From Genes to Genomes: What is New in Ancient DNA?

Johannes Krause
Institut für Naturwissenschaftliche Archäologie
Rümelinstraße 23
D-72070 Tübingen
krause@eva.mpg.de

Abstract: *Despite being a rather young research field, the study of long dead and often extinct organisms' DNA, called ancient DNA research, has received great public attention since its beginnings and has been published in the highest profile journals. At the same time, many of the high-profile publications in this field have been shown to be erroneous, for example, alleged extinct dinosaur DNA (Woodward et al. 1994) that was later found to be artifacts derived from modern human DNA contamination (Zischler et al. 1995). In the last two decades, it has become clear that DNA survival is not unlimited and that, even under ideal conditions such as a cold and dry environment, DNA does not survive for more than a million years in fossil samples (Pääbo et al. 2004). Strict guidelines were established to ensure the authenticity of obtained results. This resulted in a growing number of high-quality studies. Nevertheless, ancient DNA research remained limited due to several reasons: first, the small amount of available material, second, the highly degraded and fragmented nature of the ancient DNA, and third, problems of contamination with contemporary DNA. Only recently, after new technologies and methodological approaches became available, is the field of ancient DNA experiencing a new era. Now studies once thought to be impossible are possible, such as large-scale population genetic studies (Shapiro et al. 2004; Debruyne et al. 2008), functional analysis of extinct genes (Römpler et al. 2006; Campbell et al. 2010), genetic studies of early modern humans (Krause et al. 2010a), as well as whole genome sequencing of extinct organisms (Miller et al. 2008; Green et al. 2010; Rasmussen et al. 2010). In this review some of the latest methodological developments in ancient DNA research will be discussed, in particular how new approaches allow us to overcome previous limitations while working with ancient hominin DNA.*
Keywords: *ancient DNA, Neandertal, human evolution, early modern humans, high throughput DNA sequencing*

Von Genen zu Genomen: Was ist neu an der alten DNA?

Zusammenfassung: Obwohl die alte DNA Forschung, die sich der Erforschung der DNA aus lange verstorbenen und häufig ausgestorbenen Organismen widmet, ein eher junges Forschungsfeld ist, stand sie von Anfang an im Fokus der Öffentlichkeit und wurde in angesehenen Fachzeitschriften publiziert. Gleichzeitig konnte gezeigt werden, dass zahlreiche dieser Beiträge fehlerhaft oder gar falsch waren, z.B. die potentielle DNA aus einem Dinosaurierknochen (Woodward et al. 1994), die sich als Kontamination mit DNA eines modernen Menschen herausstellte (Zischler et al. 1995), höchst wahrscheinlich von der Oberfläche des Knochens. In den letzten 20 Jahren zeigte sich, dass selbst unter idealen Bedingungen, wie konstant niedriger Temperatur und geringer Feuchtigkeit, DNA nicht mehr als eine Million Jahre überdauert (Pääbo et al. 2004). Um die Authentizität alter DNA zu gewährleisten, etablierten Wissenschaftler strikte Regeln, die zunehmend in unverfälschten und hoch qualitativen Studien mündeten. Dennoch war die Erforschung alter DNA limitiert, da nur geringe Mengen an DNA die Zeit überdauerten, diese zudem stark fragmentiert und chemisch verändert ist; zusätzlich stellt moderne DNA Kontamination ein großes Problem für die Analysen dar. Durch die kürzlich erfolgte Entwicklung neuer Technologien und Methoden erlebt die alte DNA Forschung eine neue Ära. Studien, die bis vor kurzem nicht durchführbar waren, können nun durchgeführt werden, z.B. groß angelegte Populationsstudien (Shapiro et al. 2004; Debruyne et al. 2008), funktionale Studien von Genen ausgestorbener Organismen (Römpler et al. 2006; Campbell et al. 2010), die Analyse kompletter mtDNAs früher moderner Menschen (Krause et al. 2010a) oder die Sequenzierung und Analyse kompletter Genome ausgestorbener Tiere und Urmenschen (Miller et al. 2008; Green et al. 2010; Rasmussen et al. 2010). Dieser Beitrag befasst sich mit den neuesten technologischen Entwicklungen in der alten DNA Forschung, die zu diesen spektakulären Ergebnissen geführt haben und im Besonderen damit, wie es diese neuen Methoden möglich machen, Beschränkungen bei der Untersuchung alter DNA aus menschlichen Überresten zu überwinden.
Schlagwörter: alte DNA, Neandertaler, menschliche Evolution, frühe moderne Menschen, Hochdurchsatz-DNA-Sequenzierung

Introduction

DNA sequence analyses from fossil remains generally involve several steps. First the DNA must be isolated by dissolving the fossil tissue. This is followed by the extraction of the DNA from the suspension, which results in a product usually referred to as ancient DNA extract. Since the amount of DNA in fossils is usually quite low, the entire DNA extract or parts of it are subsequently multiplied in a process called amplification. There are various ways of amplifying fossil DNA, which result in a variety of biases and limitations as will be discussed later. Only after amplification of the fossil DNA is it possible to determine the DNA sequence using various sequencing techniques. Before going into the technical details it is interesting to review the history of ancient DNA research and to observe how this research field evolved.

Ancient DNA research started in the mid 1980s by successful histological staining of DNA in soft tissue from Egyptian mummies (Pääbo 1984) as well as extracting DNA from museum samples of the extinct close relative of zebras called quagga *(Equus quagga)* (Higuchi et al. 1984). Subsequent transfer of the extracted DNA into bacterial genomes, often referred to as cloning, enabled researchers to amplify the ancient DNA by simply culturing the bacteria that divide their genome rapidly (Fig. 1a). DNA sequencing of the amplified DNA fragments showed that they contained a high degree of similarity to modern horse DNA and therefore likely represented DNA sequences of the extinct quagga (Higuchi et al. 1984). The same bacterial cloning approach was used to sequence the first ancient modern human DNA fragment from an Egyptian mummy (Pääbo 1985). However, it was nearly impossible to reproduce the ancient DNA sequencing results using bacterial cloning, as will be discussed later. Only with targeted methods such as the polymerase chain reaction (PCR) was it possible to easily replicate DNA sequencing results (Mullis et al. 1986). PCR makes use of constant regions of the DNA sequence to copy variable regions of the DNA sequence. It typically uses two short pieces of synthesized DNA, called primers (Fig. 1b), that serve as starting points for the copying of a region of DNA. In a few hours and several cycles of PCR, a single DNA molecule can be copied millions or even billions of times. These copied DNA fragments can subsequently be sequenced and analyzed. Using this approach it is possible to replicate DNA sequencing results such as the ancient quagga DNA and to make sure that the DNA sequences obtained do not contain errors (Pääbo and Wilson 1988). As a result a growing number of studies appeared in the early 1990s that used PCR to determine DNA sequences from extinct organisms, such as the marsupial wolf, the moa (a giant flightless bird from New Zealand), and Pleistocene mega fauna such as cave bear, ground sloth and mammoth (Thomas et al. 1989; Hagelberg et al. 1991; Cooper et al. 1992; Hoss et al. 1994). This helped to clarify the phylogenetic relationship of these organisms to closely related contemporary species. These studies focused mostly on one part of the genome, the mitochondrial DNA (mtDNA) that can be found in most body cells in hundreds of copies and is passed on maternally from mother to offspring. Besides showing high copy numbers, mtDNA displays a rather high rate of evolutionary change that is useful when comparing closely related species or populations.

Besides being a very powerful tool in the analysis of small amounts of DNA, PCR also turned out to be a tempting method to produce older and older DNA sequences going back in time to more than 200 million years ago. These sequences, retrieved from fossil

plant remains as well as DNA sequences from dinosaurs (Woodward et al. 1994) were later shown to be results of modern DNA contamination most likely on the surface of the fossil (Zischler et al. 1995).

This highlighted a problem with PCR, that in the absence or presence of only tiny amounts of DNA from the target organism (endogenous DNA) there is a high chance that small amounts of contemporary DNA contamination present on or in the sample (exogenous DNA) will be amplified and erroneously identified as endogenous DNA. As a consequence, already in the first years of ancient DNA research, criteria for authenticity were suggested to avoid misinterpretation of results (Pääbo 1989). Over the years researchers converged on a number of criteria that should be strictly followed when ancient DNA is analyzed (Cooper and Poinar 2000; Hofreiter et al. 2001; Pääbo et al. 2004). These criteria include independent replication of results within a laboratory and in the case of unexpected results even in a second laboratory. There should be a spatial separation of ancient DNA extraction and pre-PCR work with all post-PCR experiments. The ancient DNA extraction should be carried out in specific clean room facilities that use filtered air, UV radiation, sterile clothing and often positive air pressure. No contemporary DNA samples should be used in such facilities. As a result, at the end of the 1990s, ancient DNA analysis became a routine approach used mostly to study specimens from museum collections and well preserved Pleistocene and Holocene fauna.

First studies on ancient humans

What remains far from trivial is the analysis of hominin DNA from modern human fossils as well as fossils from extinct hominins such as Neandertals. Already the first ancient modern human DNA sequence was criticized as potential modern human contamination, since no differences can be expected in a DNA fragment from a 2,000 year old mummy and a contemporary person working in the lab or in the museum (Pääbo 1999). It is rather unlikely to have modern elephant DNA present on a mammoth sample, but it is almost impossible to exclude the presence of modern human DNA on a fossil sample from an archaeological excavation or on laboratory equipment that comes into contact with a sample (Pääbo et al. 2004; Serre et al. 2004). Studies on modern human remains should therefore only be carried out if a retrieved DNA sequence from the sample is different from all modern humans that potentially came into contact with the fossil and the laboratory equipment. This is hard to achieve for most ancient modern human samples, such as for the Cro-Magnons, the earliest modern humans in Europe, since their mtDNA sequence is expected to be similar or identical to the one found in contemporary modern humans who could have potentially contaminated the sample (Krause et al. 2010a).

For other hominins, such as Neandertals, the situation is more controllable. In their mtDNA they show several hundred differences to modern humans and fall outside modern human mtDNA variation. This was first shown by analyzing the most variable part of the mtDNA, the so called hyper variable region 1 (HVR1), from the Neandertal type specimen that was found in *Kleine Feldhofer Grotte* in 1856 in the Neander valley near Düsseldorf, Germany. The 370 base pairs (bp) long HVR1 mtDNA sequence retrieved from the right humerus of the eponymous Neandertal from *Kleine Feldhofer Grotte* was different from all known modern human mtDNA sequences (Krings et al. 1997). Using a molecular clock model that assumes a constant rate of mutations, it was

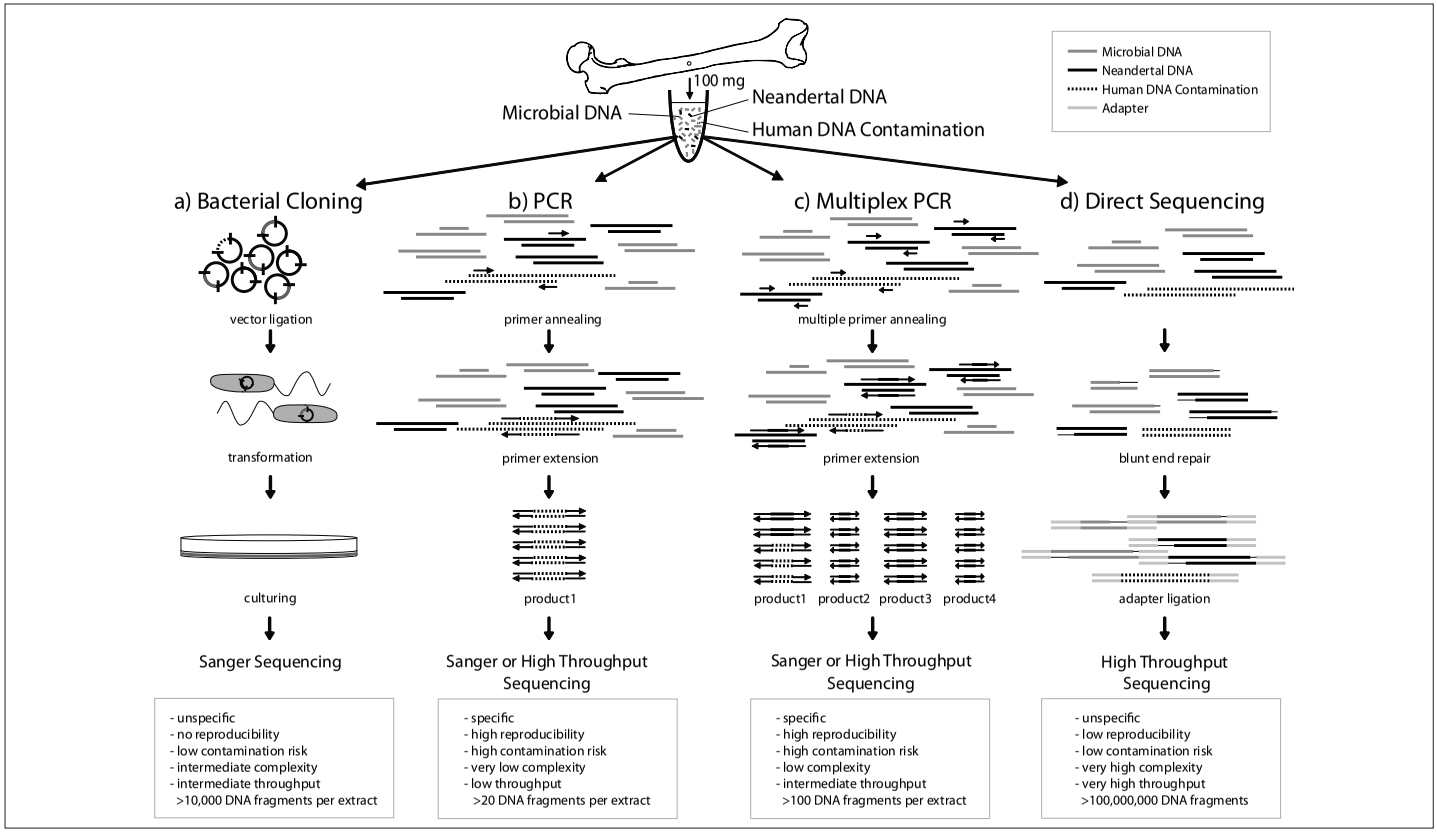

Fig. 1: *Diagram showing schematically four different approaches to obtain DNA sequences from fossil remains, here depicted as a Neandertal bone.* ***The DNA extracted from the bone serves as starting material for all four approaches a)-d). For bacterial cloning a), the ancient DNA fragments*** *are inserted into bacterial plasmids, transformed into bacteria, cultured as colonies on agar plates and subsequently sequenced using Sanger sequencing. For PCR and Multiplex PCR b) and c), ancient DNA is incubated with short oligonucleotides called primers, specific DNA fragments embraced by a pair of primers are subsequently copied in several cycles of PCR, doubling the product in each cycle. The PCR products can afterwards be sequenced using Sanger- or high throughput sequencing. For direct sequencing d), the ancient DNA first gets blunt end repaired then DNA adaptors are added to each end. The final product, called sequencing library, serves as product for various high throughput sequencing strategies and technologies (see also Fig. 3).*

possible to correlate time and number of mutations, thus it was calculated that Neandertals and modern humans shared their last common ancestor for their mtDNA around 600,000 years ago, suggesting that Neandertals were not ancestral to anatomically modern humans. Currently Neandertal-like mtDNA has been retrieved from 16 different Neandertals (Krings et al. 1997; Krings et al. 2000; Ovchinnikov et al. 2000; Schmitz et al. 2002; Serre et al. 2004; Beauval et al. 2005; Lalueza-Fox et al. 2005; Caramelli et al. 2006; Orlando et al. 2006; Krause et al. 2007b; Briggs et al. 2009a), confirming that Neandertals represent, on the mtDNA level, a genetically distinct form of human. It has also been shown that the Neandertal-like mtDNA was not present in Cro-Magnons from Europe, suggesting at most a low level of admixture between Neandertals and early modern Europeans (Serre et al. 2004). The fact that Neandertal mtDNA is genetically distinct can be used to identify morphologically ambiguous hominin bones as Neandertal. This led to the identification of Neandertal bones from Okladnikov cave in the Altai Mountains in Southern Siberia, suggesting a much wider range of Neandertals, reaching more than 2000 km further East than was previously assumed (Krause et al. 2007b).

Multiplex 2-step PCR

Despite the huge success of ancient DNA studies on faunal remains and extinct hominins such as Neandertals, many limitations remained. Although overlapping short pieces of DNA can be used to reconstruct longer DNA sequence stretches, this approach is limited by the small abundance of template molecules in DNA extracts, which limits the amount of DNA sequence that can be reconstructed. Fossils, especially hominid remains, are rare and highly valuable, therefore limiting the amount of material available for DNA studies. Thus, until five years ago, only relatively short mtDNA segments, no more than 300-600 bp in length, could be reconstructed from Pleistocene fossil remains. However, these short DNA sequences are often not sufficient to address biological questions such as phylogenetic relationships between closely related species (Cummings et al. 1995; Krause et al. 2006; Rohland et al. 2007).

An approach to retrieve longer ancient DNA sequences from small amounts of fossil material is the multiplex 2-step PCR for ancient DNA (Krause et al. 2006). In this method more than 40 ancient DNA fragments can be amplified together in a single PCR (Rohland et al. 2007). In the first step DNA extract is incubated with up to several hundred primers in a PCR, called multiplex PCR. This reaction serves as a pre-amplification of the DNA. The product is subsequently diluted and serves as a template for individual PCRs with just a single primer pair, in a singleplex PCR (Fig. 1c). Thus, instead of adding valuable DNA extract to hundreds of reactions only a low number of initial multiplex PCRs is needed. This approach was first used to reconstruct the complete mtDNA of the extinct mammoth using DNA extract from less than 200 mg of bone (Krause et al. 2006). The comparison of the mammoth mtDNA sequence to that of its closest living relatives, African and Asian elephants, resolved a long-standing controversy, identifying the mammoth's closest living relative as the Asian and not the African elephant. Other studies used this approach to reconstruct the mtDNA of the extinct Mastodon (Rohland et al. 2007) and to reconstruct complete mtDNAs from the extinct cave bear and short faced bear to resolve phylogenetic relationships within the bear family (Krause et al. 2008).

Going Nuclear

Although ancient mtDNA is quite useful in identifying and reconstructing phylogenetic relationships between extinct and extant species, it represents only a single part of the genome (genetic locus) that gets passed on from mother to offspring. However a single genetic locus does not necessarily reflect the phylogenetic relationship of the entire organism. It is in principle possible that different genetic loci such as parts of the DNA from the cell nucleus, called nuclear DNA, and the mtDNA show different phylogenetic relationships, due to a process called incomplete lineage sorting. This could be, for example, shown by comparing almost a hundred genetic loci of the nuclear genome from modern humans, chimpanzees and gorillas and reconstructing for each region the phylogenetic relationship between the three primates. For about 2/3 of the genetic loci modern humans show a closer relationship to chimpanzees, whereas for 1/3 of the loci either gorilla is closer to chimpanzee or gorilla is closer to modern humans. Thus, only analyzing a larger number of genetic loci of nuclear DNA allows us to determine the phylogenetic relationship of a species (Chen and Li 2001). Furthermore, nuclear DNA can give information about the sex of an individual as well as genes related to phenotypic properties.

Unfortunately, as mentioned above, nuclear ancient DNA is often not well preserved and is more difficult to analyze due to a low copy number compared to mtDNA. To overcome the problem of low copy number the first targeted studies on nuclear ancient DNA were done by PCR, analyzing short fragments of genes that occur in many copies in the genome, such as ribosomal RNA genes from the extinct ground sloth and mammoth (Greenwood et al. 1999). To study single copy nuclear genes the multiplex 2-step PCR proved to be useful since lower amounts of fossil extract were needed. Nevertheless, it took substantial amounts of bone extract to analyze even very short fragments of nuclear genes such as the MC1R gene in mammoth (Römpler et al. 2006) and the speech and language gene FOXP2 in Neandertals (Krause et al. 2007a). In the latter study Neandertals were found to carry the same version of the FOXP2 gene that modern humans do, which is different from the FOXP2 version found in all other mammals. This suggests that Neandertals may have been capable of language. In this study it was also found that the analyzed bone fragment came from a male Neandertal that carried a Y-chromosome sequence different from contemporary modern humans. The analysis of the MC1R gene in mammoth also marks the first functional ancient DNA study. It was shown that within the mammoth population some individuals carried a version of the MC1R gene that shows a lower activity level of the protein for which it codes. This variant of the MC1R gene would cause a fair-skinned and bright hair phenotype in most mammals, suggesting phenotypic variation of coat color within mammoths (Römpler et al. 2006), similar results could be obtained for the MC1R gene of some Neandertals (Lalueza-Fox et al. 2007).

In spite of the success of Multiplex PCR to study single copy nuclear genes, it remained impossible to sequence larger parts of genes and genomes from extinct animals. Even if up to 100 fragments can be analyzed together in a 2-step Multiplex PCR, to simply amplify one chromosome of a Neandertal with PCR one would need more than 3 kg of bone to be extracted and the costs would be tremendous.

An alternative to PCR was the approach already used in the first ancient DNA studies in the 1980s: a shot gun approach using bacterial cloning. Shot gun here refers to an approach where random DNA sequences from the organism are determined. For this the ancient DNA fragments extracted from the bone are inserted into bacterial plasmids. These plasmids are subsequently transformed into bacteria, are then cultured over night and are traditionally sequenced with a sequencing method called Sanger sequencing (Fig 1a). By screening through hundreds of such plasmids, sequences derived from ancient DNA that carry endogenous DNA of the target organism can be identified, e.g. horse-like DNA sequences that represent endogenous quagga DNA (Higuchi et al. 1984). In this approach no PCR is needed, the bacteria that grow overnight will divide themselves millions of times, thereby amplifying the inserted ancient DNA fragments. The biggest advantage of such a shot gun approach is that large amounts of ancient DNA fragments are transformed into a bacterial clone library. These clones can be repeatedly cultured and therefore allow a sort of immortalization of the fossil DNA. This classic approach was therefore used again in 2005 in an automated high throughput fashion using robots to screen more than 14,000 bacterial clone sequences from a cave bear extract as well as circa 30,000 clone sequences from a Neandertal DNA extract in 2006 (Noonan et al. 2005, 2006). Doing so, 389 fragments from the cave bear were identified that showed similar DNA sequences to nuclear dog DNA, the only closely related genome sequence that was available at that time. The authors concluded these sequences should represent nuclear cave bear DNA. They also concluded that only 1-6% of all DNA fragments from the cave bear bones are bear-like, whereas the vast majority of DNA sequences show no similarity to known mammal DNA but most likely represent DNA from microorganisms that thrived in the bones after the death of that organism (Noonan et al. 2005). Similar results were obtained for the Neandertals where only 3% of all DNA sequences inserted into the bacterial plasmids were human like, whereas the vast majority of sequences were most likely microbial DNA (Green et al. 2006; Noonan et al. 2006). These results suggest that it is rather difficult to use shot gun sequencing from bacterial clones to sequence specific regions of a genome. Large numbers of bacterial clones would have to be screened to observe every position of an extinct genome just once or a certain position of a gene several times. If only 1-5% of the inserts in the bacterial plasmids with an average length of 50 bp are endogenous to the organism studied and the rest represent bacterial DNA, one would need to sequence more than a billion clone sequences to find the same nuclear DNA fragment at least twice. Using bacterial cloning in combination with Sanger sequencing to obtain an entire ancient genome would therefore need at least a hundred times more resources than were available for the modern human genome project, which took 13 years for completion and cost several hundred million dollars.

High throughput sequencing of Ancient human DNA

Regardless of shot gun sequencing being a tedious endeavor, sequencing several thousand bp of nuclear DNA from an extinct organism marks a drastic change in the ancient DNA field. This change was additionally stimulated by the development and application of new sequencing technologies, a result of the ongoing efforts to sequence whole genomes of living organisms. The new technologies, called ‘next generation sequencers’ due to their radically different technique, were developed to rapidly produce extremely

large amounts of DNA sequences. Using this approach on ancient DNA, no bacterial cloning or targeted PCR amplification is necessary, instead all DNA molecules from an ancient DNA extract are turned into what is called a 'sequencing library' by adding artificial DNA molecules, so called adapters, to both sites of each individual ancient DNA molecule (Fig. 1d). In principle all ancient DNA library molecules can be subsequently sequenced.

Until now only two of the 'next generation sequencing technologies', also called high throughput sequencers, are used to sequence ancient DNA: the Roche/454 platform (Margulies et al. 2005) and the Illumina/Solexa platform (Bentley et al. 2008). Both technologies have similarities in that they use DNA sequencing libraries prepared from ancient DNA, to generate millions of DNA sequences in a single run of the machine. The first high throughput sequencer used for ancient DNA was the Roche/454 platform, in 2006. More than 13 million bp of mammoth nuclear DNA were produced on this platform as well as more than 1 million bp of Neandertal DNA (Green et al. 2006; Poinar et al. 2006).

Even though these represent two ground-breaking studies that eventually resulted in a draft version of the mammoth genome (Miller et al. 2008) and the genome of the first extinct human form, the Neandertal (Green et al. 2010), both studies also highlight problems of high throughput sequencing of ancient DNA. Depending on the technology used, several million DNA sequences are produced in a high throughput sequencer. These sequences present a mixture of DNA from the target organism, from microbes that once lived in the bone as well as exogenous contamination. To identify and sort the fragments, each DNA sequence is compared to known DNA sequences from publically available databases and genome sequences. Only if the ancient DNA fragments show a high similarity to a database sequence can they be identified by computer programs designed to detect similarities between DNA sequences, called alignment programs. They produce data files called alignments that present the ancient DNA sequence fragments aligned to their reference sequence from a database (Fig. 2). One problem in high throughput studies is that the genome or database sequences used to identify the ancient DNA sequences have to be sufficiently similar and complete (Prüfer et al. 2010). In the case of the cave bear the closest related species where a complete genome would be available is the dog. It was shown that using two distantly related genomes like dog and bear, which are separated by more than 100 million years of evolutionary time, can lead to massive amounts of DNA sequences that can not be identified from the high throughput DNA sequence data, due to evolutionary changes in the DNA sequence. As a result more than 80% of all endogenous cave bear DNA sequences would not be identified correctly (Prüfer et al. 2010). Additionally the fragments that can be identified are regions of the genome that evolve slowly, therefore biasing all subsequent analysis such as in measuring divergence time. To obtain a larger part of an extinct genome with high throughput sequencing, it is therefore absolutely necessary to have a closely related reference DNA sequence. Another problem are incomplete reference sequences, e.g. only a partial sequence of the African elephants genome is available that can be used to identify mammoth DNA. It is therefore not possible to identify and align all mammoth genome fragments that are, at this point, not sequenced from the modern elephants (Miller et al. 2008).

This is less of a problem when sequencing ancient human DNA, such as the Neandertal genome, since several complete primate genome sequences like human, chimpanzee

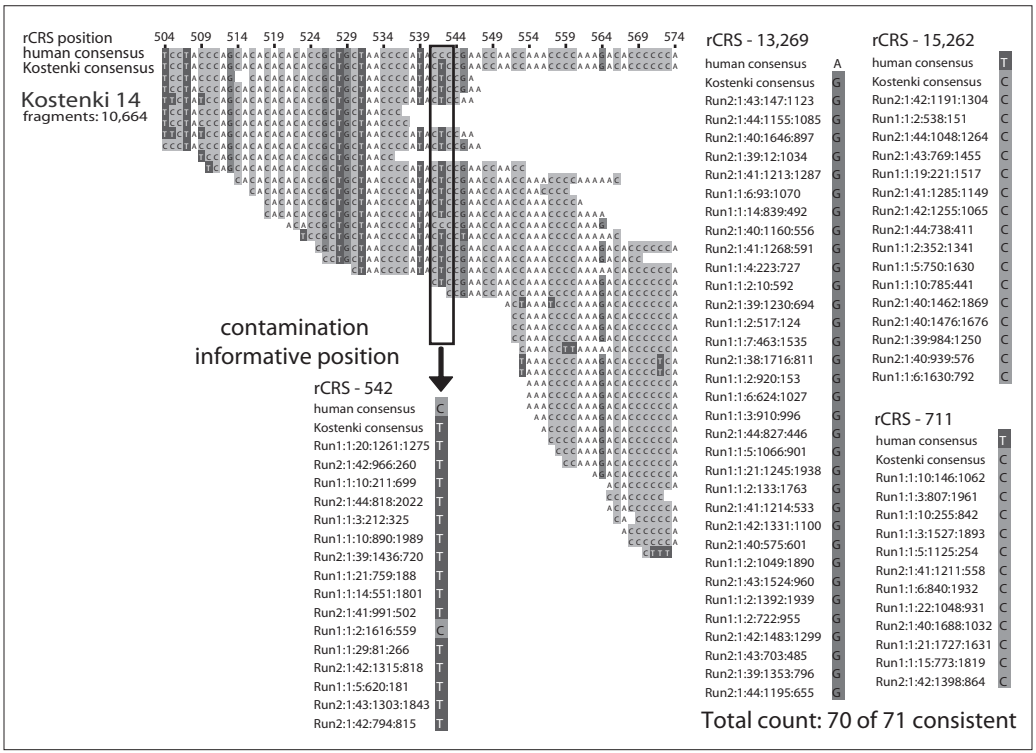

Fig. 2: *Section 504 – 574 of a sequence alignment for reconstructing the complete mtDNA genome (total 16,570 positions) of an early modern human from the Kostenki site, Russia. The positions are based on the revised Cambridge reference sequence (rCRS). The first line of the alignment shows a consensus sequence of 311 modern human mtDNAs. The second line shows the consensus sequence for 10,664 mtDNA fragments retrieved from the Kostenki early modern human bone. To get an estimate of contamination with modern human DNA, positions were identified where more than 99% of 311 modern human mtDNAs are different from the Kostenki consensus sequence. All fragments that overlap such a position (here 542) and are different from the Kostenki consensus base, are likely to be modern human contamination. In total there are four positions of that kind in the Kostenki mtDNA (542; 711; 13,269; 15,262) and 71 mtDNA fragments that overlap these positions. Only one fragment is inconsistent, suggesting a very low level of modern human contamination.*

and orangutan are available for comparison and identification. Nevertheless, the identification of contemporary modern human contamination, present in high throughput sequencing data from Neandertals, is much more difficult. Neandertals and modern humans are so closely related that one expects on average to see only one position in 500 bp to be different, therefore only one in ten Neandertal fragments would contain a different DNA sequence to a modern human. Furthermore, the Neandertal nuclear genome falls within the diversity of modern humans, hence for some parts of a Neandertal's genome he can be more closely related to a modern human than to another Neandertal individual (Pääbo 1999), making it even harder to identify whether a particular fragment is a modern human contaminant. However, it is possible to measure the amount of modern human contamination in a Neandertal extract at diagnostic positions of the nuclear genome or mtDNA where almost all modern humans and Neandertals are

found to be different. For the mtDNA this can be done by PCR amplification of diagnostic positions where all modern humans are different from Neandertals (Green et al. 2006). Subsequently the amount of modern human mtDNA contamination in the extract can be determined. Assuming that modern human contamination also contains mtDNA, it is possible to extrapolate from the mtDNA contamination level the level of contamination in regions of the genome in which no fixed differences between Neandertals and contemporary modern humans occur (Green et al. 2009).

However, analysis of contamination in DNA extracts cannot show contamination in subsequent laboratory steps. In particular, sequencing libraries from Neandertal DNA extracts contain very little Neandertal DNA, therefore contamination with small amounts of high copy number modern human DNA libraries will greatly affect the contamination level. The first Neandertal high throughput study most likely suffered from such contamination introduced from another modern human sequencing library (Wall and Kim 2007; Green et al. 2008, 2009). This can be minimized by adding project specific sequencing adapters in the clean room facility to the end of each ancient sequencing library molecule (Briggs et al. 2007). This allows detection of all contamination from other sequencing libraries outside the clean room.

An alternative to looking at mtDNA contamination in the extract using PCR is detection of mtDNA contamination directly in the high throughput sequencing data. This is possible not only with mtDNA but also with nuclear DNA from positions that were previously identified to be informative or entire regions that are not present in the target organism, e.g. the Y-chromosome if the studied specimen is female. All fragments that can be identified to be derived from such a region should represent modern human contamination and allow nuclear contamination estimates (Green et al. 2009, 2010). It was furthermore shown that contamination estimates retrieved directly from high throughput sequencing data show lower contamination counts than obtained with PCR contamination assays (Krause et al. 2010a). It was speculated that this discrepancy may be explained by the nature of contamination being less fragmented and chemically modified and therefore representing a better template for PCR amplification, but not to the same extent during DNA library preparation. This suggests that PCR based analyses of ancient human DNA are much more effected by modern human contamination than methods based on DNA sequencing libraries (Krause et al. 2010a).

Another problem that was highlighted by the high throughput sequencing of ancient DNA was the varying amount of total DNA and the highly varying amount of microbial DNA present in a particular sample or even sub-sample. Whereas well preserved permafrost samples such as mammoth bones and hair were found to have high copy numbers of DNA as well as less than 10% of microbial DNA (Gilbert et al. 2007; Miller et al. 2008; Rasmussen et al. 2010), non–permafrost samples such as Neandertal bones were found to have more than 96% microbial DNA (Green et al. 2006, 2010), in some cases such as the Spanish Neandertals from the Sidrón site more than 99.5% of the DNA come from microbes (see also Fig. 3c). This greatly increases the amount of sequencing to reconstruct large parts of a Neandertal genome. To get to a first version of the Neandertal genome it was therefore necessary to sequence almost 1.5 billion DNA fragments (Green et al. 2010) whereas only 30 million fragments were sequenced for the draft version of the mammoth genome (Miller et al. 2008).

Targeted enrichment of Ancient human DNA

Shot gun sequencing of ancient human DNA to retrieve a certain gene or genomic region of interest remains rather elaborate and costly due to the non-targeted nature of the approach and the high amounts of non-target-organism-DNA present in the fossil. The latest development of high throughput sequencing ancient DNA was therefore a targeted enrichment of certain regions of the genome from ancient DNA sequencing libraries. First, these methods make use of the fact that every molecule in an ancient DNA sequencing library has the same adapter. It is therefore possible to do whole library amplification with PCR, using primers that sit directly on the adapter (Fig. 3). Therefore thousands of copies can be made from every single molecule present in a fossil DNA extract that was turned into a DNA sequencing library. This reduces problems with small amounts of fossil DNA and allows the immortalization of the DNA present in the ancient DNA extract. Secondly, to enrich for certain regions of the genome one needs some sort of fishing device to capture the library molecules of interest. There are several targeted enrichment approaches for ancient DNA available such as array hybridization capture (Burbano et al. 2010) or primer extension capture (PEC) (Briggs et al. 2009b), the latter is similar to PCR but uses only a single primer to bind and fish out target molecules from the ancient DNA sequencing library (Fig. 3c). The fished-out molecules are subsequently sequenced on a high throughput sequencer. The PEC approach was first presented to retrieve complete mtDNAs from Neandertal DNA sequencing libraries. Using the PEC approach the amount of mtDNA fragments in the sequencing library was up to 80,000 times higher after PEC than before (Briggs et al. 2009b). Complete mtDNA sequences from five Neandertals could be retrieved in a single sequencing run using PEC, whereas 147 non-targeted shot gun sequencing runs were needed to complete the first Neandertal mtDNA sequence (Green et al. 2008). PEC also turned out to be very useful when measuring contamination in sequencing libraries by capturing mtDNA fragments from a Neandertal DNA library that fall into regions where modern humans and Neandertals carry diagnostic differences (Green et al. 2010; Krause et al. 2010a). This results in a much more comparable contamination estimate to direct shot gun sequencing than to the estimate obtained with direct PCR of the DNA extract (Krause et al. 2010a). Another inherent advantage of the PEC approach compared to other targeted approaches such as PCR is that a synthesized PEC primer is used as a capture device and will not be incorporated into the molecule that gets later on sequenced like in PCR (Fig. 1b, 1c). Therefore the entire original DNA fragment can be sequenced (Fig. 1d). Furthermore only one primer is needed to fish out a library molecule, not two like in PCR. Thus if one designs primers in a dense tiling along the DNA sequence of interest it is possible to capture almost all mtDNA fragments from the sequencing library even though they might carry differences to some of the designed primer sequences. Primers can therefore be designed based on a sequence of a related organism that is rather divergent compared to the target DNA that should be captured (Krause et al. 2010b).

Despite being a very powerful approach to sequence longer stretches of DNA such as mtDNA genomes from ancient human DNA, or to measure the amount of contamination in ancient Neandertal sequencing libraries, PEC is still limited by the number of primers that can be used and, therefore, the size of the region that can be studied. An alternative to PEC is a targeted capture approach using micro arrays; these are small glass slides that carry millions of DNA sequences attached to their surface (Fig. 3a). They allow tens

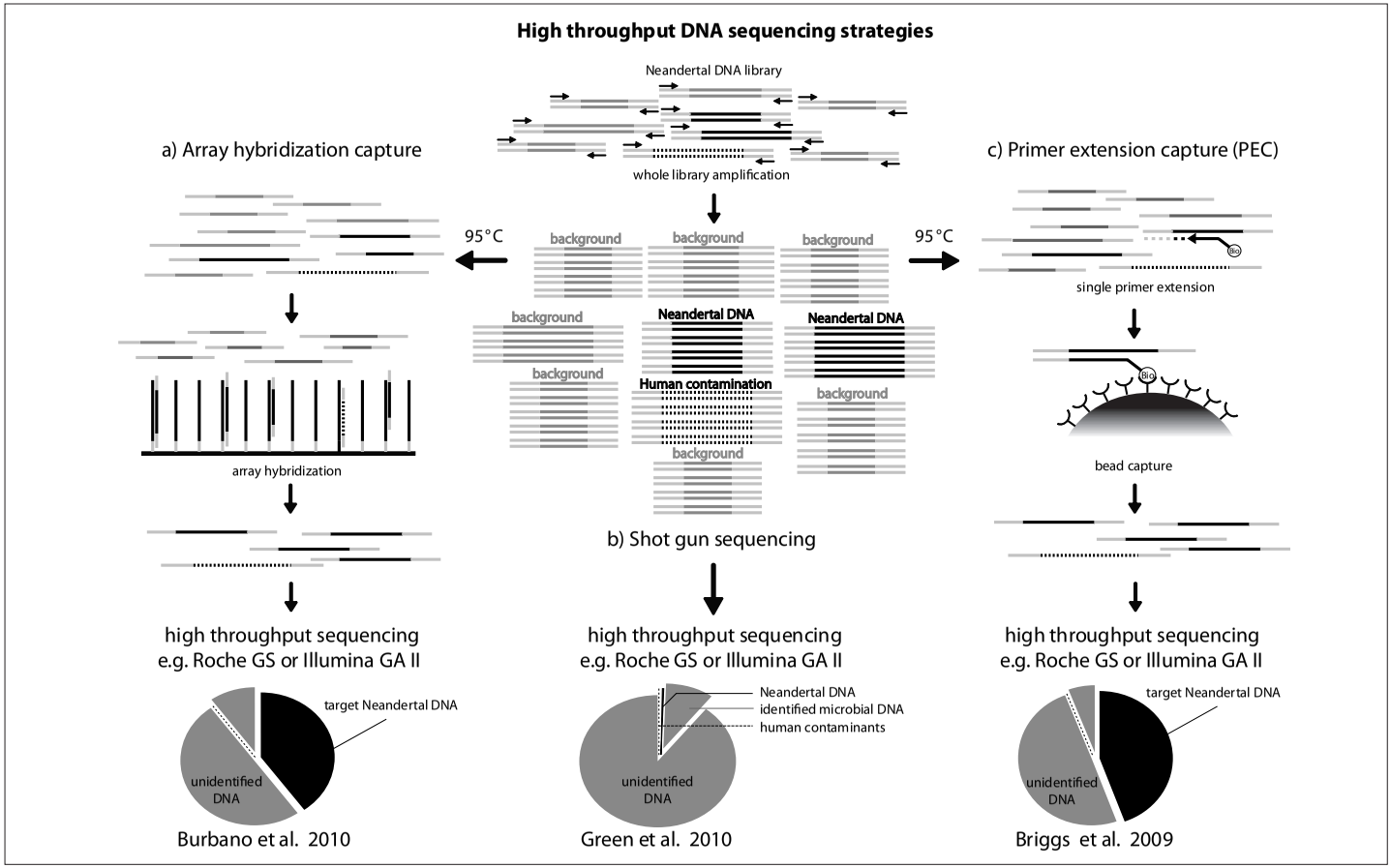

Fig. 3: *Schematic overview of high throughput sequencing strategies for ancient DNA from a Neandertal bone. As illustrated in Fig. 1d, first a DNA sequencing library is made from the ancient DNA extract. All ancient DNA molecules in the library will be first amplified using the adapters as priming sites in a PCR. Aliquots that contain copies of all original ancient DNA molecules can be directly sequenced on a high throughput sequencer b) or used in a targeted enrichment using array- a) or primer extension capture c). The pie charts illustrate the percentage of Neandertal DNA in direct shot gun sequencing b) array a) and PEC enrichment c).*

of thousands of genetic loci to be captured and sequenced in parallel from an ancient DNA sequencing library (Burbano et al. 2010). Such an approach was used on a Spanish Neandertal to capture all DNA positions that have caused amino acids to change in modern humans since we diverged from our ancestor with the chimpanzee. Through this method it was shown that there are less than 150 amino acids different between all modern humans and a Spanish Neandertal, whereas more than 14,000 amino acid differences emerged on the human lineage after we diverged from the chimpanzee (Burbano et al. 2010).

Differentiating contamination from endogenous DNA

High throughput sequencing can also help to overcome the contamination problem in ancient DNA studies on modern human remains. As mentioned above, it is rather difficult to exclude all possible sources of modern human contamination that could flaw studies on ancient modern human samples. The authenticity of the obtained results was therefore often questioned (Handt et al. 1994; Hofreiter et al. 2001; Pääbo et al. 2004; Krause et al. 2010a, b). Only well-preserved samples from the permafrost such as the

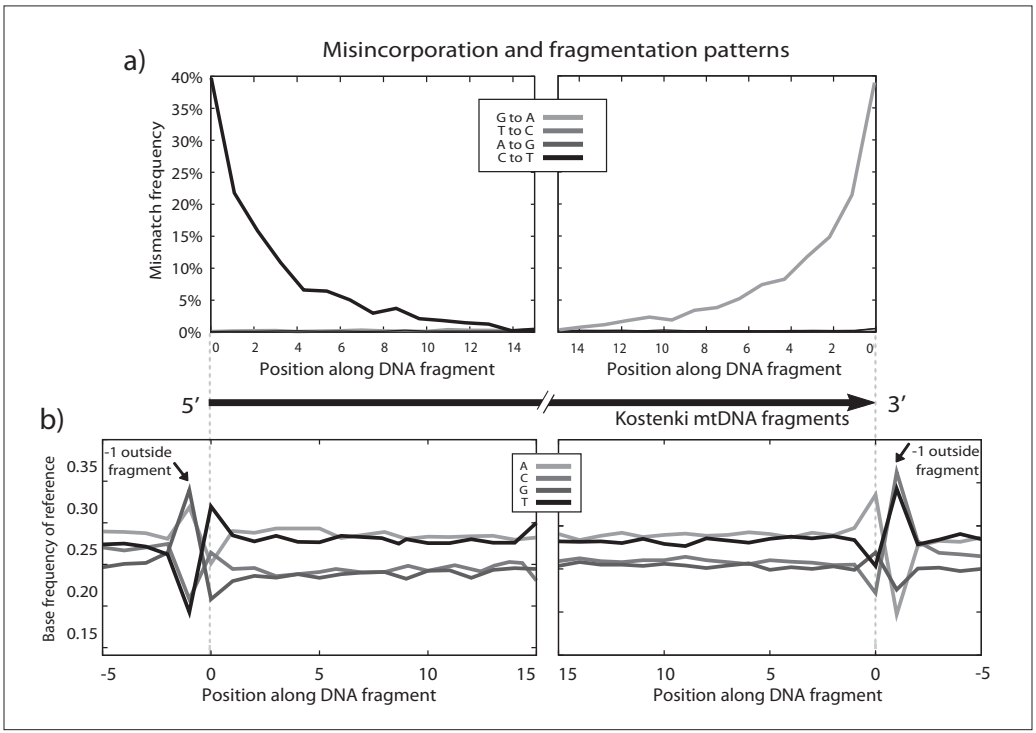

Fig. 4: *Spatial distribution of DNA degradation patterns typical for ancient DNA for mtDNA fragments from the Kostenki early modern human. a) DNA mismatches to a reference sequence for all ancient mtDNA fragments, more than 40% of Cs are seen as Ts at the 5' end of the mtDNA fragments and more than 40% Gs are seen as As at the 3' end. b) base frequency of the reference sequence showing that one base pair upstream of the 5' end of the mtDNA fragments, purines (A and G) are in high frequency and one base pair downstream of the 3' end pyrimidines (C and T) are in high frequency.*

ice man (Handt et al. 1994; Ermini et al. 2008) or Eskimos from Greenland (Gilbert et al. 2008b; Rasmussen et al. 2010) have enough DNA present to minimize the risk of modern human contamination influencing sequencing results. In some cases it is also possible to work on populations that are genetically distinct from more likely potential contamination, such as native Americans studied by a group of scientists with non-native American ancestry (Kuch et al. 2007; Gilbert et al. 2008a). Nevertheless this limits research to exceptional cases and circumvents the usage of thousands of modern human remains. It would therefore be quite useful to have other properties of the DNA that would allow distinguishing ancient endogenous DNA from modern human exogenous contamination.

A typical feature of ancient DNA is fragmentation as well as chemical modification. Several studies have suggested that modern contamination tends to have a longer average fragment length than the ancient endogenous DNA (Wall and Kim 2007; Krause et al. 2010a). This feature might therefore represent a useful tool to show that a fragment is rather old. The only problem is that variation in fragment length distribution in non-permafrost DNA samples, from Neandertals with low amounts of modern human contamination, varies almost three fold, from 35 bp to more than 85 bp average fragment length (Briggs et al. 2009b). Furthermore well preserved Pleistocene cave bear samples preserve endogenous DNA up to 500 bp in length (Krause et al. 2008). But even a few-year-old DNA sample from dried muscle tissue was shown to be in average less than 500 bp long (Pääbo 1989). It is therefore hard to find a clear cut-off for contamination given a range of length variation in contamination and endogenous DNA.

Chemical modifications however seem to be more promising. The most common type of chemical modification that can be detected in ancient DNA sequences is the loss of a methyl group in cytosine bases called deamination. This results in a nucleotide misincorporation causing a cytosine being misidentified as a thymine and guanine being misidentified as an adenine on the complementary strand (Hofreiter et al. 2001). As mentioned before, high throughput sequencing data from DNA sequencing libraries allows the entire DNA molecule to be sequenced, which is not possible with PCR (see also Fig. 1b, c, and d). Thus modifications of the DNA that result in nucleotide misincorporations can be observed over the entire length of the ancient molecule. Thus it could be shown that in Neandertal DNA deamination occurs 20 times more frequent at the end of the ancient DNA fragments than in the middle (Briggs et al. 2007). It was confirmed in other organisms, such as mammoth, that up to 40% of cytosines at each end of Pleistocene ancient DNA fragments are deaminated and read as thymine, and guanine as adenine respectively on the complementary strand. This results in a characteristic nucleotide misincorporation pattern typical for ancient DNA (Fig. 4a). In a direct comparison of these patterns in endogenous Neandertal DNA fragments and modern human DNA contaminants present in the same Neandertal DNA extract, it could be shown that the typical pattern of deamination is almost completely absent or much lower in the modern contaminants compared to the endogenous Neandertal fragments (Krause et al. 2010a). Chemical modifications therefore potentially allow distinguishing contamination from endogenous DNA. It could be furthermore shown that the ancient DNA fragmentation is often a result of losing a purine base (adenine or guanine) causing DNA strands to break. This could be observed by analyzing which type of bases, purines or pyrimidines, can be found just one position outside of the alignment of the ancient DNA fragment and the modern human reference sequence. Depending on the various types of samples analyzed

one finds that up to 80% of the bases, one position outside the ancient DNA fragment, are purines (Fig. 4b). Hence DNA breakage happens more often after a purine than a pyrimidine, most likely caused by a process called depurination. Moreover modern human

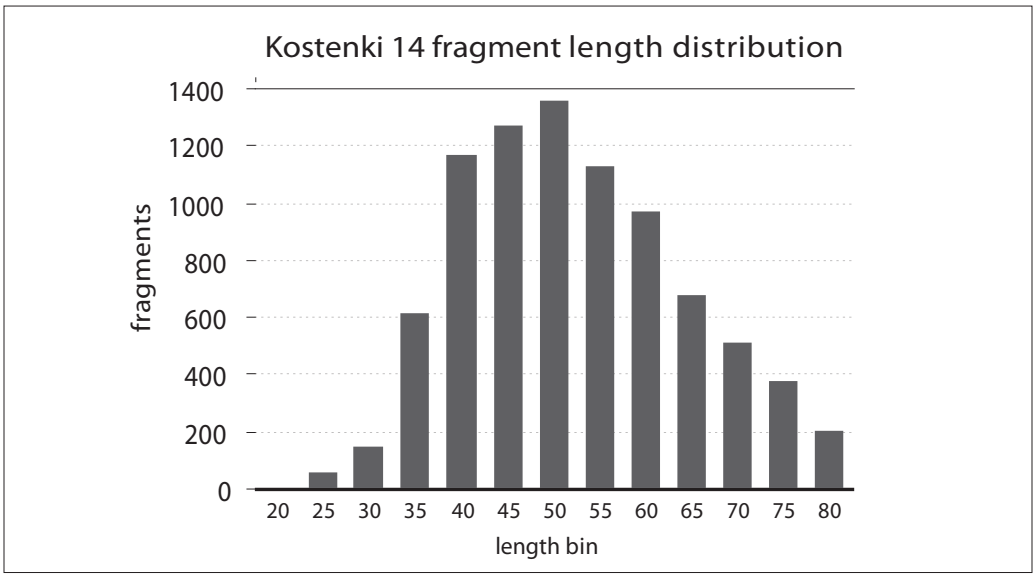

Fig. 5: *Distribution of fragment length for mtDNA fragments from the Kostenki early modern human. The average length is 54 base pairs.*

DNA contamination in a Neandertal does not show this fragmentation pattern caused by depurination. Thus, there are several DNA degradation related properties that can be used to distinguish ancient DNA and modern contamination. However given that not every molecule starts after a purine and carries a cytosine at the first position and given that only 40% of cytosines are chemically modified, it is rather difficult to identify every modern human contaminant based on the presence of the degradation signal. Only if it can be shown that the ancient DNA comes from a single biological source and that the pattern typical of ancient DNA is present in that DNA is it highly likely to represent endogenous ancient modern human DNA (Krause et al. 2010a).

To show whether DNA comes from a single human individual, one can use single DNA positions that are in low frequency or absent in other individuals from a population, such as mtDNA positions that are in very low frequency or absent in modern humans. The combination of ancient DNA properties and identification of a single biological source to study early modern human DNA was first used on a Cro-Magnon sample from the Kostenki site in Western Russia (Krause et al. 2010a). The DNA extract from this early modern European was turned into a sequencing library, enriched with PEC for mtDNA and high throughput sequenced on the Illumina high throughput sequencer, resulting in more than 10,000 human mtDNA fragments. It could be shown that the human mtDNA fragments from this fossil mostly derived from a single modern human (Fig. 2) and that the typical ancient DNA nucleotide misincorporation patterns were present (Fig 4a). Furthermore it was argued that the retrieved DNA fragments were rather short

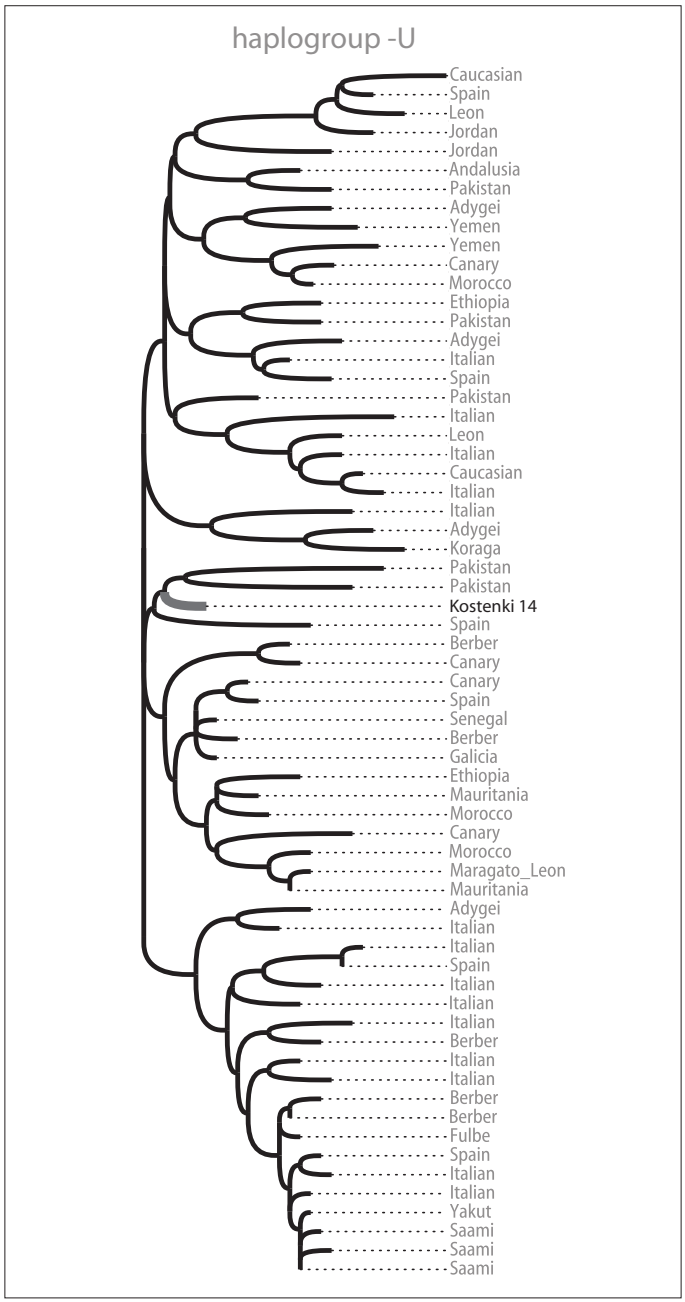

Fig. 6: *Phylogenetic tree of 63 contemporary modern human mtDNAs from a worldwide dataset of 311 sequences as well as the Kostenki mtDNA sequence. The mtDNA sequences all fall into the mtDNA haplogroup U. It can be observed that the Kostenki mtDNA shows the shortest branch length.*

in average length (Fig. 5) and fragmented in a way typical for ancient DNA (Fig. 4b), supporting the authenticity of early modern human DNA. It could be moreover shown that in a phylogenetic tree, comparing similar mtDNAs from different European and Asian populations, the Kostenki mtDNA displays a rather short branch length (Fig. 6). Applying a molecular clock approach, the age of the Kostenki mtDNA was determined with a best fit age of ca. 32,000 years and a minimum age of ca. 10,000 years, giving additional evidence for the authenticity of this early modern human DNA sequence (Krause et al. 2010a) and presenting the first DNA based molecular dating on a modern human sample.

This approach therefore seems to offer a chance to study ancient DNA from modern human remains that are up to 30,000 years old, if the fossil DNA shows no or very little DNA contamination, and if the degradation pattern typical for ancient DNA is present. It remains to be seen if degradation patterns present on ancient DNA from the Pleistocene are also present in historical samples less than 1000 years old. It is furthermore important to identify the rate of the degradation patterns that arise, to show if contamination can also present these patterns, in case the contamination is itself rather old, e.g. from a historical excavation.

Perspectives on ancient human DNA studies

Despite the vast amount of genetic information from extant animals that have become rapidly available over the last years, there are still a rather limited number of complete genome sequences from vertebrates available. Ancient genome projects on animals, such as the woolly mammoth, therefore suffer from the lack of available modern genome sequence data that can be used as a scaffold to assemble an extinct genome. Despite great efforts to reconstruct the mammoth genome it is hard to even find out the size of its genome (Miller et al. 2008).

This is much easier for humans where more than 100 modern human genome sequences are already available, as well as those of close related primates, and it will increase rapidly in the near future with the ongoing efforts of the 1000 genomes project. But even having a high quality genome sequence such as the modern human genome, it is nearly impossible to use it to put all ancient DNA fragments back together to reconstruct a full ancient human genome. Even for specimens where DNA is well preserved and every position of the genome is sequenced on average 20 times, as with the recently sequenced genome of an ancient Eskimo, it is still not possible to put together more than 85% of the genome (Rasmussen et al. 2010). Complete ancient genome sequences will therefore remain impossible to achieve. Nevertheless sequencing ancient human genomes offers a great chance to study interactions between extinct and extant human populations, to detect signals of gene flow as observed for Neandertals and modern humans from outside Africa (Green et al. 2010). Furthermore, genome sequences of extinct hominins such as Neandertals or the recently described hominin from Denisova cave in Siberia (Krause et al. 2010b), give insights into the recent evolutionary history of modern humans. They provide information about which regions of the modern human genome have changed after divergence from the ancient human forms, providing information about what genes and regions of the genome might be most relevant in recent modern human evolution.

However, contemporary modern human genomic data provides enough information about human diversity, hence little if anything can be learned from an ancient modern human genome sequence about modern human evolution. An early modern human genome from Europe is genetically much closer to a contemporary European than to contemporary Africans. An ancient modern human genome provides therefore little new genetic information (Shapiro and Hofreiter 2010). Nevertheless there are interesting population genetic questions that can be addressed by studying ancient modern human genomes, such as the cultural transition vs. replacement hypothesis during the Neolithic expansion (Bramanti et al. 2009) in Europe or the settlement and relationship of paleo-populations in the Americas and contemporary native American populations (Gilbert et al. 2008a; Rasmussen et al. 2010). Despite being a major achievement, the sequencing of the ancient Eskimo genome highlighted that less than 100,000 positions of the genome were needed to provide insights into the population history of its Eskimo population called Saqqaq. Only a small fraction of positions in the Eskimo genome were not seen in other human genomes previously. This is in sharp contrast to the recently sequenced modern human genome from Bishop Desmond Tutu that revealed almost a million positions that were so far unknown to vary within humans (Schuster et al. 2010). Therefore ancient modern human genome sequences will in most cases not give much new information about modern human diversity. The phenotypic information that could be referred from the Eskimo genome was furthermore rather limited and not unexpected for a Native American, such as inferred brown hair and eye color and dry ear wax. This is mostly due to the fact that even for a modern human genome sequence the amount of phenotype-genotype correlations is still rather limited. One can therefore conclude that whole genome sequencing of ancient modern humans is needless for obtaining information about the population genetic history. Future ancient modern human studies will rather concentrate on positions in the genome that are known to be variable within modern human populations, called single nucleotide polymorphisms (SNPs). Instead of sequencing billions of bases, several hundred thousand such SNPs are sufficient to give information about phenotypic properties of interest and population genetic questions. Such an approach, also called genotyping or re-sequencing, could be carried out for many individuals from an ancient modern human population. To allow such high throughput genotyping studies on ancient modern human remains the above mentioned capture approaches, such as array hybridization capture (Burbano et al. 2010), provide an adequate tool. Additionally the recently described approach to differentiate between modern human contamination and endogenous early modern human DNA, based on DNA degradation patterns, will allow detecting and minimizing contemporary modern human contamination in large scale ancient modern human genotyping approaches.

Conclusions

High throughput sequencing technologies have tremendously lowered time and costs of ancient DNA studies over the last few years, allowing almost complete ancient genomes of extinct species and populations to be sequenced. This provides not only new information about phylogenetic relationships between extant and extinct organisms, or phenotypic information that cannot be recovered from bone morphology, it furthermore provides large scale information about evolutionary changes, such as regions of the modern human genome that were favored by selection after modern humans and

Neandertals diverged (Green et al. 2010). By that the genomes of extinct hominins, such as the Neandertal genome, provide a unique source of data, a sort of outer view on our own modern human genome. Complete genome sequences of other extinct hominins will allow in the future getting more information on genetic relationships of extinct hominins and their evolutionary adaptations to different environments. This should not only complement paleoanthropological and archaeological information about our closest extinct relatives, it should also help us to explain biological differences between different human forms and give insights into the emergence and biological evolution of our own species.

Acknowledgements

I thank Jesse Dabney, Christopher Miller, and Susanna Rankin for helpful comments.

References

Beauval, C., Maureille, B., Lacrampe-Cuyaubère, F., Serre, D., Peressinotto, D., Bordes, J.-G., Cochard, D. Couchoud, I., Dubrasquet, D., Laroulandie, V., Lenoble, A., Mallye, J.-B., Pasty, S., Primault, J., Rohland, N., Pääbo, S., and Trinkaus, E. 2005: A late Neandertal femur from Les Rochers-de-Villeneuve, France. Proceedings of the National Academy of Sciences of the U.S.A. 102, 7085-7090.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., et al. 2008: Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456, 53-59.

Bramanti, B., Thomas, M. G., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M. N., Jankauskas, R., Kind, C.-J., Lueth, F., Terberger, T., Hiller, J., Matsumura, S., Forster, P., and Burger, J. 2009: Genetic Discontinuity Between Local Hunter-Gatherers and Central Europe's First Farmers. Science 326, 137-140.

Briggs, A. W., Good, J. M., Green, R. E., Krause, J., Maricic, T., Stenzel, U., Lalueza-Fox, C., Rudan, P., Brajković, D., Kućan, Ž., Gušić, I., Schmitz, R., Doronichev, V. B., Golovanova, L. V., de la Rasilla, M., Fortea, J., Rosas, A., and Pääbo, S. 2009 (Briggs et al. 2009a): Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes. Science 325, 318-321.

Briggs, A. W., Good, J. M., Green, R. E., Krause, J., Maricic, T., Stenzel, U., and Pääbo, S. 2009 (Briggs et al. 2009b): Primer Extension Capture: Targeted Sequence Retrieval from Heavily Degraded DNA Sources. Journal of Visualized Experiments 31. doi: 10.3791/1573.

Briggs, A. W., Stenzel, U., Johnson, P. L. F., Green, R. E., Kelso, J., Prüfer, K., Meyer, M. Krause, J., Ronan, M. T., Lachmann, M., and Pääbo, S. 2007: Patterns of damage in genomic DNA sequences from a Neandertal. Proceedings of the National Academy of Sciences of the U.S.A. 104, 14616-14621.

Burbano, H. A., Hodges, E., Green, R. E., Briggs, A. W., Krause, J., Meyer, M., Good, J. M., Maricic, P., Johnson, P. L. F., Xuan, Z., Rooks, M., Bhattacharjee, A., Brizuela, L., Albert, F. W., de la Rasilla, M., Fortea, J., Rosas, A., Lachmann, M., Hannon, G. J., and Pääbo, S. 2010: Targeted Investigation of the Neandertal Genome by Array-Based Sequence Capture. Science 328, 723-725.

Campbell, K. L., Roberts, J. E. E., Watson, L. N., Stetefeld, J., Sloan, A. M., Signore, A. V., Howatt, J. W., Tame, J. R. H., Rohland, N., Shen, T.-J., Austin, J. J., Hofreiter, M., Ho, C., Weber, R. E., and Cooper, A. 2010: Substitutions in woolly mammoth hemoglobin confer biochemical properties adaptive for cold tolerance. Nature Genetics 42, 536-540.

Caramelli, D., Lalueza-Fox, C., Condemi, S., Longo, L., Milani, L., Manfredini, A., de Saint Pierre, M., Adoni, F., Lari, M., Giunti, P., Ricci, S., Casoli, A., Calafell, F., Mallegni, F., Bertranpetit, J., Stanyon, R., Bertorelle, G., and Barbujani, G. 2006: A highly divergent mtDNA sequence in a Neandertal individual from Italy. Current Biology 16, R630-R632.

Chen, F. C. and Li, W. H. 2001: Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. American Journal of Human Genetics 68, 444-456.

Cooper, A., Mourer-Chauviré, C., Chambers, G. K., von Haeseler, A., Wilson, A. C., and Pääbo, S. 1992: Independent origins of New Zealand moas and kiwis. Proceedings of the National Academy of Sciences of the U.S.A. 89, 8741-8744.

Cooper, A. and Poinar, H. N. 2000: Ancient DNA: do it right or not at all. Science 289, 1139.

Cummings, M. P., Otto, S. P., and Wakeley, J. 1995: Sampling properties of DNA sequence data in phylogenetic analysis. Molecular Biology and Evolution 12, 814-822.

Debruyne, R., Chu, G., King, C. E., Bos, K., Kuch, M., Schwarz, C., Szpak, P., Gröcke, D. R., Matheus, P., Zazula, G., Guthrie, D., Froese, D., Buigues, B., de Marliave, C., Flemming, C., Poinar, D., Fisher, D., Southon, J., Tikhonov, A. N., MacPhee, R. D. E., and Poinar, H. N. 2008: Out of America: Ancient DNA Evidence for a New World Origin of Late Quaternary Woolly Mammoths. Current Biology 18, 1320-1326.

Ermini, L., Olivieri, C., Rizzi, E., Corti, G., Bonnal, R., Soares, P., Luciani, S., Marota, I., De Bellis, G., Richards, M. B., and Rollo, F. 2008: Complete Mitochondrial Genome Sequence of the Tyrolean Iceman. Current Biology 18, 1687-1693.

Gilbert, M. T. P., Jenkins, D. L., Götherström, A., Naveran, N., Sanchez, J. J., Hofreiter, M., Thomsen, P. F., Binladen, J., Higham, T. F. G., Yohe II, R. M., Parr, R., Scott Cummings, L., and Willerslev, E. 2008 (Gilbert et al. 2008a): DNA from Pre-Clovis Human Coprolites in Oregon, North America. Science 320, 786-789.

Gilbert, M. T. P., Kivisild, T., Grønnow, B., Andersen, P. K., Metspalu, E., Reidla, M., Tamm, E., Axelsson, E., Götherström, A., Campos, P. F., Rasmussen, M., Metspalu, M., Higham, T. F. G., Schwenninger, J.-L., Nathan, R., De Hoog, C.-J., Koch, A., Nukaaraq Møller, L., Andreasen, C., Meldgaard, M., Villems, R., Bendixen, C., and Willerslev, E. 2008 (Gilbert et al. 2008b): Paleo-Eskimo mtDNA Genome Reveals Matrilineal Discontinuity in Greenland. Science 320, 1787-1789.

Gilbert, M. T. P., Tomsho, L. P., Rendulic, S., Packard, M., Drautz, D. I., Sher, A., Tikhonov, A., Dalén, L., Kuznetsova, T., Kosintsev, P., Campos, P. F., Higham, T., Collins, M. J., Wilson, A. S., Shidlovskiy, F., Buigues, B., Ericson, P. G. P., Germonpré, M., Götherström, A., Iacumin, P., Nikolaev, V., Nowak-Kemp, M., Willerslev, E., Knight, J. R., Irzyk, G. P., Perbost, C. S., Fredrikson, K. M., Harkins, T. T., Sheridan, S., Miller, W., and Schuster, S. C. 2007: Whole-Genome Shotgun Sequencing of Mitochondria from Ancient Hair Shafts. Science 317, 1927-1930.

Green, R. E., Briggs, A. W., Krause, J., Prüfer, K., Burbano, H. A., Siebauer, M., Lachmann, M., and Pääbo, S. 2009: The Neandertal genome and ancient DNA authenticity. EMBO Journal 28, 2494-2502.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., Good, J. M., Schultz, R., Aximu-Petri, A., Butthof, A., Höber, B., Höffner, B., Siegemund, M, Weihmann, A, Nusbaum, C., Lander, E. S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Ž., Gušic, I., Doronichev, V. B., Golovanova, L. V., Lalueza-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R., Johnson, P. L. F., Eichler, E. E., Falush, D., Birney, E., Mullikin, J. C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., and Pääbo, S. 2010: A Draft Sequence of the Neandertal Genome. Science 328, 710-722.

Green, R. E., Krause, J., Ptak, S. E., Briggs, A. W., Ronan, M. T., Simons, J. F., Du, L., Egholm, M., Rothberg, J. M., Paunovic, M., and Pääbo, S. 2006: Analysis of one million base pairs of Neanderthal DNA. Nature 444, 330-336.

Green, R. E., Malaspinas, 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., Kućan, Ž., Gušić, 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, 416-426.

Greenwood, A. D., Capelli, C., Possnert, G., and Pääbo, S. 1999: Nuclear DNA sequences from late Pleistocene megafauna. Molecular Biology and Evolution 16, 1466-1473.

Hagelberg, E., Bell, L. S., Allen, T., Boyde, A., Jones, S. J., and Clegg, J. B. 1991: Analysis of ancient bone DNA: techniques and applications. Philosophical Transactions of the Royal Society London B: Biological Sciences 333, 399-407.

Handt, O., Richards, M., Trommsdorff, M., Kilger, C., Simanainen, J., Georgiev, O., Bauer, K., Stone, A., Hedges, R., Schaffner, W., Utermann, G., Sykes, B., and Pääbo, S. 1994: Molecular Genetic Analyses of the Tyrolean Ice Man. Science 264, 1775-1778.

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A., and Wilson, A. C. 1984: DNA sequences from the quagga, an extinct member of the horse family. Nature 312, 282-284.

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, 4793-4799.

Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M., and Pääbo, S. 2001: Ancient DNA. Nature Reviews Genetics 2, 353-359.

Hoss, M., Pääbo, S., and Vereshchagin, N. K. 1994: Mammoth DNA sequences. Nature 370, 333.

Krause, J., Briggs, A. W., Kircher, M., Maricic, T., Zwyns, N., Derevianko, A., and Pääbo, S. 2010 (Krause et al. 2010a): A Complete mtDNA Genome of an Early Modern Human from Kostenki, Russia. Current Biology 20, 231-236.

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, 724-727.

Krause, J., Fu, Q., Good, J. M., Viola, B., Shunkov, M. V., Derevianko, A. P., and Pääbo, S. 2010 (Krause et al. 2010b): The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. Nature 464, 894-897.

Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R. E., Burbano, H. A., Hublin, J.-J., Hänni, C., Fortea, J., de la Rasilla, M., Betranpetit, J., Rosas, A., and Pääbo, S. 2007 (Krause et al. 2007a): The Derived FOXP2 Variant of Modern Humans Was Shared With Neandertals. Current Biology 17, 1908-1912.

Krause, J., Orlando, L., Serre, D., Viola, B., Prüfer, K., Richards, M. P., Hublin, J.-J., Hänni, C., Derevianko, A. P., and Pääbo, S. 2007 (Krause et al. 2007b): Neanderthals in central Asia and Siberia. Nature 449, 902-904.

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.

Krings, M., Capelli, C., Tschentscher, F., Geisert, H., Meyer, S., von Haeseler, A., Grossschmidt, K., Possnert, G., Paunovic, M., and Pääbo, S. 2000: A view of Neandertal genetic diversity. Nature Genetics 26, 144-146.

Krings, M., Stone, A., Schmitz, R. W., Krainitzki, H., Stoneking, M., and Pääbo, S. 1997: Neandertal DNA Sequences and the Origin of Modern Humans. Cell 90, 19-30.

Kuch, M., Gröcke, D. R., Knyf, M. C., Gilbert, M. T., Younghusband, B., Young, T., Marshall, I., Willerslev, E., Stoneking, M., and Poinar, H. 2007: A preliminary analysis of the DNA and diet of the extinct Beothuk: a systematic approach to ancient human DNA. American Journal of Physical Anthropology 132, 594-604.

Lalueza-Fox, C., Römpler, H., Caramelli, D., Stäubert, C., Catalano, G., Hughes, D., Rohland, N., Pilli, E., Longo, L., Condemi, S., de la Rasilla, M., Fortea, J., Rosas, A., Stoneking, M., Schöneberg, T., Bertranpetit, J., and Hofreiter, M. 2007: A Melanocortin 1 Receptor Allele Suggests Varying Pigmentation Among Neanderthals. Science 318, 1453-1455.

Lalueza-Fox, C., Sampietro, M. L., Caramelli, D., Puder, Y., Lari, M., Calafell, F., Martínez-Maza, C., Bastir, M., Fortea, J., de la Rasilla, M., Bertranpetit, J., and Rosas, A. 2005: Neandertal Evolutionary Genetics: Mitochondrial DNA Data from the Iberian Peninsula. Molecular Biology and Evolution 22, 1077-1081.

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., 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, 376-380.

Miller, W., Drautz, D. I., Ratan, A., Pusey, B., Qi, J., Lesk, A. M., Tomsho, L. P., Packard, M. D., Zhao, F., Sher, A., Tikhonov, A., Raney, B., Patterson, N., Lindblad-Toh, K., Lander, E. S., Knight, J. R., Irzyk, G. P., Fredrikson, K. M., Harkins, T. T., Shweridan, S., Pringle, T., and Schuster, S. C. 2008: Sequencing the nuclear genome of the extinct woolly mammoth. Nature 456, 387-390.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. 1986: Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbor Symposia on Quantitative Biology 51 Pt. 1, 263-273.

Noonan, J. P., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J., Chen, F., Platt, D., Pääbo, S., Pritchard, J. K., and Rubin, E. M. 2006: Sequencing and Analysis of Neanderthal Genomic DNA. Science 314, 1113-1118.

Noonan, J. P., Hofreiter, M., Smith, D., Priest, J. R., Rohland, N., Rabeder, G., Krause, J., Detter, J. C., Pääbo, S., and Rubin, E. M. 2005: Genomic Sequencing of Pleistocene Cave Bears. Science 309, 597-599.

Orlando, L., Darlu, P., Toussaint, M., Bonjean, D., Otte, M., and Hänni, C. 2006: Revisiting Neandertal diversity with a 100,000 year old mtDNA sequence. Current Biology 16, R400-R402.

Ovchinnikov, I. V., Götherström, A., Romanova, G. P., Kharitonov, V. M., Lidén, K., and Goodwin, W. 2000: Molecular analysis of Neanderthal DNA from the northern Caucasus. Nature 404, 490-493.

Pääbo, S. 1985: Molecular cloning of Ancient Egyptian mummy DNA. Nature 314, 644-645.

Pääbo, S. 1989: Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proceedings of the National Academy of Sciences of the U.S.A 86, 1939-1943.

Pääbo, S. 1999: Human evolution. Trends in Cell Biology 9, M13-M16.

Pääbo, S. 1984: Über den Nachweis von DNA in altägyptischen Mumien. Das Altertum 30, 213-218.

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Després, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., and Hofreiter M. 2004: Genetic Analyses from Ancient DNA. Annual Review of Genetics 38, 645-679.

Pääbo, S. and Wilson, A. C. 1988: Polymerase chain reaction reveals cloning artefacts. Nature 334, 387-388.

Poinar, H. N., Schwarz, C., Qi, J., Shapiro, B., MacPhee, R. D. E., Buigues, B., Tikhonov, A., Huson, D. H., Tomsho, L. P., Auch, A., Rampp, M., Miller, W., and Schuster, S. C. 2006: Metagenomics to Paleogenomics: Large-Scale Sequencing of Mammoth DNA. Science 311, 392-394.

Prüfer, K., Stenzel, U., Hofreiter, M., Pääbo, S., Kelso, J., and Green, R. E. 2010: Computational challenges in the analysis of ancient DNA. Genome Biology 11, R47.

Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J. S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., Bertalan, M., Nielsen, K., Gilbert, M. T. P., Wang, Y., Raghavan, M., Campos, P. F., Munkholm Kamp, H., Wilson, A. S., Gledhill, A., Tridico, S., Bunce, M., Lorenzen, E. D., Binladen, J., Guo, X., Zhao, J., Zhang, X., Zhang, H., Li, Z., Chen, M., Orlando, L., Kristiansen, K., Bak, M., Tommerup, N., Bendixen, C., Pierre, T. L., Grønnow, B., Meldgaard, M., Andreasen, C., Fedorova, S. A., Osipova, L. P., Higham, T. F. G., Bronk Ramsey, C., v. O. Hansen, T., Nielsen, F. C., Crawford, M. H., Brunak, S., Sicheritz-Pontén, T., Villems, R., Nielsen, R., Krogh, A., Wang, J., and Willerslev, E. 2010: Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature 463, 757-762.

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, 1663-1671.

Römpler, H., Rohland, N., Lalueza-Fox, C., Willerslev, E., Kuznetsova, T., Rabeder, G., Bertranpetit, J., Schöneberg, T., and Hofreiter, M. 2006: Nuclear Gene Indicates Coat-Color Polymorphism in Mammoths. Science 313, 62.

Schmitz, R. W., Serre, D., Bonani, G., Feine, S., Hillgruber, F., Krainitzki, H., Pääbo, S., and Smith, F. H. 2002: The Neandertal type site revisited: Interdisciplinary investigations of skeletal remains from the Neander Valley, Germany. Proceedings of the National Academy of Sciences of the U.S.A. 99, 13342-13347.

Schuster, S. C., Miller, W., Ratan, A., Tomsho, L. P., Giardine, B., Kasson, L. R., Harris, R. S., Petersen, D. C., Zhao, F., Qi, J., Alkan, C., Kidd, J. M., Sun, Y., Drautz, D. I., Bouffard, P., Muzny, D. M:, Reid, J. G., Nazareth, L. V., Wang, Q., Burhans, R., Riemer, C., Wittekindt, N. E., Moorjani, P., Tindall, E. A., Danko, C. G., Teo, W. S., Buboltz, A. M., Zhang, Z., Ma, Q., Oosthuysen. A., Steenkamp, A. W., Oostuisen, H., Venter, P., Gajewski, J., Zhang, Y., Pugh, B. F., Makova, K. D., Nekrutenko, A., Mardis, E. R., Patterson, N., Pringle, T. H., Chiaromonte, F., Mullikin, J. C., Eichler, E. E., Hardison, R. C., Gibbs, R. A., Harkins, T. T., and Hayes, V. M. 2010: Complete Khoisan and Bantu genomes from southern Africa. Nature 463, 943-947.

Serre, D., Langaney, A., Chech, M., Teschler-Nicola, M., Paunovic, M., Mennecier, P., Hofreiter, M., Possnert, G., and Pääbo, S. 2004: No Evidence of Neandertal mtDNA Contribution to Early Modern Humans. PLoS Biology 2, 313-317.

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 306, 1561-1565.

Shapiro, B. and Hofreiter, M. 2010: Analysis of ancient human genomes. Using next generation sequencing, 20-fold coverage of the genome of a 4,000-year-old human from Greenland has been obtained. Bioessays 32, 388-391.

Thomas, R. H., Schaffner, W., Wilson, A. C., and Pääbo, S. 1989: DNA phylogeny of the extinct marsupial wolf. Nature 340, 465-467.

Wall, J. D. and Kim, S. K. 2007: Inconsistencies in Neanderthal Genomic DNA Sequences. PLoS Genetics 3, 1862-1866.

Woodward, S. R., Weyand, N. J., and Bunnell, M. 1994: DNA Sequence from Cretaceous Period Bone Fragments. Science 266, 1229-1232.

Zischler, H., Geisert, H., von Haeseler, A., and Pääbo, S. 1995: A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. Nature 378, 489-492.