicate contamination through modern DNA molecules. The 5' and 3' terminal substitution patterns were assessed in mapDamage 2.0 (Jonsson et al., 2013). The 5' C to T substitution frequencies of 16-30% in non UDG-treated aDNA from tooth samples, but also 5-9% in partially UDG-treated aDNA from bone samples were consistent with the deamination frequency expected in authentic aDNA (fig. 2.13; Supplementary File 2, Part 1). Partial UDG-treatment removes uracils within the ancient molecules in the same way as full UDG-treatment, but preserves a small fraction of the terminal uracil substitutions, so that population genetic analyses can be performed while the signal of authenticity of aDNA is retained (Rohland et al., 2015).

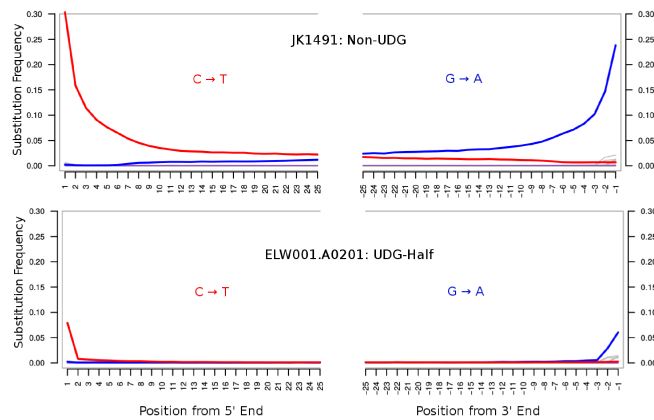


Figure 2.13: Example of 5' terminal C→T and 3' terminal G→A nucleotide substitutions observed in the human DNA of two Ellwangen plague victims. Above: a tooth sample that has not been treated with UDG (non UDG) and therefore contains the full substitution patterns. Below: a petrous bone sample partially treated with UDG (UDG-half) showing only partial terminal substitutions.

Terminal substitution patterns can be used to automatically estimate contamination in mitochondrial data. *Schmutzi* is a software that provides contamination estimates based

on deamination patterns and fragment length distributions of human mitochondrial sequences, and by utilizing a sequence database of present day human mitochondrial DNA which can be a potential source of contamination (Renaud et al., 2015). Shotgun sequencing data of the Ellwangen plague samples was mapped to the human mitochondrial reference sequence (revised Cambridge Reference Sequence, rCRS) and *Schmutzi* was run in order to obtain mitochondrial contamination estimates. Because of low sequence coverage, contamination estimation on the mitochondrial level could be retrieved only for 23 out of 44 samples. However, for these 23 samples, the mitochondrial contamination estimates suggested a low contamination level of 1-3% (Supplementary File 2, Part 1). In addition, in males, contamination can be detected by observing heterogeneity on the X-chromosome. Since males have only one copy of the X-chromosome, homogeneity is expected. Therefore, contamination on the X-chromosome in males can be estimated from the level of heterogeneity of the X-chromosomal sequences. Male X-chromosomal contamination was assessed in *ANGSD* (Korneliussen et al., 2014). Also here, contamination was generally low (Supplementary File 2, Part 1). However, in the tooth sample JK1519 the X-contamination estimate reached 32%, indicating that this sample was heavily contaminated. JK1519 was excluded from further analyses.

Testing for Population Continuity

In order to assess whether the modern and the ancient populations are comparable on the immunogenetic level, the genetic continuity between both populations was assessed in Principal Component Analysis (PCA) and Admixture analysis. The PCA is a statistical method that makes use of linear combinations to reduce variance of data to the most crucial components, the so called principal components (Pearson, 1901). Given a set of n samples the PCA can be used to project them into a two-dimensional space, representing data redundancy as correlation. The first component thereby contains the highest proportion of the variation within the data and the second component the second highest proportion. Applied to population genetic studies, ancestry differences between samples can be modeled along continuous axes of variation, where a closer genetic affinity can be assumed between samples that fall closer together. In order to run a PCA, the sequence data was transformed into the EIGENSTRAT format (Price et al., 2006) and merged with 29 West Eurasian populations on a set of 1.233.013 genetic markers (Haak et al., 2015, Mathieson et al., 2015). A PCA was performed using the *EIGENSOFT* program *smartpca* (Patterson et al., 2006) to calculate a basemap defined by the principal components of the 29 selected populations and to project the ancient

(plague) and modern Ellwangen individuals onto this map (fig. 2.14). Both populations fell on top of each other indicating genetic continuity. Since the dataset did not contain German individuals, the Ellwangen samples were merged with the *Affymetrix Human Origins* dataset, which contains 203 modern human populations genotyped at 594.924 SNPs (Lazaridis et al., 2014), including Germans. A PCA was performed using 67 selected West Eurasian populations (Lazaridis et al., 2016). Both Ellwangen populations overlapped and fell next to German individuals as expected (fig. 2.15). However, the reduced amount of SNPs caused a strong scattering within the ancient Ellwangen samples. Especially, the tooth samples JK1510, JK1519 and JK1542 had too little SNPs and fell outside the rest of the ancient Ellwangen population (fig. 2.16). Consequently, a threshold of at least 10.000 SNPs pulled down from the 1.233.013 or 594.924 SNPs genetic markers, resp., was defined for every sample to be included into further analyses.

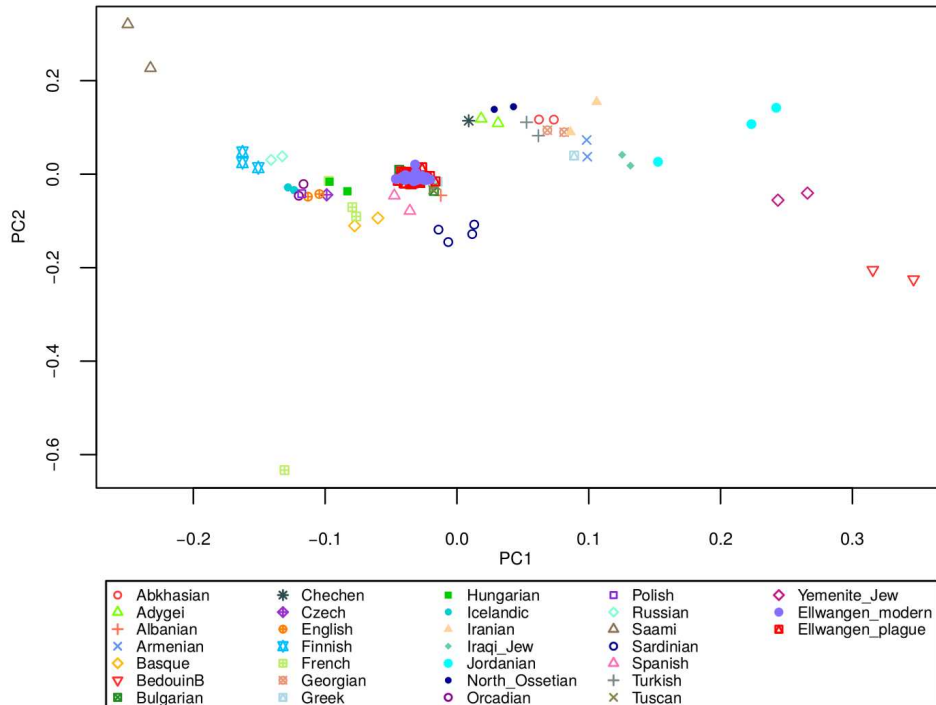


Figure 2.14: Principal component analysis of both the ancient and the modern Ellwangen individuals projected onto a basemap of 29 West Eurasian populations. The overlap between the ancient and modern Ellwangen population indicates genetic continuity.

Another approach to investigate the genetic continuity is by investigating the proportion of genetic ancestry between individuals. The software *ADMIXTURE* uses maximum likelihood estimates of the underlying admixture coefficients for a model-based

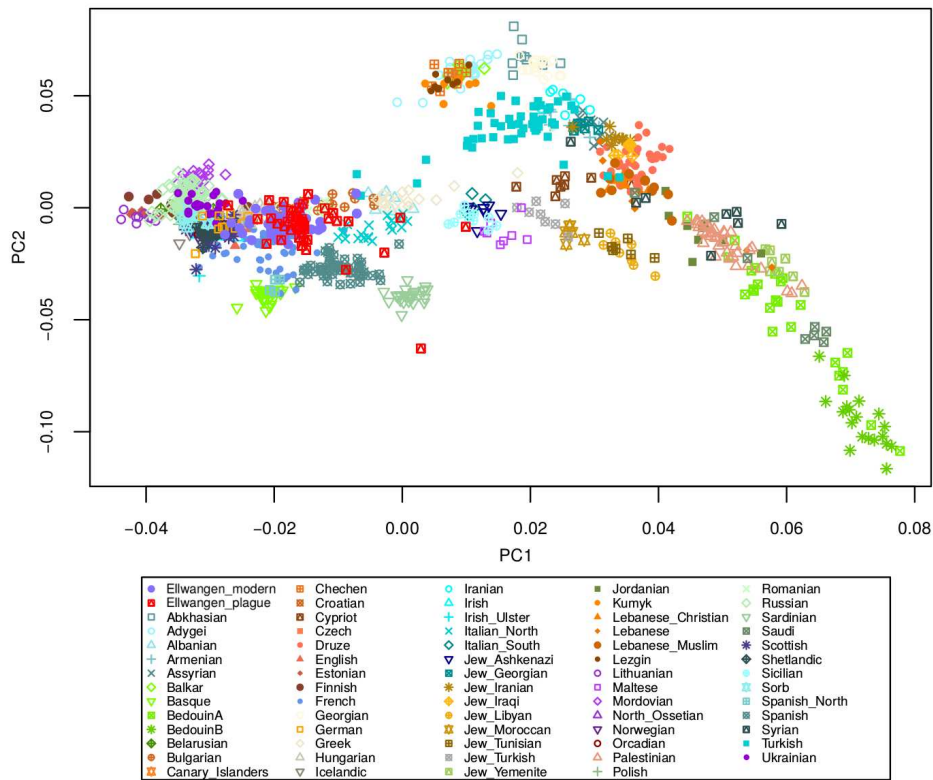


Figure 2.15: Principal component analysis of both Ellwangen populations projected onto a basemap of 67 West Eurasian populations. Except for outliers mainly caused by too little SNPs, samples from both populations cluster in close proximity to Germans and on top each other indicating genetic continuity between both populations.

estimation of ancestry in unrelated individuals (Alexander et al., 2009). *ADMIXTURE* can be used with different numbers of ancestral components (K) to operate a model of ancestral admixture proportions. Cross-validation is performed for every model and the model with the highest accuracy is determined by the lowest cross-validation error. The ancestral component composition of the model with the lowest cross-validation error ($K=4$) showed similar proportions of the ancestral components in the modern and ancient Ellwangen populations (fig. 2.17). The similarity in the ancestral genetic components observed in the admixture analysis and the coherent positioning of samples from both Ellwangen populations demonstrated by the PCA analysis confirmed genetic continuity between both populations, despite potential admixture that might have happened in the last 400 years and particularly since Ellwangen was occupied during the Thirty Years' War by the Swedish in 1632 (Ellwangen, 1886).

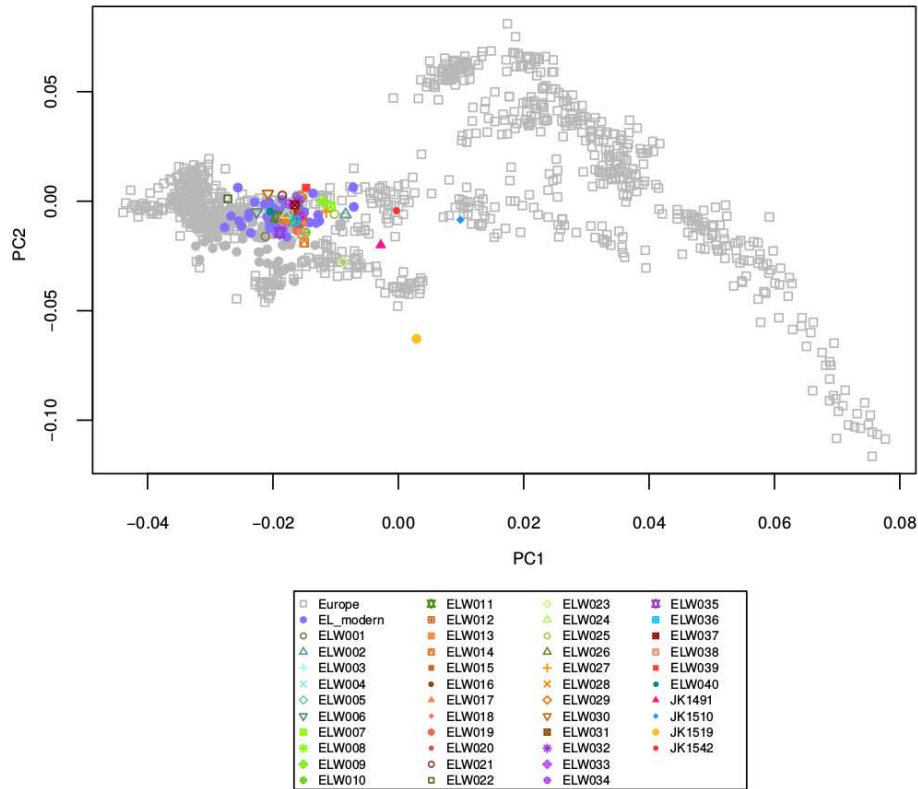


Figure 2.16: Principal component analysis of the same populations as in fig. 2.15. Modern Ellwangen individuals (EL_modern) are highlighted in purple. Plague victims are accentuated by different colors and symbols.

Kinship Analysis

Directly related individuals share common (HLA-) genotypes which can bias the allelic frequency distributions in a population. Therefore relatedness needs to be assessed in order to exclude directly related individuals from subsequent analyses. Relatedness was assessed using the software packages *READ* (Monroy et al., 2017) and *lcMLkin* (Lipatov et al., 2015) which were designed for low coverage sequence data as it is often given by aDNA. *READ* (Relationship Estimation from Ancient DNA) can infer first degree relatives (parent-child, siblings) and second degree relatives (nephew/niece-uncle/aunt, grandparent-grandchild or half-siblings) based on the proportion of shared alleles. *lcMLkin* (low coverage Maximum Likelihood estimation of kinship) calculates genotype likelihoods from the data in order to infer relatedness between a pair of individuals down to the fifth degree. In addition, *outgroup f3 statistics* were applied as a third independent measure of kinship. *Outgroup f3 statistics* measure the amount of shared genetic drift between two populations from a common ancestor (Patterson et al.,

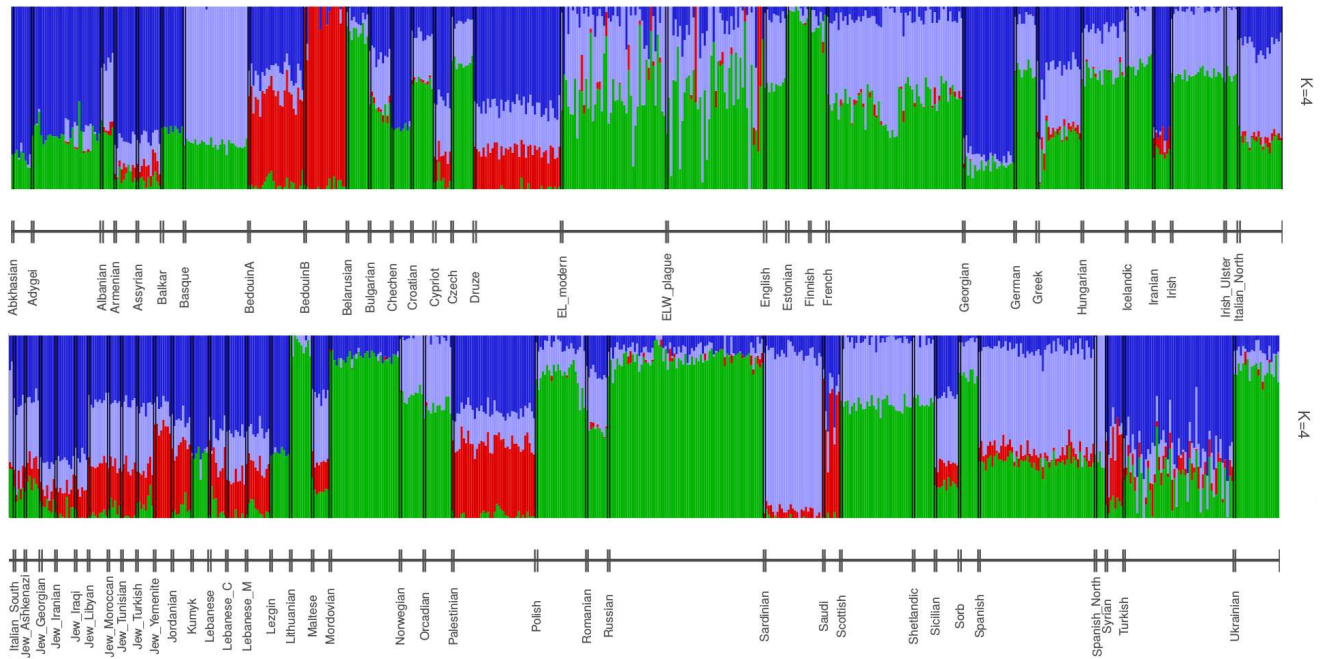


Figure 2.17: Admixture analysis of 69 West Eurasian populations including both Ellwangen populations (EL_modern, ELW_plague) with K=4 ancestral components. The proportion of ancestral genetic components appears to be similar in the modern and the ancient Ellwangen populations indicating genetic continuity.

2012) and can also be applied to infer the degree of relatedness between two individuals. For *READ*, *lcMLkin* and *outgroup f3 statistics* a minimal number of 10.000 SNPs was set as a threshold for any sample to be regarded for kinship analyses, so that the tooth samples JK1510, JK1519 and JK1542 had to be excluded due to insufficient SNP counts. All three programs independently supported a first degree relatedness for the following pairs of individuals: ELW015 and ELW037, ELW016 and ELW017, and ELW036 and ELW039 (fig. 2.18; Supplementary File 2, Part 2). A second degree relatedness was observed by *lcMLkin* and *f3 statistics* between ELW021 and ELW030 and a higher degree relatedness between ELW030 and ELW034. Moreover, a second degree relatedness between ELW007 and ELW039 was supported by *READ* and *outgroup f3 statistics*. Furthermore, second and higher degree relatedness estimates were observed between nine other pairs of individuals. Since the latter kinship estimates were supported by only one program, they were regarded as unreliable. In the modern Ellwangen population only one pair of individuals - EL1 and EL57 were found to be related (first degree; Supplementary File 3, Part 2).

For the subsequent HLA analysis, the individuals EL57, ELW017, ELW030, ELW037

and ELW039 were excluded, because they either had been constituting major "linking nodes" in the genealogy, or contained less SNPs than their counterpart (e.g. ELW015 and ELW037; EL57 and EL1).

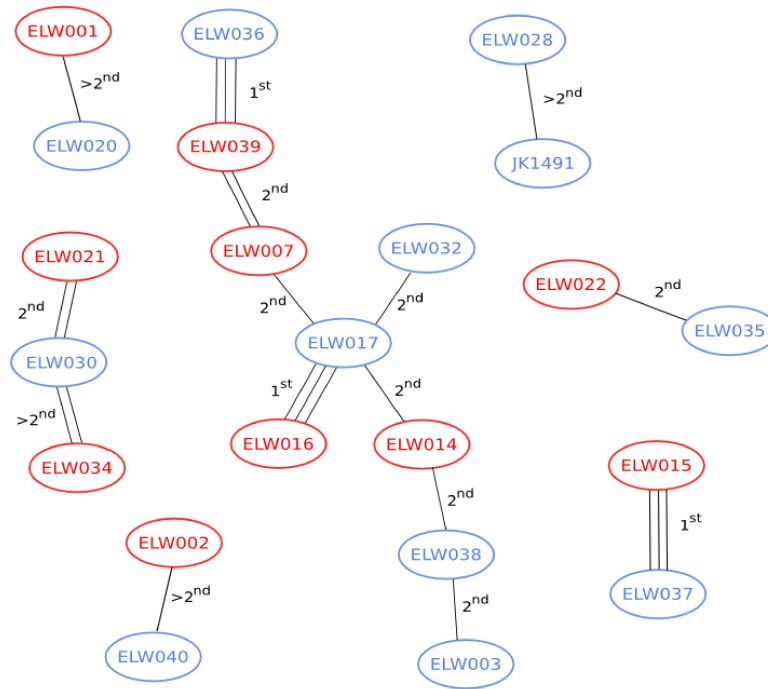


Figure 2.18: Kinship estimates of the Ellwangen plague victims. Kinship was assessed using three different software packages. Male individuals are shown in blue, female individuals in red. The number of lines indicates the number of programs supporting the kinship observations between a pair of individuals. For three pairs a first degree kinship was observed independently by all programs. Only individuals found to be related are shown.

2.4.7 HLA Typing and Allele Frequency Comparisons

After reaffirming continuity between the plague and the modern Ellwangen populations, HLA alleles were called with *OptiType* (Szolek et al., 2014), using a reference set of 11179 present day allele sequences and a sequence identity of at least 97 % for every alignment to be taken into account. The number of possible best matches was set to infinity. The allele calls are presented in Supplementary File 2, Part 3 and Supplementary File 3, Part 3, respectively. A visual inspection was conducted on the coverage of the HLA regions for each sample. The tooth samples JK1491, JK1510, JK1519 and JK1542 showed shallow and fragmentary coverage for all alleles and were therefore excluded from further analysis.

Statistical Power and Effect Size Estimation

Limited sample size is a typical problem of aDNA studies, hence, because after excluding contaminated, directly related and shallow covered samples, a total sample size of 86 (36 ancient and 50 contemporary) individuals was left, statistical power and the size of the observable effects had to be estimated prior to any statistical testing. While statistical power is the probability of correctly choosing the alternative hypothesis, the effect size here refers to the observed shift in allelic frequencies. An effect size of 0.5 can be regarded as a large effect, 0.3 corresponds to a medium effect and 0.1 can be seen as a small effect using Cohen's w as a measure of the effect size (Cohen, 1988). Power and effect size analysis were conducted in *G*Power* (Faul et al., 2009). Since the number of degrees of freedom (df) corresponds to the number of observed different alleles and this number was unknown before the subsequent HLA allele frequency analysis, df was set to 1. With a current sample size of 36 and 50 individuals only large to medium effects ($w = 0.45 - 0.4$) can be identified at an error probability $\alpha=0.05$ and a power of 0.8 (fig. 2.19). The more frequent an allele (e.g. 15 % frequency in a population), the more likely it is to observe a significant frequency shift, while in order to detect significant frequency shifts in rare alleles (e.g. appearing at 1 % frequency), it is recommended to increase the sample size to 500 (personal communication, Jill Hollenbach, UCSF), which was not possible for this aDNA study. Thus, the given sample size allowed to identify effects only in very common but not in rare alleles.

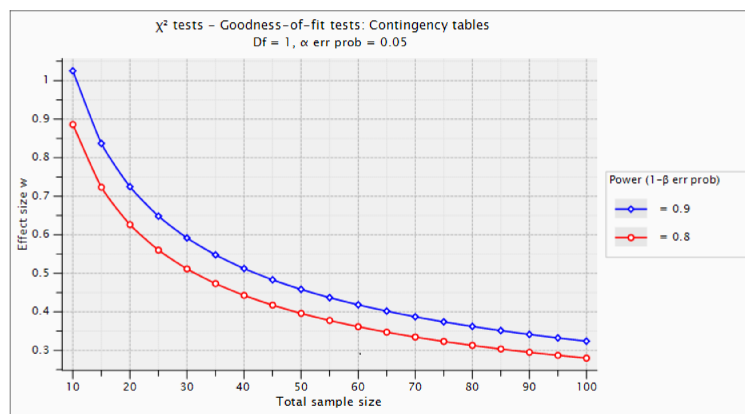


Figure 2.19: Achieved effect size as a function of sample size at a statistical power of 0.9 (blue line) and 0.8 (red line). With the current sample size of 36 and 51 individuals only large effects (effect size (w) = 0.4 - 0.45) are likely to be observed at $\alpha=0.05$ and a power of 0.8 in common (frequent) alleles. The plot was generated using *G*Power* 3.1.9.2.

HLA Allele Frequency Analyses

After exclusion of contaminated samples, directly related individuals and individuals with insufficient HLA coverage, HLA allele frequencies were calculated for both populations and compared using a Chi-squared test, which investigates whether the allele frequency distributions vary significantly between two populations. The Chi-squared test was only performed on alleles that were present in both populations. Assuming a type-I error probability of 5 % for erroneously rejecting a true Null hypothesis that the allele frequency distributions are similar in both populations, no significant difference between the HLA frequency distributions was found (Supplementary fig. S1). For counting the HLA-class II alleles, the HLA-DPA and HLA-DPB, HLA-DQA and HLA-DQB as well as the HLA-DRB1, DRB3, DRB4 and DRB5 alleles, were treated together, respectively. No significant frequency differences were found by the Chi-squared test for the distribution of class II alleles, either.

In addition, frequencies were obtained for the 30 most frequent HLA-A, -B, -C and DRB1 allele occurrences from all over Germany using a sample of 8862 German stem cell donors from DKMS (Deutsche Knochenmarkspenderdatei - German Bone Marrow Donor Register, (Schmidt et al., 2009)). This dataset also contained allele frequencies for 1811 donors from Ellwangen, 581 donors from Heidelberg, 233 donors from Lauchheim, as well as 1918 donors from Lübeck. The Ellwangen HLA allele frequencies were grouped together with the corresponding DKMS allele frequencies (Supplementary fig. S2) and their distributions were cross evaluated using a Chi-squared test as described above. Again, no significant differences could be observed. However, after performing a two-proportion z-test to assess pairwise allele frequency differences between the populations for all common alleles, HLA-B*51:01 showed significant differences in all five cases (tab. 2.3).

2.4.8 CCR5- Δ 32 Mutation in Plague Victims and Contemporary Ellwangen

One of the captured human immunity genes involved into Human Immunodeficiency Virus (HIV) infection is the chemokine receptor CCR5 (Deng et al., 1996). A known 32 bp deletion disrupts the coding region of CCR5 causing a loss-of-function receptor that does not allow HIV to enter cells such as macrophages or monocytes (Dean et al., 1996). Δ 32-homozygous individuals were shown to resist HIV-1 despite repeated exposure, and Δ 32-heterozygotes could slow down the infection for 2-3 years until the onset

Table 2.3: Significant differences in pairwise HLA-allele frequencies between the ancient Ellwangen population compared to each of the DKMS populations. A two-proportion z-test was assessed under the null hypothesis that a pair of allele frequencies does not differ significantly. Only alleles are shown where significant frequency differences ($p \leq 0.05$, after Benjamini-Hochberg correction) between at least one pair of populations were observed. plague = Ellwangen plague population, GER = Germany DKMS, ELW = Ellwangen DKMS, HEI = Heidelberg DKMS, LAU = Lauchheim DKMS, LUE = Lübeck DKMS.

	plague vs GER	plague vs ELW	plague vs HEI	plague vs LAU	plague vs LUE
A*23:01				0.0289	
A*24:02		0.0051	0.0141	0.0077	
B*50:01		0.0288			0.0410
B*51:01	0.0054	8.47E-007	0.0007	0.0002	0.0002

of the acquired immunodeficiency syndrome (AIDS) compared to individuals lacking the mutation (Dean et al., 1996). A CCR5- Δ 32 north-to-south gradient was observed with the highest CCR5- Δ 32 frequencies in the Swedish followed by the Russians and the lowest frequencies, resp. complete absence in Near Easterners and Chinese (Stephens et al., 1998). The origin of the CCR5- Δ 32 mutation has been estimated to around 700 years ago (ya) making the occurrence of the mutation a recent event (Reich and Goldstein, 1998). The recent origin of the mutation and moreover, a high level of non-synonymous mutations in the CCR5 gene as well as its prevalence in high frequencies in Caucasian populations suggest that the CCR5- Δ 32 was caused by strong selective pressure potentially exerted through a pathogen in recent historic times. Apart from viruses such as smallpox and influenza, bacterial agents such as *Mycobacterium tuberculosis*, *Shigella*, *Treponema pallidum* and *Salmonella* have been proposed as causative agents (Stephens et al., 1998). More prominent, the plague agent *Yersinia pestis* that caused the Black Death in the 14th century, which approximately coincides with the estimated origin and rise in frequencies of the CCR5- Δ 32 allele, is debated as the driving agent for the onset of the mutation (Lenski, 1988).

To investigate whether *Y. pestis* exerted selective pressure on the CCR5 gene, e.g. by infecting host macrophages through their CCR5 receptors, the frequencies of the CCR5- Δ 32 alleles between the Ellwangen plague victims and the contemporary Ellwangen individuals were compared. The sequencing data had been remapped to the human reference sequence hg19 using a gap-sensitive split-read mapper (BWA-mem) and the CCR5- Δ 32 region (chr3:46414947-46414978) was visually inspected in IGV for every

sample (fig. 2.20). No filtering for mapping quality was applied. In addition, the number of reads with positions absent in comparison to the reference (deletions and soft-clipped positions) was obtained. Two samples showed a clear gap and six samples exposed a partial coverage, whereas the remaining samples were completely covered at the CCR5- Δ 32 region. Moreover, these eight samples had above 20 % of mapped reads in the CCR5- Δ 32 region containing either deletions or having been soft-clipped, whereas the rest of the samples had below 10 % of soft-clipped reads and reads with a deletion. Based on these observations two ancient Ellwangen individuals (4.8 %) were found to be homozygous for the CCR5- Δ 32 mutation, whereas six (23.8 %) were found to be heterozygous. Using the same criteria, 11 modern day Ellwangen individuals (21.6 %) were found to be heterozygous for the mutation (tab. 2.4; Supplementary File 2, Part 3 and Supplementary File 3, Part 3). In contrast to the ancient Ellwangen samples, no complete lack of coverage could be detected in the CCR5- Δ 32 region for any of the modern day Ellwangen inhabitants. The CCR5- Δ 32 allele appears at a frequency of 16.6 % among the plague victims and is present at 10.8 % in the contemporary population (tab. 2.4).

Table 2.4: Genotype frequencies (GF) of CCR5-wildtype homozygous (wt/wt), heterozygous (wt/ Δ 32) and CCR5- Δ 32 homozygous individuals (Δ 32/ Δ 32), and allele frequencies (AF) among the Ellwangen plague victims (Ellwangen plague) and contemporary Ellwangeners (Ellwangen modern).

GF	wt/wt	wt/ Δ 32	Δ 32/ Δ 32	AF	wt	Δ 32
Ellwangen plague	71.4	23.8	4.8	Ellwangen plague	83.4	16.6
Ellwangen modern	78.4	21.6	0	Ellwangen modern	89.2	10.8

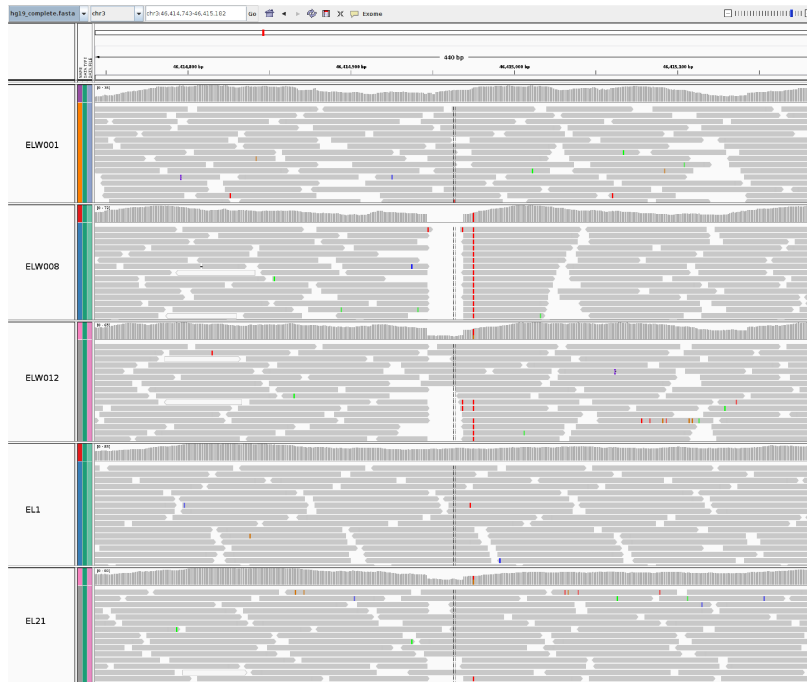


Figure 2.20: Coverage of the CCR5- Δ 32 region in 3 example Ellwangen plague victims (ELW001, ELW008, ELW012) and 2 contemporary Ellwangen individuals (EL1, EL21). ELW001 and EL1 (wt/wt) show no deletion. ELW008 (Δ 32/ Δ 32) shows a consistent gap indicating the presence of the CCR5- Δ 32 deletion. ELW012 and EL21 (het) show a clear drop in the coverage consistent with partial coverage due to heterozygosity. Visualization was performed in IGV (Robinson et al., 2011).

3 Discussion

Although sequencing whole (ancient) genomes has become more affordable compared to a decade ago, DNA capture allows to circumvent high sequencing costs, that would inevitably accumulate if ancient DNA was genome-wide sequenced to an extent that provides the required high coverage of particular genetic regions of interest. Here I have presented three studies involving successful aDNA capture (Immel et al., 2015, Immel et al., 2016, Immel et al., 2020).

3.1 Giant Deer and Megafauna

A decade ago target amplification through PCR with multiple primers (multiplex amplification) allowed to reconstruct and study mitochondrial genomes of several extinct megafaunal species such as the woolly mammoth (Krause et al., 2006, Rogaev et al., 2006), mastodon (Rohland et al., 2007) and cave bear (Krause et al., 2008). With the appearance of capture through hybridization techniques it became possible to reconstruct the mitochondrial genomes of a 300 kyo cave bear from Spain (Dabney et al., 2013), 67 Late Pleistocene mammoth specimens (Enk et al., 2016) and 13 ancient bison and an ancient wisent mitochondrial genomes (Soubrier et al., 2016). In my first study I was able to reconstruct two almost complete mitochondrial genome sequences of the extinct giant deer *M. giganteus* after capture of its mtDNA obtained from megafaunal bone remains that were excavated in the Swabian Alb. This not only added another extinct animal to the list of published mitochondrial genomes: with two complete mt genomes the resolution was high enough to resolve the phylogenetic position of the giant deer, confirming its previously suggested closer relatedness to the fallow deer *Dama dama* (Freudenberg, 1914, Lister, 1984, Lister et al., 2005), in contrast to a previously assumed closer affinity to the red deer *Cervus elaphus*, which had been proposed based on post-cranial skeletal morphology (Pfeiffer, 1999, Pfeiffer, 2002). The authenticity of the sequences originating from *Megaloceros giganteus* and its phylogenetic placement as a sister clade to the fallow deer were reaffirmed by using previously published *M. gigan-*

teus cytb sequences together with our new reconstructed cytb sequences and published cytb sequences of other cervids such as red deer, fallow deer and roe deer. Three different methods of phylogenetic reconstruction - Maximum Likelihood, Maximum Parsimony and Bayesian Inference independently confirmed the placement of all giant deer cytb sequences together as one clade next to the fallow deer clade comprising *Dama dama* and *Dama mesopotomica*.

Constructing phylogenies based on whole genome data might provide a different point of view into the evolutionary scenario of the giant deer. Experience has shown that mitochondrial and nuclear DNA lineages do not necessarily tell the same evolutionary story: ntDNA showed that the European Bison or wisent (*Bison bonasus*) is closer related to the American bison (*Bison bison*), while its mtDNA indicated a closer genetic affinity to modern cattle (Soubrier et al., 2016). In case of the Middle Pleistocene hominins from Sima de los Huesos (Spain) the mtDNA suggested a closer genetic affinity to the Denisovan individual (Meyer et al., 2014) while on ntDNA level they were closer related to the Neandertals (Meyer et al., 2016).

Moreover, morphological classification is not unequivocal and therefore many faunal remains stored in museums might have been misinterpreted - such as giant deer bones erroneously being ascribed to elk. Thus, genetic investigation of dubious faunal remains may shed new light on the previous assumptions of megafaunal dispersals and evolution.

3.2 X-rays and aDNA

The preservation of morphology plays a key role in palaeontological, archaeozoological and anthropological work. 3D-imaging techniques such as computed tomography (CT) are therefore routinely applied to precious remains prior to destructive sampling for aDNA, stable isotope analyses, or radiocarbon dating. However, it has been highly questionable that CT does not affect aDNA, since X-rays can have a detrimental effect on modern DNA.

In my second study I investigated the effect of X-rays on ancient mtDNA previously extracted and captured from Late Pleistocene megafaunal specimens, such as cave bear, bison, giant deer as well as roe deer. While previous studies used either bones from recently butchered pigs (Goetherstrom et al., 1995, Grieshaber et al., 2008) or dried bird skins (Paredes et al., 2012), our study was the first to use aDNA extracted from ancient megafaunal bone material in order to assess the impact of X-rays. Combining the results of our three experiments, we could define 200 Gy as the upper limit absorbed

radiation dose that can be reached without any detectable effects on aDNA molecules. This dose exceeds the dosage of a medical CT scan of e.g. abdomen and pelvis (20 mGy) by factor 10.000. Although this dose can be reached by a classical CT scan especially when working with voxel sizes below 5 μm , current CT settings make use of metallic filters that filter out low energy X-rays which would be absorbed by the sample and cause ions accumulating subsurface, detrimental for aDNA molecules. Moreover, the hydration and the mineralization state of the sample might play a key role in the outgoing of a CT scan (Schwarz et al., 2009, Campos et al., 2012). Therefore it is likely, that through mineralization endogenous ancient DNA can be better protected from penetrating X-ray radiation and secondary chemical reactions, e.g. caused by water radicals.

A later study assessed the effect of X-rays on shotgun sequenced aDNA from prehistoric and historic humans and did not find any significant correlation between ionizing radiation and the recovery and amplification of DNA (Fehren-Schmitz et al., 2016). Interestingly, Fehren-Schmitz *et al.* do not define the exact absorbed radiation dose, but refer to "T1 = clinical/low irradiation dosage" and "T2 = high irradiation dosage" which corresponds to 20 times T1. Supposed that T1 corresponds to the average absorbed X-ray dose after an abdomen or pelvis scan (20 mGy), T2 would correspond to 400 mGy, which is 500 times lower than our detection limit for any X-ray induced effects on aDNA. It is therefore not surprising, that Fehren-Schmitz *et al.* do not observe any significant effects.

3.3 In-Solution Immunity Capture

Tracing back the evolutionary history of host-pathogen interactions has become possible through the combination of target enrichment and HTS. In 2015, using hybridization capture, 394 577 (390K) target SNPs could be enriched from 69 ancient Eurasian individuals who lived 8 kya to 3 kya, while reducing the amount of sequencing required to obtain genome-wide coverage by at least 45 times (Haak et al., 2015). In the same year a larger set including 1 233 013 (1240K) target SNPs for capture was applied to 230 ancient West Eurasians, who lived 6.5 kya to 300 BCE, using the same in-solution target enrichment strategy in combination with petrous bones (Mathieson et al., 2015). So far, however, enrichment of immunity-related genes directly involved into host-pathogen interaction has not been conducted in aDNA research. Target enrichment in combination with DNA extraction from petrous bones has allowed me to conduct an immunogenetic

investigation of plague victims from the 16th century and their modern descendants on particular genes of interest, while reducing the amount of sequencing by 370 times on average.

In comparison to the 1240K capture and both versions of the *Illumina ImmunoChip* the novel in-solution immunity capture described here allows to simultaneously capture whole alleles and call diagnostic markers for almost 500 immunity-relevant genes. While the *Illumina HumanImmuno v1.0 BeadChip* contains 196 806 markers whereof the majority are intergenic and the *Illumina ImmunoArray24 v2.0 BeadChip* contains 253 702 markers with a strong focus on autoimmune disorders, alleles required to predict haplotypes for important gene families such as HLA and KIR need to be imputed from SNP data (Cortes and Brown, 2011). Our in-solution immunity capture allows to reconstruct complete allelic haplotypes of exonic and intronic regions without imputations and therefore has been used to successfully reconstruct HLA alleles and to detect the CCR5- Δ 32 deletion using ancient as well as modern DNA.

3.4 Stability of the CCR5- Δ 32 Mutation

In contrast to the assumption that bubonic plague has elevated CCR5- Δ 32 frequencies in European populations (Stephens et al., 1998) the allele frequency in Ellwangen seems to have decreased since the plague outbreak in the 16th century. Evidence exists, that CCR5- Δ 32 allele frequencies have been stable in Central Europe since the Middle Ages (Bouwman et al., 2017). The CCR5- Δ 32 allele frequency of 16.6 % in the Ellwangen plague victims corresponds to a frequency of 14.2 % in Black Death victims from Lübeck, previously discovered using PCR (Hummel et al., 2005), despite that here the only evidence of plague refers to a "well-documented medieval plague mass grave" since no *Y. pestis* DNA could be detected there. Moreover, 16.6 % is in accordance with what is expected in modern northern European populations (Galvani and Novembre, 2005). So far the given data does not allow to infer selective pressure acting on CCR5 through *Y. pestis*. Evidence exists that not the *Y. pestis*-caused Black Death could have caused sufficient selection pressure to drive the CCR5- Δ 32 allele to a current frequency of 10 % in Europeans, but the variola virus causing smallpox (Galvani and Novembre, 2005). Smallpox and HIV both are viruses that infect lymphocytes by use of Chemokine receptors (Lalani et al., 1999). Indeed, a relatively strong selection coefficient ($s = 0.17$) was found to be consistent with selection on CCR5- Δ 32 exerted through smallpox (Galvani and Novembre, 2005).

3.5 Frequency Shifts in particular HLA alleles

Although the all-over distribution of alleles between the plague victims and the modern day Ellwangen population did not differ significantly, a significant frequency shift could be detected in the HLA-B*51:01 allele. While HLA-B*51:01 is associated with papillary thyroid carcinoma (Shuxian et al., 2014) and, more generally, HLA-B*51 is associated with Behçets Disease (Wallace, 2014), it is not associated with plague susceptibility and not known to be in linkage with any HLA-class II allele involved into the presentation of peptides of gram-negative bacteria such as *Y. pestis*. However, HLA-B*51:01 provides anti-viral immunity (Kawashima et al., 2010) and co-evolves with the natural killer (NK) cell receptor KIR3DL1 (Norman et al., 2013). A possible tendency to mount innate immune responses against virus infections via cell-destruction through NK cells may have released intracellular *Y. pestis* into the bloodstream causing lethal septicemic plague. This could have caused strong negative selection on carriers of the HLA-B*51:01 allele. HLA-DQB1*06:03, HLA-DRB1*11:04, DRB1*13:01, DRB1*13:02 and DRB1*16:01 were found to be at least twice as frequent in the contemporary population compared to the plague victims. Among them, DRB1*13:01 and DRB1*13:02 differ in only one base from each other, thus summarizing them into one field (two digit) resolution yields a frequency shift from 5.8% in the ancient population to 16% in the modern one, which might suggest a protective feature associated with this allele family. Indeed, HLA-DRB1*13 has been shown in association with resistance to *Mycobacterium tuberculosis* in a Polish population (Dubaniewicz et al., 2000). *M. tuberculosis*, similar to *Y. pestis*, can invade and survive within macrophages (Pieters, 2008). It is therefore likely that the same signalling pathway triggered by HLA-DRB1*13, that provides resistance to *M. tuberculosis*, may also provide resistance to *Y. pestis*. In general, however, no evidence of strong selection through *Y. pestis* on human immunity genes could be observed. The sample size of 36 ancient and 50 modern samples might be too small to detect significant shifts in frequencies of less common HLA alleles, which was already indicated by the effect size calculations suggesting only medium to large detectable effects with our given sample size. Nevertheless, Lindo *et al.* could detect significant allele and SNP frequency changes in the HLA-DQA1 gene based on only 25 ancient and 25 modern exomes from ancient and modern Native Americans, that they attributed to negative selection likely caused by smallpox (Lindo et al., 2016).

The alternative scenario is that plague caused by *Y. pestis* in the 16th century was not novel and therefore not strong enough to leave any signals of selection on HLA genes of

the Ellwangen plague victims. Plague has been shown to prevail in human populations since the Late Neolithic (Rascovan et al., 2019, Andrades Valtuena et al., 2017) and probably came into Europe during the expansion of Yamnaya steppe pastoralists (Andrades Valtuena et al., 2017). Therefore, the plague outbreak in Ellwangen may have not been strong enough to leave any signs of selection on this European sub-population, whose immunogenetic makeup may have already been shaped by pathogens such as *Y. pestis* even long before the Black Death in the 14th century.

Y. pestis, such as any biological causative agent of disease, becomes especially efficient if the immune system is weakened. The pandemics caused by plague, such as the Justinian Plague or the infamous Black Death that supposedly eradicated two thirds of medieval Europe, were predated by other catastrophes, such as wars or famines: tree ring analysis has shown an extended period of cold between 536 and 545 AD followed by drought that could have caused crop failures and famine and also could have driven rats inside homes (Keys, 1999), before the outbreak of the Justinian Plague in the 6th century AD. Moreover, a multiyear period of severe rainfalls made food production nearly impossible and led to the Great Famine of 1315-1317, as well as subsequent famines, such as of 1330-1334 (Ruiz, 2001, Cook et al., 2015). Malnutrition then weakened the European population and promoted its susceptibility to infectious diseases (Ruiz, 2001). Thus, it is possible that also the late medieval Ellwangen population had a weakened immune system facilitating the plague outbreak, which had then predominantly affected children (ca. 60 % of all individuals) and affected juveniles and adults to the same extent (ca. 20 %) (Wahl, 2014).

4 Conclusion

In this work I have demonstrated the application of different capture methods to ancient DNA in combination with Next-Generation-Sequencing and bioinformatic analyses. Beginning with the mitochondrial DNA capture applied to aDNA extracted from giant deer remains excavated in Southern Germany, it was possible to reconstruct the giant deer's complete mitochondrial genome and resolve its questionable phylogenetic position within the deer clade. Moreover, the results indicate that the giant deer recolonized parts of Central Europe after the Last Glacial Maximum.

Mitochondrial DNA capture was applied to aDNA extracted from Pleistocene Megafaunal skeletal material that was exposed to different Computed Tomography scanning experiments. Since previous studies provided contradictory results, it was questionable whether CT is really a non-invasive method that does not affect the contained ancient DNA. Our combined results allowed for the first time to investigate the effects of X-rays on aDNA extracted from Pleistocene bones, and to define an absorbed-radiation-limit dose of 200 Gy, up to which no destructive effects on aDNA could be observed. However, current CT methods rarely reach this dose. Nevertheless, simple precautionary measures should be followed thus CT scans can be performed without affecting the integrity of aDNA in sub-fossilized archaeological or palaeontological findings.

The main focus of my dissertation was the development of a specific immunity capture that allows to track selection on large immunity-relevant candidate gene families, such as the Human Leukocyte Antigen, in ancient and modern populations. Using this novel in-solution immunity capture it was possible to capture complete HLA alleles and the CCR5- Δ 32 allele from aDNA of 16th century plague victims and a contemporary reference population. Applying bioinformatic analyses to the captured DNA data revealed potential signals of selection through plague on particular HLA genes, even though the general evidence of selection was rather marginal.

Supplementary Material and Appendix

Supplementary File 1

488 Genes involved in the Design of the Immunity Capture

ADPRHL2	CD180	CD86	HLA-DRA	IL22RA1	MYO5C	PSMC2	TNF
AIM2	CD19	CD8A	HLA-DRB	IL22RA2	NCAPD2	PSMC3	TNFAIP1
AIP	CD1B	CD9	HLA-E	IL23A	NFKB1	PSMC3IP	TNFAIP2
APBA2	CD1D	CD93	HLA-F	IL23R	NFKB2	PSMC4	TNFAIP3
APMAP	CD1E	CD96	HLA-G	IL24	NFKBIA	PSMC5	TNFAIP6
APOH	CD2	CD97	HLA-H	IL25	NFKBIB	PSMC6	TNFAIP8
ARID3B	CD200	CDKN2B	HLA-J	IL26	NFKBID	PSMD1	TNFAIP8L1
ARL5A	CD200R1	CFH	HLA-K	IL27RA	NFKBIE	PSMD10	TNFAIP8L2
ARPC1A	CD200R1L	CFHR5	HLA-L	IL28RA	NFKBIL1	PSMD10P	TNFAIP8L3
ARPC1B	CD207	CFI	HLA-V	IL2RA	NFKBIL2	PSMD11	TNFRSF10A
ATP5J2	CD209	CLEC10A	HSPD1	IL2RB	NFKBIZ	PSMD12	TNFRSF10B
ATP5J2-PTCD1	CD22	CLEC12A	IFI27	IL2RG	NLRB1	PSMD13	TNFRSF10C
BMP3	CD226	CLEC12B	IFIH1	IL3	NLRC3	PSMD14	TNFRSF11A
BPI	CD244	CLEC1B	IFIT1	IL31	NLRC4	PSMD2	TNFRSF11B
BUD31	CD247	CLEC2A	IFIT2	IL31RA	NLRC5	PSMD3	TNFRSF12A
C2	CD248	CLEC4A	IFIT3	IL33	NLRP1	PSMD4	TNFRSF13B
C3	CD27	CLEC4C	IFIT5	IL34	NLRP10	PSMD5	TNFRSF13C
CARD15	CD274	CLEC4D	IFNA8	IL4	NLRP11	PSMD6	TNFRSF14
CASP1	CD276	CLEC4E	IFNAR1	IL4I1	NLRP12	PSMD7	TNFRSF17
CCL1	CD28	CLEC5A	IFNAR2	IL4R	NLRP13	PSMD9	TNFRSF18
CCL11	CD2AP	CLEC6A	IFNB1	IL5	NLRP14	PSME1	TNFRSF19
CCL14	CD2BP2	CLEC7A	IFNE	IL5RA	NLRP2	PSME2	TNFRSF1B
CCL15	CD300A	CLEC9A	IFNG	IL6	NLRP3	PSME3	TNFRSF21
CCL16	CD300C	COL3A	IFNGR1	IL6R	NLRP4	PSME4	TNFRSF25
CCL17	CD300E	COL8A2	IFNGR2	IL6ST	NLRP5	PSMF1	TNFRSF4
CCL18	CD302	COQ4	IFNK	IL7	NLRP6	PSMG1	TNFRSF6B
CCL19	CD320	CPSF3L	IL10	IL7R	NLRP7	PSMG4	TNFRSF8
CCL2	CD33	CPSF4	IL10RA	IL8	NLRP8	PTCD1	TNFRSF9
CCL20	CD34	CR1	IL10RB	IL8RA	NLRP9	PTX3	TNFSF10
CCL21	CD36	CTLA4	IL11RA	IL8RB	NLRX1	PYCARD	TNFSF13
CCL22	CD37	CUX2	IL12A	IL9	NOD1	RIPK1	TNFSF13B
CCL23	CD38	CXCR3	IL12B	IRAK1	NOD2	RIPK2	TNFSF14
CCL24	CD3D	CXCR4	IL12RB1	IRAK1BP1	ODF2L	RSPO4	TNFSF15
CCL25	CD3E	CXCR5	IL12RB2	IRAK2	OXSM	SARM1	TNFSF18
CCL27	CD3G	CXCR6	IL13	IRAK3	PARD3B	SCNN1D	TNFSF4
CCL28	CD4	CXCR7	IL13RA1	IRAK4	PARK2/PACRG	SENP8	TNFSF8
CCL5	CD40	DC-SIGN/CD209	IL13RA2	KIR2DL1	PCMTD1	SLC11A1	TNFRSF1A
CCL7	CD44	DCTN1	IL15	KIR2DL4	PDAP1	SLC12A1	TOLLIP
CCL8	CD46	DDX58	IL15RA	KIR2DS4	PGL1	SLC24A5	TRAF1
CCNL2	CD47	DEFB1	IL16	KIR3DL2	PI3K	SLC27A4	TRAF2
CCR1	CD48	DUT	IL17A	KIR3DL3	PPARD	STK17B	TRAF3
CCR10	CD5	DVL1	IL17B	L1CAM	PPP2R5C	SVIL	TRAF3IP1
CCR3	CD52	EDNRA	IL17C	LAMA2	PSMA1	TAP1	TRAF3IP2
CCR4	CD53	EEF2K	IL17D	LBP	PSMA2	TAP2	TRAF3IP3
CCR5	CD55	ERAP2	IL17F	LILRA5	PSMA3	TEKT2	TRAF4
CCR6	CD59	FAM200A	IL17RA	LILRA6	PSMA4	TGFB1	TRAF5
CCR7	CD5L	FCN2	IL17RB	LILRB4	PSMA5	TICAM1	TRAF6
CCR8	CD6	GOLGA2	IL18	LILRB5	PSMA6	TICAM2	TRAF7
CCR9	CD63	GS1-259H13_2	IL18R1	LINS	PSMA7	TIRAP	TRAFD1
CCRK	CD68	HIF1A	IL19	LRP3	PSMA8	TLR1	TRAPPC3
CCRL1	CD69	HLA-A	IL1A	LTA	PSMB1	TLR10	TRUB2
CCRL2	CD7	HLA-B	IL1B	LTA4H	PSMB10	TLR2	UNC93B1
CCRN4L	CD70	HLA-C	IL1R1	MAVS	PSMB2	TLR3	URM1
CD109	CD72	HLA-DMA	IL1R2	MBL2	PSMB3	TLR4	VDR
CD14	CD74	HLA-DMB	IL2	MCP1	PSMB4	TLR5	VWA3A
CD151	CD79B	HLA-DOA	IL20	MD2	PSMB5	TLR6	WDR88
CD160	CD80	HLA-DOB	IL20RA	MIC-A	PSMB6	TLR7	ZBP1
CD163	CD81	HLA-DPA	IL20RB	MIC-B	PSMB7	TLR8	ZKSCAN5
CD163L1	CD82	HLA-DPB	IL21	MRC2	PSMB8	TLR9	ZNF655
CD164	CD83	HLA-DQA	IL21R	MXRA8	PSMB9	TMEM173	ZNF789
CD164L2	CD84	HLA-DQB	IL22	MYD88	PSMC1	TMEM30B	ZSCAN25

Supplementary File 2

Ellwangen Plague Victims (Part 1)

Sample	Raw Reads (CP)	Mapped Reads After rmdup (CP)	Endogenous DNA (%) (CP)	CF (CP)	Mean Cov (SG)	Avg Read Length (SG)	MT/NUC (SG)	DMG 1 st Base 5' (SG)	Sex (SG)	MT-Cont (SG rCDS mapped)	X Cont (SG)	Reads in Target (CP)	% Reads in Target (CP)	Reads in Target After rmdup (CP)	% Reads in Target After rmdup (CP)	Per Region Avg Readcov (CP)
ELW001	10466553	1860240	35.411	1.4	0.0084	65.87	132.41	0.0789	F	NA	NA	878908	8.4	393924	3.8	44.92
ELW002	10359441	3185057	61.025	1.3	0.0302	59.41	126.36	0.0774	F	0.02	NA	1148670	11.1	571413	5.5	54.73
ELW003	10036348	3193155	57.488	1.2	0.0263	63.75	143.06	0.0789	M	NA	NA	1070611	10.7	669259	6.7	52.49
ELW004	9161046	4041857	76.925	1.2	0.0626	64.66	90.59	0.0531	F	0.01	NA	1154544	12.6	739766	8.1	56.6
ELW005	9133695	4331467	81.576	1.2	0.1164	64.42	141.2	0.0537	F	0.01	NA	1262988	13.8	801929	8.8	62.46
ELW006	10798124	2933986	51.843	1.2	0.0216	63	106.82	0.0949	F	NA	NA	1057764	9.8	637681	5.9	51.85
ELW007	10420126	2384575	43.466	1.2	0.0107	62.9	137.13	0.0773	F	0.01	NA	907796	8.7	527880	5.1	43.81
ELW008	9799328	4343226	76.607	1.2	0.0728	65.59	109.04	0.0496	M	NA	NA	1243659	12.7	785290	8.0	61.26
ELW009	8479678	2888121	69.832	1.2	0.0654	64.54	105.53	0.0666	F	0.01	NA	1029471	13.8	668723	8.9	52.4
ELW010	8508824	4273374	86.563	1.2	0.0831	64.5	89.62	0.0614	F	0.01	NA	1145330	13.5	756735	8.9	57.08
ELW011	8364303	3412822	69.585	1.2	0.0539	66.25	105.61	0.0566	F	NA	NA	1033137	12.4	647354	7.7	52.75
ELW012	10198477	4155526	71.612	1.2	0.0614	65.65	104.24	0.0674	M	0.01	0.029227	1308928	12.8	796916	12.8	65.9
ELW013	8423826	2849734	60.408	1.2	0.0311	64.35	113.32	0.0717	M	NA	NA	964249	11.4	586007	7.0	47.75
ELW014	9545804	976007	18.429	1.2	0.0032	67.01	101.08	0.0721	F	NA	NA	377948	4.0	214756	2.2	19.46
ELW015	8486436	3635055	75.126	1.2	0.0482	63.17	102.35	0.0659	F	0.01	NA	1092049	12.9	671270	7.9	52.78
ELW016	9573389	3595998	64.59	1.2	0.0615	66.18	154.62	0.0683	F	0.01	NA	1140841	11.9	661537	6.9	57.98
ELW017	1390967	385959	46.053	1.1	0.0128	65.09	121.39	0.0639	M	NA	NA	148550	10.7	129170	9.3	7.72
ELW018	8028617	3444876	72.751	1.2	0.053	66.44	106.6	0.0447	M	NA	NA	1054092	13.1	684728	8.5	54.24
ELW019	8669279	3632411	73.033	1.2	0.0527	64.67	83.06	0.0613	F	0.01	NA	1057432	12.2	689035	7.9	53.25
ELW020	8224392	3736956	74.924	1.2	0.0529	66.92	141.96	0.0527	M	NA	0.002003	1041762	12.7	674453	8.2	54.51
ELW021	8734310	3794052	73.651	1.2	0.0444	66.1	111.34	0.0583	F	NA	NA	674488	12.0	674488	7.7	52.96
ELW022	9508207	1109891	20.953	1.2	0.004	64.83	112.67	0.0611	F	NA	NA	497156	5.2	320121	3.4	25.11
ELW026	9331311	3151110	60.393	1.2	0.0054	64.31	87.08	0.064	M	0.01	NA	1128750	12.1	722900	7.7	56.87
ELW027	8976014	2511458	49.549	1.2	0.0265	65.86	101.63	0.0572	M	0.03	NA	834012	9.3	559568	6.2	42.67
ELW028	9084000	3787149	73.303	1.2	0.0638	64.82	137.51	0.0517	M	0.01	0.004904	1126249	12.4	716715	7.9	56.18
ELW029	9227039	2857767	58.744	1.3	0.0405	63.85	132.89	0.0677	M	0.01	NA	1058350	11.5	540062	5.9	32.97
ELW030	8451222	1732774	38.671	1.2	0.0064	61.18	94.9	0.0875	M	NA	NA	681247	8.1	415048	4.9	33.76
ELW031	9092591	1872198	37.131	1.2	0.0109	63.4	71.22	0.0792	M	NA	NA	680945	5.2	474558	5.2	33.76
ELW032	9027378	2157420	44.671	1.2	0.0077	61.89	91.26	0.0871	M	NA	NA	791335	8.8	478282	5.3	38.52
ELW033	8029890	2956927	68.352	1.3	0.0463	65.85	87.36	0.0616	M	0.01	NA	1132729	14.1	648933	8.1	58.21
ELW034	9637130	3455168	66.459	1.2	0.0327	64.64	84.13	0.0599	F	NA	NA	1224791	12.7	703983	7.3	62.09
ELW035	10926629	2997186	53.245	1.2	0.0284	63.44	72.31	0.0886	M	NA	NA	1170271	10.7	671707	6.1	58.14
ELW036	9286357	3370113	70.704	1.3	0.0463	64	98.5	0.0576	M	0.01	NA	1108583	11.9	620672	6.7	56.91
ELW037	10488558	3776394	70.1	1.2	0.0471	65.5	119.33	0.0474	M	0.01	0.003447	1126947	10.7	662809	6.3	58.76
ELW038	8548352	1846634	43.972	1.2	0.0071	60.73	117.29	0.1156	M	NA	NA	695266	8.1	405663	4.7	33.48
ELW039	10260944	2274561	42.44	1.3	0.0094	62.83	121.45	0.0836	F	NA	NA	969823	9.5	553788	5.4	48.23
ELW040	11341241	2748546	53.9	1.5	0.0171	64.48	106.72	0.067	M	0.01	NA	945105	8.3	613114	5.4	48.23
JK1491	5131824	543918	56.418	4.0	0.0082	61.31	1161.9	0.3036	M	0.02	NA	1495752	29.1	80381	1.6	75.53
JK1510	6121683	177538	39.043	10.8	0.0017	58.58	2468.81	0.2041	M	0.02	NA	1618311	26.4	26583	0.4	123.22
JK1519	5954784	328134	44.27	6.3	0.0034	58.59	1115.98	0.1892	M	0.01	0.322985	1673438	28.1	82202	1.4	65.83
JK1542	6590156	283139	33.964	6.4	0.0013	51.05	1896.01	0.1896	F	0.02	NA	1274460	19.3	75320	1.1	87.14
EXB011.A0101	77524	1698	42.904	12.0	0.0003	65.14	138.17	0.0864	NA	NA	NA	4453	5.7	40	0.1	112.05
EXB011.A0102	17539	494	30.384	6.0	0.0001	63.21	102.49	0.0635	NA	NA	NA	180	1.0	20	0.1	8.77
LIB001.A0106	26259	188	49.484	62.7	0.0001	64.73	132.95	0.0985	NA	NA	NA	10270	39.1	24	0.1	395.44
LIB001.A0107	15940	346	34.059	10.0	0.0001	64.1	66.36	0.0696	NA	NA	NA	13	0.1	10	0.1	1.29
LIB001.A0108	21331	962	23.224	3.6	0.0001	64.25	78.81	0.1187	NA	NA	NA	1141	5.3	125	0.6	4.34
LIB001.A0109	15662	310	39.625	13.4	0.0001	63.82	25.67	0.0846	NA	NA	NA	2468	15.8	26	0.2	102.08

Compiled results. **Raw Reads:** No. of unprocessed reads after sequencing. **CP:** based on capture data. **SG:** based on shotgun data. **Mapped Reads after rmdup:** No. of unique reads after duplicate removal, mapping to hg19 (mapped with BWA 0.7.2 with n=0.01 and MQ=30 in EAGER). **Endogenous DNA %:** percentage of raw reads mapping to hg19. **CF:** Cluster Factor (comp. Table 3.1). **Mean Cov (SG):** Average coverage of hg19 based on shotgun data. **MT/NUC:** Mitochondrial to nuclear sequence ratio. **DMG 1st base 5' (SG):** frequency of C→T substitutions at the 1st position from 5' end (*mapDamage2.0*). **Mt Cont:** contamination estimation based on mtDNA using *Schmutzi*. **X Cont:** contamination estimation in males based on X-chromosome using *ANGSD*. **Reads in Target:** No. of reads mapping to the targeted regions of interest (*samtools*). **Per Region Avg Readcov:** average coverage of target regions. **ELW*:** petrous bone samples. **JK*:** tooth samples. **EXB*:** extraction blanks. **LIB*:** library blanks. **NA:** not available/missing data.

Supplementary File 2

Ellwangen Plague Victims (Part 2)

Sample	Per Region Avg Readcov after rmdup (CP)	% target region covered>=1x After rmdup (CP)	% target region covered>=1x After rmdup (CP)	% target region covered>=5x After rmdup (CP)	% target region covered>=12x After rmdup (CP)	% target region covered>=20x After rmdup (CP)	SNPs called (from 1240k, SG+CP)	Kinship READ (SG+CP, 1240K)	Kinship icMLkin (SG+CP, 1240K; pi_HAT>=-0.2)	Kinship β_3 (SG+CP, 1240K; β_3 =-0.35)
ELW001	19.68	87.15	78.17	59.43	37.18	24176	JK1542 (2 nd degree)	unrelated	unrelated	ELW020 (above 2 nd degree, β_3 =0.33)
ELW002	27.27	86.84	64.9	79.1	48.53	70330	unrelated	unrelated	ELW040 (above 2 nd degree, β_3 =0.33)	
ELW003	33.08	87.75	82.81	72.63	59.48	63097	ELW038 (2 nd degree)	unrelated	unrelated	
ELW004	36.75	87.96	83.2	74.34	63.11	133173	unrelated	unrelated	unrelated	
ELW005	40.34	87.74	82.85	74.37	64.54	206751	unrelated	unrelated	unrelated	
ELW006	31.51	87.66	81.47	69.84	56.45	51078	unrelated	unrelated	unrelated	
ELW007	25.57	87.43	81.42	68.42	50.91	30529	ELW017, ELW039 (2 nd degree)	unrelated	ELW039 (2 nd degree, β_3 =0.35)	
ELW008	39.13	88.26	84.36	76.82	66.7	150158	unrelated	unrelated	unrelated	
ELW009	34.22	87.49	81.8	71.52	59.59	108555	unrelated	unrelated	unrelated	
ELW010	38.25	87.76	82.64	73.88	63.2	166774	unrelated	unrelated	unrelated	
ELW011	33.22	87.42	81.65	70.99	58.63	112607	unrelated	unrelated	unrelated	
ELW012	40.53	83.88	83.88	76.02	65.82	129787	unrelated	unrelated	unrelated	
ELW013	29.07	87.98	82.8	72.37	57.26	70527	unrelated	unrelated	unrelated	
ELW014	10.89	85.47	67.74	32.92	6.66	11058	ELW017, ELW038, JK1542 (2 nd degree)	unrelated	unrelated	
ELW015	32.74	88.02	83.46	74.72	62.35	108189	ELW037 (1 st degree)	ELW037 (1 st degree, piHat=0.43)	ELW037 (1 st degree, β_3 =0.40)	
ELW016	33.56	88.08	83.67	74.99	63.03	122411	ELW017 (1 st degree)	ELW017 (1 st degree, piHat=0.41)	ELW017 (1 st degree, β_3 =0.41)	
ELW017	6.72	80.66	48.27	10.32	0.61	23353	ELW016, JK1510 (1 st degree)	ELW016 (1 st degree, piHat=0.41)	ELW016 (1 st degree, β_3 =0.41)	
ELW018	35.39	87.75	82.55	73.01	61.14	113543	ELW007, ELW014, ELW032 (2 nd degree)	unrelated	unrelated	
ELW019	34.92	87.7	82.4	72.63	60.66	111703	unrelated	unrelated	unrelated	
ELW020	35.3	87.64	82.18	72.49	60.78	119647	unrelated	unrelated	ELW001 (above 2 nd degree, β_3 =0.33)	
ELW021	34.36	87.77	82.57	73	60.9	105082	unrelated	ELW030 (2 nd degree, piHat=0.20)	ELW030 (2 nd degree, β_3 =0.35)	
ELW022	16.05	86.29	74.55	51.16	24.84	13058	ELW035 (2 nd degree)	unrelated	unrelated	
ELW023	26.63	87.32	81.59	69.23	52.33	35582	unrelated	unrelated	unrelated	
ELW024	37.57	87.99	83.66	75.4	64.55	106086	unrelated	unrelated	unrelated	
ELW025	34.65	88.08	83.77	75.2	63.39	71490	unrelated	unrelated	unrelated	
ELW026	36.8	87.53	82.55	73.28	61.96	108183	unrelated	unrelated	unrelated	
ELW027	28.75	87.42	79.96	66.67	52.35	59696	unrelated	unrelated	unrelated	
ELW028	36.06	87.99	83.19	74.08	62.39	133526	JK1542 (2 nd degree)	JK1491 (below 2 nd degree, piHat=0.12)	unrelated	
ELW029	27.18	87.59	81.36	68.98	52.91	84096	unrelated	unrelated	unrelated	
ELW030	19.99	86.86	77.51	59.17	37.21	21587	unrelated	ELW021 (2 nd degree, piHat=0.20)	ELW021 (2 nd degree, β_3 =0.35)	
ELW031	23.63	87.3	83.98	63.98	45.59	30084	unrelated	ELW034 (above 2 nd degree, piHat=0.1)	ELW034 (above 2 nd degree, β_3 =0.33)	
ELW032	23.07	87.69	81.02	66.27	45.6	27356	ELW017, JK1510 (2 nd degree)	unrelated	unrelated	
ELW033	33.32	87.52	81.23	69.75	56.88	98318	JK1510 (identical)	unrelated	unrelated	
ELW034	35.73	87.78	82.51	72.91	61.4	81324	unrelated	ELW030 (below 2 nd degree, piHat=0.1)	ELW030 (above 2 nd degree, β_3 =0.33)	
ELW035	33.41	88	82.64	72.67	59.94	63417	ELW022 (2 nd degree)	unrelated	unrelated	
ELW036	31.74	86.88	79.21	66.87	53.96	102332	ELW039 (1 st degree)	ELW039 (1 st degree, piHat=0.46)	ELW039 (1 st degree, β_3 =0.44)	
ELW037	34.5	87.73	81.4	70.15	58.08	109549	ELW015 (1 st degree)	ELW015 (1 st degree, piHat=0.43)	ELW015 (1 st degree, β_3 =0.40)	
ELW038	19.37	86.72	76.36	56.8	34.87	22562	ELW003, ELW014, JK1542 (2 nd degree)	unrelated	unrelated	
ELW039	27.63	86.68	78.12	64.31	50.15	28959	ELW036, ELW007 (1 st degree)	ELW036 (1 st degree, piHat=0.46)	ELW036 (1 st degree, β_3 =0.44)	
ELW040	31.36	88.02	82.37	71.48	57.72	49356	JK1549 (2 nd degree)	unrelated	ELW007 (2 nd degree, β_3 =0.35)	
JK1491	3.52	79.59	20.75	0.14	0	22430	JK1519 (identical)	ELW028 (above 2 nd degree, piHat=0.12)	ELW002 (2 nd degree, β_3 =0.33)	
JK1510	1.65	49.02	0.99	0	0	4838	JK1542 (2 nd degree)	unrelated	unrelated	
JK1519	3.69	77.68	23.09	0.1	0	9562	ELW017 (1 st degree)	ELW033 (identical)	unrelated	
JK1542	3.24	69.01	15.97	0.23	0	4512	ELW004 (identical), JK1510 (2 nd degree)	unrelated	unrelated	
EXB011.A0101	0.99	0.17	0	0	0	NA	ELW001, ELW014, ELW032	unrelated	NA	
EXB011.A0102	0.99	0.08	0	0	0	NA	ELW038, ELW039, JK1491 (2 nd degree)	unrelated	NA	
LIB001.A0106	1	0.09	0	0	0	NA	NA	unrelated	NA	
LIB001.A0107	0.99	0.04	0	0	0	NA	NA	unrelated	NA	
LIB001.A0108	0.99	0.53	0	0	0	NA	NA	unrelated	NA	
LIB001.A0109	0.99	0.09	0	0	0	NA	NA	unrelated	NA	

Compiled results: % target region covered>=1 (5,12,20)x after rmdup: percentage of the targeted regions covered by unique reads at least once or 5, 12, 20 times, respectively (samtools). SNPs called: No. of single nucleotide polymorphisms. 1240K: merged with the 1233013 (1240K) SNPs dataset. SG+CP: based on merged shogun and capture data. Kinship READ (icMLkin, β_3): relatedness estimations using the programs READ (icMLkin, β_3 outgroup statistics). pi_Hat: coefficient of relatedness. β_3 : measure of shared genetic drift. NA: not available/missing data.

Supplementary File 2 Ellwangen Plague Victims (Part 3)

Sample	HLA-A(1)	HLA-A(2)	HLA-B(1)	HLA-B(2)	HLA-C(1)	HLA-C(2)	DPA1(1)	DPA1(2)	DPB1(1)	DPB1(2)	DQA1(1)	DQA1(2)	DQB1(1)	DQB1(2)	DRB1(1)	DRB1(2)	DRB3/4/5(1)	DRB3/4/5(2)	CCR5-Delta32	
ELW001	33:03	02:74	07:02	18:01	07:02	07:01	01:03	01:03	03:01	04:02	01	02	03:03	06:02	07:01	15:01	4*01:03	5*01:01	wt	
ELW002	24:02	01:01	08:01	44:05	07:01	02:02	01:03	02:01	01:01	04:01	02	05	02:01	03:01	03:01	11:01	3*01:01	3*02:02	het	
ELW003	01:01	03:01	08:01	07:02	14:02	07:02	01:03	02:01	04:01	09:01	01	06	02:01	06:02	08:02	15:01	3*01:01	5*01:01	wt	
ELW004	02:01	03:01	51:01	44:05	14:02	02:02	01:03	02:01	04:01	10:01	01	05	03:01	05:03	08:03	14:01	3*02:02	NA	wt	
ELW005	24:02	32:01	40:02	38:01	12:03	02:02	01:03	01:03	03:01	04:01	01	05	03:01	05:02	11:03	16:02	3*02:02	5*02:02	wt	
ELW006	23:01	01:01	27:05	50:01	06:02	06:02	01:03	01:03	04:01	04:01	02	03	02:02	03:02	04:01	07:01	4*01:03	NA	wt	
ELW007	24:02	24:02	35:03	49:01	07:01	04:01	01:03	01:03	04:01	04:01	01	05	03:01	06:02	11:01	15:01	3*02:02	5*01:01	wt	
ELW008	02:01	01:01	35:03	35:03	04:01	04:01	01:03	01:03	04:01	04:02	01	01	05:01	06:04	01:01	13:02	3*03:01	NA	delta32	
ELW009	24:02	01:01	08:01	56:01	07:01	01:02	01:03	02:01	02:01	14:01	05	05	02:01	03:01	03:01	11:01	3*01:01	3*02:02	wt	
ELW010	24:02	68:01	51:01	15:01	16:02	03:04	01:03	01:03	04:01	04:01	01	03	03:02	06:02	04:01	15:01	4*01:03	5*01:01	wt	
ELW011	02:01	33:03	51:01	44:02	16:02	05:01	01:03	01:03	04:01	04:02	04	05	03:01	03:01	11:01	11:03	3*02:02	NA	wt	
ELW012	02:01	01:01	51:01	18:01	06:02	01:02	02:01	02:02	05:01	14:01	01	05	03:01	05:01	01:01	11:04	3*02:02	NA	het	
ELW013	03:01	03:01	07:02	07:02	07:02	07:02	01:03	01:03	04:01	04:01	01	01	05:01	06:02	01:01	15:01	5*01:01	NA	wt	
ELW014	02:01	03:01	51:01	07:02	07:02	07:02	01:03	01:03	03:01	04:02	01	06	03:01	06:02	08:03	15:01	5*01:01	NA	wt	
ELW015	26:01	29:02	40:01	58:01	07:01	03:04	01:03	02:01	04:02	14:01	01	04	05:01	04:02	01:01	08:04	NA	NA	het	
ELW016	68:01	02:05	51:01	50:01	15:02	06:02	01:03	02:01	04:01	17:01	01	02	02:02	05:02	07:01	16:01	4*01:03	5*02:02	wt	
ELW017	30:01	02:74	35:08	40:01	04:01	03:04	02:01	01:03	02:01	14:01	01	05	03:01	06:04	11:01	13:02	3*02:02	3*03:01	het	
ELW018	02:74	03:01	40:01	57:01	06:02	03:04	01:03	02:02	04:01	05:01	01	02	03:03	06:04	07:01	13:02	3*03:01	4*01:03	wt	
ELW019	24:02	68:01	40:01	38:01	03:04	12:03	01:03	01:03	04:01	04:02	03	05	02:01	03:03	03:01	09:01	3*01:01	4*01:03	wt	
ELW020	23:01	29:02	37:01	07:05	06:02	15:05	01:03	01:03	04:01	04:02	01	05	03:01	06:02	12:01	15:01	3*02:02	5*01:01	wt	
ELW021	03:01	03:01	07:02	13:02	06:02	07:02	01:03	01:03	03:01	03:01	01	01	06:02	06:03	13:01	15:01	3*02:02	5*01:01	wt	
ELW022	02:01	01:01	51:01	15:01	04:01	15:02	01:03	02:01	02:01	17:01	02	03	03:02	03:03	04:02	07:01	4*01:03	NA	het	
ELW023	02:01	01:01	08:01	08:01	07:01	07:01	02:01	02:02	01:01	01:01	05	01	02:01	02:01	03:01	03:01	3*01:01	NA	wt	
ELW024	11:01	11:01	35:01	35:01	04:01	04:01	01:03	01:03	04:01	04:01	01	03	03:01	05:03	04:01	NA	3*02:02	4*01:03	wt	
ELW025	01:01	25:01	57:01	57:01	06:02	12:03	01:03	02:01	04:01	04:01	01	02	03:03	06:02	07:01	15:01	4*01:03	5*01:01	wt	
ELW026	24:02	24:02	52:01	27:05	03:03	12:02	01:03	02:02	02:01	05:01	01	03	03:01	06:01	04:07	15:02	4*01:03	5*01:08N	wt	
ELW027	34:03	02:01	51:01	07:02	07:02	15:02	01:03	02:01	04:01	09:01	02	05	03:01	03:03	07:01	11:01	3*02:02	4*01:03	wt	
ELW028	02:01	33:01	51:01	14:02	14:02	08:02	01:??	02:02	05:01	15:01	01	01	05:01	06:03	01:02	13:01	3*01:01	NA	wt	
ELW029	33:03	03:01	35:01	44:02	07:04	04:01	01:03	01:03	04:01	04:01	01	05	03:01	05:01	01:01	11:01	3*02:02	NA	wt	
ELW030	01:01	03:01	13:02	57:01	06:02	06:02	01:03	01:03	03:01	03:01	01	01	06:02	06:03	13:01	15:01	3*02:02	5*01:01	wt	
ELW031	11:01	03:01	51:01	07:02	07:02	03:03	01:03	01:03	04:01	04:01	01	03	03:02	06:02	04:01	15:01	4*01:03	5*01:01	wt	
ELW032	02:01	25:01	27:02	18:01	07:02	12:03	01:03	01:03	02:01	03:01	01	03	03:02	05:03	04:01	15:01	4*01:03	5*01:01	wt	
ELW033	25:01	23:01	27:05	44:02	05:01	01:02	01:03	02:02	04:01	05:01	01	03	03:02	05:01	04:01	04:01	4*01:03	NA	wt	
ELW034	02:01	01:01	27:02	57:01	02:09	06:02	01:03	01:03	04:01	04:01	01	01	05:01	06:02	01:01	15:01	5*01:01	NA	wt	
ELW035	03:01	01:01	35:24	50:01	04:01	06:02	01:03	01:03	02:01	04:01	02	03	02:02	03:02	04:02	07:01	4*01:03	NA	wt	
ELW036	31:01	24:02	51:01	15:01	15:02	03:03	01:03	02:01	04:01	10:01	03	05	03:01	NA	04:08	11:01	3*02:02	4*01:03	wt	
ELW037	26:01	29:01	40:01	44:03	16:01	03:04	01:03	02:01	04:02	11:01	01	02	02:02	05:01	01:01	07:01	4*01:01	NA	delta32	
ELW038	02:13	29:01	08:01	15:01	07:01	03:04	01:03	01:03	03:01	04:02	03	05	02:01	03:02	03:01	04:01	3*01:01	4*01:03	wt	
ELW039	31:01	24:02	51:01	15:01	15:02	03:03	01:03	02:01	04:01	10:01	03	05	03:01	NA	04:08	11:01	3*02:02	4*01:03	het	
ELW040	24:02	01:01	08:01	07:02	07:02	07:01	01:03	01:03	02:01	04:01	01	05	02:01	06:02	03:01	15:01	3*01:01	5*01:01	wt	
JK1491	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	wt	
JK1510	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	wt
JK1519	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	wt
JK1542	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	wt

Compiled results. **HLA*:** HLA-allele calls using *OptiType1* and *OptiType2*. **CCR5-Delta 32:** genotype of the CCR5 chemokine receptor gene being either homozygous for the 32bp deletion ($\Delta 32/\Delta 32$), heterozygous (het) or homozygous wildtype (wt/wt). **NA:** not available/missing data.

Supplementary File 3

Ellwangen Contemporary Individuals (Part 1)

Sample	Raw Reads (CP)	Mapped Reads RMDup (CP)	Endogenous DNA (%) (CP)	CF (CP)	Mean Cov (SG)	Avg Read Length (SG)	MT/NUC (SG)	DMG 1st Base 5' (SG)	Sex (SG)	Mt Cont (SG rCRS mapped)	X Cont (SG)	Reads in Target (CP)	% Reads in Target (CP)	Reads in Target After rmdup (CP)	% Reads in Target After rmdup (CP)	Per Region Avg Readcov (CP)	Per Region Avg Readcov After rmdup (CP)
EL1	10890388	5776848	89.997	1.255	0.0776	74.69	100.98	NA	XY	NA	0.002262028	1936754	17.7	891075	8.2	113.47	51.35
EL2	10561192	6286347	90.902	1.234	0.083	75.63	98.8	NA	NA	NA	NA	1813426	17.2	893876	8.5	106.55	51.52
EL3	10615620	621021	92.841	1.202	0.167	75.03	35.92	NA	XY	NA	NA	1715767	16.2	892853	8.4	97.38	49.97
EL4	8930933	5373003	92.959	1.177	0.1262	74.62	39.3	NA	XY	NA	0.02912387	1283537	14.5	729357	8.2	76.46	42.45
EL5	10691356	6385041	92.538	1.238	0.0924	75.07	70.48	NA	XX	NA	NA	1785921	16.7	852832	8.0	106.29	49.49
EL6	9328363	5547176	89.355	1.215	0.1177	75.02	11.3	NA	XY	0.01	NA	1630475	17.5	874961	17.5	95.49	50.46
EL7	10641390	6609146	94.947	1.212	0.016	75.34	44.9	NA	XY	NA	NA	1775301	16.7	878223	8.3	102.21	49.74
EL8	10082294	6451103	91.569	1.169	0.0559	75.04	73.34	NA	XX	NA	NA	1534731	15.2	874097	8.7	88.95	50.14
EL9	10446621	6203686	93.305	1.288	0.1562	74.97	19.17	NA	XY	NA	NA	2191772	9.3	968927	21.0	127.22	55.41
EL10	9690627	6116607	93.449	1.204	0.0891	75.07	95.54	NA	XX	NA	NA	1708172	17.6	924104	9.5	98.83	52.84
EL11	8733592	5468738	93.444	1.2	0.1175	75.31	84.02	NA	XX	0.01	NA	1502574	17.2	814258	9.3	86.74	46.28
EL12	9242687	5468722	91.55	1.215	0.118	74.39	82.07	NA	NA	NA	NA	1469895	15.3	719189	7.8	81.42	40.83
EL13	10147319	6013222	90.814	1.211	0.102	65	73.35	NA	NA	NA	NA	1600953	15.8	75639	7.6	94.63	44.62
EL14	10638403	6005278	87.474	1.235	0.0983	74.78	94.72	NA	XY	NA	0.0152411	1757040	16.5	843105	7.9	103.69	48.55
EL15	8591432	5002150	89.346	1.197	0.0716	74.8	97.23	NA	XY	NA	NA	1258127	14.6	676727	14.6	74.27	39.35
EL16	8909477	4762431	81.473	1.211	0.0683	75.01	192.45	NA	XY	NA	NA	1287047	14.4	667763	7.5	73.38	37.33
EL17	10540030	5892624	89.95	1.267	0.0414	75.11	132.08	NA	XX	NA	NA	1938550	18.4	881479	8.4	112.89	50.19
EL18	11172214	6784656	91.16	1.23	0.0298	74.29	120.06	NA	XY	NA	NA	2023947	18.1	987678	8.8	114.67	54.96
EL19	11204310	6905373	92.366	1.231	0.0185	74.94	69.91	NA	XY	NA	NA	1885060	16.8	844851	7.5	106.09	46.37
EL21	11800016	7455170	95.037	1.253	0.1521	75.32	69.11	NA	XX	NA	NA	2119743	18.0	870540	7.4	121.46	48.24
EL22	11248614	7435966	94.8	1.194	0.1197	74.32	51.3	NA	XX	NA	NA	1959866	17.4	1073003	9.5	106.12	57.95
EL24	10708071	6785292	92.998	1.215	0.0978	74.05	101.12	NA	XX	NA	NA	1888562	17.6	928338	8.7	105.99	51.09
EL27	12836002	7971566	93.921	1.21	0.0887	74.94	97.59	NA	XX	NA	NA	1937550	15.5	915982	7.1	115.27	51.87
EL30	10112396	6453249	92.898	1.193	0.1116	74.96	70.16	NA	XY	0.01	0.01261083	1654432	16.4	926459	9.2	93.26	51.62
EL32	10431095	6376691	94.796	1.276	0.0729	74.8	69.13	NA	XY	NA	NA	2265564	21.7	1039645	10.0	130.78	59
EL33	10136754	6482842	94.781	1.231	0.0787	75.17	67.85	NA	XY	NA	NA	2098596	20.7	1059314	10.5	118.17	58.96
EL43	10946992	6532149	92.229	1.213	0.0713	75.05	67.51	NA	NA	NA	NA	1779522	16.3	882747	8.1	104.13	50.64
EL44	8918114	5761616	93.821	1.194	0.0889	75.18	87.56	NA	XX	NA	NA	1538186	16.8	859789	9.6	86.66	48.06
EL45	9792334	5868316	92.316	1.207	0.0164	75.2	135.67	NA	XY	NA	NA	1645256	16.8	905204	9.2	94.55	51.33
EL46	12227276	6882243	90.907	1.279	0.1205	74.87	133.35	NA	XY	NA	0.001223986	2220089	18.2	965967	7.8	131.25	54.85
EL47	1015455	5715801	89.648	1.254	0.0691	75.05	28.78	NA	XY	NA	NA	1811651	17.8	889552	8.8	105.59	50.95
EL48	10916812	6135295	85.691	1.224	0.1019	74.43	199.1	NA	XY	NA	NA	1908705	14.1	972485	9.3	110.03	43.77
EL49	10951766	6541663	93.896	1.29	0.1005	75.44	83.63	NA	XY	NA	NA	2371715	21.7	1075492	9.8	139.51	61.54
EL50	8935593	5213544	91.877	1.27	0.1237	75.26	65.41	NA	XX	NA	NA	1946568	21.8	937394	10.5	113.21	53.44
EL52	10972913	5523429	86.914	1.236	0.0241	73.87	77.65	NA	XY	NA	0.07691957	1655367	15.1	760567	6.9	100.14	45.1
EL53	10229478	6241107	92.11	1.211	0.0586	74.78	77.8	NA	XX	NA	NA	1767286	17.3	910697	8.9	103.45	52.43
EL54	10627073	5497456	89.324	1.255	0.0178	74.76	148.39	NA	XY	NA	NA	1700422	16.0	756197	7.1	102.65	44.24
EL55	12742608	6446617	89.995	1.252	0.0884	73.92	115.73	NA	XX	NA	NA	1851664	14.5	795412	6.2	113.32	47.3
EL56	11787959	5799193	80.478	1.298	0.0127	74.93	190.48	NA	XY	NA	NA	2077528	17.6	876560	7.4	122.43	50.05
EL57	9201657	5109984	93.58	1.389	0.03	75.33	166.71	NA	XX	NA	NA	1906378	20.7	480738	5.2	110.26	26.44
EL58	9790901	5424990	90.206	1.279	0.0257	75.11	166.71	NA	XY	NA	NA	2064050	21.1	935757	9.6	122.68	54.48
EL59	11102518	6375440	92.604	1.266	0.1064	74.86	111.06	NA	XY	0.01	0.01867077	2157254	19.4	1026106	9.2	125.5	58.61
EL60	10267687	5877225	93.712	1.282	0.1065	75.29	74.31	NA	XX	NA	NA	2045414	19.9	889918	8.7	120.97	51.37
EL61	16957578	236578	51.699	27.667	0.0057	65.9	56.56	NA	XY	NA	NA	2284313	13.5	519	0.0	2979.4	1.38
EL62	9116532	5018467	91.688	1.237	0.022	74.99	100.41	NA	XX	NA	NA	1524102	16.7	733125	8.0	92.15	43.12
EL63	10447719	5939215	85.306	1.238	0.0613	74.71	308.06	NA	XX	NA	NA	174196	17.0	857258	8.2	105.22	49.58
EL64	10362339	5972012	92.351	1.256	0.0793	74.85	81.15	NA	XY	NA	NA	1946405	18.8	880943	8.5	116.64	51.39
EL65	10865204	4439423	70.987	1.336	0.0435	74.94	710.05	NA	XY	0	0.006043569	1961891	17.9	843566	7.7	114.18	47.96
EL66	11030209	6412446	92.902	1.25	0.085	74.87	116.26	NA	XY	NA	NA	2105002	19.1	992624	9.0	121.46	56.13
EL74	13090671	9293819	92.988	1.25	0.0236	74.69	98.95	NA	XY	NA	0.003492648	2479527	18.9	1054049	8.8	139.72	63.88
EL106	12307231	7461627	92.397	1.232	0.0143	74.84	93.21	NA	XY	NA	NA	2191349	17.8	1054446	8.6	125.11	59.05
EL107	1172	501	66.297	1.108	0.1115	74.6	97.65	NA	XY	NA	NA	147	12.5	138	11.8	1.06	0.99

Sample	Raw Reads (CP)	Mapped Reads RMDup (CP)	Endogenous DNA (%) (CP)	CF (CP)	Mean Cov (SG)	Avg Read Length (SG)	MT/NUC (SG)	DMG 1st Base 5' (SG)	Sex (SG)	Mt Cont (SG rCRS mapped)	X Cont (SG)	Reads in Target (CP)	% Reads in Target (CP)	Reads in Target After rmdup (CP)	% Reads in Target After rmdup (CP)	Per Region Avg Readcov (CP)
EB1	182016	17287	34.249	2.607	0.0021	68.39	84.11	NA	XX	NA	NA	11291	6.2	179	0.1	63.98
EB2	146367	20681	43.464	2.512	0.0038	72.95	67.48	NA	XX	NA	NA	18688	12.8	423	0.3	44.1
EB3	83492	11465	48.103	2.796	0.0025	69.61	112.01	NA	XY	NA	NA	11681	14.0	183	0.2	65.3
EB5.1	NA	NA	NA	NA	0.0001	68.83	0	NA	NA	NA	NA	0	NA	0	0.0	0
EB5.2	44651	1294	9.276	2.379	0.0004	67.4	94.24	NA	XX	NA	NA	1037	2.3	45	0.1	22.77
EB6	77825	3973	18.827	2.914	0.0008	64.11	40.59	NA	NA	NA	NA	5360	6.9	132	0.2	39.43
LB1	269443	34908	60.299	3.552	0.0023	71.7	62.21	NA	XX	NA	NA	27378	10.2	284	0.1	90.21
LB2	380194	62106	81.153	3.95	0.0038	70.28	62.35	NA	XX	NA	NA	78595	20.7	518	0.1	151.62
LB3	23469	1844	57.74	5.924	0.0003	64.5	40.65	NA	NA	NA	NA	3363	14.3	27	0.1	132.85
LB5.1	NA	NA	NA	NA	0.0003	71.11	73.25	NA	XY	NA	NA	0	NA	0	0.0	0
LB5.2	57785	4210	66.983	7.629	0.0007	70.08	96.28	NA	XX	NA	NA	13615	23.6	51	0.1	233.5
LB6	43704	4849	78.277	5.378	0.0008	64.3	77.52	NA	NA	NA	NA	10511	24.1	57	0.1	184.3
LB7	103439	7982	73.42	7.4	0.0012	66.38	29.73	NA	NA	NA	NA	20555	19.9	78	0.1	267.24

Compiled results. Raw Reads: No. of unprocessed reads after sequencing. **CP:** based on capture data. **SG:** based on shotgun data. **Mapped Reads after rmdup:** No. of unique reads after duplicate removal, mapping to hg19 (mapped with BWA 0.7.2 with n=0.01 and MQ=30 in EAGER). **Endogenous DNA %:** percentage of raw reads mapping to hg19. **CF:** Cluster Factor (comp. Table 3.1). **Mean Cov (SG):** Average coverage of hg19 based on shotgun data. **MT/NUC:** Mitochondrial to nuclear sequence ratio. **DMG 1st base 5' (SG):** frequency of C → T substitutions at the 1st position from 5' end (*mapDamage2.0*). **Mt Cont:** contamination estimation based on mtDNA using *Schmitzi*. **X Cont:** contamination estimation in males based on X-chromosome using *ANGSD*. **Reads in Target:** No. of reads mapping to the targeted regions of interest (*samtools*). **Per Region Avg Readcov:** average coverage of target regions. **EL*:** modern saliva samples. **EB*:** extraction blanks. **LB*:** library blanks. **NA:** not available/missing data.

Supplementary File 3

Ellwangen Contemporary Individuals (Part 2)

Sample	Per Region Avg Readcov after rmdup (CP)	% target region covered==1x After rmdup (CP)	% target region covered==5x After rmdup (CP)	% target region covered==12x After rmdup (CP)	% target region covered==20x After rmdup (CP)	SNPs called (from 1233013, SG+CP)	Kinship Read (SG+CP, 1240K) EL57 (1° degree)	Kinship IcMLkin (SG+CP, 1240K; pI_HAT>=-0.2) EL57 (1° degree; piHat=0.394)	Kinship f3 (SG+CP, 1240K; f3>=-0.35) EL57 (1° degree; f3=0.40)
EL1	51.35	88.68	85.92	81.2	75.33	197289	unrelated	unrelated	unrelated
EL2	51.52	88.75	84.97	79.16	72.57	113829	unrelated	unrelated	unrelated
EL3	49.97	88.9	86.3	82.2	76.73	321413	unrelated	unrelated	unrelated
EL4	42.45	88.23	83.5	76.2	67.61	246337	unrelated	unrelated	unrelated
EL5	49.49	88.65	85.17	79.38	72.57	228218	unrelated	unrelated	unrelated
EL6	50.46	88.62	84.81	78.88	72.38	250016	unrelated	unrelated	unrelated
EL7	49.74	88.76	85.71	80.96	75.12	153087	unrelated	unrelated	unrelated
EL8	50.14	88.55	84.96	79.31	73.31	193906	unrelated	unrelated	unrelated
EL9	55.41	88.78	85.35	80.62	75.07	315849	unrelated	unrelated	unrelated
EL10	52.84	88.81	85.45	80.42	74.52	228559	unrelated	unrelated	unrelated
EL11	46.28	88.93	85.77	80.74	73.92	243895	unrelated	unrelated	unrelated
EL12	40.83	88.7	84.94	78.59	70.7	108878	unrelated	unrelated	unrelated
EL13	44.62	88.86	85.55	79.79	72.03	110947	unrelated	unrelated	unrelated
EL14	48.55	88.86	85.63	80.2	73.57	229111	unrelated	unrelated	unrelated
EL15	39.35	88.4	83.33	74.98	65.74	180756	unrelated	unrelated	unrelated
EL16	37.33	88.66	85.57	80.11	71.65	171367	unrelated	unrelated	unrelated
EL17	50.19	88.83	85.46	80.32	74.3	163321	unrelated	unrelated	unrelated
EL18	54.96	89.09	86.79	83.04	78.08	169602	unrelated	unrelated	unrelated
EL19	46.37	89.1	86.63	83.03	77.89	158549	unrelated	unrelated	unrelated
EL21	48.24	89.28	86.78	82.96	77.87	318911	unrelated	unrelated	unrelated
EL22	57.95	89.33	87.17	84.46	81.39	270257	unrelated	unrelated	unrelated
EL24	51.09	89.1	86.68	83.25	78.73	241460	unrelated	unrelated	unrelated
EL27	51.87	88.22	84.6	79.11	72.63	248115	unrelated	unrelated	unrelated
EL30	51.62	89.25	86.66	82.91	77.83	257696	unrelated	unrelated	unrelated
EL32	59	88.79	86.43	82.98	78.47	209610	unrelated	unrelated	unrelated
EL33	58.96	89.2	86.9	83.73	79.72	227975	unrelated	unrelated	unrelated
EL43	50.64	88.79	85.49	80.19	73.57	118469	unrelated	unrelated	unrelated
EL44	48.06	88.91	85.79	80.96	74.93	221632	unrelated	unrelated	unrelated
EL45	51.33	88.86	86.09	81.74	75.93	129333	unrelated	unrelated	unrelated
EL46	54.85	89	85.85	81.08	75.45	288241	unrelated	unrelated	unrelated
EL47	50.95	88.72	85.27	79.82	73.51	187972	unrelated	unrelated	unrelated
EL48	43.77	88.31	85.07	79.52	71.49	219141	unrelated	unrelated	unrelated
EL49	61.54	88.82	86.17	82.4	77.82	251998	unrelated	unrelated	unrelated
EL50	53.44	88.81	85.8	81.1	75.6	248191	unrelated	unrelated	unrelated
EL51	55.18	88.63	85.98	81.93	76.71	145343	unrelated	unrelated	unrelated
EL52	45.1	87.59	80.94	72.08	63.46	89882	unrelated	unrelated	unrelated
EL53	52.43	88.45	84.79	79.08	72.76	197339	unrelated	unrelated	unrelated
EL54	44.24	88.42	83.76	76.42	67.87	118560	unrelated	unrelated	unrelated
EL55	47.3	87.97	82.51	74.34	65.91	198350	unrelated	unrelated	unrelated
EL56	50.05	89.02	86.07	81.68	75.74	115111	unrelated	unrelated	unrelated
EL57	26.44	88.44	84.3	75.27	61.55	153011	unrelated	unrelated	unrelated
EL58	54.48	88.78	84.75	78.56	71.8	134498	unrelated	unrelated	unrelated
EL59	58.61	89	86.14	82.03	76.8	253965	unrelated	unrelated	unrelated
EL60	51.37	88.41	84.6	78.83	72.08	234699	unrelated	unrelated	unrelated
EL61	1.5	1.5	0.03	0	0	9844	unrelated	unrelated	unrelated
EL62	43.12	88.29	83.82	76.44	67.12	114550	unrelated	unrelated	unrelated
EL63	49.58	88.73	85.33	79.99	73.3	166787	unrelated	unrelated	unrelated
EL64	51.39	88.56	85.1	79.53	72.77	205378	unrelated	unrelated	unrelated
EL65	47.96	88.93	85.86	81.04	74.64	122990	unrelated	unrelated	unrelated
EL66	56.13	88.86	86.24	82.12	77.03	220245	unrelated	unrelated	unrelated
EL74	63.88	89.17	87.03	84.33	81.04	176760	unrelated	unrelated	unrelated
EL106	59.05	89.11	86.91	83.81	79.68	159451	unrelated	unrelated	unrelated
EL107	0.99	0.58	0	0	0	153553	unrelated	unrelated	unrelated

Sample	Per Region Avg Readcov after rmdup (CP)	% target region covered>=1x After rmdup (CP)	% target region covered>=5x After rmdup (CP)	% target region covered>=12x After rmdup (CP)	% target region covered>=20x After rmdup (CP)	SNPs called (from 1233013, SG+CP)	Kinship Read (SG+C, 1240K)	Kinship icMLkin (SG+C, 1240K; pi_HAT>=-0.2)	Kinship f3 (SG+C, 1240K; f3>=-0.95)
EB1	1.01	0.74	0	0	0	NA	NA	NA	NA
EB2	1.01	1.87	0	0	0	NA	NA	NA	NA
EB3	1	0.82	0	0	0	NA	NA	NA	NA
EB5.1	0	0	0	0	0	NA	NA	NA	NA
EB5.2	1.02	0.2	0	0	0	NA	NA	NA	NA
EB6	1.02	0.53	0	0	0	NA	NA	NA	NA
LB1	1	1.25	0	0	0	NA	NA	NA	NA
LB2	1.02	2.24	0	0	0	NA	NA	NA	NA
LB3	1.06	0.1	0	0	0	NA	NA	NA	NA
LB5.1	0	0	0	0	0	NA	NA	NA	NA
LB5.2	1.03	0	0	0	0	NA	NA	NA	NA
LB6	1.02	0.23	0	0	0	NA	NA	NA	NA
LB7	1.03	0.32	0	0	0	NA	NA	NA	NA

Compiled results: % **target region covered**>=1 (5,12,20)x **after rmdup**: percentage of the targeted regions covered by unique reads at least once or 5, 12, 20 times, respectively (*samtool/s*). **SNPs called**: No. of single nucleotide polymorphisms. **1240K**: merged with the 1233013 (1240K) SNPs dataset. **SG+CP**: based on merged shotgun and capture data. **Kinship READ (icMLkin, f3)**: relatedness estimations using the programs *READ (icMLkin, f3 outgroup statistics)*. **pi_Hat**: coefficient of relatedness. **f3**: measure of shared genetic drift. **NA**: not available/missing data.

Supplementary File 3 Ellwangen Contemporary Individuals (Part 3)

Sample	HLA-A(1)	HLA-A(2)	HLA-B(1)	HLA-B(2)	HLA-C(1)	HLA-C(2)	DPA1(1)	DPA1(2)	DPB1(1)	DPB1(2)	DQA1(1)	DQA1(2)	DQB1(1)	DQB1(2)	DRB1(1)	DRB1(2)	DRB3/4/5(1)	DRB3/4/5(2)	CRF5-Delta32
EL1	24:02	11:01	14:02	35:02	04:01	08:02	01:03	02:01	04:01	05:01	01	05	03:01	06:09	11:04	13:02	3*02:02	3*03:02	wt
EL2	24:02	02:01	15:07	14:01	03:03	05:01	01:03	01:03	04:02	06:01	02	04	02:02	03:02	04:04	07:01	NA	NA	het
EL3	01:01	25:01	08:01	40:01	07:01	03:04	01:03	01:03	02:01	04:01	05	05	02:01	04:02	03:01	08:01	3*01:01	NA	wt
EL4	23:01	02:01	13:02	57:01	06:02	06:02	01:03	01:03	04:01	04:01	02	02	02:02	03:03	07:01	07:01	4*01:03	NA	wt
EL5	11:01	03:01	35:01	49:01	07:01	04:01	01:03	01:03	03:01	04:01	01	03	03:02	06:02	04:05	15:01	4*01:03	5*01:01	wt
EL6	02:01	11:01	13:02	51:01	06:02	15:02	01:03	01:03	04:01	04:02	02	05	02:02	03:01	07:01	07:01	3*02:02	4*01:03	wt
EL7	03:01	11:01	14:02	07:02	07:02	08:02	01:03	01:03	02:01	04:01	01	01	06:03	06:09	13:01	13:02	3*01:01	3*03:01	wt
EL8	23:01	23:01	07:02	07:02	07:02	07:02	01:03	01:03	03:01	04:01	01	04	04:02	05:01	01:01	08:01	NA	NA	wt
EL9	02:01	68:01	07:02	44:02	07:02	05:01	01:03	01:03	02:01	04:01	01	03	03:01	06:02	04:03	15:01	4*01:03	5*01:01	het
EL10	02:01	02:13	07:02	07:02	07:02	07:02	01:03	01:03	02:01	04:01	01	03	05:03	06:02	14:54	15:01	3*02:02	5*01:01	wt
EL11	31:01	01:02	15:01	35:02	04:01	03:03	01:03	02:01	01:01	04:01	05	05	03:01	03:01	11:04	11:04	3*02:02	NA	wt
EL12	11:01	01:01	08:01	35:08	07:01	04:01	01:03	01:03	03:01	04:01	01	05	03:01	06:01	11:04	15:01	3*02:02	5*01:01	het
EL13	26:01	23:01	18:01	41:01	17:01	07:01	01:03	01:03	02:01	02:01	02	03	02:02	03:03	09:01	09:01	4*01:03	NA	wt
EL14	31:01	68:02	07:02	44:02	07:02	02:02	01:03	01:03	02:01	04:01	01	01	06:02	06:03	13:01	15:01	3*01:01	5*01:01	wt
EL15	02:01	03:01	35:01	44:02	07:04	04:01	01:03	01:03	02:01	04:01	01	01	05:01	05:01	01:01	16:01	5*02:02	NA	wt
EL16	68:01	01:01	08:01	38:01	07:01	12:03	01:03	02:01	01:01	04:01	05	05	02:01	06:03	03:01	13:01	3*01:01	NA	wt
EL17	24:02	26:01	15:01	55:01	03:03	03:03	01:03	02:01	04:01	17:01	05	05	02:01	03:01	03:01	11:03	3*01:01	3*02:02	wt
EL18	02:01	03:01	13:02	35:01	06:53	04:01	01:03	01:03	02:01	03:01	01	01	06:02	06:03	13:01	15:01	3*02:02	5*01:01	wt
EL19	24:02	02:01	40:01	46:01	03:03	03:03	01:03	02:01	03:01	17:01	03	05	03:01	03:01	04:08	11:03	3*02:02	4*01:03	het
EL20	01:01	03:01	08:01	07:02	07:01	04:01	01:03	02:01	03:01	14:01	01	01	05:01	06:02	01:01	15:01	5*01:01	NA	wt
EL21	31:01	03:01	08:01	39:01	07:01	12:03	01:03	01:03	03:01	04:02	01	05	02:01	05:02	03:01	16:01	3*01:01	5*02:02	het
EL22	24:02	03:01	08:01	07:02	07:02	04:01	01:03	02:01	03:01	04:01	01	01	06:02	06:03	13:01	15:01	3*02:02	5*01:01	wt
EL23	30:01	25:01	41:01	18:01	17:01	12:03	01:03	01:03	02:01	04:01	03	05	03:03	03:02	04:04	11:01	4*01:01	4*01:03	het
EL24	26:01	33:03	50:01	27:05	02:02	06:02	01:03	01:03	02:01	04:02	03	05	03:01	03:02	04:04	11:01	3*02:02	4*01:03	het
EL25	01:01	03:01	08:01	07:02	07:01	07:46	01:03	01:03	03:01	04:01	01	01	06:02	06:03	13:01	15:01	3*02:02	5*01:01	wt
EL26	03:01	03:01	15:01	07:02	07:02	03:03	01:03	01:03	02:01	04:01	01	02	02:02	06:02	07:01	15:01	4*01:01	5*01:01	wt
EL27	02:01	01:01	27:05	44:02	05:01	02:02	01:03	02:01	04:01	13:01	03	03	03:01	03:01	04:01	04:07	4*01:03	NA	het
EL28	11:01	03:01	15:01	18:01	03:03	12:03	01:03	02:01	04:02	10:01	01	01	05:01	05:01	10:01	11:13	3*02:02	NA	het
EL29	02:01	03:01	08:01	07:02	07:01	07:02	01:03	02:02	04:01	19:01	01	01	06:02	06:03	13:01	15:01	3*02:02	5*01:01	het
EL30	02:13	02:01	15:01	15:01	03:04	03:04	01:03	01:03	03:01	04:01	01	05	02:01	03:01	01:01	03:01	3*01:01	NA	wt
EL31	11:01	23:01	35:03	44:03	04:01	12:03	01:03	01:04	02:01	15:01	02	03	02:02	03:04	04:08	07:01	4*01:??	NA	wt
EL32	02:13	01:01	08:01	08:01	07:01	07:01	01:03	02:01	01:01	03:01	05	05	02:01	02:01	03:01	03:01	3*01:01	NA	wt
EL33	02:01	01:01	08:01	07:02	07:01	03:03	01:03	01:03	03:01	04:01	01	05	03:01	03:01	01:01	03:01	3*01:01	NA	wt
EL34	11:01	03:01	15:01	15:01	03:04	03:04	01:03	01:03	04:01	04:01	03	03	03:02	03:02	04:01	04:01	4*01:03	NA	wt
EL35	02:01	02:01	35:03	44:03	04:01	12:03	01:03	01:04	02:01	15:01	02	03	02:02	03:04	04:08	07:01	4*01:??	NA	wt
EL36	02:13	01:01	08:01	08:01	07:01	07:01	01:03	02:01	01:01	03:01	05	05	02:01	02:01	03:01	03:01	3*01:01	NA	wt
EL37	24:03	01:01	08:01	51:01	07:01	07:01	01:03	01:03	04:01	04:01	01	05	03:01	06:02	11:04	15:01	3*02:02	5*01:01	wt
EL38	02:01	01:01	08:01	15:01	07:01	03:04	01:03	02:01	01:01	04:01	01	05	02:01	06:02	03:01	15:01	3*01:01	5*01:01	wt
EL39	11:01	03:01	15:01	15:01	07:01	15:02	01:03	01:03	03:01	04:02	05	05	03:01	03:01	11:01	11:01	3*02:02	NA	wt
EL40	33:03	03:01	51:01	15:01	15:02	03:03	01:03	02:01	03:01	04:01	01	05	03:01	06:03	11:01	13:01	3*02:02	NA	wt
EL41	24:02	02:01	07:02	41:02	07:02	17:01	01:03	02:01	02:01	04:01	01	01	06:02	06:03	13:01	15:01	3*02:02	5*01:01	wt
EL42	02:01	01:01	15:01	57:01	06:02	03:03	02:01	02:02	13:01	19:01	01	02	03:03	06:03	07:01	13:01	3*02:02	4*01:03	wt
EL43	02:13	02:01	44:02	53:01	04:01	05:01	01:03	01:03	03:01	04:01	01	01	05:01	06:04	10:01	13:02	3*03:01	NA	wt
EL44	24:02	01:01	40:02	35:02	04:01	02:02	01:03	01:03	04:01	04:02	01	05	03:01	05:02	11:04	16:01	NA	5*02:02	het
EL45	02:01	11:01	27:05	53:01	04:01	01:30	01:03	02:01	11:01	20:01	01	01	05:01	05:01	01:01	01:02	NA	NA	het
EL46	66:01	02:01	51:01	41:02	17:01	01:02	01:03	02:02	04:02	05:01	05	05	03:01	03:03	03:01	13:03	3*01:01	NA	wt
EL47	29:02	01:01	08:01	44:03	07:01	16:01	01:03	01:03	04:01	04:01	02	05	02:02	03:03	03:01	07:01	3*01:01	4*01:03	wt
EL48	24:02	02:01	14:02	44:03	16:01	08:02	01:03	01:03	04:01	04:01	01	02	02:02	05:01	01:02	07:01	NA	4*01:01	wt
EL49	01:01	03:01	35:03	55:02	04:14	01:02	01:03	01:03	02:01	04:01	01	01	06:03	06:04	13:02	NA	3*01:01	NA	wt
EL50	24:02	02:01	52:01	15:17	07:01	12:02	01:03	01:03	04:01	NA	01	01	06:01	06:04	13:02	15:02	3*01:01	5*01:08N	wt
EL51	02:01	68:02	41:01	53:01	17:01	04:01	01:03	02:01	13:01	13:01	01	02	02:02	06:04	07:01	13:02	3*01:01	4*01:01	wt
EL52	23:01	33:01	14:02	44:03	04:01	08:02	01:03	01:03	02:01	23:01	05	05	02:01	03:01	11:01	11:01	3*02:02	NA	wt
EL53	24:02	01:01	35:01	44:02	04:01	05:01	01:03	02:01	01:01	04:02	01	05	02:01	06:03	03:01	13:01	3*01:01	NA	wt
EL54	24:02	03:01	51:01	35:01	04:01	05:01	01:03	01:03	03:01	04:01	01	03	02:01	03:03	03:01	15:01	5*01:01	5*01:01	wt
EL55	02:01	01:01	15:01	57:01	06:02	03:03	02:01	02:02	13:01	19:01	01	02	03:03	06:03	07:01	13:01	3*02:02	4*01:03	wt
EL56	02:13	02:01	44:02	53:01	04:01	05:01	01:03	01:03	03:01	04:01	01	01	05:01	06:04	10:01	13:02	3*03:01	NA	wt
EL57	24:02	01:01	40:02	35:02	04:01	02:02	01:03	01:03	04:01	04:02	01	05	03:01	05:02	11:04	16:01	NA	5*02:02	het
EL58	02:01	11:01	27:05	53:01	04:01	01:30	01:03	02:01	11:01	20:01	01	01	05:01	05:01	01:01	01:02	NA	NA	het
EL59	66:01	02:01	51:01	41:02	17:01	01:02	01:03	02:02	04:02	05:01	05	05	03:01	03:03	03:01	13:03	3*01:01	NA	wt
EL60	29:02	01:01	08:01	44:03	07:01	16:01	01:03	01:03	04:01	04:01	02	05	02:02	03:03	03:01	07:01	3*01:01	4*01:03	wt
EL61	24:02	02:01	14:02	44:03	16:01	08:02	01:03	01:03	04:01	04:01	01	02	02:02	05:01	01:02	07:01	NA	4*01:01	wt
EL62	01:01	03:01	35:03	55:02	04:14	01:02	01:03	01:03	02:01	04:01	01	01	06:03	06:04	13:02	NA	3*01:01	NA	wt
EL63	01:01	03:01	35:03	55:02	04:14	01:02	01:03	01:											

Sample	HLA-A(1)	HLA-A(2)	HLA-B(1)	HLA-B(2)	HLA-C(1)	HLA-C(2)	HLA-DPA1(1)	HLA-DPA1(2)	HLA-DPB1(1)	HLA-DPB1(2)	HLA-DQA1(1)	HLA-DQA1(2)	HLA-DQB1(1)	HLA-DQB1(2)	HLA-DRB1(1)	HLA-DRB1(2)	HLA-DRB3(4/5)(1)	HLA-DRB3(4/5)(2)	CCR5-Delta 32
EB1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EB2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EB3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EB5.1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EB5.2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EB6	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
LB1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
LB2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
LB3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
LB5.1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
LB5.2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
LB6	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
LB7	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Compiled results. HLA-*: HLA-allele calls using *OptiType1* and *OptiType2*. CCR5-Delta 32: genotype of the CCR5 chemokine receptor gene being either homozygous for the 32bp deletion ($\Delta 32/\Delta 32$), heterozygous (het) or homozygous wildtype (wt/wt). NA: not available/missing data.

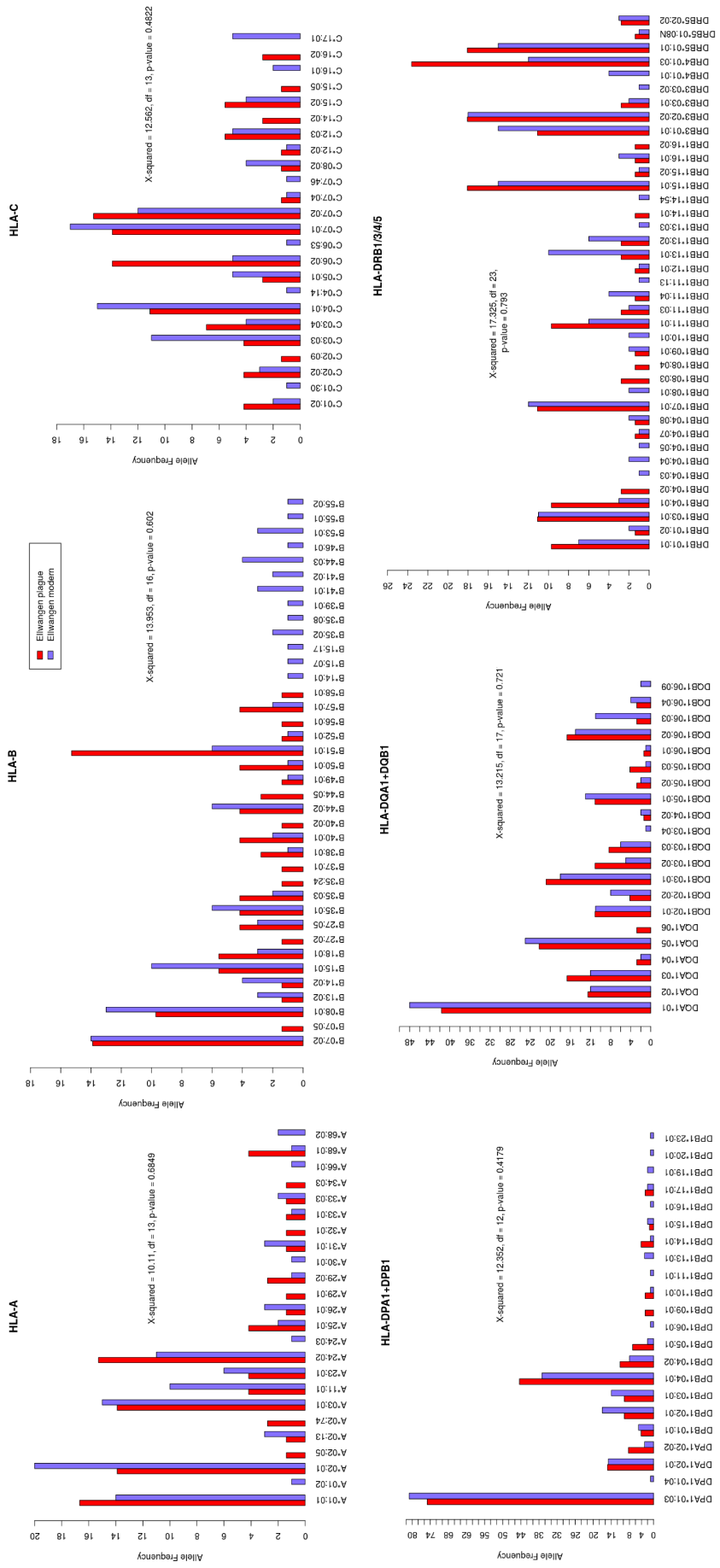


Figure S1: HLA-class I and II allele frequencies for the Ellwangen plague population (red) and the contemporary Ellwangen population (blue). Chi-squared test for the independence of the allele frequency distributions ($\alpha=0.05$) was conducted only on alleles common in both populations.

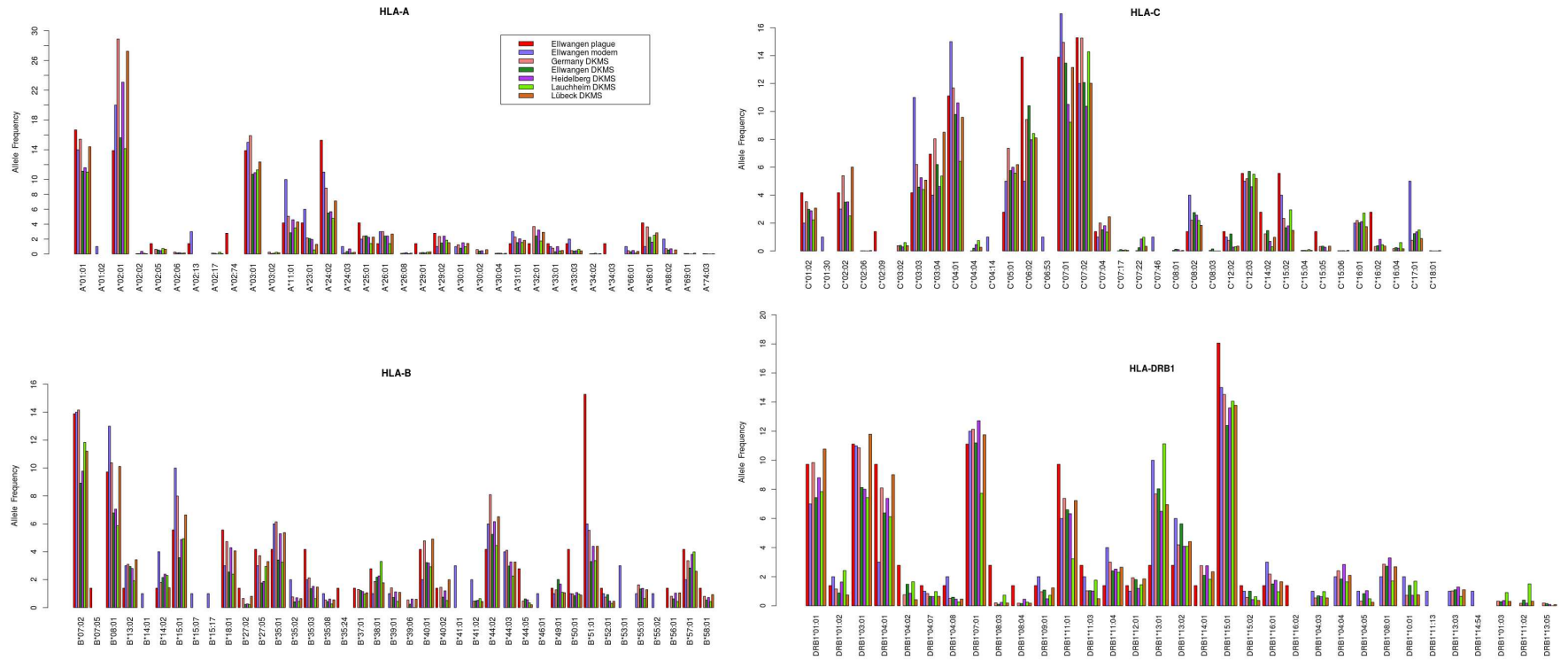


Figure S2: Barplots showing allele frequencies for HLA-A, B, C and DRB1 for both the ancient and modern Ellwangen populations and the top 30 most frequent alleles obtained from DKMS for Germany, Ellwangen, Heidelberg, Lauchheim and Lübeck.

Literature

- Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A., and Carniel, E. (1999). *Yersinia pestis*, the cause of plague, is a recently emerged clone of *yersinia pseudotuberculosis*. *Proc Natl Acad Sci U S A*, 96(24):14043–14048.
- Aken, B. L., Achuthan, P., Akanni, W., Amode, M. R., Bernsdorff, F., Bhai, J., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Gil, L., Girón, C. G., Gordon, L., Hourlier, T., Hunt, S. E., Janacek, S. H., Juettemann, T., Keenan, S., Laird, M. R., Lavidas, I., Maurel, T., McLaren, W., Moore, B., Murphy, D. N., Nag, R., Newman, V., Nuhn, M., Ong, C. K., Parker, A., Patricio, M., Riat, H. S., Sheppard, D., Sparrow, H., Taylor, K., Thormann, A., Vullo, A., Walts, B., Wilder, S. P., Zadissa, A., Kostadima, M., Martin, F. J., Muffato, M., Perry, E., Ruffier, M., Staines, D. M., Trevanion, S. J., Cunningham, F., Yates, A., Zerbino, D. R., and Flicek, P. (2017). Ensembl 2017. *Nucleic acids research*, 45:635–642.
- Albert, E. D. and Baur, M. P. and Mayr, W. R. (1984). *Histocompatibility Testing 1984*. Springer, Berlin.
- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, 19(9):1655–1664.
- Anderson, D. A., Murphy, K. M., and Briseño, C. G. (2017). Development, diversity, and function of dendritic cells in mouse and human. *Cold Spring Harbor perspectives in biology*, 10:a028613.
- Andrades Valtuena, A., Mittnik, A., Massy, K., Allmae, R., Daubaras, M., Jankauskas, R., Torv, M., Pfrengle, S., Spyrou, M. A., Feldman, M., Haak, W., Bos, K. I., Stockhammer, P. W., Herbig, A., and Krause, J. (2017). The stone age plague: 1000 years of persistence in eurasia. *Current Biology*, 27:3683–3691.
- Augusto, D. G. and Petzl-Erler, M. L. (2015). Kir and hla under pressure: evidences of coevolution across worldwide populations. *Human genetics*, 134(9):929–940.
- Auld, D. S. (1995). Removal and replacement of metal ions in metallopeptidases. *Methods in enzymology*, 248:228–242.
- Bardina, S. V., Michlmayr, D., Hoffman, K. W., Obara, C. J., Sum, J., Charo, I. F., Lu, W., Pletnev, A. G., and Lim, J. K. (2015). Differential roles of chemokines ccl2 and ccl7 in monocytosis and leukocyte migration during west nile virus infection. *Journal of immunology (Baltimore, Md. : 1950)*, 195(9):4306–4318.

- Baum, D. and Smith, S. (2013). *Tree thinking: an introduction to phylogenetic biology*. Roberts.
- Benedictow, O. J. (2004). *The Black Death, 1346-1353: The Complete History*. Boydell & Brewer.
- Bennett, S. T., Barnes, C., Cox, A., Davies, L., and Brown, C. (2005). Toward the 1,000 dollars human genome. *Pharmacogenomics*, 6(4):373–382.
- Benson, D. A., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2014). Genbank. *Nucleic acids research*, 42(Database issue):32–37.
- Bergsten, J. (2005). A review of long-branch attraction. *Cladistics*, 21(2):163–193.
- Bi, K., Vanderpool, D., Singhal, S., Linderoth, T., Moritz, C., and Good, J. M. (2012). Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC genomics*, 13:403.
- Bitam, I., Dittmar, K., Parola, P., Whiting, M. F., and Raoult, D. (2010). Fleas and flea-borne diseases. *International journal of infectious diseases*, 14(8):667–676.
- Boland, A. and Cornelis, G. R. (1998). Role of yopp in suppression of tumor necrosis factor alpha release by macrophages during yersinia infection. *Infection and immunity*, 66(5):1878–1884.
- Bos, K. I., Herbig, A., Sahl, J., Waglechner, N., Fourment, M., Forrest, S. A., Klunk, J., Schuenemann, V. J., Poinar, D., Kuch, M., Golding, G. B., Dutour, O., Keim, P., Wagner, D. M., Holmes, E. C., Krause, J., and Poinar, H. N. (2016). Eighteenth century yersinia pestis genomes reveal the long-term persistence of an historical plague focus. *eLife*, 5:e12994.
- Bos, K. I., Schuenemann, V. J., Golding, G. B., Burbano, H. A., Waglechner, N., Coombes, B. K., McPhee, J. B., DeWitte, S. N., Meyer, M., Schmedes, S., Wood, J., Earn, D. J. D., Herring, D. A., Bauer, P., Poinar, H. N., and Krause, J. (2011). A draft genome of yersinia pestis from victims of the black death. *Nature*, 478(7370):506–510.
- Bouwman, A., Shved, N., Akguel, G., Ruehli, F., and Warinner, C. (2017). Ancient dna investigation of a medieval german cemetery confirms long-term stability of ccr5- δ 32 allele frequencies in central europe. *Human Biology*, 89:119–124.
- Bradley, M. O. and Kohn, K. W. (1979). X-ray induced dna double strand break production and repair in mammalian cells as measured by neutral filter elution. *Nucleic Acids Res*, 7(3):793–804.
- Briggs, A. W., Good, J. M., Green, R. E., Krause, J., Maricic, T., Stenzel, U., Lalueza-Fox, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Schmitz, R., Doronichev, V., Golovanova, L., de la Rasilla, M., Fortea, J., Rosas, A., and Paabo, S. (2009).

- Targeted retrieval and analysis of five neandertal mtdna genomes. *Science*, 325:318–321.
- Briggs, A. W., Stenzel, U., Johnson, P. L., Green, R. E., Kelso, J., Prufer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M., and Paabo, S. (2007). Patterns of damage in genomic dna sequences from a neandertal. *Proc Natl Acad Sci U S A*, 104(37):14616–21.
- Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., and Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient dna. *Nucleic acids research*, 38(6):e87.
- Bruckner, S., Rhamouni, S., Tautz, L., Denault, J.-B., Alonso, A., Becattini, B., Salvesen, G. S., and Mustelin, T. (2005). Yersinia phosphatase induces mitochondrially dependent apoptosis of t cells. *The Journal of biological chemistry*, 280(11):10388–10394.
- Brundin, M., Figdor, D., Sundqvist, G., and Sjogren, U. (2013). Dna binding to hydroxyapatite: a potential mechanism for preservation of microbial dna. *J Endod*, 39(2):211–6.
- Butler, T. (1983). *Plague and Other Yersinia Infections*. Springer US.
- Campos, P. F., Craig, O. E., Turner-Walker, G., Peacock, E., Willerslev, E., and Gilbert, M. T. (2012). Dna in ancient bone - where is it located and how should we extract it? *Ann Anat*, 194(1):7–16.
- Cano, R. J., Poinar, H. N., Pieniazek, N. J., Acra, A., and Poinar, G. O. (1993). Amplification and sequencing of dna from a 120-135-million-year-old weevil. *Nature*, 363(6429):536–538.
- Champlot, S., Berthelot, C., Pruvost, M., Bennett, E. A., Grange, T., and Geigl, E.-M. (2010). An efficient multistrategy dna decontamination procedure of pcr reagents for hypersensitive pcr applications. *PloS one*, 5(9).
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Earlbaum Associates.
- Collins, R. W. M. (2004). Human mhc class I chain related (mic) genes: their biological function and relevance to disease and transplantation. *European journal of immunogenetics : official journal of the British Society for Histocompatibility and Immunogenetics*, 31(3):105–114.
- Comer, J. E., Sturdevant, D. E., Carmody, A. B., Virtaneva, K., Gardner, D., Long, D., Rosenke, R., Porcella, S. F., and Hinnebusch, B. J. (2010). Transcriptomic and innate immune responses to yersinia pestis in the lymph node during bubonic plague. *Infection and immunity*, 78(12):5086–5098.

- Cook, E. R., Seager, R., Kushnir, Y., Briffa, K. R., Büntgen, U., Frank, D., Krusic, P. J., Tegel, W., van der Schrier, G., Andreu-Hayles, L., Baillie, M., Baittinger, C., Bleicher, N., Bonde, N., Brown, D., Carrer, M., Cooper, R., Čufar, K., Dittmar, C., Esper, J., Griggs, C., Gunnarson, B., Günther, B., Gutierrez, E., Haneca, K., Helama, S., Herzig, F., Heussner, K.-U., Hofmann, J., Janda, P., Kontic, R., Köse, N., Kyncl, T., Levanič, T., Linderholm, H., Manning, S., Melvin, T. M., Miles, D., Neuwirth, B., Nicolussi, K., Nola, P., Panayotov, M., Popa, I., Rothe, A., Seftigen, K., Seim, A., Svarva, H., Svoboda, M., Thun, T., Timonen, M., Touchan, R., Trotsiuk, V., Trouet, V., Walder, F., Ważny, T., Wilson, R., and Zang, C. (2015). Old world megadroughts and pluvials during the common era. *Science Advances*, 1(10).
- Cooper, A. and Poinar, H. (2000). Ancient dna: Do it right or not at all. *Science*, 289:1139.
- Cornelis, G. R. (2002). The yersinia ysc-yop virulence apparatus. *International journal of medical microbiology : IJMM*, 291(6-7):455–462.
- Cortes, A. and Brown, M. A. (2011). Promise and pitfalls of the immunochip. *Arthritis research & therapy*, 13(1):101.
- Cui, Y., Yu, C., Yan, Y., Li, D., Li, Y., Jombart, T., Weinert, L., Wang, Z., Guo, Z., Xu, L., Zhang, Y., Zheng, H., Qin, N., Xiao, X., Wu, M., Wang, X., Zhou, D., Qi, Z., Du, Z., Wu, H., Yang, X., Cao, H., Wang, H., Wang, J., Yao, S., Rakin, A., Li, Y., Falush, D., Balloux, F., Achtman, M., Song, Y., Wang, J., and Yang, R. (2013). Historical variations in mutation rate in an epidemic pathogen, yersinia pestis. *Proc Natl Acad Sci U S A*, 110:577–582.
- Dabney, J., Knapp, M., Glocke, I., Gansauge, M. T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Paabo, S., Arsuaga, J. L., and Meyer, M. (2013). Complete mitochondrial genome sequence of a middle pleistocene cave bear reconstructed from ultrashort dna fragments. *Proc Natl Acad Sci U S A*, 110(39):15758–63.
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jmodeltest 2: more models, new heuristics and parallel computing. *Nature methods*, 9(8):772.
- Dean, M., Carrington, M., Winkler, C., Huttley, G., Smith, M., Allikmets, R., and Goedert, J. e. a. (1996). Genetic restriction of hiv-1 infection and progression to aids by a deletion allele of the ckr5 structural gene. hemophilia growth and development study, multicenter aids cohort study, multicenter hemophilia cohort study, san francisco city cohort, alive study. *Science*, 273:1856–1862.
- Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., and Di Marzio, P. e. a. (1996). Identification of a major co-receptor for primary isolates of hiv-1. *Nature*, 381:661–666.

- Drummond, A. J., Ho, S. Y. W., Phillips, M. J., and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS biology*, 4(5):e88.
- Dubaniewicz, A., Lewko, B., Moszkowska, G., Zamorska, B., and Stepinski, J. (2000). Molecular subtypes of the hla-dr antigens in pulmonary tuberculosis. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases*, 4(3):129–133.
- Ellwangen (1886). Beschreibung des oberamts ellwangen. band i. https://de.wikisource.org/wiki/Beschreibung_des_Oberamts_Ellwangen. Accessed: 2020-01-07.
- Enk, J., Devault, A., Widga, C., Saunders, J., Szpak, P., Southon, J., Rouillard, J.-M., Shapiro, B., Golding, G. B., Zazula, G., Froese, D., Fisher, D. C., MacPhee, R. D. E., and Poinar, H. (2016). Mammuthus population dynamics in late pleistocene north america: Divergence, phylogeography, and introgression. *Frontiers in Ecology and Evolution*, 4:42.
- Erlich, H. A., Opelz, G., and Hansen, J. (2001). Hla dna typing and transplantation. *Immunity*, 14(4):347–356.
- Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using g*power 3.1: tests for correlation and regression analyses. *Behavior research methods*, 41(4):1149–1160.
- Fehren-Schmitz, L., Kapp, J., Ziegler, K., Harkins, K., Aronsen, G., and Conlogue, G. (2016). An investigation into the effects of x-ray on the recovery of ancient dna from skeletal remains. *J. Archaeol. Sci.*, 76:1–8.
- Feldman, M., Harbeck, M., Keller, M., Spyrou, M. A., Rott, A., Trautmann, B., Scholz, H. C., Pääfgen, B., Peters, J., McCormick, M., Bos, K., Herbig, A., and Krause, J. (2016). A high-coverage yersinia pestis genome from a sixth-century justinianic plague victim. *Molecular biology and evolution*, 33(11):2911–2923.
- Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, 39:783–791.
- Freudenberg, W. (1914). Die säugetiere des älteren quartärs von mitteleuropa. *Geologische und Paläontologische Abhandlungen*, 12:455–670.
- Fricke, S. P. (2015). Targeting chemokine receptors for hiv: past, present and future. *Future medicinal chemistry*, 7(17):2311–2315.
- Frisch, T., Sørensen, M. S., Overgaard, S., Lind, M., and Bretlau, P. (1998). Volume-referent bone turnover estimated from the interlabel area fraction after sequential labeling. *Bone*, 22(6):677–682.

- Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H. A., Kelso, J., and Pääbo, S. (2013). Dna analysis of an early modern human from tianyuan cave, china. *Proc Natl Acad Sci U S A*, 110(6):2223–2227.
- 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, A. P., Drozdov, N., Slavinsky, V., Tsybankov, A., Cremonesi, R. G., Mallegni, F., Gély, B., Vacca, E., Morales, M. R. G., Straus, L. G., Neugebauer-Maresch, C., Teschler-Nicola, M., Constantin, S., Moldovan, O. T., 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, M. A., Cupillard, C., Bocherens, H., Conard, N. J., Harvati, K., Moiseyev, V., Drucker, D. G., Svoboda, J., Richards, M. P., Caramelli, D., Pinhasi, R., Kelso, J., Patterson, N., Krause, J., Pääbo, S., and Reich, D. (2016). The genetic history of ice age europe. *Nature*, 534(7606):200–205.
- Furtwaengler, A., Reiter, E., Neumann, G., Siebke, I., Steuri, N., Hafner, A., Loesch, S., Anthes, N., Schuenemann, V., and Krause, J. (2018). Ratio of mitochondrial to nuclear dna affects contamination estimates in ancient dna analysis. *Scientific Reports*, 8:s41598–018–32083–0.
- Galvani, A. P. and Novembre, J. (2005). The evolutionary history of the ccr5-delta32 hiv-resistance mutation. *Microbes and infection*, 7(2):302–309.
- Gamba, C., Jones, E., Teasdale, M., McLaughlin, R., Gonzalez-Fortes, G., Mattiangeli, V., Domboroczki, L., Kovari, I., Pap, I., Anders, A., Whittle, A., Dani, J., Raczky, P., Higham, T., Hofreiter, M., Bradley, D., and Pinhasi, R. (2014). Genome flux and stasis in a five millennium transect of european prehistory. *Nature Communications*, 5:ncomms6257.
- Gansauge, M.-T. and Meyer, M. (2013). Single-stranded dna library preparation for the sequencing of ancient or damaged dna. *Nature protocols*, 8(4):737–748.
- Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S., and Nusbaum, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature biotechnology*, 27(2):182–189.
- Goetherstrom, A., Fischer, C., and Linden, K. (1995). X-raying ancient bone: a destructive method in connection with dna analysis. *Laborativ Arkeologi*, 8:26–28.
- Golenberg, E. M., Giannasi, D. E., Clegg, M. T., Smiley, C. J., Durbin, M., Henderson, D., and Zurawski, G. (1990). Chloroplast dna sequence from a miocene magnolia species. *Nature*, 344(6267):656–658.

- Goritzka, M., Pereira, C., Makris, S., Durant, L. R., and Johansson, C. (2015). T cell responses are elicited against respiratory syncytial virus in the absence of signalling through tlrs, rlrns and il-1r/il-18r. *Scientific reports*, 5:18533.
- Gould, S. (1977). *The misnamed, mistreated, and misunderstood Irish elk*, book section 9. W.W. Norton, New York.
- Green, R. E., Malaspina, A.-S., Krause, J., Briggs, A. W., Johnson, P. L. F., Uhler, C., Meyer, M., Good, J. M., Maricic, T., Stenzel, U., Prüfer, K., Siebauer, M., Burbano, H. A., Ronan, M., Rothberg, J. M., Egholm, M., Rudan, P., Brajković, D., Kučan, Z., Gusić, I., Wikström, M., Laakkonen, L., Kelso, J., Slatkin, M., and Pääbo, S. (2008). A complete neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell*, 134(3):416–426.
- Grieshaber, B., Osborne, D., Doubleday, A., and Kaestle, F. (2008). A pilot study into the effects of x-ray and computed tomography exposure on the amplification of dna from bone. *J. Archaeol. Sci.*, 35:681–687.
- Grossovsky, A. J., de Boer, J. G., de Jong, P. J., Drobetsky, E. A., and Glickman, B. W. (1988). Base substitutions, frameshifts, and small deletions constitute ionizing radiation-induced point mutations in mammalian cells. *Proc Natl Acad Sci U S A*, 85(1):185–8.
- Gutaker, R. M., Reiter, E., Furtwängler, A., Schuenemann, V. J., and Burbano, H. A. (2017). Extraction of ultrashort dna molecules from herbarium specimens. *BioTechniques*, 62(2):76–79.
- Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., Fu, Q., Mittnik, A., Bánffy, E., Economou, C., Francken, M., Friederich, S., Pena, R. G., Hallgren, F., Khartanovich, V., Khokhlov, A., Kunst, M., Kuznetsov, P., Meller, H., Mochalov, O., Moiseyev, V., Nicklisch, N., Pichler, S. L., Risch, R., Rojo Guerra, M. A., Roth, C., Szécsényi-Nagy, A., Wahl, J., Meyer, M., Krause, J., Brown, D., Anthony, D., Cooper, A., Alt, K. W., and Reich, D. (2015). Massive migration from the steppe was a source for indo-european languages in europe. *Nature*, 522(7555):207–211.
- HannonLab (2009). Fastx-toolkit. http://hannonlab.cshl.edu/fastx_toolkit/index.html.
- Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial dna. *Journal of molecular evolution*, 22(2):160–174.
- Hayashi, F., Smith, K. D., Ozinsky, A., Hawn, T. R., Yi, E. C., Goodlett, D. R., Eng, J. K., Akira, S., Underhill, D. M., and Aderem, A. (2001). The innate immune response to bacterial flagellin is mediated by toll-like receptor 5. *Nature*, 410(6832):1099–1103.

- Higuchi, R., Bowman, B., Freiburger, M., Ryder, O. A., and Wilson, A. C. (1984). Dna sequences from the quagga, an extinct member of the horse family. *Nature*, 312(5991):282–284.
- Hinnebusch, B. J. and Erickson, D. L. (2008). *Yersinia pestis* biofilm in the flea vector and its role in the transmission of plague. *Current topics in microbiology and immunology*, 322:229–248.
- Hinnebusch, B. J., Rudolph, A. E., Cherepanov, P., Dixon, J. E., Schwan, T. G., and Forsberg, A. (2002). Role of yersinia murine toxin in survival of yersinia pestis in the midgut of the flea vector. *Science (New York, N.Y.)*, 296(5568):733–735.
- hla.alleles.org (2020). Nomenclature for factors of the hla system. <http://hla.alleles.org/nomenclature/stats.html>. Accessed: 2020-01-06.
- Hodges, E., Rooks, M., Xuan, Z., Bhattacharjee, A., Benjamin Gordon, D., Brizuela, L., Richard McCombie, W., and Hannon, G. J. (2009). Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nature protocols*, 4(6):960–974.
- Hofreiter, M., Jaenicke, V., Serre, D., von Haeseler, A., and Pääbo, S. (2001). Dna sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient dna. *Nucleic acids research*, 29(23):4793–4799.
- Hughes, S., Hayden, T. J., Douady, C. J., Tougaard, C., Germonpré, M., Stuart, A., Lbova, L., Carden, R. F., Hänni, C., and Say, L. (2006). Molecular phylogeny of the extinct giant deer, megaloceros giganteus. *Molecular phylogenetics and evolution*, 40(1):285–291.
- Hummel, S., Schmidt, D., Kremeyer, B., Herrmann, B., and Oppermann, M. (2005). Detection of the ccr5-delta32 hiv resistance gene in bronze age skeletons. *Genes and immunity*, 6(4):371–374.
- Illumina (2016). *An Introduction to Next-Generation Sequencing Technology*. https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf.
- Immel, A., Drucker, D. G., Bonazzi, M., Jahnke, T. K., Münzel, S. C., Schuenemann, V. J., Herbig, A., Kind, C.-J., and Krause, J. (2015). Mitochondrial genomes of giant deers suggest their late survival in central europe. *Scientific reports*, 5:10853.
- Immel, A., Key, F., Szolek, A., Barquera, R., Robinson, M., Spyrou, M., Susat, J., Krause-Kyora, B., Bos, K., Forrest, S., Hernandez-Zaragoza, D., Sauter, J., Solloch, U., Schmidt, A., Schuenemann, V., Reiter, E., Kairies, M., Weiss, R., Arnold, S., Wahl, J., Hollenbach, J., Kohlbacher, O., Herbig, A., Norman, P., and Krause, J. (2020). Analysis of genomic dna from late medieval plague victims suggests effect of yersinia pestis on human immunity genes. *Manuscript submitted for publication*.

- Immel, A., Le Cabec, A., Bonazzi, M., Herbig, A., Temming, H., Schuenemann, V. J., Bos, K. I., Langbein, F., Harvati, K., Bridault, A., Pion, G., Julien, M.-A., Krotova, O., Conard, N. J., Münzel, S. C., Drucker, D. G., Viola, B., Hublin, J.-J., Tafforeau, P., and Krause, J. (2016). Effect of x-ray irradiation on ancient dna in sub-fossil bones - guidelines for safe x-ray imaging. *Scientific reports*, 6:32969.
- Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P. L., and Orlando, L. (2013). map-damage2.0: fast approximate bayesian estimates of ancient dna damage parameters. *Bioinformatics*, 29(13):1682–4.
- Jukes, T. and Cantor, C. (1969). *Evolution of Protein Molecules*, pages 21–123. Academic Press, New York.
- Kang, X., Kim, J., Deng, M., John, S., Chen, H., Wu, G., Phan, H., and Zhang, C. C. (2016). Inhibitory leukocyte immunoglobulin-like receptors: Immune checkpoint proteins and tumor sustaining factors. *Cell cycle*, 15(1):25–40.
- Kawashima, Y., Kuse, N., Gatanaga, H., Naruto, T., Fujiwara, M., Dohki, S., Akahoshi, T., Maenaka, K., Goulder, P., Oka, S., and Takiguchi, M. (2010). Long-term control of hiv-1 in hemophiliacs carrying slow-progressing allele hla-b*5101. *Journal of virology*, 84(14):7151–7160.
- Keller, M., Spyrou, M. A., Scheib, C. L., Neumann, G. U., Kröpelin, A., Haas-Gebhard, B., Pfüffgen, B., Haberstroh, J., Ribera I Lacomba, A., Raynaud, C., Cessford, C., Durand, R., Stadler, P., Nägele, K., Bates, J. S., Trautmann, B., Inskip, S. A., Peters, J., Robb, J. E., Kivisild, T., Castex, D., McCormick, M., Bos, K. I., Harbeck, M., Herbig, A., and Krause, J. (2019). Ancient genomes from across western europe reveal early diversification during the first pandemic (541-750). *Proc Natl Acad Sci U S A*, 116(25):12363–12372.
- Kemp, B. M. and Smith, D. G. (2005). Use of bleach to eliminate contaminating dna from the surface of bones and teeth. *Forensic science international*, 154(1):53–61.
- Key, F. M., Posth, C., Krause, J., Herbig, A., and Bos, K. I. (2017). Mining metagenomic data sets for ancient dna: Recommended protocols for authentication. *Trends in genetics : TIG*, 33(8):508–520.
- Keys, D. (1999). *Catastrophe: an investigation into the origins of the modern world*. Ballentine Books, New York.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of molecular evolution*, 16(2):111–120.
- Kircher, M., Heyn, P., and Kelso, J. (2011). Addressing challenges in the production and analysis of illumina sequencing data. *BMC Genomics*, 12:382.

- Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the illumina platform. *Nucleic Acids Res*, 40(1):e3.
- Kitamura, H., Iwamoto, C., Sakairi, N., Tokura, S., and Nishi, N. (1997). Marked effect of dna on collagen fibrillogenesis in vitro. *International journal of biological macromolecules*, 20(3):241–244.
- Kopitar-Jerala, N. (2015). Innate immune response in brain, nf-kappa b signaling and cystatins. *Frontiers in molecular neuroscience*, 8:73.
- Korneliussen, T. S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, 15(1):356.
- Krause, J., Dear, P. H., Pollack, J. L., Slatkin, M., Spriggs, H., Barnes, I., Lister, A. M., Ebersberger, I., Pääbo, S., and Hofreiter, M. (2006). Multiplex amplification of the mammoth mitochondrial genome and the evolution of elephantidae. *Nature*, 439(7077):724–727.
- Krause, J., Fu, Q., Good, J. M., Viola, B., Shunkov, M. V., Derevianko, A. P., and Paabo, S. (2010). The complete mitochondrial dna genome of an unknown hominin from southern siberia. *Nature*, 464(7290):894–897.
- Krause, J., Unger, T., Noçon, A., Malaspinas, A.-S., Kolokotronis, S.-O., Stiller, M., Soibelzon, L., Spriggs, H., Dear, P. H., Briggs, A. W., Bray, S. C. E., O’Brien, S. J., Rabeder, G., Matheus, P., Cooper, A., Slatkin, M., Pääbo, S., and Hofreiter, M. (2008). Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the miocene-pliocene boundary. *BMC evolutionary biology*, 8:220.
- Lalani, A. S., Masters, J., Zeng, W., Barrett, J., Pannu, R., Everett, H., Arendt, C. W., and McFadden, G. (1999). Use of chemokine receptors by poxviruses. *Science (New York, N.Y.)*, 286(5446):1968–1971.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson,

D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kasprzyk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y. J., and International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007). Clustal w and clustal x version 2.0. *Bioinformatics (Oxford, England)*, 23(21):2947–2948.

Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D. C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., Connell, S., Stewardson, K., Harney, E., Fu, Q., Gonzalez-Fortes, G., Jones, E. R., Roodenberg, S. A., Lengyel, G., Bocquentin, F., Gasparian, B., Monge, J. M., Gregg, M., Eshed, V., Mizrahi, A.-S., Meiklejohn, C., Gerritsen, F., Bejenaru, L., Blüher, M., Campbell, A., Cavalleri, G., Comas, D., Froguel, P., Gilbert, E., Kerr, S. M., Kovacs, P., Krause, J., McGettigan, D., Merrigan, M., Merriwether, D. A., O'Reilly, S., Richards, M. B., Semino, O., Shamoony-Pour, M., Stefanescu, G., Stumvoll, M., Tönjes, A., Torroni, A., Wil-

- son, J. F., Yengo, L., Hovhannisyan, N. A., Patterson, N., Pinhasi, R., and Reich, D. (2016). Genomic insights into the origin of farming in the ancient near east. *Nature*, 536(7617):419–424.
- Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P. H., Schraiber, J. G., Castellano, S., Lipson, M., Berger, B., Economou, C., Bollongino, R., Fu, Q., Bos, K. I., Nordenfelt, S., Li, H., de Filippo, C., Prüfer, K., Sawyer, S., Posth, C., Haak, W., Hallgren, F., Fornander, E., Rohland, N., Delsate, D., Francken, M., Guinet, J.-M., Wahl, J., Ayodo, G., Babiker, H. A., Bailliet, G., Balanovska, E., Balanovsky, O., Barrantes, R., Bedoya, G., Ben-Ami, H., Bene, J., Berrada, F., Bravi, C. M., Brisighelli, F., Busby, G. B. J., Cali, F., Churnosov, M., Cole, D. E. C., Corach, D., Damba, L., van Driem, G., Dryomov, S., Dugoujon, J.-M., Fedorova, S. A., Gallego Romero, I., Gubina, M., Hammer, M., Henn, B. M., Hervig, T., Hodoglugil, U., Jha, A. R., Karachanak-Yankova, S., Khusainova, R., Khusnutdinova, E., Kittles, R., Kivisild, T., Klitz, W., Kučinskis, V., Kushniarevich, A., Laredj, L., Litvinov, S., Loukidis, T., Mahley, R. W., Melegh, B., Metspalu, E., Molina, J., Mountain, J., Näkkäläjärvi, K., Nesheva, D., Nyambo, T., Osipova, L., Parik, J., Platonov, F., Posukh, O., Romano, V., Rothhammer, F., Rudan, I., Ruizbakiev, R., Sahakyan, H., Sajantila, A., Salas, A., Starikovskaya, E. B., Tarekegn, A., Toncheva, D., Turdikulova, S., Uktveryte, I., Utevska, O., Vasquez, R., Villena, M., Voevoda, M., Winkler, C. A., Yepiskoposyan, L., Zalloua, P., Zemunik, T., Cooper, A., Capelli, C., Thomas, M. G., Ruiz-Linares, A., Tishkoff, S. A., Singh, L., Thangaraj, K., Vilems, R., Comas, D., Sukernik, R., Metspalu, M., Meyer, M., Eichler, E. E., Burger, J., Slatkin, M., Pääbo, S., Kelso, J., Reich, D., and Krause, J. (2014). Ancient human genomes suggest three ancestral populations for present-day europeans. *Nature*, 513(7518):409–413.
- Lenski, R. E. (1988). Evolution of plague virulence. *Nature*, 334(6182):473–474.
- Leonard, J. A., Wayne, R. K., and Cooper, A. (2000). Population genetics of ice age brown bears. *Proc Natl Acad Sci U S A*, 97(4):1651–1654.
- Li, B. and Yang, R. (2008). Interaction between yersinia pestis and the host immune system. *Infection and immunity*, 76(5):1804–1811.
- Li, C., Hofreiter, M., Straube, N., Corrigan, S., and Naylor, G. J. P. (2013). Capturing protein-coding genes across highly divergent species. *BioTechniques*, 54(6):321–326.
- Li, H. and Durbin, R. (2010). Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics*, 26(5):589–95.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–9.

- Liber, H. L., Leong, P. M., Terry, V. H., and Little, J. B. (1986). X-rays mutate human lymphoblast cells at genetic loci that should respond only to point mutagens. *Mutat Res*, 163(1):91–7.
- Lindahl, T. (1993). Instability and decay of the primary structure of dna. *Nature*, 362(6422):709–15.
- Lindo, J., Huerta-Sánchez, E., Nakagome, S., Rasmussen, M., Petzelt, B., Mitchell, J., Cybulski, J. S., Willerslev, E., DeGiorgio, M., and Malhi, R. S. (2016). A time transect of exomes from a native american population before and after european contact. *Nature communications*, 7:13175.
- Lipatov, M., Sanjeev, K., Patro, R., and Veeramah, K. (2015). Maximum likelihood estimation of biological relatedness from low coverage sequencing data. *bioRxiv*, 10.1101/023374.
- Lister, A. M., Edwards, C. J., Nock, D. A. W., Bunce, M., van Pijlen, I. A., Bradley, D. G., Thomas, M. G., and Barnes, I. (2005). The phylogenetic position of the 'giant deer' megaloceros giganteus. *Nature*, 438(7069):850–853.
- Lister, T. (1984). Evolutionary and ecological origins of british deer. *P. Roy. Soc. Edinb.*, 82(B):205–229.
- López Puga, J., Krzywinski, M., and Altman, N. (2015). Points of significance: Bayes' theorem. *Nature methods*, 12(4):277–278.
- Lu, Y.-C., Yeh, W.-C., and Ohashi, P. S. (2008). Lps/tlr4 signal transduction pathway. *Cytokine*, 42(2):145–151.
- Mahdi, B. M. (2013). A glow of hla typing in organ transplantation. *Clinical and translational medicine*, 2(1):6.
- Man, S. M. and Kanneganti, T.-D. (2015). Regulation of inflammasome activation. *Immunological reviews*, 265(1):6–21.
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J.-B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F., and Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057):376–380.

- Maricic, T., Whitten, M., and Paabo, S. (2010). Multiplexed dna sequence capture of mitochondrial genomes using pcr products. *PLoS One*, 5(11):e14004.
- Mason, V. C., Li, G., Helgen, K. M., and Murphy, W. J. (2011). Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. *Genome research*, 21(10):1695–1704.
- Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., Sirak, K., Gamba, C., Jones, E. R., Llamas, B., Dryomov, S., Pickrell, J., Arsuaga, J. L., de Castro, J. M. B., Carbonell, E., Gerritsen, F., Khokhlov, A., Kuznetsov, P., Lozano, M., Meller, H., Mochalov, O., Moiseyev, V., Guerra, M. A. R., Roodenberg, J., Vergès, J. M., Krause, J., Cooper, A., Alt, K. W., Brown, D., Anthony, D., Lalueza-Fox, C., Haak, W., Pinhasi, R., and Reich, D. (2015). Genome-wide patterns of selection in 230 ancient eurasians. *Nature*, 528(7583):499–503.
- Maxam, A. M. and Gilbert, W. (1977). A new method for sequencing dna. 1977. *Biotechnology (Reading, Mass.)*, 24:99–103.
- McBride, H. M., Neuspiel, M., and Wasiak, S. (2006). Mitochondria: more than just a powerhouse. *Current biology : CB*, 16(14):R551–R560.
- McCluskey, J., Rossjohn, J., and Purcell, A. W. (2004). Tap genes and immunity. *Current opinion in immunology*, 16(5):651–659.
- Mellins, E. D. and Stern, L. J. (2014). Hla-dm and hla-do, key regulators of mhc-ii processing and presentation. *Current opinion in immunology*, 26:115–122.
- Meyer, M., Arsuaga, J.-L., de Filippo, C., Nagel, S., Aximu-Petri, A., Nickel, B., Martinez, I., Gracia, A., de Castro, J. M. B., Carbonell, E., Viola, B., Kelso, J., Pruefer, K., and Paabo, S. (2016). Nuclear dna sequences from the middle pleistocene sima de los huesos hominins. *Nature*, 531:504–507.
- Meyer, M., Fu, Q., Aximu-Petri, A., Glocke, I., Nickel, B., Arsuaga, J.-L., Martinez, I., Gracia, A., de Castro, J. M. B., Carbonell, E., and Paabo, S. (2014). A mitochondrial genome sequence of a hominin from sima de los huesos. *Nature*, 505:403–406.
- Meyer, M. and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc*, 2010(6):pdb.prot5448.
- Minnich, S. and Rohde, H. (2007). A rationale for repression and/or loss of motility by pathogenic yersinia in the mammalian host. *Adv Exp Med Biol.*, 603:298–310.
- Monroy, K., Jose, M., Jakobsson, M., and Günther, T. (2017). Estimating genetic kin relationships in prehistoric populations. *bioRxiv*, 10.1101/100297.

- Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics*, 38.
- Muller, H. J. (1927). Artificial transmutation of the gene. *Science*, 66(1699):84–7.
- NIH (2019). The cost of sequencing a human genome. <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>. Accessed: 2019-11-26.
- Norman, P., Hollenbach, J., Nemat-Gorgani, N., Guethlein, L., Hilton, H., Pando, M., Koram, K., Riley, E., Abi-Rached, L., and Parham, P. (2013). Co-evolution of human leukocyte antigen (hla) class i ligands with killer-cell immunoglobulin-like receptors (kir) in a genetically diverse population of sub-saharan africans. *PLoS Genet*, 9:10.1371/journal.pgen.1003938.
- Nymo, S., Gustavsen, A., Nilsson, P. H., Lau, C., Espevik, T., and Mollnes, T. E. (2016). Human endothelial cell activation by escherichia coli and staphylococcus aureus is mediated by tnf and il-1 β secondarily to activation of c5 and cd14 in whole blood. *Journal of immunology (Baltimore, Md. : 1950)*, 196(5):2293–2299.
- Pääbo, S. (1985). Molecular cloning of ancient egyptian mummy dna. *Nature*, 314(6012):644–645.
- Pääbo, S. (1989). Ancient dna: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc Natl Acad Sci U S A*, 86(6):1939–43.
- Pääbo, S., R.G., H., and A.C., W. (1989). Ancient dna and the polymerase chain reaction. *J Biol Chem*, 264:9709–9712.
- Pamer, E. and Cresswell, P. (1998). Mechanisms of mhc class I-restricted antigen processing. *Annu. Rev. Immunol.*, 16:593–617.
- Pandey, V., Nutter, R. C., and Prediger, E. (2008). Applied biosystems solid™ system: Ligation-based sequencing. *Next Generation Genome Sequencing: Towards Personalized Medicine*, pages 29–42.
- Paredes, U., Prys-Jones, R., Adams, M., Groombridge, J., Kundu, S., Agapow, P.-M., and Abel, R. (2012). Micro-ct x-rays do not fragment dna in preserved bird skins. *J. Zool. Sys. Evol. Res.*, 50:247–250.
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. *Genetics*, 192(3):1065–1093.
- Patterson, N., Price, A. L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS genetics*, 2(12):e190.
- Pearson, K. (1901). On lines and planes of closest fit to a system of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 6:559–572.

- Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., and Nieselt, K. (2016). Eager: efficient ancient genome reconstruction. *Genome biology*, 17:60.
- Peñalba, J. V., Smith, L. L., Tonione, M. A., Sass, C., Hykin, S. M., Skipwith, P. L., McGuire, J. A., Bowie, R. C. K., and Moritz, C. (2014). Sequence capture using PCR-generated probes: a cost-effective method of targeted high-throughput sequencing for nonmodel organisms. *Molecular ecology resources*, 14(5):1000–1010.
- Perry, R. D. and Fetherston, J. D. (1997). *Yersinia pestis* - etiologic agent of plague. *Clinical microbiology reviews*, 10(1):35–66.
- Pfeiffer, T. (1999). Die stellung von dama (cervidae, mammalia) im system pleisometacarpaler hirsche des pleistozäns. phylogenetische rekonstruktion - metrische analyse. *Cour. Forsch. Inst. Senck.*, 211:1–218.
- Pfeiffer, T. (2002). The first complete skeleton of megaloceros verticornis (dawkins, 1868) cervidae, mammalia, from bilshausen (lower saxony, germany): description and phylogenetic implications. *Fossil Record*, 5:289–308.
- Pieters, J. (2008). Mycobacterium tuberculosis and the macrophage: maintaining a balance. *Cell host & microbe*, 3(6):399–407.
- Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S., Alpaslan-Roodenberg, S., Gerritsen, F., Moiseyev, V., Gromov, A., Raczky, P., Anders, A., Pietrusewsky, M., Rollefson, G., Jovanovic, M., Trinhhoang, H., Bar-Oz, G., Oxenham, M., Matsumura, H., and Hofreiter, M. (2015). Optimal ancient dna yields from the inner ear part of the human petrous bone. *PloS one*, 10(6):e0129102.
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, 38(8):904–909.
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Rascovan, N., Sjögren, K.-G., Kristiansen, K., Nielsen, R., Willerslev, E., Desnues, C., and Rasmussen, S. (2019). Emergence and spread of basal lineages of yersinia pestis during the neolithic decline. *Cell*, 176(1-2):295–305.e10.
- Rasmussen, S., Allentoft, M. E., Nielsen, K., Orlando, L., Sikora, M., Sjögren, K.-G., Pedersen, A. G., Schubert, M., Van Dam, A., Kapel, C. M. O., Nielsen, H. B., Brunak, S., Avetisyan, P., Epimakhov, A., Khalyapin, M. V., Gnuni, A., Kriiska, A., Lasak, I., Metspalu, M., Moiseyev, V., Gromov, A., Pokutta, D., Saag, L., Varul, L., Yepiskoposyan, L., Sicheritz-Pontén, T., Foley, R. A., Lahr, M. M., Nielsen, R., Kristiansen, K., and Willerslev, E. (2015). Early divergent strains of yersinia pestis in eurasia 5,000 years ago. *Cell*, 163(3):571–582.

- Reich, D. E. and Goldstein, D. B. (1998). *Estimating the age of mutations using the variation at linked markers*. Oxford University Press, Oxford.
- Renaud, G., Slon, V., Duggan, A. T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient dna. *Genome biology*, 16:224.
- Richards, G. D., Jabbour, R. S., Horton, C. F., Ibarra, C. L., and MacDowell, A. A. (2012). Color changes in modern and fossil teeth induced by synchrotron microtomography. *Am J Phys Anthropol*, 149(2):172–80.
- Rizzi, E., Lari, M., Gigli, E., De Bellis, G., and Caramelli, D. (2012). Ancient dna studies: new perspectives on old samples. *Genetics, selection, evolution : GSE*, 44:21.
- Robinson, J., Halliwell, J. A., McWilliam, H., Lopez, R., and Marsh, S. G. E. (2013). Ipd—the immuno polymorphism database. *Nucleic acids research*, 41(Database issue):1234–1240.
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., and Mesirov, J. P. (2011). Integrative genomics viewer. *Nature biotechnology*, 29(1):24–26.
- Rogaev, E. I., Moliaka, Y. K., Malyarchuk, B. A., Kondrashov, F. A., Derenko, M. V., Chumakov, I., and Grigorenko, A. P. (2006). Complete mitochondrial genome and phylogeny of pleistocene mammoth *mammuthus primigenius*. *PLoS biology*, 4(3):e73.
- Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil-dna-glycosylase treatment for screening of ancient dna. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 370(1660):20130624.
- Rohland, N. and Hofreiter, M. (2007). Ancient dna extraction from bones and teeth. *Nat Protoc*, 2(7):1756–62.
- Rohland, N., Malaspinas, A.-S., Pollack, J. L., Slatkin, M., Matheus, P., and Hofreiter, M. (2007). Proboscidean mitogenomics: chronology and mode of elephant evolution using mastodon as outgroup. *PLoS biology*, 5(8):e207.
- Ronquist, F. and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxford, England)*, 19(12):1572–1574.
- Roots, R. and Okada, S. (1975). Estimation of life times and diffusion distances of radicals involved in x-ray-induced dna strand breaks of killing of mammalian cells. *Radiat Res*, 64(2):306–20.
- Ruiz, T. F. (2001). *Medieval Europe: Crisis and Renewal*. The Teaching Company.

- Rusk, N. (2011). Torrents of sequence. *Nature Methods*, 8:nmeth.f.330.
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977). Dna sequencing with chain-terminating inhibitors. 1977. *Biotechnology (Reading, Mass.)*, 24:104–108.
- Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., and Paabo, S. (2012). Temporal patterns of nucleotide misincorporations and dna fragmentation in ancient dna. *PLoS One*, 7(3):e34131.
- Schmidt, A. H., Baier, D., Solloch, U. V., Stahr, A., Cereb, N., Wassmuth, R., Ehninger, G., and Rutt, C. (2009). Estimation of high-resolution hla-a, -b, -c, -drb1 allele and haplotype frequencies based on 8862 german stem cell donors and implications for strategic donor registry planning. *Human immunology*, 70(11):895–902.
- Schmidt, H. and Haeseler, V. (2009). *Phylogenetic inference using maximum likelihood methods*, pages 181–198. Cambridge University Press, Cambridge.
- Schwarz, C., Debruyne, R., Kuch, M., McNally, E., Schwarcz, H., Aubrey, A. D., Bada, J., and Poinar, H. (2009). New insights from old bones: Dna preservation and degradation in permafrost preserved mammoth remains. *Nucleic Acids Res*, 37(10):3215–29.
- Shapiro, B., Drummond, A. J., Rambaut, A., Wilson, M. C., Matheus, P. E., Sher, A. V., Pybus, O. G., Gilbert, M. T. P., Barnes, I., Binladen, J., Willerslev, E., Hansen, A. J., Baryshnikov, G. F., Burns, J. A., Davydov, S., Driver, J. C., Froese, D. G., Harington, C. R., Keddie, G., Kosintsev, P., Kunz, M. L., Martin, L. D., Stephenson, R. O., Storer, J., Tedford, R., Zimov, S., and Cooper, A. (2004). Rise and fall of the beringian steppe bison. *Science (New York, N.Y.)*, 306(5701):1561–1565.
- Shapiro, B. and Hofreiter, M. (2014). A paleogenomic perspective on evolution and gene function: new insights from ancient dna. *Science (New York, N.Y.)*, 343(6169):1236573.
- Shuxian, J., Xiaoyun, C., Zhihui, F., Xiaohua, L., Zhanhui, D., Bin, H., and Lin, Z. (2014). Association of hla-b*51:01 with papillary thyroid carcinoma in the chinese han population of the shandong coastal areas. *Thyroid : official journal of the American Thyroid Association*, 24(5):867–871.
- Sing, A., Rost, D., Tvardovskaia, N., Roggenkamp, A., Wiedemann, A., Kirschning, C. J., Aepfelbacher, M., and Heesemann, J. (2002). Yersinia v-antigen exploits toll-like receptor 2 and cd14 for interleukin 10-mediated immunosuppression. *The Journal of experimental medicine*, 196(8):1017–1024.
- Skoglund, P., Storå, J., Goetherstroem, A., and Jakobsson, M. (2013). Accurate sex identification of ancient human remains using dna shotgun sequencing. *Journal of Archaeological Science*, 40:4477–4482.

- Smith, C. I., Chamberlain, A. T., Riley, M. S., Cooper, A., Stringer, C. B., and Collins, M. J. (2001). Neanderthal dna. not just old but old and cold? *Nature*, 410(6830):771–772.
- Soubrier, J., Gower, G., Chen, K., Richards, S. M., Llamas, B., Mitchell, K. J., Ho, S. Y. W., Kosintsev, P., Lee, M. S. Y., Baryshnikov, G., Bollongino, R., Bover, P., Burger, J., Chivall, D., Crégut-Bonnoure, E., Decker, J. E., Doronichev, V. B., Douka, K., Fordham, D. A., Fontana, F., Fritz, C., Glimmerveen, J., Golovanova, L. V., Groves, C., Guerreschi, A., Haak, W., Higham, T., Hofman-Kaminska, E., Immel, A., Julien, M.-A., Krause, J., Krotova, O., Langbein, F., Larson, G., Rohrlach, A., Scheu, A., Schnabel, R. D., Taylor, J. F., Tokarska, M., Tosello, G., van der Plicht, J., van Loenen, A., Vigne, J.-D., Wooley, O., Orlando, L., Kowalczyk, R., Shapiro, B., and Cooper, A. (2016). Early cave art and ancient dna record the origin of european bison. *Nature communications*, 7:13158.
- Spyrou, M. A., Tikhbatova, R. I., Feldman, M., Drath, J., Kacki, S., Beltrán de Heredia, J., Arnold, S., Sitdikov, A. G., Castex, D., Wahl, J., Gazimzyanov, I. R., Nurgaliev, D. K., Herbig, A., Bos, K. I., and Krause, J. (2016). Historical y.pestis genomes reveal the european black death as the source of ancient and modern plague pandemics. *Cell host & microbe*, 19(6):874–881.
- Spyrou, M. A., Tikhbatova, R. I., Wang, C.-C., Valtueña, A. A., Lankapalli, A. K., Kondrashin, V. V., Tsybin, V. A., Khokhlov, A., Kühnert, D., Herbig, A., Bos, K. I., and Krause, J. (2018). Analysis of 3800-year-old yersinia pestis genomes suggests bronze age origin for bubonic plague. *Nature communications*, 9(1):2234.
- Stephens, J. C., Reich, D. E., Goldstein, D. B., Shin, H. D., Smith, M. W., Carrington, M., Winkler, C., Huttley, G. A., Allikmets, R., Schriml, L., Gerrard, B., Malasky, M., Ramos, M. D., Morlot, S., Tzetzis, M., Oddoux, C., di Giovine, F. S., Nasioulas, G., Chandler, D., Aseev, M., Hanson, M., Kalaydjieva, L., Glavac, D., Gasparini, P., Kanavakis, E., Claustres, M., Kambouris, M., Ostrer, H., Duff, G., Baranov, V., Sibul, H., Metspalu, A., Goldman, D., Martin, N., Duffy, D., Schmidtke, J., Estivill, X., O’Brien, S. J., and Dean, M. (1998). Dating the origin of the ccr5-delta32 aids-resistance allele by the coalescence of haplotypes. *American journal of human genetics*, 62(6):1507–1515.
- Stuart, A. J., Kosintsev, P. A., Higham, T. F. G., and Lister, A. M. (2004). Pleistocene to holocene extinction dynamics in giant deer and woolly mammoth. *Nature*, 431(7009):684–689.
- Sun, J., Luan, Y., Xiang, D., Tan, X., Chen, H., Deng, Q., Zhang, J., Chen, M., Huang, H., Wang, W., Niu, T., Li, W., Peng, H., Li, S., Li, L., Tang, W., Li, X., Wu, D., and Wang, P. (2016). The 11s proteasome subunit psme3 is a positive feedforward regulator of nf- κ b and important for host defense against bacterial pathogens. *Cell reports*, 14(4):737–749.

- Szolek, A., Schubert, B., Mohr, C., Sturm, M., Feldhahn, M., and Kohlbacher, O. (2014). Optitype: precision hla typing from next-generation sequencing data. *Bioinformatics (Oxford, England)*, 30(23):3310–3316.
- Takahashi, K. and Kawai, T. (2007). Pathogen recognition by toll-like receptor. *Nihon rinsho. Japanese journal of clinical medicine*, 65 Suppl 2 Pt. 1:53–57.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). Mega6: Molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution*, 30(12):2725–2729.
- Trowsdale, J. (2011). The mhc, disease and selection. *Immunol. Lett*, 137:1–8.
- Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J. A. (2007). Primer3plus, an enhanced web interface to primer3. *Nucleic Acids Res*, 35(Web Server issue):W71–4.
- Velan, B., Bar-Haim, E., Zauberman, A., Mamroud, E., Shafferman, A., and Cohen, S. (2006). Discordance in the effects of yersinia pestis on the dendritic cell functions manifested by induction of maturation and paralysis of migration. *Infection and immunity*, 74(11):6365–6376.
- Wagner, D. M., Klunk, J., Harbeck, M., Devault, A., Waglechner, N., Sahl, J. W., Enk, J., Birdsell, D. N., Kuch, M., Lumibao, C., Poinar, D., Pearson, T., Fourment, M., Golding, B., Riehm, J. M., Earn, D. J. D., Dewitte, S., Rouillard, J.-M., Grupe, G., Wiechmann, I., Bliska, J. B., Keim, P. S., Scholz, H. C., Holmes, E. C., and Poinar, H. (2014). Yersinia pestis and the plague of justinian 541-543 ad: a genomic analysis. *The Lancet. Infectious diseases*, 14(4):319–326.
- Wahl, J. (2014). Ellwangen: Von massengraeberen, arthrose und wurzelfrass - erste ergebnisse zum anthropologischen befund. Ellwangen.
- Wallace, G. R. (2014). Hla-b*51 the primary risk in behçet disease. *Proc Natl Acad Sci U S A*, 111(24):8706–8707.
- Wang, F., Mao, L., Hou, H., Wu, S., Huang, M., Yin, B., Huang, J., Zhu, Q., Pan, Y., and Sun, Z. (2016). The source of mycobacterium tuberculosis-specific ifn- γ production in peripheral blood mononuclear cells of tb patients. *International immunopharmacology*, 32:39–45.
- Watts, C. (1997). Capture and processing of exogenous antigens for presentation on mhc molecules. *Annu. Rev. Immunol.*, 15:821–850.
- Weese, D., Holtgrewe, M., and Reinert, K. (2012). Razers 3: faster, fully sensitive read mapping. *Bioinformatics (Oxford, England)*, 28(20):2592–2599.

- Welkos, S., Friedlander, A., McDowell, D., Weeks, J., and Tobery, S. (1998). V antigen of yersinia pestis inhibits neutrophil chemotaxis. *Microbial pathogenesis*, 24(3):185–196.
- Wetterstrand, K. A. (2016). Dna sequencing costs: Data from the nhgri genome sequencing program (gsp) available at:. www.genome.gov/sequencingcostsdata. Accessed: 2016-11-13.
- WHO (2019). Plague. <http://www.who.int/mediacentre/factsheets/fs267/en/>.
- Wolff, S. (1967). Radiation genetics. *Annu. Rev. Genet.*, 1:221–224.
- Woodward, S. R., Weyand, N. J., and Bunnell, M. (1994). Dna sequence from cretaceous period bone fragments. *Science (New York, N.Y.)*, 266(5188):1229–1232.
- Yang, Z. and Rannala, B. (2012). Molecular phylogenetics: principles and practice. *Nature reviews. Genetics*, 13(5):303–314.
- Zimble, D. L., Schroeder, J. A., Eddy, J. L., and Lathem, W. W. (2015). Early emergence of yersinia pestis as a severe respiratory pathogen. *Nature communications*, 6:7487.

Veröffentlichungen und Vorträge

Publikationen

- (1) **Immel, A.**, Drucker, D.G., Bonazzi, M., Jahnke, T.K., Münzel, S.C., Schuenemann, V.J., Herbig, A. Kind, C.-J., and Krause, J. (2015). Mitochondrial Genomes of Giant Deers suggest their Late Survival in Central Europe. *Scientific Reports*, 5:10853
- (2) **Immel, A.**, Le Cabec, A., Bonazzi, M., Herbig, A., Temming, H., Schuenemann, V.J., Bos, K.I., Langbein, F., Harvati, K., Bridault, A., Pion, G., Julien, M.-A., Krotova, O., Conard, N.J., Münzel, S.C., Drucker, D.G., Viola, B., Hublin, J.-J., Tafforeau, P., and Krause, J. (2016). Effect of X-ray Irradiation on ancient DNA in sub-fossil Bones - Guidelines for safe X-ray Imaging. *Scientific Reports*, 6:32969.
- (3) **Immel, A.**, Key, F.M., Szolek, A., Barquera, R., Robinson, M.K., Harrison, G.F., Palmer, W.H., Spyrou, M.A., Susat, J., Krause-Kyora, B., Bos, K.I., Forrest, S., Hernández-Zaragoza, D.I., Sauter, J., Solloch, U., Schmidt, A.H., Schuenemann, V.J., Reiter, E., Kairies, M.S., Weiß, R., Arnold, S., Wahl, J., Hollenbach, J.A., Kohlbacher, O., Herbig, A., Norman, P.J., and Krause, J. (2020). Analysis of genomic DNA from medieval plague victims suggests effect of *Yersinia pestis* on human immunity genes. *Molecular Biology and Evolution*, submitted.
- (4) **Immel, A.**, Terna, S., Simalcsik, A., Susat, J., Sarov, O., Sirbu, G., Hofmann, R., Müller, J., Nebel, A., and Krause-Kyora, B. (2020). Gene-flow from steppe individuals into Cucuteni-Trypillia associated populations indicates long-standing contacts and gradual admixture. *Scientific Reports*, 10:4253.
- (5) Soubrier, J., Gower, G., Chen, K., Richards, S.M., Lllamas, B., Mitchell, K.J., Ho, S.Y.W., Kosintsev, P., Lee, M.S.Y., Baryshnikov, G., Bollongino, R., Bover, P., Burger, J., Chivall, D., Crégut-Bonnoure, E., Decker, J.E., Doronichev, V.B., Douka, K., Fordham, D.A., Fontana, F., Fritz, C., Glimmerveen, J., Golovanova, L.V., Groves, C., Guerreschi, A., Haak, W., Higham, T., Hofman-Kaminska, E., **Immel, A.**, Julien, M.-A., Krause, J., Krotova, O., Langbein, F., Larson, G., Rohrlach, A., Scheu, A., Schnabel, R.D., Taylor, J.F., Tokarska, M., Tosello, G., van der Plicht, J., van Loenen, A., Vigne, J.-D., Wooley, O., Orlando, L., Kowalczyk, R., Shapiro, B., and Cooper, A. (2016). Early Cave Art and ancient DNA record the Origin of European Bison. *Nature Communications*, 7:13158
- (6) Krause-Kyora, B., Susat, J., Key, F.M., Kühnert, D., Bosse, E., **Immel, A.**, Rinne, C., Kornell, S.C., Yepes, D., Franzenburg, S., Heyne, H.O., Meier, T., Lösch, S., Meller, H., Friederich, S., Nicklisch, N., Alt, K.W., Schreiber, S., Tholey, A., Herbig, A., Nebel, A., and Krause, J. (2018). Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. *Elife*, doi: 10.7554/eLife.36666

- (7) Jeong, C., Balanovsky, O., Lukianova, E., Kahbatkyzy, N., Flegontov, P., Zaporozhchenko, V., **Immel, A.**, Wang, C., Ixan, O., Khussainova, E., Bekmanov, B., Zaibert, V., Lavryashina, M., Pocheshkhova, E., Yusupov, Y., Agdzhoyan, A., Koshel, S., Bukin, A., Nymadawa, P., Turdikulova, S., Dalimova, D., Churnosov, M., Skhalyakho, R., Daragan, D., Bogunov, Y., Bogunova, A., Shtrunov, A., Dubova, N., Zhabagin, M., Yepiskoposyan, L., Churakov, V., Pislegin, N., Damba, L., Saroyants, L., Dibirova, K., Atramentova, L., Utevska, O., Idrisov, E., Kamenshchikova, E., Evseeva, I., Metspalu, M., Outram, A., Robbeets, M., Djansugurova, L., Balanovska, E., Schiffels, S., Haak, W., Reich, D., and Krause, J. (2019). The genetic history of admixture across inner Eurasia. *Nature Ecology & Evolution*, 3:966-976

Vorträge

- (1) **ISBA 2014:** 6th International Symposium on Biomolecular Archaeology; August 27-29, 2014; Location: Basel, Switzerland; Title: *A complete mitochondrial Genome of the Giant Deer*
- (2) **Meeting StEvE 2014:** The meeting of Students in Evolution and Ecology; October 23-24, 2014; Location: Tübingen, Germany; Title: *A complete mitochondrial Genome of the Giant Deer*
- (3) **Plenary Talk:** April 30, 2015; Location: Ellwangen, Germany; Title: *Paläogenetische Untersuchungen der Pest in Ellwangen*
- (4) **ISBA 2016:** 7th International Symposium on Biomolecular Archaeology; September 14-16, 2016; Location: Oxford, Great Britain; Title: *Effect of X-ray Irradiation on ancient DNA in sub-fossil Bones - Guidelines for safe X-ray Imaging*
- (5) **Plenary Talk:** June 13, 2019; Location: Natural History Museum Berlin, Germany; Title: *Effect of X-ray Irradiation on ancient DNA in sub-fossil Bones - Guidelines for safe X-ray Imaging*

Poster

- (1) **SMBE 2015:** Society for Molecular Biology and Evolution; July 12-16, 2015; Location: Vienna, Austria; Title: *Quantitative Evaluation of ancient DNA losses during Computed Tomography Scanning of Palaeontological Remains*

Ehrenwörtliche Erklärung

Hiermit erkläre ich, dass

- mir die Promotionsordnung der Pharmazeutisch-Biologischen Fakultät der Friedrich-Schiller-Universität Jena bekannt ist,
- ich die Dissertation selbst angefertigt habe und alle von mir benutzten Hilfsmittel, persönlichen Mitteilungen und Quellen in meiner Arbeit angegeben sind,
- mich folgende Personen bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts unterstützt haben: Prof. Dr. Johannes Krause, Dr. Alexander Herbig
- die Hilfe eines Promotionsberaters nicht in Anspruch genommen wurde,
- Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen,
- ich die gleiche, eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung nicht bei einer anderen Hochschule als Dissertation eingereicht habe.

Alexander Immel

Jena, den

Prozentuale Anteile der eigens erbrachten Leistung

1. Studie:

Mitochondrial Genomes of Giant Deers suggest their Late Survival in Central Europe; publiziert in *Scientific Reports* am 08.06.2015

Author Contributions: *A.I. and J.K. conceived and designed the research. T.K.J., S.C.M. and C.J.K. provided the bone samples and conducted initial morphological analyses. M.B. and V.J.S. performed the extraction of mtDNA and the preparation of sequencing libraries. A.I. and A.H. performed the bioinformatic analyses and D.G.D. conducted the stable isotope analyses. A.I. wrote the manuscript and J.K. and D.G.D. mostly contributed to the discussion. All authors reviewed the manuscript.*

Eigene Beiträge:

- bioinformatische Prozessierung der Sequenzdaten der beiden Riesenhirsch-DNA Proben
- Rekonstruktion der mitochondrialen Sequenzen der Riesenhirschgenome
- Vergleich mit mitochondrialen Genomsequenzen rezenter Hirsche
- Rekonstruktion der Phylogenien
- Verfassen des Manuskripts

Beiträge von Co-Autoren:

- Bereitstellung von Knochenproben und morphologische Untersuchung
- DNA-Extraktion und Vorbereitung der Sequenzierbibliotheken
- Analyse der stabilen Isotope
- Korrektur, Beiträge zum Manuskript sowie Verbesserungsvorschläge

Eigener Anteil: 50 %

2. Studie:

Effect of X-ray irradiation on ancient DNA in subfossil bones - Guidelines for safe X-ray imaging; publiziert in *Scientific Reports* am 12.09.2016

Author Contributions: *J.K., P.T., M.B., A.I. and A.L.C. conceived and designed the general research. A.I., M.B. and J.K. further designed the genetic analyses while P.T. and A.L.C. designed and conducted the irradiation-experiments at the ESRF. P.T., A.L.C. and H.T. designed the conventional scanner dosimetry measurements that were performed by A.L.C. and H.T. at the MPI-EVA. M.B., F.L. and V.J.S. performed mtDNA-extraction and the preparation of sequencing libraries at the University of Tübingen. N.J.C., S.C.M., A.B., G.P., D.G.D., M.-A.J. and O.K. provided samples. B.V., K.H., P.T. and J.-J.H. facilitated the research and provided access to infrastructure. A.I. and A.H. performed bioinformatic analyses at the MPI-SHH. A.I., P.T. and J.K. wrote the manuscript and A.L.C. and K.B. contributed to it. All authors reviewed the manuscript.*

Eigene Beiträge:

- bioinformatische Prozessierung der DNA-Sequenzdaten
- Auswertung der qPCR Ergebnisse
- Auswertung der Experimente und statistische Erfassung der Ergebnisse in R
- Verfassen des Manuskripts

Beiträge von Co-Autoren:

- Bereitstellung von Proben
- Bereitstellung der Infrastruktur und Durchführung der Bestrahlungsexperimente am European Synchrotron Research Facility in Grenoble
- DNA-Extraktion und Vorbereitung der Sequenzierbibliotheken
- Korrektur, Beiträge zum Manuskript sowie Verbesserungsvorschläge

Eigener Anteil: 50 %

3. Studie:

Analysis of genomic DNA from medieval plague victims suggests effect of Yersinia pestis on human immunity genes; eingereicht bei Molecular Biology and Evolution (under review)

Author Contributions: *AI, AS, PJN, AH, KIB, SF, OK and JK designed the immunity capture. AI and JK collected modern saliva samples. AI, MAS, VJS and ER conducted lab work. AI, RB, AS and DIHZ performed HLA genotyping, and FMK conducted Danc and Fst analysis. JS and BKK performed pathogen screening. PJN and MKR performed KIR genotyping. JH and PJN performed data analyses. PJN, WP and GH performed simulations for natural selection. JS, US and AHS contributed the DKMS allele frequencies. RW and SA led the Excavation in Ellwangen. JW and RW provided the 16 th century Ellwangen samples. MSK and JW conducted anthropological analyses. AI, JH, PJN and JK wrote the manuscript. JK led the study.*

Eigene Beiträge:

- Literaturrecherche und Sammeln der Gensequenzen von Immungenen
- Design des Immunity Capture
- Testen des Immunity Capture an Kontrollproben
- Sammeln von Speichelproben
- DNA-Extraktion und Vorbereitung der Sequenzierbibliotheken aus Speichelproben und Knochenproben von Pestopfern aus dem 16. Jahrhundert
- bioinformatische Prozessierung der DNA-Sequenzdaten
- Bestimmung der Klasse I HLA Allele
- Vergleich der HLA-Allelfrequenzen
- Verfassen des Manuskripts

Beiträge von Co-Autoren:

- Bereitstellung von Knochenproben von Pestopfern und anthropologische Untersuchungen
- Beiträge zum Design des Immunity Capture
- Bereitstellung von DKMS HLA-Frequenzen
- Screening der DNA Daten nach Pestsequenzen
- Genotypisierung von HLA und KIR Genen

- Simulation von neutraler und natürlicher Selektion
- Korrektur, Beiträge zum Manuskript sowie Verbesserungsvorschläge

Eigener Anteil: 60 %

Hiermit bestätige ich, dass Alexander Immel die oben genannten Arbeiten geleistet hat.

Prof. Dr. Johannes Krause
(Erstbetreuer)

Jena, den

Danksagung

Zuerst möchte ich Prof. Johannes Krause danken für die Zeit, in der er mich betreut hat und dafür, dass er mir den Weg einer wissenschaftlichen Laufbahn beginnend in seiner Arbeitsgruppe für Paläo- und Archäogenetik an der Eberhard Karls Universität Tübingen sowie später am Max-Planck-Institut für Menschheitsgeschichte in Jena ermöglicht hat. Ich freue mich auf weitere Zusammenarbeit mit ihm.

Ich bin Prof. Martin S. Fischer sehr dankbar für die Bereitschaft meine Zweitbetreuung an der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität Jena zu übernehmen. Weiterhin danke ich Prof. Oliver Kohlbacher für die Drittbetreuung.

Ich danke herzlichst Prof. Ben Krause-Kyora und Prof. Almut Nebel für das Angebot einer Stelle als wissenschaftlicher Mitarbeiter in ihrer Arbeitsgruppe für alte DNA am Institut für Klinische Molekularbiologie an der Christian-Albrechts-Universität zu Kiel. Mein großer Dank gilt Dr. Alexander Herbig, der mich seit Beginn meiner Promotion in Sachen Bioinformatik betreut hat.

Ich danke meinen Kollegen Cosimo Posth, Alexander Peltzer, Rodrigo Barquera, Julian Susat, Katharina Fuchs, Maria Spyrou und Michal Feldman, die mir mit Rat und Tat zur Seite gestanden haben.

Weiterhin danke ich allen meinen Kollegen am Max-Planck-Institut für Menschheitsgeschichte in Jena und am Institut für Naturwissenschaftliche Archäologie in Tübingen, sowie meinen Kollegen am Institut für Klinische Molekularbiologie in Kiel für die schöne Zeit miteinander.

Besonderen Dank richte ich an Lilly und Friedrich Oettinger, sowie Heinrich Wedel, bei denen ich zu Beginn meiner Promotion gewohnt habe und die mir in meiner damaligen Situation sehr geholfen haben.

Meinen lieben Dank richte ich an meine Freundin Swetlana Wall für ihre langjährige Unterstützung seit Beginn meines Studiums.

Nicht zuletzt gilt mein ganz besonderer Dank meiner Schwester Katharina und meinem Schwager Andrej Henning, meinen Eltern Svetlana und Wladimir Immel und meiner Großmutter Katharina Wedel für ihre stetige Liebe und Unterstützung und dafür, dass sie die Rahmenbedingungen geschaffen und es mir möglich gemacht haben, meinen Interessen nachgehen zu können.

Tabellarischer Lebenslauf

Persönliche Daten

Name: Alexander Immel
Geburtsdatum: 22.03.1986
Geburtsort: Bischkek (Kyrgyzstan)
Staatsangehörigkeit: deutsch

Schulbildung

1993 – 1995: Grundschule in Kaliningrad
1995 – 1997: Grundschule in Detmold
1997 – 2006: Stadtgymnasium Detmold (Abitur)

Grundwehrdienst

2006 – 2007: Augustdorf, NRW

Studium und Beruf

ab 10.2007: Studium der Bioinformatik und Genomforschung an der Universität Bielefeld
22.10.2010: Abschluss Bachelor of Science
ab 10.2010: Studium der Genombasierten Systembiologie an der Universität Bielefeld
13.05.2013: Abschluss Master of Science
06.2013 - 05.2015: wiss. Angestellter am Institut für Naturwissenschaftliche Archäologie der Universität Tübingen
06.2015 - 11.2016: wiss. Angestellter am Max-Planck-Institut für Menschheitsgeschichte, Jena
ab 12.2016: wiss. Angestellter am Institut für Klinische Molekularbiologie, UKSH, Kiel

Alexander Immel