Insights into Neandertals and Denisovans from Denisova Cave

Der Fakultät für Biowissenschaften, Pharmazie und Psychologie

der Universität Leipzig

genehmigte

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

Doctor rerum naturalium

Dr. rer. nat.

vorgelegt

von Master of Science Susanna Sawyer

geboren am 23.09.1985 in Trostberg, Deutschland

Dekan:

Gutachter: Professor Svante Pääbo und Dr. Beth Shapiro

Tag der Verteidigung 08.04.2016

BIBLIOGRAPHISCHE DARSTELLUNG

Susanna Sawyer

Insights into Neandertals and Denisovans from Denisova cave

Fakultät für Biowissenschaften, Pharmazie und Psychologie

Universität Leipzig

Dissertation

139 Seiten, 153 Literaturangaben, 30 Abbildungen, 32 Tabellen

Denisova Cave is located in the Altai mountains of Russia. Excavations from this cave have yielded two large hominin molars and three hominin phalanxes from the Pleistocene. One of the phalanxes (*Denisova 3*) had extraoridinary DNA preservation allowing the sequencing of high quality nuclear and mitochondrial DNA (mtDNA) genomes and has been shown to belong to a young girl from hereto unknown sister group of Neandertals, called Denisovans. The mtDNA of *Denisova 3* surpirisingly split from the mtDNA ancestor of modern humans and Neandertals twice as long ago as the split of modern humans and Neandertals. The mtDNA of one of the molars (*Denisova 4*) was also sequenced and differs at only two positions from the mtDNA of *Denisova 3.* A second phalanx (*Altai 1*) also yielded a high quality genome, and was a Neandertal. While Neandertals show an admixture signal of 1-4% into present-day non-Africans, Denisovans show an admixture of up to 5% in present-day Oceanians, and to a much lesser extent East Asians.

This thesis encompasses two studies. In the first study, we sequenced the complete mtDNA genome of the additional molar (*Denisova 8*), as well as a few megabases of nuclear DNA from *Denisova 4* and *Denisova 8*. While the mtDNA of *Denisova 8* is clearly of the Denisova type, its branch to the most recent common ancestor of Denisovans is half as long as the branch leading to *Denisova 3* or *Denisova 4*, indicating that *Denisova 8* lived many millenia before the other two. Both *Denisova 4* and *8* fall together with *Denisova 3* based on nuclear DNA, bringing the number of known Denisovans from one to three.

In the second study, we sequenced an almost complete mtDNA and a few megabases of nuclear DNA from the third hominin phalanx from Denisova Cave, *Altai 2*. Both the mtDNA and the nuclear DNA show *Altai 2* to be a Neandertal. The mtDNA also showed the presence of substantial Pleistocene spotted hyena contamination. Low levels of spotted hyena contamination were also found in *Altai 1*, *Denisova 3* and *Denisova 4*. Partial mtDNA genomes of the contaminating spotted hyenas from these four hominins were compared to mtDNA genomes of other extant and extinct spotted hyenas. We show that the spotted hyenas that contaminated the two Denisovans come from a population of spotted hyenas found in Pleistocene Europe as well as present-day Africa, while the spotted hyenas that contaminated *Altai 2*, and possibly *Altai 1*, come from a population of spotted hyenas found in Pleisticene eastern Russia and northern China. This indicates that Denisova Cave was a meeting point of eastern and western hominins as well as eastern and western spotted hyena populations.

For Emma.

Table of Contents

1. Thesis Summary

Denisova Cave is located in the Siberian Altai Mountains on the Anui River close to the borders with Kazakhstan, China and Mongolia. Although there are other archaeological sites in the region, including Okladnikov Cave to the North where Neandertal remains have been found, the stone tools excavated in the Altai Mountains show little change in culture from 300,000 to 30,000 years ago, which has prompted some to argue for a multi-regional human evolution model in central Asia (1, 2). In addition, the region has yielded few hominin remains (3). The hominin remains discovered are often small pieces, with no complete or even partially complete skeletons (3). The geographically closest complete skeletons are from two early modern human children in Mal'ta on the southern tip of Lake Baikal (4).

Excavations of Denisova Cave, led by Professor Anatoly Derevianko, have been ongoing since 1984. To date these excavations have yielded seven hominin remains from the Pleistocene (5, 6). One of these remains, a small piece of a terminal finger phalanx of a young child (*Denisova 3*), was found in 2008 in the East Gallery. Ancient DNA (aDNA) preservation in this bone was remarkably good and yielded not only a complete mitochondrial DNA (mtDNA) genome (7), but also a high quality nuclear genome (8). While the mtDNA of the *Denisova 3* split off twice as early as the split between Neandertals and early modern humans (7), the nuclear DNA showed that *Denisova 3* belonged to a sister group of Neandertals (8, 9) which was named Denisovans. In 2000, an unusually large molar (*Denisova 4*) was found in the South Gallery of the Cave. Before the present study, only mtDNA was retrieved from *Denisova 4* (9). The mtDNA has only two differences to the mtDNA of *Denisova 3*, indicating that the molar may have belonged to a Denisovan (9). In 2010, an intact toe phalanx was found in the East Gallery. Again the DNA preservation was good enough to produce high quality nuclear and mtDNA genomes, which revealed that the toe bone belonged to a Neandertal (10). I refer to this individual as *Altai 1*.

From previous Neandertal data, it was known that non-Africans carry 1-4% DNA from Neandertals (11). Denisovans show an admixture signal to present-day Oceanians (of 3-6%), and a small admixture signal of ~0.2% to East Asians (10). Additionally, *Denisova 3* carries >0.5% of Neandertal DNA that is more closely related to *Altai 1* than to more western

Neandertals. *Denisova 3* also carries DNA from another archaic hominin, from which it may have gotten its deeply diverged mtDNA genome (10).

While none of the hominin remains from Denisova Cave are directly C^{14} dated, dating based on branch shortening due to a lack of accumulated mutations estimates that *Altai 1* is older than *Denisova 3*, and that both lived between 120,000 and 50,000 years ago (10). During this time Pleistocene spotted hyenas lived in the Altai Mountains (12). Based on cytochrome b data from Pleistocene hyenas, they fall into the genetic variation of spotted hyenas in sub-Saharan Africa today (13, 14) They were larger in size (12), and have been shown to have eaten hominin remains in the area (3).

This thesis is composed of two studies. First I discuss two additional Denisovans from Denisova Cave. We sequenced a small amount of nuclear DNA sequences from *Denisova 4*, the large molar from the South Gallery. We then calculated the divergence of *Denisova 4* from *Denisova 3*, *Altai 1* and ten present-day humans from around the world, on the lineage to the human and chimpanzee ancestor. *Denisova 4* diverged from *Denisova 3* 2.9% back on this lineage, which is $1/3^{rd}$ of its divergence from *Altai 1* and $1/5^{th}$ of its divergence from the present-day humans. Therefore *Denisova 4* is a Denisovan.

We sequenced 24 Megabases (Mb) of the nuclear genome as well as the complete mtDNA genome from an additional third molar from the East Gallery of Denisova Cave (*Denisova 8*). The mtDNA genome of *Denisova 8* falls together with the two previously described Denisovans. However, its mtDNA is quite diverged, carrying almost twice as many differences to the other two mtDNA genomes as seen between pairs of Neandertal mtDNAs ranging from Spain to Siberia. The number of mutations leading to the *Denisova 8* mtDNA from the most recent common ancestor of Denisovans is almost half the number of mutations to the other two Denisovans. This translates to a 60,000 to 100,000 age difference between these two Denisovan groups. The nuclear divergence of *Denisova 8* is lower to *Denisova 3* than either Neandertals or present-day humans, therefore this additional molar also belonged to a Denisovan. However, the divergence of *Denisova 8* from *Denisova 3* is higher than that of *Denisova 4* from *Denisova 3*, and higher than the divergences between Neandertals.

The mtDNA branch shortening of *Denisova 8* suggests that Denisovans inhabited the Denisova Cave region at least twice over a very long time period, interrupted or possibly coexisting with Neandertals for some time. Since both of the molars from the Denisovans are

unusually large, it suggests that this morphology was present in the Denisovans of the Altai Mountains for a long time, and that is may aid in identifying Denisovans in the future.

The second study presented in this thesis centers on a fifth hominin remain from Denisova Cave. *Altai 2* is a complete finger phalanx found in the deepest layer of the East Gallery excavated to date. It is a hominin bone based on morphology. We sequenced an almost complete mtDNA genome from this specimen. It is of the Neandertal type and clusters closest with *Altai 1*, with only ten differences. We also sequenced 18.4 Mb of the nuclear genome. On the lineage to the human chimpanzee ancestor, the divergence of *Altai 2* is lowest to *Altai 1*, second lowest to *Denisova 3* and highest to ten present-day humans. However *Altai 2* has the highest divergence to *Altai 1* when compared to other Neandertals from Spain, Croatia and the Caucasus.

We found a large amount of spotted hyena DNA in the *Altai 2* bone (32.4% of total), which could explain the deep divergence of *Altai 2* to *Altai 1* on the nuclear level. We were able to sequence a partial mtDNA genome of the main contaminating hyena. It falls outside the variation of the four Pleistocene and extant spotted hyenas for whom complete mtDNA genomes exist. Based on data from cytochrome b from 57 present-day and Pleistocene spotted hyenas, the *Altai 2* spotted hyena contaminant is most closely related to Pleistocene spotted hyenas from eastern Russia and China.

We sequenced the almost complete mtDNA of a spotted hyena bone from the same layer and gallery in Denisova Cave. In addition we looked for spotted hyena mtDNA sequences among the DNA sequences determined from other Denisova Cave hominins. We found low level spotted hyena contamination in *Denisova 3* and *4* and *Altai 1*, but not in *Denisova 8*. The spotted hyena contaminants of *Denisova 3* and *4*, as well as the spotted hyena from Denisova Cave fall together with Pleistocene spotted hyenas from Europe and extant spotted hyenas from Africa. The *Altai 1* spotted hyena contaminant is more diverged, but not as diverged as the *Altai 2* hyena contaminant.

We washed the bone powder of *Altai 2* with a phosphate wash, which has been shown to be effective at washing off microbial DNA that has colonized a bone after the death of an individual (15). The microbial DNA content was higher in the phosphate wash than in the subsequent DNA extraction, however the hyena DNA did not preferentially wash off. In fact

the amount of hyena DNA present in the subsequent extraction was higher than the endogenous Neandertal DNA.

How did the hyena contamination end up in these remains? It is possible that spotted hyenas hunted hominins and left only small phalanxes and teeth behind, a practice not uncommon for spotted hyenas today with their prey (12, 16). Such an idea has been suggested for the nearby Okladnikov Cave, a small cave, more suited in size to a hyena than to a human (3). Okladnikov Cave has large amounts of hyena remains, and the hominin remains that were found there are believed to have been dragged into the cave (3). The Pleistocene hyenas may have scavenged the carcasses of hominins, possibly by digging up graves, again a behavior seen in spotted hyenas today (17).

It is possible, however, that the contaminating hyenas had no interaction with the hominins. The Denisovans and Neandertals may have left their remains some other way in Denisova Cave, and over the thousands of years until spotted hyenas died out 13-14,000 years ago (12), hyenas were digging in the cave and either eating the old bones or urinating and defecating in the cave, thus allowing their DNA to leech into the soil (18) and into the remaining bones and teeth. Even if Denisova Cave was not often used by either Denisovans or Neandertals, and was instead mostly a spotted hyena den, these hominins must have lived within the home range of the hyenas using the Caves, within 100 km of the Cave (17).

Denisova Cave was a meeting point of hominids from the east (Denisovans) and west (Neandertals) as well as Pleistocene spotted hyena populations from the east and west. Interestingly the western spotted hyenas are associated with Denisovans while the eastern spotted hyenas are associated with Neandertals. More sequencing and dating of Pleistocene spotted hyenas from Eurasia could potentially shed light on Neandertal and Denisovan movements in the area.

2. Zusammenfassung

Die Denisova-Höhle befindet sich im sibirischen Altaigebirge am Fluss Anui nahe der Grenze zu Kasachstan, China und der Mogolei. Trotz anderer archäologischer Ausgrabungsstätten in der Region – unter anderem die Okladnikov-Höhle im Norden, eine Fundstelle für Neandertalerknochen – zeigen die Steinwerkzeuge aus dem Altai-Gebirge vor 300.000-30.000 Jahren wenig kultuelle Veränderung, was bei einigen Wissenschaftler zur Annahme eines multiregionalen Evolutionmodells in Zentralasien führte (1 ,2).

Zusätzlich wurden in der Region wenige menschliche Überreste gefunden. Die ausgegrabenen menschlichen Fossilien sind oft nur kleine Fragmente, welche keinen vollständige Rekonstruktion des Skeletts zulassen (3). Die am besten rekonstruierten Skelette stammen von zwei früh-modernen Kindern aus Mal'ta von der südlichen Spitze des Baikalsees (4).

Seit 1984 finden, geleitet von Professor Anatoly Derevianko, Ausgrabungen in der Denisova Höhle statt. Dabei konnten bisher sieben menschliche Fossilien aus dem Pleistozän geborgen werden (5). Eines dieser Überreste – ein kleiner Teil von einem Fingerglied eines jungen Kindes (*Denisova 3*) – wurde 2008 in der Ost-Galerie gefunden. Die DNA-Erhaltung in diesem Knochen war bemerkenswert gut und lieferte nicht nur ein vollständiges Mitochnodrien DNA Genom (mtDNA) (6), sondern auch ein hochqualitatives nukleares Genom (7).

Anhand der mtDNA fand eine Abspaltung von *Denisova 3* zweimal früher statt als die Abspaltung des Neandertalers von der menschlichen Linie (6). Das nukleare Genome weist hingegen darauf hin, dass *Denisova 3* eine Schwestergruppe zu den Neandertalern bildet (7, 8), welche als Denisovaner oder Denisova-Menschen bezeichnet wird.

Im Jahr 2000 wurde ein ungewöhnlich großer Backenzahn in der Süd-Galerie der Höhle gefunden (*Denisova 4*). Anhand von Analysen der mtDNA von *Denisova 4*, die einzige genetische Information vor der Veröffentlichung der hier präsentierten Studie, wurde festgestellt, dass *Denisova 4* nur zwei Unterschiede zu *Denisova 3* aufweist, ein Indikator dafür, dass *Denisova 4* zu den Denisova-Menschen gehört (8).

2010 wurde dann ein intakter Zehenknochen in der Ost-Galerie ausgegraben, dessen DNA-Erhaltung erneut gut genug war um hochqualifizierte mt and nukleare Genome ervorzubringen. Analysen der DNA ergaben eine Zugehörigkeit zu den Neandertalern (9). Dieses Individuum wird fortlaufend als *Altai 1* bezeichnet.

Von früheren Neandertal-Daten ist bekannt, dass Nicht-Afrikaner 1-4% Neandertaler-DNA in sich tragen (10). Denisova-Menschen zeigen eine Vermischung mit heute lebenden Ozeaniern von 3-6%, mit Ostasiaten hingehen nur ~0.2% (9). *Denisova 3* weist >0.5% an Neandertaler-DNA auf und ist somit näher mit *Altai 1* als mit westlicheren Neandertalern verwandt. Zusätzlich zeigt *Denisova 3* Anzeichen von Vermischung mit einem weiteren unbekannten archaischen Menschen, von welchem es sein stark divergentes mtGenome haben könnte (9).
Das Alter der homininen Überreste wurde bisher weder durch C14-Datierung bestimmt, noch kann die Stratigraphie der Höhle zu Bestimmung herangezogen werden. Mit Hilfe von *branch shortening*, wobei das Alter an Hand fehlender Mutationenanhäufung nach dem Tod des Organismus bestimmt werden kann, wird geschätzt, dass *Altai 1* älter als *Denisova 3* ist und beide vor 120.000 - 50.000 Jahren gelebt haben (9).
Während dieser Zeit lebten außerdem Tüpfelhyänen des Pleistozän im Altai-Gebirge (11), welche zwar keine separate Spezies zu den heute lebenden Tüpfelhyänen der afrikanischen Subsahara darstellten (12, 13), die aber deutlich größer waren und sich von homininen Überresten ernährten (3).

Diese Doktorarbeit ist in zwei Teile unterteilt. Zuerst werden zwei neue Denisova-Menschen aus der Denisova-Höhle diskutiert. Teile des nuklearen Genoms des großen Backenzahns aus der Süd-Galerie (*Denisova 4*) wurden sequenziert. Außerdem wurde die Divergenz von *Denisova 4* zu *Denisova 3*, *Altai 1* und zehn modernen Menschen aus aller Welt auf der Linie zum Vorfahren des Menschen und Schimpanzen bestimmt. *Denisova 4* und *Denisova 3* divergieren zu 2.9 %, dies entspricht einem Drittel seiner Divergenz zu *Altai 1* und einem Fünftel seiner Divergenz zu allen modernen Menschen und kategorisiert *Denisova 4* als einen Denisova-Menschen.
Außerdem wurden insgesamt 24 Megabasen (Mb) des nuklearen Genoms und das komplette mtDNA Genom eines dritten Backenzahns aus der Ost-Galerie der Denisova-Höhle sequenziert (*Denisova 8*). Das mtGenome fällt taxonomisch zusammen mit den bereits beschriebenen Denisova-Menschen. Nichtsdestotrotz ist die mtDNA mit fast doppelt so vielen Unterschieden zu den beiden anderen mtGenomen so divergent wie beispielsweise die mtDNA zwischen einem Neandertalern aus Spanien und einem aus Sibieren. Die Anzahl der Mutationen zwischen *Denisova 8* und dem jünsgten Vorfahren der Denisova-Menschen ist fast halb so groß wie die

Anzahl der Mutationen von *Denisova 4* und *Denisova 3,* dies lässt sich in einen Altersunterschied der beiden Denisova-Gruppen von 60.000-100.000 Jahren übersetzen.
Die Divergenz von *Denisova 8* und *Denisova 3* ist niedriger als zu irgendeinem Neandertaler oder modernen Menschen, was den zusätzlich gefundenen Backenzahn ebenfalls zu den Denisova-Menschen gehören lässt. Dennoch ist die Divergenz zwischen *Denisova 8* und *Denisova 3* größer als zwischen *Denisova 4* und *Denisova 3,* sowie der Divergenz innerhalb der Neandertaler.
Auf Grund des mtDNA *branch shortening* von *Denisova 8* kann angenommen werden, dass die Denisova-Menschen die Denisova-Höhle über zwei lange Zeitspannen bewohnt haben, mit Unterbrechung oder möglicher zeitlichen Überlappung durch die Existenz von Neandertalern.
Auf Grund der ungewöhlichen Größer der Denisova-Backenzähne wird angenommen, dass diese Morphologie über lange Zeit bei den Altai-Denisovanern vorzufinden war und dass es als Unterstützung zur Identifizierung von weiteren Denisova-Menschen herangezogen werden kann.
Man geht davon aus, dass das divergente mtDNA Genom von *Denisova 3* durch Genfluss von einer unbekannten archaischen Population hervorgegangen ist, welche sich vor 1-4 Millionen Jahren von den Denisova-Menschen abgespalten hat (9). Die diversere mtDNA von *Denisova 8* könnte hingegen von einer divergenteren archaischen Gruppe stammen, was aufgrund der geringen DNA-Erhaltung des Fossils aber momentan leider nicht beantwortet werden kann.

Die zweite Studie aus dieser Doktorarbeit beschäftifgt sich mit einem fünften homininen Fossil aus der Denisova-Höhle. *Altai 2* ist ein kompletter Fingergliedknochen aus der bisher am tiefsten erschlossenen Schicht der Ost-Galerie. Basierend auf der Morpholigie ist der Knochen homininer Herkunft. Das mtDNA Genom der Spezies wurde fast vollständig sequenziert. Demnach gehört es den Neandertalern an und fällt, mit nur zehn Unterschieden, taxonomisch mit *Altai 1* zusammen. Desweiteren wurden 18.4 Mb des nuklearen Genoms seuqenziert. Auf der Linie zum gemeinsamen Vorfahren von Mensch und Schimpanze ist die Divergenz von *Altai 2* am geringsten zu *Altai 1,* am zweit geringsten zu *Denisova 3* und am höchsten zu den zehn modernen Menschen. Dennoch hat *Altai 2* verglichen zu Neandetalern aus Spanien, Kroatien und dem Kaukasus die höhste Divergenz zu *Altai 1*. Analysen ergaben einen hohen Anteil an DNA von der Tüpfelhyäne in *Altai 2* (32.4%), was die tiefe Divergenz zwischen *Altai 1* und *Altai 2* auf nuklearem Level erklären könnte. Außerdem gelang es einen Teil des mtDA Genoms zu

sequenzieren, welches von der Kontamination durch die Hyäne stammt. Die mtDNA fällt außerhalb der Variation der vier Pleistozän und rezenten Tüpfelhyänen, von welchen vollständige mtDNA Genome existieren. Basierend auf Daten für Cytochrom B von 57 heute lebenden und Pleistozän Tüpfelhyänen ist die *Altai 2* kontaminierende Tüpfelhyäne am nähesten mit den Pleistozän Tüpfelhyänen aus Ostrussland und China verwandt. Fast das gesamte mtDNA Genom einer weiteren Tüpfelhyäne aus derselben Schicht und derselben Galerie der Denisova-Höhle wurde sequenziert. Zusätzlich wurde nach weiteren von Tüpfelhyäne stammenden mtDNA Sequenzen in allen Sequenzen der Homininen aus der Höhle geschaut. Es wurde eine geringe Kontamination an Tüpfelhyäne in *Denisova 3*, *Denisova 4* und *Altai 1*, allerdings keine in *Denisova 8* gefunden. Die Kontaminaten von *Denisova 3* und *Denisova 4*, sowie die sequenzierte Tüpfelhyäne aus der Denisova-Höhle fallen taxonomisch mit den Pleistozän Tüpfelhyänen aus Europa und rezenten Tüpfelhyänen aus Afrika zusammen. Die Tüpfelhyänen-Kontamination von *Altai 1* ist divergenter aber nicht so divergent wie die von *Altai 2*.

Eine erst kürzlich veröffentlichte Phosphatwasch-Methode wurde auf das Knochenpulver angewendet. Es konnte gezeigt werden, dass die Methode von Bakterien stammende DNA, welche den Organismus nach dessen Tod besiedeln, effizient beseitigt (16). Der Anteil mikrobieller DNA im mit Phosphat gewaschenen Überstand war höher als in der anschließenden Extraktion. Dennoch ließ sich die Hyänen-Kontamination nicht effektiv entfernen. Der Anteil der Hyänen DNA im finalen Extrakt war sogar höher als der Anteil an endogener Neandertaler-DNA.

Wie kam es zur Kontamination der menschlichen Fossilien durch die Hyänen? Ist es möglich, dass die Tüpfelhyänen die Homininen jagdten und nur kleine Metapodien und Zähne zurückließen, ein nicht ungewöhnliches Verhalten auch heute lebender Hyänen im Umgang mit ihrer Beute (11, 14). Dies wurde auch für die nahegelegende Okladnikov Höhle angenommen, die auf Grund ihrer schmalen Größe mehr für Hyänen als Menschen geeigent schein (3). In der Okladinikov Höhle wurden viele Überreste von Hyänen gefunden. Von den menschlichen Fossilen glaubt man, dass diese durch Hyänen in die Höhle gebracht wurden (3). Die Kadaver wurden wahrscheinlich gründlich abgenagt und vergraben, ein Verhalten, wie es auch bei heute lebenden Tüpfelhyänen zu beobachten ist (17). Aber es ist auch möglich, dass die kontaminierenden Hyänen keinen Kontakt mit den Homininen hatten. Denisova-Menschen und Neandertaler könnten in der Höhle gelebt und ihr Überreste hinterlassen haben, welche dann von

den Tüpfelhyänen in den tausenden von Jahren bis zu ihren Aussterben vor 13.000-14.000 Jahren, ausgegraben und gefressen oder mit Kot und Urin verunreinigt wurden, was die Tüpfelhyänen-DNA im Boden (18) und auf den menschlichen Fossilien erkären würde.

Auch wenn die Denisova-Höhle wahrscheinlich vorwiegend als Hyänenbau und weniger als Unterschlupf für Denisova-Menschen und Neandertaler herhielt, mussten die Homininen dennoch in einem Umkreis von circa 100 km von der Höhle und damit im Habitat der Hyänen gelebt haben (17).

Zusammenfassend kann man also annehmen, dass das rätselhafte Fehlen von homininen Überresten im Altai-Gebirge höchstwahrscheinlich auf eine Kombination aus kleinen Populationsgrößen der Homininen und der Aktivität von Hyänen in der Region zurückzuführen ist.

3. Introduction

3.1 Human evolution

As humans we have always been fascinated by our origins, as evidenced by the countless religious beliefs that detail the formation of humans. Historically, humans were seen as special and were classed as a unique being (19). The belief that we as humans are special continued until 1863, when Thomas Huxley claimed that the difference between man and other great apes was not as great as the difference between great apes and lower apes (20). In 1871, Charles Darwin confirmed this idea (21). Already in 1856, however, with the discovery of the Neandertal, it was becoming clear that the origin of man is more complex than was formerly believed (22). It took developments in biochemistry and immunology, and most importantly in DNA sequencing, to be able to better describe our relationship with other primates. Using DNA, it became clear that we as humans are in fact most closely related to chimpanzees and bonobos (Figure 1) (23-25).

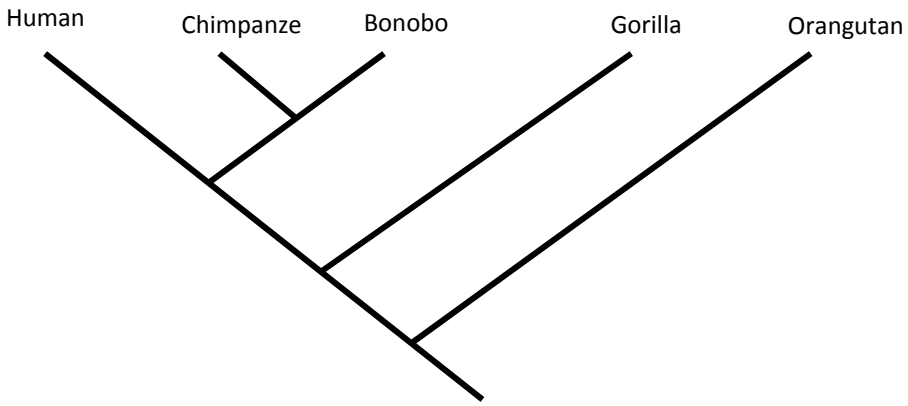

Figure 1. Tree of relationships between humans, chimpanzee, bonobos, gorillas and orangutans based on DNA. Tree is not drawn to scale.

There are many fossils that branch from or fall on the human branch after the split from the chimpanzee/bonobo common ancestor that range from primitive to modern (Figure 2). I will refer to this branch as the hominin branch and the fossils on this branch as hominins in this thesis. The relationship between these fossils is not always clear, specifically who our direct ancestors are and who represents an offshoot from our genetic history, although the fact that both the oldest fossils and our closest living relatives are only found in Africa indicates that hominins evolved in Africa (26). However already 1.8 million years ago, hominin fossils start appearing

outside of Africa (27). Thus there are multiple hypotheses about the evolution of present-day humans, such as the theory that early *Homo* left Africa, and then over the next million years evolved separately into Asians and Europeans, while the hominins left in Africa evolved into Africans, with some exchange of genes throughout this time (28). The out-of-Africa model claims that the ancestors of present-day humans evolved into anatomically modern humans in Africa, and that some of these anatomically modern humans left Africa recently, colonized Europe and Asia very rapidly and replaced the previous hominins living there (29). Genetic evidence from present-day humans made a clear case for the out-of-Africa model (30, 31), showing that present-day humans shared a common ancestor in Africa about 200,000 years ago (32) and that a small group of these early modern humans left Africa only 60,000 years ago (33). Archeological and linguistic data support this model as well (34-36).

Bones and teeth with Neandertal-type morphological features first appear in the fossil record about 400,000 years ago (37, 38). Neandertals have been found across Europe (22), the Middle East (39, 40), the Caucasus (41), and as far east as central Asia and Siberia (42-44). They disappeared around 30,000 years ago (45). Since the oldest early modern human fossils are found in the Middle East already 120,000 years ago (46, 47) there existed the possibility that Neandertals and early modern humans met.

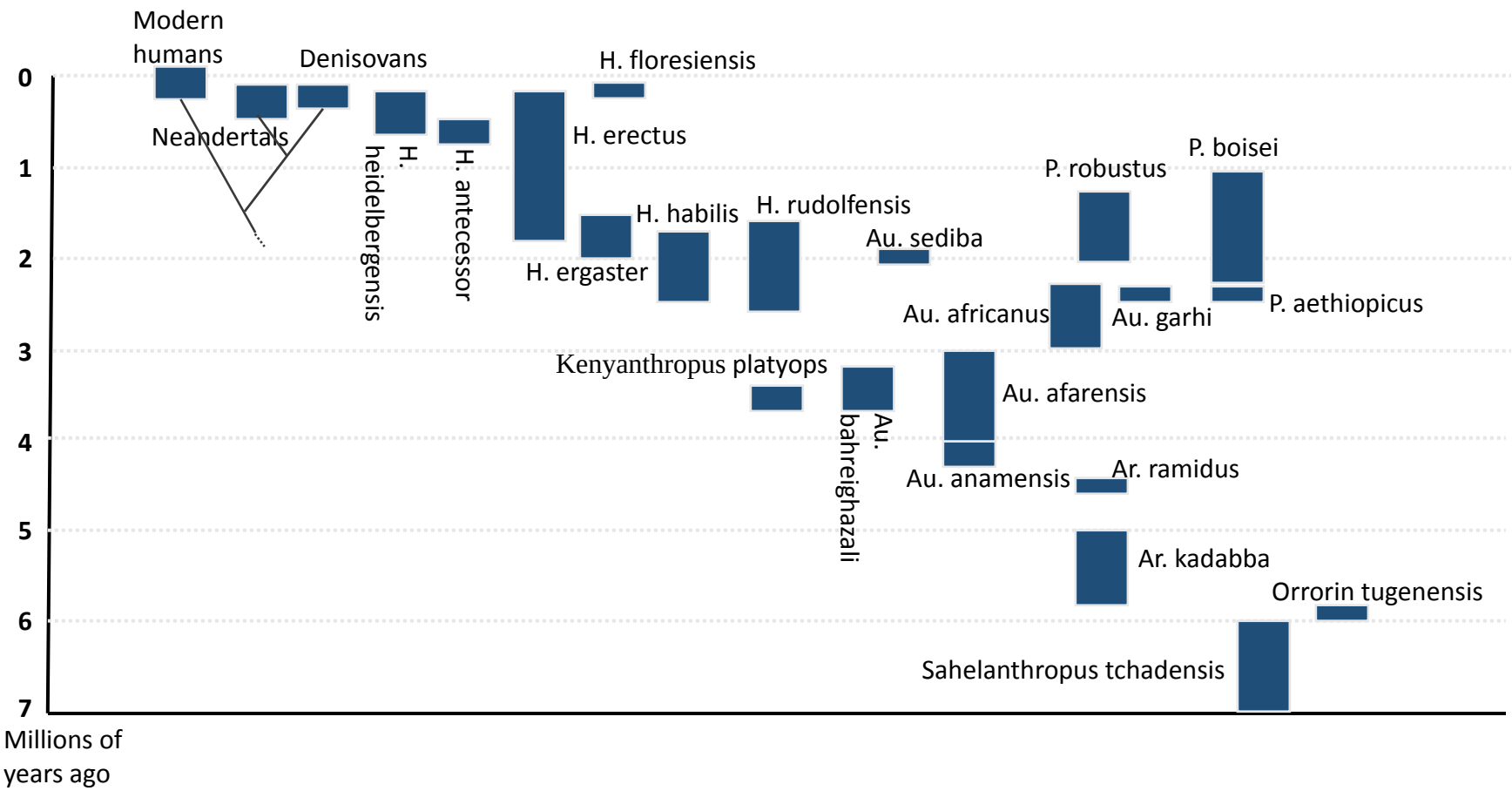

Figure 2. A depiction of hominin taxonomy. Relationships are only indicated if they are based on DNA evidence. Modified from (26). 'H.' refers to *Homo*, 'Au.' to *Australopithecus*, 'P.' to *Paranthropus*, and 'Ar.' to *Ardipithecus*.

3.2 Insights into hominin evolution using ancient DNA

While DNA from present-day humans is able to answer many questions about human evolution, far more questions can be answered if we can compare our genomes to the genomes of our closest known relatives, the Neandertals, as well as possible other unknown relatives. The field of ancient DNA was born with the sequencing of a few base pairs (bps) of an extinct zebra, the quagga (48), and an Egyptian mummy (49). The field of ancient DNA possesses many hurdles though. After death, cells rupture and DNA is no longer protected from the environment (50). Thus, the DNA that is extractable from the bones and teeth, is only present in a very small amount (51). In addition, the amount of DNA that is endogenous to the animal often represents less than 1% of the already small amount of DNA retrievable (51). Therefore any contamination from a present-day source will overwhelm the small amounts of DNA present, and can cause considerable problems if the contaminating DNA is closely related to the endogenous DNA, as is the case with present-day humans and Neandertals. Since the DNA has been exposed to the environment, it accumulates damage from the removal of the amine groups off of bases adenine (A), guanine (G) and cytosine (C) to produce hypoxanthine, xanthine and uracil (U) respectively, a process called deamination (52). The deamination of C to U is particularly relevant for this thesis. Due to uracils being read as a thymine (T) after sequencing, this form of deamination is read as a change in the sequence from a C (in the original, pre-damaged, fragment) to a T in the sequenced and damaged fragment. I refer to this form of damage as C to T changes in this thesis. C to T changes are the most visible of the types of deamination in ancient DNA sequences, and cluster at the ends of fragments (53, 54). Over ten percent of the Cs at fragment ends have been deaminated in samples older than 500 years, while Neandertals have at least 20 percent of their Cs deaminated (55). Thus, cytosine deamination has been used as a criterion for DNA authenticity (56, 57).

Despite these hurdles, it was still possible to sequence DNA from the hypervariable region of the 16,500 bp circular genome of the mitochondria of the Neandertal type specimen from Germany (58). Since each cell contains up to thousands of mitochondria, and therefore thousands of copies of the mitochondrial DNA (mtDNA), mtDNA is often a better target in ancient DNA than the >3 billion bp nuclear genome, of which only two copies per cell exist. These first sequences of the Neandertal were enough to conclude that Neandertal mtDNA falls outside the variation of human

mtDNA (58). Sequencing technologies during these times were very expensive, and therefore the sequencing of large amounts of ancient DNA was seen as unrealistic. The advent of high-throughput sequencing technologies allowed for much cheaper sequencing and for the complete mtDNA genome of a Neandertal from Vindija Cave, Croatia to be sequenced in 2008, which confirmed that this Neandertal mtDNA falls outside the variation of present-day humans, and that the Neandertal mtDNA diverged from modern humans 660,000 years ago (59). In 2010, the nuclear genomes of three Neandertals, also from Vindija Cave, were sequenced to low coverage (combined, each base was on average covered 1.5 times, referred henceforth as 1.5-fold coverage) (11). This draft genome of the Neandertal revealed that the humans leaving Africa did not completely replace the Neandertals they encountered, and instead admixed with them, as all non-Africans have a Neandertal ancestry of 1-4%. The draft genome confirmed the relationship shown in the mtDNA genome between present-day humans and the Neandertal (11).

In 2010, a small piece of a finger phalanx, belonging to a young child, was found in Denisova Cave in the northwest Altai Mountains in Siberia. The phalanx is called *Denisova 3*. The mtDNA genome of this finger bone showed a divergence twice as deep as the divergence between present-day humans and Neandertals (7). Due to the excellent DNA preservation in the bone, a low-coverage nuclear genome of 1.9-fold coverage quickly followed, and revealed that *Denisova 3* belonged to a girl from a population of hominins that are a sister group to Neandertals, subsequently named Denisovans (9). Denisovans also admixed with early modern humans who left Africa, but unlike the Neandertal admixture signal seen in all non-Africans, Denisovan admixture is seen in present-day humans living in Oceania (Australia, Melanesia and the Philippines) (60). A large third molar, *Denisova 4*, was found in Denisova Cave in 2000. The mtDNA from this tooth has two differences (out of 16,595 positions) to the mtDNA of *Denisova 3* (9).

Further advances in methodology, specifically the development of a new method in the library preparation of ancient DNA for sequencing, called the single-stranded library method, allowed for the production of a 30-fold nuclear genome of *Denisova 3* (8, 61). The high coverage of this genome means that the genome is of high quality, on par with the quality of genomes from present-day humans (8). In 2008, a further phalanx, this time from a toe from an adult, was found in Denisova Cave and again revealed excellent DNA preservation. A 50-fold high quality nuclear genome was produced in 2013, and showed that this toe phalanx belonged to a Neandertal (10).

The toe phalanx is officially called *Denisova 5*, however I refer to it as *Altai 1* in this thesis to avoid confusion.

The presence of two high-quality genomes, one each from a Neandertal and a Denisovan, as well as four low-coverage genomes from other Neandertals (the three Vindija Neandertals mentioned earlier and a 0.5-fold genome from *Mezmaiskaya1* from Mezmaiskaya Cave in the Caucasus of Russia (10, 62)) allow for a unique insight into hominin evolution. *Altai 1* is 1.3 times older than *Denisova 3*. Early modern humans split from the ancestor of Denisovans and Neandertals 1.4 times earlier than the split between Denisovans and Neandertals and six times earlier than the earliest split within modern humans (10). Denisovans and Neandertals had low levels of heterozygosity and thus small population sizes (about 30% of the heterozygosity in non-Africans), with long stretches of homozygosity in *Altai 1*, indicative of very recent inbreeding (10). Using the *Altai 1* genome, the admixture signal in non-Africans could be narrowed down to 1.5-2.1%, as well as to the introgressing Neandertal population, which is closest related to the *Mezmaiskaya1* Neandertal (10). A small amount of Denisovan admixture (~0.2%) was found in mainland Asian and Native American populations. *Denisova 3* has at least 0.5% admixture from Neandertals, specifically a Neandertal that was more closely related to *Altai 1* than *Mezmaiskaya1*, as well as 0.5-8.0% admixture from an unknown archaic hominin that split 1-4 million years ago from other hominins. The introgressing Denisovan population split from *Denisova 3* 3.5 times earlier than the introgressing Neandertal split from *Altai 1* (10).

3.3 The Altai Mountains and Denisova Cave

The Altai Mountains are a continuous mountain range that span from the borders of Russia, Mongolia, China and Kazakhstan to the southern tip of Lake Baikal. The north-west portion of the Altai, present in Siberia, make up 'Gorny-Altai' and are mostly foothills and lower mountains, which are divided by longitudinal and narrow river valleys (63) (Figure 3). Today this region has cool summers (+15°C average) and cold winters (-15°C average) (63). From 130-120kya, the climate in Gorny Altai was warmer than today and humid (2, 63) with widespread forests of pine, birch, spruce, cedar and broad-leafed trees (2). Over time the climate became slightly cooler and dryer (63), until 30kya when it became much cooler, which caused the forest to retreat and more and more steppe and meadows to appear (2). The climate in the Altai Mountains was buffered from the extreme temperature changes seen the surrounding low lands making the mountains a possible refugia for animals (63).

Hominins arrived in the Altai Mountains by at least 120kya (63), and left behind evidence of their presence in sites such as Ust-Karakol-1, Okladnikov cave and Anui-2, through the production of stone tools (lithics) (2) (Figure 3). However, the assemblages of lithics in the Altai Mountains may indicate sporadic occupations and high mobility (63). The caves in the region show a high frequency of micromammal and carnivore activity, which again argues for low-intensity human occupation (2, 63).

Denisova Cave is located at 51°22'50" N 84°41'20" E, 28 meters above the Anui River (2, 3). The cave was first excavated by Nikolai Ovodov in 1977 (3) and has been an active site of excavation since 1984, led by Anatoly Derevianko (2). It is a large cave made up of one main chamber and two galleries (East and South) (Figure 4). The main chamber and the cave entrance have been excavated, while the East Gallery is an active excavation site since 2005 (Figure 4) and the South Gallery has been excavated periodically. A chimney is located above the main chamber and and the cave is cold and damp, even in the warm summer months. It is unclear when the chimney appeared. There are over 8000 bone pieces excavated from layers 7-22 in the Main Chamber alone (3). Based on a sample of 116 of these bones, almost 3/4ths of these are morphologically indeterminable and small (5.2cm mean size) (3). Over 95% of the bones show peri-mortem damage, often in the form of marked bone processing or size reduction from humans or animals (3). Cut marks on bones are rare, which is in line with the indication that the region had low-intensity human occupation (63). The bones in the cave encompass 27 taxa of

large animals, including hyena, wolf, fox, cave bear, ibex, horse, deer, wooly rhinoceros, cave bear, yak, bison and saiga antelope (2). Denisova Cave shows a higher lithic assemblage than other excavations in the Gorny Altai region, including possible longer-term tool production, based on calculations from layer 12 in the main chamber, which calculates 250 artifacts over 1000 years for layer 12 in the Main Chamber (63).

During the excavation of the South Gallery in 2000, *Denisova 4* was found in layer 11.1 in the southern tip of the Gallery. The nuclear DNA extracted from this tooth will be discussed in this thesis. Unfortunately the stratigraphy of the South gallery is not published in detail. The East Gallery has produced four hominin remains (Figure 5), including *Altai 1* from layer 11.4 and *Denisova 3* from layer 11.2. In addition *Denisova 8*, another large third molar, was found on the border of layer 11.4 and layer 12 in 2010. During the excavation of the summer of 2011, a fourth hominin remain was found, an intact finger phalanx from an adult, *Denisova 10* (called *Altai 2* in this thesis), found in layer 12. Both mtDNA and nuclear DNA sequences from *Denisova 8* and *Altai 2* will be discussed in this thesis. The East Gallery shows significant disturbance in layer 11 (Figure 5; 'dist' in red), likely caused by water from the chimney, Pleistocene spotted hyenas digging dens or a combination of these and other factors. It is unclear if this disturbance continues into layer 12, but it may make the stratigraphy unreliable in this gallery.

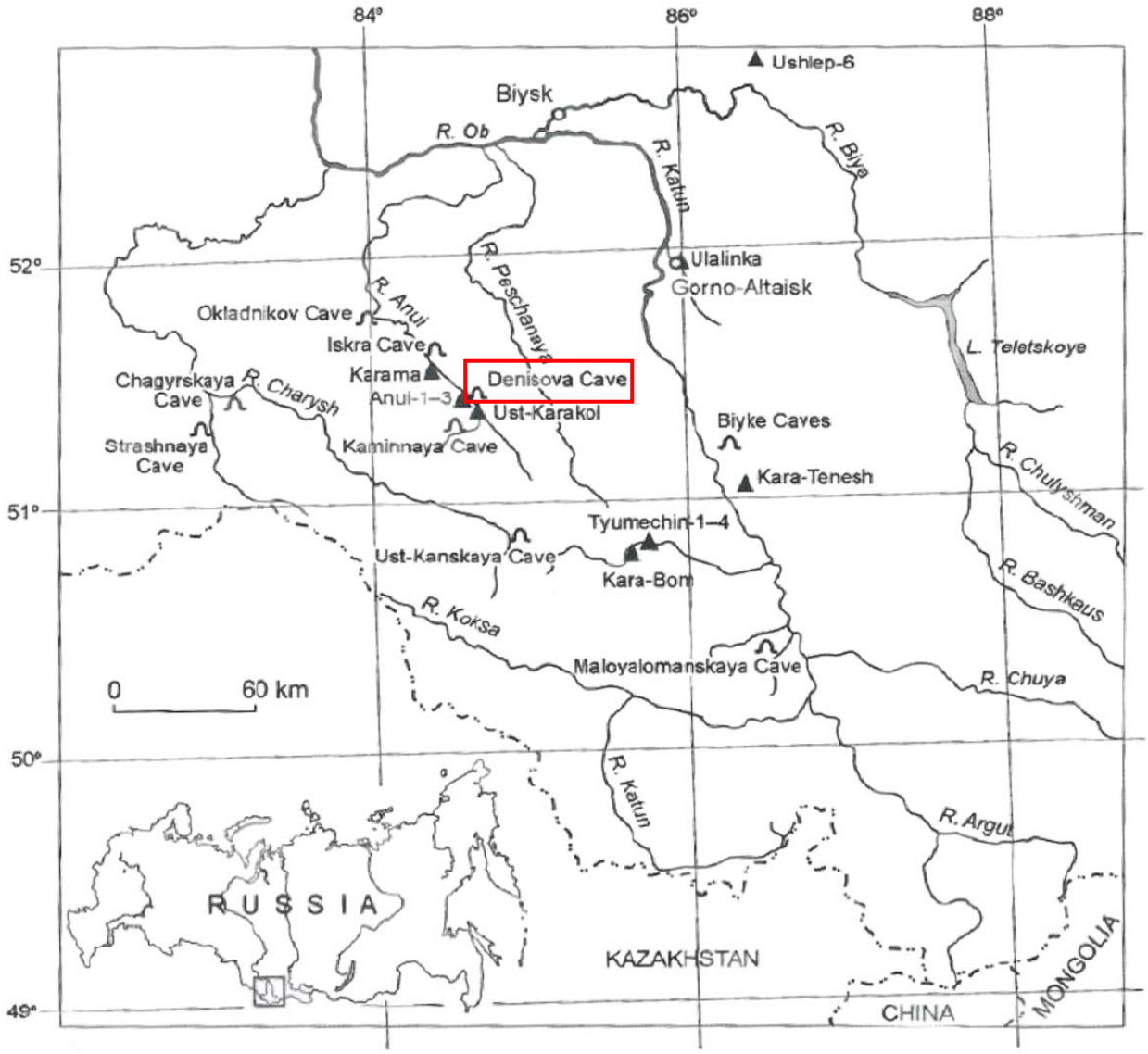

Figure 3. Map of the Altai region (Gorny Altai) of the Siberian Altai Mountains with Middle and Upper Paleolithic sites shown. Denisova Cave is marked with a red square (modified from (1)).

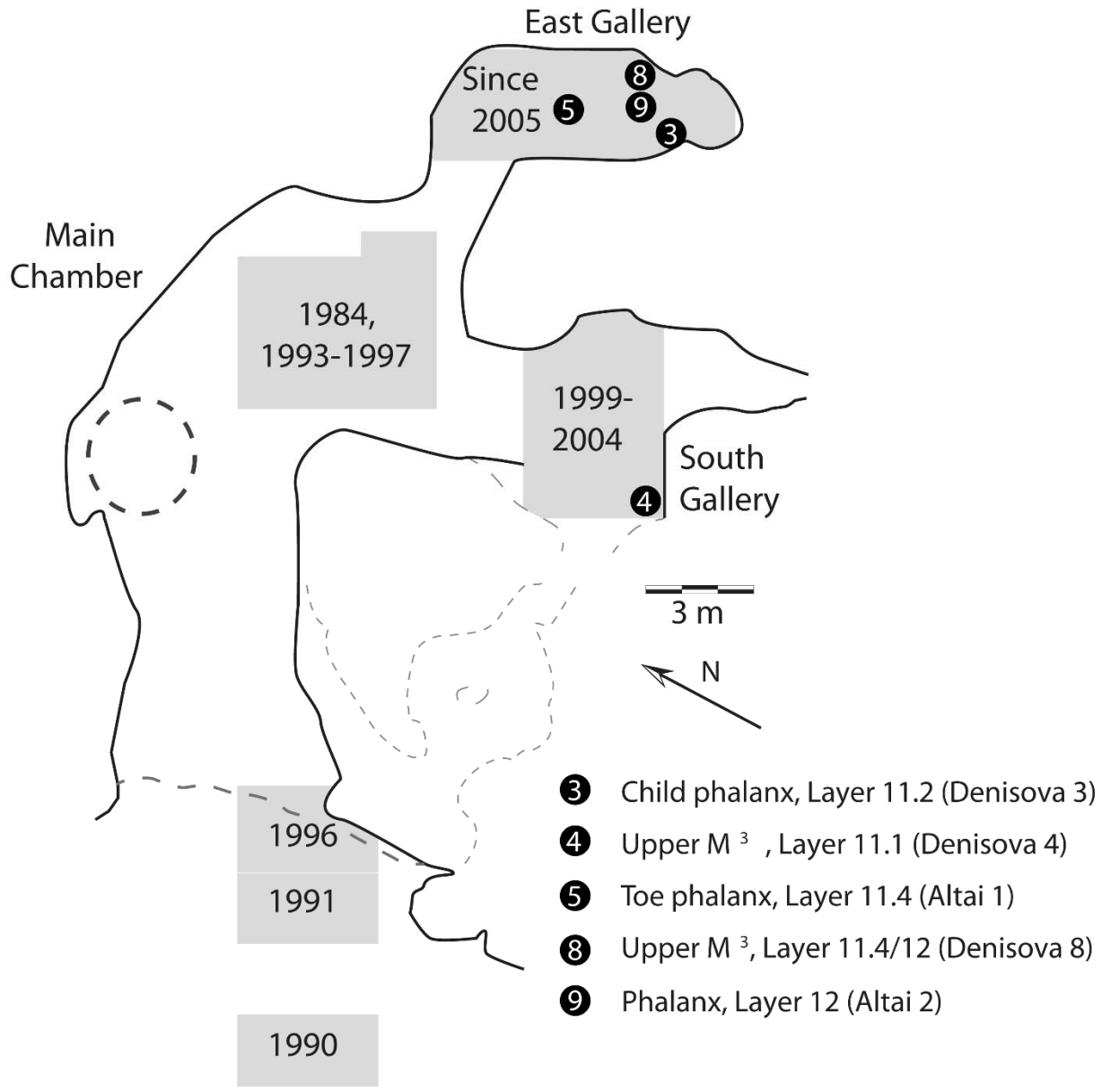

Figure 4. Layout of Denisova Cave (modified from original from Anastasy Abdulmanova and Bence Viola). Excavation sites and years are shown in the gray boxes. The locations of the five individuals discussed in this thesis are shown with their respective Denisova number. The dashed circle shows an approximate location for the chimney in the cave.

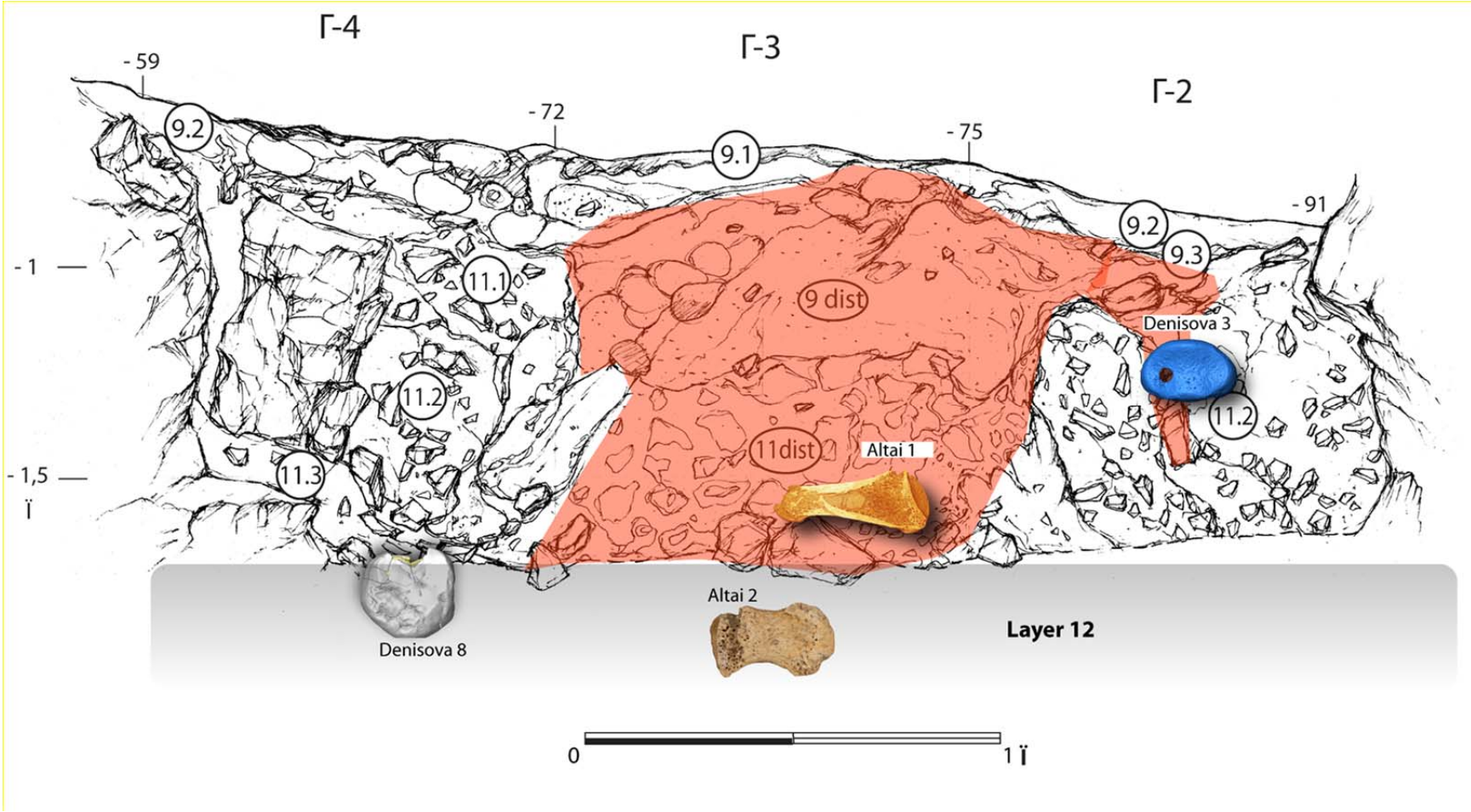

Figure 5. A representation of layer 11 in the East Gallery of Denisova Cave. The four bones/teeth found in the East Gallery that are discussed in this thesis are depicted along with their location in the layer. Layer 12 has no detailed representation, but is indicated. Modified from the original from Bence Viola and Anastasiya Abdulmanova.

3.4 Hyenas

The Hyaenidae family is a remnant of a once prolific family, which at its peak seven to eight million years ago had over 80 species spanning from Africa to Europe and east Asia (64). Today there are four hyena species left. The brown hyena (*Hyaena brunnea*) is a shaggy omnivore, which inhabits southern Africa (65). The striped hyena (*Hyaena hyaena*) looks similar to the brown hyena, but is slightly smaller, less shaggy and has a more distinct striped pattern (65). They inhabit northern Africa and are the only hyena to also inhabit areas outside of Africa, namely the Middle East and India. The aardwolf (*Proteles cristata*) is the smallest type of hyena and is also the only insectivore (65). It lives in southern and eastern Africa. The spotted hyena (*Crocuta crocuta*), which I will focus on in this thesis, is the largest of the hyenas. As indicated by its name, it has a spotted pattern and is a pure meat eater (65). Its range today encompasses much of sub-Saharan Africa (Figure 6). All four types of hyena are nocturnal (65).

Spotted hyenas are extraordinary carnivores. They have jaws and teeth built to crush and digest entire skeletons of large animals (66), except for hair, hooves and horn (3). They move an average of 27 km at night, and move at 10km/hr when searching for prey, but can run at 50km/hr over 0.5-2.5 km when chasing prey (66). They usually hunt in groups, but can also hunt alone (65, 66). Although spotted hyenas are famous for their scavenging behavior, they also hunt between 50-90% of their food depending on prey densities (17, 66). Their clan and territory sizes also range based on prey density, with clans varying in size between 8-80 individuals, and territories varying in size between 10-1000km^2. Spotted hyenas are matrilineal, with females taking on very masculine traits and showing marked aggression (17, 66). They dig extensive dens for their young (3).

The cave hyena of the Pleistocene had a vast range over most of Europe and Asia to Africa, with a northern limit of the 56th parallel (Figure 6) (3, 12). They were larger than the spotted hyena today (12) and were therefore often grouped as a separate species (3, 12, 64). It has been shown in previous studies of cave hyena mtDNA that these hyenas fall into the variation of present-day spotted hyenas (13, 14), and therefore will be referred to as spotted hyenas in this study. Spotted hyenas went extinct in Europe and Asia 13-14kya (3) at the end of the Pleistocene. Older specimens (>400kya) were smaller, about the same size as spotted hyenas today, but they gradually increased in size (12). The origin of the spotted hyenas has been theorized to be in either Asia (12, 14) or Africa (13).

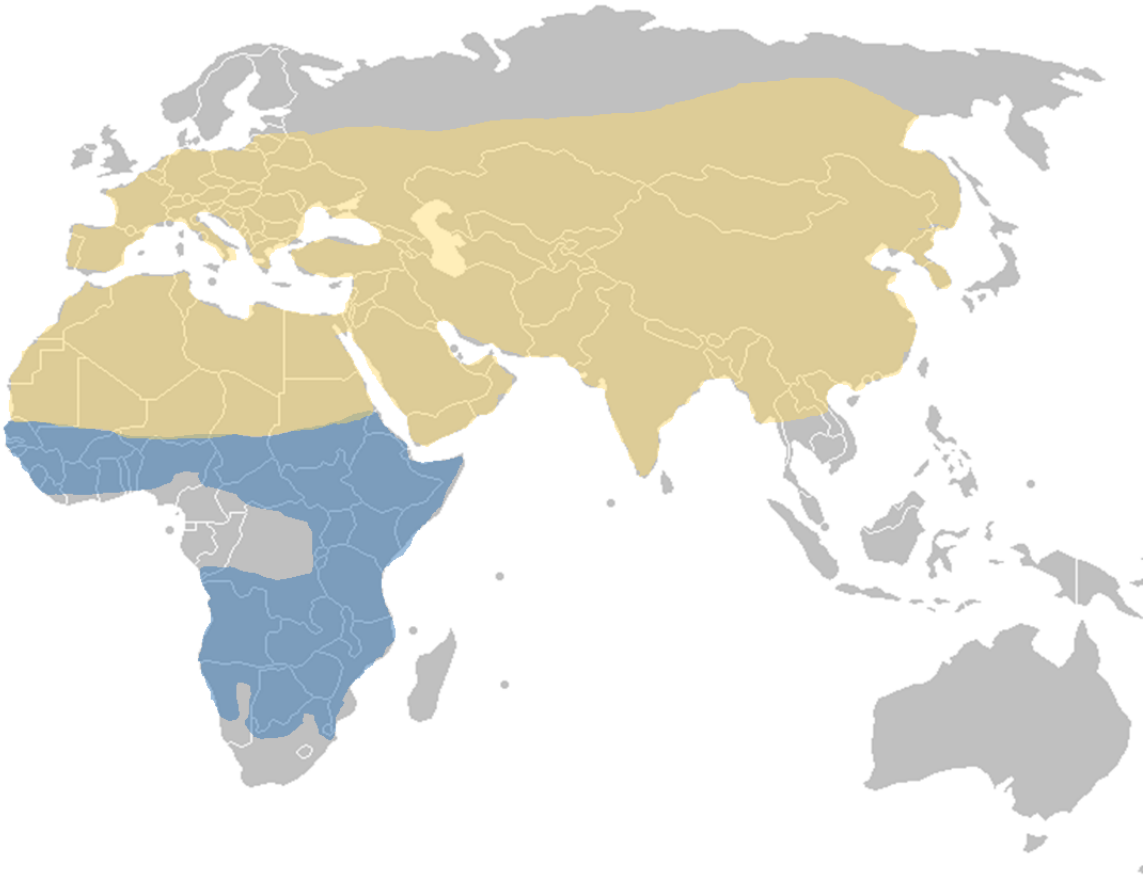

Figure 6. Map of spotted hyena ranges. Current spotted hyena range is shown in blue, the maximum range in the Pleistocene is shown in yellow (after (12)).

4. Materials and Methods

4.1 Nuclear and mitochondrial DNA sequences from two Denisovan individuals

4.1.1 DNA extraction, library preparation, amplification, mtDNA capture, and sequencing

Thirty six millligrams (mg) of dentin were removed from the inside of the enamel cusp of *Denisova 8* using a dentistry drill and used to produce 100 microliters (uL) of extract as described (67). From $1/20^{th}$ of this extract, as well as from $1/10^{th}$ of a previous 100uL extract made from 40mg of *Denisova 4* (9), we produced Illumina libraries, using a single-stranded library preparation protocol that maximizes the yield of sequences from ancient DNA (8). The libraries were treated with *E. coli* Uracil DNA Glycosylase (UDG) and endonuclease VIII to remove uracils (U) (68). UDG does not effectively excise terminal Us (8). The *Denisova 4* library (L9234, see Table 1) had a final volume of 40uL in EBT (10mM Tris-HCl, pH 8.0; 0.05% Tween-20), while *Denisova 8* (B1113) had a final volume of 20uL in EBT.

The concentrations of L9234 and B1113 were measured by qPCR. L9234 from *Denisova 4* was split into two equal parts and used as template for an indexing PCR using two distinct indexing primers per library. The indexing PCR was performed using AccuPrime Pfx DNA polymerase (Life Technologies) and purified with the MinElute purification system as described (8). The purified and indexed libraries were each eluted in 30uL of EB (Qiagin MinElute Kit) to produce L9243 and L9250. An indexing PCR was also performed on B1113 from *Denisova 8* as described above except that all of B1113 was used in one indexing reaction to produce L9108.

To produce larger amounts of amplified library for the mtDNA enrichment, 5□L of L9243 from *Denisova 4* and of L9108 from *Denisova 8* were further amplified with Herculase II Fusion using adapter primers IS5 and IS6 (8, 69), purified with MinElute and eluted into 20□L of EB. DNA concentration was measured on a Nanodrop (ND-1000) and 500ng of the amplified DNA were enriched for human mtDNA via a bead-based protocol where PCR products are sheared, ligated to biotinylated linkers and immobilized on streptavidin-coated beads (70). The enriched libraries were quantified by qPCR and amplified with Herculase II Fusion, taking care not to reach PCR plateau. After measuring DNA concentration on a Bioanalyzer 2100 (Agilent) the *Denisova*

4 capture product (L9320) was sequenced on 1/7th of an Illumina MiSeq lane and the *Denisova 8* capture product (L9126) on 1/10th of an Illumina GAII lane.

For shotgun sequencing, the two libraries from *Denisova 4*, L9243 and L9250 (see Table 1), were amplified with Herculase II Fusion. Molecules with insert sizes between 35 and 450 bp were isolated using gel electrophoresis as described to produce L9349 and L9350 (8). L9108 from *Denisova 8* was also size fractionated to isolate molecules of lengths between 40 and 200 bp using gel electrophoresis without prior amplification to produce L9133. This library was amplified and quantified on the Bioanalyzer 2100 (Agilent) along with L9349 and L9350. The two *Denisova 4* libraries (L9349 and L9350) were pooled in equimolar amounts and sequenced on two Illumina HiSeq 2500 High Output flowcells, while the *Denisova 8* library (L9133) was sequenced on one High Output flowcell.

Table 1. Extraction and library IDs. IDs of *Denisova 4* and *8* after each processing step are given. The *Denisova 4* single-stranded (ss) library was split into two aliquots for the indexing amplification.

	Extract ID	(ss)Lib ID	Lib ID after Indexing	Lib ID after mtDNA capture	Lib ID after gel excision for shotgun seq
Denisova 4	E324	L9234	L9243	L9320	L9349
			L9250	-	L9350
Denisova 8	E652	B1113	L9108	L9126	L9133

4.1.2 Sequence processing and mapping

Ibis v1.1.6 (71) was used for base calling and sequence processing was carried out as described (72). Briefly, after base-calling, reads were demultiplexed allowing a single mismatch in the indexes; Illumina adapters were identified and removed, and overlapping read-pairs merged when the overlap was at least 11 bp. For all sequences the following basic filters were applied:

- Sequences with more than 5 bases with base qualities less than 15 (phred score) were removed

- Sequences having a base with a quality less than 10 (phred score) in the index reads were removed
- Sequences shorter than 35 bp were removed
- PCR duplicates were identified based on the same beginning and end coordinates and collapsed

MtDNA sequences were aligned to the mitochondrial sequence of the high coverage *Denisova 3* phalanx (NC_013993.1) using MIA (parameters: -c, -i) ((73), https://github.com/udo-stenzel/mapping-iterative-assembler) which was also used to generate what approximates a 75% consensus sequence.

The shotgun-sequenced fragments were aligned to hg19 (74) using BWA v.0.5.10 (75) with a maximum edit distance (-n option) of 0.01, a maximum of 2 gap openings (-o 2), and without a seed (-l 16500).

4.1.3 Present-day human mtDNA contamination estimate

We identified 183 and 174 "diagnostic positions" in *Denisova 4* and *Denisova 8, respectively*, where their consensus mtDNA sequences as estimated by MIA differ from every individual in a panel of 311 present-day humans from around the world.

We then re-aligned all captured sequences from the two molars to the human mtDNA reference sequence (76) using BWA version 0.5.10 (75) with relaxed parameters (-n 0.01, -o 2, -l 16500). This allows modern human mtDNA fragments that differ from the Denisovan mtDNA to be identified. Fragments carrying present-day human variants at the diagnostic sites were counted as contaminants, while fragments carrying consensus variants were counted as endogenous. 95% confidence intervals were calculated using a Wilson score interval.

The shotgun sequences were aligned to the human mtDNA reference sequence as described above, and, using the same diagnostic positions as above, mtDNA contamination estimated for the shotgun data.

4.1.4 Present-day human nuclear contamination estimate

To estimate present-day human contamination in the nuclear sequence data, we calculated the divergences of two French individuals to each other as well as two Sardinian individuals to each other (see Figure 13 for explanation of divergence calculation) and used these divergences as a

hypothetical contamination of 100% (*c,* Figure 13). Similarly, we used the divergence of the *Denisova 3* phalanx sequences to the four Europeans as a proxy for 0% contamination (*a*, Figure 13). We then calculated the divergence of *Denisova 4* and *Denisova 8* to the French and Sardinians using sequences that had not been filtered for a terminal C to T change (*b*, Figure 13). The percent contamination in the *Denisova 4* and *Denisova 8* sequences were then calculated as (a-b/a-c)x100.

4.1.5 C to T substitutions and aDNA authenticity

To determine whether different populations of molecules that differ in their extent of cytosine deamination-induced C to T substitutions occur in the libraries, we calculated the apparent C to T substitution rate at the 5'- and 3'-ends of DNA fragments. We then calculated the 5' C to T rate of fragments that have a 3' C to T and *vice versa*. Since deamination-induced misincorporations are rare in modern DNA that contaminates ancient DNA preparations (55, 56), it is unlikely that such DNA fragments carry C to T changes on both ends. In contrast, DNA molecules that carry a C to T change at one end are likely to be ancient and the C to T rate at the other end of such molecules can thus be taken to approximate the deamination rate in ancient, endogenous molecules (under the assumption that deamination at the two ends of molecules is independent). By comparing the C to T rates of all sequences to those that carry C to T at one end we can thus gauge if two or more populations of molecules that differ in their rates of deamination occur in the libraries and thus if contamination may exist in a library. 95% CIs were calculated using Wilson score intervals. Although this approach may be affected by factors that we do not fully understand, it yields contamination estimates for *Denisova 4* of 54-69% and 1.3-6.1% for *Denisova 8* (Table 6) which are qualitatively compatible with ones based on divergence above. For the mtDNA the 95% CIs of the C to T rates of the two populations of molecules overlap (Table 6).

4.1.6 Sex determination

For sex determination, we used sequences that passed the filters described in section 4.1.3 have a minimum map quality of 37 (phred scale).

We identified regions on the sex chromosomes that are $\geq$500 bps long and pass the mappability filter. The mappability filter removes positions where at least one overlapping window of 35bp length maps to a different position in the genome with up to one mismatch (10). On the Y-chromosome we in addition excluded positions that overlap with sequences from four females

from the 1000 Genomes Project (NA12878, NA12892, NA19240, NA19238) (10). This left us with 627,426 bp on the Y chromosome and 40,661,238 bp on the X chromosome.

The number of sequenced fragments expected to fall in these regions if the individuals were male is: (Number of fragments aligned to the whole genome) × (the number positions in the X or Y-chromosome) / (genome size), where genome size is: 2 × (autosomal positions) + (X-chromosomal positions) + (Y-chromosomal positions).

We then determined the number of fragments that actually fall within these regions using either (*i*) all fragments or (*ii*) only those that carry putative deamination-induced C to T substitutions. We determined if the observed and expected numbers are significantly different from the male expectation using a Chi-square test (chisq.test) in the R package 3.1.0 (77). For the X-chromosomal fragments carrying C to T substitutions, we also determined if there is a significant difference under the female expectation. Both *Denisova 4* and *8* are more likely to come from males than from females. See Table 7.

4.1.7 Sex chromosome present-day human contamination estimate

Because the molars come from male individuals, we can estimate the fraction of fragments due to female contamination using the number of "extra" fragments mapped to the X-chromosome relative to the expected number if the individual is male and all Y-chromosome fragments are assumed to be endogenous. The contamination rate is then the difference between the number of fragments mapped to the X chromosome and the number expected if the individual is male divided by number expected if the individual is male. A Wilson score interval was used to calculate 95% CIs.

4.1.8 mtDNA phylogenetics

. The mtDNA sequences of the three Denisovan individuals, seven Neandertals (Altai – KC879692, *Mezmaiskaya1* – FM865411.1, Feldhofer 1 – FM865407.1, Feldhofer 2 – FM865408.1, Vindija 33.16 – AM948965, Vindija 33.25 – FM865410.1 and Sidron 1253 – FM865409.1) (10, 73), five present-day humans (San – AF347008, Yoruba – AF347014, Han Chinese – AF346972, French – AF346981 and Papuan – AF347004) (78) and the chimpanzee (X93335.1) (79) were aligned using the software MAFFT v6.708b (80, 81). Pairwise mtDNA differences among the seven Neandertals and three Denisovans were calculated using MEGA 6.06 (82). In addition, the three Denisovan

mtDNAs were aligned with 311 modern human mtDNAs and the pairwise differences among these individuals were calculated.

To estimate phylogenetic relationships, Modeltest 3.7 (83) was used to identify an appropriate substitution model (GTR+G+I) and MrBayes 3.2 (84, 85) was run with default MCMC parameters for 5,000,000 generations, sampling every 1,000 generations, using a burn-in of 1,000,000 generations. The 4,000 resulting trees were combined to a consensus using TreeAnnotator v1.6.2 from the BEAST package (86) (Figure 15A).

A tree including the partial mtDNA sequence of a hominid from Sima de los Huesos, Spain (KF683087.1) (87) was estimated as above (Figure 16).

4.1.9 mtDNA dating

The most recent common ancestor (MRCA) of the three Denisovans was estimated using parsimony and a Yoruba mtDNA (AF347014). There were two positions where the MRCA was not resolvable. The MRCA of the seven Neandertals was calculated in the same way, with five unresolvable positions. The pairwise differences between the MRCAs and each individual were then calculated (Table 10).

We estimated the age of the two molars and the divergence times between the three Denisovans, five radiocarbon-dated Neandertals (18), ten radiocarbon-dated ancient modern humans (88) and the five present-day humans used for tree estimations (Fig. 2) using BEAST v1.6.2. The age of *Denisova 3* date was set to either 50,000 years or 100,000 years as in ref. (10). A strict as well as a relaxed uncorrelated lognormal molecular clock was used with a normally distributed substitution rate prior of 2.67×10^{-8} per site per year (88) (standard deviation 1.0×10^{-8}), a Bayesian skyline coalescent tree prior with a uniform population size prior of 1,000 to 1,000,000 individuals, and a TN93 substitution model (89) . MCMC runs were carried out for 100,000,000 generations, sampling every 10,000 generations, with a burn-in of 10,000,000 generations. As expected, the relaxed clock is a better fit to the data and was used for the estimates presented in Table 12.

4.1.10 Watterson's estimator θw

θw was calculated for the three Denisovan individuals and the seven Neandertal, 31 Europeans (Italians, Germans, Spanish, Saami, English, Dutch, Finnish and French) and 311 present-day humans (including the Europeans) (Table 11). θw was calculated as follows: $K/a_n/16{,}595$, where

K is the number of segregating sites, and a_n is $\sum_{i=1}^{n-1} \frac{1}{i}$. The numbers of segregating sites were ascertained using DNA Sequence Polymorphism (DnaSP) version 5.10.01 (90).

4.1.11 Autosomal data filtering

The following filters were implemented for the *Denisova 4* and *Denisova 8* autosomal analyses:

- Filters outlined in section 4.1.3
- A minimum map quality of 37 (PHRED scale)
- Base quality set to 2 (phred scale) for Ts at the first or last two positions of fragments (to avoid errors induced by cytosine deamination)
- A minimum base quality of 30 (PHRED scale) (results in removal thymines with low base quality from step above)
- mapability filter that retains all positions where all possible overlapping 35-mers do not have match elsewhere in the genome allowing for one mismatch (10)
- Removal of triallelic sites
- Removal of CpG sites if the CpG occurs in either human, chimpanzee, gorilla or orangutan
- Removal of sites with a coverage higher than 2-fold
- When estimating nucleotide misincorporations due to cytosine deamination positions where the human reference (hg19) carries a C but one or more present-day human from the 1000 Genomes carries a T were excluded.

For high-coverage genomes the following filters were used:

- mapability filter that retains all positions where all possible overlapping 35-mers do not have match elsewhere in the genome allowing for one mismatch (10)
- Root mean square of the map quality >= 30
- Coverage cut-off of 2.5% on each side of the coverage distribution; corrected for GC content for the *Denisova 3* and the Altai Neandertal (10)

4.1.12 Autosomal divergence calculation

We estimate the divergence for *Denisova 4* and *Denisova 8* to ten present-day humans (French - HGDP00521, Sardinian - HGDP00665, Han - HGDP00778, Dai - HGDP01307, Papuan - HGDP00542, Australian - SS6004477, Dinka - DNK02, Mbuti - HGDP0456, Yoruba -

HGDP00927, San - HGDP01029) (8, 10)), the high-coverage *Denisova 3* genome (8) and the high-coverage Altai Neandertal genome (10). The variant call format (VCF) files for the present-day humans as well as the *Denisova 3* and the Altai Neandertal were filtered as stated above.

Divergences between low-coverage and high-coverage genomes are estimated as the percentages of substitutions from the human-chimp ancestor to high-coverage genomes that occurred after the split of the low-coverage genomes from high-coverage genomes (see Figure 17A). Ancestral states for the human-chimpanzee ancestor was taken from the 6-way primate EPO alignments from Ensembl version 69 (genome-wide alignments of human, chimpanzee, gorilla, orangutan, macaque, marmoset) (91, 92) and substitutions were parsimoniously assigned to one of the three lineages. Random alleles were picked at heterozygous sites in the high-coverage genomes while for the low-coverage Denisovan molars a random fragment was picket to represent each site analyzed. Standard errors for the divergence estimates (Table 13-16) were estimated by running 5,000 jackknife replicates of the divergences in 5 Mb windows. Standard errors were multiplied by 1.96 to generate 95% CIs.

We similarly estimated divergences to the high-coverage Altai Neandertal genome (10) for low-coverage data from Vindija Cave, Croatia (Vindija 33.16, Vindija 33.25, Vindija 33.26), from El Sidron Cave, Spain (Sidron 1253), from Feldhofer Cave, Germany (Feldhofer 1) (all available from ERP000119, (11)), and from Mesmaiskaya Cave, Russia (*Mezmaiskaya1*) (10). We excluded regions with a coverage higher than 2-fold for Feldhofer 1, 3-fold for the Vindija Neandertals and 4-fold for the *Mezmaiskaya1* Neandertal. We removed putative deamination-induced C to T substitutions at first and last two positions of the fragments from the *Mezmaiskaya1* Neandertal, as a double-stranded library preparation method and *E. coli* UDG was used, which does not remove uracils efficiently at these positions. For the other low-coverage Neandertals, which were not UDG treated, we removed putative deamination-induced C to T substitutions at the first and last five bases. We calculated the divergence of these six low-coverage Neandertals to the Altai Neandertal along with a 95% CI as above (Table 16).

4.1.13 D-statistics studies on autosomal data

D-statistics (93) were calculated from genotype calls for high-coverage genomes, picking random alleles at heterozygous positions, or from random fragments for low-coverage genomes. Ancestral states were from the EPO alignment (91, 92) (Ensembl v69).

When the low-coverage *Mezmaiskaya1* genome was analyzed together with the high-coverage Altai Neandertal genome, random DNA sequences were picked from both genomes to avoid problems resulting from the difference in sequence quality between the two genomes.

Errors in the low coverage genome sequences contribute apparently derived alleles. To test if derived alleles in DNA sequences determined from *Denisova 8* tend match derived allele in one present-day person more than another, we used *Denisova 8* fragments and asked if derived alleles in *Denisova 8* match derived alleles in one or the other of two individuals from different African populations. This is not the case (D=0.01, Z=0.73).

Table 17 shows that *Denisova 8* tends to share more derived alleles with the Papuan or Australian genomes using all sites (*D*:-0.03 to -0.08, Z-score: -1.9 to -4.3). However, the amount of data limits the power, as can be seen for similar comparisons using the whole high-coverage *Denisova 3* genome (*D*:-0.05 to -0.07, Z-score: -4.2 to -10.1).

To see if the amount of data determined from *Denisova 8* is enough to detect the excess sharing of derived alleles with the Altai relative to the *Mezmaiskaya1* previously described (8), we restrict the analysis to positions in the *Denisova 3* genome covered by the *Denisova 8* fragments and failed to detect the extra sharing (Table 18). As expected from this, we fail to detect any excess sharing of derived alleles between *Denisova 8* and the Altai genome *(*Table 18) when we restricted the analysis to transversions in order to avoid aberrant results due to errors in the low-coverage *Mezmaiskaya1* genome (not shown).

4.2 A Neandertal from Denisova Cave with ancient spotted hyena contamination

4.2.1 DNA extraction and library preparation

Bone powder was obtained by drilling from 3 locations in the *Altai 2* bone (Figure 7). DNA was extracted from bone powder from two of the locations as described in Dabney *et al* (94). Bone powder from the third location was treated with phosphate prior to extraction (15).

Four libraries were prepared from between 10 and 20% of the extracts using single-stranded DNA library preparation with and without UDG treatment. In addition, two libraries were prepared using a previously described U selection method (95).

Library yields were quantified by qPCR. Libraries were amplified and barcoded with two sample-specific indices as described elsewhere using AccuPrime Pfx polymerase (94, 96). For some libraries (L9467, L9366 and L9367) an optimal number of amplification cycles was determined based on the results of qPCR. The other libraries were fully amplified into plateau using 35 PCR cycles. These libraries were amplified for one further cycle to remove heteroduplices since the single melting and hyrbridization allows misaligned sequences to again align correctly. See Figure 9 and Table 2 for more details.

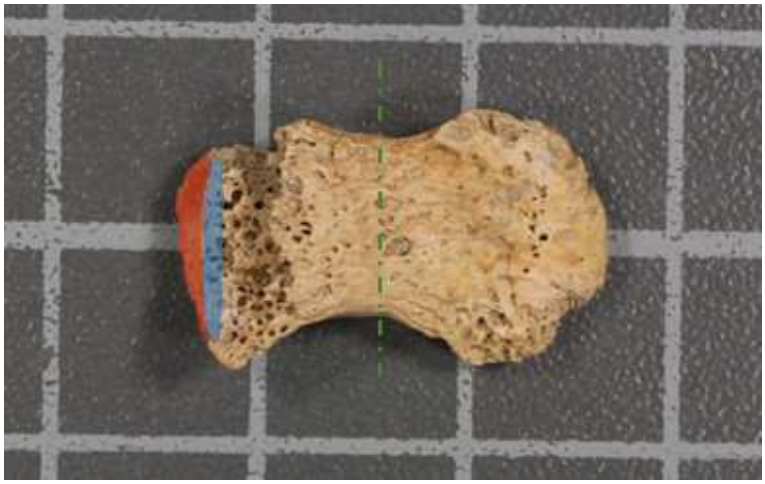

Figure 7. Locations of drilling for SP2990 (*Altai 2*). E1114 came from bone powder drilled from the red area and E1269 came from the blue area. The bone was cut down the middle as indicated by the green dotted line. E3000/E3001 was then drilled from inside the bone after cutting.

The Denisova spotted hyena SP3388 bone fragment was drilled once with a dentistry drill to produce 52 mg of bone powder (Figure 8). All of this bone powder was turned into a DNA extract following the method from Dabney *et al* (94). A single-stranded library was produced from 30% of the extract (61). The entire library was then indexed (96) and amplified into plateau to create library A2396.

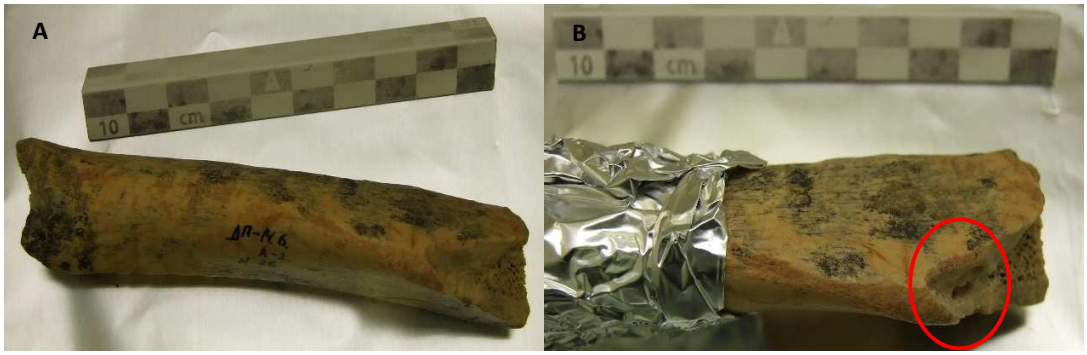

Figure 8. Picture of the Denisova Cave spotted hyena bone, SP3388. **A.** The bone in its entirety. **B.** The bone after drilling: area shown by a red circle.

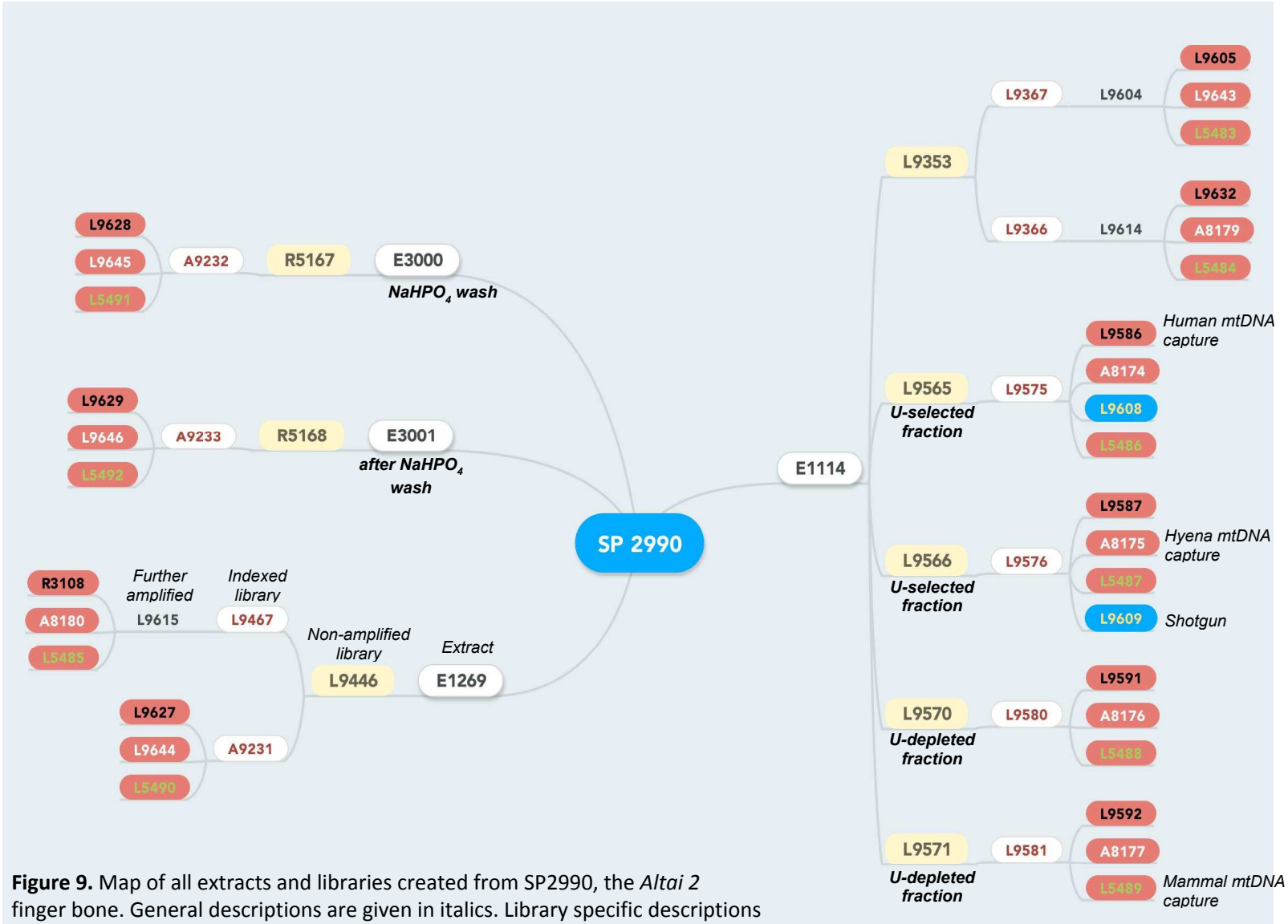

Figure 9. Map of all extracts and libraries created from SP2990, the *Altai 2* finger bone. General descriptions are given in italics. Library specific descriptions are given in bold italics. See Table 2 for more details.

Table 2. Extract and library names with descriptions for *Altai 2*. Input refers to the amount of the previous library/extract that went into the reaction to form the present library/extract. Output refers to the amount the present library was eluted in. Elution was done in either EBT (10mM Tris-HCL, 0.05% tween-20) or TET (1mM EDTA, 10mM tris-HCL, 0.05% tween-20). See Figure 9 for a map of how the libraries/extracts are related to each other.

Name	Description	Input	Output
E1114	Regular Extract	30 mg bone powder	100 uL
E1269	Regular Extract	23 mg bone powder	50 uL
E3000	Extract; $NaHPO_4$ wash	20 mg bone powder	50 uL
E3001	Extract; regular after $NaHPO_4$ wash	Same 20 mg as E3000	50 uL
A8174	Spotted spotted hyena mtDNA capture	1 ug	20 uL
A8175	Spotted spotted hyena mtDNA capture	1 ug	20 uL
A8177	Spotted spotted hyena mtDNA capture	1 ug	20 uL
A8178	Spotted spotted hyena mtDNA capture	1 ug	20 uL
A8179	Spotted spotted hyena mtDNA capture	1 ug	20 uL
A8180	Spotted spotted hyena mtDNA capture	1 ug	20 uL
A9231	Indexing PCR, 35 cycles	20 uL	30 uL
A9232	Indexing PCR, 35 cycles	50 uL	30 uL
A9233	Indexing PCR, 35 cycles	50 uL	30 uL
L9353	ssLIb prep, UDG treated	10 uL	40 uL
L9366	Indexing PCR, 8 cycles	20 uL	30 uL
L9367	Indexing PCR, 8 cycles	20 uL	30 uL
L9446	ssLIb prep, UDG treated	5 uL	40 uL
L9467	Indexing PCR, 9 cycles	20 uL	30 uL
L9486	Herculase amp, 10 cycles	5 uL	20 uL
L9521	Human mtDNA capture	1 ug	20 uL
L9565	ssLib prep, u-selection, u-selected fraction	15 uL	50 uL
L9566	ssLib prep, u-selection, u-selected fraction	15 uL	50 uL
L9570	ssLib prep, u-selection, u-depted fraction	15 uL	50 uL
L9571	ssLib prep, u-selection, u-depted fraction	15 uL	50 uL
L9575	Indexing PCR, 35 cycles	50 uL	40 uL
L9576	Indexing PCR, 35 cycles	50 uL	40 uL

L9580	Indexing PCR, 35 cycles	50 uL	40 uL
L9581	Indexing PCR, 35 cycles	50 uL	40 uL
L9586	Human mtDNA capture	2 ug	20 uL
L9587	Human mtDNA capture	2 ug	20 uL
L9591	Human mtDNA capture	2 ug	20 uL
L9592	Human mtDNA capture	2 ug	20 uL
L9604	Herculase amp, 15 cycles	3 uL	20 uL
L9605	Human mtDNA capture	1 ug	10 uL
L9608	Herculase amp, 1 cycle	1 uL	15 uL
L9609	Herculase amp, 1 cycle	1 uL	15 uL
L9614	Herculase amp, 20 cycles	3 uL	15 uL
L9615	Herculase amp, 20 cycles	3 uL	15 uL
L9627	Human mtDNA capture	1 ug	20 uL
L9628	Human mtDNA capture	1 ug	20 uL
L9629	Human mtDNA capture	1 ug	20 uL
L9632	Human mtDNA capture	1 ug	20 uL
L9643	Spotted spotted hyena mtDNA capture	1 ug	20 uL
L9644	Spotted spotted hyena mtDNA capture	1 ug	20 uL
L9645	Spotted spotted hyena mtDNA capture	1 ug	20 uL
L9646	Spotted spotted hyena mtDNA capture	1 ug	20 uL
L5483	All-mammalian mtDNA capture	1 ug	20 uL
L5484	All-mammalian mtDNA capture	1 ug	20 uL
L5485	All-mammalian mtDNA capture	1 ug	20 uL
L5486	All-mammalian mtDNA capture	1 ug	20 uL
L5487	All-mammalian mtDNA capture	1 ug	20 uL
L5488	All-mammalian mtDNA capture	1 ug	20 uL
L5489	All-mammalian mtDNA capture	1 ug	20 uL
L5490	All-mammalian mtDNA capture	1 ug	20 uL
L5491	All-mammalian mtDNA capture	1 ug	20 uL
L5492	All-mammalian mtDNA capture	1 ug	20 uL
R5167	ssLib prep, no UDG	10 uL	50 uL
R5168	ssLib prep, no UDG	10 uL	50 uL

4.2.2 MtDNA capture and sequencing

For the *Altai 2* finger bone, each of the ten indexed libraries, or final amplified offspring thereof, were captured with three probe sets: human mtDNA (rCRS used for design, NC012920 (76)), spotted hyena mtDNA (NC020670 used for design, (97)) and all-mammalian capture probes (242 mammals, (98)). See Figure 9 and Table 2 for more details of which libraries were captured. The capture method is based on the bead-based capture method described in Maricic *et al* (70) with modifications described in (99).

The human, spotted hyena and all-mammalian captures were sequenced on either the MiSeq or HiSeq Rapid platform for 76+7+76+7 cycles.

L9608 and L9609, the non-heteroduplex containing U-selected fractions, were pooled in equimolar amounts and sequenced together on a single rapid HiSeq lane for 76+7+76+7 cycles. The indexed libraries L9580, L9581, A9233, A9232, L9467, A9231, L9367 and L9366 were also sequenced on a Miseq for 76+7+76+7 cycles.

A2396, the amplified and indexed library from SP3388 (the spotted hyena bone from Denisova Cave) was captured using the same spotted hyena mtDNA probes as the *Altai 2* finger bone libraries. The captured library was pooled with other project-unrelated libraries and sequenced on a MiSeq for 76+7+76+7 cycles.

4.2.3 Sequence processing and mapping

Sequence processing:

After sequencing finished, basecalling, adapter trimming and index demultiplexing were performed. Basecalling for MiSeq runs was done with Bustard (Illumina Corp.), while for HiSeq runs freeIbis was used (100). For adapter trimming, Illumina adapters were removed and putative chimeric sequences were flagged as failing quality (leeHom option "—ancientdna, (101)). Sequences were demultiplexed by assigning them to their sample of origin using deML with default quality thresholds (102). Read-pairs were merged if they had an 11 bp overlap.

The following basic filters were used for all sequenced libraries:

Removal of:

1. sequences with a length less than 35 basepairs
2. sequences that do not match the index combinations used to produce each library
3. sequences with more than 5 bases with base qualities less than 15 (phred score)

4. sequences having a base with a quality less than 10 (phred score) in the index reads

Mapping:

L9629, L9628, L9521, L9627, L9605, L9632, L9586, L9587, L9591 and L9592 from the *Altai 2* finger bone were captured using human mtDNA probes (see Figure 9 and Table 2). Each of these libraries were first filtered as described above. They were then aligned to the human rCRS mitochondrial genome (76) as well as the Vindija 33.16 Neandertal mtDNA genome (59) using BWA version 0.5.10 (75) with relaxed parameters (-n 0.01, -o 2, -l 16500). After mapping, the following analyses were conducted for each library: (i) Percent in target: the number of sequences that mapped with a map quality over zero divided by the number of total sequences (including unmapped). Sequences were then filtered for a map quality 37 (phred score) and PCR duplicates were collapsed (collapsing sequences with the same beginning and end coordinates). (ii) Percent unique: the number of duplicate collapsed sequences were divided by the number of mapped sequences before duplicate removal. (iii) Average coverage across the mtDNA of duplicate collapsed sequences.

The ten *Altai 2* libraries captured with spotted hyena probes (L9643-L9646, A8174-A8177, A8179), as well as the SP3388 Denisova Cave spotted hyena library (L5497) were filtered using the basic filters described above. They were then aligned to the spotted hyena mtDNA (NC020670 (97)) using BWA version 0.5.10 (75) with relaxed parameters (-n 0.01, -o 2, -l 16500). After alignment, percent in target, percent unique and average coverage were calculated as for the human mtDNA captures. Thus, sequences with a map quality less than 37 (phred score) and PCR duplicates were removed. After these individual analyses and filters, the ten *Altai 2* libraries were merged into one.

The all-mammalian captures of the *Altai 2* (L5483 to L5492) were filtered using the basic filters described above and then aligned to the 242 mammalian genomes used to make the capture bait as described in Slon *et al* (98). After alignment, sequences were filtered for a map quality of 37 (phred scale) and PCR duplicates were collapsed. These sequences were then aligned to the Nucleotide database of NCBI using the Basic Local Alignment Search Tool (BLAST) and BLAST hits were ranked by taxonomic identification number (98).

4.2.4 Present-day human mtDNA contamination estimate

After alignment to the human rCRS, *Altai 2* sequences from each human mtDNA captured library were compared to 69 human-Neandertal diagnostic positions. These diagnostic positions are positions where ten Neandertals (73, 95, 103) differ from 311 present-day humans from around the world. Sequences that carry the Neandertal allele are deemed clean, while sequences that carry the present-day human allele are deemed contaminating. The analysis was repeated looking only at sequences containing thymine residues at the first and/or last two positions at sites where the rCRS sequence carries cytosine residues.

4.2.5 Spotted hyena contamination estimate

Spotted hyena diagnostic positions were determined as follows. First the spotted hyena (NC020670 (97)) mtDNA was chopped into 100 bp sequences with 1 bp tiling. These 100 bp sequences were then aligned to the human rCRS mtDNA (76) using BWA version 0.5.10 (75) with relaxed parameters (-n 0.01, -o 2, -l 16500). Sequences that mapped with a map quality above 0, were then used to make a consensus, requiring at least 1-fold coverage and 80% consensus support. Sequences mapped in regions 2,888-3,108 and 5,705-5,805 (in rCRS coordinates). The consensus was aligned to the same 311 humans used for the present-day human contamination estimate, as well as three spotted hyenas (97), one striped hyena (97) and one mongoose (NC006835) using MAFFT (81). The alignments were checked by eye in BioEdit v. 7.1.3 (104). In the two regions where the spotted hyena sequences mapped, a diagnostic position was called where the three spotted hyenas differed from all 311 humans. Thus nine diagnostic positions were called where there was a difference between the three spotted hyenas and the 311 humans, striped hyena and mongoose. Spotted hyena contamination was estimated based on these nine diagnostic positions as was done in section 4.2.4.

4.2.6 Assembly of Neandertal mtDNA

As significantly more sequences match the Neandertal state, the alignments to the Altai Neandertal mtDNA from the ten libraries were combined into one file. Due to the high level of contamination, only sequences containing thymine residues at the first and/or last two positions where the reference sequence has a C were used. A consensus was made of the sequences using a cutoff of 80% consensus support and 5-fold coverage.

In order to resolve as many positions as possible, we included two additional filters. We first required that in order for a C to T difference to be used for separating out a deaminated sequence, the minority of sequences at the positions needed to have a T. This aims to exclude the C to T difference existing due to a mutation instead of a deamination event. Second we took out sequences that align better to the spotted hyena (NC020670 (97)) than they do to the Neandertal.

After these filters, seven unresolved positions remain. One N remains in the C-stretch (position 310) due to the difficulty on mapping fragments that begin and end in the C-stretch. This region was resolved by calling a consensus of sequences that span the C-stretch region. Three positions fall into region 185-195 (rCRS coordinates). When examined by eye, two haplotypes are evident (see Table 3). After a megablast of both haplotypes, it became evident that both haplotypes are seen in present-day humans. The last three unresolved positions fall within the 16S and aspartate tRNA (Asn tRNA) genes. The two Ns that fall into the 16S gene (positions 2951 and 2964, rCRS coordinates) are close enough together that they combine to show two distinct haplotypes. After a megablast (105) of both haplotypes, haplotype one is seen in humans, while haplotype two is seen in individuals on the cat branch (cats, spotted hyenas, mongooses, civets) (Table 3). The last unresolved position (5767, rCRS coordinates), in the Asn tRNA gene. After a megablast (105) of sequences carrying both types of positions, one sequence occurs in humans, while the other occurs in members of the Cervid family, namely sheep, goats and deer. Thus the final mtDNA sequence for the *Altai 2* Neandertal has six missing positions.

Table 3. Unresolved positions in the Neandertal alignments on *Altai 2*. The two seqeunces at each position are shown. The unresolved position is shown in bold, while the number of sequences containing each position are shown in italics. The asterix (*) denotes a deletion.

Position	Coverage	Consensus support	Majority base	Base in rCRS	Base in Vindija 33.16	Base in CC8 Crocuta	Sequence 1	Sequence 2
185	24	58.3	G	G	G	A	**G**CGA**A**CATAC**T** present-day human *4 sequences*	**A**CGA**G**CATAC**C** Neandertal and present-day human *7 sequences*
189	26	57.7	G	A	G	A		
195	19	78.9	C	T	T	G		
2951 (16S)	152	74.3	C	C	C	T	**C**TAGAGTCCATATC**A** human, CC8 crocuta *82 sequences*	**T**TAGAGTCCATATC**G** Cat branch, not in CC8 crocuta *28 sequences*
2964 (16S)	152	63.8	A	A	A	A		
5767 (12S)	157	68.2	*	*	*	C	GGCAG*****GTTTG human *84 sequences*	GGCAG**A**GTTTG Deer, sheep, goat *69 sequences*

4.2.7 Assembly of spotted hyena mtDNA

After combining the ten *Altai 2* libraries captured and aligned to the spotted hyena, coverage and consensus support across the spotted hyena mtDNA genome were plotted. The coverage spikes by almost two-fold in areas with high conservation in the mtDNA genome (especially the 16S region, Figure 24A), due to the large amount of hominid DNA present in the bone. To be sure that we reconstruct a spotted hyena mtDNA with no influence from the human sequences, we removed regions that mapped to human mtDNA. This was done by chopping the human rCRS mtDNA (76) into 35-100 bp sequences in increments of 5 bps with 1 bp tiling. Then these sequences were mapped to the spotted hyena mtDNA (NC020670 (97)) using BWA version 0.5.10 (75) with relaxed parameters (-n 0.01, -o 2, -l 16500). Regions where the human sequences mapped to the spotted hyena were removed. After human regions were removed, the extreme coverage peaks disappear (Figure 24B). A consensus was called of the sequences that cover the non-human-mapping regions, by requiring at least 5-fold coverage and a consensus support of 80%.

4.2.8 Spotted hyena mtDNA in Denisova Cave specimins

The unmapped sequences of the high coverage Altai Neandertal (*Altai 1*) (10), high coverage Denisovan (*Denisova 3*) (8), two low-coverage Denisovans (*Denisova 4* and *8*), a high coverage early modern human (Ust-Ishim) (106), a low coverage Neandertal (*Mezmaiskaya1*) (10) and a present-day human were filtered as in section 4.2.3 and then mapped to thespotted hyena (NC020670 (97)), the ringed seal (NC008428, (107)), the cave bear (NC011112, (108)) and the rCRS human (NC012920, (76)) mtDNA using BWA version 0.5.10 (75) with relaxed parameters (-n 0.01, -o 2, -l 16500). After mapping, duplicates were collapsed and sequences with a map quality less than 37 (phred scale) were removed.

Regions of the seal and cave bear mtDNA that do not map human mtDNA were determined as the non-human mapping spotted hyena regions were calculated in section 4.2.7. Sequences that fall completely within these non-human mapping regions were kept for further analyses.

For the *Denisova 3* and 4 individuals as well as the *Altai 1* Neandertal, a consensus of the filtered sequences was called requiring 80% consensus support and at least 3-fold coverage.

The number of sequences that mapped perfectly (no insertions, deletions or mutations) were also calculated. A two-tailed fisher exact test was done in R v3.2.0 (fisher.test, (77)) between each of two individuals and two of the mapped mtDNAs (e.g. present-day human seal and hyena sequences mapped versus *Denisova 4* seal and hyena sequences mapped).

4.2.9 MtDNA phylogenetics (human and spotted hyena)

Human

MtDNAs of seven published Neandertals (10, 73, 103), five present-day humans (San – AF347008, Yoruba – AF347014, Han Chinese – AF346972, French – AF346981 and Papuan – AF347004) (78), one Denisovan (7) and a chimpanzee (X93335.1) (79) were aligned to the *Altai 2* Neandertal consensus sequence using MAFFT (80, 81). A Modeltest was done as described in section 4.1.9, best model: GTR+I+G. A Bayesian tree was then produced using BEAST (86, 109) with the following parameters: GTR+I+G model, uncorrelated log normal clock, set to 2.7e-8 with a standard deviation of 1e-8; a Bayesian skyline tree prior, initial 1000, distribution 0 to 100000; run for 1000000000 generations and sample every 1000. A pairwise comparison of the same mtDNAs was done using Mega6 (82). The number of differences were also calculated to the most common recent ancestor of Neandertals as calculated in section 4.1.10.

Spotted hyena

The complete mtDNAs of three spotted hyenas (NC020670, JF894379 and JF894377, (97)) a striped hyena (NC020669, (97)) and a mongoose (NC006835) were aligned to the consensus sequences of the SP3388 Denisova Cave spotted hyena and the *Altai 2* finger bone using MAFFT as above. A Modeltest was also done as above, result: GTR+I. A Bayesian tree run with BEAST with the same parameters as above except the GTR+I model was used. A pairwise comparison of the same mtDNAs was done using Mega6 (82).

The consensus sequences of the SP338 Denisova Cave spotted hyena and the *Altai 2* finger bone were also aligned to the cytochrome b sequences of 55 spotted hyenas (see Appendix Table 1 for accession numbers and references), two Aardwolves (AY928679 and AY048791, (110)), six brown hyenas (DQ157588-591, AY048789-90, (13, 111)), 18 striped hyenas (DQ157576-DQ157587, AY048788,AY048787,AY928678,AF153055,AF153054,EF107524,NC020669, (13, 97, 110-112)) and one mongoose (same as above) using MAFFT as above. A pairwise

comparison of these mtDNAs was done using Mega6 (82). Based on the pairwise comparison, individuals with the same sequences were collapsed and given letters (see Table 27). For the phylogenetic analysis with BEAST, only the sequences groups were used. Modeltest was run as above, best model: TrN+I. BEAST was run as above but using the TN93 model, which is closest to the TrN+I model suggested.

Spotted hyena in high coverage archaics

The complete mtDNAs of the same three spotted hyenas, striped hyena and mongoose as above were aligned separately to the spotted hyena consensus sequences of the *Denisova 3*, *Denisova 4* and *Altai 1* individuals using MAFFT as above. Modeltest and BEAST were again run as described above. Modeltest suggested the GTR+I model for the alignments including *Denisova 3*, the HKY+I model for alignments including *Denisova 4*, and the TVM+I model for alignments using *Altai 1*. BEAST was run as above, and the TN93+I model was used for the *Altai 1* alignments.

4.2.10 Autosomal divergence calculation

All of the shotgun libraries were filtered using the basic filters described in section 4.2.3 and then aligned to the human genome (74) using BWA version 0.5.10 (75) with relaxed parameters (-n 0.01, -o 2, -l 16500). Each library was then filtered for mapped sequences (sequences with a map quality over 0). The percent endogenous for each library was calculated by dividing the number of mapped sequences by the number of unmapped sequences.

After mapping, the sequences from L9608 and L9609 had PCR duplicates collapsed, were combined and sequences with a map quality less than 37 (phred score) were removed. In addition sequences not containing thymine residues at the first and/or last two positions were removed as described in section 4.1.12. This left 18.4 Mb. Divergence was calculated as described in section 4.1.13 for the *Altai 2* to ten present-day humans, the high coverage *Altai 1* and *Denisova 3* (Figure 22). Divergence of the *Altai 2* to the high coverage *Altai 1*, was also compared to divergences of six published Neandertals to the high coverage Altai Neandertal (Figure 23), using the same values used in section 4.1.13.

Lineage attribution was calculated as described in Meyer *et al* (113). In brief, *Denisova 3*, *Altai 1* and an Mbuti present-day human were used to determine positions that are ancestral or derived when compared to the human-chimpanzee ancestor. Then the *Altai 2* sequences were compared to these positions and counted.

5. Nuclear and mitochondrial DNA sequences from two Denisovan individuals

This manuscript is published at the Proceedings of the National Academy of Sciences of the United States of America, doi: 10.1073/pnas.1519905112

Authors: Susanna Sawyer[a], Gabriel Renaud[a], Bence Viola [b,c,d], Jean-Jacques Hublin [c], Marie-Theres Gansauge [a], Michail V. Shunkov [d,e], Anatoly P. Derevianko [d,f], Kay Prüfer [a], Janet Kelso [a], Svante Pääbo [a].

[a] Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany

[b] Department of Anthropology, University of Toronto, Toronto, ON M5S 2S2, Canada

[c] Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany

[d] Institute of Archaeology and Ethnography, Russian Academy of Sciences, Siberian Branch, Lavrentieva Avenue, 17 Novosibirsk, RU-630090, Russia.

[e] Novosibirsk National Research State University, Pirogova st., 2, Novosibirsk, RU-630090, Russia

[f] Altai State University, Pr. Lenina, 61, Barnaul, RU-656049, Russia

Author contributions:

Contributed materials: Michail Shunkov, Anatoly Derevianko

Performed experiments: Susanna Sawyer and Marie-Theres Gansauge

Analyzed morphological data: Bence Viola

Analyzed mitochondrial data: Susanna Sawyer

Analyzed nuclear data: Susanna Sawyer, Gabriel Renaud, Kay Prüfer, Janet Kelso

Wrote the paper: Susanna Sawyer, Bence Viola and Svante Pääbo

5.1 Abstract

Denisovans, a sister-group of Neandertals, have been described based on a nuclear genome sequence from a finger phalanx (*Denisova 3*) found in Denisova Cave in the Altai Mountains. The only other Denisovan specimen described to date is a molar (*Denisova 4)* found at the same site. This tooth carries a mitochondrial (mt) DNA sequence similar to that of *Denisova 3*. Here we present nuclear DNA sequences from *Denisova 4* and a morphological description, as well as mitochondrial and nuclear DNA sequence data, from another molar (*Denisova 8)* found in Denisova Cave in 2010. This new molar is similar to *Denisova 4* in being very large and lacking traits typical of Neandertals and modern humans. Nuclear DNA sequences from the two molars form a clade with *Denisova 3*. The mtDNA of *Denisova 8* is more diverged and has accumulated fewer substitutions than the mtDNAs of the other two specimens suggesting that Denisovans were present in the cave over an extended period of time. The nuclear DNA sequence diversity among the three Denisovans is comparable to that among six Neandertals but lower than that among present-day humans.

5.2 Introduction

In 2008, a finger phalanx from a child (*Denisova 3*) was found in Denisova Cave in the Altai Mountains in southern Siberia. The mitochondrial (mt) genome shared a common ancestor with present-day human and Neandertal mtDNAs about one million years ago (7), *i.e*. about twice as long ago as the shared ancestor of present-day human and Neandertal mtDNAs. However, the nuclear genome revealed that this individual belonged to a sister group of Neandertals. This group was named Denisovans after the site where the bone was discovered (8, 9). Analysis of the Denisovan genome showed that Denisovans have contributed on the order of five percent of the DNA to the genomes of present-day people in Oceania (8, 9, 60) and about 0.2 percent to the genomes of Native Americans and mainland Asians (10).

In 2010, continued archaeological work in Denisova Cave resulted in the discovery of a toe phalanx (*Denisova 5*), identified based on its genome sequence as Neandertal. The genome sequence allowed detailed analyses of the relationship of Denisovans and Neandertals to each other and to present-day humans. Although divergence times in terms of calendar years are unsure due to uncertainty about the human mutation rate (114), it showed that Denisovan and Neandertal populations split from each other in the order of four times further back in time than the deepest divergence among present-day human populations occurred, while the ancestors of the two archaic groups split from the ancestors of present-day humans in the order

of six times as long ago as present-day populations (10). In addition, a minimum of 0.5 percent of the genome of the *Denisova 3* individual was derived from a Neandertal population more closely related to the Neandertal from Denisova Cave than to Neandertals from more western locations (10).

Although Denisovan remains have, to date, only been recognized in Denisova Cave, the fact that Denisovans contributed DNA to the ancestors of present-day populations across Asia and Oceania suggests that, in addition to the Altai Mountains, they may have lived in other parts of Asia. Besides the finger phalanx, a molar (*Denisova 4*) was found in the cave in 2000. Although less than 0.2% of the DNA in the tooth derive from a hominin source, the mtDNA was sequenced and differed from the finger phalanx mtDNA at only two positions suggesting that it too may be from a Denisovan (8, 9). This molar has several primitive morphological traits different from both late Neandertals and modern humans. In 2010, another molar (*Denisova 8*) was found in Denisova Cave. Here we describe the morphology and mtDNA of *Denisova 8* and present nuclear DNA sequences from both molars.

5.3 Results

5.3.1 *Denisova 8* morphology

Denisova 8. The *Denisova 8* molar (Figure 10) was found at the interface between layers 11.4 and 12 in the East Gallery of Denisova Cave, slightly below the Neandertal toe phalanx (*Denisova 5*, Layer 11.4) and the Denisovan finger (*Denisova 3*, Layer 11.2). Radiocarbon dates for layer 11.2 as well as for the underlying 11.3 layer yield ages over ~50,000 years (OxA-V-2359-16 & -14) (2). *Denisova 8* is thus older than *Denisova 3* which is at least 50,000 years old. It is reassembled from four fragments which fit well together, although a piece of enamel and most of the root is missing (Figure 11B).

The *Denisova 4* molar was found in Layer 11.1 in the South Gallery, a different part of the cave. Radiocarbon dates for layer 11.2 of the South Gallery are over 50,000 years (OxA-V-2359-17 & -18) and 48.6 $\pm$ 2.3 thousand years before present (KIA 25285) (2). Although the lack of direct stratigraphic connection between the different parts of the cave makes relative ages difficult to assess it is likely that *Denisova 4* is younger than *Denisova 8.*

Based on crown shape and the presence of a marked *Crista obliqua*, a feature unique to maxillary molars, we identify *Denisova 8* as an upper molar, despite having five major cusps. The mesial half of the crown is worn, with a small dentine exposure on the protocone, while

there is no wear on the distal part. The lack of a distal interproximal facet indicates that the tooth is a third molar, or a second molar without the eruption of the M^3. Usually, when Neandertal and *H. heidelbergensis* upper M^2s reach wear levels to the extent seen here, the adjacent M^3 is already erupted and an interproximal facet is visible. One possibility is that the *Denisova 8* is a second molar of an individual with M^3 agenesis. Despite being common in modern humans, this is rare in archaic hominins, but it does occur in Asian late *Homo erectus* and Middle Pleistocene hominins.

The previously described *Denisova 4* molar is characterized by its large size, flaring buccal and lingual sides, strong distal tapering and massive and strongly diverging roots (2). Not all of these characteristics can be assessed in *Denisova 8*, but it is clear that it lacks the strong flare of the lingual and buccal surfaces and distal tapering of *Denisova 4*.

The length of *Denisova 8* is more than three standard deviations larger than the means of Neandertal and modern human molars and in the range of Pliocene hominins (Figure 10 and 11). Both *Denisova 8* and *4* are very large compared to Neandertal and early modern human molars, and *Denisova 8* is even larger than *Denisova 4*. Only two Late Pleistocene third molars are comparable in size, those of the inferred early Upper Paleolithic modern human *Oase 2* in Romania and *Obi-Rakhmat 1* in Uzbekistan (42, 115).

The morphology of third molars is variable, and thus not very diagnostic. Nevertheless, Neandertal third molars differ from *Denisova 8* in that they frequently show a reduction or absence of the hypocone, reduction of the metacone and generally lack a continuous *Crista obliqua* (115, 116). This applies also to Middle Pleistocene European hominins which also only rarely show a Cusp 5 (116). The massive and diverging roots of *Denisova 4* are very unlike the root morphology of Neandertals and Middle Pleistocene hominins in Europe. East Asian *Homo erectus* and Middle Pleistocene *Homo* frequently show massive roots similar to *Denisova 4*, but in these groups crown size become strongly reduced starting around one million years ago (117). The recently described Xujiayao teeth from China (118) have massively flaring roots and relatively large and complex crowns, similar to the Denisova teeth, but have reduced hypocones and metacones.

Early and recent modern humans show the most morphological variability of third molars, and there are specimens that have large hypocones, metacones or continuous *Cristae obliquae* (116). The combination of an unreduced metacone and hypocone, continuous *Crista obliqua*, a large fifth cusp, and large over-all size is reminiscent of earlier *Homo*, but *Denisova 8* lacks the multiple distal accessory cusps frequently seen in early *Homo* and Australopithecines.

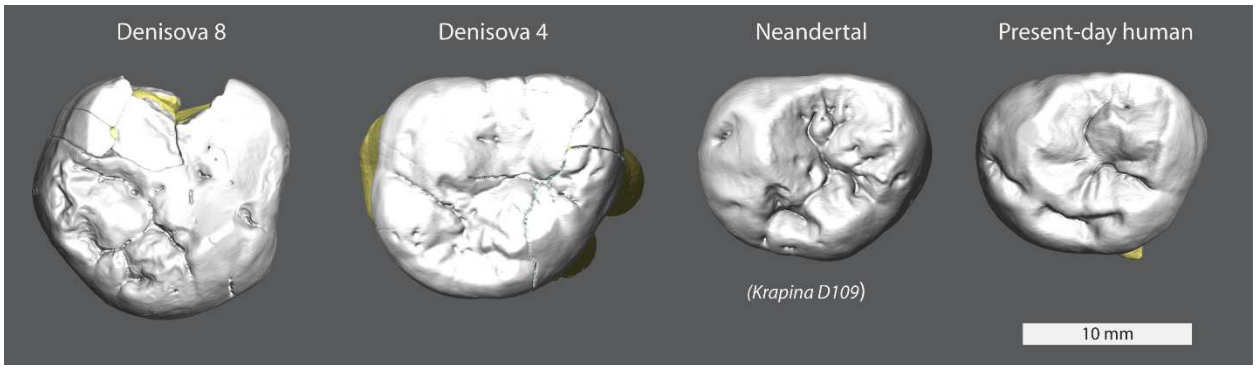

Figure 10. Occlusal surfaces of the *Denisova 4* and *8* molars and third molars of a Neandertal and a present-day European.

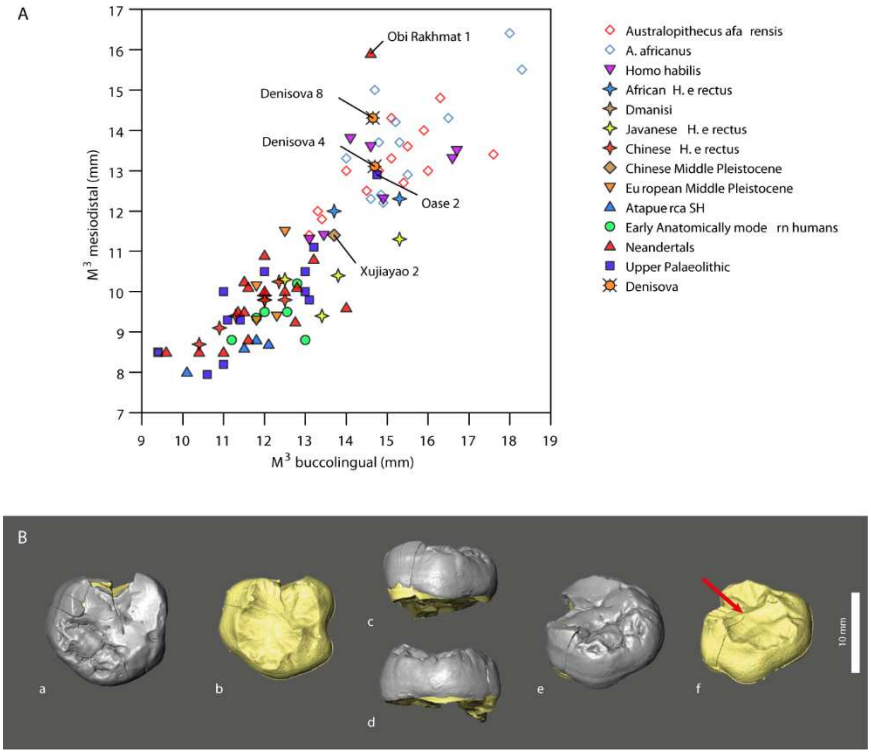

Figure 11. **Morphology of *Denisova 8* molar**. . **a:** occlusal view (surface model from µCT scan); **b:** enamel dentine junction in occlusal view, The arrow indicates the marked *Crista obliqua* on the enamel-dentine junction; **c**: Biplot of the mesiodistal (md) and labiolingual (bl) diameters of *Denisova 8* and other hominin M^3s. For comparative sample used and sources for data see Table 4. **d:** Biplot of the mesiodistal (md) and labiolingual (bl) diameters of *Denisova 8* and other hominin M^2s. For comparative sample used and sources for data see Table 4.

Table 4. Metric comparisons of M^2 and M^3 length and breadth in various fossil hominins and the Denisova remains.

	M^2 md[1]	M^2 bl[2]	M^3 md	M^3 bl
A. afarensis	13.7±1.4 (13)[3]	14.7±0.9 (13)	13.1±1 (14)	15±1.3 (14)
A. africanus	13.9±1 (12)	15.3±1.1 (12)	13.8±1.3 (12)	15.6±1.4 (12)
Homo habilis	12.6±0.6 (6)	14±1.1 (6)	12.7±1.1 (7)	14.8±1.4 (7)
Dmanisi	12.3 (12.05-12.5; 2)[4]	12.7 (12.1-13.2; 2)	9.8 (1)	12 (1)
***H. erectus* (Africa)**	12.7 (11.7-13.7; 4)	13.5 (12.15-14.7; 4)	12.2 (12-12.3; 2)	14.5 (13.7-15.3; 2)
***H. erectus* (Indonesia)**	12.3 (11.2-13.6; 3)	14 (12.8-15.4; 3)	10.4 (9.4-11.3; 4)	13.8 (12.5-15.3; 4)
***H. erectus* (China)**	11.3±0.9 (8)	13.2±1.1 (8)	9.6±0.5 (7)	11.6±0.8 (7)
Atapuerca SH	10.6±0.7 (6)	12.9±0.9 (6)	8.5±0.4 (4)	11.4±0.9 (4)
***H. heidelbergensis* (Europe)**	11.6 (11.4-12.1; 4)	12.7 (11.9-13.7; 4)	10.1 (9.3-11.5; 4)	12.1 (11.8-12.5; 4)
Neandertals	11±1.4 (21)	12.7±1.2 (21)	10.1±1.8 (17)	12±1.3 (17)
Neandertals (w/o Obi-Rakhmat)	10.7±0.8 (20)	12.6±1.1 (20)	9.8±1 (16)	11.8±1.1 (16)
Early AMH	10.8±1.2 (10)	12.7±1.1 (10)	9.4±0.5 (6)	12.2±0.7 (6)
Upper Palaeolithic	10.4±1 (21)	12.3±1.2 (21)	9.8±1.4 (12)	12±1.5 (12)
Denisova 4	13.1	14.7	13.1	14.7
Denisova 8	-	-	14.3	14.65

1. Mesiodistal length measured following the definition of (119)
2. Buccoligual breadth measured following the definition of (119)
3. Mean+-standard deviation (N)
4. Mean (range; N)

Sources of metric data:

A. afarensis: Hadar, Omo (own measurements)
A. africanus: Stekfontein, Makapansgat (120)
Homo habilis: Olduvai (121), East Turkana (120)
Dmanisi (122)
H. erectus (Africa): East Turkana (120), Nariokotome (123), Konso (124), Swartkrans (120)
H. erectus (China): Zhoukoudian (125), Hexian (126)
H. erectus (Indonesia): Trinil (120), Sangiran (own measurements, (127))
Atapuerca SH (116)
H. heidelbergensis (Europe): La Chaise (128), Biache (129), Arago (130), Petralona (128)
Neandertals: Amud (131), Châteauneuf (132), St. Brelade (119), Krapina (133), La Croze de Dua (119), La Quina (119), Le Moustier (119), Obi-Rakhmat (own measurements), Saccopastore (119), Shanidar (134), Spy (119), Tabun (119), Vergisson la Falaise (119)
Early AMH: Skhul (135), Qafzeh (136), Temara (137)
Upper Paleolithic: Brno (119), Changwu (126), Cro-Magnon (119), Dolni Vestonice (138), Grotte des Enfants (119), Kostenki (own measurements), La Rochette (119), Leuca (119), Mladec (119), Oase (139), Predmosti (119), Sungir (own measurements)

5.3.2 DNA isolation and sequencing

DNA was extracted from 36 mg of dentine from *Denisova 8* in our clean room facility (67) and DNA libraries from this specimen as well as from a previously prepared extract of *Denisova 4* were prepared as described (8, 61) (see Table 1). From both teeth, random DNA fragments were sequenced and mapped to the human reference genome (hg19). In addition, mtDNA fragments were isolated from the libraries (70) and sequenced.

Of the DNA fragments sequenced from *Denisova 4* and *Denisova 8* 0.05% and 0.9%, respectively, could be confidently mapped to the human genome sequence, yielding 54.6 and 265 million base pairs (Mb) of nuclear DNA sequences for *Denisova 4* and *Denisova 8*, respectively (see Table 9 for overview). MtDNA sequences from the two specimens were aligned to the mtDNA of *Denisova3* (NC_013993.1). For *Denisova 4,* the average mtDNA coverage is 72.1-fold. The lowest support for the majority base at any position is 89% (Figure 12) and the consensus sequence is identical to the previously published mtDNA sequence from this specimen (2). For *Denisova 8,* the mtDNA coverage is 118.9-fold and the lowest support for the majority base is 86% (Figure 12).

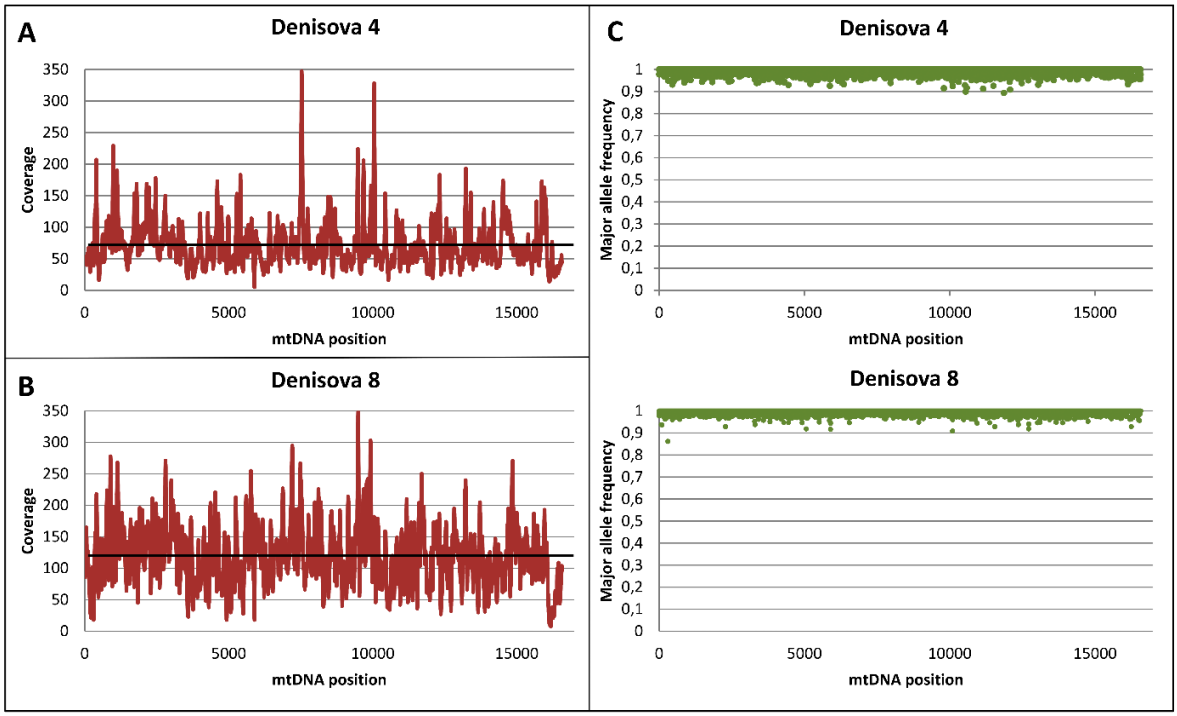

Figure 12. Quality of mtDNA sequences from *Denisova 4* and *8*. A, B: Coverage across the mitochondrial genomes. Black lines denote the average coverage. **C, D**: Consensus support across the genomes.

5.3.3 DNA sequence authenticity

We used three approaches to estimate present-day human DNA contamination in the two libraries. First, for each library, we used all unique DNA fragments that aligned to the present-day human reference mtDNA (76) and counted as contaminating those that carried a nucleotide different to the majority mtDNA sequence determined from the molar at positions where the endogenous majority consensus differed from all of 311 present-day human mtDNAs. The mtDNA contamination thus estimated was 5.2% (95% confidence interval (CI): 4.5-6.0%) for *Denisova 4* and 3.2% (95% CI: 2.9-3.6%) for *Denisova 8*.

Second, we estimate contamination by present-day nuclear DNA by estimating DNA sequence divergence (as described below and in Figure 17A) of the two molars to present-day humans. We assume that the divergence of two present-day European individuals to each other represent 100% contamination while the divergence of the high quality genome determined from

Denisova 3 to present-day humans represents zero percent contamination. By this approach, we estimate the *Denisova 4* DNA contamination of *Denisova 4* to 65.2-67.0% and *Denisova 8* to 14.6-15.4% (Table 5). That the nuclear DNA contamination is high, particularly of *Denisova 4,* is compatible with an estimate based on cytosine deamination patterns at the 3'- and 5'- ends of the aligned sequences (Supplementary methods).

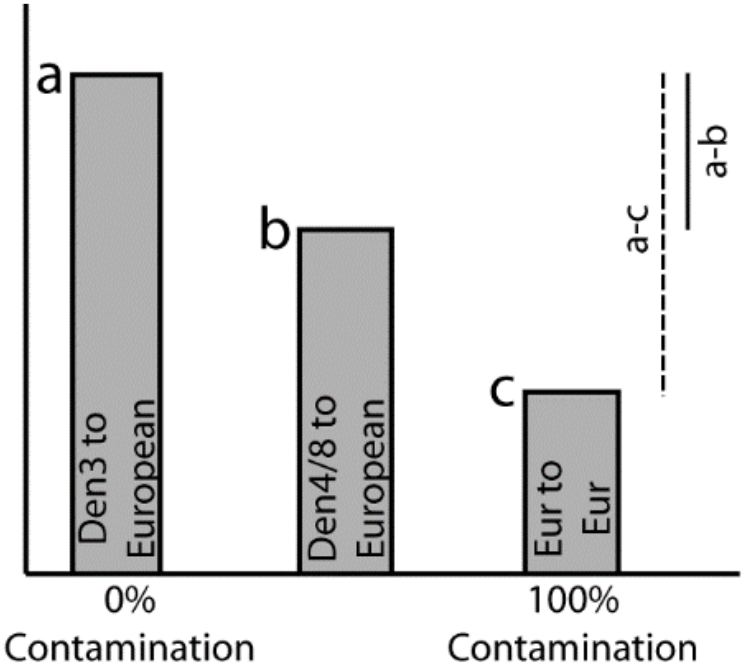

Figure 13. Divergence-based contamination estimates. The divergence of the *Denisovan 3* to two French and two Sardinians (left bar, *a*) is assumed to represent 0 % present-day human contamination. The divergence of French-French and Sardinian-Sardinian (right bar, *c*) is assumed to represent 100 % contamination. The divergence of *Denisova 4* or *8* to the French and Sardinians (middle bar, *b*) is then gauged as the reduction in divergence to the present-day humans as a fraction of the divergence among the present-day humans ((*a*-*b*) / (*a*-*c*)).

Table 5. Nuclear contamination estimate. An estimate of the nuclear contamination using the method described in Figure 13 applied to fragments without filtering for deamination.

European used to calc div[a]	% Divergence European[b]	% Divergence *Denisova 3*[c]	% Divergence *Denisova 4*	% Divergence *Denisova 8*	Div Den3 – Div human[d]	Div Den3 – Div *Den4*	Div Den3 – Div *Den8*	% contamination Den4[e]	% contamination Den8
French1	6.36 *(to Fr2)*	11.85	8.22	11.02	5.49	3.63	0.83	**66.1**	**15.1**
French2	6.09 *(to Fr1)*	11.62	7.98	10.81	5.53	3.64	0.81	**65.8**	**14.6**
Sardinian1	6.34 *(to Sa2)*	11.86	8.26	11.05	5.52	3.6	0.81	**65.2**	**14.7**
Sardinian2	6.06 *(to Sa1)*	11.64	7.9	10.78	5.58	3.74	0.86	**67.0**	**15.4**

a. The European present-day humans to whom divergence is calculated and whose mutations are used to calculate divergence
b. Divergence calculation using pairs of Europeans. Thus: French2 to French 1, and vice versa, as well as Sardinian2 to Sardinian1 and vice versa. As an example French2 to French1 uses the mutations on the branch to French1 to calculate the divergence and gives a result of 6.36%.
c. Divergence of *Denisova 3* to each of the European present-day humans listed.
d. Differences in divergence, calculated e.g. divergence of Den3 to French1 minus the divergence of French2 to French1 (in this case *a* – *c* in Figure 13).
e. Percent contamination, calculated e.g. (divergence of Den3 to Fr1 – divergence of Den8 to Fr1) / (divergence of Den3 to Fr1 – divergence of Fr2 to Fr1)*100. In this case this would be (a-b)/(a-c)*100 in Figure 13.

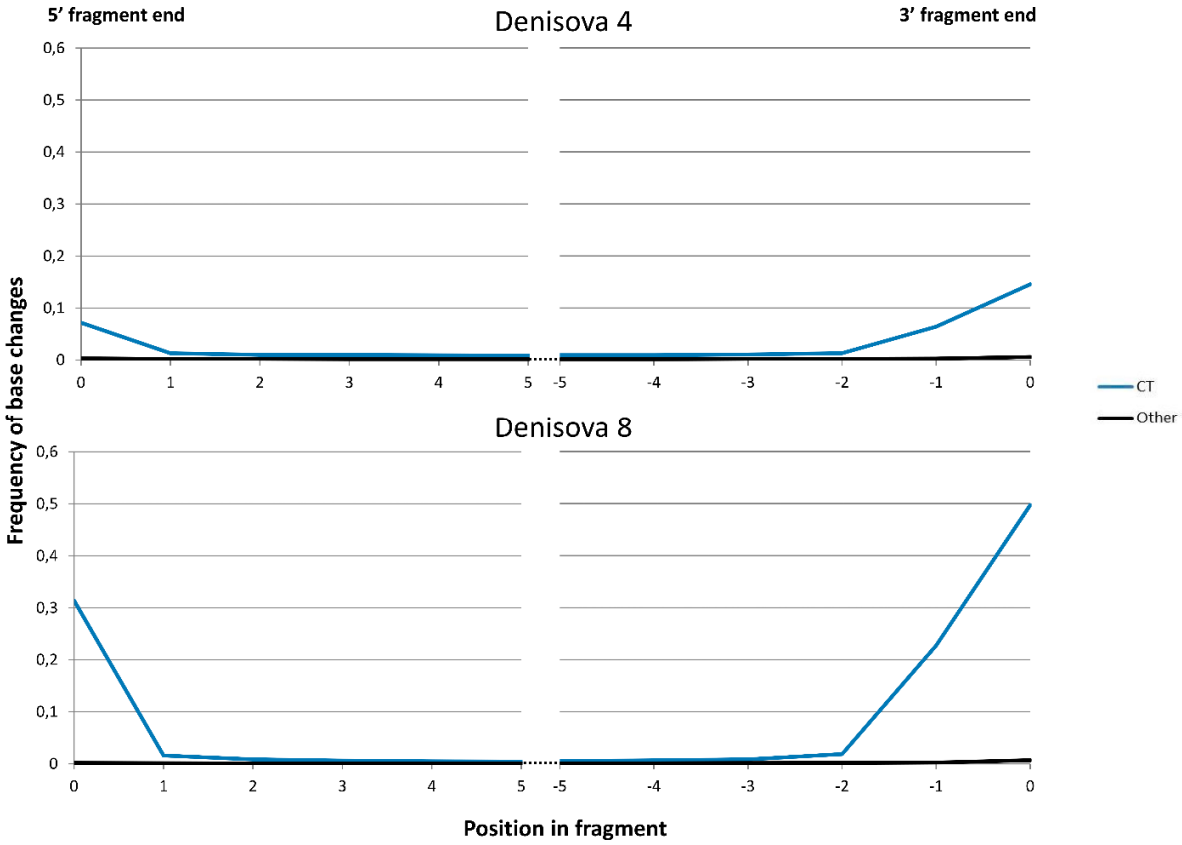

Figure 14. Nucleotide differences to the human reference genome as a function of distance from fragment ends. Differences are given as percent of a base in the reference genome that occurs as a different base in the sequenced DNA fragments. C to T differences are largely due to deamination of cytosine residues in ancient DNA fragments. Libraries were treated with *E.coli* uracil DNA glycosylase, which is not efficient at the first, the last and second to last bases.

Table 6. Terminal C to T substitutions nuclear and mtDNA fragments. C to T substitutions relative to the corresponding mtDNA consensus sequences are shown for mtDNA and nuclear DNA fragments sequenced from *Denisova 4* and *Denisova 8*, respectively. "3' filtered" and "5' filtered" refer to fragments that carry C to T substitutions at their 3'- and 5'-ends, respectively. The 95% CI is given in parenthesis.

		5 prime	3 prime
***Denisova 4* mtDNA**	***No filter***	11.3 *(9.7-13.0)*	22.4 *(20.9-24.1)*
	3' filtered	17 *(9.7-27.8)*	100
	5' filtered	100	30.5 *(22.2-40.4)*
***Denisova 4* nuclear**	***No filter***	7.2 *(6.9-7.4)*	14.6 *(14.3-14.8)*
	3' filtered	18.9 *(16.0-22.2)*	100
	5' filtered	100	35.7 *(32.6-39.1)*
***Denisova 8* mtDNA**	***No filter***	23.7 *(21.9-25.6)*	46.0 *(44.5-47.6)*
	3' filtered	20.8 *(16.2-26.3)*	100
	5' filtered	100	46.9 *(39.9-54.2)*
***Denisova 8* nuclear**	***No filter***	31.4 *(31.2-31.6)*	49.8 *(49.7-49.9)*
	3' filtered	32.5 *(32.0-33.2)*	100
	5' filtered	100	52.3 *(51.8-52.8)*

In the third approach, we first determined the sex of the individuals from which the molars derive by counting the numbers of DNA fragments that map to the X chromosome and autosomes, respectively. To limit the influence of present-day DNA contamination in this part of the analysis, we restrict to DNA fragments that at their 5'- and/or 3'-ends carry thymines (T) at positions where the human reference nuclear genome carries cytosines (C). Such apparent C to T substitutions are frequently caused by deamination of cytosine to uracil towards the ends of ancient DNA fragments (55, 56). We find that both teeth come from males ($p\sim0.4$) rather than females ($p<<0.01$) (Table 7). We then estimated the amount of female DNA contamination among the aligned sequences as the fraction of DNA fragments that match the X chromosome in excess of what is expected for a male bone. This yields a female DNA contamination rate of 28.4% (95 CI: 27.3-29.5%) for *Denisova 4* and 8.6% (95 CI: 8.3-8.9%) for *Denisova 8*.

Table 7. Sex determination and female contamination. The number of X- and Y-chromosomal sequences mapped and expected to map if the molars are from males. DNA sequences carrying terminal C to T substations as well as all sequences were analyzed.

		Y-chromosome			X-chromosome			
Denisova	**Analysis/ Sequences**	# of sequences mapped	# sequences expected to map if male	χ^2*-test p-value*	# of sequences mapped	# sequences expected to map if male	χ^2*-test p-value*	Percent female contamination
4	**Sex determination (Terminal C->T seqs)**	8	3	-	231	222	*0.42 (5.9e-14 if female)*	-
8		94	86	*0.26*	5,535	5,576	*0.43 (<2.2e-16 if female)*	-
4	**Contamination estimate (all seqs)**	75	93	*0.006*	7,764	6,048	*<2.2e-16*	28.4% *(27.3-29.5)*
8		617	599	*0.32*	42,175	38,829	*<2.2e-16*	8.6% *(8.3-8.9)*

The estimates based on mtDNA and nuclear DNA differ drastically (Table 8) presumably because the ratios of mitochondrial to nuclear DNA differ between the endogenous and the contaminating source(s) of DNA while the two estimates based on nuclear DNA suggest that more males than females are among the contaminating individuals. It is clear that although these methods yield different contamination estimates, they all suggest that the nuclear DNA contamination in both libraries is substantial, particularly in *Denisova 4* where it is likely to exceed 50%. To reduce the influence of DNA contamination (87, 103) we therefore restrict the analyses of nuclear DNA to fragments that carry thymine residues at the first and/or last two positions at sites where the human reference sequence carries cytosine residues (but remove these C/T sites themselves in the analyses). Using these criteria, a total of 1.0 Mb of nuclear DNA sequences for *Denisova 4* and 24.1 Mb for *Denisova 8* (Tables 8 and 9) can be analyzed.

Table 8. Overview of DNA sequences produced, contamination estimates, and amount of nuclear sequences used for analyses.

	Denisova 4	*Denisova 8*
Amount of mapped sequences	54.6Mb	265Mb
MtDNA coverage	72-fold	119-fold
Autosomal contamination	~66%	~15%
mtDNA contamination	~5.2%	~3.2%
X chr. contamination	~28%	~9%
Nuclear sequences used	1Mb	24Mb

Table 9. DNA sequences yields.

	Mg of bone powder for extract[a]	**% of extract used for library**	**% endogenous[b]**	**Mb aligned to human genome[c]**	**Mb aligned after duplicate removal**	**% unique[d]**	**Mb aligned after deamination filter[e]**
Denisova 4	40	20%	0.05%	80.7 Mb	54.6 Mb	67.6%	1.0
Denisova 8	36	10%	0.9%	1,128 Mb	265 Mb	23.5%	24.1

a. Milligrams of bone powder used to make 100uL of extract
b. Percent endogenous is calculated as the Mb aligned to the human genome (after filtering for mapped sequences with a length above 35) divided by the total Mb sequenced (after filtering for a length above 35) times 100.
c. Mb aligned to hg19 after passing the following filters: length > 35, map quality > 37, merging of paird reads with minimum 11 bp overlap, fewer than 5 bases with base quality below 15, index reads with base qualities above 10.
d. Percent unique is Mb aligned with filters to the human genome after duplicate removal divided by aligned Mb before duplicate removal times 100
e. For deamination filter see the supplemental text.

5.3.4 MtDNA relationships

A phylogenetic tree relating the mtDNAs from the *Denisova 3, 4* and *8*, seven Neandertals from Spain, Croatia, Germany, the Russian Caucasus and the Altai Mountains (10, 73), and five present-day humans (Figure 15A, B) shows that the mtDNAs of the two Denisovan molars form a clade with *Denisova 3* to the exclusion of the Neandertals. The largest number of differences seen among the three Denisovan mtDNAs is 86 while the largest number of differences seen among seven Neandertal mtDNAs is 51 and among 311 present-day humans, 118 (Figure 15C). When comparing Watterson's estimator θw, which to some extent takes the numbers of samples into account, among the populations the mtDNA diversity of the three Denisovans is 3.5×10^{-3}, that of Neandertals 1.8×10^{-3} while that of present-day Europeans is 4.0×10^{-3} and present-day humans world-vide is 16.1×10^{-3}. Thus, mtDNA diversity among late Neandertals seems to be low relative to Denisovans as well as present-day humans.

The number of nucleotide changes inferred to have occurred from the most recent common ancestor (MRCA) of the three Denisovan mtDNAs to the *Denisova 4* molar, the *Denisova 3* phalanx and the *Denisova 8* molar are 55, 57 and 29 respectively (Figure 15B, Table 10). The corresponding number of substitutions from the MRCA of the seven Neandertal mtDNAs to each of the Neandertal mtDNAs varies between 17 and 25 (Table 10). This suggests that the time back to the mtDNA MRCA from the *Denisova 3* and the *Denisova 4* mtDNAs was almost twice as long as that from the *Denisova 8* mtDNA.

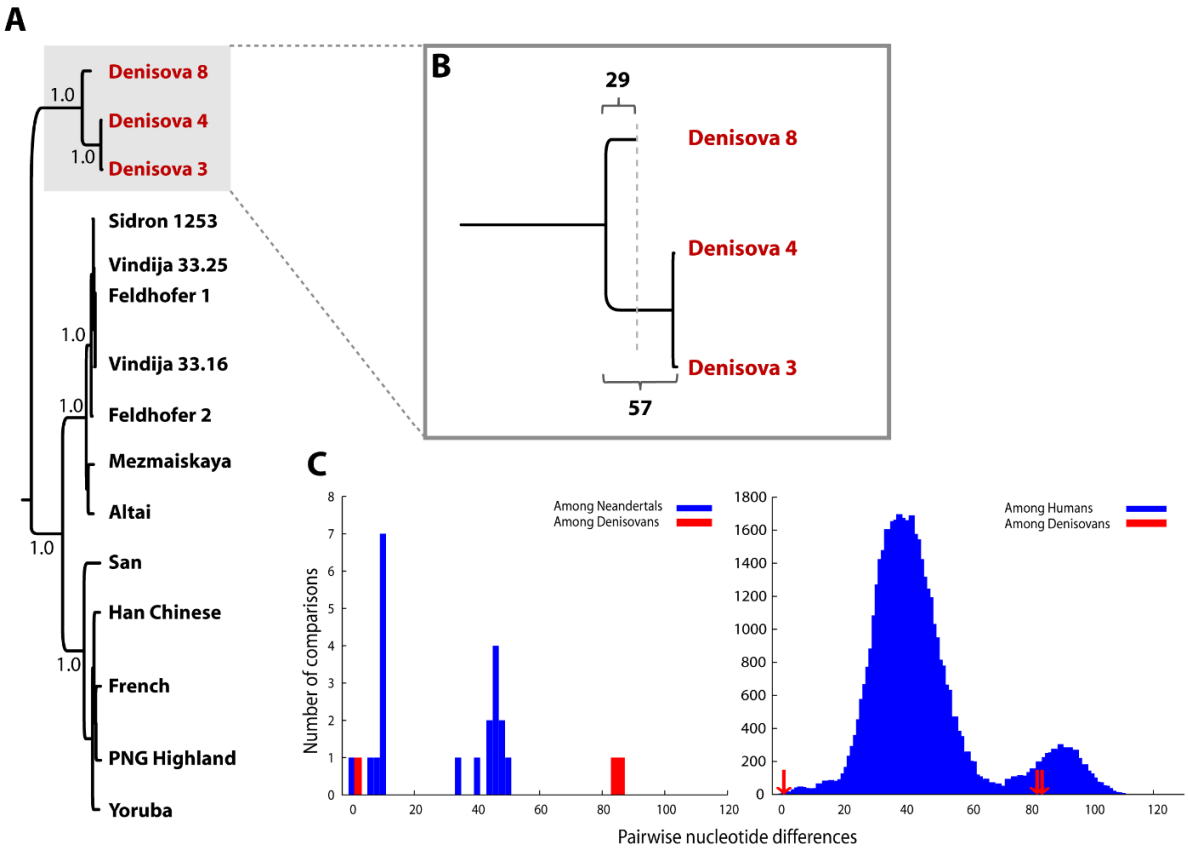

Figure 15. Evolutionary relationships of Denisovan mtDNAs. A. Bayesian tree relating the mtDNAs of three Denisovans, seven Neandertals and five present-day humans. Posterior probabilities are indicated. A chimpanzee mtDNA was used to root the tree. **B.** Numbers of differences between the two molar mtDNAs and the inferred common mtDNA ancestor of the three Denisovan mtDNA. **C.** Pairwise nucleotide differences among the Denisovans and Neandertals (left panel) and among the Denisovans and 311 present-day human mtDNAs (right panel).

Table 10. Number of differences to mtDNA MRCAs. The number of differences between each Denisovan mtDNA and their inferred MRCA as well as between each Neandertal mtDNA and their inferred MRCA.

Denisovan	Number of diffs to MRCA of Denisovans	Neandertal	Number of diffs to MRCA of Neandertals
Denisova 3	57	***Mezmaiskaya1***	25
Denisova 4	55	**Altai 1**	24
Denisova 8	29	**Feldhofer 1**	21
		Feldhofer 2	17
		Sidron 1253	19
		Vi33.16	23
		Vi33.25	21

Table 11. Watterson's estimator *(θw)* for mtDNA.

Population	# segregating sites	n (# indv)	*θw*
Denisovans	86	3	3.46E-03
Neandertals	73	7	1.80E-03
Present-day humans	1,689	311	16.1E-03
Present-day Europeans	262	31	3.96E-03

Table 12. Age estimates of the two molars and mtDNA lineages divergences based on mtDNA. Estimates using dates of 50,000 years as well as 100,000 years for *Denisova 3* and 95% upper and lower highest posterior densities (HPD) are given in thousand years (kyr).

	Age of *Denisova 3* set to 50,000 years BP			Age *of Denisova 3* set to 100,000 years BP		
Mitochondrial lineage	Estimate	95% HPD lower	95% HDP upper	Estimate	95% HPD lower	95% HDP upper
Denisova 8 age	177 kyr	97 kyr	265 kyr	226 kyr	143 kyr	313 kyr
Denisova 4 age	56 kyr	45 kyr	69 kyr	106 kyr	094 kyr	121 kyr
Denisova-Human/Neandertal	808 kyr	622 kyr	1,016 kyr	846 kyr	652 kyr	1056 kyr
Den8 – Den4/*Den3*	262 kyr	187 kyr	343 kyr	314 kyr	238 kyr	393 kyr
Human-Neandertal	405 kyr	312 kyr	511 kyr	413 kyr	318 kyr	522 kyr
San-rest of humans	173 kyr	128 kyr	223 kyr	176 kyr	128 kyr	225 kyr
Mezmaiskaya1-rest of Neandertals	128 kyr	101 kyr	155 kyr	129 kyr	103 kyr	157 kyr

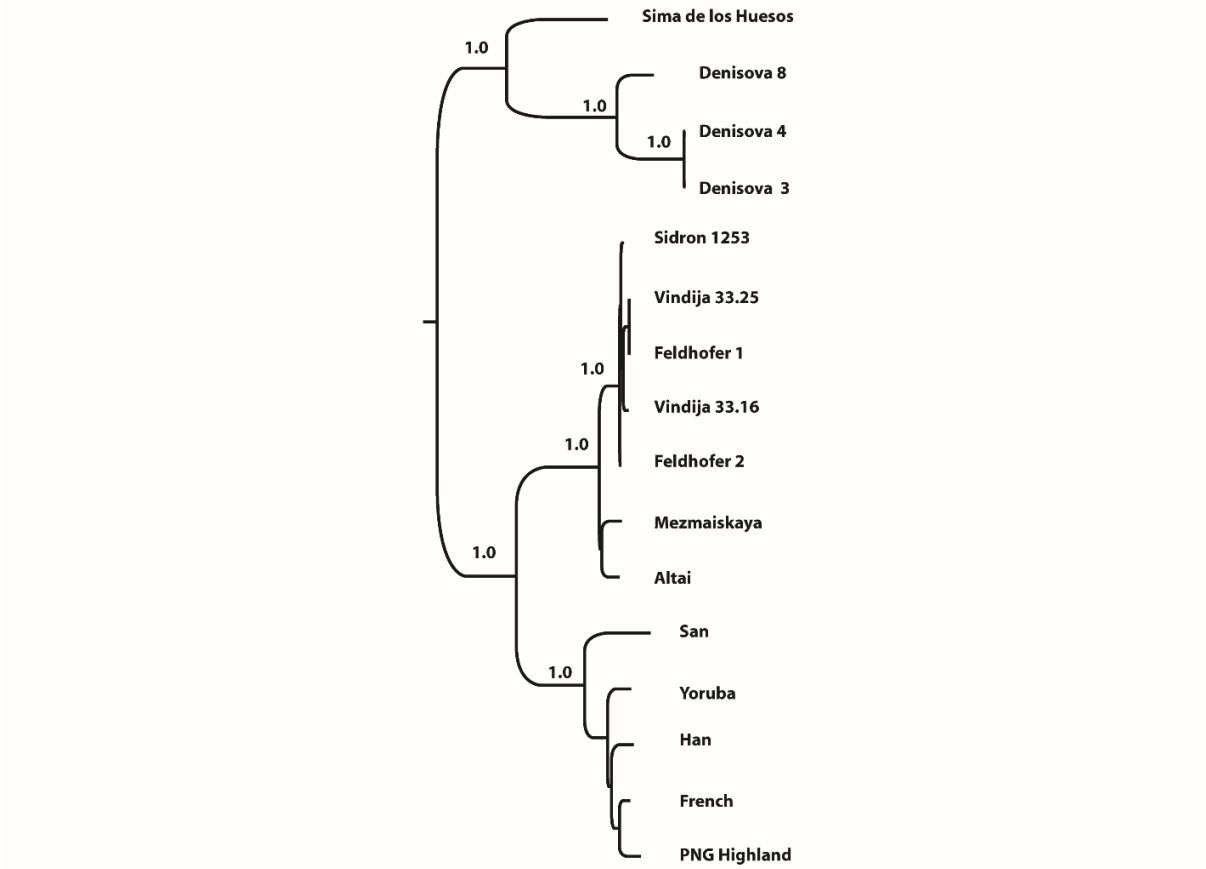

Figure 16. MtDNA tree of three Denisovans, seven Neandertals, a hominin from Sima de los Huesos (87), and five present-day humans. The Bayesian tree was computed using 16,286 mtDNA positions and a chimpanzee mtDNA (X93335.1) as outgroup (not shown). Important posterior probabilities are shown.

5.3.5 Autosomal Analyses

To estimate the divergence of the low coverage DNA sequences retrieved from *Denisova 4* and *8* to the high quality genomes of *Denisova 3* (3) as well as to the Neandertal from Denisova Cave and to ten present-day humans (10), we first counted nucleotide substitutions inferred to have occurred on the lineages from the human-chimpanzee ancestor to each of the high-coverage genomes (Figure 17A, a + b). We then used the low coverage molar sequences to estimate the fraction of those substitutions that occurred after their divergence from the high coverage lineages, *i.e*. the fraction of such substitutions not seen in the molars (Figure 17A, b). To the Denisovan high coverage genome, these fractions are 2.9% (95% CI: 2.28-3.44) and 3.4% (95% CI: 3.25-3.53) for *Denisova 4* and *Denisova 8*, respectively. Divergences of *Denisova 4* and *Denisova 8* are 8.9% (CI: 8.01-9.83%) and 8.3% (CI: 8.01-8.48 %) to the high coverage Neandertal genome and 10.9 - 12.9% to ten present-day humans (Figure 17B; Tables 10 and 11). These results show that the two teeth come from Denisovans and confirm that Denisovans were a sister group of Neandertals.

The average pairwise divergence among six low-coverage Neandertals to the Altai Neandertal genome is 2.5% (range 2.5% to 2.6%) (Table 14). This is slightly lower than the divergence of 2.9% and 3.4% of the two Denisovan molars from the Denisova genome and shows that the individuals from whom the two molars derive are almost as closely related to the *Denisova 3* genome as are the Neandertals to the Altai Neandertal genome. By comparison, the range of divergences among ten present-day human genomes is 4.2% to 9.5%, among the four Europeans 6.0 to 6.4% and between the two individuals from the South American tribal group Karitiana 4.2%. Thus, nuclear DNA diversity appears low among the archaic individuals, especially the Neandertals.

Using the high coverage *Denisova 3* genome it was shown that Denisovans have contributed DNA to present-day people in Oceania (8-10, 60). As expected, we found that *Denisova 8* also shares more derived alleles with Papuans and Australians than with other non-Africans (*D*: -0.04 to -0.07, [Z]=1.8–3.0, excluding CpG sites, Table 16). However, when we subsample from the high coverage Denisovan genome the DNA segments covered by fragments sequenced from *Denisova 4* we find that there are not enough data to similarly detect gene flow from *Denisova 4* to Oceanians (data not shown). This precludes us from asking whether either *Denisova 4* or *Denisova 8* are more closely related to the introgressing Denisovan than *Denisova 3*. Similarly, there are not enough data to determine whether gene flow from Neandertals at the

level detected in the high-coverage *Denisova 3* genome (10) is present in *Denisova 4* and *8* (Table 17).

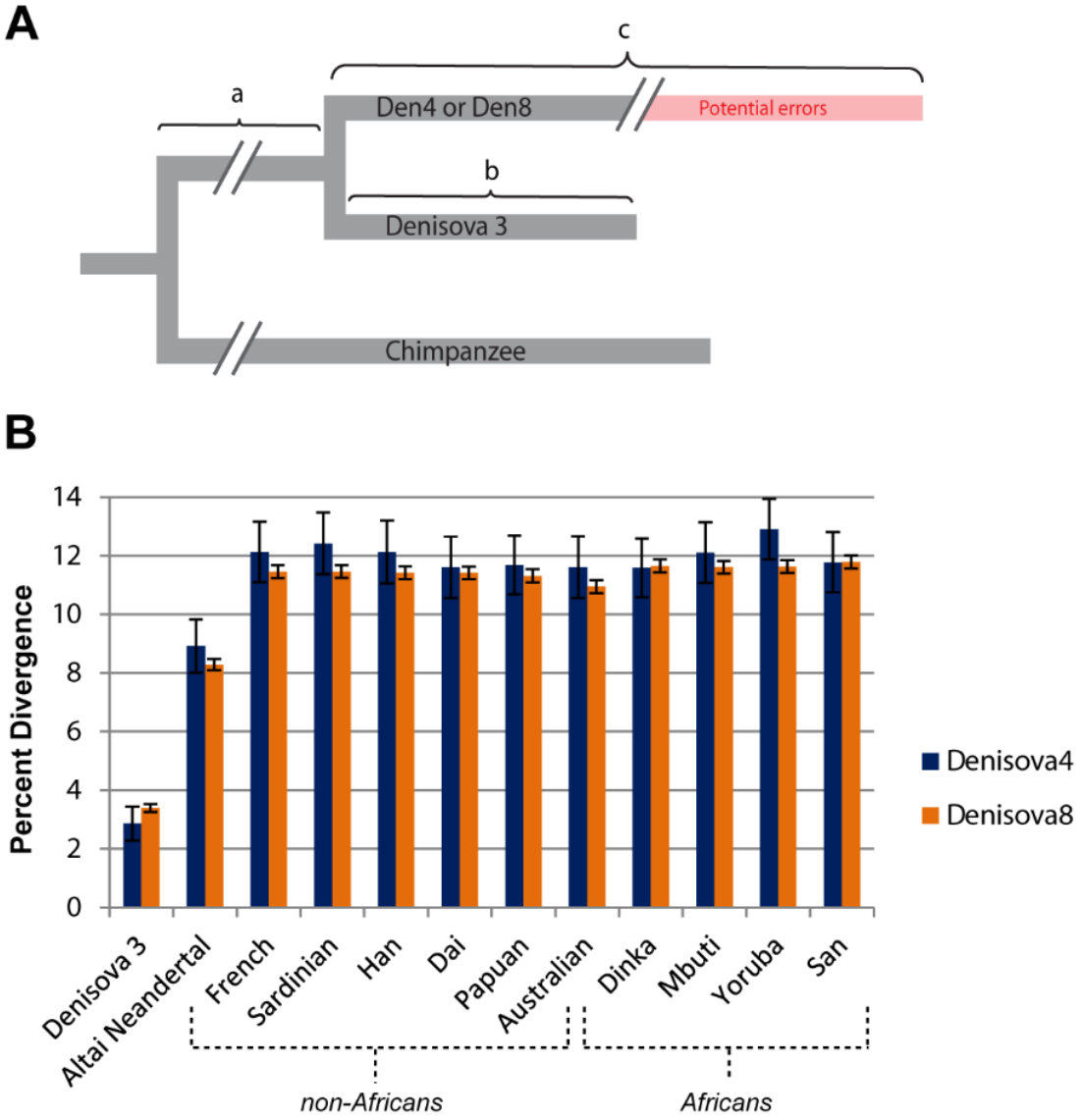

Figure 17. **Nuclear DNA divergence between *Denisova 4* and *8* and the Denisovan genome. A:** DNA sequences from *Denisova 4* and *8* were each compared to the genomes of *Denisova 3* (8) and the inferred human-chimpanzee ancestor (91, 92). The differences from the human-chimpanzee ancestor common to the two Denisovans (a) as well as differences unique to each Denisovan are shown (b and c). Errors in the low coverage Denisova genomes result in artificially long branches (c). Divergences of the molar genomes to *Denisova 3* are therefore calculated as the percent of all differences between *Denisova 3* and the human-chimpanzee ancestor that are not shared with the molar genomes, b/(a+b)x100. **B.** Autosomal divergences of *Denisova 4* and *Denisova 8* to the *Denisova 3* genome, the Neandertal genome, and ten present-day human genomes calculated as in A. All estimates are based on DNA fragments from the two molars that carry putative deamination-induced C to T substitutions. Bars indicate 95% confidence intervals.

Table 13. Divergences for *Denisova 4*. Divergences for the deaminated sequences, not deaminated sequences as well as all sequences combined are shown. Divergence is given the percent divergence of *Denisova 4* along the branch to the human-chimpanzee ancestor from the high-coverage genomes given in the first column. Percent divergences and 95% CI are given.

	Deaminated fragments				**Not deaminated fragments**				**All fragments**			
High-coverage genomes	Shared[1]	Genome[2]	*Den4*[3]	%	Shared	Genome	*Den4*	%	Shared	Genome	*Den4*	%
Denisova 3	3,699	109	3,767	**2.86** *2.28-3.44*	121,663	11,775	77,551	**8.82** *8.66-8.99*	126,716	11,990	81,920	**8.64** *8.48-8.81*
Altai Neandertal	3,471	340	4,029	**8.92** *8.01-9.83*	120,142	13,796	79,546	**10.30** *10.11-10.48*	124,952	14,290	84,303	**10.26** *10.08-10.44*
French	3,482	481	4,164	**12.14** *11.10-13.17*	126,237	11,133	76,306	**8.10** *7.94-8.27*	131,123	11,749	80,963	**8.22** *8.05-8.39*
Sardinian	3,448	489	4,095	**12.42** *11.37-13.47*	124,622	11,049	75,208	**8.14** *7.97-8.31*	129,262	11,634	80,055	**8.26** *8.09-8.42*
Han	3,455	477	4,111	**12.13** *11.06-13.2*	125,153	11,464	76,061	**8.39** *8.21-8.57*	129,955	11,919	80,724	**8.40** *8.23-8.57*
Dai	3,442	452	4,120	**11.61** *10.56-12.66*	124,793	11,519	75,590	**8.45** *8.28-8.62*	129,623	11,993	80,407	**8.47** *8.31-8.63*
Papuan	3,445	456	4,087	**11.69** *10.69-12.69*	124,182	11,617	75,444	**8.55** *8.38-8.73*	129,005	12,275	80,101	**8.69** *8.53-8.85*
Australian	3,418	449	4,098	**11.61** *10.56-12.67*	124,613	11,252	75,620	**8.28** *8.11-8.45*	129,368	11,845	80,360	**8.39** *8.23-8.55*

Dinka	3,418	448	4,159	**11.59** *10.58-12.59*	123,200	12,939	77,631	**9.50** *9.32-9.69*	127,989	13,397	82,318	**9.48** *9.3-9.66*
Mbuti	3,433	473	4,129	**12.11** *11.08-13.14*	122,769	13,726	78,122	**10.06** *9.87-10.24*	127,615	14,241	82,765	**10.04** *9.86-10.22*
Yoruba	3,473	515	4,146	**12.91** *11.88-13.95*	123,623	13,188	78,107	**9.64** *9.46-9.82*	128,425	13,890	82,882	**9.76** *9.57-9.95*
San	3,407	455	4,095	**11.78** *10.76-12.81*	121,951	13,989	77,901	**10.29** *10.10-10.48*	126,739	14,558	82,650	**10.30** *10.11-10.49*

1. The number of allelic states shared by the genome and *Densiova 4* but not shared with the human-chimpanzee ancestor.
2. Allelic states specific to the genome analyzed.
3. Allelic states specific to *Denisova 4*.

Table 14. **Divergences for *Denisova 8***. See Table 13 for explanations.

	Denisova8 deaminated				Denisova8 not deaminated				Denisova8 all			
Individual#1	Shared	Genome	*Den8*	%	Shared	Genome	*Den8*	%	Shared	Genome	*Den8*	%
Denisova 3	88,315	3102	33,574	**3.39** *3.25-3.53*	507,405	26,224	210,931	**4.91** *4.83-5*	637,505	31,657	261,670	**4.73** *4.64-4.82*
Altai Neandertal	84,101	7598	38,370	**8.29** *8.09-8.48*	486,591	47,274	234,493	**8.86** *8.73-8.97*	611,034	58,838	292,030	**8.78** *8.66-8.9*
French	82,999	10741	40,898	**11.46** *11.23-11.69*	486,909	60,026	243,442	**10.97** *10.86-11.09*	609,735	75,858	303,855	**11.02** *10.95-11.17*
Sardinian	82,188	10641	40,463	**11.46** *11.24-11.68*	481,113	59,575	240,320	**11.02** *10.9-11.13*	602,610	74,671	299,982	**11.05** *10.92-11.13*
Han	82,694	10661	40,505	**11.42** *11.2-11.64*	483,764	60,157	242,418	**11.06** *10.95-11.17*	606,187	75,989	302,355	**11.13** *11.03-11.24*
Dai	82,488	10633	40,676	**11.42** *11.2-11.64*	482,321	59,659	242,036	**11.01** *10.89-11.12*	604,506	75,249	302,505	**11.10** *10.97-11.17*
Papuan	82,423	10515	40,375	**11.31** *11.1-11.54*	481,045	59,090	240,568	**10.94** *10.83-11.05*	602,992	74,518	300,472	**11.00** *10.89-11.11*
Australian	82,513	10150	40,374	**10.95** *10.73-11.18*	482,594	57,825	240,792	**10.70** *10.59-10.81*	604,910	72,637	300,738	**10.76** *10.61-10.83*
Dinka	82,250	10846	40,385	**11.65** *11.43-11.87*	480,376	61,308	243,261	**11.32** *11.21-11.43*	601,643	76,990	303,706	**11.31** *11.24-11.45*
Mbuti	82,646	10858	40,571	**11.61**	480,838	62,446	244,989	**11.49**	603,063	78,469	305,286	**11.51**

				11.4-11.82				*11.37-11.61*				*11.4-11.63*
Yoruba	82,598	10875	40,745	**11.63** *11.42-11.85*	482,785	62,201	244,267	**11.41** *11.29-11.53*	604,950	77,960	304,739	**11.41** *11.31-11.52*
San	82,173	10985	40,645	**11.79** *11.57-12.01*	478,377	62,644	243,639	**11.58** *11.46-11.69*	599,764	79,290	304,396	**11.65** *11.57-11.78*

Table 15. Divergences for *Denisova 3*. See Table 13 for explanations.

	***Denisova 3* deaminated**				***Denisova 3* all**			
Individual#1	Shared	Genome	*Den3*	%	Shared	Genome	*Den3*	%
Denisova 3	-	-	-	-	-	-	-	-
Altai Neandertal	4531663	418624	1180396	**8.46** *8.37-8.54*	6040420	560355	1424811	**8.49** *8.4-8.57*
French	4439597	591694	1303961	**11.76** *11.68-11.84*	5908484	793950	1585350	**11.85** *11.76-11.93*
Sardinian	4391458	584609	1288694	**11.75** *11.67-11.82*	5842629	786441	1568908	**11.86** *11.78-11.94*
Han	4421887	587375	1295063	**11.73** *11.64-11.81*	5882753	788594	1576327	**11.82** *11.73-11.9*
Dai	4431008	587058	1299791	**11.70** *11.62-11.78*	5893395	788793	1581617	**11.80** *11.72-11.89*
Papuan	4410486	577448	1283146	**11.58** *11.49-11.66*	5867117	774918	1559910	**11.67** *11.58-11.75*
Australian	4433269	565793	1285982	**11.32** *11.23-11.4*	5899300	759006	1564065	**11.40** *11.31-11.49*
Dinka								

				11.69-11.87				*11.92-12.27*
Mbuti	4427808	593721	1301891	**11.82** *11.74-11.9*	5889250	795352	1585013	**11.90** *11.82-11.98*
Yoruba	4422950	592266	1297910	**11.81** *11.72-11.89*	5884572	794419	1581895	**11.89** *11.81-11.98*
San	4413422	595874	1297860	**11.90** *11.81-11.98*	5870882	798906	1580382	**11.98** *11.89-12.06*

Table 16. Divergences for Neandertals to the high coverage Altai Neandertal genome. See Table 13 for explanations of labels. All *Mezmaiskaya1* fragments were used for this analysis, because UDG treatment left C to T substitutions at only 4% of fragment ends.

	Neandertal deaminated				Neandertal all			
Neandertal	Shared	AltaiNea	Neandertal	%	Shared	AltaiNea	Neandertal	%
Feldhofer 1	447	6	576	1.32 *0.28-2.37*	2,581	67	3,446	2.53 *1.96-3.1*
Sidron 1253	893	29	1026	3.15 *2.00-4.29*	2,716	73	3,158	2.62 *1.97-3.26*
Vindija33.16	569,284	14,610	750,801	2.50 *2.44-2.57*	1,611,437	42,324	1,991,958	2.56 *2.5-2.61*
Vindija33.25	500,325	12,729	560,651	2.48 *2.41-2.55*	1,730,545	43,780	1,918,680	2.47 *2.41-2.52*
Vindija33.26	477,869	12,296	585,208	2.51 *2.44-2.58*	1,591,266	40,910	1,829,657	2.51 *2.45-2.56*
Mezmaiskaya1	-	-	-	-	2,331,784	59,473	772,431	2.49 *2.43-2.54*

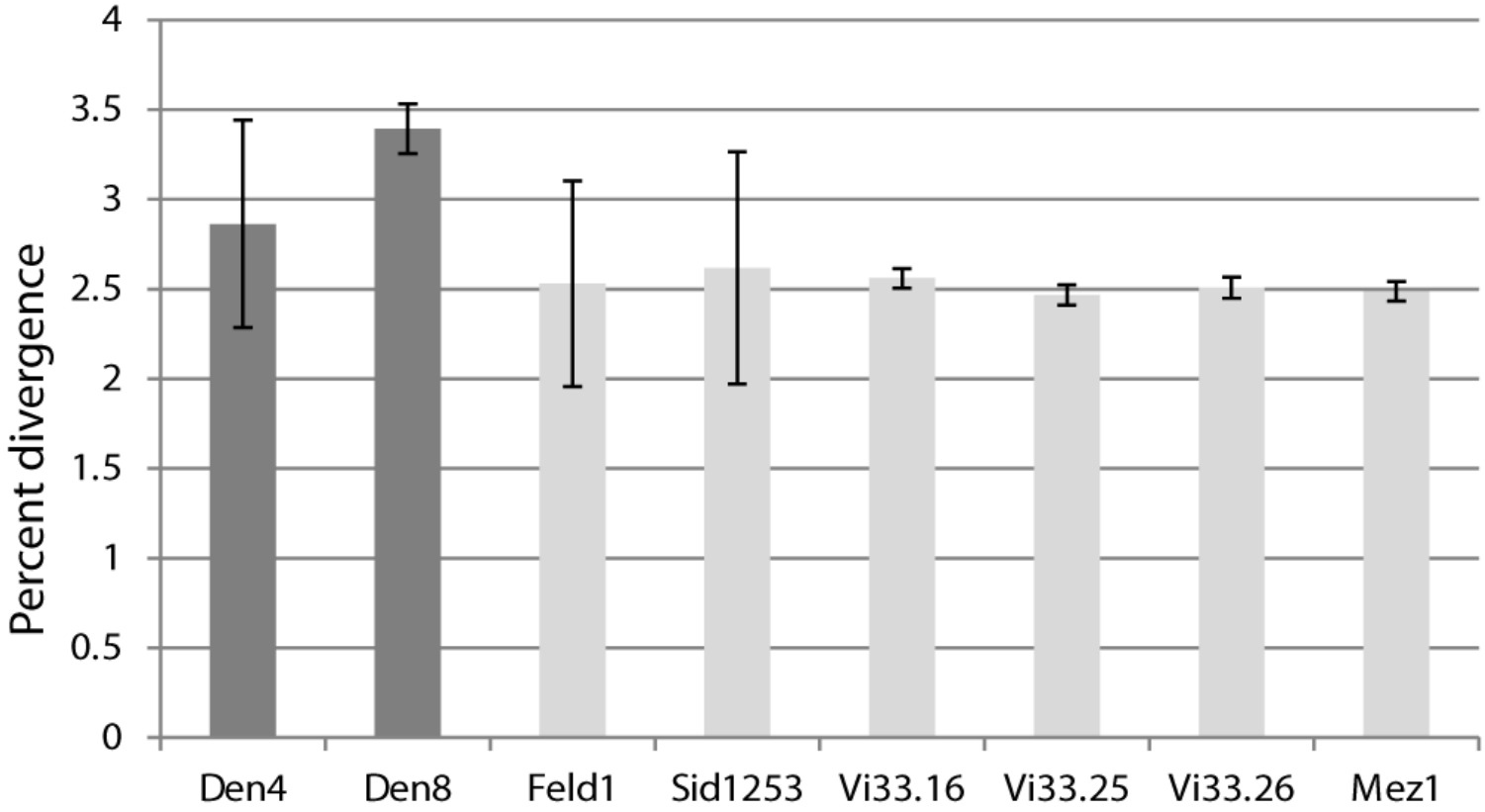

Figure 18. Divergences to *Denisova 3* and Altai Neandertal reference genomes. The percent divergence of the *Denisova 4* and *8* genomes to the *Denisova 3* genome (dark gray) and of six low-coverage Neandertal genomes to the Altai Neandertal genome (light gray) estimated as in main text Fig. 3A. Error bars indicate 95% CIs.

Table 17. Sharing of derived alleles between *Denisova 8* and Eurasian populations. Only *Denisova 8* fragments carrying a C to T substitutions at the first or last two positions are used.

	Type of sites	$AADA_a$	ADDA	DADA	DDDA	(ADDA-DADA)/ (ADDA+ADDA)	Z_b
Papuan, French, *Den8*, Chimp	all sites	43,502	1,311	1,473	205,735	-0.06	-3.03
	no cpg sites	36,640	906	1,022	179,687	-0.06	**-2.55**
	only cpg sites	6,862	405	451	26,048	-0.05	-1.57
	transitions	25,093	913	1,004	136,322	-0.05	-2.03
	transversions	18,409	398	469	69,413	-0.08	-2.33
Papuan, Sardinian, *Den8*, Chimp	all sites	43,387	1,358	1,454	205,685	-0.03	-1.90
	no cpg sites	36,519	930	1,023	179,680	-0.05	**-2.24**
	only cpg sites	6,868	428	431	26,005	0.00	-0.10
	transitions	25,031	944	1,010	136,224	-0.03	-1.56
	transversions	18,356	414	444	69,461	-0.03	-1.02
Papuan, Han, *Den8*, Chimp	all sites	43,255	1,232	1,352	204,023	-0.05	-2.32
	no cpg sites	36,435	832	951	178,188	-0.07	**-2.84**
	only cpg sites	6,820	400	401	25,835	0.00	-0.03
	transitions	24,989	855	913	135,233	-0.03	-1.36
	transversions	18,266	377	439	68,790	-0.08	-2.09
Papuan, Dai, *Den8*, Chimp	all sites	43,215	1,199	1,356	204,110	-0.06	-3.31
	no cpg sites	36,360	833	956	178,239	-0.07	**-3.01**
	only cpg sites	6,855	366	400	25,871	-0.04	-1.32
	transitions	24,981	816	927	135,238	-0.06	-2.81

	transversions	18,234	383	429	68,872	-0.06	-1.66
Australian, French, *Den8*, Chimp	all sites	43,027	1,224	1,451	204,126	-0.08	-4.32
	no cpg sites	36,314	861	966	178,308	-0.06	**-2.43**
	only cpg sites	6,713	363	485	25,818	-0.14	-4.19
	transitions	24,847	865	979	135,092	-0.06	-2.66
	transversions	18,180	359	472	69,034	-0.14	-3.82
Australian, Sardinian, *Den8*, Chimp	all sites	43,118	1,313	1,482	204,409	-0.06	-3.19
	no cpg sites	36,335	915	1,023	178,581	-0.06	**-2.40**
	only cpg sites	6,783	398	459	25,828	-0.07	-2.13
	transitions	24,892	896	1,009	135,205	-0.06	-2.60
	transversions	18,226	417	473	69,204	-0.06	-1.91
Australian, Han, *Den8*, Chimp	all sites	43,016	1,228	1,389	202,806	-0.06	-3.06
	no cpg sites	36,281	844	944	177,174	-0.06	**-2.34**
	only cpg sites	6,735	384	445	25,632	-0.07	-2.07
	transitions	24,852	875	927	134,265	-0.03	-1.20
	transversions	18,164	353	462	68,541	-0.13	-3.92
Australian, Dai, *Den8*, Chimp	all sites	42,767	1,243	1,391	202,727	-0.06	-2.98
	no cpg sites	36,047	894	969	177,058	-0.04	**-1.80**
	only cpg sites	6,720	349	422	25,669	-0.09	-2.75
	transitions	24,757	848	929	134,146	-0.05	-2.04
	transversions	18,010	395	462	68,581	-0.08	-2.20
Papuan, Han, *Den3*, Chimp	all sites	71,720	8,606	9,909	1,397,467	-0.07	-9.5

	no cpg sites	60,186	6,052	7,040	1,225,758	-0.08	**-8.60**
	only cpg sites	11,534	2,554	2,869	171,709	-0.06	-4.211
	transitions	48,439	5,944	6,801	927,866	-0.07	-7.52
	transversions	23,281	2,662	3,108	469,601	-0.08	-5.87
Papuan, French, *Den3*, Chimp	all sites	71,440	8,886	10,258	1,397,118	-0.07	-10.0
	no cpg sites	59,920	6,284	7,224	1,225,378	-0.07	**-8.30**
	only cpg sites	11,520	2,602	3,034	171,740	-0.08	-5.69
	transitions	48,215	6,168	7,094	927,573	-0.07	-8.08
	transversions	23,225	2,718	3,164	469,545	-0.08	-6.00

a. 'A' refers to an ancestral state and 'D' refers to a derived state. Thus, this column shows the number of sites where populations 1 and 2 share the ancestral allele with population 4 (Ancestral), and population 3 (Derived) has a derived state.
b. The *Z*-score is the difference between the *D*-statistics using all data and the mean of the same statistics for bootstrap replicates divided by the standard deviation for those replicates.

Table 18. Sharing of derived alleles between *Denisova 8* and Neandertals. *Denisova 8* fragments carrying a C to T substitutions at the first or last two positions (*Den8*_deaminated) as well as all fragments (*Den8*_all) are used. Only estimates based on transversions can be used due to errors in the low coverage *Mezmaiskaya1* genome.

	Type of sites	**AADA**	**ADDA**	**DADA**	**DDDA**	**(ADDA-DADA)/ (ADDA+ADDA)**	**Z**
Mez, AltaiNea, *Den8*_deam, Chimp	all sites	15,245	511	376	77,110	0.15	4.49
	no cpg sites	12,142	179	139	64,649	0.13	2.27
	only cpg sites	3,103	332	237	12,461	0.17	4.00
	transitions	8,898	431	313	52,358	0.16	4.34
	transversions	**6,347**	**80**	**63**	**24,752**	**0.12**	**1.44**
Mez, AltaiNea, *Den8*_all, Chimp	all sites	104,707	3,586	2,532	521,739	0.17	13.88
	no cpg sites	87,986	1,382	1,138	441,125	0.10	4.92
	only cpg sites	16,721	2,204	1,394	80,614	0.23	14.12
	transitions	56,272	3,063	2,041	354,226	0.20	14.75
	transversions	**48,435**	**523**	**491**	**167,513**	**0.03**	**1.02**
Mez, AltaiNea, *Den3*, Chimp	all sites	3,392	296	498	77,271	-0.25	-7.15
(sites covered by *Den8*_deaminated)	no cpg sites	2,655	121	177	64,648	-0.19	-3.21
	only cpg sites	737	175	321	12,623	-0.29	-6.75
	transitions	2,371	234	420	52,530	-0.28	-7.44
	transversions	**1,021**	**62**	**78**	**24,741**	**-0.11**	**-1.32**
Mez, AltaiNea, *Den3*, Chimp	all sites	23,573	3,463	2,024	523,579	0.26	20.23
(sites covered by *Den8*_all)	no cpg sites	18,757	1,348	957	441,914	0.17	8.38
	only cpg sites	4,816	2,115	1,067	81,665	0.33	20.29
	transitions	16,333	2,977	1,599	355,784	0.30	21.29
	transversions	**7,240**	**486**	**425**	**167,795**	**0.07**	**2.04**
Mez, AltaiNea, *Den3*, Chimp	all sites	295,159	42,000	24,746	6,550,020	0.26	70.25
(all *Den3* sites, not conditioned on	no cpg sites	232,239	16,149	11,699	5,547,171	0.16	27.62
***Den8*)**	only cpg sites	62,920	25,851	13,047	1,002,849	0.33	67.79
	transitions	205,685	36,111	19,517	4,490,038	0.30	76.13
	transversions	**89,474**	**5,889**	**5,229**	**2,059,982**	**0.06**	**6.35**

5.4 Discussion

The nuclear DNA sequences retrieved from *Denisova 4* and *8* are more closely related to the *Denisova 3* genome used to define the Denisovans as a hominin group than to present-day human or Neandertal genomes. Furthermore, the mtDNA of the two molars form a clade with *Denisova 3*. Thus, the present work extends the number of Denisovan individuals identified by mitochondrial and nuclear DNA from one to three. Although the number of Denisovan individuals is small, restricted to one locality, and differ in age, it is nevertheless interesting to note that the nuclear DNA sequence diversity among the three Denisovans is slightly higher than that found among seven Neandertals although these are widely geographically distributed, but lower than that seen among present-day humans world-wide or among Europeans.

Although the three Denisovans come from a single cave, they may differ significantly in age as indicated by the branch-length of the mtDNA of the *Denisova 8* molar which is shorter than those of *Denisova 4* and the *Denisova 3*, an observation that is congruent with the stratigraphy. If we assume that the mtDNA mutation rate of ~2.5 x 10^{-8}/site/year (95% confidence interval 1.8-3.2) estimated for modern humans (106) applies also to Denisovan mtDNA, *Denisova 8* is in the order of 60,000 years older than *Denisova 3* and *Denisova 4*. A similar or even larger age difference between *Denisova 8* and the other two teeth is suggested by a Bayesian analysis (Suppl. Material; Table 12). Although it is unclear if the mtDNA mutation rate in archaic humans is similar to that in modern humans and thus if the difference in age is as large as this, it is clear that *Denisova 8* is substantially older than *Denisova 4* and *Denisova 3*. This is of interest from several perspectives.

First, the two molars are very large and their morphology is unlike what is typical for either Neandertals or modern humans. Since they differ substantially in age this reinforces the view that Denisovan dental morphology was not only distinct from that of both Neandertals and modern humans, but also a feature typical of Denisovans over an extended period of time, at least in the Altai region. This may prove useful for the identification of potential Denisovan teeth at other sites.

Second, the difference in age between the two Denisovan molars as well as their similar morphology suggests that Denisovans used Denisova Cave at least twice and possibly over a long time, perhaps interrupted by Neandertal occupation(s) (10). Denisovans may therefore have been present in southern Siberia over an extended period. Alternatively, they may have been present in neighboring regions from where they may have periodically extended their range to the Altai.

Third, the *Denisova 8* molar is not only older than *Denisova 4* and *Denisova 3*, its mtDNA differs substantially from the other two. The mtDNA diversity among the three Denisovan individuals is larger than that among seven Neandertals from which complete mtDNA sequences are available (Figure 15C), despite the fact that the Denisovans all come from the same site while the Neandertals are broadly distributed across western and central Eurasia. Notably, the nuclear genome of *Denisova 8* also shows a tendency to be more deeply diverged from the genome of *Denisova 3* than is *Denisova 4* (Figure 17B). Given that the high-coverage genome from the *Denisovan 3* phalanx carries a component derived from an unknown hominin who diverged 1-4 million years ago from the lineage leading to Neandertals, Denisovans and present-day humans (10), it is possible that this component differs among the three Denisovan individuals. In particular, it may be that the older Denisovan population living in the cave carried a larger or different such component. It is also possible that the two diverged mtDNA lineages seen in *Denisova 8* on the one hand and *Denisova 3* and *4* on the other were both introduced into the Denisovans from this unknown hominin as has been suggested for the mtDNA of *Denisova 3* (8, 9). However, more nuclear DNA sequences from Denisovan specimens of ages similar to *Denisova 4* and *8* are needed to address this question fully.

6. A Neandertal from Denisova Cave with ancient spotted hyena contamination

This manuscript is in preparation for publication

Authors: Susanna Sawyer[a], Bence Viola [b,c,d], Michael V. Shunkov [d,e], Anatoly P. Derevianko [d,f], Matthias Meyer[a], Kay Prüfer [a], Svante Pääbo [a].

[a] Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany

[b] Department of Anthropology, University of Toronto, Toronto, ON M5S 2S2, Canada

[c] Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany

[d] Institute of Archaeology and Ethnography, Russian Academy of Sciences, Siberian Branch, Lavrentieva Avenue, 17 Novosibirsk, RU-630090, Russia.

[e] Novosibirsk National Research State University, Pirogova st., 2, Novosibirsk, RU-630090, Russia

[f] Altai State University, Pr. Lenina, 61, Barnaul, RU-656049, Russia

Author contributions:

Contributed materials: Michael Shunkov, Anatoly Derevianko

Performed experiment: Susanna Sawyer

Analyzed morphological data: Bence Viola

Analyzed mitochondrial data: Susanna Sawyer, Kay Prüfer, Matthias Meyer

Analyzed nuclear data: Susanna Sawyer

Wrote the paper: Susanna Sawyer, Bence Viola and Svante Pääbo

6.1 Abstract

To date, four hominin remains have been published from Denisova Cave. A piece of a finger phalanx (*Denisova 3*) and two upper molars (*Denisova 4* and 8) belong to a recently discovered sister group of Neandertals, the Densiovans. The fourth remain, a toe phalanx, belongs to a Neandertal (*Altai 1*). Here we present an almost complete mitochondrial genome as well as 18.4 Mb of the autosomal genome of a second Neandertal from Denisova Cave (*Altai 2*). Based on mtDNA, *Altai 2* is most closely related to *Altai 1* compared to all other published Neandertal mtDNA genomes. We also present partial mtDNA genomes of spotted hyenas from the Pleistocene that contaminated the *Altai 1* and *Altai 2* Neandertals as well as *Denisova 3* and *4*. The hyena contamination is heaviest in *Altai 2*, and belonged to a hyena population that until now has only been found in eastern Russia and China. The hyenas that contaminated the Denisovans, came from a population found in Pleistocene Europe as well as northern Africa today.

6.2 Introduction

Denisova Cave is located in the Siberian Altai Mountains on the Anui River. In 2008, a small piece of a child's finger phalanx (*Denisova 3*) was found in the east Gallery of the cave. Ancient DNA was extracted from the bone, and the nuclear and mtDNA genomes were sequenced to high coverage (7, 8). The nuclear DNA showed that this bone belonged to a new hominin group, the Denisovans, a sister group to Neandertals. Surprisingly, the mtDNA is far deeper diverged, with a divergence from the early modern human-Neandertal ancestor twice as deep as the Neandertal divergence from early modern humans (7). The low-coverage genomes from two molars (*Denisova 4* and 8) showed that an additional two Denisovans left remains in the cave. *Denisova 8* belonged to a Denisovan that lived many millennia earlier than *Denisova 3* or *4*, showing that Denisovan presence in the region occurred at least twice over a long period of time (Chapter 5). In 2010, a second phalanx was found in the East Gallery of Denisova Cave, this time from an adult toe. The DNA preservation proved to be remarkable enough to produce another high coverage genome, which revealed that the individual to whom the toe phalanx belonged was a Neandertal (10), called *Altai 1* here. *Denisova 3* shows more than 0.5% Neandertal admixture from a population more closely related to *Altai 1* than other Neandertals (10). Present-day non-African humans also show Neandertal admixture (1.5-2.1%), from an admixture event that

happened 37-86kya (140) from a Neandertal population that was more closely related to *Mezmaiskaya1*, a 65,000 year old Neandertal from the Russian Caucasus (10).
Although none of the hominin remains from Denisova Cave are dated, Denisovan and Neandertal occupation of the region most likely happened over a long time period in the Pleistocene, possibly between 150,000 and 50,000 years ago (Chapter 5). During this time, the Altai Mountains were also home to a large number of large carnivores, especially the spotted hyena (2, 3, 63). The Pleistocene spotted hyena had a vast range from Europe to eastern Asia (12). Due to its larger size than the current spotted hyena found in sub-Saharan Africa, it was often hypothesized to be a separate species (12, 64), however genetic data shows that the Pleistocene spotted hyenas fall within the variation of present-day spotted hyenas (13, 14). Spotted hyenas are impressive carnivores, with a diet consisting completely of meat (65), and jaws able to crush and digest entire skeletons, leaving behind only teeth, hair, horns and often smaller metapodials (16). They live in clans, and dig large dens, and both hunt and scavenge for food (17). Spotted hyenas today are not fearful of humans (3), suggesting that the Pleistocene spotted hyena was a terrifying opponent to hominins of the time.
In 2011, a second finger phalanx from an adult was found in layer 12 of the East Gallery of Denisova Cave. Here we present the almost complete mtDNA genome and 18.4 Mb of the nuclear genome from this phalanx, called *Altai 2*. We also present the partial mtDNA genome of the spotted hyena contaminant of *Altai 1*, *Altai 2*, *Denisova 3* and *Denisova 4*.

6.3 Results

6.3.1 *Altai 2* morphology

Altai 2 is a distal manual phalanx found in 2011 in the East Gallery of Denisova Cave. The phalanx derives from Layer 12 in square G-3. The official excavation name of this phalanx is *Denisova 9*, however I refer to this phalanx as *Altai 2* in this thesis, to avoid confusion.
The specimen is well preserved, with a complete distal end. Proximally, the palmar aspect of the articular surface is missing. The presence of a large and flattened apical tuft, as opposed to a claw indicates that the specimen is a human (Figure 19). Similarly broad and rounded apical tufts are usually present in Neanderthals, while modern humans usually have narrower and oval apical tufts. There is no indication of acid etching, thuis no indication of the bone having been exposed to stomach acids.

Figure 19. The manual distal phalanx from *Altai 2*. Each square represents 1 cm.

6.3.2 Neandertal mtDNA analyses

Ten libraries were produced from three DNA extractions (Figure 9, Table 2) of bone powder from three areas of the *Altai 2* bone (Figure 7). After enrichment for human mtDNA, sequences from each of the ten libraries were filtered as described in section 4.2.3, and aligned to the human reference mtDNA genome (rCRS, (76)). The sequences were then compared to 69 diagnostic positions where 10 Neandertals differ from 311 present-day humans from around the world. Each of the ten libraries has more sequences containing Neandertal diagnostic bases than human bases (54.3 to 93.3 percent Neandertal, Table 19). The same analysis was done for the combined libraries for Denisovan versus human diagnostic positions. 99.5% of the sequences carry human-like bases. After filtering for sequences with a putative C to T change at the ends of molecules, the percentage of sequences that carry the present-day human base decreases to 0.46 to 8.8% in the ten libraries, indicating that this bone belongs to a Neandertal. After merging of the putative C to T filtered libraries the present-day human contamination rate is 1.9% (95%CI: 1.6-2.3%) (Table 19).

Table 19. Contamination estimates of human mtDNA captured and aligned *Altai 2* libraries. Estimates are given per library as well as after merging of the ten libraries (TOTAL). Present-day human contamination (PD-human cont) is given for all sequences as well as sequences filtered for terminal C to Ts. Spotted hyena contamination is after terminal C to T filtering. 95 % CI in brackets.

			PD-human cont (All seq)			PD-human cont (terminal C>T seqs)			Spotted hyena contamination		
Parent Library	Library	Description	Human-like seqs	Neandertal-like seq	% contamination	Human-like seqs	Neandertal-like seq	% contamination	Hyena-like seqs	Human-like seqs	% contamination
L9366	L9632	UDG	2844	8404	25.3 (24.5-26.1)	12	866	1.4 (0.8-2.4)	413	1968	17.3 (15.9-18.9)
L9367	L9605	UDG	1201	3483	25.6 (24.4-26.9)	3	287	1.0 (0.4-3.0)	33	109	23.2 (17.1-30.8)
L9575	L9586	UDG, u-selec fraction	131	1821	6.7 (5.7-7.9)	3	644	0.46 (0.2-1.4)	50	239	17.3 (13.4-22.1)
L9576	L9587	UDG, u-selec fraction	138	1863	6.9 (5.9-8.1)	4	688	0.57 (0.2-1.5)	48	259	15.6 (12.0-20.1)
L9580	L9591	UDG, u-dep fraction	2470	8957	21.6 (20.9-22.4)	3	104	2.8 (1.0-7.9)	5	31	13.9 (6.1-28.7)
L9581	L9592	UDG, u-dep fraction	2224	7886	22.0 (21.2-22.8)	2	80	2.4 (0.7-8.5)	14	26	35.0 (22.1-50.5)
L9467	R3108	UDG	278	1353	17.0 (15.8-19.8)	4	121	3.2	347	392	47.0 (43.4-50.6)
A9231	L9627	UDG	245	1139	17.7 (15.8-19.8)	5	73	6.4 (2.8-14.1)	269	338	44.3 (40.4-48.3)
A9232	L9628	$NaHPO_4$ wash, no UDG	77	126	37.9 (31.5-44.8)	3	77	3.8 (1.3-10.5)	6	31	16.2 (7.7-31.1)
A9233	L9629	after $NaHPO_4$ wash, no UDG	768	913	45.7 (43.3-48.1)	53	549	8.8 (6.8-11.3)	476	559	46.0 (43.0-49.0)
	TOTAL		11,405	37,661	23.2 (22.9-23.6)	104	5370	1.9 (1.6-2.3)	584	1216	32.4 (30.3-34.6)

The sequences of these ten libraries were subsequently mapped to the Vindija 33.16 Neandertal mtDNA (AM948965, (59)). The average coverage across the mtDNA genome ranged in the ten libraries from 3.9 to 211.2-fold, with a total coverage of 946.3-fold. After C to T filtering the coverage was reduced to 102.8-fold (Table 20).

We calculated the spotted hyena contamination level in the conserved regions of the mtDNA genome as described in section 4.2.5. The spotted hyena contamination levels varied from 13.8 to 46.9% in the ten libraries with a combined contamination level of 32.4% (95%CI: 30.3-34.6%) (Table 19).

The Neandertal mtDNA genome was built as described in section 4.2.6, leaving an average coverage of 98-fold and six unresolved positions (positions 185, 189, 195, 2951, 2964 and 5767 in rCRS coordinates) (Figure 20). A Bayesian phylogenetic mtDNA tree, based on 16,446 positions, shows that the *Altai 2* Neandertal falls together with the *Altai 1* Neandertal (posterior probability of 1) (Figure 21). The *Altai 1* and *Altai 2* Neandertals have ten differences between their mtDNAs. All ten differences are at positions where the *Altai 2* alingment has a coverage above 44 and a consensus support above 94%. Six of these differences are unique to *Altai 2*. Both *Altai 1* and *Altai 2* carry more differences to the other Neandertals (Table 21). There are 30 differences between the *Altai 2* and the MRCA of Neandertals (calculated exactly as in section 4.1.10), which is the highest number of differences to the MRCA when compared to seven other Neandertals (Table 10), five more than to the next highest, *Mezmaiskaya1*.

Table 20. Coverage of the Neandertal and spotted hyena mtDNA genome of the ten libraries from *Altai 2*.

				Human mtDNA capture			Spotted hyena mtDNA capture		
Extract	Parent Library	Description	mg bone used for lib	Avg coverage (all;CT filtered)	Capture efficiency	% uniq	Avg coverage, no human regions (all;CT filtered)	Capture efficiency	% uniq
E1114	L9366	UDG	3	211.2 ; 16.0	14.2	30.1	78.4; 10.5	17.4	19.1
	L9367	UDG	3	168.3 ; 12.3	5.2	32.9	22.4; 3.1	6.6	78.5
	L9575	UDG, u-selec fraction	4.5	35.9 ; 28.7	5.1	15.4	18.7; 15.7	14.4	
	L9576	UDG, u-selec fraction	4.5	36.9 ; 30.0	5.4	15.2	20.5; 17.5	16.8	4.8
	L9580	UDG, u-dep fraction	4.5	209.2 ; 2.7	8.5	44.1	98.1; 3.5	15.6	19.2
	L9581	UDG, u-dep fraction	4.5	186.2 ; 2.2	4.8	59.5	103.7; 3.7	14.7	22.4
E1269	L9467	UDG	1.2	32.1 ; 2.5	13.6	4.6	59.4; 6.8	22.6	8.8
	A9231	UDG	1.2	26.9 ; 2.0	7.9	4.0	43.1; 4.9	17.5	17.5
E3000	A9232	$NaHPO_4$ wash, no UDG	4	3.9 ; 0.6	0.1	80.5	8.4; 1.7	0.8	26.5
E3001	A9233	after $NaHPO_4$ wash, no UDG		35.7 ; 5.9	4.1	9.5	90.0; 17.6	17.0	18.7
	TOTAL	-	-	946.3 ; 102.8	-	-	543; 85.0	-	-

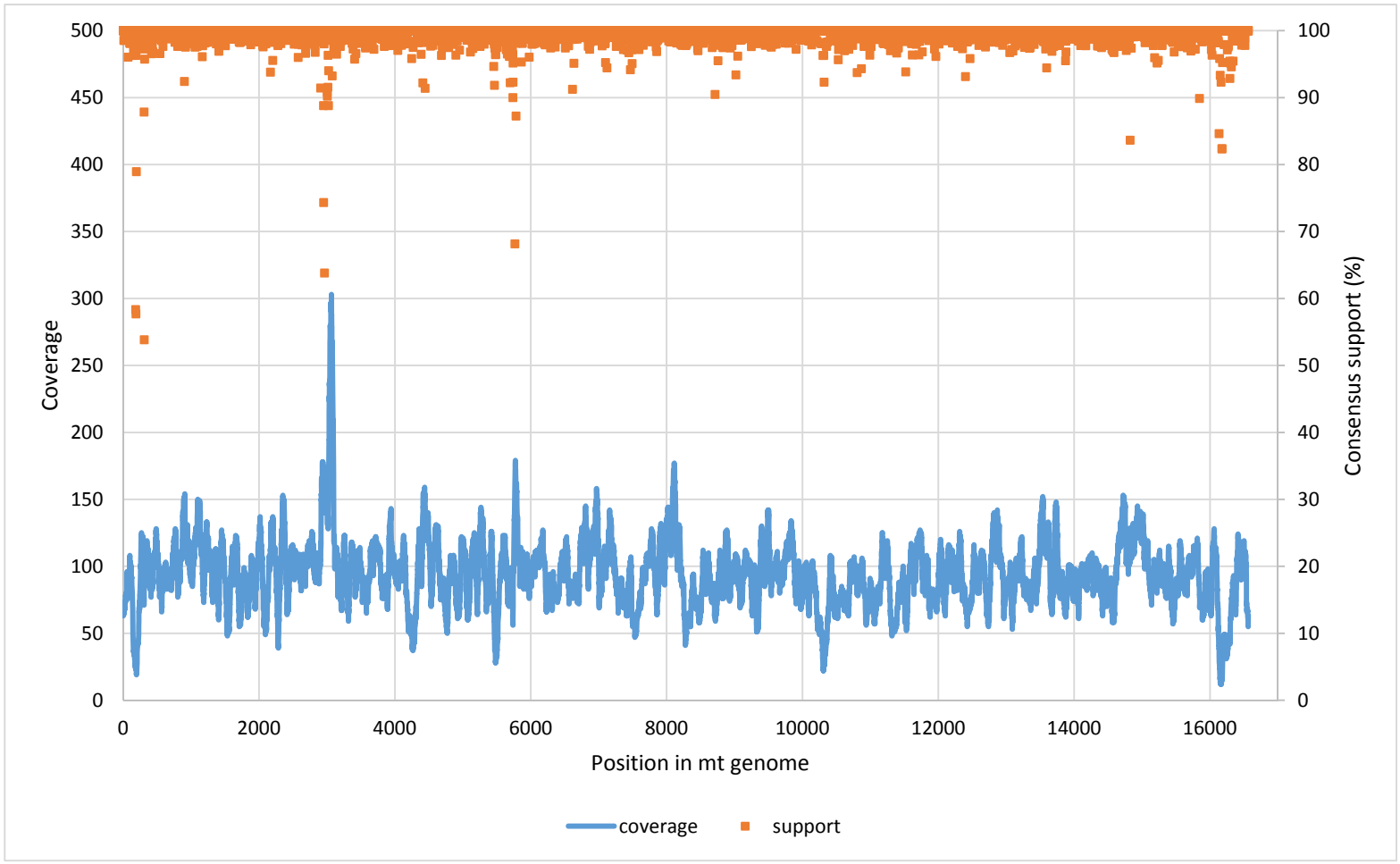

Figure 20. Coverage and consensus support of the merged *Altai 2* libraries after human mtDNA capture, alignment to Neandertal mtDNA and filtering (see section 4.2.6).

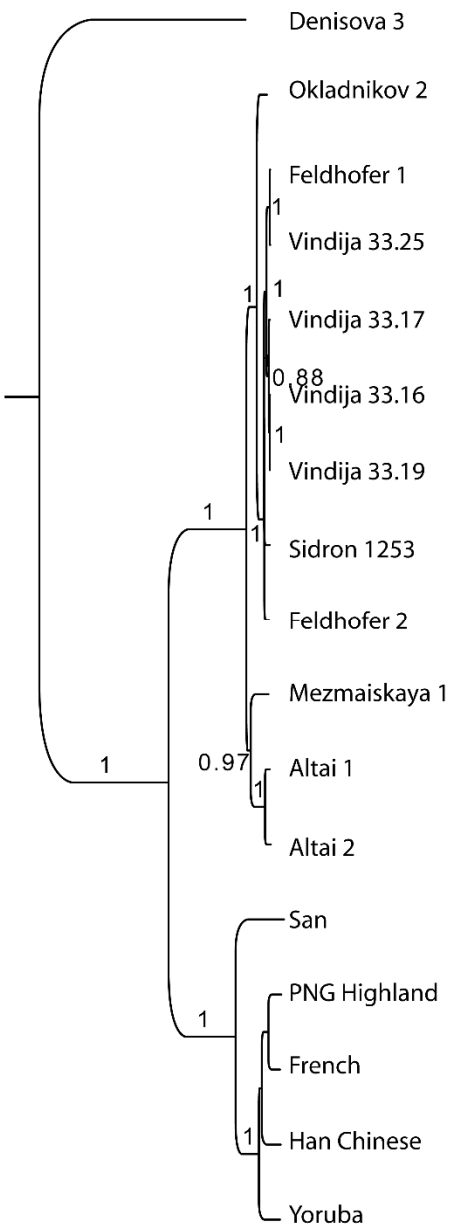

Figure 21. Bayesian tree of eleven Neandertal mtDNAs (including the *Altai 2*), five present-day humans, one Denisovan and a Chimpanzee as an outgroup (not shown). Posterior values are shown.

Table 21. Pairwise differences among ten Neandertals, including the *Altai 2* Neandertal.

	Mez1	**Feld1**	**Vindija 33.25**	**Feld2**	**Sidron**	**Vindija 33.16**	**Vindija 33.19**	**Vindija 33.17**	***Altai 1***
Feldhofer1	43								
Vindija 33.25	43	0							
Feldhofer2	38	9	9						
Sidron1253	41	10	10	5					
Vindija 33.16	44	9	9	8	9				
Vindija 33.19	44	9	9	8	9	0			
Vindija 33.17	43	8	8	7	8	1	1		
Altai 1	32	45	45	40	43	46	46	45	
Altai 2	38	53	53	48	51	54	54	53	10

6.3.3 Autosomal analyses

The ten *Altai 2* libraries were each sequenced without enrichment. Percent endogenous was calculated for each of the ten libraries, after alignment to the human genome (see section 4.2.10 for details) (Table 22). The percent endogenous for the regular libraries of E1114 is highest at 1.2%. The U-selected libraries, L9575 and L9576, have a slight decrease in percent endogenous, but are still double than that of the U-depleted libraries (L9580 and L9581). The two libraries from E1269 have a percent endogenous of 1.0%. The $NaHPO_4$ wash (A9232) has the lowest percent endogenous, half that of the extraction after the $NaHPO_4$ wash (A9233).

Table 22. Percent endogenous (% end) for each of the ten libraries.

Extract	E1114						E1269		E3000	E3001
Library	L9367	L9366	L9575	L9576	L9580	L9581	L9467	A9231	A9232	A9233
% end	1.2	1.2	0.84	0.82	0.46	0.46	1.0	1.0	0.26	0.56

The two U-selected libraries (L9575 and L9576) were pooled and sequenced on one lane of a HiSeq. After filtering for sequences with putative C to T changes (see section 4.1.12) there exist 18.4 Mb of data. Divergence was calculated to *Altai 1*, *Denisova 3* and ten present-day humans on the lineage to the human-chimpanzee ancestor as described in sections 4.1.13 and Figure 17a. The divergence of *Altai 2* is lowest to *Altai 1* (3.31, 95% CI 3.13-3.49), second lowest to *Denisova 3* (10.11, 95% CI: 9.86-10.36), and highest to the ten present-day humans (12.7-13.22) (Figure 22 and Table 23).

In addition the divergence of the *Altai 2* to the *Altai 1* was also compared to the divergences of seven low-coverage Neandertals to *Altai 1* (as described in sections 4.1.13 and 5.3.5). The divergence of *Altai 2* to *Altai 1* is highest (3.31, 95% CI 3.13-3.49) than that of the other Neandertals to *Altai 1* (2.47-2.62, Table 16).

The *Altai 2* sequences were also compared to derived positions on the Denisovan, Neandertal, present-day human (based on a single Mbuti individual), Denisovan-Neandertal, human-Neandertal, and human-Denisovan lineages. The fraction of sequences that share the derived states of each lineage were calculated as described in Meyer *et al* (113). The *Altai 2* shares the

highest fraction of derived positions with the Neandertal, Denisovan-Neandertal and human-Neandertal lineages (>0.7), while it shares less with the other three lineages (<0.1) (Table 24).

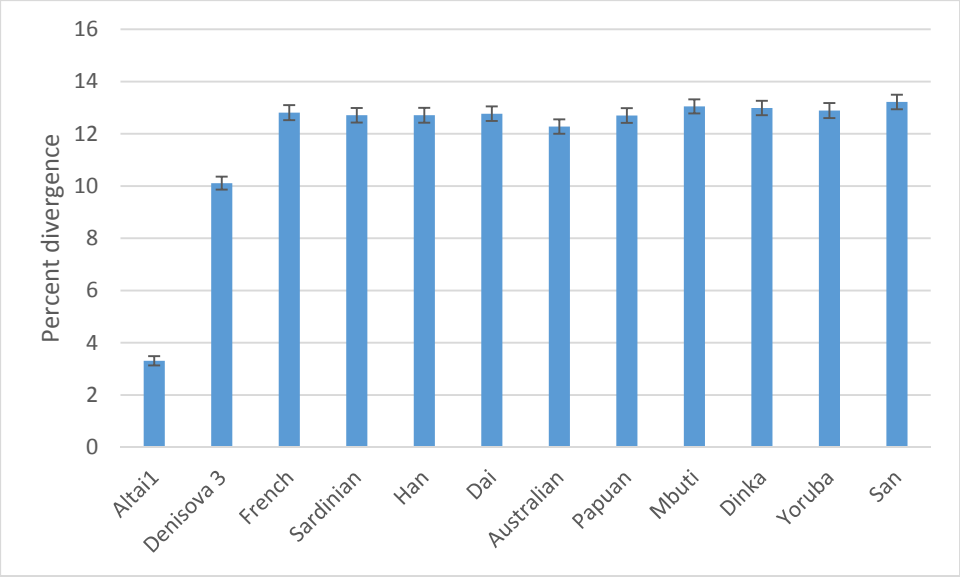

Figure 22. Autosomal divergences of *Altai 2* to the *Altai 1* Neandertal genome, the *Denisova 3* genome, and ten present-day human genomes calculated as in Figure 17A. All estimates are based on DNA fragments that carry putative deamination-induced C to T substitutions. Bars indicate 95% confidence intervals.

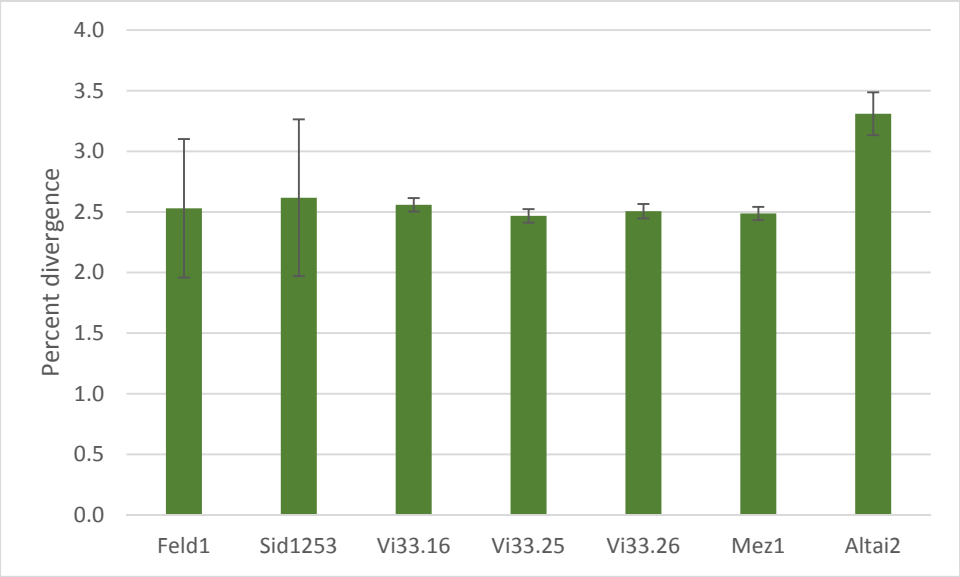

Figure 23. Autosomal divergences of seven low-coverage Neandertal genomes to the high-coverage *Altai 1* Neandertal genome calculated as in Figure 17A. All estimates are based on DNA fragments that carry putative deamination-induced C to T substitutions, except for *Mezmaiskaya1*. Bars indicate 95% confidence intervals.

Table 23. Divergences for the deaminated sequences. Divergence is given as the percent divergence of *Altai 2* along the branch to the human-chimpanzee ancestor from the high-coverage genomes given in the first column. Percent divergences and 95% CI (low and high) are given.

High-coverage genomes	Shared	Genome	*Altai 2*	% div	% div low	% div high
Altai 1	66250	2268	66250	3.31	3.13	3.49
Denisova3	61154	6876	61154	10.11	9.86	10.36
French	60777	8930	60777	12.81	12.52	13.10
Sardinian	60242	8772	60242	12.71	12.43	12.99
Han	60653	8829	60653	12.71	12.42	13.00
Dai	60738	8893	60738	12.77	12.49	13.05
Australian	60724	8499	60724	12.28	12.00	12.56
Papuan	60356	8782	60356	12.7	12.42	12.98
Mbuti	60636	9102	60636	13.05	12.78	13.32
Dinka	60171	8985	60171	12.99	12.71	13.27
Yoruba	60519	8956	60519	12.89	12.60	13.18
San	60487	9215	60487	13.22	12.94	13.50

Table 24. Number of nuclear sequences from the *Altai 2* that look derived or ancestral on various lineages. The total number of sequences covering the derived positions for each lineage are also given, as well as the fraction of sequences that are derived for each lineage.

Lineage	Ancestral	Derived	Total	Fraction derived
Denisova	2250	51	2301	0.02
PD Human - Denisova	403	35	438	0.08
Neandertal	453	1611	2064	0.78
Neandertal - Denisova	111	1104	1215	0.91
Neandertal - PD Human	43	539	582	0.93
All	541	23948	24489	0.98
PD Human	2941	41	2982	0.01

6.3.4 Animal contamination analyses based on mtDNA

A coverage distribution of the combined libraries aligned to the Vindija 33.16 Neandertal mtDNA reveals 3 major spikes in coverage: position 850-900, which falls into the 16S rRNA gene, 2200-3000, which falls into the 12S rRNA gene and 5600-5750, which falls into the tRNA-Asn gene (Figure 20). Both the 12S and 16S gene of the mtDNA are well known for being highly conserved, so this is a possible signal of animal contamination. We therefore recaptured the ten libraries using probes designed from 242 mammalian genomes (98). After capture, the filtered sequences were aligned using BLAST and ranked by taxanomic identification number. The ranking of each library as well as the combined sequences of all ten libraries shows that most sequences align to either modern humans or Neandertals (Table 25). In all ten libraries the taxon with the third most sequences aligned is the spotted hyena, which has 2.4 to 5.6 times more aligned sequences than the next mammal. When the ten libraries are combined, 6392 sequences align to either Neandertal or modern human. 477 sequences align to the spotted hyena, 4.8 times more than the 100 sequences that align to the woolly rhinoceros.

Table 25. BLAST alignment of each library from *Altai 2* after capture with 242 mammalian mtDNA genomes. The six taxons with the most hits are shown. Results after combining the libraries are also given.

L5483		**L5484**		**L5485**		**L5486**		**L5487**		**L5488**		**L5489**		**L5490**		**L5491**		**L5492**		**All combined**	
Sp.[1]	#[2]	Sp.	#	Sp.	#	Sp.	#	Sp.	#	Sp.	#	Sp.	#	Sp.	#	Sp.	#	Sp.	#	**Sp.**	**#**
Nea	439	Nea	690	Nea	118	Nea	750	Nea	410	Nea	1254	Nea	460	Nea	80	Nea	93	MH	72	**Nea**	**4344**
MH	286	MH	403	MH	44	MH	156	MH	108	MH	622	MH	236	MH	36	MH	85	Nea	50	**MH**	**2048**
Cro	42	Cro	45	Cro	42	Cro	78	Cro	34	Cro	112	Cro	35	Cro	33	Cro	6	Cro	50	**Cro**	**477**
Bos	8	Coe	12	Equ	8	Vul	32	Coe	14	Bos	20	Bos	10	Coe	11	Cer	2	Ovi	12	**Coe**	**100**
Urs	8	Bos	10	Equ	7	Cer	22	Cer	8	Coe	18	Urs	5	Equ	10	Vul	2	Equ	9	**Bos**	**81**
Cer	5	Cer	8	Urs	6	Cer	11	Cer	7	Cer	15	Vul	5	Bos	8	Bos	2	Coe	9	**Vul**	**71**

1. Sp. – Species Name:
   Nea: Neandertal, MH: Modern human, Cro: Spotted hyena (*Crocuta crocuta*), Bos: Cow (*Bos taurus*), Urs: Brown bear (*Ursus arctos*), Cer: Deer (*Cervus elaphus* or *albirostris*, or *Dama dama*), Coe: Woolly rhinoceros (*Coelodonta antiquitatis*), Equ: Horse (*Equus caballus* or *ovodovi*), Vul: Fox (*Vulpes vulpes*), Ovi: Sheep (*Ovis ammon hodgsoni*)
2. # – Number of sequences that aligned in the BLAST search to the specific species

6.3.5 Spotted hyena mtDNA analyses in *Altai 2*

Based on the all mammalian capture, the spotted hyena contributed the most DNA (aside from humans) to the sequences in the *Altai 2* bone. Therefore the ten libraries of the *Altai 2* were also captured with and aligned to the spotted hyena mtDNA genome (NC020670, (97)) as described in section 4.2.7. After filtering (see section 4.2.3), the coverage across the spotted hyena genome showed significant peaks in conserved regions, due to the large amount of Neandertal mtDNA (Figure 24). Therefore regions that map to the human mtDNA genome were removed (see section 4.2.7), reducing the spotted hyena mtDNA genome from 17,138 to 14,478 positions. After human-mappable regions were removed, the average coverage for each of the ten libraries varied from 8.4 to 103.7-fold, with a combined coverage of 543-fold (Table 20, Figure 24). After filtering for of C to T sequences, the coverage drops to 85-fold (Table 20).

Sequences from the ten libraries were combined and a consensus of the spotted hyena sequences of both before and after C to T filtering were called, requiring at least 5-fold coverage and 80% consensus support. 13,724 positions of the 14,478 possible were used to call a consensus for all sequences, while 12,485 positions were used to call a consensus of the C to T filtered sequences.

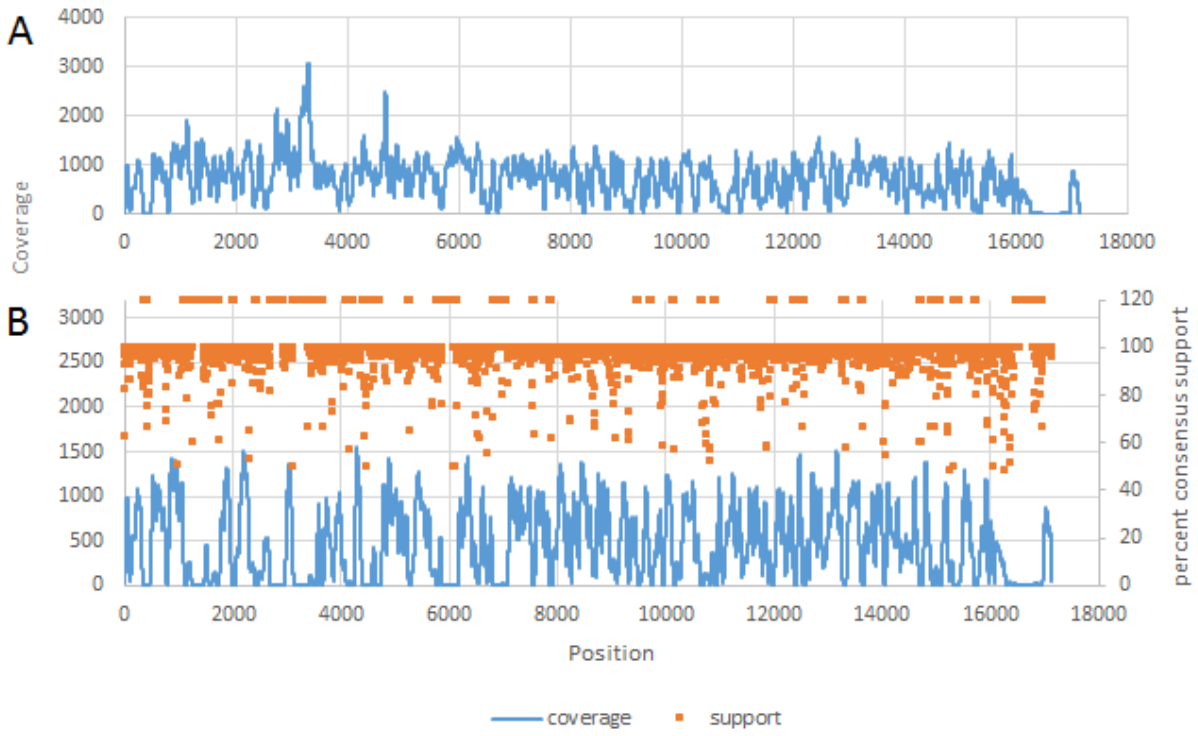

Figure 24. Coverage and consensus support across the spotted hyena captured and aligned sequences of *Altai 2*. Positions with (A) and without (B) human-mapped regions are shown. Positions with a coverage below 5-fold are shown with a consensus support of 120%.

The library made from SP3388, the spotted hyena bone from the same gallery and layer of Denisova Cave as *Altai 2*, was captured using the same spotted hyena probes used to capture the libraries from *Altai 2*. After sequencing, the sequences were filtered and aligned and used to create a consensus sequence as described in section 4.2.7. The average coverage is 1596-fold (Figure 25). The consensus is made up of 17,093 of the 17,138 positions possible when a minimum of 5-fold coverage and consensus support of 80% are required. The 3' C to T percent is 35.0% (95% CI: 34.6-35.4%) and the 5' C to T percent is 31.0% (95% CI: 30.7-31.3%).

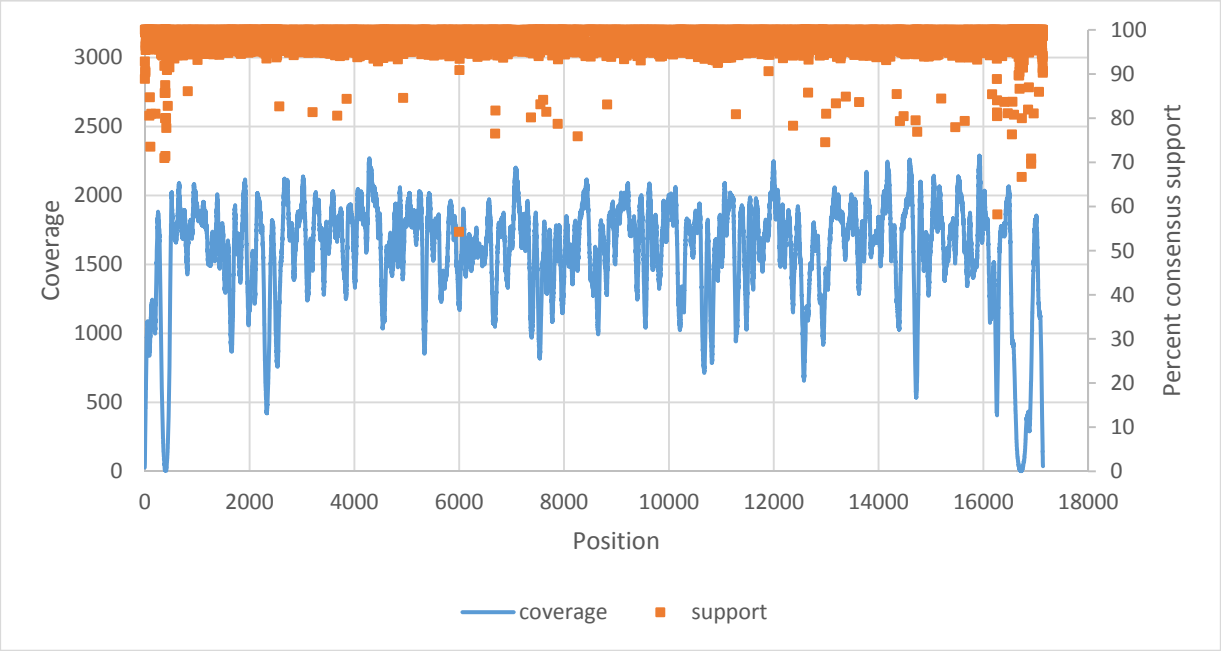

Figure 25. Coverage and consensus support across the spotted hyena captured and aligned sequences of SP3388.

The complete mtDNA genomes of two Pleistocene spotted hyenas from Coumere Cave, France (CC8 (NC020670) and CC9 (JF894379), (97)), an extant spotted hyena (Kira (zoo name), JF894377, (97)), an extant striped hyena (Cerza (zoo name), NC020669, (97)), a mongoose (NC006835), and the partial mtDNAs of the *Altai 2* spotted hyena sequences and the Denisova Cave spotted hyena from SP3388 (see section 4.2.7 for more details), were used to calculate the pairwise differences among hyenas (Table 26). The two Pleistocene spotted hyenas, CC8 and CC9, have identical mtDNA genomes. The Denisova Cave spotted hyena from bone SP3388 is the closest to the two French Pleistocene hyenas with 21 differences to both, while carrying 82 differences to the extant spotted hyena, JF894377. The spotted hyena sequence consensus gained

from the *Altai 2* Neandertal bone carries between 382 and 404 differences to the other spotted hyenas, about five times higher than the second largest number of differences within spotted hyenas and three times less than the number of differences between striped and spotted hyena. The number of differences between striped and spotted hyenas does not differ greatly between the spotted hyena individuals, including the *Altai 2* spotted hyena. Filtering for C to T sequences does not have a significant effect.

Table 26. Pairwise differences among five spotted hyenas, one striped hyena and one mongoose. Number of differences for all sequences are shown. The number of differences for sequences filtered for C to T differences are shown in parentheses. The accession number of the individual is given where available.

#	Individual	Group	1	2	3	4	5	6
1	*Altai 2*	Spotted hyena						
2	NC020670 (CC8 ref genome)	Spotted hyena	383 (349)					
3	JF894379 (CC9)	Spotted hyena	383 (349)	0				
4	SP3388	Spotted hyena	382 (350)	21 (19)	21 (19)			
5	JF894377 (Kira)	Spotted hyena	404 (370)	81 (59)	81 (59)	82 (62)		
6	NC020669 (Cerza)	Striped hyena	1297 (1191)	1292 (1167)	1292 (1167)	1299 (1174)	1294 (1176)	
7	NC006835	Mongoose	2245 (2062)	2239 (2045)	2239 (2045)	2233 (2041)	2233 (2043)	2204 (2011)

A Bayesian phylogenetic tree was constructed of these five spotted hyenas, the striped hyena and the mongoose as an outgroup (see section 4.2.9 for more details) (Figure 26). The SP3388 Denisova Cave spotted hyena falls together with the CC8 and CC9 spotted hyenas from France (posterior probability of 1). The spotted hyena sequences from the *Altai 2* falls together with the other spotted hyenas with a posterior probability of 1, but is the most diverged (Figure 26).

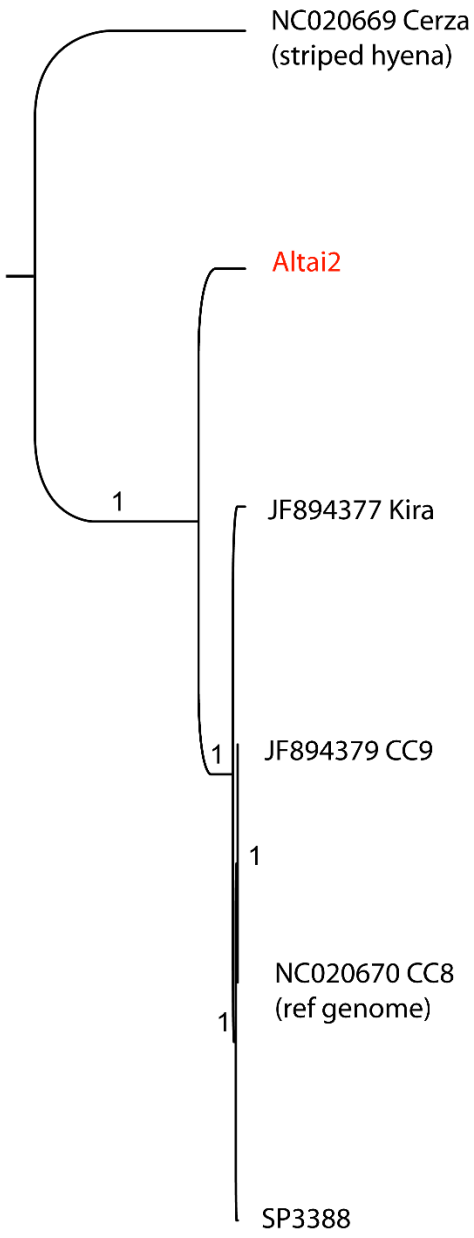

Figure 26. A Bayesian mtDNA tree of the *Altai 2* spotted hyena consensus (not filtered for C to T changes), the spotted hyena from Denisova Cave (SP3388), three spotted hyenas (CC8, CC9, Kira), one striped hyena (Cerza) and a mongoose as outgroup (not shown). Posterior values are given. The tree is based on 12,849 positions.

A pairwise comparison was done of 234 bps of the cytochrome b gene of the mtDNA genomes of 57 spotted hyenas, 2 aardwolves, 18 striped hyenas, 6 brown hyenas and one mongoose (see section 4.2.9 for method and accession numbers). The 57 spotted hyenas are represented by 13 sequences as many of the individuals carry the same 234 bp sequence. They have been labeled as sequences A-M (Table 27). The most differences (9 to 11 differences) are between sequence I (the sequence shared by three from teeth from Da'an Cave, China (14) and one individual from the Geographical Society cave near Vladivostok, Russia, (13)) and sequences B-G and J-M. The *Altai 2* Neandertal spotted hyena is the only individual with sequence H and only differs from sequence I at one position. The two aardwolves each have a distinct sequence (N and O), the 6 brown hyenas all share one sequence (sequence P), and the 18 striped hyenas reduce to 3 sequences (Q, R and S) (see Table 27).

A Bayesian tree was constructed of the 234 bps of the cytochrome b gene of the mtDNA genomes of the 13 spotted hyenas sequences (A-M), the 2 Aardwolf sequences (N and O), the brown hyenas sequence (P), the 3 striped hyena sequences (Q, R and S) and one mongoose as an outgroup (see section 4.2.9 for method and Appendix Table 1 for accession numbers) (Figure 27). Clades within the larger spotted hyena clade are grouped by color if the posterior probability is above 0.94. Thus there are four major clades (green, yellow, pink and blue). Spotted hyena samples from Africa are from living individuals or from museum samples that are at most 150 years old, and therefore represent present-day spotted hyenas. Samples from outside Africa are from spotted hyenas from the Pleistocene.

The pink clade is found exclusively in southern Africa in present-day spotted hyenas with one individual coming from farther north, in Sudan. The yellow clade is found only in Pleistocene spotted hyenas from Europe (Figure 27). The blue clade has the most individuals and is present in both present-day northern-Africa as well as Pleistocene Europe (Figure 27). Two of the spotted hyena individuals from Denisova Cave (the individual sequenced here, SP3388, and the ~42k year old AJ809327, (141)) fall into the larger blue clade. The green clade is made up of five individuals, three from the teeth from Da'an Cave, China (14), one from the Geographical Society cave near Vladivostok, Russia (13), and the consensus sequence of the spotted hyena sequences from the *Altai 2* bone from Denisova Cave.

Table 27. Pairwise differences in 234 bps of the cytochrome b gene among the mtDNA genomes of 57 spotted hyenas, 2 aardwolves, 18 striped hyenas, 6 brown hyenas and one mongoose. Individuals with no sequence differences between them are grouped and highlighted with the same color.

| | Accession # | Group | Sequence Type | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 |
|---|---|---

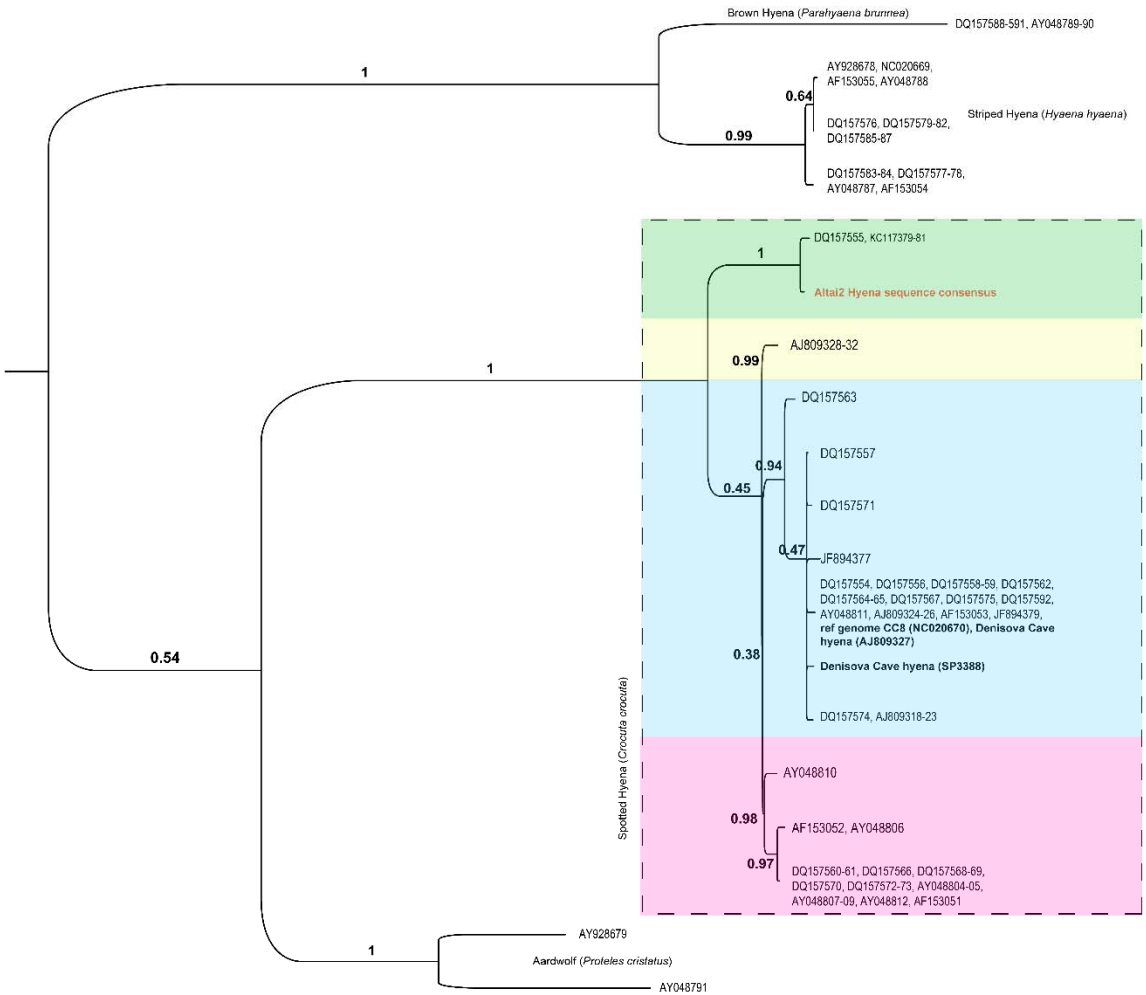

Figure 27. A Bayesian tree of the 234 bps of the cytochrome b gene in the mtDNA of 13 spotted hyena sequences (based on 57 individuals), two aardwolves, three striped hyena sequences (based on 18 individuals), one brown hyenas sequence (based on six individuals) and one mongoose (outgroup: not shown). Posterior values are shown if above 0.8 or at important nodes. Spotted hyenas were grouped together by color if their clade had a posterior probability above 0.94. The colors correspond to the colors in the map in Figure 28. Accession numbers of all individuals are given.

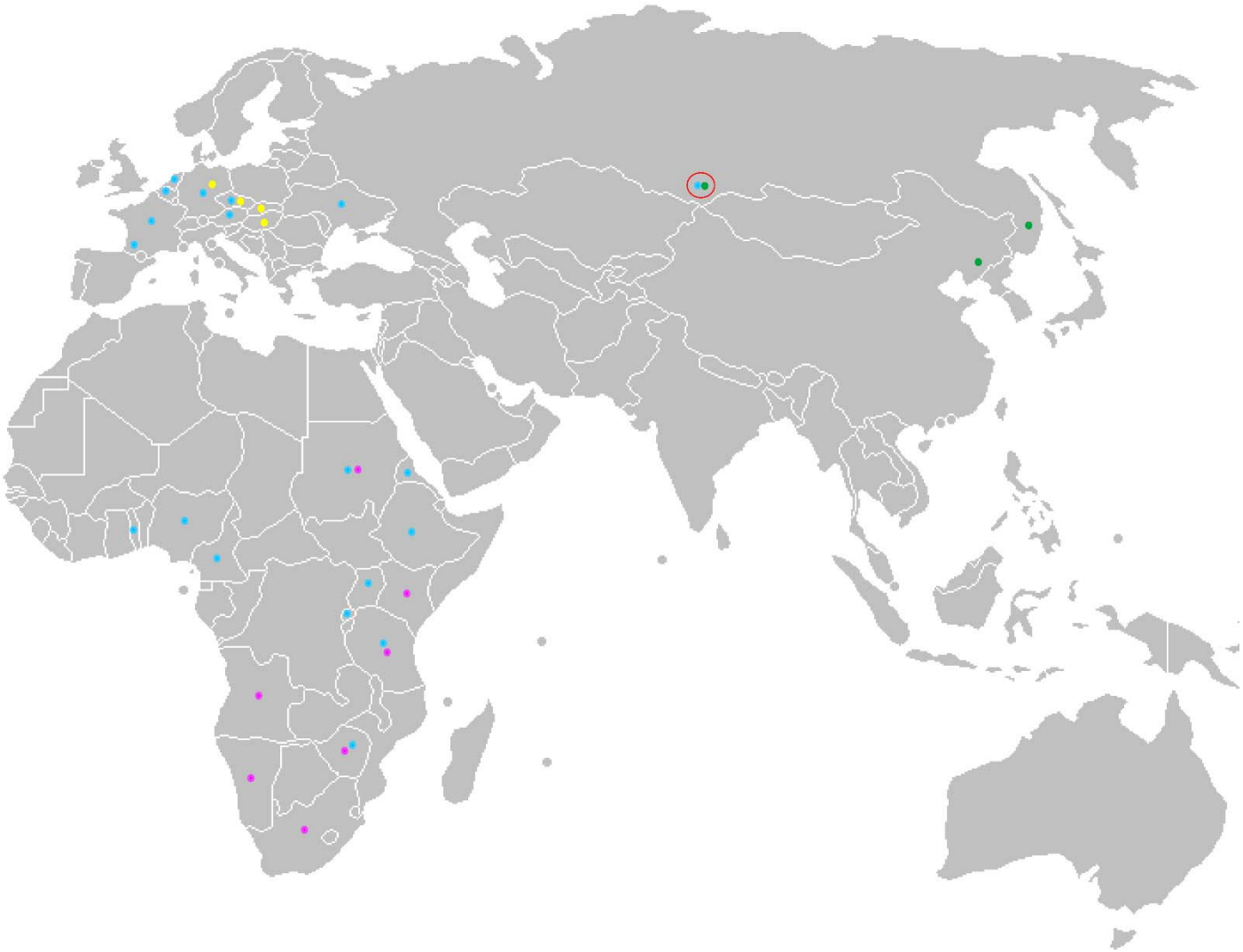

Figure 28. A map of the old world showing the spotted hyena for which the 234 bps of the cytochrome b gene in the mtDNA of 13 sequences from 57 individuals were determined. Colors are based on the groupings in Figure 27. Denisova Cave is marked with a large red circle.

We used a newly described method (15) to wash bone powder from the *Altai 2* bone with a phosphate wash and subsequently extracted after the phosphate wash. The sequences gained in the phosphate wash have almost the same average coverage for both the human capture/alignment and spotted hyena capture/alignment (Table 28). The sequences extracted after the phosphate wash on the other hand have twice as many spotted hyena sequences as human sequences (Table 28). The percent spotted hyena contamination is also significantly higher in the extract after phosphate washing than in the phosphate wash (Table 28). These observations are confirmed in the all-mammalian capture (section 6.3.4), where the number of sequences that align to the spotted hyena are less than one tenth the amount aligned to human or Neandertal for the sequences from the phosphate wash. After the phosphate wash, the same number of sequences align to spotted hyena and Neandertal, with ~25% more aligning to modern human (Table 28).

The spotted hyena mtDNA captures have a higher C to T rate than the human mtDNA captures, even when the human mtDNA captured C to T rate is filtered for C to Ts on the opposite fragment ends (3' filtered) (Table 29). We compared the sequence lengths of sequences captured and aligned to the spotted hyena mtDNA genome and sequences captured with human mtDNA probes and aligned to the Neandertal mtDNA genome. While the two sequences sets have approximately the same peak at 45 bps, the Neandertal sequences are longer and the spotted hyena sequences are shorter (Figure 29).

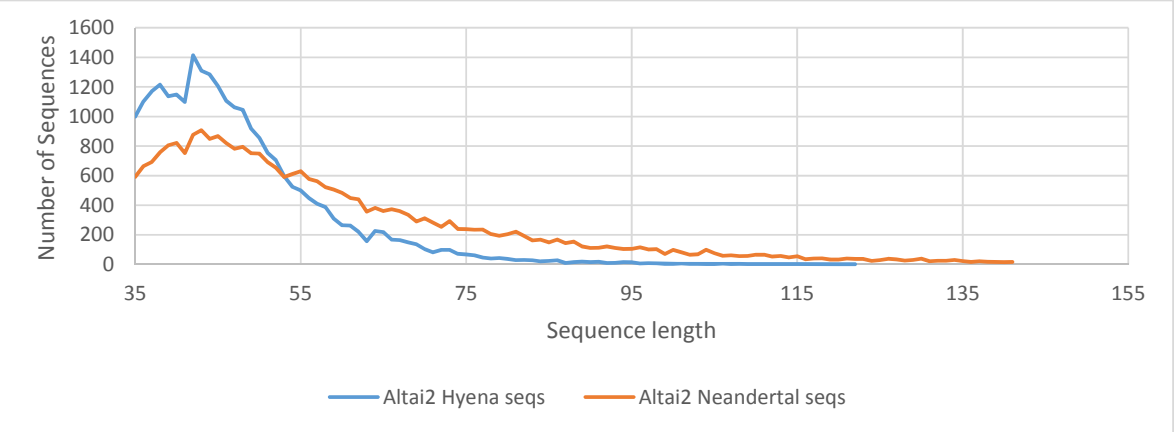

Figure 29. Length distribution differences between spotted hyena and Neandertal sequences in the *Altai 2* bone. Only sequences with C to T differences to the reference genome are shown here.

Table 28. Summary of relevant data from phosphate treatment of bone powder from the *Altai 2* bone. 'Avg cov exhausted' refers to the average coverage in this library for either human/Neandertal or spotted hyena capture/alignment if the library were sequenced to exhaustion. PD-human refers to present-day human. Percent spotted hyena contamination was calculated for the Neandertal sequences as described in section 4.2.5; 95% CI are shown in italics. The top hits in BLAST were calculated after the libraries were captured using the all-mammalian probe set as described in section 4.2.3.

Parent library	**Treatment**	**Avg cov exhausted Neandertal**	**Avg cov exhausted spotted hyena**	**% PD-human contamination in Nea seqs**	**% spotted hyena contamination in Nea seqs**	**Top hits in BLAST**	
						Species	**# seqs align**
A9232	$NaHPO_4$ wash	10.5	9.2	37.9 (*31.5-44.8*)	16.2 *(7.7-31.1)*	Neandertal	93
						Modern human	85
						Crocuta crotuta	6
						Dama dama	2
A9233	After $NaHPO_4$ wash	36.9	95.8	45.7 (*43.3-48.1*)	45.9 *(42.9-49.0)*	Modern human	72
						Neandertal	50
						Crocuta crotuta	50
						Ovis ammon	12

Table 29. Terminal C to T substitutions. "3' filtered" and "5' filtered" refer to fragments that carry C to T substitutions at their 3'- and 5'-ends, respectively. Note that libraries with UDG treatment have a lower 5' C to T rate than 3' rate as described in (8).

		Human mtDNA captures				Spotted spotted hyena mtDNA captures	
		All		**5' filtered**	**3' filtered**	**All (regions with no human seq)**	
Parent Library	**Treatment**	**5' C>T percent**	**3' C>T percent**	**3' C>T percent**	**5' C>T percent**	**5' C>T percent**	**3' C>T percent**
L9366	UDG	8.1 (*7.8-8.6*)	31.9 (*31.0-32.8*)	8.9 (*7.1-11.2*)	37.9 (*31.1-45.3*)	18.6 (*17.4-19.8*)	47.4 (*45.9-48.9*)
L9367	UDG	7.9 (*7.4-8.5*)	31.5 (*30.5-32.6*)	10.5 (*8.2-13.3*)	38.4 (*31.0-46.4*)	19.8 (*17.6-22.2*)	47.8 (*45.0-50.5*)
L9575	U-selected	34.9 (*33.3-36.6*)	94.7 (*94.1-95.2*)	NA	NA	51.6 (*49.1-54.0*)	97.2 (*96.6-97.7*)
L9576	U-selected	37.6 (*36.0-39.3*)	94.9 (*94.3-95.4*)	NA	NA	51.2 (*48.9-53.5*)	97.0 (*96.4-97.5*)
L9580	U-depleted	3.8 (*3.5-4.2*)	3.9 (*3.4-4.5*)	NA	NA	9.7 (*8.9-10.6*)	18.1 (*16.5-19.8*)
L9581	U-depleted	3.2 (*2.9-3.6*)	3.4 (*2.9-4.0*)	NA	NA	9.3 (*8.5-10.1*)	14.3 (*13.0-15.7*)
L9467	UDG	9.9 (*8.6-11.4*)	52.5 (*49.2-55.7*)	14.0 (*8.7-21.8*)	60.0 (*40.7-76.6*)	16.5 (*15.2-17.9*)	66.3 (*64.2-68.3*)
A9231	UDG	10.5 (*9.0-12.2*)	52.4 (*48.7-56.0*)	7.8 (*3.6-16.0*)	60.0 (*31.3-83.2*)	16.5 (*15.0-18.2*)	66.6 (*64.1-69.0*)
A9232	non-UDG NaHPO4 wash	22.5 (*17.8-28.1*)	32.5 (*25.7-40.1*)	21.4 (*7.6-47.6*)	37.9 (*10.8-60.3*)	41.6 (*37.6-45.8*)	50.7 (*45.9-55.5*)
A9233	non-UDG after NaHPO4 wash	27.6 (*25.7-29.6*)	38.0 (*35.5-40.6*)	31.9 (*24.2-40.8*)	38.4 (*29.4-48.2*)	41.5 (*40.2-42.8*)	52.8 (*51.2-54.4*)

6.3.6 Spotted hyena mtDNA analyses in other Denisova Cave individuals

Shotgun sequences from three Denisovans (*Denisova 3*, *4* and *8*) (8), three Neandertals (*Altai 1*, *Altai 2* and *Mezmaiskaya1*) (10), one early modern human (Ust Ishim) (106) and one present-day human were mapped to the mtDNA genomes of the spotted hyena, an arctic ringed seal (NC008428), a cave bear (NC011112) and human. The ratio of hyena to human mapped sequences for *Denisova 3*, *Denisova 4*, *Altai 1* and *Altai 2* are at least one order of magnitude higher than the ratios of seal or cave bear to human (Table 30). There is also a significant difference between the number of sequences that align to hyena and seal or cave bear for *Denisova 3*, *Denisova 4*, *Altai 1* and *Altai 2* when compared to the present-day human or Ust-Ishim. None of the other individuals show this significant difference (Table 31). When looking only at perfectly aligned sequences, *Denisova 3*, *4*, *Altai 1* and *Altai 2* again have at least an order of magnitude more sequences align to hyena than to seal or cave bear. The other individuals have either no sequences or very few, as is the case with Mezmaiskaya where four sequences align to the hyena (Table 32).

We aligned the consensus of the mtDNA spotted hyena sequences of *Denisova 3*, *4* and *Altai 1* each to three published spotted hyenas, a striped hyena, a mongoose as an outgroup as well as the partial mtDNAs of the *Altai 2* spotted hyena sequences and the Denisova Cave spotted hyena from SP3388 (see section 4.2.7 for more details) (Figure 30). The *Altai 2* spotted hyena consensus falls outside the variation of the other spotted hyena mtDNA genomes in every case. Based on 7,507 positions, *Altai 1* also falls outside the variation of the other spotted hyenas (with the exception of *Altai 2*). Both of the Denisovans fall within the variation of the western Pleistocene spotted hyenas, the spotted hyena from Denisova Cave and an extant spotted hyena (based on 8,592 positions for *Denisova 3* and 468 positions for *Denisova 4*) (Figure 30).

The *Denisova 3* hyena sequence covers 143 of the 234 positions of the cytochrome b gene discussed in section 6.3.5. These 143 positions are identical to the JF894377 and CC8 spotted hyenas. The *Altai 1* hyena sequence covers only 62 of these positions and is also identical to the JF894377 and CC8 spotted hyenas at these positions. The *Denisova 4* hyena sequence does not overlap with any of the 243 cytochrome b positions.

Table 30. Number of sequences that align to spotted hyena, arctic ribbon seal, cave bear or human mtDNA genomes for three Denisovans (*Denisova 3*,*4* and *8*), three Neandertals (*Altai1*, *2 Mezmaiskaya1*), one early modern human (Ust Ishim) and one present-day human.

Individual	# seq align to hyena	# seq align to seal	# seq align to cave bear	# seq align to human	ratio hyena/human	ratio seal/human	ratio cave bear/human
Present-day human	33	43	29	391,582	8.43E-05	1.10E-04	7.41E-05
Denisova 3	1262	120	82	481,784	2.62E-03	2.49E-04	1.70E-04
Denisova 4	142	8	8	11,001	1.29E-02	7.27E-04	7.27E-04
Denisova 8	2	5	3	17,503	1.14E-04	2.86E-04	1.71E-04
Altai 1	1583	112	104	465,621	3.40E-03	2.41E-04	2.23E-04
Altai 2	1009	25	31	940	1.07	2.66E-02	3.30E-02
Mezmaiskaya1	6	2	2	64,746	9.27E-05	3.09E-05	3.09E-05
Ust Ishim	42	68	47	130,820	3.21E-04	5.20E-04	3.59E-04

Table 31. P-values of a fisher exact test of number of sequences shown in Table 30. Comparison was between either a present-day human and archaic individuals or the Ust Ishim early modern human and archaic individuals.

Individual 1	Individual 2	Hyena vs. Seal	Hyena vs. Cave bear	Seal vs. Cave bear
Present-day human	*Denisova 3*	< 0.0009	< 0.0009	1
Present-day human	*Denisova 4*	< 0.0009	< 0.0009	0.58
Present-day human	*Denisova 8*	0.69	0.66	1
Present-day human	*Altai 1*	< 0.0009	< 0.0009	0.28
Present-day human	*Altai 2*	< 0.0009	< 0.0009	0.11
Present-day human	*Mezmaiskaya1*	0.14	0.14	1
Present-day human	Ust Ishim	0.54	0.51	1
Ust Ishim	*Denisova 3*	< 0.0009	< 0.0009	1
Ust Ishim	*Denisova 4*	< 0.0009	< 0.0009	0.59
Ust Ishim	*Denisova 8*	0.71	1	1
Ust Ishim	*Altai 1*	< 0.0009	< 0.0009	0.25
Ust Ishim	*Altai 2*	< 0.0009	< 0.0009	0.1
Ust Ishim	*Mezmaiskaya1*	0.06	0.16	1

Table 32. As Table 30, except the number of sequences reported here are only sequences that align perfectly.

Individual	# seq align to hyena	# seq align to seal	# seq align to cave bear	# seq align to human	ratio hyena/human	ratio seal/human	ratio ursus/human
Present-day human	0	0	0	294,538	0	0	0
Denisova 3	900	1	1	94,588	9.51E-03	1.06E-05	1.06E-05
Denisova 4	92	0	0	3,255	2.83E-02	0	0
Denisova 8	0	0	0	5,417	0	0	0
Altai 1	541	1	11	162,907	3.32E-03	6.14E-06	6.75E-05
Altai 2	29	0	1	95	0.31	0	1.05E-02
Mezmaiskaya1	4	0	0	27,748	1.44E-04	0	0
Ust Ishim	0	0	0	111,780	0	0	0

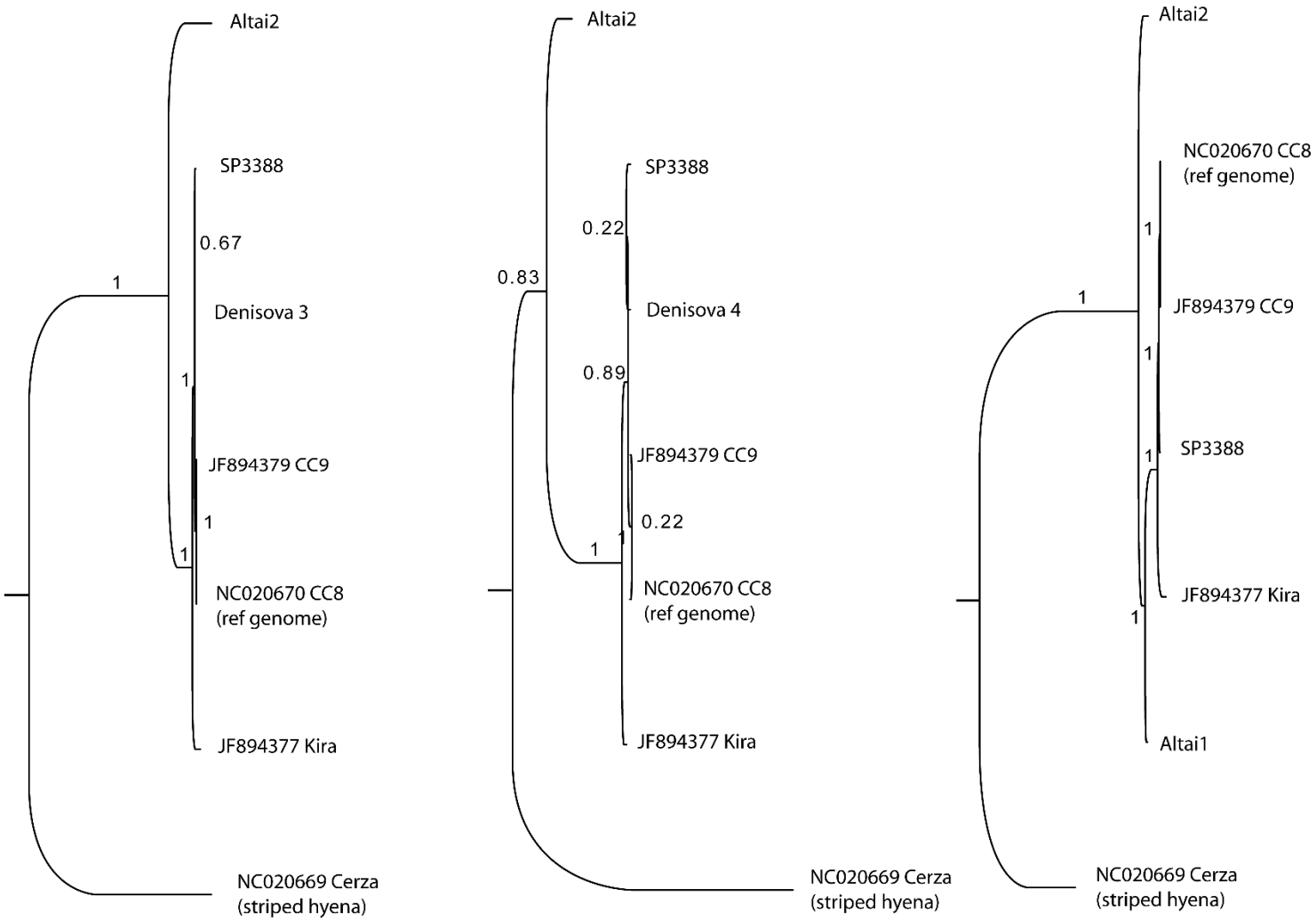

Figure 30. Separate bayesian phylogenetic mtDNA trees of the spotted hyena sequences of *Denisova 3*, *4* and *Altai 1*, each with the *Altai 2* spotted hyena consensus, the spotted hyena from Denisova Cave (SP3388), three additional spotted hyenas, one striped hyena and a mongoose as outgroup (not shown). The number of positions used to build each tree were 8,592 for *Denisova 3*, 468 for *Denisova 4* and 7,507 for *Altai 1*.

6.4 Discussion

The finger bone from the *Altai 2* individual shows clear morphological signs of being human. When aligned to the present-day human mtDNA and nuclear genomes, the *Altai 2* is clearly a Neandertal.

This Neandertal bone is substantially contaminated with ancient spotted hyena DNA. Nevertheless we were able to build a Neandertal mtDNA genome free of hyena and present-day human contamination with only six positions missing, which can therefore be used for comparative mtDNA analyses. The *Altai 2* falls together with the *Altai 1* Neandertal from the same cave. The *Altai 2* finger bone was found in a deeper layer of the same gallery as the *Altai 1* (layer 12 versus layer 11.4), however the two mtDNA genomes only have 10 differences between them, three to five times less than the differences they carry to other Neandertal individuals. So far, in the nine Neandertal mtDNAs published, all of the western European Neandertals cluster together, while the eastern Neandertals from Mezmaiskaya, Okladnikov and Denisova fall outside the variation of the European Neandertals (Figure 21).

The nuclear genome data generated from the *Altai 2* finger bone is not enough to effectively eliminate hyena or present-day human contamination. We examine only sequences with putative cytosine deamination at the fragment ends, which reduces contamination (see Chapter 5). Filtering for deaminated Cs does not however eliminate the ancient hyena sequences. The data generated shows that based on nuclear DNA, the divergence of the *Altai 2* is lower to the high-coverage *Altai 1* Neandertal than to the Denisovans or present-day humans. When comparing divergences of other low-coverage Neandertals and the *Altai 2* to the *Altai 1*, the divergences of the other Neandertals are all significantly lower. It is possible that either residual present-day human contamination, or more likely, substantial hyena contamination is inflating the divergence of the *Altai 2* to the *Altai 1* Neandertal.

We had hoped to reduce the hyena contamination from our extract by first washing the bone powder with sodium phosphate before extracting DNA from the now washed bone powder. It has been shown that sodium phosphate washing preferentially removes the microbial contamination in ancient bones, allowing endogenous DNA to be extracted after washing, possibly due to the microbial DNA being bound to the outer Ca^{2+} of a hydroxyapatite compound, while the endogenous DNA, which bound first, is bound farther inside and harder to wash off (15). Present-day human contamination is not preferentially washed off however, maybe due to

the fragments being protected in skin flakes (15). Contrary to our expectations the sodium phosphate wash did not reduce the hyena contamination. In fact the extraction after sodium phosphate washing has significantly more hyena contamination as well as higher present-day human contamination (Table 28). One explanation for this could be that the hyena DNA was already present in large amounts while the Neandertal cells were lysing, maybe as the hyena was eating a freshly killed Neandertal, and so both populations of DNA had equal chances to bond to the hydroxyapatite. This extract may have also come from a pocket of the bone where there happened to be more contaminant hyena DNA than endogenous Neandertal DNA, as is also the case for E1269 (Table 20), explaining the higher amounts of hyena DNA. However the hyena DNA does seem to be more degraded, since the hyena DNA is more fragmented and has a higher C to T deamination at the fragment ends than the Neandertal DNA (Table 29 and Figure 29). This degradation could be an indication that the DNA was more exposed than the endogenous Neandertal DNA and could also be due to the source of the hyena DNA, from saliva or feces/urine. The sodium phosphate wash does contain more microbial background than the extraction after washing (0.26% endogenous in the $NaHPO_4$ wash to 0.56% endogenous afterwards).

The substantial amount of hyena contamination in the *Altai 2* finger bone gives us the opportunity to study the hyena that contaminated the bone. We are able to reconstruct a mostly complete mtDNA of a hyena, although the low consensus support across some of the mtDNA genome indicates that there were possibly more hyenas or other animals that also contributed to the contamination (Figure 24). The contaminating hyena is clearly a spotted hyena, as expected since the Pleistocene hyenas that inhabited Europe and Asia have been shown to have mtDNAs that fall within the variation of present-day spotted hyenas (13, 14). However this spotted hyena sequence is quite diverged from the four other complete mtDNA genomes of extant and Pleistocene spotted hyenas, including the mtDNA genome of a spotted hyena bone from the same layer and gallery from Denisova Cave.

A comparison between the cytochrome b gene of 57 spotted hyenas both recent and ancient, including a second spotted hyena from Denisova Cave (AJ809327, (141)), shows that the *Altai 2* hyena contaminant only has one difference to the cytochrome b gene of Pleistocene hyenas from the far east, from caves near Vladivostok and north eastern China.

When we examined the other individuals from Denisova Cave, *Denisova 3*, *4*, *8* and *Altai 1*, we find that all except *Denisova 8* also show detectable levels of hyena contamination (Tables 30-32). A closer look at these sequences reveals that the two Denisovans were contaminated with spotted hyenas closer to the western European and African hyenas, while the *Altai 1* contaminating hyena falls outside the variation of these spotted hyenas, although it is not as diverged as the *Altai 2* hyena.

These results could stem from various scenarios. First it is possible that Pleistocene hyenas scavenged or ate both Denisovans and Neandertals living in the area and brought the remains into Denisova Cave. Since none of the bones have any detectable bite marks or acid etching from stomach acids, these small bones may have been regurgitated by the hyenas in hair balls (16). Hyenas are known for leaving behind both teeth and metapodials after eating large prey (12, 16). This could be a possible explanation for why only such small remains of Denisovans and Neandertals are found in the cave. The larger bones would have been easily crushed by the hyenas while eating, leaving behind only morphologically indistinct remains. Denisova Cave has thousands of such small bone fragments (3).

It is also possible however that the remains from Denisovans and Neandertals were deposited in other ways, and that the hyena contamination occurred afterwards. The hyenas could have found or dug up the remains, as spotted hyenas in Africa are known for grave robbing (3). Hyenas could have peed and defecated in the cave as well, and this excrement could have mixed with water to form a soup in which the archaic human remains sat and absorbed the hyena DNA. The differing amount of hyena contamination in the various bones could have been a result of either differing amounts of hyena contact or of the remains lying in parts of the cave with more or less hyena defecation. However, since it is possible that more than one hyena contaminated the *Altai 2* bone, it is not possible to predict which scenario occurred.

Regardless of how the archaic human remains were contaminated with hyena DNA, this study is the first example of such extreme animal DNA contamination in a Neandertal or Denisovan. While Denisova Cave does show evidence of human occupation (2) some of the homins may have been dragged into the cave by hyenas. Based on spotted hyenas in Africa today, their ranges can vary widely based on prey density, anywhere between 40 and 1000 km^2 (66), but they do not often drag prey very far from a kill site (17), and thus the hominds were most likely within a few kilometers of the cave. It is already known that Neandertals lived in the region (Okladnikov and

Chagyrskaya Caves (15, 44)), but this indicates that Denisovans also must have been close to the cave if they did not occupy it.

Denisova Cave was a meeting point of both eastern and western hominids (Denisovans and Neandertals) as well as eastern and western populations of Pleistocene spotted hyenas. Of note is that the hyena DNA found in the eastern hominid group (Denisovans) comes from a western clade of hyenas, while the hyena DNA in the western hominid group (Neandertals) comes from an eastern clade of hyenas. Direct dating of hyenas from the Denisova cave could reveal when these hyena populations and, by extension, when Neandertals and Denisovans inhabited the region. One of the hyenas from Denisova Cave (AJ809327 (141)), whose cytochrome b gene falls into the blue clade (Figures 27 and 28) has been directly C^{14} dated to 42,300 +940/-840 years old. In addition, one of the three hyena teeth from Da'an Cave in northern China was dated to 35,520 +/- 230 years old, while the hyena from the Geographical Society Cave from eastern Russia was dated to 48,650 +2380/-1840 years old. The cytochrome b genes of both of these eastern hyenas fall into the green clade (Figures 27 and 28) with *Altai 2*. These initial dates indicate that the two populations overlapped in time, however a direct date of a hyena from Denisova Cave that stems from the far eastern hyena population would be needed to shed light on this question. More genetic data, including nuclear DNA data, from Pleistocene hyena populations across Eurasia would also be of interest.

7. Discussion

7.1 Hominin occupation of the Denisova Cave region

Denisovans are a new hominin group that are the first to be identified based solely on DNA (7-9). The only known remains before the work presented here were a small morphologically indistinct piece of a finger phalanx from a young child and possibly a large upper third molar, although only the mtDNA was known (9). With this work we can confirm that the large molar did indeed belong to a Denisovan. We can also confirm that the large size of this molar is not a morphological aberration. Instead it was a constant Denisovan feature over a long period of time since an additional large upper third molar has been described here that belonged to an individual that lived much earlier. No other morphology can be associated with Denisovans. It is possible that they were otherwise not morphologically distinct from Neandertals or even more archaic hominins, and have been misclassified based on morphology, but this can only be confirmed with further DNA analysis.

Denisovans populated the Denisova Cave region on at least two occasions over a span of about 60,000 to 100,000 years (Table 12, as well as calculations in section 5.4). It is possible that they lived in the region continuously during this time, and that none of their remains have to date been found. Or they only occasionally populated the Altai Mountains and used them as a refugium during extreme climate changes in the lower steppe lands.

Denisovans show a higher diversity in their autosomal DNA than Neandertals (Figure 18). This higher diversity could be due to the great difference in age between the *Denisova 8* and the two other Denisovans. The Neandertals to which we compared the Denisovans range in age from 65,000 years for *Mezmaiskaya1* (62) to 38,000 for Vindija 33.16 (73), and show no significant difference in divergence to the *Altai 1* Neandertal (Table 16, Figure 18). However it should be noted the Neandertals cover a geographical range from Spain to Siberia, while the three Denisovans are all from the same cave.

Denisovans show an even higher degree of diversity in their mtDNA genomes compared to Neandertals (Figure 15, Table 10). The number of differences between *Denisova 8* and *Denisova 3* is 73% of the largest number of differences measured between 311 present-day humans (Figure 15). Again this could be due to the large age difference among the Denisovans, however both the Neandertals and humans come from large geographic ranges.

Denisova 8 is much older than *Denisova 3*, and shows a higher divergence both in autosomal and mtDNA data to *Denisova 3* compared to divergences among Neandertals. The Denisovan population that introgressed with the ancestors of present-day Oceanians diverged from the *Denisova 3* population 100 to 400 kya, depending on which mutation rate is used (10). Since *Denisova 8* could be up to 100ky older than *Denisova 3*, it is possible that *Denisova 8* belonged to a population that is more closely related to the population of Denisovans that introgressed with Oceanians. Unfortunately there are not enough data from *Denisova 8* to answer this question. There are also not enough data to ascertain whether *Denisova 8* had an admixture signal from the *Altai 1* Neandertal, as is shown in *Denisova 3*. Since this admixture signal comes from a recent event, post-dating almost all genetic drift in the population to which *Denisova 3* belonged (10), it is possible that such an admixture signal would not be seen in *Denisova 8*. *Denisova 8* may also have more admixture from more archaic hominins, but again the lack of data prevents an answer to these questions.

Altai 2 is a second Neandertal from Denisova Cave. This new Neandertal comes from the deepest layer of all of the remains in the East Gallery, however the disturbance of stratigraphy of the East Gallery could have occurred in layer 12 as well. Based on the mtDNA, there is no branch shortening to be seen in *Altai 2*. In fact *Altai 2* has the longest branch from the MRCA of Neandertals when compared to seven other Neandertals by five differences (section 6.3.2 and Table 10). Therefore it is unlikely that *Altai 2* is significantly older than *Altai 1*; if anything it is more likely younger.

The region around Denisova Cave also has other caves in which Neandertal remains have been found. One of these is Okladnikov cave (44, 103), from which an almost complete mtDNA exists from an individual which has been dated multiple times and produced dates from 29,990 $\pm$ 500 to 37,800 $\pm$ 450 years before present (uncalibrated C^{14} dates) (44). Another cave, Chagyrskaya Cave (15, 142), also has Neandertal bones which are beyond the C^{14} dating boundary (so greater than 50,000 years old) (143), however no published DNA exists from Neandertals from this site.

We now have mtDNA sequences from four Neandertals outside of Western Europe: *Mezmaiskaya1* (dated ~65ky old (62)), Okladnikov (dated ~30-38ky old (44)) and two undated Neandertals from Denisova Cave. All four of these Neandertals fall basal to the Neandertals from Western Europe (Figure 21). Based on this data, it seems that Eastern Neandertals fall

outside the variation of European Neandertals, as has been previously shown (103). Previous work with the hypervariable region of the mtDNA has shown that some older European Neandertals also fall outside the variation of the European Neandertals used in this study (144), however the use of the hypervariable region for Okladnikov has been shown to result in unreliable placement in the tree (103). While it is tempting to start speculating about Neandertal population movements based on these results, it would take a large scale autosomal DNA study of these Neandertals to more accurately look at this question.

7.2 The deposition of hominin remains in Denisova Cave

Many of the hominin bone pieces found in the Altai Mountains are small, morphologically indistinct pieces (3). No complete or even semi-complete skeletons have been excavated to date (3). One reason for only small pieces of remains being left is that the population sizes of hominins in the region could have been small. If only small groups of hominins existed, they would have left few dead behind. If we follow the hypothetical calculations done by Tuner *et al* in 2013, then 100 bands with 20 people each would leave only 300 dead to be discovered over a 30,000 year time frame (3). Both Denisovans and Neandertals had very small population sizes (8, 10). In addition, the *Altai 1* Neandertal was the product of extreme and recent inbreeding (10). Based on coding regions, the *Altai 1* has 79% and 89% of the heterozygosity of a Neandertal from El Sidron, Spain or Vindija Cave, Croatia, respectively (145). Therefore it is possible that both the Denisovan and Neandertal populations in the Altai Mountains were small. Turner *et al* 2013 (3) and Wrinn 2010 (63) both argue that the number of lithics found in the Altai mountains are small. Based on the calculations from Wrinn, layer 12 of the main chamber of Denisova Cave has the most lithics compared to surrounding sites such as Okladnikov and Anui 3, with about 600 artifacts per m^3 over 10,000 years, compared to the other sites that rarely reach 200 artifacts per m^3 over 10,000 years (63). The authors argue that these results suggest a sporadic occupation of the Altai Mountains. However no comprehensive comparison of the Altai region with other regions has been done to show that the lithic assemblages are indeed smaller than in other regions. A study from Mellers and French in 2011 compared archeological evidence from Neandertal and anatomically modern human sites in the Aquitaine region of southwestern France (146). They calculated stone tool densities between 6.6 to 17.6 tools per m^2 per 1000 years (146), however it is not clear if their methods for calculations are similar to the

calculations done by Wrinn 2010, and it would thus not be prudent to draw conclusions based on these data. In addition it has been cautioned that using stone tool densities to try to calculate population size is extremely complex (147).

The spotted hyena of the Pleistocene was a formidable opponent and obstacle to hominins. They were larger than spotted hyenas today (12), who are not fearful of humans (3). Reports have shown that the number of attacks on humans by hyenas are higher than by other terrestrial carnivores (3). They hunt in packs and are easily able to take down large prey (17). Since spotted hyenas dig dens for their young, they, along with other cave-dwelling animals such as the cave bear, would have been harsh competitors for hominins for cave-use. It is hard to imagine small hominin groups expelling an established hyena den from a large cave with ample digging room such as Denisova Cave.

The only hominin remains from the Pleistocene in Denisova Cave are phalanxes and teeth. Studies of bones left behind at dens of African spotted hyena today show that hyenas often leave behind metapodials and teeth, as well as occasional phalanxes (12, 16). Although hyena dens can show large assemblages of other bones, most of the unmodified bones are metopodials, possibly an indication of being swallowed whole and then regurgitated (16). We see hyena contamination in four of the five hominin remains in Denisova Cave that have been published so far. Only *Denisova 8* shows no detectable amount of hyena contamination. *Denisova 8* is much older than the other Denisovans and Neandertals, so it is possible that this individual was not eaten by hyenas, or that the location in the cave was somehow not tainted by hyena refuse. However, lack of detection of contamination does not mean that this tooth was not also disturbed by hyenas.

There are various possible reasons for the hyena contamination on the hominin remains in Denisova Cave. First it is possible that the remains were deposited during occupation of the cave by Denisovans or Neandertals. It is unclear whether Neandertals buried their dead (148, 149), but if remains were left in an occupied cave, it is likely that they were put under the ground in some manner to avoid the smell. The cave could have also been used as a site of deposition of dead bodies, while the hominins lived outside the cave. After the hominins left the cave (or were chased out), hyenas could have moved in. It has been speculated, based on hyena and animal bones left at Okladnikov cave, that hominins only used the cave in spring and fall, and lived in the forests during the winter, where they were closer to firewood, while the hyenas then moved into the caves in the winter as a protection against the cold (3).

It is unclear how much later the remains at Denisova Cave could have been disturbed, whether it was a season later or thousands of years later. Whether hyenas eat old bones is not mentioned in the literature. However it has been observed that hyenas very occasionally chew on dried out skeletons (personal communication with Dr. Gus Mills, Hyaena Specialist Group) and are not averse to chewing on crunchy material such as wood or plastic car tail lights (17). Hyenas can easily go many days without food and large numbers of bleached bones have been seen in areas where hyena density is low and food density is high, meaning that they are not eaten by the hyenas (17). In addition, spotted hyenas in Africa today will take off with a chunk of a kill, store it in water for a few days to preserve (17), and return for it later on. Thus they most likely prefer well preserved meat that has not begun to degrade in the hot African sun. But it is possible that a starving or bored hyena may chew on a bone that is thousands of years old.

Thus hyenas may have come across recent carcasses and dug them up, or even have stolen carcasses from where they were placed by the hominins. This phenomenon is observed countless times in the wild today, as hyenas, like other carnivores, often scavenge for food and will steal food away from other animals as well as each other (17). As the hyena DNA did not wash off more than the endogenous Neandertal DNA during the phosphate washing of the *Altai 2* bone (Table 28), it is possible that the hyena and Neandertal DNA were present at the same time, while the cells were lysing, and thus both populations of DNA had the same chances to bind to the hydroxyapatite in the bone. However there is an indication that more than one individual contaminated the *Altai 2* bone, making this scenario less likely.

It is entirely possible that the hyenas never directly interacted with the remains in Denisova Cave. Hyenas probably used the cave as a den multiple times over the span of thousands of years, causing disturbance of the layers in the East Gallery in the process, and may have defecated there. Spotted hyenas use latrines, specific areas often on the outer perimeter of their ranges, where they will defecate (17). The young, however, must defecate in the dens, especially when there is danger and they cannot leave the cave. Either Denisova Cave was used as a latrine or as a den or both, many times over the thousands of years after the remains were deposited. The feces then decomposed and released the DNA into the soil (18). The combination of the wet environment and hyena waste could have caused the hyena DNA to leech into the Neandertal and Denisovan remains. If such a situation occurred, it is unclear why the hyena DNA did not preferentially release off of the bone during the phosphate washing as did the microbial DNA.

Since remains from both the East and South Gallery show hyena contamination, the hyenas must have used both galleries for such activities at some point in time.

The last possibility is that the large Pleistocene hyenas hunted the hominins, as they have been known to do in Africa today (3, 17). There is evidence that hyenas ate hominins in Siberia from signs of acid erosion on bones (3), as would be expected from such a large predator. Hyenas are nocturnal (65) and pose a great threat to humans that sleep out in the open in Africa today. Hominins in the Pleistocene were not agriculturalists with established protected buildings as most Africans are today, and unless they were sleeping in a cave, they would have been exposed to hyena attacks at night. Therefore if hyenas were using Denisova Cave as a den or were otherwise occupying it, making it inaccessible to the hominins as a place of shelter, the hominins may have lived out in the open. While hyena territories today span between 40 and 1000 km^2 (66), they do not necessarily drag their food back to their den for their young, since spotted hyenas exclusively nurse their young for many months (17). The killing or scavenging of the hominin would therefore have happened within a few km of the cave, and during the feeding frenzy a hyena could have taken an arm foot or head with them to eat it in the shelter of the cave in peace, as is common behavior for spotted hyenas today (17). Therefore even if the Denisovan and Neandertal remains that we find today did not belong to individuals that occupied the cave, they must have been within a few km of the cave.

Due to the phosphate washing results, it is likely that the hyena and Neandertal DNA in the *Altai 2* bone bonded to the hydroxyapatite at the same time, so for the *Altai 2* at least the Neandertal and hyena may have had direct contact, either before or right after death. This could also explain why the amount of hyena DNA is so much higher in the *Altai 2* bone than in the others. However since the *Altai 2* hyena mtDNA sequences show a potential mixture of contributers, it is very possible that more than one of these scenarios happened in combination, since Denisova Cave was used heavily by hyenas as is evidenced by the large amounts of hyena bones and even coprolites in the cave (2, 3).

7.3 The Pleistocene spotted hyenas of Denisova Cave

Based on the mtDNA cytochrome b data, there are three clades of Pleistocene hyenas (Figure 27). Of the two that are no longer found in Africa today, one occurred exclusively in Europe

while the other occurred in eastern Asia and Denisova Cave. The third Pleistocene hyena clade occurs in African spotted hyenas today as well as in Pleistocene hyenas in Europe and Denisova Cave. Since there are no hyena sequences from hyenas that lived between western Ukraine and Denisova Cave, or between Denisova Cave and Vladivostok, it is unclear how much these two clades overlap. However it is clear that they overlap at Denisova Cave, making Denisova Cave a possible meeting point of two Pleistocene hyena populations. Whether these populations existed simultaneously is unclear. Of the 24 Pleistocene hyena sequences, 11 come from individuals that have been dated with C^{14} dating. The far-east hyenas have dates between 35kya (Chinese) (14) and 48kya (eastern Russian) (13). This age range overlaps with the one dated hyena from Denisova Cave, 42 kya (141), which is one of the Denisova Cave hyenas that fall into the clade found also in Europe and in extant African spotted hyenas. The age range also overlaps with the ages of the European hyenas, which range from 38kya to 51kya (13, 141). The exclusively European clade only has two dated individuals, and both are the oldest dated hyenas in Europe at 51kya and older than 48kya, so it is possible that this clade was made up of older hyenas. However the difference between eastern and western Eurasian hyenas does not seem to be due to age.

Previous work on the cytochrome b gene of the mtDNA of Pleistocene hyenas has been used to hypothesize about the origin of the spotted hyena. Rohland *et al* concluded that spotted hyenas populated Eurasia from an African origin in three waves. The first wave left southern Africa, the site of the some of the earliest spotted hyena fossils in Africa (3.46 million years ago (150, 151)), crossed northern Africa, then exited Africa and went across Asia to East Asia. They must have traveled through Pakistan, where the oldest Asian spotted hyena fossils exist (2.6-3.7 million years ago (152)). They could have also traveled over the Altai Mountains, where a group could have stayed. Since this migration happened up to 3.5 million years ago, it is likely though that the hyena sequence consensus found in *Altai 2*, which carries only one difference to the eastern hyena sequences, was either part of a large and often mixing eastern population, encompassing the far-east and Denisova Cave, or represents a back migration later in time. The second wave left Africa and came to Europe, where the earliest hyena fossils are dated to 0.8 million years ago (153). This group then died out in Africa, leaving only a European clade. Around this time, the hyena populations were separating roughly into northern and southern African populations. The

last wave left northern Africa and populated both Europe and Siberia (13). Our additional results do not refute this hypothesis.

With the addition of three more eastern Pleistocene hyena cytochrome b sequences (14), a new hypothesis arose, where all four clades were panmictic in Eurasia. First the eastern clade separated 400 to 230 kya, then the exclusively African clade as well as the African and European clades entered Africa between 145 and 50 kya, and subsequently they separated into northern and southern populations. The eastern hyena contamination of *Altai 2* does not negate this hypothesis, since the eastern clade of hyenas could have migrated back to Denisova Cave, however the 3.46 million year old spotted hyena fossils from Africa (150, 151) complicate the picture.

Both hypotheses state that the East Asian spotted hyena clade was the earliest to split off. We now show that this clade also existed much farther west than previously thought; thus it is possible that this eastern population had a much larger geographical distribution than previously thought, encompassing both eastern Russia and central Siberia.

Denisova Cave was not only a possible meeting point for Pleistocene hyena populations, but also for hominin groups. Denisova Cave represents the farthest eastern location with Neandertal remains found to date, and is possibly on the far East of the Neandertal range in Europe. No Denisovans have been found outside of Denisova Cave. An admixture signal from Denisovans is seen in Oceania and to a lesser degree in East Asia, so it is tempting to hypothesize that Denisovans lived in the East. However the split between the introgressing Denisovan population and *Denisova 3* is quite deep, as they diverged at least 140kya (10). Therefore it is possible that the introgressing Denisovans were from a very diverged or archaic population, and the population to which the three Denisovans from Denisova Cave belonged never came into contact with the ancestors of present-day Oceanians. Thus it is unclear how far East or West Denisovans lived, although it is less likely they were widespread in Europe.

Of interest is that the Pleistocene hyenas that contaminated the Denisovans came from the western hyena clade, while the hyena that contaminated the *Altai 2* and possibly *Altai 1* Neandertals came from an eastern population.

These hyena data give a glimpse into possible populations of Pleistocene hyenas. However it is important to note that the results here are mostly based on a small part of the mtDNA. Even the complete mtDNA data are only giving information on this one maternally-inherited locus. As

shown in Denisovans, the mtDNA alone can show a very different picture from autosomal data (9). Therefore it would be of great interest to sequence the autosomal genomes of these Pleistocene hyenas, as well as more populations across Russia and the Middle East. Before such sequencing is done, the question of spotted hyena origin and movement cannot be fully answered.

8. Appendix

8.1 Index of Figures

Figure: Page:

8.2 Index of Tables

Table: Page:

8.4 Appendix of Tables

Appendix Table 1. List of spotted hyena sequences used in section 6.3.5. Clade color refers to the color of the clade the individual falls into, see Figure 27. Age is for samples greater than 1000 years old, collection date is given for samples collected in the 19th and 20th century. Samples from live individuals were taken in the last 20 years.

Accession number	Museum number/Name	Location	Clade color	Age/collection date	Reference
AJ809318	Teufel 2	Teufelslucke (Austria)	blue	38,060	(141)
AJ809320	Winden	Winden cave (Austria)	blue	38,680	(141)
AJ809319	Irpfel 4	Irpfel cave (Germany)	blue	-	(141)
AJ809324	Igric (V10529)	Igric (Romania)	blue	41,800	(141)
AJ809331	Kiske M1 (V14484)	Kiskevelyi (Hungary)	yellow	>48,500	(141)
AJ809321	Vypustek p	Vypustek (Czech Rep.)	blue	46,000	(141)
AJ809330	Linde 1	Lindenthal cave (Germany)	yellow	-	(141)
AJ809327	Altai D19	Denisova Cave (Russia) Asia	blue	42,300 +940/-840	(141)
AJ809326	Bukovinka	Bukovinka cave (Ukraine)	blue	41,300	(141)
AJ809328	Certova 1	Certova pec (Slovakia)	yellow	51,200	(141)
AJ809329	Sveduv 2	Sveduv stul (Czech Rep.)	yellow	-	(141)
AJ809332	Tmava (TS 250)	Tmava skala (Slovakia)	yellow	-	(141)
AJ809322	PLU 681 E3 II	Les Plumettes (France)	blue	-	(141)
AJ809325	RDV 01H10 23	Les Roches de Villeneuve (France)	blue	40,700	(141)
AJ809323	Niederlande 1	North Sea (The Netherlands)	blue	-	(141)
DQ157554	2812 37	Goyet cave (Belgium)	blue	-	(13)
DQ157555	895 (34490 (1))	Geographical Society cave (Vladivostok, Russia) Asia	green	48,650 +2380/-1840	(13)
NC020670	CC8	Coumere Cave, France	blue	-	(97)

JF894379	CC9	Coumere Cave, France	blue	-	(97)
KC117379	DARD1	Da'an cave, Tonghua county, Jilin province, China	green	35,520 +/- 230	(14)
KC117380	DARD2	Da'an cave, Tonghua county, Jilin province, China	green	-	(14)
KC117381	DARD3	Da'an cave, Tonghua county, Jilin province, China	green	-	(14)
DQ157556	41225	Dagana (Senegal)	blue	1925	(13)
DQ157557	46123	Dire Dawa (Ethiopia)	blue	1928	(13)
DQ157558	82399	(N-Cameroon)	blue	1913	(13)
DQ157559	82400	Kete Kratshi (Togo)	blue	1899	(13)
DQ157563	82404	Ikoma (Tanzania)	blue	1906	(13)
DQ157568	82412	Windhuk (Namibia)	pink	1898	(13)
DQ157562	82423	Lake Kivu (Rwanda)	blue	1902	(13)
DQ157567	82432	Ikoma (Tanzania)	blue	1913	(13)
DQ157564	82455	Sefane (Eritrea)	blue	1913	(13)
DQ157573	82467	Singa (Sudan)	pink	1912	(13)
DQ157566	82472	Malindi (Zimbabwe)	pink	1911	(13)
DQ157592	82477	(NE-Rwanda)	blue	1907	(13)
DQ157560	82482	Masai steppe (Tanzania)	pink	1905	(13)
DQ157561	82533	Otawi (Namibia)	pink	1902	(13)
DQ157570	82537	Loanda (Angola)	pink	1901	(13)
DQ157569	1244	Welkom (South Africa)	pink	1880s	(13)
DQ157571	1755	Somaliland (Somalia)	blue	1898	(13)
DQ157572	3919	Stony Athi (Kenia)	pink	1908	(13)

DQ157565	6061	upper course of Dinder (Sudan)	blue	1925	(13)
DQ157575	7397	S shore of lake Tschad (Cameroon)	blue	1928	(13)
DQ157574	21495	Mutir (Uganda)	blue	1882	(13)
JF894377	Kira	French zoo	blue	From live individual	(97)
AF153053	Ibo	Berlin zoo	blue	From live individual	(112)
AY048811	Nigeria-TPIbo	Berlin zoo	blue	From live individual	(111)
AF153052	Kr.L.097	Ngorongor crater, wild	pink	From live individual	(112)
AF153051	M119	serengeti, wild	pink	From live individual	(112)
AY048812	Suedafrika-Muen	Munich zoo, poss South African origin	pink	From live individual	(111)
AY048807	KrX012	northwest tanzania, wild (ngorongoro?) (GPS: 03 10 80/35 34 76)	pink	From live individual	(111)
AY048808	Lemuta-Z056	Serengeti, no GPS	pink	From live individual	(111)
AY048809	Seronera-Z098	Serengeti, no GPS	pink	From live individual	(111)
AY048805	Central-M119	Serengeti, wild (GPS: 02 27 30 / 34 46 00)	pink	From live individual	(111)
AY048806	Ngorongoro-KrM17	Ngorongoro, wild (GPS: 03 09 64 / 35 36 63)	pink	From live individual	(111)
AY048804	Camsite-C004	Serengeti, wild (GPS: 02 24 82 / 34 50 48)	pink	From live individual	(111)
AY048810	Ruaha-Z095	Ruaha park, no GPS	pink	From live individual	(111)

9. References

1. Derevianko A (2011) The upper paleolithic in Africa and Eurasia and the origin of anatomically modern humans. *Novosibirsk, Institute of Archeology and Ethnography SB RAS Press, MV Shunkov (Ed)*.
2. Derevianko A*, et al.* (2003) Paleoenvironmnet and paleolithic human occupation of Gorny Altai. *Novosibirk, Institute of Archeology and Ethnography, SB RAS Press, AP derevianko and MV Shunkov (Eds)*.
3. Turner C, Ovodov N, & Pavlova O (2013) Animal teeth and Human tools: A taphonomic Odyssey in ice age Siberia. *Cambridge University Press*.
4. Turner C (1990) Upper Paleolithic Mal'ta child (Siberia) *News of siberian Branch, USSR Academy of Sciences* 2:70-71.
5. Viola B (2009) New hominid remains from central Asia and Siberia: the eastern-most Neanderthals? *PhD Dissertation: Vienna University*.
6. Mednikova MB (2011) A proximal pedal phalanx of a paleolithic hominin from Denisova cave, Altai. *Archaeology Ethnology & Anthropology of Eurasia* 39(1):129-138.
7. Krause J*, et al.* (2010) The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* 464(7290):894-897.
8. Meyer M*, et al.* (2012) A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338(6104):222-226.
9. Reich D*, et al.* (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468(7327):1053-1060.
10. Prufer K*, et al.* (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505(7481):43-49.
11. Green RE*, et al.* (2010) A draft sequence of the Neandertal genome. *Science* 328(5979):710-722.
12. Kurten B (1968) Pleistocene mammals of Europe. *Weidenfeld and Nicholson (eds), London*.
13. Rohland N*, et al.* (2005) The population history of extant and extinct hyenas. *Mol Biol Evol* 22(12):2435-2443.
14. Sheng GL*, et al.* (2014) Pleistocene Chinese cave hyenas and the recent Eurasian history of the spotted hyena, Crocuta crocuta. *Mol Ecol* 23(3):522-533.
15. Korlević P, Gerber, T., Gansauge, M.T., Hajdinjak, M., Nagel, S., Aximu-Petri, A., Meyer, M (2015) Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* (in press).
16. Hill A (1989) Bone modification by modern spotted hyenas. *In: Bonnichsen R, Sorg MH (Eds), Bone modification, Peopling of the Americas, Center for the study of the first Americans, Orono*.
17. Kruuk H (1972) The spotted hyena, A study of predation and social behavior. *In: Schaller GB (Ed), Wildlife behavior and ecology, University of Chicago press*.
18. Horwitz L & Goldberg P (1989) A study of pleistocene and holocene hyaena coprolites. *Journal of Archaeological Science* 16:71-94.
19. Saint-Hilaire IG (1859) Histoire Naturelle Generale des Regres Organiques, Tome II. *Librarie de Voctor Masson, Paris*.
20. Huxley TH (1863) Evidence as to Man's place in nature *Williams and Norgate, London*.
21. Darwin C (1871) The descent of man. *John Murray, London*.
22. King W (1864) The reputed fossil man of the Neandertal. *Quarterly Journal of Science* 1:88-97.
23. Chimpanzee S & Analysis C (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437(7055):69-87.

24. Prufer K*, et al.* (2012) The bonobo genome compared with the chimpanzee and human genomes. *Nature* 486(7404):527-531.
25. Goodman M (1999) The genomic record of Humankind's evolutionary roots. *Am J Hum Genet* 64(1):31-39.
26. Wood B (2010) Colloquium paper: reconstructing human evolution: achievements, challenges, and opportunities. *Proc Natl Acad Sci U S A* 107 Suppl 2:8902-8909.
27. Gabunia L*, et al.* (2000) Earliest Pleistocene hominid cranial remains from Dmanisi, Republic of Georgia: taxonomy, geological setting, and age. *Science* 288(5468):1019-1025.
28. Wolpoff MH*, et al.* (1988) Modern human origins. *Science* 241(4867):772-774.
29. Cavalli-Sforza LL (1966) Population structure and human evolution. *Proc R Soc Lond B Biol Sci* 164(995):362-379.
30. Cann RL, Stoneking M, & Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* 325(6099):31-36.
31. Tishkoff SA & Verrelli BC (2003) Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu Rev Genomics Hum Genet* 4:293-340.
32. Schlebusch CM*, et al.* (2012) Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *Science* 338(6105):374-379.
33. Li H & Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493-496.
34. Stringer C (2002) Modern human origins: progress and prospects. *Philos Trans R Soc Lond B Biol Sci* 357(1420):563-579.
35. Atkinson QD (2011) Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 332(6027):346-349.
36. McBrearty S & Brooks AS (2000) The revolution that wasn't: a new interpretation of the origin of modern human behavior. *J Hum Evol* 39(5):453-563.
37. Hublin JJ (2009) Out of Africa: modern human origins special feature: the origin of Neandertals. *Proc Natl Acad Sci U S A* 106(38):16022-16027.
38. Stringer CB & Hublin J (1999) New age estimates for the Swanscombe hominid, and their significance for human evolution. *J Hum Evol* 37(6):873-877.
39. Schwarcz HP, Simpson JJ, & Stringer CB (1998) Neanderthal skeleton from Tabun: U-series data by gamma-ray spectrometry. *J Hum Evol* 35(6):635-645.
40. F B & E T (2006) Middle Paleolithic Human remains from Bisitun cave, Iran. *Paleorient*:105-111.
41. Ovchinnikov IV*, et al.* (2000) Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* 404(6777):490-493.
42. Glantz M, Viola, B., Wrinn, P., Chikisheva, T., Derevianko, A., Krivoshapkin, A., Islamov, U., Suleimanov, R. & Ritzman, T. (2008) New hominin remains from Uzbekistan. *Journal of Human Evolution* 55(2):223-237.
43. Gunz P & Bulygina E (2012) The Mousterian child from Teshik-Tash is a Neanderthal: a geometric morphometric study of the frontal bone. *Am J Phys Anthropol* 149(3):365-379.
44. Krause J*, et al.* (2007) Neanderthals in central Asia and Siberia. *Nature* 449(7164):902-904.
45. Finlayson C*, et al.* (2006) Late survival of Neanderthals at the southernmost extreme of Europe. *Nature* 443(7113):850-853.
46. Grun R*, et al.* (2006) ESR and U-series analyses of enamel and dentine fragments of the Banyoles mandible. *J Hum Evol* 50(3):347-358.
47. Stringer CB, Grun R, Schwarcz HP, & Goldberg P (1989) ESR dates for the hominid burial site of Es Skhul in Israel. *Nature* 338(6218):756-758.
48. Higuchi R, Bowman B, Freiberger M, Ryder OA, & Wilson AC (1984) DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312(5991):282-284.

49. Paabo S (1985) Molecular cloning of Ancient Egyptian mummy DNA. *Nature* 314(6012):644-645.
50. Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* 362(6422):709-715.
51. Green RE*, et al.* (2006) Analysis of one million base pairs of Neanderthal DNA. *Nature* 444(7117):330-336.
52. Paabo S*, et al.* (2004) Genetic analyses from ancient DNA. *Annu Rev Genet* 38:645-679.
53. Briggs AW*, et al.* (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A* 104(37):14616-14621.
54. Brotherton P*, et al.* (2007) Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res* 35(17):5717-5728.
55. Sawyer S, Krause J, Guschanski K, Savolainen V, & Paabo S (2012) Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* 7(3):e34131.
56. Krause J*, et al.* (2010) A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr Biol* 20(3):231-236.
57. Skoglund P*, et al.* (2012) Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* 336(6080):466-469.
58. Krings M*, et al.* (1997) Neandertal DNA sequences and the origin of modern humans. *Cell* 90(1):19-30.
59. Green RE*, et al.* (2008) A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* 134(3):416-426.
60. Reich D*, et al.* (2011) Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am J Hum Genet* 89(4):516-528.
61. Gansauge MT & Meyer M (2013) Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature protocols* 8(4):737-748.
62. Skinner AR*, et al.* (2003) ESR dating at Mezmaiskaya Cave, Russia. *American Journal of Physical Anthropology*:193-193.
63. Wrinn P (2010) Middle Paleolithic settlement and land use in the Altai Mountains, Siberia. *Settlement dynamics of the Middle Paleolithic and Middle Stone Age vol. III, Tubingen kerns Publishing, N Conrad and A Delagnes (Eds)*.
64. Werdelin L & N S (1991) The hyaenidae: taxonomy, systematics and evolution. *Fossils and Strata* 30:1-104.
65. Gittleman J & Harvey P (1982) Carnivore home-range size, metabolic needs and ecology. *Behav Ecol Sociobiol* 10:57-63.
66. Mill M (1989) The comparative behavioral ecology of hyenas: the importance of diet and food dispersion *Carnivore behavior, ecology and evolution, Gittleman J (Ed), Cornell University*.
67. Rohland N & Hofreiter M (2007) Comparison and optimization of ancient DNA extraction. *Biotechniques* 42(3):343-352.
68. Briggs AW*, et al.* (2010) Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res* 38(6):e87.
69. Dabney J & Meyer M (2012) Length and GC-biases during sequencing library amplification: A comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* 52(2):87-94.
70. Maricic T, Whitten M, & Paabo S (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* 5(11):e14004.
71. Kircher M, Stenzel U, & Kelso J (2009) Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol* 10(8):R83.

72. Kircher M (2012) Analysis of high-throughput ancient DNA sequencing data. *Methods in molecular biology* 840:197-228.
73. Briggs AW*, et al.* (2009) Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* 325(5938):318-321.
74. Genomes Project Consortium*, et al.* (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491(7422):56-65.
75. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754-1760.
76. Andrews RM*, et al.* (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23(2):147.
77. R Core Team (2014) R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*.
78. Ingman M, Kaessmann H, Paabo S, & Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408(6813):708-713.
79. Arnason U, Xu X, & Gullberg A (1996) Comparison between the complete mitochondrial DNA sequences of Homo and the common chimpanzee based on nonchimeric sequences. *J Mol Evol* 42(2):145-152.
80. Katoh K, Misawa K, Kuma K, & Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30(14):3059-3066.
81. Katoh K & Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* 30(4):772-780.
82. Koichiro Tamura GS, Daniel Peterson, Alan Filipski, and Sudhir Kumar (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. . *Molecular Biology and Evolution* 30:2725-2729.
83. Posada D & Crandall K (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* 14(9):817-818.
84. Huelsenbeck JP, Ronquist F, Nielsen R, & Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294(5550):2310-2314.
85. Ronquist F & Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12):1572-1574.
86. Drummond AJ, Suchard MA, Xie D, & Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29(8):1969-1973.
87. Meyer M*, et al.* (2014) A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* 505(7483):403-406.
88. Fu Q*, et al.* (2013) A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr Biol* 23(7):553-559.
89. Kass R & Raftery A (1995) Bayes Factors. *Journal of the American Statistical Association* 90(430):773-795.
90. Librado P & Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451-1452.
91. Paten B, Herrero J, Beal K, Fitzgerald S, & Birney E (2008) Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Res* 18(11):1814-1828.
92. Paten B*, et al.* (2008) Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res* 18(11):1829-1843.
93. Patterson N*, et al.* (2012) Ancient admixture in human history. *Genetics* 192(3):1065-1093.
94. Dabney J*, et al.* (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):15758-15763.

95. Gansauge MT & Meyer M (2014) Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res* 24(9):1543-1549.
96. Kircher M, Sawyer S, & Meyer M (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* 40(1):e3.
97. Bon C*, et al.* (2012) Coprolites as a source of information on the genome and diet of the cave hyena. *Proc Biol Sci* 279(1739):2825-2830.
98. Slon V*, et al.* (2015) Mammalian mitochondrial capture, a tool for rapid screening of DNA preservation in faunal and undiagnostic remains, and its application to Middle Pleistocene specimens from Qesem Cave (Israel). *Quaternary International* In press.
99. Fu Q*, et al.* (2013) DNA analysis of an early modern human from Tianyuan Cave, China. *Proc Natl Acad Sci U S A* 110(6):2223-2227.
100. Renaud G, Kircher M, Stenzel U, & Kelso J (2013) freeIbis: an efficient basecaller with calibrated quality scores for Illumina sequencers. *Bioinformatics* 29(9):1208-1209.
101. Renaud G, Stenzel U, & Kelso J (2014) leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res* 42(18):e141.
102. Renaud G, Stenzel U, Maricic T, Wiebe V, & Kelso J (2015) deML: robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics* 31(5):770-772.
103. Skoglund P*, et al.* (2014) Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc Natl Acad Sci U S A* 111(6):2229-2234.
104. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. . *Nucl. Acids. Symp.* 41:95-98.
105. Camacho C*, et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
106. Fu Q*, et al.* (2014) Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 514(7523):445-449.
107. Arnason U*, et al.* (2006) Pinniped phylogeny and a new hypothesis for their origin and dispersal. *Mol. Phylogenet. Evol.* 41(2):345-354.
108. Krause J*, et al.* (2008) Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evol Biol* 8:220.
109. Drummond AJ & Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
110. Koepfli KP*, et al.* (2006) Molecular systematics of the Hyaenidae: relationships of a relictual lineage resolved by a molecular supermatrix. *Mol. Phylogenet. Evol.* 38(3):603-620.
111. Albert R (2002) Genstruktur und Genfluss in ausgewaehlten Populationen der Tuepfelhyaene (Crocuta crocuta). *PhD Dissertation, Freie Universitaet, Berlin*.
112. Albert R, Hofer H, East M, & Pitra C (2000) Genetic identification of the geographic origin of spotted hyaenas (Crocuta crocuta). *Zool. Gart.* 70:1-10.
113. Meyer M*, et al.* (2016) Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Submitted to Nature*.
114. Scally A & Durbin R (2012) Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet* 13(10):745-753.
115. Bailey S, Glantz, M., Weaver, T.D. & Viola, B (2008) The affinity of the dental remains from Obi-Rakhmat Grotto, Uzbekistan. *Journal of Human Evolution* 55(2):238-248.
116. Martinón-Torres M, Bermúdez de Castro, J.M., Gómez-Robles, A., Prado-Simón, L. & Arsuaga, J.L. (2012) Morphological description and comparison of the dental remains from Atapuerca-Sima de los Huesos site (Spain). *Journal of Human Evolution* 62(1):7-58.
117. Kaifu Y (2006) Advanced dental reduction in Javanese Homo erectus. *Anthropol Sci* 114:35-43.

118. Xing S, Martinon-Torres, M., Bermudez de Castro, J.M., Wu, X., Liu, W (2015) Hominin Teeth From the Early Late Pleistocene Site of Xujiayao, Northern China. *Journal of Physical Anthropology* 156(2):224-240.
119. Frayer D (1976) Evolutionary Dental Changes in Upper Palaeolithic and Mesolithic Human Populations *PhD Dissertation, University of Michigan, Ann Arbor*.
120. Wood BA (1991) Hominid cranial remains. Koobi Fora Research Project, Vol. 4. *Clarendon Press, Oxford*.
121. Tobias PV (1991) Olduvai Gorge (The skulls, endocasts and teeth of Homo habilis, Vol. 4). *Cambridge: Cambridge University Press*.
122. Martinón-Torres M, Bermúdez de Castro, J.M., Gómez-Robles, A., Margvelashvili, A., Prado, L., Lordkipanidze, D. & Vekua, A. (2008) Dental remains from Dmanisi (Republic of Georgia): Morphological analysis and comparative study. *Journal of Human Evolution* 55(2):249-273.
123. Brown B & Walker A (1993) The dentition. *The Nariokotome Homo erectus Skeleton*:161-192.
124. Suwa G, Asfaw, B., Haile-Selassie, Y., White, T., Katoh, S., WoldeGabriel, G., Hart, W.K., Nakaya, H. & Beyene, Y. (2007) Early Pleistocene Homo erectus fossils from Konso, southern Ethiopia. *Anthropological Science* 115(2):133-151.
125. Weidenreich F (1937) The dentition of Sinanthropus pekinensis: a comparative odontography of the hominids. Paleaontologia Sinica new Series D. *Chinese Academy of Sciences, Beijing*.
126. Wu X & Poirier FE (1995) Human Evolution in China. *Oxford University Press, Oxford*.
127. Grine FE & Franzen JL (1994) Fossil hominid teeth from the Sangiran dome (Java, Indonesia). *Courier Forsch Senckenberg* 171:75-103.
128. Condemi S (2001) Les Néanderthaliens de La Chaise (Abri Bourgeois-Delaunay). *Éditions du Comité des Travaux Historiques et Scientifiques, Paris*.
129. Rougier H (2003) Étude descriptive et comparative de Biache-Saint-Vaast 1 (Biache-Saint-Vaast, Pas-de-Calais, France). *PhD Thesis, University of Bordeaux I*.
130. de Lumley H, & de Lumley, M.A (1982) L'Homo erectus et sa place de l'homme de Tautavel parmi les hominidés fossiles. *Nice, Palais des expositions*.
131. Suzuki H & Takai F (1970) The Amud Man and his Cave Site. *Academic Press of Japan, Tokyo.*
132. Tillier A (1979) La dentition de l'enfant moustérien Châteauneuf 2, découvert á l'Abri de Hauteroche (Charente). *L'Anthropologie* 83(3):417-438.
133. Wolpoff MH (1979) The Krapina dental remains. *American Journal of Physical Anthropology* 50(1):67-113.
134. Trinkaus E (1983) The Shanidar Neanderthals. *Academic Press, New York*.
135. McCown T & Keith A (1939) The Stone Age of Mount Carmel II: The human remains from the Levalloiso-Mousterien. *Oxford: Clarendon Press*.
136. Vandermeersch B (1981) Les hommes fossiles de Qafzeh (Israel). *Paris: Editions du CNRS* 319.
137. Hublin J-J, Verna, C., Bailey, S., Smith, T., Olejniczak, A., Sbihi-Alaoui, F. Z., & Zouak, M. (2012) Dental evidence from the Aterian human populations of Morocco. In J.- J. Hublin & S. McPherron (Eds.). *Vertebrate Paleobiology and Paleaanthropology: Modern origins - A North African perspective. Dordrecht: Springer*:189-204.
138. Sladek V, Trinkaus, E., HIllson, S.W. & Holliday, T.W. (2000) The People of the Pavlovian: Skeletal Catalogue and Osteometrics of the Gravettian Fossil Hominids from Dolní Vestonice and Pavlov. *Inst. of Archaeology, Academy of Sciences of the Czech Republic, Brno.*
139. Trinkaus E (2010) Denisova Cave, Peştera cu Oase, and Human Divergence in the Late Pleistocene. *PaleoAnthropology* 2010:196-200.
140. Sankararaman S, Patterson N, Li H, Paabo S, & Reich D (2012) The date of interbreeding between Neandertals and modern humans. *PLoS Genet* 8(10):e1002947.

141. Hofreiter M*, et al.* (2004) Lack of phylogeography in European mammals before the last glaciation. *Proc Natl Acad Sci U S A* 101(35):12963-12968.
142. Mednikova MB (2013) AN ARCHAIC HUMAN ULNA FROM CHAGYRSKAYA CAVE, ALTAI: MORPHOLOGY AND TAXONOMY. *ARCHAEOLOGY, ETHNOLOGY & ANTHROPOLOGY OF EURASIA* 41(1):66-77.
143. Rudaya N, Vasilyev S, Viola B, Talamo S, & Markin S (in press) Paleoenvorinments during the period of the Neanderthals settlement in Chagyrskaya cave (Altai mountains, Russia). *Paleogeography, Paleoclimatology, Paleaoecology*.
144. Dalen L*, et al.* (2012) Partial genetic turnover in neandertals: continuity in the East and population replacement in the West. *Mol Biol Evol* 29(8):1893-1897.
145. Castellano S*, et al.* (2014) Patterns of coding variation in the complete exomes of three Neandertals. *Proc Natl Acad Sci U S A* 111(18):6666-6671.
146. Mellars P & French JC (2011) Tenfold population increase in Western Europe at the Neandertal-to-modern human transition. *Science* 333(6042):623-627.
147. Dogandzic T & McPherron SP (2013) Demography and the demise of Neandertals: a comment on 'Tenfold population increase in Western Europe at the Neandertal-to-modern human transition'. *J Hum Evol* 64(4):311-313.
148. Gargett R (1999) Middle Palaeolithic burial is not a dead issue: the view from Qafzeh, Saint-Césaire, Kebara, Amud, and Dederiyeh. *Journal of Human Evolution* 37:27-90.
149. Belfer-Cohen A & Hovers E (1992) In the Eye of the Beholder: Mousterian and Natufian Burials in the Levant. *Curr Anthropol* 33(4):463-471.
150. Barry J (1987) Larger Carnivores (Canidae, Hyaenidae, Feidae) from Laetoli. *In: MD Leakey and JM Harris (eds). Laetoli: A Pliocene site in Tanzania. Oxford Clarendon Press*:235-258.
151. Turner A (1990) The evolution of the guild fo larger terrestrial carnivores during the Pliopleistocene in Africa. *Geobios* 23:349-368.
152. de Vos J, Leinders J, & Hussain S (1987) A historical review of the Siwalik Hyaenidae (Mammalia, Carnivora) and description of two new finds from the Upper Siwalik of Pakistan. *Palaentology, Proc B* 90:333-369.
153. Garcia N & Arsuaga JL (1999) Carnivores from the Early Pleistocene hominid-bearing Trinchera Dolina 6 (Sierra de Atapuerca, Spain). *J Hum Evol* 37(3-4):415-430.

Acknowledgements

I would like to thank my husband, Spencer, who never really understood what I did, but nodded his head and pretended to listen very loyally. Without him I would not be at this point in my life. We have both had tremendously stressful lives the past few years, not helped by the addition of our daughter, but we stuck together through all the hard times. Nothing can make me forget the craziness of our lives more than spending a few hours with him and Emma.

I would like to thank my sister, Johanna, who has always been my crutch in life. Without her I would never have come this far. Regardless of what she is doing, she listened to me, helped me deal with hundreds of hurdles and life crises, and has always been at my side. Throughout my life she has always been My Person. I cannot thank her enough.

I would like to thank my mother and father, Sabine and Haywood, who gave me such an amazing childhood. They nurtured my interest for science despite the fact that neither of them are science-inclined. They encouraged my various science hobbies and fascinations, from gorillas to blowing up chemicals with my chemistry set. They stood by side and cheered me on, they prodded me when I needed it, but also knew when to stand aside and let me make own mistakes.

I would like to thank my mentor, Svante Pääbo. He gave the amazing opportunity to pursue my PhD in his group. He expected me to do well and was therefore tough on me at times. This has given me the opportunity to grow as a person and a scientist and has been an invaluable experience. I would like to especially thank him for being so incredibly understanding towards me in regards to having a child during my PhD.

I would like to thank Aida and Janet for being there in times of crises, and always having an open door and listening while I poured my heart out. I would also like to thank Janet, Kay and Matthias for helping me with this thesis. A big thanks to the Ancient DNA group, both past and present members, especially Qiaomei, without whom I would have been totally lost those first few years.

I would like to thank my two favorite office mates, Tomi and Felix. You two have had to listen to my inane babbling for five years now, and I am sure you are happy to finally be rid of me. Our impromptu discussions have always been the best!

And of course I would like to thank Shona and Ana. Our weekly Friday night drinks, our much needed coffee breaks, our various forays into the Leipzig nightlife, they were all so special.

Curriculum Vitae

Susanna Katherine Sawyer
susanna_sawyer@eva.mpg.de
Karl-Liebknecht-Str. 63, 04275 Leipzig; Tel: +4917680216758

Academic Career

June 2010 -present	**PhD**, Department of Genetics, Max Planck Institute for Evolutionary Anthropology, Germany Supervisor: Svante Pääbo Dissertation focus: Nuclear and Mitochondrial DNA relationships within Denisovans and Neandertals of the Altai Mountains Relevant skills: Ancient DNA lab work (extraction, library preparation using Illumina platform, targeted capture, sequencing), genome analysis (mtDNA, low-coverage nuclear genomes, contamination assesment), basic population genetics, basic programing skills (perl, awk, bash, R)
August 2008 – June 2010	**M.S. in Evolutionary Biology** at Uppsala Universitet, Uppsala, Sweden Masters Project: "*Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA*" done at the Max Planck Institute for Evolutionary Anthropology September 2009-June 2010
August 2004 – May 2008	**Bachelors degree** at North Carolina State University, Raleigh, NC, USA B.S. in Zoology; GPA: 3.953/4.33 (total), 4.00/4.33 (major)
January 2005 – June 2008	**Undergraduate research on behavior of zebrafish** with Dr. John Godwin, NCSU Relevant skills: use of Open Field tests, developed own behavioral tests, Microarrays, Cortisol extractions and measurements
March 2004 – June 2004	**Internship**: Max Delbrück Center for Molecular Medicine, Berlin, Germany Relevent skills: Immunohistology of cells, Maintenance of small lab animals, Measurement of tumor growth, Mouse xenotransplantation

Publications

Research Papers

Sawyer S, Viola B, Shunkov M, Derevianko A, Meyer M, Prüfer K, Pääbo S. A Neandertal from Denisova Cave with ancient spotted hyena contamination. *In preperation.*

Sawyer S*, Renaud G* , Viola B, Gansauge MT, Shunkov M, Derevianko A, Prüfer K, Kelso J, Pääbo S (2015) Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *PNAS.*
* These authors contributed equally

Benazzi S, Slon V, Talamo S, Negrino F, Peresani M, Bailey SE, **Sawyer S**, et al. (2015) The makers of the Protoaurignacian and implications for Neandertal extinction. *Science.*

Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, **Sawyer S** et al. (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature.*

Castellano S, Parra G, Sanchez-Quinto FA, Racimo F, Kuhlwilm M, Kircher M, **Sawyer S** et al. (2014) Patterns of coding variation in the complete exomes of three Neandertals. *Proc Natl Acad Sci U S A.*

Lazaridis I, [18 people] **Sawyer S** et al. (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature.*

S. T. D. Consortium (2014) Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature.*

Thalmann O, Shapiro B, Cui P, Schuenemann VJ, **Sawyer S**, et al. (2013)., Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science*

Guschanski K, Krause J, **Sawyer S**, Valente LM, Bailey S, et al. (2013) Next-generation museomics disentangles one of the largest primate radiations. *Systematic biology.*

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

Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S (2012) Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One.*

Conferences

Sawyer S, Renaud G , Viola B, Gansauge MT, Shunkov M, Derevianko A, Prüfer K, Kelso J, Pääbo S. An Analysis of Three Denisovan Individuals Using Nuclear and Mitochondrial DNA. Abstract for **Poster**. The Biology of Genomes at CSHL, Cold Spring Harbor, NY, USA (2013)

Sawyer S, Viola B, Shunkov M, Derevianko A, Pääbo S. Neandertal and Denisovan Genomes from the Altai. Abstract for **Plenary Talk**. European Society for the study of Human Evolution, Bordeaux, France (2012)

Sawyer S, Viola B, Gansauge MT, Shunkov M, Derevianko A, Pääbo S. The complete mitochondrial genome of third individual from Denisova Cave. Abstract for **Poster**. Society for Molecular Biology and Evolution, Dublin, Ireland (2012), received a

Sawyer S. Denisovans, Neandertals and Modern Humans in the Altai. Abstract for **Talk**. International Symposium in Denisova Cave, Altai Mountains, Russia (2011)

Awards and Scholarships

2012	Graduate Student Poster Prize, Society for Molecular Biology and Evolution
2004-2008	Park Scholarship: merrit based full scholarship for undergraduate studies at North Carolina State University, $40,000
2006-2008	Howard Hughes Medical Institute Scholar, two year research grant, $9000
2004-2008	North Carolina State University Honors Program

Public Understanding of Science

2013	Documentary: *Neandethals Decoded* (Nova)
2012	Documentary: *Sex in the Stone Age* (National Geographic)

Selbständigkeitserklärung

Ich versichere, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Ich reiche sie erstmals als Prüfungsleistung ein. Mir ist bekannt, dass ein Betrugsversuch mit der Note "nicht ausreichend" (5,0) geahndet wird und im Wiederholungsfall zum Ausschluss von der Erbringung weiterer Prüfungsleistungen führen kann.

Name: Sawyer

Vorname: Susanna

Leipzig

14.12.2015

Insights into Neandertals and Denisovans from Denisova Cave

Der Fakultät für Biowissenschaften, Pharmazie und Psychologie

der Universität Leipzig

eingereichte

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

Doctor rerum naturalium

Dr. rer. nat.

vorgelegt

von Master of Science Susanna Sawyer

geboren am 23.09.1985 in Trostberg, Deutschland

Leipzig, den 14.12.2015

BIBLIOGRAPHISCHE DARSTELLUNG

Susanna Sawyer

Insights into Neandertals and Denisovans from Denisova cave

Fakultät für Biowissenschaften, Pharmazie und Psychologie

Universität Leipzig

Dissertation

139 Seiten, 153 Literaturangaben, 30 Abbildungen, 32 Tabellen

Denisova Cave is located in the Altai mountains of Russia. Excavations from this cave have yielded two large hominin molars and three hominin phalanxes from the Pleistocene. One of the phalanxes (*Denisova 3*) had extraoridinary DNA preservation allowing the sequencing of high quality nuclear and mitochondrial DNA (mtDNA) genomes and has been shown to belong to a young girl from hereto unknown sister group of Neandertals, called Denisovans. The mtDNA of *Denisova 3* surpirisingly split from the mtDNA ancestor of modern humans and Neandertals twice as long ago as the split of modern humans and Neandertals. The mtDNA of one of the molars (*Denisova 4*) was also sequenced and differs at only two positions from the mtDNA of *Denisova 3.* A second phalanx (*Altai 1*) also yielded a high quality genome, and was a Neandertal. While Neandertals show an admixture signal of 1-4% into present-day non-Africans, Denisovans show an admixture of up to 5% in present-day Oceanians, and to a much lesser extent East Asians.

This thesis encompasses two studies. In the first study, we sequenced the complete mtDNA genome of the additional molar (*Denisova 8*), as well as a few megabases of nuclear DNA from *Denisova 4* and *Denisova 8*. While the mtDNA of *Denisova 8* is clearly of the Denisova type, its branch to the most recent common ancestor of Denisovans is half as long as the branch leading to *Denisova 3* or *Denisova 4*, indicating that *Denisova 8* lived many millenia before the other two. Both *Denisova 4* and *8* fall together with *Denisova 3* based on nuclear DNA, bringing the number of known Denisovans from one to three.

In the second study, we sequenced an almost complete mtDNA and a few megabases of nuclear DNA from the third hominin phalanx from Denisova Cave, *Altai 2*. Both the mtDNA and the nuclear DNA show *Altai 2* to be a Neandertal. The mtDNA also showed the presence of substantial Pleistocene spotted hyena contamination. Low levels of spotted hyena contamination were also found in *Altai 1*, *Denisova 3* and *Denisova 4*. Partial mtDNA genomes of the contaminating spotted hyenas from these four hominins were compared to mtDNA genomes of other extant and extinct spotted hyenas. We show that the spotted hyenas that contaminated the two Denisovans come from a population of spotted hyenas found in Pleistocene Europe as well as present-day Africa, while the spotted hyenas that contaminated *Altai 2*, and possibly *Altai 1*, come from a population of spotted hyenas found in Pleisticene eastern Russia and northern China. This indicates that Denisova Cave was a meeting point of eastern and western hominins as well as eastern and western spotted hyena populations.

For Emma.

Table of Contents

1. Thesis Summary

Denisova Cave is located in the Siberian Altai Mountains on the Anui River close to the borders with Kazakhstan, China and Mongolia. Although there are other archaeological sites in the region, including Okladnikov Cave to the North where Neandertal remains have been found, the stone tools excavated in the Altai Mountains show little change in culture from 300,000 to 30,000 years ago, which has prompted some to argue for a multi-regional human evolution model in central Asia (1, 2). In addition, the region has yielded few hominin remains (3). The hominin remains discovered are often small pieces, with no complete or even partially complete skeletons (3). The geographically closest complete skeletons are from two early modern human children in Mal'ta on the southern tip of Lake Baikal (4).

Excavations of Denisova Cave, led by Professor Anatoly Derevianko, have been ongoing since 1984. To date these excavations have yielded seven hominin remains from the Pleistocene (5, 6). One of these remains, a small piece of a terminal finger phalanx of a young child (*Denisova 3*), was found in 2008 in the East Gallery. Ancient DNA (aDNA) preservation in this bone was remarkably good and yielded not only a complete mitochondrial DNA (mtDNA) genome (7), but also a high quality nuclear genome (8). While the mtDNA of the *Denisova 3* split off twice as early as the split between Neandertals and early modern humans (7), the nuclear DNA showed that *Denisova 3* belonged to a sister group of Neandertals (8, 9) which was named Denisovans. In 2000, an unusually large molar (*Denisova 4*) was found in the South Gallery of the Cave. Before the present study, only mtDNA was retrieved from *Denisova 4* (9). The mtDNA has only two differences to the mtDNA of *Denisova 3*, indicating that the molar may have belonged to a Denisovan (9). In 2010, an intact toe phalanx was found in the East Gallery. Again the DNA preservation was good enough to produce high quality nuclear and mtDNA genomes, which revealed that the toe bone belonged to a Neandertal (10). I refer to this individual as *Altai 1*.

From previous Neandertal data, it was known that non-Africans carry 1-4% DNA from Neandertals (11). Denisovans show an admixture signal to present-day Oceanians (of 3-6%), and a small admixture signal of ~0.2% to East Asians (10). Additionally, *Denisova 3* carries >0.5% of Neandertal DNA that is more closely related to *Altai 1* than to more western

Neandertals. *Denisova 3* also carries DNA from another archaic hominin, from which it may have gotten its deeply diverged mtDNA genome (10).

While none of the hominin remains from Denisova Cave are directly C^{14} dated, dating based on branch shortening due to a lack of accumulated mutations estimates that *Altai 1* is older than *Denisova 3*, and that both lived between 120,000 and 50,000 years ago (10). During this time Pleistocene spotted hyenas lived in the Altai Mountains (12). Based on cytochrome b data from Pleistocene hyenas, they fall into the genetic variation of spotted hyenas in sub-Saharan Africa today (13, 14) They were larger in size (12), and have been shown to have eaten hominin remains in the area (3).

This thesis is composed of two studies. First I discuss two additional Denisovans from Denisova Cave. We sequenced a small amount of nuclear DNA sequences from *Denisova 4*, the large molar from the South Gallery. We then calculated the divergence of *Denisova 4* from *Denisova 3*, *Altai 1* and ten present-day humans from around the world, on the lineage to the human and chimpanzee ancestor. *Denisova 4* diverged from *Denisova 3* 2.9% back on this lineage, which is $1/3^{rd}$ of its divergence from *Altai 1* and $1/5^{th}$ of its divergence from the present-day humans. Therefore *Denisova 4* is a Denisovan.

We sequenced 24 Megabases (Mb) of the nuclear genome as well as the complete mtDNA genome from an additional third molar from the East Gallery of Denisova Cave (*Denisova 8*). The mtDNA genome of *Denisova 8* falls together with the two previously described Denisovans. However, its mtDNA is quite diverged, carrying almost twice as many differences to the other two mtDNA genomes as seen between pairs of Neandertal mtDNAs ranging from Spain to Siberia. The number of mutations leading to the *Denisova 8* mtDNA from the most recent common ancestor of Denisovans is almost half the number of mutations to the other two Denisovans. This translates to a 60,000 to 100,000 age difference between these two Denisovan groups. The nuclear divergence of *Denisova 8* is lower to *Denisova 3* than either Neandertals or present-day humans, therefore this additional molar also belonged to a Denisovan. However, the divergence of *Denisova 8* from *Denisova 3* is higher than that of *Denisova 4* from *Denisova 3*, and higher than the divergences between Neandertals.

The mtDNA branch shortening of *Denisova 8* suggests that Denisovans inhabited the Denisova Cave region at least twice over a very long time period, interrupted or possibly coexisting with Neandertals for some time. Since both of the molars from the Denisovans are

unusually large, it suggests that this morphology was present in the Denisovans of the Altai Mountains for a long time, and that is may aid in identifying Denisovans in the future.

The second study presented in this thesis centers on a fifth hominin remain from Denisova Cave. *Altai 2* is a complete finger phalanx found in the deepest layer of the East Gallery excavated to date. It is a hominin bone based on morphology. We sequenced an almost complete mtDNA genome from this specimen. It is of the Neandertal type and clusters closest with *Altai 1*, with only ten differences. We also sequenced 18.4 Mb of the nuclear genome. On the lineage to the human chimpanzee ancestor, the divergence of *Altai 2* is lowest to *Altai 1*, second lowest to *Denisova 3* and highest to ten present-day humans. However *Altai 2* has the highest divergence to *Altai 1* when compared to other Neandertals from Spain, Croatia and the Caucasus.

We found a large amount of spotted hyena DNA in the *Altai 2* bone (32.4% of total), which could explain the deep divergence of *Altai 2* to *Altai 1* on the nuclear level. We were able to sequence a partial mtDNA genome of the main contaminating hyena. It falls outside the variation of the four Pleistocene and extant spotted hyenas for whom complete mtDNA genomes exist. Based on data from cytochrome b from 57 present-day and Pleistocene spotted hyenas, the *Altai 2* spotted hyena contaminant is most closely related to Pleistocene spotted hyenas from eastern Russia and China.

We sequenced the almost complete mtDNA of a spotted hyena bone from the same layer and gallery in Denisova Cave. In addition we looked for spotted hyena mtDNA sequences among the DNA sequences determined from other Denisova Cave hominins. We found low level spotted hyena contamination in *Denisova 3* and *4* and *Altai 1*, but not in *Denisova 8*. The spotted hyena contaminants of *Denisova 3* and *4*, as well as the spotted hyena from Denisova Cave fall together with Pleistocene spotted hyenas from Europe and extant spotted hyenas from Africa. The *Altai 1* spotted hyena contaminant is more diverged, but not as diverged as the *Altai 2* hyena contaminant.

We washed the bone powder of *Altai 2* with a phosphate wash, which has been shown to be effective at washing off microbial DNA that has colonized a bone after the death of an individual (15). The microbial DNA content was higher in the phosphate wash than in the subsequent DNA extraction, however the hyena DNA did not preferentially wash off. In fact

the amount of hyena DNA present in the subsequent extraction was higher than the endogenous Neandertal DNA.

How did the hyena contamination end up in these remains? It is possible that spotted hyenas hunted hominins and left only small phalanxes and teeth behind, a practice not uncommon for spotted hyenas today with their prey (12, 16). Such an idea has been suggested for the nearby Okladnikov Cave, a small cave, more suited in size to a hyena than to a human (3). Okladnikov Cave has large amounts of hyena remains, and the hominin remains that were found there are believed to have been dragged into the cave (3). The Pleistocene hyenas may have scavenged the carcasses of hominins, possibly by digging up graves, again a behavior seen in spotted hyenas today (17).

It is possible, however, that the contaminating hyenas had no interaction with the hominins. The Denisovans and Neandertals may have left their remains some other way in Denisova Cave, and over the thousands of years until spotted hyenas died out 13-14,000 years ago (12), hyenas were digging in the cave and either eating the old bones or urinating and defecating in the cave, thus allowing their DNA to leech into the soil (18) and into the remaining bones and teeth. Even if Denisova Cave was not often used by either Denisovans or Neandertals, and was instead mostly a spotted hyena den, these hominins must have lived within the home range of the hyenas using the Caves, within 100 km of the Cave (17).

Denisova Cave was a meeting point of hominids from the east (Denisovans) and west (Neandertals) as well as Pleistocene spotted hyena populations from the east and west. Interestingly the western spotted hyenas are associated with Denisovans while the eastern spotted hyenas are associated with Neandertals. More sequencing and dating of Pleistocene spotted hyenas from Eurasia could potentially shed light on Neandertal and Denisovan movements in the area.

2. Zusammenfassung

Die Denisova-Höhle befindet sich im sibirischen Altaigebirge am Fluss Anui nahe der Grenze zu Kasachstan, China und der Mogolei. Trotz anderer archäologischer Ausgrabungsstätten in der Region – unter anderem die Okladnikov-Höhle im Norden, eine Fundstelle für Neandertalerknochen – zeigen die Steinwerkzeuge aus dem Altai-Gebirge vor 300.000-30.000 Jahren wenig kultuelle Veränderung, was bei einigen Wissenschaftler zur Annahme eines multiregionalen Evolutionmodells in Zentralasien führte (1 ,2).

Zusätzlich wurden in der Region wenige menschliche Überreste gefunden. Die ausgegrabenen menschlichen Fossilien sind oft nur kleine Fragmente, welche keinen vollständige Rekonstruktion des Skeletts zulassen (3). Die am besten rekonstruierten Skelette stammen von zwei früh-modernen Kindern aus Mal'ta von der südlichen Spitze des Baikalsees (4).

Seit 1984 finden, geleitet von Professor Anatoly Derevianko, Ausgrabungen in der Denisova Höhle statt. Dabei konnten bisher sieben menschliche Fossilien aus dem Pleistozän geborgen werden (5). Eines dieser Überreste – ein kleiner Teil von einem Fingerglied eines jungen Kindes (*Denisova 3*) – wurde 2008 in der Ost-Galerie gefunden. Die DNA-Erhaltung in diesem Knochen war bemerkenswert gut und lieferte nicht nur ein vollständiges Mitochnodrien DNA Genom (mtDNA) (6), sondern auch ein hochqualitatives nukleares Genom (7).

Anhand der mtDNA fand eine Abspaltung von *Denisova 3* zweimal früher statt als die Abspaltung des Neandertalers von der menschlichen Linie (6). Das nukleare Genome weist hingegen darauf hin, dass *Denisova 3* eine Schwestergruppe zu den Neandertalern bildet (7, 8), welche als Denisovaner oder Denisova-Menschen bezeichnet wird.

Im Jahr 2000 wurde ein ungewöhnlich großer Backenzahn in der Süd-Galerie der Höhle gefunden (*Denisova 4*). Anhand von Analysen der mtDNA von *Denisova 4*, die einzige genetische Information vor der Veröffentlichung der hier präsentierten Studie, wurde festgestellt, dass *Denisova 4* nur zwei Unterschiede zu *Denisova 3* aufweist, ein Indikator dafür, dass *Denisova 4* zu den Denisova-Menschen gehört (8).

2010 wurde dann ein intakter Zehenknochen in der Ost-Galerie ausgegraben, dessen DNA-Erhaltung erneut gut genug war um hochqualifizierte mt and nukleare Genome ervorzubringen. Analysen der DNA ergaben eine Zugehörigkeit zu den Neandertalern (9). Dieses Individuum wird fortlaufend als *Altai 1* bezeichnet.

Von früheren Neandertal-Daten ist bekannt, dass Nicht-Afrikaner 1-4% Neandertaler-DNA in sich tragen (10). Denisova-Menschen zeigen eine Vermischung mit heute lebenden Ozeaniern von 3-6%, mit Ostasiaten hingehen nur ~0.2% (9). *Denisova 3* weist >0.5% an Neandertaler-DNA auf und ist somit näher mit *Altai 1* als mit westlicheren Neandertalern verwandt. Zusätzlich zeigt *Denisova 3* Anzeichen von Vermischung mit einem weiteren unbekannten archaischen Menschen, von welchem es sein stark divergentes mtGenome haben könnte (9).
Das Alter der homininen Überreste wurde bisher weder durch C14-Datierung bestimmt, noch kann die Stratigraphie der Höhle zu Bestimmung herangezogen werden. Mit Hilfe von *branch shortening*, wobei das Alter an Hand fehlender Mutationenanhäufung nach dem Tod des Organismus bestimmt werden kann, wird geschätzt, dass *Altai 1* älter als *Denisova 3* ist und beide vor 120.000 - 50.000 Jahren gelebt haben (9).
Während dieser Zeit lebten außerdem Tüpfelhyänen des Pleistozän im Altai-Gebirge (11), welche zwar keine separate Spezies zu den heute lebenden Tüpfelhyänen der afrikanischen Subsahara darstellten (12, 13), die aber deutlich größer waren und sich von homininen Überresten ernährten (3).

Diese Doktorarbeit ist in zwei Teile unterteilt. Zuerst werden zwei neue Denisova-Menschen aus der Denisova-Höhle diskutiert. Teile des nuklearen Genoms des großen Backenzahns aus der Süd-Galerie (*Denisova 4*) wurden sequenziert. Außerdem wurde die Divergenz von *Denisova 4* zu *Denisova 3*, *Altai 1* und zehn modernen Menschen aus aller Welt auf der Linie zum Vorfahren des Menschen und Schimpanzen bestimmt. *Denisova 4* und *Denisova 3* divergieren zu 2.9 %, dies entspricht einem Drittel seiner Divergenz zu *Altai 1* und einem Fünftel seiner Divergenz zu allen modernen Menschen und kategorisiert *Denisova 4* als einen Denisova-Menschen.
Außerdem wurden insgesamt 24 Megabasen (Mb) des nuklearen Genoms und das komplette mtDNA Genom eines dritten Backenzahns aus der Ost-Galerie der Denisova-Höhle sequenziert (*Denisova 8*). Das mtGenome fällt taxonomisch zusammen mit den bereits beschriebenen Denisova-Menschen. Nichtsdestotrotz ist die mtDNA mit fast doppelt so vielen Unterschieden zu den beiden anderen mtGenomen so divergent wie beispielsweise die mtDNA zwischen einem Neandertalern aus Spanien und einem aus Sibieren. Die Anzahl der Mutationen zwischen *Denisova 8* und dem jünsgten Vorfahren der Denisova-Menschen ist fast halb so groß wie die

Anzahl der Mutationen von *Denisova 4* und *Denisova 3,* dies lässt sich in einen Altersunterschied der beiden Denisova-Gruppen von 60.000-100.000 Jahren übersetzen.
Die Divergenz von *Denisova 8* und *Denisova 3* ist niedriger als zu irgendeinem Neandertaler oder modernen Menschen, was den zusätzlich gefundenen Backenzahn ebenfalls zu den Denisova-Menschen gehören lässt. Dennoch ist die Divergenz zwischen *Denisova 8* und *Denisova 3* größer als zwischen *Denisova 4* und *Denisova 3,* sowie der Divergenz innerhalb der Neandertaler.
Auf Grund des mtDNA *branch shortening* von *Denisova 8* kann angenommen werden, dass die Denisova-Menschen die Denisova-Höhle über zwei lange Zeitspannen bewohnt haben, mit Unterbrechung oder möglicher zeitlichen Überlappung durch die Existenz von Neandertalern.
Auf Grund der ungewöhlichen Größer der Denisova-Backenzähne wird angenommen, dass diese Morphologie über lange Zeit bei den Altai-Denisovanern vorzufinden war und dass es als Unterstützung zur Identifizierung von weiteren Denisova-Menschen herangezogen werden kann.
Man geht davon aus, dass das divergente mtDNA Genom von *Denisova 3* durch Genfluss von einer unbekannten archaischen Population hervorgegangen ist, welche sich vor 1-4 Millionen Jahren von den Denisova-Menschen abgespalten hat (9). Die diversere mtDNA von *Denisova 8* könnte hingegen von einer divergenteren archaischen Gruppe stammen, was aufgrund der geringen DNA-Erhaltung des Fossils aber momentan leider nicht beantwortet werden kann.

Die zweite Studie aus dieser Doktorarbeit beschäftifgt sich mit einem fünften homininen Fossil aus der Denisova-Höhle. *Altai 2* ist ein kompletter Fingergliedknochen aus der bisher am tiefsten erschlossenen Schicht der Ost-Galerie. Basierend auf der Morpholigie ist der Knochen homininer Herkunft. Das mtDNA Genom der Spezies wurde fast vollständig sequenziert. Demnach gehört es den Neandertalern an und fällt, mit nur zehn Unterschieden, taxonomisch mit *Altai 1* zusammen. Desweiteren wurden 18.4 Mb des nuklearen Genoms seuqenziert. Auf der Linie zum gemeinsamen Vorfahren von Mensch und Schimpanze ist die Divergenz von *Altai 2* am geringsten zu *Altai 1,* am zweit geringsten zu *Denisova 3* und am höchsten zu den zehn modernen Menschen. Dennoch hat *Altai 2* verglichen zu Neandetalern aus Spanien, Kroatien und dem Kaukasus die höhste Divergenz zu *Altai 1*. Analysen ergaben einen hohen Anteil an DNA von der Tüpfelhyäne in *Altai 2* (32.4%), was die tiefe Divergenz zwischen *Altai 1* und *Altai 2* auf nuklearem Level erklären könnte. Außerdem gelang es einen Teil des mtDA Genoms zu

sequenzieren, welches von der Kontamination durch die Hyäne stammt. Die mtDNA fällt außerhalb der Variation der vier Pleistozän und rezenten Tüpfelhyänen, von welchen vollständige mtDNA Genome existieren. Basierend auf Daten für Cytochrom B von 57 heute lebenden und Pleistozän Tüpfelhyänen ist die *Altai 2* kontaminierende Tüpfelhyäne am nähesten mit den Pleistozän Tüpfelhyänen aus Ostrussland und China verwandt. Fast das gesamte mtDNA Genom einer weiteren Tüpfelhyäne aus derselben Schicht und derselben Galerie der Denisova-Höhle wurde sequenziert. Zusätzlich wurde nach weiteren von Tüpfelhyäne stammenden mtDNA Sequenzen in allen Sequenzen der Homininen aus der Höhle geschaut. Es wurde eine geringe Kontamination an Tüpfelhyäne in *Denisova 3*, *Denisova 4* und *Altai 1*, allerdings keine in *Denisova 8* gefunden. Die Kontaminaten von *Denisova 3* und *Denisova 4*, sowie die sequenzierte Tüpfelhyäne aus der Denisova-Höhle fallen taxonomisch mit den Pleistozän Tüpfelhyänen aus Europa und rezenten Tüpfelhyänen aus Afrika zusammen. Die Tüpfelhyänen-Kontamination von *Altai 1* ist divergenter aber nicht so divergent wie die von *Altai 2*.

Eine erst kürzlich veröffentlichte Phosphatwasch-Methode wurde auf das Knochenpulver angewendet. Es konnte gezeigt werden, dass die Methode von Bakterien stammende DNA, welche den Organismus nach dessen Tod besiedeln, effizient beseitigt (16). Der Anteil mikrobieller DNA im mit Phosphat gewaschenen Überstand war höher als in der anschließenden Extraktion. Dennoch ließ sich die Hyänen-Kontamination nicht effektiv entfernen. Der Anteil der Hyänen DNA im finalen Extrakt war sogar höher als der Anteil an endogener Neandertaler-DNA.

Wie kam es zur Kontamination der menschlichen Fossilien durch die Hyänen? Ist es möglich, dass die Tüpfelhyänen die Homininen jagdten und nur kleine Metapodien und Zähne zurückließen, ein nicht ungewöhnliches Verhalten auch heute lebender Hyänen im Umgang mit ihrer Beute (11, 14). Dies wurde auch für die nahegelegende Okladnikov Höhle angenommen, die auf Grund ihrer schmalen Größe mehr für Hyänen als Menschen geeigent schein (3). In der Okladinikov Höhle wurden viele Überreste von Hyänen gefunden. Von den menschlichen Fossilen glaubt man, dass diese durch Hyänen in die Höhle gebracht wurden (3). Die Kadaver wurden wahrscheinlich gründlich abgenagt und vergraben, ein Verhalten, wie es auch bei heute lebenden Tüpfelhyänen zu beobachten ist (17). Aber es ist auch möglich, dass die kontaminierenden Hyänen keinen Kontakt mit den Homininen hatten. Denisova-Menschen und Neandertaler könnten in der Höhle gelebt und ihr Überreste hinterlassen haben, welche dann von

den Tüpfelhyänen in den tausenden von Jahren bis zu ihren Aussterben vor 13.000-14.000 Jahren, ausgegraben und gefressen oder mit Kot und Urin verunreinigt wurden, was die Tüpfelhyänen-DNA im Boden (18) und auf den menschlichen Fossilien erkären würde.

Auch wenn die Denisova-Höhle wahrscheinlich vorwiegend als Hyänenbau und weniger als Unterschlupf für Denisova-Menschen und Neandertaler herhielt, mussten die Homininen dennoch in einem Umkreis von circa 100 km von der Höhle und damit im Habitat der Hyänen gelebt haben (17).

Zusammenfassend kann man also annehmen, dass das rätselhafte Fehlen von homininen Überresten im Altai-Gebirge höchstwahrscheinlich auf eine Kombination aus kleinen Populationsgrößen der Homininen und der Aktivität von Hyänen in der Region zurückzuführen ist.

3. Introduction

3.1 Human evolution

As humans we have always been fascinated by our origins, as evidenced by the countless religious beliefs that detail the formation of humans. Historically, humans were seen as special and were classed as a unique being (19). The belief that we as humans are special continued until 1863, when Thomas Huxley claimed that the difference between man and other great apes was not as great as the difference between great apes and lower apes (20). In 1871, Charles Darwin confirmed this idea (21). Already in 1856, however, with the discovery of the Neandertal, it was becoming clear that the origin of man is more complex than was formerly believed (22). It took developments in biochemistry and immunology, and most importantly in DNA sequencing, to be able to better describe our relationship with other primates. Using DNA, it became clear that we as humans are in fact most closely related to chimpanzees and bonobos (Figure 1) (23-25).

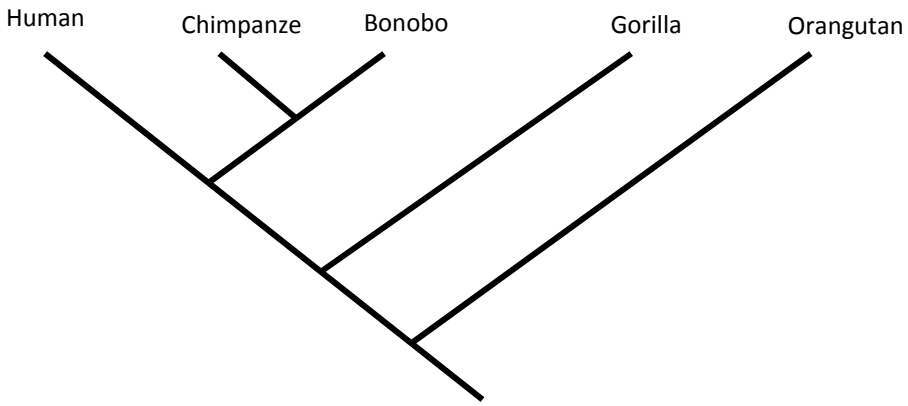

Figure 1. Tree of relationships between humans, chimpanzee, bonobos, gorillas and orangutans based on DNA. Tree is not drawn to scale.

There are many fossils that branch from or fall on the human branch after the split from the chimpanzee/bonobo common ancestor that range from primitive to modern (Figure 2). I will refer to this branch as the hominin branch and the fossils on this branch as hominins in this thesis. The relationship between these fossils is not always clear, specifically who our direct ancestors are and who represents an offshoot from our genetic history, although the fact that both the oldest fossils and our closest living relatives are only found in Africa indicates that hominins evolved in Africa (26). However already 1.8 million years ago, hominin fossils start appearing

outside of Africa (27). Thus there are multiple hypotheses about the evolution of present-day humans, such as the theory that early *Homo* left Africa, and then over the next million years evolved separately into Asians and Europeans, while the hominins left in Africa evolved into Africans, with some exchange of genes throughout this time (28). The out-of-Africa model claims that the ancestors of present-day humans evolved into anatomically modern humans in Africa, and that some of these anatomically modern humans left Africa recently, colonized Europe and Asia very rapidly and replaced the previous hominins living there (29). Genetic evidence from present-day humans made a clear case for the out-of-Africa model (30, 31), showing that present-day humans shared a common ancestor in Africa about 200,000 years ago (32) and that a small group of these early modern humans left Africa only 60,000 years ago (33). Archeological and linguistic data support this model as well (34-36).

Bones and teeth with Neandertal-type morphological features first appear in the fossil record about 400,000 years ago (37, 38). Neandertals have been found across Europe (22), the Middle East (39, 40), the Caucasus (41), and as far east as central Asia and Siberia (42-44). They disappeared around 30,000 years ago (45). Since the oldest early modern human fossils are found in the Middle East already 120,000 years ago (46, 47) there existed the possibility that Neandertals and early modern humans met.

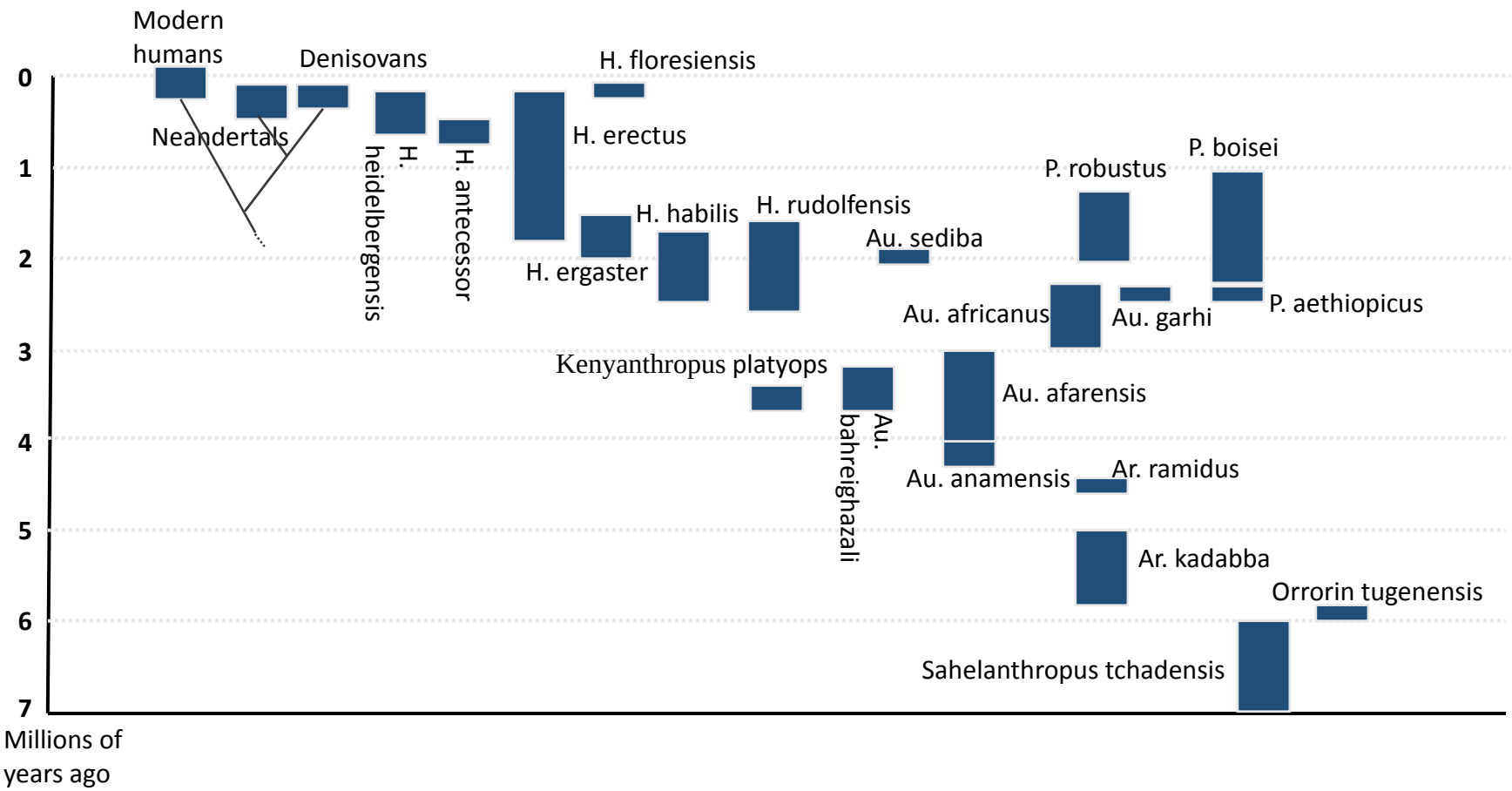

Figure 2. A depiction of hominin taxonomy. Relationships are only indicated if they are based on DNA evidence. Modified from (26). 'H.' refers to *Homo*, 'Au.' to *Australopithecus*, 'P.' to *Paranthropus*, and 'Ar.' to *Ardipithecus*.

3.2 Insights into hominin evolution using ancient DNA

While DNA from present-day humans is able to answer many questions about human evolution, far more questions can be answered if we can compare our genomes to the genomes of our closest known relatives, the Neandertals, as well as possible other unknown relatives. The field of ancient DNA was born with the sequencing of a few base pairs (bps) of an extinct zebra, the quagga (48), and an Egyptian mummy (49). The field of ancient DNA possesses many hurdles though. After death, cells rupture and DNA is no longer protected from the environment (50). Thus, the DNA that is extractable from the bones and teeth, is only present in a very small amount (51). In addition, the amount of DNA that is endogenous to the animal often represents less than 1% of the already small amount of DNA retrievable (51). Therefore any contamination from a present-day source will overwhelm the small amounts of DNA present, and can cause considerable problems if the contaminating DNA is closely related to the endogenous DNA, as is the case with present-day humans and Neandertals. Since the DNA has been exposed to the environment, it accumulates damage from the removal of the amine groups off of bases adenine (A), guanine (G) and cytosine (C) to produce hypoxanthine, xanthine and uracil (U) respectively, a process called deamination (52). The deamination of C to U is particularly relevant for this thesis. Due to uracils being read as a thymine (T) after sequencing, this form of deamination is read as a change in the sequence from a C (in the original, pre-damaged, fragment) to a T in the sequenced and damaged fragment. I refer to this form of damage as C to T changes in this thesis. C to T changes are the most visible of the types of deamination in ancient DNA sequences, and cluster at the ends of fragments (53, 54). Over ten percent of the Cs at fragment ends have been deaminated in samples older than 500 years, while Neandertals have at least 20 percent of their Cs deaminated (55). Thus, cytosine deamination has been used as a criterion for DNA authenticity (56, 57).

Despite these hurdles, it was still possible to sequence DNA from the hypervariable region of the 16,500 bp circular genome of the mitochondria of the Neandertal type specimen from Germany (58). Since each cell contains up to thousands of mitochondria, and therefore thousands of copies of the mitochondrial DNA (mtDNA), mtDNA is often a better target in ancient DNA than the >3 billion bp nuclear genome, of which only two copies per cell exist. These first sequences of the Neandertal were enough to conclude that Neandertal mtDNA falls outside the variation of human

mtDNA (58). Sequencing technologies during these times were very expensive, and therefore the sequencing of large amounts of ancient DNA was seen as unrealistic. The advent of high-throughput sequencing technologies allowed for much cheaper sequencing and for the complete mtDNA genome of a Neandertal from Vindija Cave, Croatia to be sequenced in 2008, which confirmed that this Neandertal mtDNA falls outside the variation of present-day humans, and that the Neandertal mtDNA diverged from modern humans 660,000 years ago (59). In 2010, the nuclear genomes of three Neandertals, also from Vindija Cave, were sequenced to low coverage (combined, each base was on average covered 1.5 times, referred henceforth as 1.5-fold coverage) (11). This draft genome of the Neandertal revealed that the humans leaving Africa did not completely replace the Neandertals they encountered, and instead admixed with them, as all non-Africans have a Neandertal ancestry of 1-4%. The draft genome confirmed the relationship shown in the mtDNA genome between present-day humans and the Neandertal (11).

In 2010, a small piece of a finger phalanx, belonging to a young child, was found in Denisova Cave in the northwest Altai Mountains in Siberia. The phalanx is called *Denisova 3*. The mtDNA genome of this finger bone showed a divergence twice as deep as the divergence between present-day humans and Neandertals (7). Due to the excellent DNA preservation in the bone, a low-coverage nuclear genome of 1.9-fold coverage quickly followed, and revealed that *Denisova 3* belonged to a girl from a population of hominins that are a sister group to Neandertals, subsequently named Denisovans (9). Denisovans also admixed with early modern humans who left Africa, but unlike the Neandertal admixture signal seen in all non-Africans, Denisovan admixture is seen in present-day humans living in Oceania (Australia, Melanesia and the Philippines) (60). A large third molar, *Denisova 4*, was found in Denisova Cave in 2000. The mtDNA from this tooth has two differences (out of 16,595 positions) to the mtDNA of *Denisova 3* (9).

Further advances in methodology, specifically the development of a new method in the library preparation of ancient DNA for sequencing, called the single-stranded library method, allowed for the production of a 30-fold nuclear genome of *Denisova 3* (8, 61). The high coverage of this genome means that the genome is of high quality, on par with the quality of genomes from present-day humans (8). In 2008, a further phalanx, this time from a toe from an adult, was found in Denisova Cave and again revealed excellent DNA preservation. A 50-fold high quality nuclear genome was produced in 2013, and showed that this toe phalanx belonged to a Neandertal (10).

The toe phalanx is officially called *Denisova 5*, however I refer to it as *Altai 1* in this thesis to avoid confusion.

The presence of two high-quality genomes, one each from a Neandertal and a Denisovan, as well as four low-coverage genomes from other Neandertals (the three Vindija Neandertals mentioned earlier and a 0.5-fold genome from *Mezmaiskaya1* from Mezmaiskaya Cave in the Caucasus of Russia (10, 62)) allow for a unique insight into hominin evolution. *Altai 1* is 1.3 times older than *Denisova 3*. Early modern humans split from the ancestor of Denisovans and Neandertals 1.4 times earlier than the split between Denisovans and Neandertals and six times earlier than the earliest split within modern humans (10). Denisovans and Neandertals had low levels of heterozygosity and thus small population sizes (about 30% of the heterozygosity in non-Africans), with long stretches of homozygosity in *Altai 1*, indicative of very recent inbreeding (10). Using the *Altai 1* genome, the admixture signal in non-Africans could be narrowed down to 1.5-2.1%, as well as to the introgressing Neandertal population, which is closest related to the *Mezmaiskaya1* Neandertal (10). A small amount of Denisovan admixture (~0.2%) was found in mainland Asian and Native American populations. *Denisova 3* has at least 0.5% admixture from Neandertals, specifically a Neandertal that was more closely related to *Altai 1* than *Mezmaiskaya1*, as well as 0.5-8.0% admixture from an unknown archaic hominin that split 1-4 million years ago from other hominins. The introgressing Denisovan population split from *Denisova 3* 3.5 times earlier than the introgressing Neandertal split from *Altai 1* (10).

3.3 The Altai Mountains and Denisova Cave

The Altai Mountains are a continuous mountain range that span from the borders of Russia, Mongolia, China and Kazakhstan to the southern tip of Lake Baikal. The north-west portion of the Altai, present in Siberia, make up 'Gorny-Altai' and are mostly foothills and lower mountains, which are divided by longitudinal and narrow river valleys (63) (Figure 3). Today this region has cool summers (+15°C average) and cold winters (-15°C average) (63). From 130-120kya, the climate in Gorny Altai was warmer than today and humid (2, 63) with widespread forests of pine, birch, spruce, cedar and broad-leafed trees (2). Over time the climate became slightly cooler and dryer (63), until 30kya when it became much cooler, which caused the forest to retreat and more and more steppe and meadows to appear (2). The climate in the Altai Mountains was buffered from the extreme temperature changes seen the surrounding low lands making the mountains a possible refugia for animals (63).

Hominins arrived in the Altai Mountains by at least 120kya (63), and left behind evidence of their presence in sites such as Ust-Karakol-1, Okladnikov cave and Anui-2, through the production of stone tools (lithics) (2) (Figure 3). However, the assemblages of lithics in the Altai Mountains may indicate sporadic occupations and high mobility (63). The caves in the region show a high frequency of micromammal and carnivore activity, which again argues for low-intensity human occupation (2, 63).

Denisova Cave is located at 51°22'50" N 84°41'20" E, 28 meters above the Anui River (2, 3). The cave was first excavated by Nikolai Ovodov in 1977 (3) and has been an active site of excavation since 1984, led by Anatoly Derevianko (2). It is a large cave made up of one main chamber and two galleries (East and South) (Figure 4). The main chamber and the cave entrance have been excavated, while the East Gallery is an active excavation site since 2005 (Figure 4) and the South Gallery has been excavated periodically. A chimney is located above the main chamber and and the cave is cold and damp, even in the warm summer months. It is unclear when the chimney appeared. There are over 8000 bone pieces excavated from layers 7-22 in the Main Chamber alone (3). Based on a sample of 116 of these bones, almost 3/4ths of these are morphologically indeterminable and small (5.2cm mean size) (3). Over 95% of the bones show peri-mortem damage, often in the form of marked bone processing or size reduction from humans or animals (3). Cut marks on bones are rare, which is in line with the indication that the region had low-intensity human occupation (63). The bones in the cave encompass 27 taxa of

large animals, including hyena, wolf, fox, cave bear, ibex, horse, deer, wooly rhinoceros, cave bear, yak, bison and saiga antelope (2). Denisova Cave shows a higher lithic assemblage than other excavations in the Gorny Altai region, including possible longer-term tool production, based on calculations from layer 12 in the main chamber, which calculates 250 artifacts over 1000 years for layer 12 in the Main Chamber (63).

During the excavation of the South Gallery in 2000, *Denisova 4* was found in layer 11.1 in the southern tip of the Gallery. The nuclear DNA extracted from this tooth will be discussed in this thesis. Unfortunately the stratigraphy of the South gallery is not published in detail. The East Gallery has produced four hominin remains (Figure 5), including *Altai 1* from layer 11.4 and *Denisova 3* from layer 11.2. In addition *Denisova 8*, another large third molar, was found on the border of layer 11.4 and layer 12 in 2010. During the excavation of the summer of 2011, a fourth hominin remain was found, an intact finger phalanx from an adult, *Denisova 10* (called *Altai 2* in this thesis), found in layer 12. Both mtDNA and nuclear DNA sequences from *Denisova 8* and *Altai 2* will be discussed in this thesis. The East Gallery shows significant disturbance in layer 11 (Figure 5; 'dist' in red), likely caused by water from the chimney, Pleistocene spotted hyenas digging dens or a combination of these and other factors. It is unclear if this disturbance continues into layer 12, but it may make the stratigraphy unreliable in this gallery.

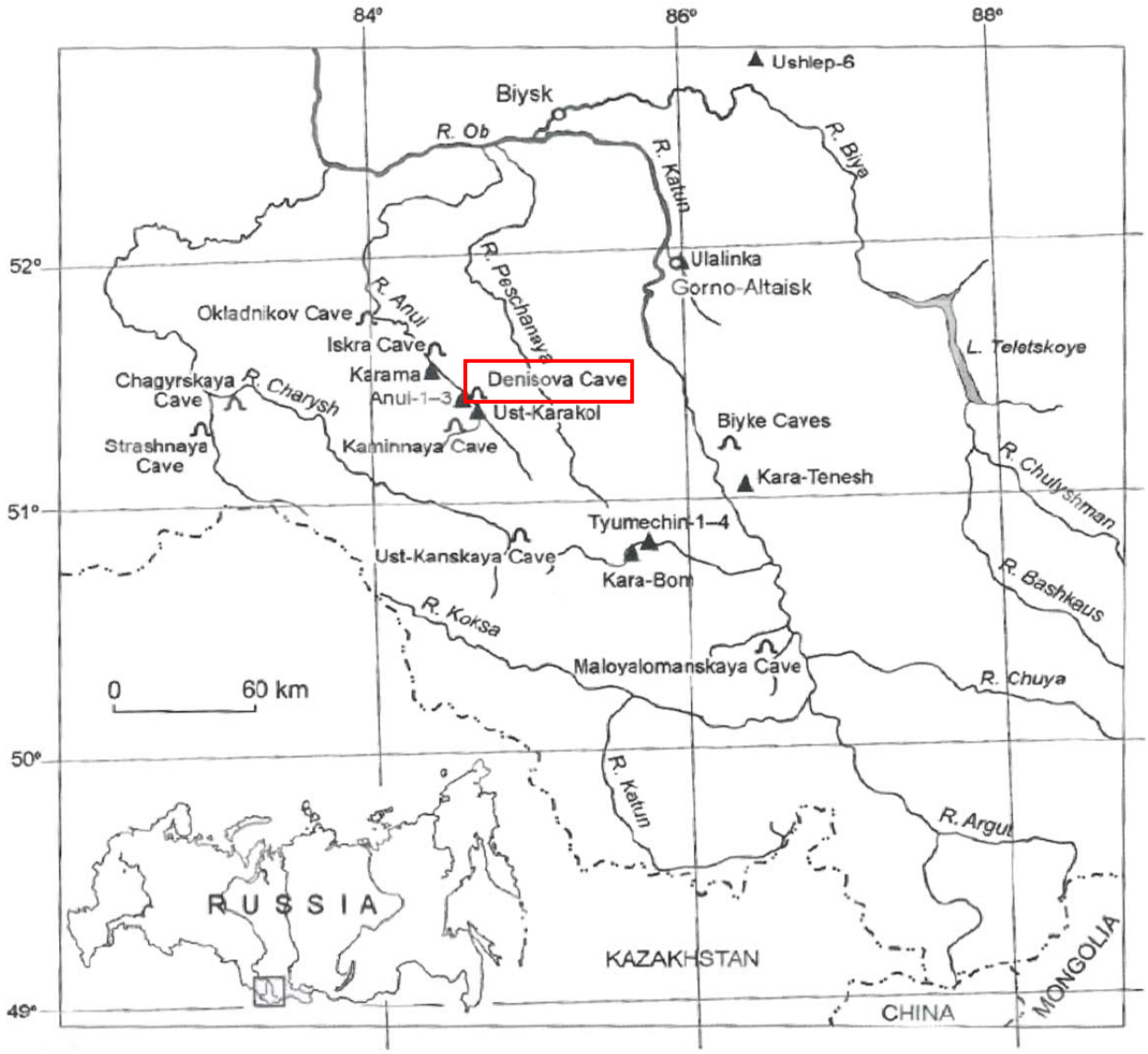

Figure 3. Map of the Altai region (Gorny Altai) of the Siberian Altai Mountains with Middle and Upper Paleolithic sites shown. Denisova Cave is marked with a red square (modified from (1)).

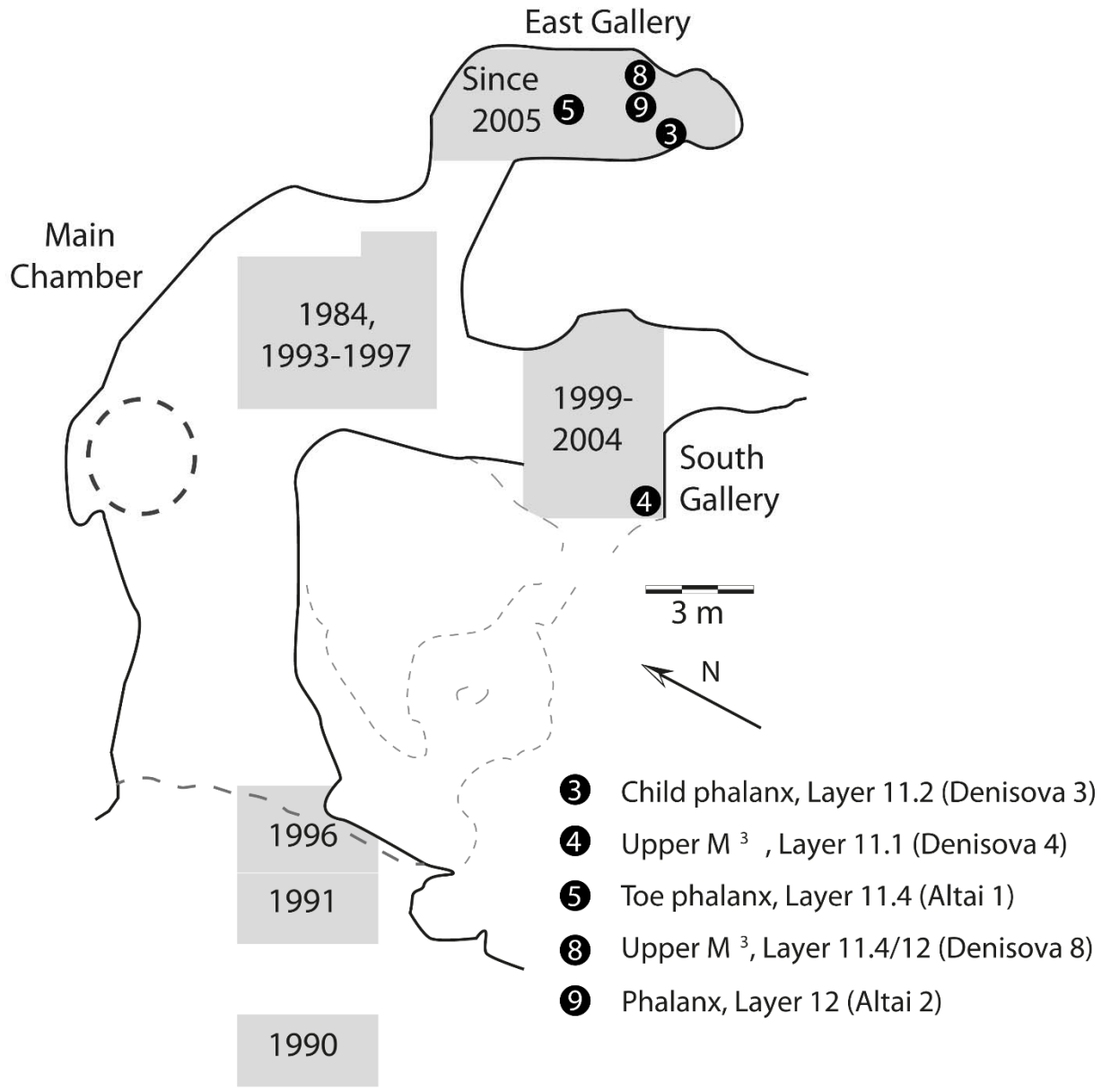

Figure 4. Layout of Denisova Cave (modified from original from Anastasy Abdulmanova and Bence Viola). Excavation sites and years are shown in the gray boxes. The locations of the five individuals discussed in this thesis are shown with their respective Denisova number. The dashed circle shows an approximate location for the chimney in the cave.

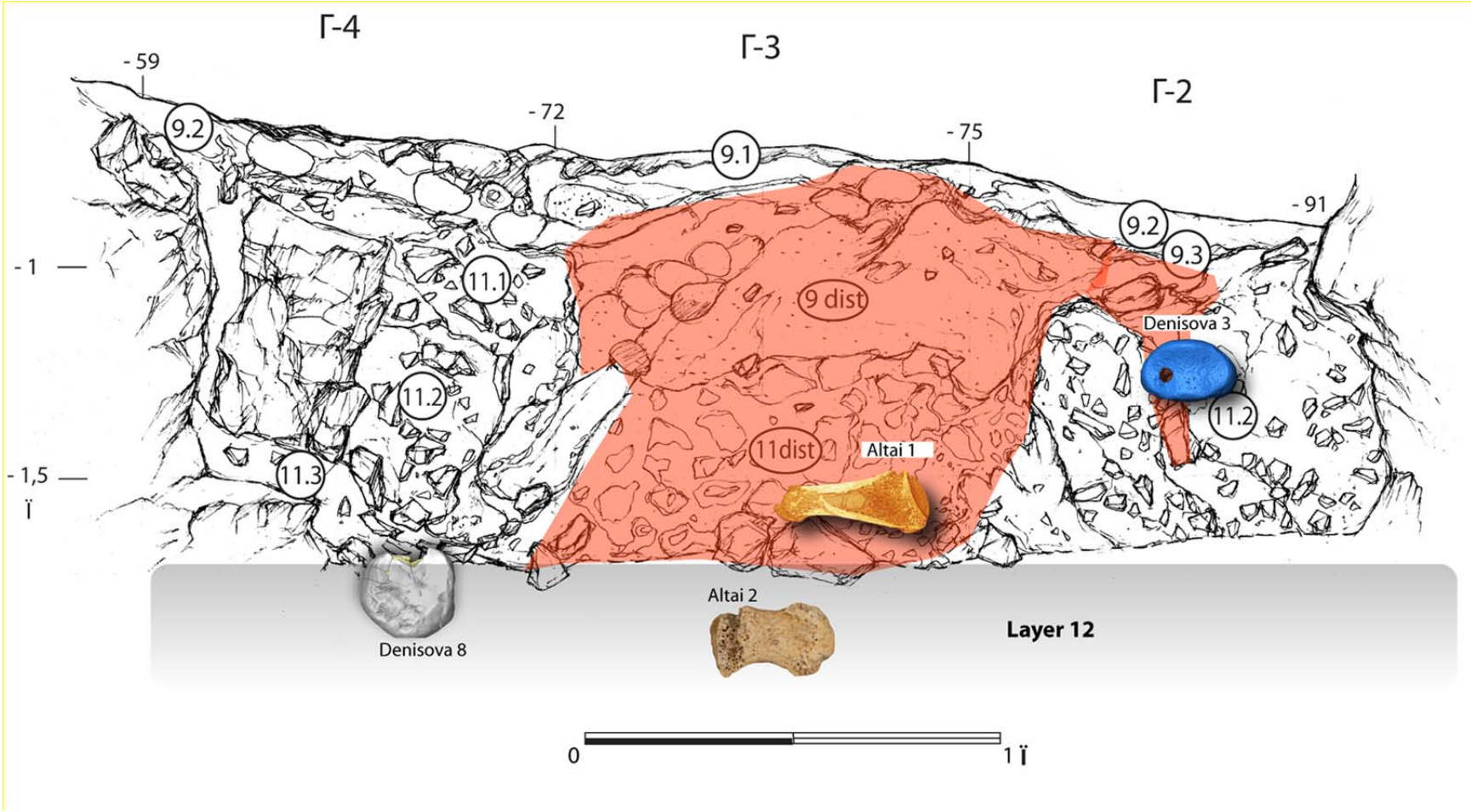

Figure 5. A representation of layer 11 in the East Gallery of Denisova Cave. The four bones/teeth found in the East Gallery that are discussed in this thesis are depicted along with their location in the layer. Layer 12 has no detailed representation, but is indicated. Modified from the original from Bence Viola and Anastasiya Abdulmanova.

3.4 Hyenas

The Hyaenidae family is a remnant of a once prolific family, which at its peak seven to eight million years ago had over 80 species spanning from Africa to Europe and east Asia (64). Today there are four hyena species left. The brown hyena (*Hyaena brunnea*) is a shaggy omnivore, which inhabits southern Africa (65). The striped hyena (*Hyaena hyaena*) looks similar to the brown hyena, but is slightly smaller, less shaggy and has a more distinct striped pattern (65). They inhabit northern Africa and are the only hyena to also inhabit areas outside of Africa, namely the Middle East and India. The aardwolf (*Proteles cristata*) is the smallest type of hyena and is also the only insectivore (65). It lives in southern and eastern Africa. The spotted hyena (*Crocuta crocuta*), which I will focus on in this thesis, is the largest of the hyenas. As indicated by its name, it has a spotted pattern and is a pure meat eater (65). Its range today encompasses much of sub-Saharan Africa (Figure 6). All four types of hyena are nocturnal (65).

Spotted hyenas are extraordinary carnivores. They have jaws and teeth built to crush and digest entire skeletons of large animals (66), except for hair, hooves and horn (3). They move an average of 27 km at night, and move at 10km/hr when searching for prey, but can run at 50km/hr over 0.5-2.5 km when chasing prey (66). They usually hunt in groups, but can also hunt alone (65, 66). Although spotted hyenas are famous for their scavenging behavior, they also hunt between 50-90% of their food depending on prey densities (17, 66). Their clan and territory sizes also range based on prey density, with clans varying in size between 8-80 individuals, and territories varying in size between 10-1000km^2. Spotted hyenas are matrilineal, with females taking on very masculine traits and showing marked aggression (17, 66). They dig extensive dens for their young (3).

The cave hyena of the Pleistocene had a vast range over most of Europe and Asia to Africa, with a northern limit of the 56th parallel (Figure 6) (3, 12). They were larger than the spotted hyena today (12) and were therefore often grouped as a separate species (3, 12, 64). It has been shown in previous studies of cave hyena mtDNA that these hyenas fall into the variation of present-day spotted hyenas (13, 14), and therefore will be referred to as spotted hyenas in this study. Spotted hyenas went extinct in Europe and Asia 13-14kya (3) at the end of the Pleistocene. Older specimens (>400kya) were smaller, about the same size as spotted hyenas today, but they gradually increased in size (12). The origin of the spotted hyenas has been theorized to be in either Asia (12, 14) or Africa (13).

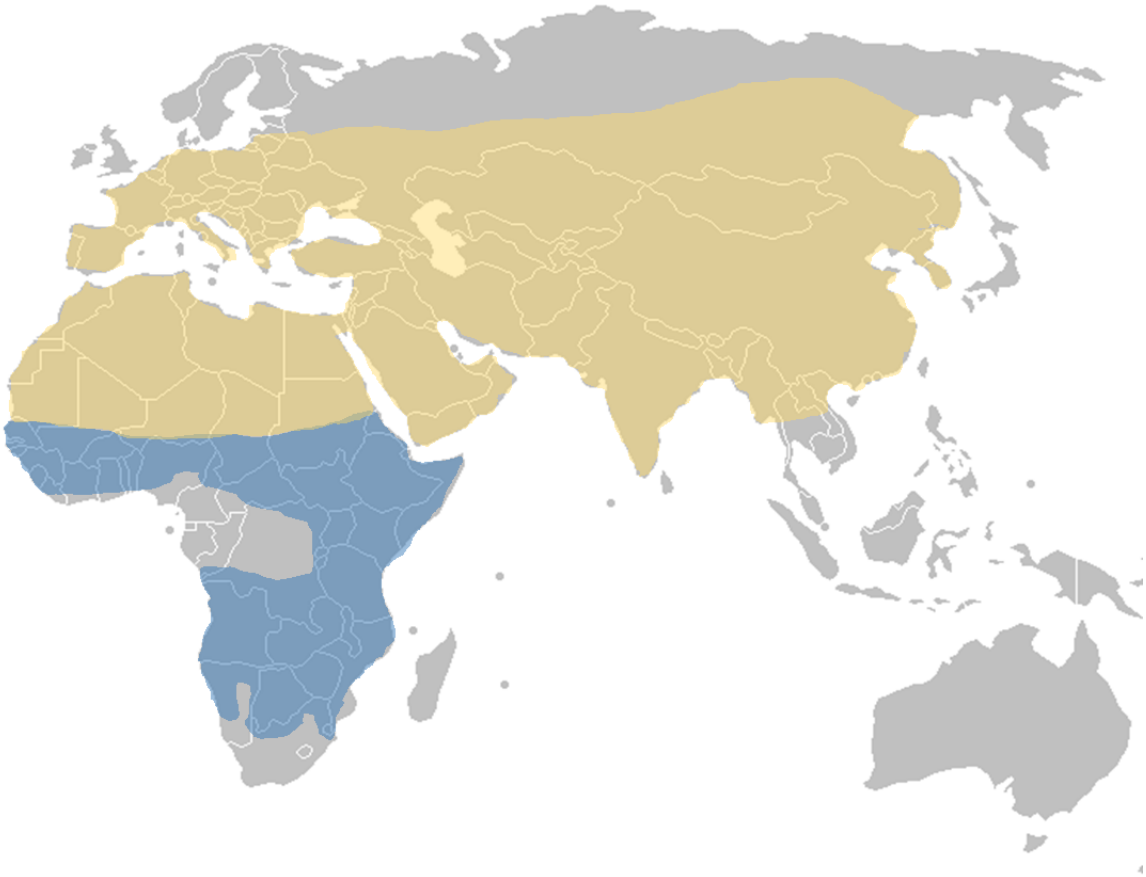

Figure 6. Map of spotted hyena ranges. Current spotted hyena range is shown in blue, the maximum range in the Pleistocene is shown in yellow (after (12)).

4. Materials and Methods

4.1 Nuclear and mitochondrial DNA sequences from two Denisovan individuals

4.1.1 DNA extraction, library preparation, amplification, mtDNA capture, and sequencing

Thirty six millligrams (mg) of dentin were removed from the inside of the enamel cusp of *Denisova 8* using a dentistry drill and used to produce 100 microliters (uL) of extract as described (67). From $1/20^{th}$ of this extract, as well as from $1/10^{th}$ of a previous 100uL extract made from 40mg of *Denisova 4* (9), we produced Illumina libraries, using a single-stranded library preparation protocol that maximizes the yield of sequences from ancient DNA (8). The libraries were treated with *E. coli* Uracil DNA Glycosylase (UDG) and endonuclease VIII to remove uracils (U) (68). UDG does not effectively excise terminal Us (8). The *Denisova 4* library (L9234, see Table 1) had a final volume of 40uL in EBT (10mM Tris-HCl, pH 8.0; 0.05% Tween-20), while *Denisova 8* (B1113) had a final volume of 20uL in EBT.

The concentrations of L9234 and B1113 were measured by qPCR. L9234 from *Denisova 4* was split into two equal parts and used as template for an indexing PCR using two distinct indexing primers per library. The indexing PCR was performed using AccuPrime Pfx DNA polymerase (Life Technologies) and purified with the MinElute purification system as described (8). The purified and indexed libraries were each eluted in 30uL of EB (Qiagin MinElute Kit) to produce L9243 and L9250. An indexing PCR was also performed on B1113 from *Denisova 8* as described above except that all of B1113 was used in one indexing reaction to produce L9108.

To produce larger amounts of amplified library for the mtDNA enrichment, 5□L of L9243 from *Denisova 4* and of L9108 from *Denisova 8* were further amplified with Herculase II Fusion using adapter primers IS5 and IS6 (8, 69), purified with MinElute and eluted into 20□L of EB. DNA concentration was measured on a Nanodrop (ND-1000) and 500ng of the amplified DNA were enriched for human mtDNA via a bead-based protocol where PCR products are sheared, ligated to biotinylated linkers and immobilized on streptavidin-coated beads (70). The enriched libraries were quantified by qPCR and amplified with Herculase II Fusion, taking care not to reach PCR plateau. After measuring DNA concentration on a Bioanalyzer 2100 (Agilent) the *Denisova*

4 capture product (L9320) was sequenced on 1/7th of an Illumina MiSeq lane and the *Denisova 8* capture product (L9126) on 1/10th of an Illumina GAII lane.

For shotgun sequencing, the two libraries from *Denisova 4*, L9243 and L9250 (see Table 1), were amplified with Herculase II Fusion. Molecules with insert sizes between 35 and 450 bp were isolated using gel electrophoresis as described to produce L9349 and L9350 (8). L9108 from *Denisova 8* was also size fractionated to isolate molecules of lengths between 40 and 200 bp using gel electrophoresis without prior amplification to produce L9133. This library was amplified and quantified on the Bioanalyzer 2100 (Agilent) along with L9349 and L9350. The two *Denisova 4* libraries (L9349 and L9350) were pooled in equimolar amounts and sequenced on two Illumina HiSeq 2500 High Output flowcells, while the *Denisova 8* library (L9133) was sequenced on one High Output flowcell.

Table 1. Extraction and library IDs. IDs of *Denisova 4* and *8* after each processing step are given. The *Denisova 4* single-stranded (ss) library was split into two aliquots for the indexing amplification.

	Extract ID	(ss)Lib ID	Lib ID after Indexing	Lib ID after mtDNA capture	Lib ID after gel excision for shotgun seq
Denisova 4	E324	L9234	L9243	L9320	L9349
			L9250	-	L9350
Denisova 8	E652	B1113	L9108	L9126	L9133

4.1.2 Sequence processing and mapping

Ibis v1.1.6 (71) was used for base calling and sequence processing was carried out as described (72). Briefly, after base-calling, reads were demultiplexed allowing a single mismatch in the indexes; Illumina adapters were identified and removed, and overlapping read-pairs merged when the overlap was at least 11 bp. For all sequences the following basic filters were applied:

- Sequences with more than 5 bases with base qualities less than 15 (phred score) were removed

- Sequences having a base with a quality less than 10 (phred score) in the index reads were removed
- Sequences shorter than 35 bp were removed
- PCR duplicates were identified based on the same beginning and end coordinates and collapsed

MtDNA sequences were aligned to the mitochondrial sequence of the high coverage *Denisova 3* phalanx (NC_013993.1) using MIA (parameters: -c, -i) ((73), https://github.com/udo-stenzel/mapping-iterative-assembler) which was also used to generate what approximates a 75% consensus sequence.

The shotgun-sequenced fragments were aligned to hg19 (74) using BWA v.0.5.10 (75) with a maximum edit distance (-n option) of 0.01, a maximum of 2 gap openings (-o 2), and without a seed (-l 16500).

4.1.3 Present-day human mtDNA contamination estimate

We identified 183 and 174 "diagnostic positions" in *Denisova 4* and *Denisova 8, respectively*, where their consensus mtDNA sequences as estimated by MIA differ from every individual in a panel of 311 present-day humans from around the world.

We then re-aligned all captured sequences from the two molars to the human mtDNA reference sequence (76) using BWA version 0.5.10 (75) with relaxed parameters (-n 0.01, -o 2, -l 16500). This allows modern human mtDNA fragments that differ from the Denisovan mtDNA to be identified. Fragments carrying present-day human variants at the diagnostic sites were counted as contaminants, while fragments carrying consensus variants were counted as endogenous. 95% confidence intervals were calculated using a Wilson score interval.

The shotgun sequences were aligned to the human mtDNA reference sequence as described above, and, using the same diagnostic positions as above, mtDNA contamination estimated for the shotgun data.

4.1.4 Present-day human nuclear contamination estimate

To estimate present-day human contamination in the nuclear sequence data, we calculated the divergences of two French individuals to each other as well as two Sardinian individuals to each other (see Figure 13 for explanation of divergence calculation) and used these divergences as a

hypothetical contamination of 100% (*c,* Figure 13). Similarly, we used the divergence of the *Denisova 3* phalanx sequences to the four Europeans as a proxy for 0% contamination (*a*, Figure 13). We then calculated the divergence of *Denisova 4* and *Denisova 8* to the French and Sardinians using sequences that had not been filtered for a terminal C to T change (*b*, Figure 13). The percent contamination in the *Denisova 4* and *Denisova 8* sequences were then calculated as (a-b/a-c)x100.

4.1.5 C to T substitutions and aDNA authenticity

To determine whether different populations of molecules that differ in their extent of cytosine deamination-induced C to T substitutions occur in the libraries, we calculated the apparent C to T substitution rate at the 5'- and 3'-ends of DNA fragments. We then calculated the 5' C to T rate of fragments that have a 3' C to T and *vice versa*. Since deamination-induced misincorporations are rare in modern DNA that contaminates ancient DNA preparations (55, 56), it is unlikely that such DNA fragments carry C to T changes on both ends. In contrast, DNA molecules that carry a C to T change at one end are likely to be ancient and the C to T rate at the other end of such molecules can thus be taken to approximate the deamination rate in ancient, endogenous molecules (under the assumption that deamination at the two ends of molecules is independent). By comparing the C to T rates of all sequences to those that carry C to T at one end we can thus gauge if two or more populations of molecules that differ in their rates of deamination occur in the libraries and thus if contamination may exist in a library. 95% CIs were calculated using Wilson score intervals. Although this approach may be affected by factors that we do not fully understand, it yields contamination estimates for *Denisova 4* of 54-69% and 1.3-6.1% for *Denisova 8* (Table 6) which are qualitatively compatible with ones based on divergence above. For the mtDNA the 95% CIs of the C to T rates of the two populations of molecules overlap (Table 6).

4.1.6 Sex determination

For sex determination, we used sequences that passed the filters described in section 4.1.3 have a minimum map quality of 37 (phred scale).

We identified regions on the sex chromosomes that are $\geq$500 bps long and pass the mappability filter. The mappability filter removes positions where at least one overlapping window of 35bp length maps to a different position in the genome with up to one mismatch (10). On the Y-chromosome we in addition excluded positions that overlap with sequences from four females

from the 1000 Genomes Project (NA12878, NA12892, NA19240, NA19238) (10). This left us with 627,426 bp on the Y chromosome and 40,661,238 bp on the X chromosome.

The number of sequenced fragments expected to fall in these regions if the individuals were male is: (Number of fragments aligned to the whole genome) × (the number positions in the X or Y-chromosome) / (genome size), where genome size is: 2 × (autosomal positions) + (X-chromosomal positions) + (Y-chromosomal positions).

We then determined the number of fragments that actually fall within these regions using either (*i*) all fragments or (*ii*) only those that carry putative deamination-induced C to T substitutions. We determined if the observed and expected numbers are significantly different from the male expectation using a Chi-square test (chisq.test) in the R package 3.1.0 (77). For the X-chromosomal fragments carrying C to T substitutions, we also determined if there is a significant difference under the female expectation. Both *Denisova 4* and *8* are more likely to come from males than from females. See Table 7.

4.1.7 Sex chromosome present-day human contamination estimate

Because the molars come from male individuals, we can estimate the fraction of fragments due to female contamination using the number of "extra" fragments mapped to the X-chromosome relative to the expected number if the individual is male and all Y-chromosome fragments are assumed to be endogenous. The contamination rate is then the difference between the number of fragments mapped to the X chromosome and the number expected if the individual is male divided by number expected if the individual is male. A Wilson score interval was used to calculate 95% CIs.

4.1.8 mtDNA phylogenetics

. The mtDNA sequences of the three Denisovan individuals, seven Neandertals (Altai – KC879692, *Mezmaiskaya1* – FM865411.1, Feldhofer 1 – FM865407.1, Feldhofer 2 – FM865408.1, Vindija 33.16 – AM948965, Vindija 33.25 – FM865410.1 and Sidron 1253 – FM865409.1) (10, 73), five present-day humans (San – AF347008, Yoruba – AF347014, Han Chinese – AF346972, French – AF346981 and Papuan – AF347004) (78) and the chimpanzee (X93335.1) (79) were aligned using the software MAFFT v6.708b (80, 81). Pairwise mtDNA differences among the seven Neandertals and three Denisovans were calculated using MEGA 6.06 (82). In addition, the three Denisovan

mtDNAs were aligned with 311 modern human mtDNAs and the pairwise differences among these individuals were calculated.

To estimate phylogenetic relationships, Modeltest 3.7 (83) was used to identify an appropriate substitution model (GTR+G+I) and MrBayes 3.2 (84, 85) was run with default MCMC parameters for 5,000,000 generations, sampling every 1,000 generations, using a burn-in of 1,000,000 generations. The 4,000 resulting trees were combined to a consensus using TreeAnnotator v1.6.2 from the BEAST package (86) (Figure 15A).

A tree including the partial mtDNA sequence of a hominid from Sima de los Huesos, Spain (KF683087.1) (87) was estimated as above (Figure 16).

4.1.9 mtDNA dating

The most recent common ancestor (MRCA) of the three Denisovans was estimated using parsimony and a Yoruba mtDNA (AF347014). There were two positions where the MRCA was not resolvable. The MRCA of the seven Neandertals was calculated in the same way, with five unresolvable positions. The pairwise differences between the MRCAs and each individual were then calculated (Table 10).

We estimated the age of the two molars and the divergence times between the three Denisovans, five radiocarbon-dated Neandertals (18), ten radiocarbon-dated ancient modern humans (88) and the five present-day humans used for tree estimations (Fig. 2) using BEAST v1.6.2. The age of *Denisova 3* date was set to either 50,000 years or 100,000 years as in ref. (10). A strict as well as a relaxed uncorrelated lognormal molecular clock was used with a normally distributed substitution rate prior of 2.67×10^{-8} per site per year (88) (standard deviation 1.0×10^{-8}), a Bayesian skyline coalescent tree prior with a uniform population size prior of 1,000 to 1,000,000 individuals, and a TN93 substitution model (89) . MCMC runs were carried out for 100,000,000 generations, sampling every 10,000 generations, with a burn-in of 10,000,000 generations. As expected, the relaxed clock is a better fit to the data and was used for the estimates presented in Table 12.

4.1.10 Watterson's estimator θw

θw was calculated for the three Denisovan individuals and the seven Neandertal, 31 Europeans (Italians, Germans, Spanish, Saami, English, Dutch, Finnish and French) and 311 present-day humans (including the Europeans) (Table 11). θw was calculated as follows: $K/a_n/16{,}595$, where

K is the number of segregating sites, and a_n is $\sum_{i=1}^{n-1} \frac{1}{i}$. The numbers of segregating sites were ascertained using DNA Sequence Polymorphism (DnaSP) version 5.10.01 (90).

4.1.11 Autosomal data filtering

The following filters were implemented for the *Denisova 4* and *Denisova 8* autosomal analyses:

- Filters outlined in section 4.1.3
- A minimum map quality of 37 (PHRED scale)
- Base quality set to 2 (phred scale) for Ts at the first or last two positions of fragments (to avoid errors induced by cytosine deamination)
- A minimum base quality of 30 (PHRED scale) (results in removal thymines with low base quality from step above)
- mapability filter that retains all positions where all possible overlapping 35-mers do not have match elsewhere in the genome allowing for one mismatch (10)
- Removal of triallelic sites
- Removal of CpG sites if the CpG occurs in either human, chimpanzee, gorilla or orangutan
- Removal of sites with a coverage higher than 2-fold
- When estimating nucleotide misincorporations due to cytosine deamination positions where the human reference (hg19) carries a C but one or more present-day human from the 1000 Genomes carries a T were excluded.

For high-coverage genomes the following filters were used:

- mapability filter that retains all positions where all possible overlapping 35-mers do not have match elsewhere in the genome allowing for one mismatch (10)
- Root mean square of the map quality >= 30
- Coverage cut-off of 2.5% on each side of the coverage distribution; corrected for GC content for the *Denisova 3* and the Altai Neandertal (10)

4.1.12 Autosomal divergence calculation

We estimate the divergence for *Denisova 4* and *Denisova 8* to ten present-day humans (French - HGDP00521, Sardinian - HGDP00665, Han - HGDP00778, Dai - HGDP01307, Papuan - HGDP00542, Australian - SS6004477, Dinka - DNK02, Mbuti - HGDP0456, Yoruba -

HGDP00927, San - HGDP01029) (8, 10)), the high-coverage *Denisova 3* genome (8) and the high-coverage Altai Neandertal genome (10). The variant call format (VCF) files for the present-day humans as well as the *Denisova 3* and the Altai Neandertal were filtered as stated above.

Divergences between low-coverage and high-coverage genomes are estimated as the percentages of substitutions from the human-chimp ancestor to high-coverage genomes that occurred after the split of the low-coverage genomes from high-coverage genomes (see Figure 17A). Ancestral states for the human-chimpanzee ancestor was taken from the 6-way primate EPO alignments from Ensembl version 69 (genome-wide alignments of human, chimpanzee, gorilla, orangutan, macaque, marmoset) (91, 92) and substitutions were parsimoniously assigned to one of the three lineages. Random alleles were picked at heterozygous sites in the high-coverage genomes while for the low-coverage Denisovan molars a random fragment was picket to represent each site analyzed. Standard errors for the divergence estimates (Table 13-16) were estimated by running 5,000 jackknife replicates of the divergences in 5 Mb windows. Standard errors were multiplied by 1.96 to generate 95% CIs.

We similarly estimated divergences to the high-coverage Altai Neandertal genome (10) for low-coverage data from Vindija Cave, Croatia (Vindija 33.16, Vindija 33.25, Vindija 33.26), from El Sidron Cave, Spain (Sidron 1253), from Feldhofer Cave, Germany (Feldhofer 1) (all available from ERP000119, (11)), and from Mesmaiskaya Cave, Russia (*Mezmaiskaya1*) (10). We excluded regions with a coverage higher than 2-fold for Feldhofer 1, 3-fold for the Vindija Neandertals and 4-fold for the *Mezmaiskaya1* Neandertal. We removed putative deamination-induced C to T substitutions at first and last two positions of the fragments from the *Mezmaiskaya1* Neandertal, as a double-stranded library preparation method and *E. coli* UDG was used, which does not remove uracils efficiently at these positions. For the other low-coverage Neandertals, which were not UDG treated, we removed putative deamination-induced C to T substitutions at the first and last five bases. We calculated the divergence of these six low-coverage Neandertals to the Altai Neandertal along with a 95% CI as above (Table 16).

4.1.13 D-statistics studies on autosomal data

D-statistics (93) were calculated from genotype calls for high-coverage genomes, picking random alleles at heterozygous positions, or from random fragments for low-coverage genomes. Ancestral states were from the EPO alignment (91, 92) (Ensembl v69).

When the low-coverage *Mezmaiskaya1* genome was analyzed together with the high-coverage Altai Neandertal genome, random DNA sequences were picked from both genomes to avoid problems resulting from the difference in sequence quality between the two genomes.

Errors in the low coverage genome sequences contribute apparently derived alleles. To test if derived alleles in DNA sequences determined from *Denisova 8* tend match derived allele in one present-day person more than another, we used *Denisova 8* fragments and asked if derived alleles in *Denisova 8* match derived alleles in one or the other of two individuals from different African populations. This is not the case (D=0.01, Z=0.73).

Table 17 shows that *Denisova 8* tends to share more derived alleles with the Papuan or Australian genomes using all sites (*D*:-0.03 to -0.08, Z-score: -1.9 to -4.3). However, the amount of data limits the power, as can be seen for similar comparisons using the whole high-coverage *Denisova 3* genome (*D*:-0.05 to -0.07, Z-score: -4.2 to -10.1).

To see if the amount of data determined from *Denisova 8* is enough to detect the excess sharing of derived alleles with the Altai relative to the *Mezmaiskaya1* previously described (8), we restrict the analysis to positions in the *Denisova 3* genome covered by the *Denisova 8* fragments and failed to detect the extra sharing (Table 18). As expected from this, we fail to detect any excess sharing of derived alleles between *Denisova 8* and the Altai genome *(*Table 18) when we restricted the analysis to transversions in order to avoid aberrant results due to errors in the low-coverage *Mezmaiskaya1* genome (not shown).

4.2 A Neandertal from Denisova Cave with ancient spotted hyena contamination

4.2.1 DNA extraction and library preparation

Bone powder was obtained by drilling from 3 locations in the *Altai 2* bone (Figure 7). DNA was extracted from bone powder from two of the locations as described in Dabney *et al* (94). Bone powder from the third location was treated with phosphate prior to extraction (15).

Four libraries were prepared from between 10 and 20% of the extracts using single-stranded DNA library preparation with and without UDG treatment. In addition, two libraries were prepared using a previously described U selection method (95).

Library yields were quantified by qPCR. Libraries were amplified and barcoded with two sample-specific indices as described elsewhere using AccuPrime Pfx polymerase (94, 96). For some libraries (L9467, L9366 and L9367) an optimal number of amplification cycles was determined based on the results of qPCR. The other libraries were fully amplified into plateau using 35 PCR cycles. These libraries were amplified for one further cycle to remove heteroduplices since the single melting and hyrbridization allows misaligned sequences to again align correctly. See Figure 9 and Table 2 for more details.

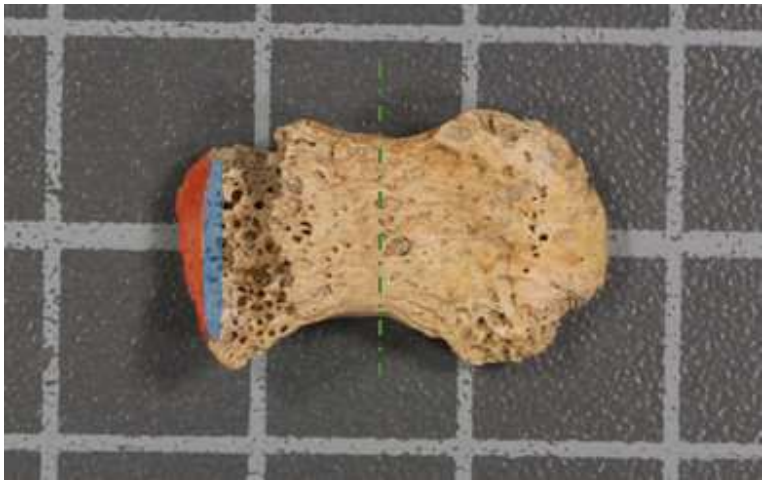

Figure 7. Locations of drilling for SP2990 (*Altai 2*). E1114 came from bone powder drilled from the red area and E1269 came from the blue area. The bone was cut down the middle as indicated by the green dotted line. E3000/E3001 was then drilled from inside the bone after cutting.

The Denisova spotted hyena SP3388 bone fragment was drilled once with a dentistry drill to produce 52 mg of bone powder (Figure 8). All of this bone powder was turned into a DNA extract following the method from Dabney *et al* (94). A single-stranded library was produced from 30% of the extract (61). The entire library was then indexed (96) and amplified into plateau to create library A2396.

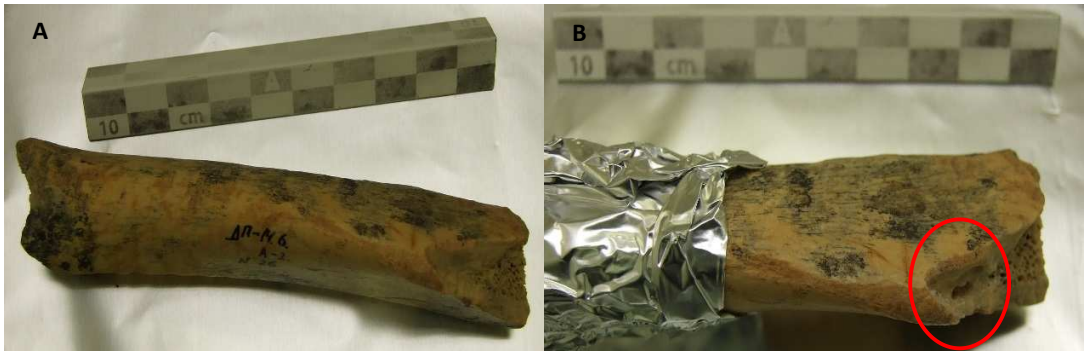

Figure 8. Picture of the Denisova Cave spotted hyena bone, SP3388. **A.** The bone in its entirety. **B.** The bone after drilling: area shown by a red circle.

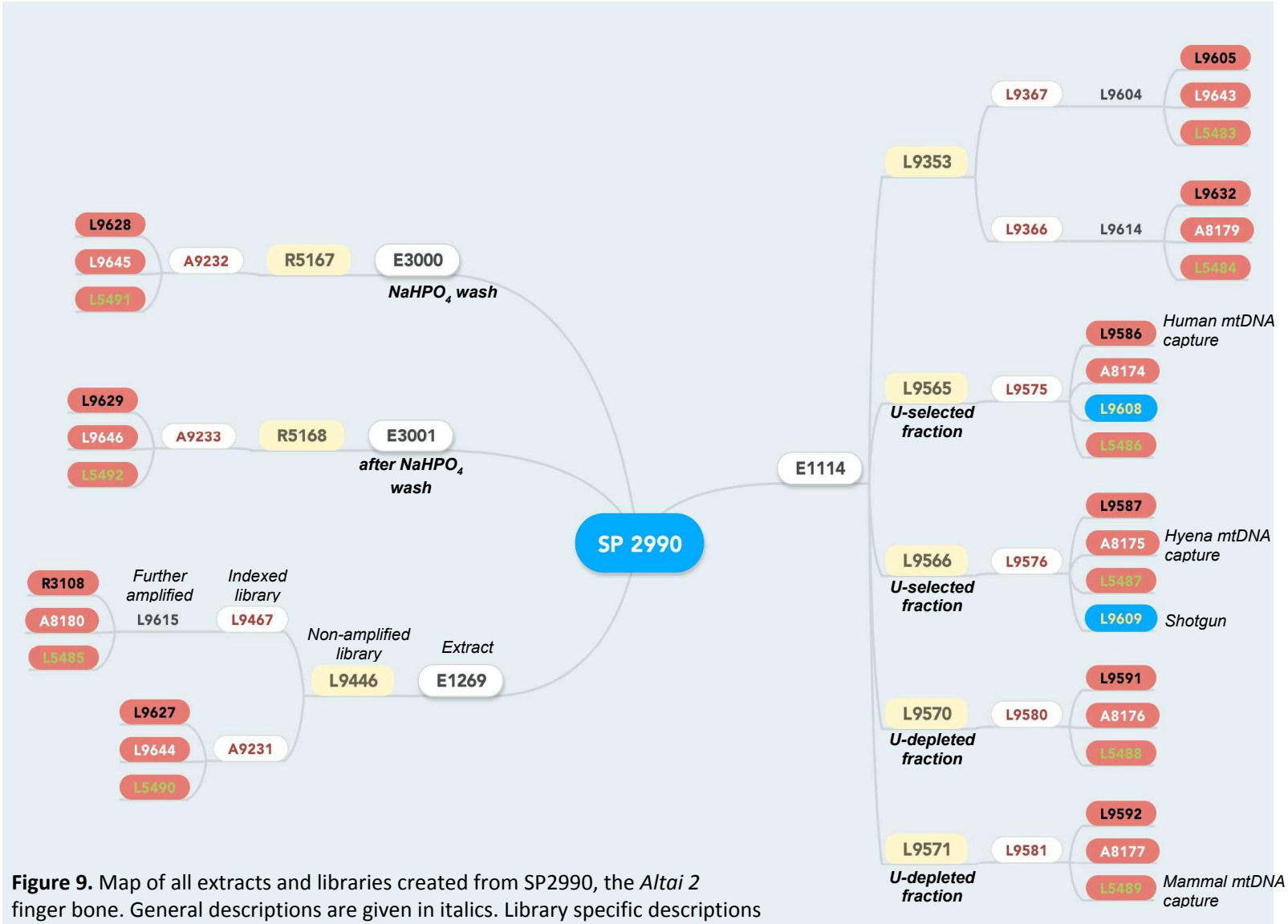

Figure 9. Map of all extracts and libraries created from SP2990, the *Altai 2* finger bone. General descriptions are given in italics. Library specific descriptions are given in bold italics. See Table 2 for more details.

Table 2. Extract and library names with descriptions for *Altai 2*. Input refers to the amount of the previous library/extract that went into the reaction to form the present library/extract. Output refers to the amount the present library was eluted in. Elution was done in either EBT (10mM Tris-HCL, 0.05% tween-20) or TET (1mM EDTA, 10mM tris-HCL, 0.05% tween-20). See Figure 9 for a map of how the libraries/extracts are related to each other.

Name	Description	Input	Output
E1114	Regular Extract	30 mg bone powder	100 uL
E1269	Regular Extract	23 mg bone powder	50 uL
E3000	Extract; $NaHPO_4$ wash	20 mg bone powder	50 uL
E3001	Extract; regular after $NaHPO_4$ wash	Same 20 mg as E3000	50 uL
A8174	Spotted spotted hyena mtDNA capture	1 ug	20 uL
A8175	Spotted spotted hyena mtDNA capture	1 ug	20 uL
A8177	Spotted spotted hyena mtDNA capture	1 ug	20 uL
A8178	Spotted spotted hyena mtDNA capture	1 ug	20 uL
A8179	Spotted spotted hyena mtDNA capture	1 ug	20 uL
A8180	Spotted spotted hyena mtDNA capture	1 ug	20 uL
A9231	Indexing PCR, 35 cycles	20 uL	30 uL
A9232	Indexing PCR, 35 cycles	50 uL	30 uL
A9233	Indexing PCR, 35 cycles	50 uL	30 uL
L9353	ssLIb prep, UDG treated	10 uL	40 uL
L9366	Indexing PCR, 8 cycles	20 uL	30 uL
L9367	Indexing PCR, 8 cycles	20 uL	30 uL
L9446	ssLIb prep, UDG treated	5 uL	40 uL
L9467	Indexing PCR, 9 cycles	20 uL	30 uL
L9486	Herculase amp, 10 cycles	5 uL	20 uL
L9521	Human mtDNA capture	1 ug	20 uL
L9565	ssLib prep, u-selection, u-selected fraction	15 uL	50 uL
L9566	ssLib prep, u-selection, u-selected fraction	15 uL	50 uL
L9570	ssLib prep, u-selection, u-depted fraction	15 uL	50 uL
L9571	ssLib prep, u-selection, u-depted fraction	15 uL	50 uL
L9575	Indexing PCR, 35 cycles	50 uL	40 uL
L9576	Indexing PCR, 35 cycles	50 uL	40 uL

L9580	Indexing PCR, 35 cycles	50 uL	40 uL
L9581	Indexing PCR, 35 cycles	50 uL	40 uL
L9586	Human mtDNA capture	2 ug	20 uL
L9587	Human mtDNA capture	2 ug	20 uL
L9591	Human mtDNA capture	2 ug	20 uL
L9592	Human mtDNA capture	2 ug	20 uL
L9604	Herculase amp, 15 cycles	3 uL	20 uL
L9605	Human mtDNA capture	1 ug	10 uL
L9608	Herculase amp, 1 cycle	1 uL	15 uL
L9609	Herculase amp, 1 cycle	1 uL	15 uL
L9614	Herculase amp, 20 cycles	3 uL	15 uL
L9615	Herculase amp, 20 cycles	3 uL	15 uL
L9627	Human mtDNA capture	1 ug	20 uL
L9628	Human mtDNA capture	1 ug	20 uL
L9629	Human mtDNA capture	1 ug	20 uL
L9632	Human mtDNA capture	1 ug	20 uL
L9643	Spotted spotted hyena mtDNA capture	1 ug	20 uL
L9644	Spotted spotted hyena mtDNA capture	1 ug	20 uL
L9645	Spotted spotted hyena mtDNA capture	1 ug	20 uL
L9646	Spotted spotted hyena mtDNA capture	1 ug	20 uL
L5483	All-mammalian mtDNA capture	1 ug	20 uL
L5484	All-mammalian mtDNA capture	1 ug	20 uL
L5485	All-mammalian mtDNA capture	1 ug	20 uL
L5486	All-mammalian mtDNA capture	1 ug	20 uL
L5487	All-mammalian mtDNA capture	1 ug	20 uL
L5488	All-mammalian mtDNA capture	1 ug	20 uL
L5489	All-mammalian mtDNA capture	1 ug	20 uL
L5490	All-mammalian mtDNA capture	1 ug	20 uL
L5491	All-mammalian mtDNA capture	1 ug	20 uL
L5492	All-mammalian mtDNA capture	1 ug	20 uL
R5167	ssLib prep, no UDG	10 uL	50 uL
R5168	ssLib prep, no UDG	10 uL	50 uL

4.2.2 MtDNA capture and sequencing

For the *Altai 2* finger bone, each of the ten indexed libraries, or final amplified offspring thereof, were captured with three probe sets: human mtDNA (rCRS used for design, NC012920 (76)), spotted hyena mtDNA (NC020670 used for design, (97)) and all-mammalian capture probes (242 mammals, (98)). See Figure 9 and Table 2 for more details of which libraries were captured. The capture method is based on the bead-based capture method described in Maricic *et al* (70) with modifications described in (99).

The human, spotted hyena and all-mammalian captures were sequenced on either the MiSeq or HiSeq Rapid platform for 76+7+76+7 cycles.

L9608 and L9609, the non-heteroduplex containing U-selected fractions, were pooled in equimolar amounts and sequenced together on a single rapid HiSeq lane for 76+7+76+7 cycles. The indexed libraries L9580, L9581, A9233, A9232, L9467, A9231, L9367 and L9366 were also sequenced on a Miseq for 76+7+76+7 cycles.

A2396, the amplified and indexed library from SP3388 (the spotted hyena bone from Denisova Cave) was captured using the same spotted hyena mtDNA probes as the *Altai 2* finger bone libraries. The captured library was pooled with other project-unrelated libraries and sequenced on a MiSeq for 76+7+76+7 cycles.

4.2.3 Sequence processing and mapping

Sequence processing:

After sequencing finished, basecalling, adapter trimming and index demultiplexing were performed. Basecalling for MiSeq runs was done with Bustard (Illumina Corp.), while for HiSeq runs freeIbis was used (100). For adapter trimming, Illumina adapters were removed and putative chimeric sequences were flagged as failing quality (leeHom option "—ancientdna, (101)). Sequences were demultiplexed by assigning them to their sample of origin using deML with default quality thresholds (102). Read-pairs were merged if they had an 11 bp overlap.

The following basic filters were used for all sequenced libraries:

Removal of:

1. sequences with a length less than 35 basepairs
2. sequences that do not match the index combinations used to produce each library
3. sequences with more than 5 bases with base qualities less than 15 (phred score)

4. sequences having a base with a quality less than 10 (phred score) in the index reads

Mapping:

L9629, L9628, L9521, L9627, L9605, L9632, L9586, L9587, L9591 and L9592 from the *Altai 2* finger bone were captured using human mtDNA probes (see Figure 9 and Table 2). Each of these libraries were first filtered as described above. They were then aligned to the human rCRS mitochondrial genome (76) as well as the Vindija 33.16 Neandertal mtDNA genome (59) using BWA version 0.5.10 (75) with relaxed parameters (-n 0.01, -o 2, -l 16500). After mapping, the following analyses were conducted for each library: (i) Percent in target: the number of sequences that mapped with a map quality over zero divided by the number of total sequences (including unmapped). Sequences were then filtered for a map quality 37 (phred score) and PCR duplicates were collapsed (collapsing sequences with the same beginning and end coordinates). (ii) Percent unique: the number of duplicate collapsed sequences were divided by the number of mapped sequences before duplicate removal. (iii) Average coverage across the mtDNA of duplicate collapsed sequences.

The ten *Altai 2* libraries captured with spotted hyena probes (L9643-L9646, A8174-A8177, A8179), as well as the SP3388 Denisova Cave spotted hyena library (L5497) were filtered using the basic filters described above. They were then aligned to the spotted hyena mtDNA (NC020670 (97)) using BWA version 0.5.10 (75) with relaxed parameters (-n 0.01, -o 2, -l 16500). After alignment, percent in target, percent unique and average coverage were calculated as for the human mtDNA captures. Thus, sequences with a map quality less than 37 (phred score) and PCR duplicates were removed. After these individual analyses and filters, the ten *Altai 2* libraries were merged into one.

The all-mammalian captures of the *Altai 2* (L5483 to L5492) were filtered using the basic filters described above and then aligned to the 242 mammalian genomes used to make the capture bait as described in Slon *et al* (98). After alignment, sequences were filtered for a map quality of 37 (phred scale) and PCR duplicates were collapsed. These sequences were then aligned to the Nucleotide database of NCBI using the Basic Local Alignment Search Tool (BLAST) and BLAST hits were ranked by taxonomic identification number (98).

4.2.4 Present-day human mtDNA contamination estimate

After alignment to the human rCRS, *Altai 2* sequences from each human mtDNA captured library were compared to 69 human-Neandertal diagnostic positions. These diagnostic positions are positions where ten Neandertals (73, 95, 103) differ from 311 present-day humans from around the world. Sequences that carry the Neandertal allele are deemed clean, while sequences that carry the present-day human allele are deemed contaminating. The analysis was repeated looking only at sequences containing thymine residues at the first and/or last two positions at sites where the rCRS sequence carries cytosine residues.

4.2.5 Spotted hyena contamination estimate

Spotted hyena diagnostic positions were determined as follows. First the spotted hyena (NC020670 (97)) mtDNA was chopped into 100 bp sequences with 1 bp tiling. These 100 bp sequences were then aligned to the human rCRS mtDNA (76) using BWA version 0.5.10 (75) with relaxed parameters (-n 0.01, -o 2, -l 16500). Sequences that mapped with a map quality above 0, were then used to make a consensus, requiring at least 1-fold coverage and 80% consensus support. Sequences mapped in regions 2,888-3,108 and 5,705-5,805 (in rCRS coordinates). The consensus was aligned to the same 311 humans used for the present-day human contamination estimate, as well as three spotted hyenas (97), one striped hyena (97) and one mongoose (NC006835) using MAFFT (81). The alignments were checked by eye in BioEdit v. 7.1.3 (104). In the two regions where the spotted hyena sequences mapped, a diagnostic position was called where the three spotted hyenas differed from all 311 humans. Thus nine diagnostic positions were called where there was a difference between the three spotted hyenas and the 311 humans, striped hyena and mongoose. Spotted hyena contamination was estimated based on these nine diagnostic positions as was done in section 4.2.4.

4.2.6 Assembly of Neandertal mtDNA

As significantly more sequences match the Neandertal state, the alignments to the Altai Neandertal mtDNA from the ten libraries were combined into one file. Due to the high level of contamination, only sequences containing thymine residues at the first and/or last two positions where the reference sequence has a C were used. A consensus was made of the sequences using a cutoff of 80% consensus support and 5-fold coverage.

In order to resolve as many positions as possible, we included two additional filters. We first required that in order for a C to T difference to be used for separating out a deaminated sequence, the minority of sequences at the positions needed to have a T. This aims to exclude the C to T difference existing due to a mutation instead of a deamination event. Second we took out sequences that align better to the spotted hyena (NC020670 (97)) than they do to the Neandertal.

After these filters, seven unresolved positions remain. One N remains in the C-stretch (position 310) due to the difficulty on mapping fragments that begin and end in the C-stretch. This region was resolved by calling a consensus of sequences that span the C-stretch region. Three positions fall into region 185-195 (rCRS coordinates). When examined by eye, two haplotypes are evident (see Table 3). After a megablast of both haplotypes, it became evident that both haplotypes are seen in present-day humans. The last three unresolved positions fall within the 16S and aspartate tRNA (Asn tRNA) genes. The two Ns that fall into the 16S gene (positions 2951 and 2964, rCRS coordinates) are close enough together that they combine to show two distinct haplotypes. After a megablast (105) of both haplotypes, haplotype one is seen in humans, while haplotype two is seen in individuals on the cat branch (cats, spotted hyenas, mongooses, civets) (Table 3). The last unresolved position (5767, rCRS coordinates), in the Asn tRNA gene. After a megablast (105) of sequences carrying both types of positions, one sequence occurs in humans, while the other occurs in members of the Cervid family, namely sheep, goats and deer. Thus the final mtDNA sequence for the *Altai 2* Neandertal has six missing positions.

Table 3. Unresolved positions in the Neandertal alignments on *Altai 2*. The two seqeunces at each position are shown. The unresolved position is shown in bold, while the number of sequences containing each position are shown in italics. The asterix (*) denotes a deletion.

Position	Coverage	Consensus support	Majority base	Base in rCRS	Base in Vindija 33.16	Base in CC8 Crocuta	Sequence 1	Sequence 2
185	24	58.3	G	G	G	A	**G**CGA**A**CATAC**T** present-day human *4 sequences*	**A**CGA**G**CATAC**C** Neandertal and present-day human *7 sequences*
189	26	57.7	G	A	G	A		
195	19	78.9	C	T	T	G		
2951 (16S)	152	74.3	C	C	C	T	**C**TAGAGTCCATATC**A** human, CC8 crocuta *82 sequences*	**T**TAGAGTCCATATC**G** Cat branch, not in CC8 crocuta *28 sequences*
2964 (16S)	152	63.8	A	A	A	A		
5767 (12S)	157	68.2	*	*	*	C	GGCAG*****GTTTG human *84 sequences*	GGCAG**A**GTTTG Deer, sheep, goat *69 sequences*

4.2.7 Assembly of spotted hyena mtDNA

After combining the ten *Altai 2* libraries captured and aligned to the spotted hyena, coverage and consensus support across the spotted hyena mtDNA genome were plotted. The coverage spikes by almost two-fold in areas with high conservation in the mtDNA genome (especially the 16S region, Figure 24A), due to the large amount of hominid DNA present in the bone. To be sure that we reconstruct a spotted hyena mtDNA with no influence from the human sequences, we removed regions that mapped to human mtDNA. This was done by chopping the human rCRS mtDNA (76) into 35-100 bp sequences in increments of 5 bps with 1 bp tiling. Then these sequences were mapped to the spotted hyena mtDNA (NC020670 (97)) using BWA version 0.5.10 (75) with relaxed parameters (-n 0.01, -o 2, -l 16500). Regions where the human sequences mapped to the spotted hyena were removed. After human regions were removed, the extreme coverage peaks disappear (Figure 24B). A consensus was called of the sequences that cover the non-human-mapping regions, by requiring at least 5-fold coverage and a consensus support of 80%.

4.2.8 Spotted hyena mtDNA in Denisova Cave specimins

The unmapped sequences of the high coverage Altai Neandertal (*Altai 1*) (10), high coverage Denisovan (*Denisova 3*) (8), two low-coverage Denisovans (*Denisova 4* and *8*), a high coverage early modern human (Ust-Ishim) (106), a low coverage Neandertal (*Mezmaiskaya1*) (10) and a present-day human were filtered as in section 4.2.3 and then mapped to thespotted hyena (NC020670 (97)), the ringed seal (NC008428, (107)), the cave bear (NC011112, (108)) and the rCRS human (NC012920, (76)) mtDNA using BWA version 0.5.10 (75) with relaxed parameters (-n 0.01, -o 2, -l 16500). After mapping, duplicates were collapsed and sequences with a map quality less than 37 (phred scale) were removed.

Regions of the seal and cave bear mtDNA that do not map human mtDNA were determined as the non-human mapping spotted hyena regions were calculated in section 4.2.7. Sequences that fall completely within these non-human mapping regions were kept for further analyses.

For the *Denisova 3* and 4 individuals as well as the *Altai 1* Neandertal, a consensus of the filtered sequences was called requiring 80% consensus support and at least 3-fold coverage.

The number of sequences that mapped perfectly (no insertions, deletions or mutations) were also calculated. A two-tailed fisher exact test was done in R v3.2.0 (fisher.test, (77)) between each of two individuals and two of the mapped mtDNAs (e.g. present-day human seal and hyena sequences mapped versus *Denisova 4* seal and hyena sequences mapped).

4.2.9 MtDNA phylogenetics (human and spotted hyena)

Human

MtDNAs of seven published Neandertals (10, 73, 103), five present-day humans (San – AF347008, Yoruba – AF347014, Han Chinese – AF346972, French – AF346981 and Papuan – AF347004) (78), one Denisovan (7) and a chimpanzee (X93335.1) (79) were aligned to the *Altai 2* Neandertal consensus sequence using MAFFT (80, 81). A Modeltest was done as described in section 4.1.9, best model: GTR+I+G. A Bayesian tree was then produced using BEAST (86, 109) with the following parameters: GTR+I+G model, uncorrelated log normal clock, set to 2.7e-8 with a standard deviation of 1e-8; a Bayesian skyline tree prior, initial 1000, distribution 0 to 100000; run for 1000000000 generations and sample every 1000. A pairwise comparison of the same mtDNAs was done using Mega6 (82). The number of differences were also calculated to the most common recent ancestor of Neandertals as calculated in section 4.1.10.

Spotted hyena

The complete mtDNAs of three spotted hyenas (NC020670, JF894379 and JF894377, (97)) a striped hyena (NC020669, (97)) and a mongoose (NC006835) were aligned to the consensus sequences of the SP3388 Denisova Cave spotted hyena and the *Altai 2* finger bone using MAFFT as above. A Modeltest was also done as above, result: GTR+I. A Bayesian tree run with BEAST with the same parameters as above except the GTR+I model was used. A pairwise comparison of the same mtDNAs was done using Mega6 (82).

The consensus sequences of the SP338 Denisova Cave spotted hyena and the *Altai 2* finger bone were also aligned to the cytochrome b sequences of 55 spotted hyenas (see Appendix Table 1 for accession numbers and references), two Aardwolves (AY928679 and AY048791, (110)), six brown hyenas (DQ157588-591, AY048789-90, (13, 111)), 18 striped hyenas (DQ157576-DQ157587, AY048788,AY048787,AY928678,AF153055,AF153054,EF107524,NC020669, (13, 97, 110-112)) and one mongoose (same as above) using MAFFT as above. A pairwise

comparison of these mtDNAs was done using Mega6 (82). Based on the pairwise comparison, individuals with the same sequences were collapsed and given letters (see Table 27). For the phylogenetic analysis with BEAST, only the sequences groups were used. Modeltest was run as above, best model: TrN+I. BEAST was run as above but using the TN93 model, which is closest to the TrN+I model suggested.

Spotted hyena in high coverage archaics

The complete mtDNAs of the same three spotted hyenas, striped hyena and mongoose as above were aligned separately to the spotted hyena consensus sequences of the *Denisova 3*, *Denisova 4* and *Altai 1* individuals using MAFFT as above. Modeltest and BEAST were again run as described above. Modeltest suggested the GTR+I model for the alignments including *Denisova 3*, the HKY+I model for alignments including *Denisova 4*, and the TVM+I model for alignments using *Altai 1*. BEAST was run as above, and the TN93+I model was used for the *Altai 1* alignments.

4.2.10 Autosomal divergence calculation

All of the shotgun libraries were filtered using the basic filters described in section 4.2.3 and then aligned to the human genome (74) using BWA version 0.5.10 (75) with relaxed parameters (-n 0.01, -o 2, -l 16500). Each library was then filtered for mapped sequences (sequences with a map quality over 0). The percent endogenous for each library was calculated by dividing the number of mapped sequences by the number of unmapped sequences.

After mapping, the sequences from L9608 and L9609 had PCR duplicates collapsed, were combined and sequences with a map quality less than 37 (phred score) were removed. In addition sequences not containing thymine residues at the first and/or last two positions were removed as described in section 4.1.12. This left 18.4 Mb. Divergence was calculated as described in section 4.1.13 for the *Altai 2* to ten present-day humans, the high coverage *Altai 1* and *Denisova 3* (Figure 22). Divergence of the *Altai 2* to the high coverage *Altai 1*, was also compared to divergences of six published Neandertals to the high coverage Altai Neandertal (Figure 23), using the same values used in section 4.1.13.

Lineage attribution was calculated as described in Meyer *et al* (113). In brief, *Denisova 3*, *Altai 1* and an Mbuti present-day human were used to determine positions that are ancestral or derived when compared to the human-chimpanzee ancestor. Then the *Altai 2* sequences were compared to these positions and counted.

5. Nuclear and mitochondrial DNA sequences from two Denisovan individuals

This manuscript is published at the Proceedings of the National Academy of Sciences of the United States of America, doi: 10.1073/pnas.1519905112

Authors: Susanna Sawyer[a], Gabriel Renaud[a], Bence Viola [b,c,d], Jean-Jacques Hublin [c], Marie-Theres Gansauge [a], Michail V. Shunkov [d,e], Anatoly P. Derevianko [d,f], Kay Prüfer [a], Janet Kelso [a], Svante Pääbo [a].

[a] Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany

[b] Department of Anthropology, University of Toronto, Toronto, ON M5S 2S2, Canada

[c] Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany

[d] Institute of Archaeology and Ethnography, Russian Academy of Sciences, Siberian Branch, Lavrentieva Avenue, 17 Novosibirsk, RU-630090, Russia.

[e] Novosibirsk National Research State University, Pirogova st., 2, Novosibirsk, RU-630090, Russia

[f] Altai State University, Pr. Lenina, 61, Barnaul, RU-656049, Russia

Author contributions:

Contributed materials: Michail Shunkov, Anatoly Derevianko

Performed experiments: Susanna Sawyer and Marie-Theres Gansauge

Analyzed morphological data: Bence Viola

Analyzed mitochondrial data: Susanna Sawyer

Analyzed nuclear data: Susanna Sawyer, Gabriel Renaud, Kay Prüfer, Janet Kelso

Wrote the paper: Susanna Sawyer, Bence Viola and Svante Pääbo

5.1 Abstract

Denisovans, a sister-group of Neandertals, have been described based on a nuclear genome sequence from a finger phalanx (*Denisova 3*) found in Denisova Cave in the Altai Mountains. The only other Denisovan specimen described to date is a molar (*Denisova 4)* found at the same site. This tooth carries a mitochondrial (mt) DNA sequence similar to that of *Denisova 3*. Here we present nuclear DNA sequences from *Denisova 4* and a morphological description, as well as mitochondrial and nuclear DNA sequence data, from another molar (*Denisova 8)* found in Denisova Cave in 2010. This new molar is similar to *Denisova 4* in being very large and lacking traits typical of Neandertals and modern humans. Nuclear DNA sequences from the two molars form a clade with *Denisova 3*. The mtDNA of *Denisova 8* is more diverged and has accumulated fewer substitutions than the mtDNAs of the other two specimens suggesting that Denisovans were present in the cave over an extended period of time. The nuclear DNA sequence diversity among the three Denisovans is comparable to that among six Neandertals but lower than that among present-day humans.

5.2 Introduction

In 2008, a finger phalanx from a child (*Denisova 3*) was found in Denisova Cave in the Altai Mountains in southern Siberia. The mitochondrial (mt) genome shared a common ancestor with present-day human and Neandertal mtDNAs about one million years ago (7), *i.e*. about twice as long ago as the shared ancestor of present-day human and Neandertal mtDNAs. However, the nuclear genome revealed that this individual belonged to a sister group of Neandertals. This group was named Denisovans after the site where the bone was discovered (8, 9). Analysis of the Denisovan genome showed that Denisovans have contributed on the order of five percent of the DNA to the genomes of present-day people in Oceania (8, 9, 60) and about 0.2 percent to the genomes of Native Americans and mainland Asians (10).

In 2010, continued archaeological work in Denisova Cave resulted in the discovery of a toe phalanx (*Denisova 5*), identified based on its genome sequence as Neandertal. The genome sequence allowed detailed analyses of the relationship of Denisovans and Neandertals to each other and to present-day humans. Although divergence times in terms of calendar years are unsure due to uncertainty about the human mutation rate (114), it showed that Denisovan and Neandertal populations split from each other in the order of four times further back in time than the deepest divergence among present-day human populations occurred, while the ancestors of the two archaic groups split from the ancestors of present-day humans in the order

of six times as long ago as present-day populations (10). In addition, a minimum of 0.5 percent of the genome of the *Denisova 3* individual was derived from a Neandertal population more closely related to the Neandertal from Denisova Cave than to Neandertals from more western locations (10).

Although Denisovan remains have, to date, only been recognized in Denisova Cave, the fact that Denisovans contributed DNA to the ancestors of present-day populations across Asia and Oceania suggests that, in addition to the Altai Mountains, they may have lived in other parts of Asia. Besides the finger phalanx, a molar (*Denisova 4*) was found in the cave in 2000. Although less than 0.2% of the DNA in the tooth derive from a hominin source, the mtDNA was sequenced and differed from the finger phalanx mtDNA at only two positions suggesting that it too may be from a Denisovan (8, 9). This molar has several primitive morphological traits different from both late Neandertals and modern humans. In 2010, another molar (*Denisova 8*) was found in Denisova Cave. Here we describe the morphology and mtDNA of *Denisova 8* and present nuclear DNA sequences from both molars.

5.3 Results

5.3.1 *Denisova 8* morphology

Denisova 8. The *Denisova 8* molar (Figure 10) was found at the interface between layers 11.4 and 12 in the East Gallery of Denisova Cave, slightly below the Neandertal toe phalanx (*Denisova 5*, Layer 11.4) and the Denisovan finger (*Denisova 3*, Layer 11.2). Radiocarbon dates for layer 11.2 as well as for the underlying 11.3 layer yield ages over ~50,000 years (OxA-V-2359-16 & -14) (2). *Denisova 8* is thus older than *Denisova 3* which is at least 50,000 years old. It is reassembled from four fragments which fit well together, although a piece of enamel and most of the root is missing (Figure 11B).

The *Denisova 4* molar was found in Layer 11.1 in the South Gallery, a different part of the cave. Radiocarbon dates for layer 11.2 of the South Gallery are over 50,000 years (OxA-V-2359-17 & -18) and 48.6 $\pm$ 2.3 thousand years before present (KIA 25285) (2). Although the lack of direct stratigraphic connection between the different parts of the cave makes relative ages difficult to assess it is likely that *Denisova 4* is younger than *Denisova 8.*

Based on crown shape and the presence of a marked *Crista obliqua*, a feature unique to maxillary molars, we identify *Denisova 8* as an upper molar, despite having five major cusps. The mesial half of the crown is worn, with a small dentine exposure on the protocone, while

there is no wear on the distal part. The lack of a distal interproximal facet indicates that the tooth is a third molar, or a second molar without the eruption of the M^3. Usually, when Neandertal and *H. heidelbergensis* upper M^2s reach wear levels to the extent seen here, the adjacent M^3 is already erupted and an interproximal facet is visible. One possibility is that the *Denisova 8* is a second molar of an individual with M^3 agenesis. Despite being common in modern humans, this is rare in archaic hominins, but it does occur in Asian late *Homo erectus* and Middle Pleistocene hominins.

The previously described *Denisova 4* molar is characterized by its large size, flaring buccal and lingual sides, strong distal tapering and massive and strongly diverging roots (2). Not all of these characteristics can be assessed in *Denisova 8*, but it is clear that it lacks the strong flare of the lingual and buccal surfaces and distal tapering of *Denisova 4*.

The length of *Denisova 8* is more than three standard deviations larger than the means of Neandertal and modern human molars and in the range of Pliocene hominins (Figure 10 and 11). Both *Denisova 8* and *4* are very large compared to Neandertal and early modern human molars, and *Denisova 8* is even larger than *Denisova 4*. Only two Late Pleistocene third molars are comparable in size, those of the inferred early Upper Paleolithic modern human *Oase 2* in Romania and *Obi-Rakhmat 1* in Uzbekistan (42, 115).

The morphology of third molars is variable, and thus not very diagnostic. Nevertheless, Neandertal third molars differ from *Denisova 8* in that they frequently show a reduction or absence of the hypocone, reduction of the metacone and generally lack a continuous *Crista obliqua* (115, 116). This applies also to Middle Pleistocene European hominins which also only rarely show a Cusp 5 (116). The massive and diverging roots of *Denisova 4* are very unlike the root morphology of Neandertals and Middle Pleistocene hominins in Europe. East Asian *Homo erectus* and Middle Pleistocene *Homo* frequently show massive roots similar to *Denisova 4*, but in these groups crown size become strongly reduced starting around one million years ago (117). The recently described Xujiayao teeth from China (118) have massively flaring roots and relatively large and complex crowns, similar to the Denisova teeth, but have reduced hypocones and metacones.

Early and recent modern humans show the most morphological variability of third molars, and there are specimens that have large hypocones, metacones or continuous *Cristae obliquae* (116). The combination of an unreduced metacone and hypocone, continuous *Crista obliqua*, a large fifth cusp, and large over-all size is reminiscent of earlier *Homo*, but *Denisova 8* lacks the multiple distal accessory cusps frequently seen in early *Homo* and Australopithecines.

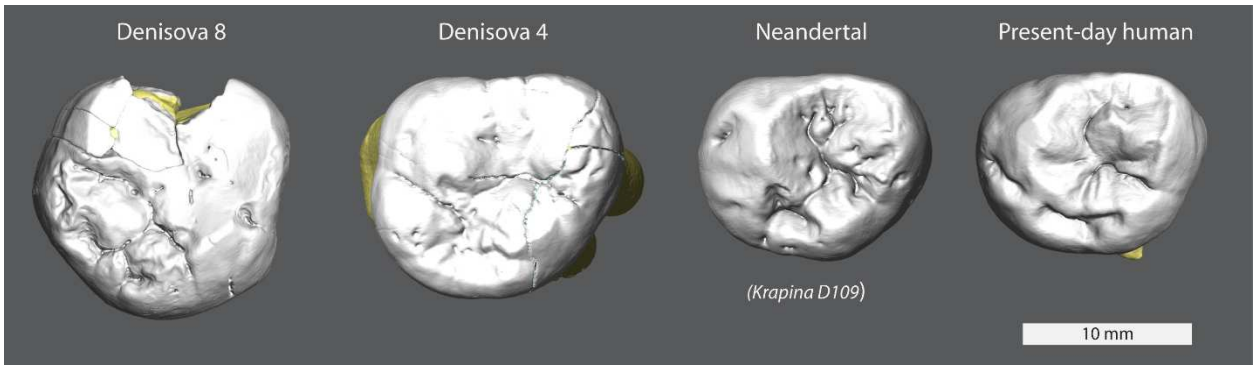

Figure 10. Occlusal surfaces of the *Denisova 4* and *8* molars and third molars of a Neandertal and a present-day European.

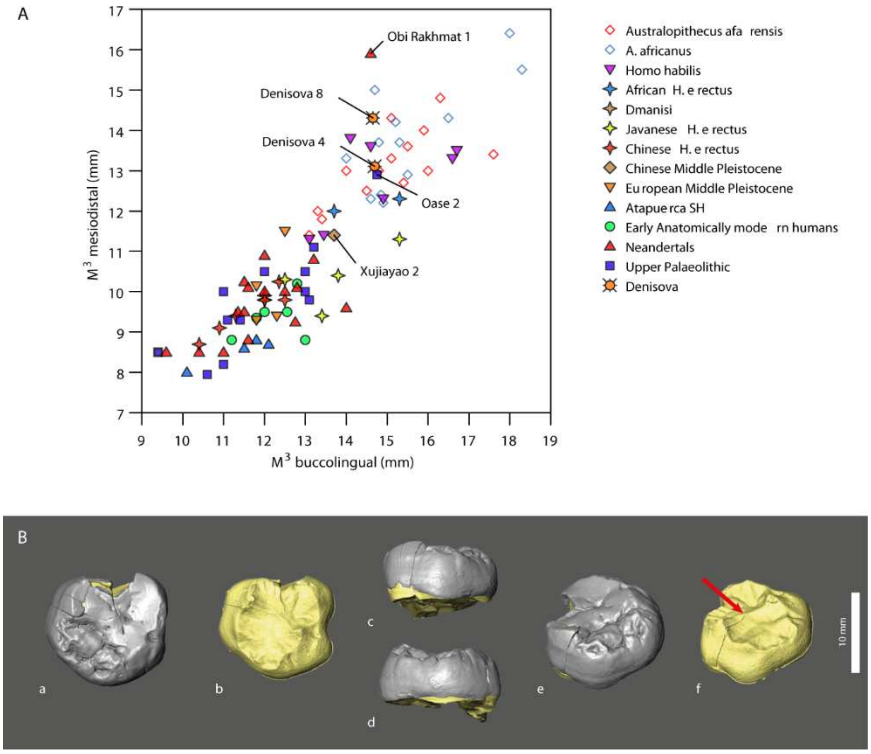

Figure 11. **Morphology of *Denisova 8* molar**. . **a:** occlusal view (surface model from µCT scan); **b:** enamel dentine junction in occlusal view, The arrow indicates the marked *Crista obliqua* on the enamel-dentine junction; **c**: Biplot of the mesiodistal (md) and labiolingual (bl) diameters of *Denisova 8* and other hominin M^3s. For comparative sample used and sources for data see Table 4. **d:** Biplot of the mesiodistal (md) and labiolingual (bl) diameters of *Denisova 8* and other hominin M^2s. For comparative sample used and sources for data see Table 4.

Table 4. Metric comparisons of M^2 and M^3 length and breadth in various fossil hominins and the Denisova remains.

	M^2 md[1]	M^2 bl[2]	M^3 md	M^3 bl
A. afarensis	13.7±1.4 (13)[3]	14.7±0.9 (13)	13.1±1 (14)	15±1.3 (14)
A. africanus	13.9±1 (12)	15.3±1.1 (12)	13.8±1.3 (12)	15.6±1.4 (12)
Homo habilis	12.6±0.6 (6)	14±1.1 (6)	12.7±1.1 (7)	14.8±1.4 (7)
Dmanisi	12.3 (12.05-12.5; 2)[4]	12.7 (12.1-13.2; 2)	9.8 (1)	12 (1)
***H. erectus* (Africa)**	12.7 (11.7-13.7; 4)	13.5 (12.15-14.7; 4)	12.2 (12-12.3; 2)	14.5 (13.7-15.3; 2)
***H. erectus* (Indonesia)**	12.3 (11.2-13.6; 3)	14 (12.8-15.4; 3)	10.4 (9.4-11.3; 4)	13.8 (12.5-15.3; 4)
***H. erectus* (China)**	11.3±0.9 (8)	13.2±1.1 (8)	9.6±0.5 (7)	11.6±0.8 (7)
Atapuerca SH	10.6±0.7 (6)	12.9±0.9 (6)	8.5±0.4 (4)	11.4±0.9 (4)
***H. heidelbergensis* (Europe)**	11.6 (11.4-12.1; 4)	12.7 (11.9-13.7; 4)	10.1 (9.3-11.5; 4)	12.1 (11.8-12.5; 4)
Neandertals	11±1.4 (21)	12.7±1.2 (21)	10.1±1.8 (17)	12±1.3 (17)
Neandertals (w/o Obi-Rakhmat)	10.7±0.8 (20)	12.6±1.1 (20)	9.8±1 (16)	11.8±1.1 (16)
Early AMH	10.8±1.2 (10)	12.7±1.1 (10)	9.4±0.5 (6)	12.2±0.7 (6)
Upper Palaeolithic	10.4±1 (21)	12.3±1.2 (21)	9.8±1.4 (12)	12±1.5 (12)
Denisova 4	13.1	14.7	13.1	14.7
Denisova 8	-	-	14.3	14.65

1. Mesiodistal length measured following the definition of (119)
2. Buccoligual breadth measured following the definition of (119)
3. Mean+-standard deviation (N)
4. Mean (range; N)

Sources of metric data:

A. afarensis: Hadar, Omo (own measurements)
A. africanus: Stekfontein, Makapansgat (120)
Homo habilis: Olduvai (121), East Turkana (120)
Dmanisi (122)
H. erectus (Africa): East Turkana (120), Nariokotome (123), Konso (124), Swartkrans (120)
H. erectus (China): Zhoukoudian (125), Hexian (126)
H. erectus (Indonesia): Trinil (120), Sangiran (own measurements, (127))
Atapuerca SH (116)
H. heidelbergensis (Europe): La Chaise (128), Biache (129), Arago (130), Petralona (128)
Neandertals: Amud (131), Châteauneuf (132), St. Brelade (119), Krapina (133), La Croze de Dua (119), La Quina (119), Le Moustier (119), Obi-Rakhmat (own measurements), Saccopastore (119), Shanidar (134), Spy (119), Tabun (119), Vergisson la Falaise (119)
Early AMH: Skhul (135), Qafzeh (136), Temara (137)
Upper Paleolithic: Brno (119), Changwu (126), Cro-Magnon (119), Dolni Vestonice (138), Grotte des Enfants (119), Kostenki (own measurements), La Rochette (119), Leuca (119), Mladec (119), Oase (139), Predmosti (119), Sungir (own measurements)

5.3.2 DNA isolation and sequencing

DNA was extracted from 36 mg of dentine from *Denisova 8* in our clean room facility (67) and DNA libraries from this specimen as well as from a previously prepared extract of *Denisova 4* were prepared as described (8, 61) (see Table 1). From both teeth, random DNA fragments were sequenced and mapped to the human reference genome (hg19). In addition, mtDNA fragments were isolated from the libraries (70) and sequenced.

Of the DNA fragments sequenced from *Denisova 4* and *Denisova 8* 0.05% and 0.9%, respectively, could be confidently mapped to the human genome sequence, yielding 54.6 and 265 million base pairs (Mb) of nuclear DNA sequences for *Denisova 4* and *Denisova 8*, respectively (see Table 9 for overview). MtDNA sequences from the two specimens were aligned to the mtDNA of *Denisova3* (NC_013993.1). For *Denisova 4,* the average mtDNA coverage is 72.1-fold. The lowest support for the majority base at any position is 89% (Figure 12) and the consensus sequence is identical to the previously published mtDNA sequence from this specimen (2). For *Denisova 8,* the mtDNA coverage is 118.9-fold and the lowest support for the majority base is 86% (Figure 12).

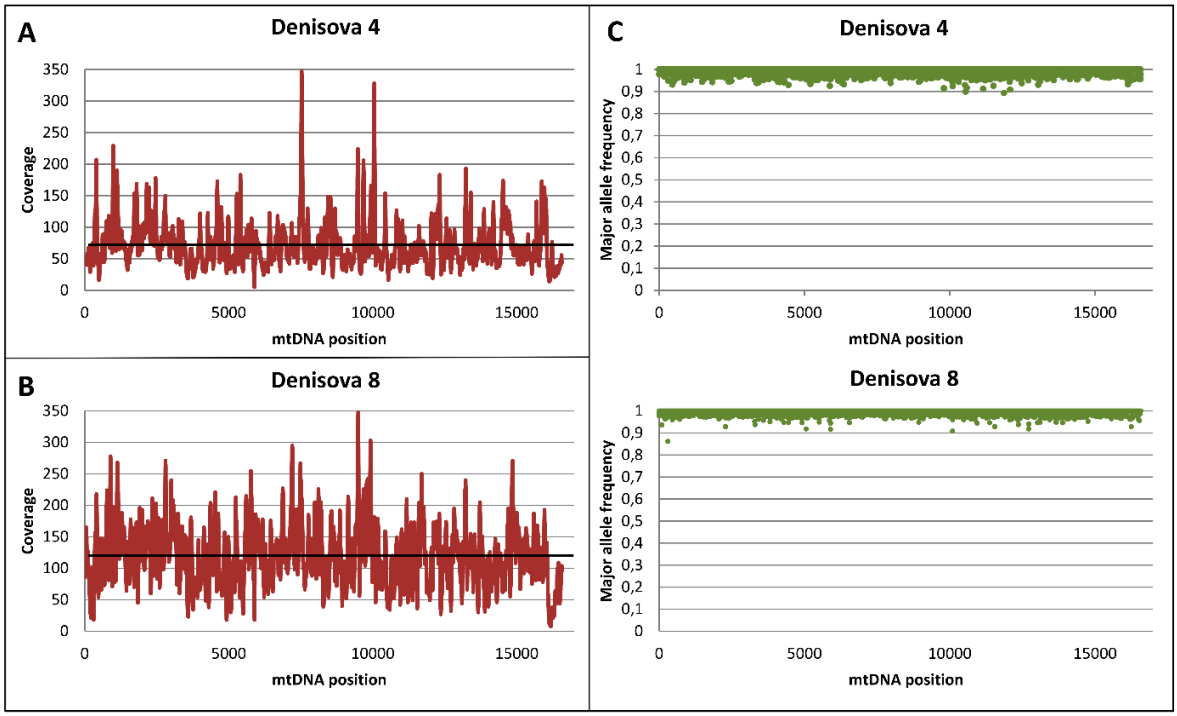

Figure 12. Quality of mtDNA sequences from *Denisova 4* and *8*. A, B: Coverage across the mitochondrial genomes. Black lines denote the average coverage. **C, D**: Consensus support across the genomes.

5.3.3 DNA sequence authenticity

We used three approaches to estimate present-day human DNA contamination in the two libraries. First, for each library, we used all unique DNA fragments that aligned to the present-day human reference mtDNA (76) and counted as contaminating those that carried a nucleotide different to the majority mtDNA sequence determined from the molar at positions where the endogenous majority consensus differed from all of 311 present-day human mtDNAs. The mtDNA contamination thus estimated was 5.2% (95% confidence interval (CI): 4.5-6.0%) for *Denisova 4* and 3.2% (95% CI: 2.9-3.6%) for *Denisova 8*.

Second, we estimate contamination by present-day nuclear DNA by estimating DNA sequence divergence (as described below and in Figure 17A) of the two molars to present-day humans. We assume that the divergence of two present-day European individuals to each other represent 100% contamination while the divergence of the high quality genome determined from

Denisova 3 to present-day humans represents zero percent contamination. By this approach, we estimate the *Denisova 4* DNA contamination of *Denisova 4* to 65.2-67.0% and *Denisova 8* to 14.6-15.4% (Table 5). That the nuclear DNA contamination is high, particularly of *Denisova 4,* is compatible with an estimate based on cytosine deamination patterns at the 3'- and 5'- ends of the aligned sequences (Supplementary methods).

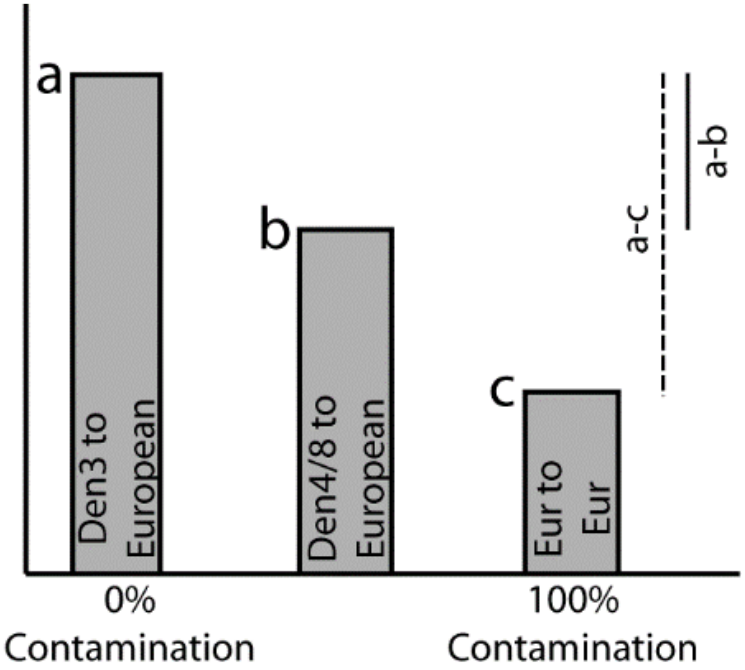

Figure 13. Divergence-based contamination estimates. The divergence of the *Denisovan 3* to two French and two Sardinians (left bar, *a*) is assumed to represent 0 % present-day human contamination. The divergence of French-French and Sardinian-Sardinian (right bar, *c*) is assumed to represent 100 % contamination. The divergence of *Denisova 4* or *8* to the French and Sardinians (middle bar, *b*) is then gauged as the reduction in divergence to the present-day humans as a fraction of the divergence among the present-day humans ((*a*-*b*) / (*a*-*c*)).

Table 5. Nuclear contamination estimate. An estimate of the nuclear contamination using the method described in Figure 13 applied to fragments without filtering for deamination.

European used to calc div[a]	% Divergence European[b]	% Divergence *Denisova 3*[c]	% Divergence *Denisova 4*	% Divergence *Denisova 8*	Div Den3 – Div human[d]	Div Den3 – Div *Den4*	Div Den3 – Div *Den8*	% contamination Den4[e]	% contamination Den8
French1	6.36 *(to Fr2)*	11.85	8.22	11.02	5.49	3.63	0.83	**66.1**	**15.1**
French2	6.09 *(to Fr1)*	11.62	7.98	10.81	5.53	3.64	0.81	**65.8**	**14.6**
Sardinian1	6.34 *(to Sa2)*	11.86	8.26	11.05	5.52	3.6	0.81	**65.2**	**14.7**
Sardinian2	6.06 *(to Sa1)*	11.64	7.9	10.78	5.58	3.74	0.86	**67.0**	**15.4**

a. The European present-day humans to whom divergence is calculated and whose mutations are used to calculate divergence
b. Divergence calculation using pairs of Europeans. Thus: French2 to French 1, and vice versa, as well as Sardinian2 to Sardinian1 and vice versa. As an example French2 to French1 uses the mutations on the branch to French1 to calculate the divergence and gives a result of 6.36%.
c. Divergence of *Denisova 3* to each of the European present-day humans listed.
d. Differences in divergence, calculated e.g. divergence of Den3 to French1 minus the divergence of French2 to French1 (in this case *a* – *c* in Figure 13).
e. Percent contamination, calculated e.g. (divergence of Den3 to Fr1 – divergence of Den8 to Fr1) / (divergence of Den3 to Fr1 – divergence of Fr2 to Fr1)*100. In this case this would be (a-b)/(a-c)*100 in Figure 13.

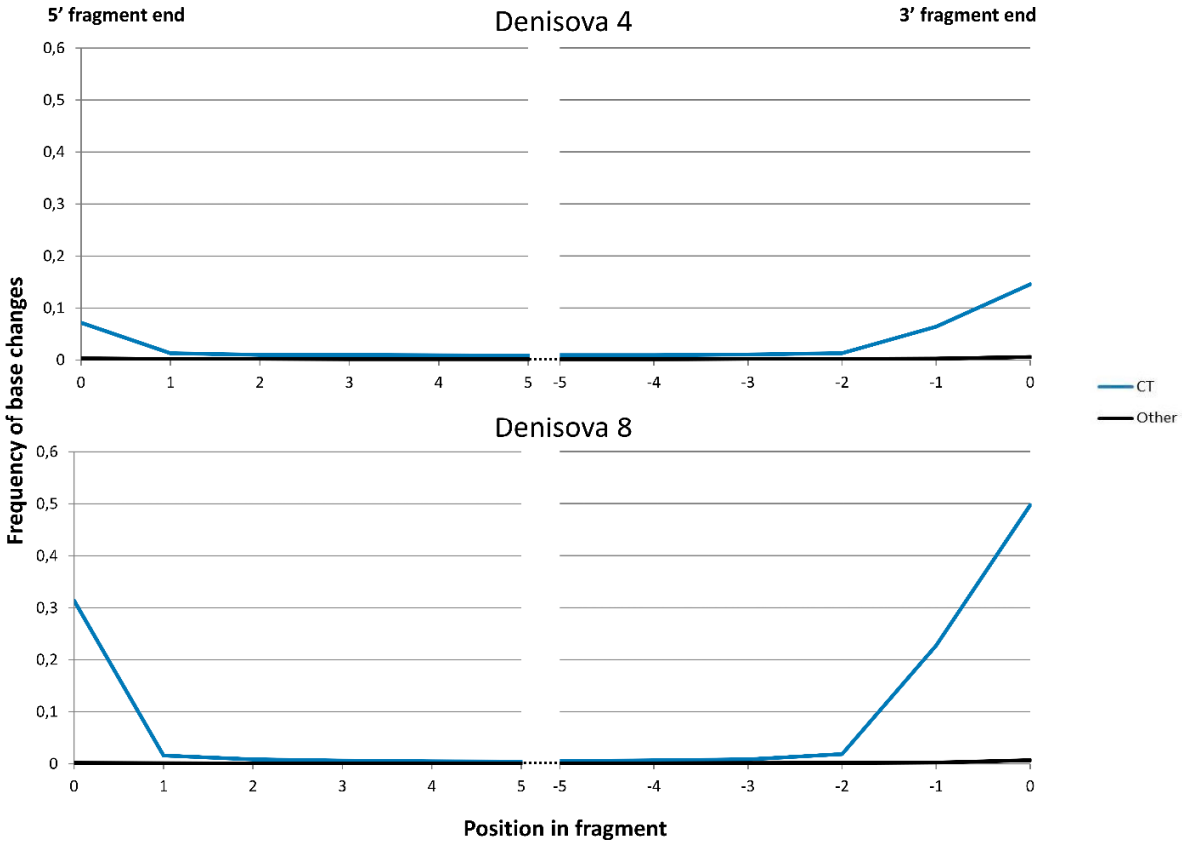

Figure 14. Nucleotide differences to the human reference genome as a function of distance from fragment ends. Differences are given as percent of a base in the reference genome that occurs as a different base in the sequenced DNA fragments. C to T differences are largely due to deamination of cytosine residues in ancient DNA fragments. Libraries were treated with *E.coli* uracil DNA glycosylase, which is not efficient at the first, the last and second to last bases.

Table 6. Terminal C to T substitutions nuclear and mtDNA fragments. C to T substitutions relative to the corresponding mtDNA consensus sequences are shown for mtDNA and nuclear DNA fragments sequenced from *Denisova 4* and *Denisova 8*, respectively. "3' filtered" and "5' filtered" refer to fragments that carry C to T substitutions at their 3'- and 5'-ends, respectively. The 95% CI is given in parenthesis.

		5 prime	3 prime
***Denisova 4* mtDNA**	***No filter***	11.3 *(9.7-13.0)*	22.4 *(20.9-24.1)*
	3' filtered	17 *(9.7-27.8)*	100
	5' filtered	100	30.5 *(22.2-40.4)*
***Denisova 4* nuclear**	***No filter***	7.2 *(6.9-7.4)*	14.6 *(14.3-14.8)*
	3' filtered	18.9 *(16.0-22.2)*	100
	5' filtered	100	35.7 *(32.6-39.1)*
***Denisova 8* mtDNA**	***No filter***	23.7 *(21.9-25.6)*	46.0 *(44.5-47.6)*
	3' filtered	20.8 *(16.2-26.3)*	100
	5' filtered	100	46.9 *(39.9-54.2)*
***Denisova 8* nuclear**	***No filter***	31.4 *(31.2-31.6)*	49.8 *(49.7-49.9)*
	3' filtered	32.5 *(32.0-33.2)*	100
	5' filtered	100	52.3 *(51.8-52.8)*

In the third approach, we first determined the sex of the individuals from which the molars derive by counting the numbers of DNA fragments that map to the X chromosome and autosomes, respectively. To limit the influence of present-day DNA contamination in this part of the analysis, we restrict to DNA fragments that at their 5'- and/or 3'-ends carry thymines (T) at positions where the human reference nuclear genome carries cytosines (C). Such apparent C to T substitutions are frequently caused by deamination of cytosine to uracil towards the ends of ancient DNA fragments (55, 56). We find that both teeth come from males ($p\sim0.4$) rather than females ($p<<0.01$) (Table 7). We then estimated the amount of female DNA contamination among the aligned sequences as the fraction of DNA fragments that match the X chromosome in excess of what is expected for a male bone. This yields a female DNA contamination rate of 28.4% (95 CI: 27.3-29.5%) for *Denisova 4* and 8.6% (95 CI: 8.3-8.9%) for *Denisova 8*.

Table 7. Sex determination and female contamination. The number of X- and Y-chromosomal sequences mapped and expected to map if the molars are from males. DNA sequences carrying terminal C to T substations as well as all sequences were analyzed.

		Y-chromosome			X-chromosome			
Denisova	**Analysis/ Sequences**	# of sequences mapped	# sequences expected to map if male	χ^2*-test p-value*	# of sequences mapped	# sequences expected to map if male	χ^2*-test p-value*	Percent female contamination
4	**Sex determination (Terminal C->T seqs)**	8	3	-	231	222	*0.42 (5.9e-14 if female)*	-
8		94	86	*0.26*	5,535	5,576	*0.43 (<2.2e-16 if female)*	-
4	**Contamination estimate (all seqs)**	75	93	*0.006*	7,764	6,048	*<2.2e-16*	28.4% *(27.3-29.5)*
8		617	599	*0.32*	42,175	38,829	*<2.2e-16*	8.6% *(8.3-8.9)*

The estimates based on mtDNA and nuclear DNA differ drastically (Table 8) presumably because the ratios of mitochondrial to nuclear DNA differ between the endogenous and the contaminating source(s) of DNA while the two estimates based on nuclear DNA suggest that more males than females are among the contaminating individuals. It is clear that although these methods yield different contamination estimates, they all suggest that the nuclear DNA contamination in both libraries is substantial, particularly in *Denisova 4* where it is likely to exceed 50%. To reduce the influence of DNA contamination (87, 103) we therefore restrict the analyses of nuclear DNA to fragments that carry thymine residues at the first and/or last two positions at sites where the human reference sequence carries cytosine residues (but remove these C/T sites themselves in the analyses). Using these criteria, a total of 1.0 Mb of nuclear DNA sequences for *Denisova 4* and 24.1 Mb for *Denisova 8* (Tables 8 and 9) can be analyzed.

Table 8. Overview of DNA sequences produced, contamination estimates, and amount of nuclear sequences used for analyses.

	Denisova 4	*Denisova 8*
Amount of mapped sequences	54.6Mb	265Mb
MtDNA coverage	72-fold	119-fold
Autosomal contamination	~66%	~15%
mtDNA contamination	~5.2%	~3.2%
X chr. contamination	~28%	~9%
Nuclear sequences used	1Mb	24Mb

Table 9. DNA sequences yields.

	Mg of bone powder for extract[a]	**% of extract used for library**	**% endogenous[b]**	**Mb aligned to human genome[c]**	**Mb aligned after duplicate removal**	**% unique[d]**	**Mb aligned after deamination filter[e]**
Denisova 4	40	20%	0.05%	80.7 Mb	54.6 Mb	67.6%	1.0
Denisova 8	36	10%	0.9%	1,128 Mb	265 Mb	23.5%	24.1

a. Milligrams of bone powder used to make 100uL of extract
b. Percent endogenous is calculated as the Mb aligned to the human genome (after filtering for mapped sequences with a length above 35) divided by the total Mb sequenced (after filtering for a length above 35) times 100.
c. Mb aligned to hg19 after passing the following filters: length > 35, map quality > 37, merging of paird reads with minimum 11 bp overlap, fewer than 5 bases with base quality below 15, index reads with base qualities above 10.
d. Percent unique is Mb aligned with filters to the human genome after duplicate removal divided by aligned Mb before duplicate removal times 100
e. For deamination filter see the supplemental text.

5.3.4 MtDNA relationships

A phylogenetic tree relating the mtDNAs from the *Denisova 3, 4* and *8*, seven Neandertals from Spain, Croatia, Germany, the Russian Caucasus and the Altai Mountains (10, 73), and five present-day humans (Figure 15A, B) shows that the mtDNAs of the two Denisovan molars form a clade with *Denisova 3* to the exclusion of the Neandertals. The largest number of differences seen among the three Denisovan mtDNAs is 86 while the largest number of differences seen among seven Neandertal mtDNAs is 51 and among 311 present-day humans, 118 (Figure 15C). When comparing Watterson's estimator θw, which to some extent takes the numbers of samples into account, among the populations the mtDNA diversity of the three Denisovans is 3.5×10^{-3}, that of Neandertals 1.8×10^{-3} while that of present-day Europeans is 4.0×10^{-3} and present-day humans world-vide is 16.1×10^{-3}. Thus, mtDNA diversity among late Neandertals seems to be low relative to Denisovans as well as present-day humans.

The number of nucleotide changes inferred to have occurred from the most recent common ancestor (MRCA) of the three Denisovan mtDNAs to the *Denisova 4* molar, the *Denisova 3* phalanx and the *Denisova 8* molar are 55, 57 and 29 respectively (Figure 15B, Table 10). The corresponding number of substitutions from the MRCA of the seven Neandertal mtDNAs to each of the Neandertal mtDNAs varies between 17 and 25 (Table 10). This suggests that the time back to the mtDNA MRCA from the *Denisova 3* and the *Denisova 4* mtDNAs was almost twice as long as that from the *Denisova 8* mtDNA.

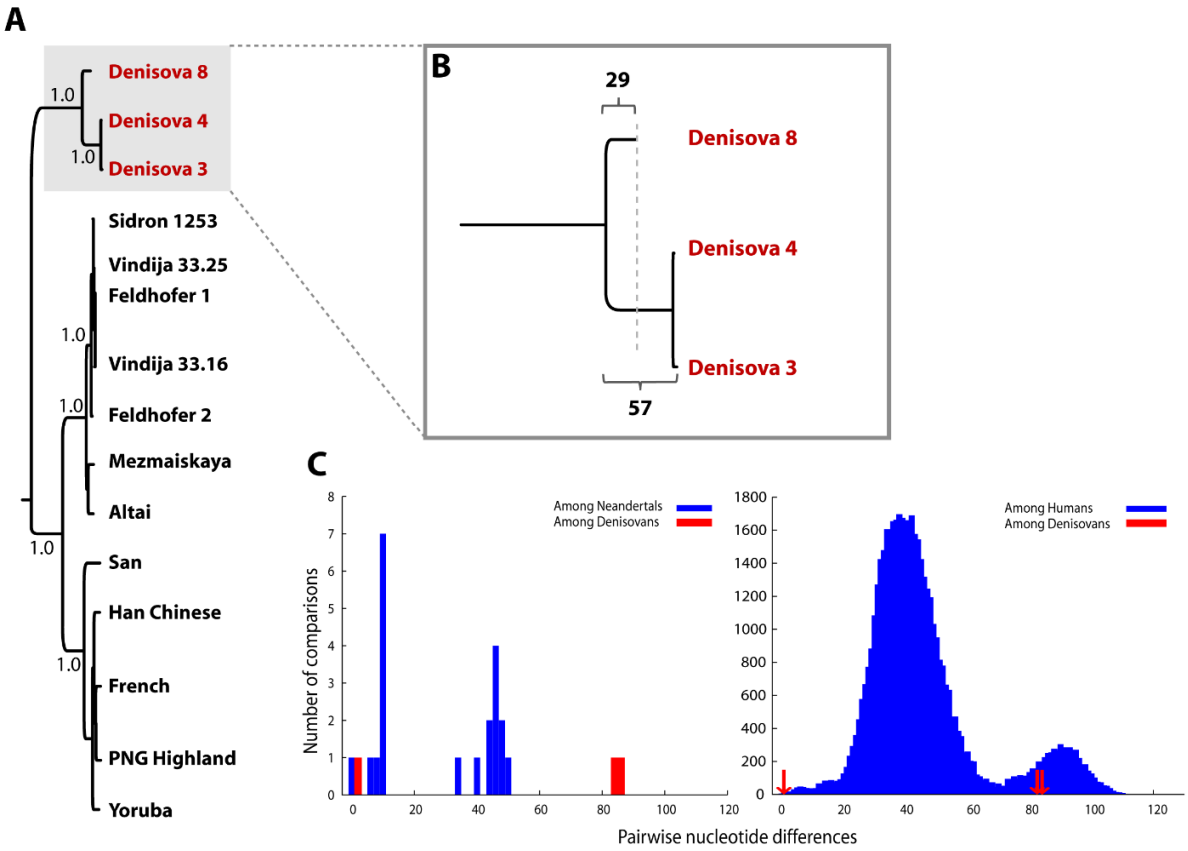

Figure 15. Evolutionary relationships of Denisovan mtDNAs. A. Bayesian tree relating the mtDNAs of three Denisovans, seven Neandertals and five present-day humans. Posterior probabilities are indicated. A chimpanzee mtDNA was used to root the tree. **B.** Numbers of differences between the two molar mtDNAs and the inferred common mtDNA ancestor of the three Denisovan mtDNA. **C.** Pairwise nucleotide differences among the Denisovans and Neandertals (left panel) and among the Denisovans and 311 present-day human mtDNAs (right panel).

Table 10. Number of differences to mtDNA MRCAs. The number of differences between each Denisovan mtDNA and their inferred MRCA as well as between each Neandertal mtDNA and their inferred MRCA.

Denisovan	Number of diffs to MRCA of Denisovans	Neandertal	Number of diffs to MRCA of Neandertals
Denisova 3	57	***Mezmaiskaya1***	25
Denisova 4	55	**Altai 1**	24
Denisova 8	29	**Feldhofer 1**	21
		Feldhofer 2	17
		Sidron 1253	19
		Vi33.16	23
		Vi33.25	21

Table 11. Watterson's estimator *(θw)* for mtDNA.

Population	# segregating sites	n (# indv)	*θw*
Denisovans	86	3	3.46E-03
Neandertals	73	7	1.80E-03
Present-day humans	1,689	311	16.1E-03
Present-day Europeans	262	31	3.96E-03

Table 12. Age estimates of the two molars and mtDNA lineages divergences based on mtDNA. Estimates using dates of 50,000 years as well as 100,000 years for *Denisova 3* and 95% upper and lower highest posterior densities (HPD) are given in thousand years (kyr).

	Age of *Denisova 3* set to 50,000 years BP			Age *of Denisova 3* set to 100,000 years BP		
Mitochondrial lineage	Estimate	95% HPD lower	95% HDP upper	Estimate	95% HPD lower	95% HDP upper
Denisova 8 age	177 kyr	97 kyr	265 kyr	226 kyr	143 kyr	313 kyr
Denisova 4 age	56 kyr	45 kyr	69 kyr	106 kyr	094 kyr	121 kyr
Denisova-Human/Neandertal	808 kyr	622 kyr	1,016 kyr	846 kyr	652 kyr	1056 kyr
Den8 – Den4/*Den3*	262 kyr	187 kyr	343 kyr	314 kyr	238 kyr	393 kyr
Human-Neandertal	405 kyr	312 kyr	511 kyr	413 kyr	318 kyr	522 kyr
San-rest of humans	173 kyr	128 kyr	223 kyr	176 kyr	128 kyr	225 kyr
Mezmaiskaya1-rest of Neandertals	128 kyr	101 kyr	155 kyr	129 kyr	103 kyr	157 kyr

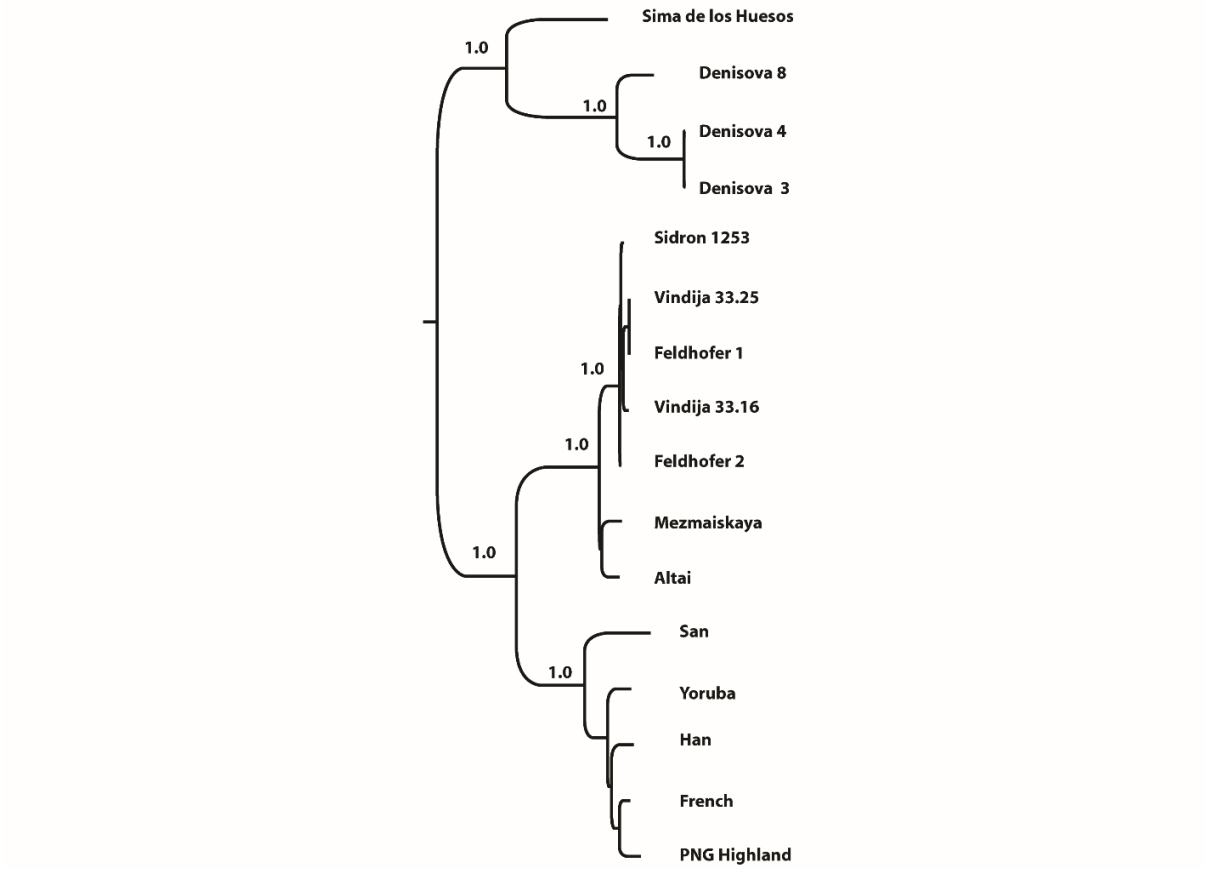

Figure 16. MtDNA tree of three Denisovans, seven Neandertals, a hominin from Sima de los Huesos (87), and five present-day humans. The Bayesian tree was computed using 16,286 mtDNA positions and a chimpanzee mtDNA (X93335.1) as outgroup (not shown). Important posterior probabilities are shown.

5.3.5 Autosomal Analyses

To estimate the divergence of the low coverage DNA sequences retrieved from *Denisova 4* and *8* to the high quality genomes of *Denisova 3* (3) as well as to the Neandertal from Denisova Cave and to ten present-day humans (10), we first counted nucleotide substitutions inferred to have occurred on the lineages from the human-chimpanzee ancestor to each of the high-coverage genomes (Figure 17A, a + b). We then used the low coverage molar sequences to estimate the fraction of those substitutions that occurred after their divergence from the high coverage lineages, *i.e*. the fraction of such substitutions not seen in the molars (Figure 17A, b). To the Denisovan high coverage genome, these fractions are 2.9% (95% CI: 2.28-3.44) and 3.4% (95% CI: 3.25-3.53) for *Denisova 4* and *Denisova 8*, respectively. Divergences of *Denisova 4* and *Denisova 8* are 8.9% (CI: 8.01-9.83%) and 8.3% (CI: 8.01-8.48 %) to the high coverage Neandertal genome and 10.9 - 12.9% to ten present-day humans (Figure 17B; Tables 10 and 11). These results show that the two teeth come from Denisovans and confirm that Denisovans were a sister group of Neandertals.

The average pairwise divergence among six low-coverage Neandertals to the Altai Neandertal genome is 2.5% (range 2.5% to 2.6%) (Table 14). This is slightly lower than the divergence of 2.9% and 3.4% of the two Denisovan molars from the Denisova genome and shows that the individuals from whom the two molars derive are almost as closely related to the *Denisova 3* genome as are the Neandertals to the Altai Neandertal genome. By comparison, the range of divergences among ten present-day human genomes is 4.2% to 9.5%, among the four Europeans 6.0 to 6.4% and between the two individuals from the South American tribal group Karitiana 4.2%. Thus, nuclear DNA diversity appears low among the archaic individuals, especially the Neandertals.

Using the high coverage *Denisova 3* genome it was shown that Denisovans have contributed DNA to present-day people in Oceania (8-10, 60). As expected, we found that *Denisova 8* also shares more derived alleles with Papuans and Australians than with other non-Africans (*D*: -0.04 to -0.07, [Z]=1.8–3.0, excluding CpG sites, Table 16). However, when we subsample from the high coverage Denisovan genome the DNA segments covered by fragments sequenced from *Denisova 4* we find that there are not enough data to similarly detect gene flow from *Denisova 4* to Oceanians (data not shown). This precludes us from asking whether either *Denisova 4* or *Denisova 8* are more closely related to the introgressing Denisovan than *Denisova 3*. Similarly, there are not enough data to determine whether gene flow from Neandertals at the

level detected in the high-coverage *Denisova 3* genome (10) is present in *Denisova 4* and *8* (Table 17).

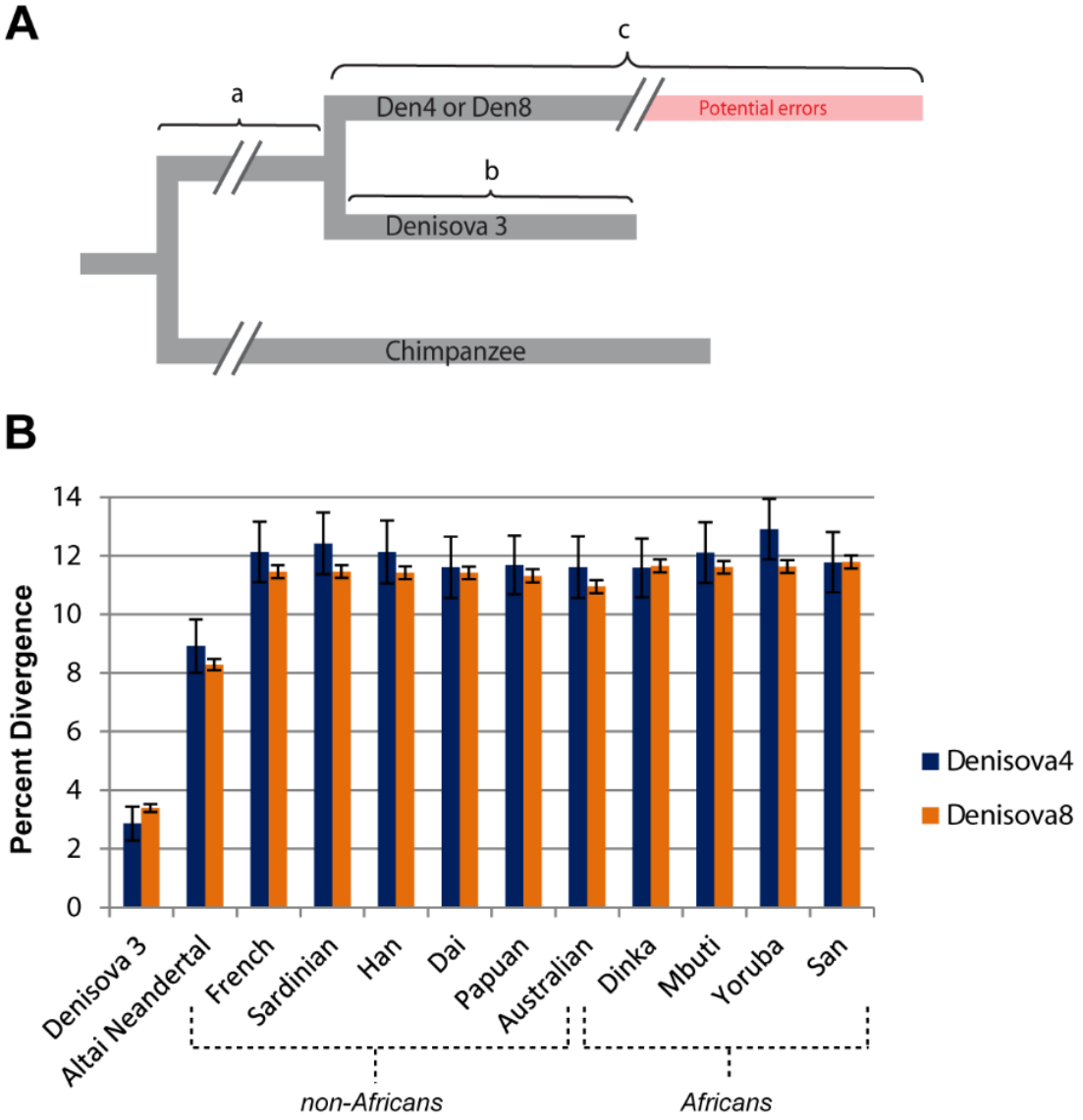

Figure 17. **Nuclear DNA divergence between *Denisova 4* and *8* and the Denisovan genome. A:** DNA sequences from *Denisova 4* and *8* were each compared to the genomes of *Denisova 3* (8) and the inferred human-chimpanzee ancestor (91, 92). The differences from the human-chimpanzee ancestor common to the two Denisovans (a) as well as differences unique to each Denisovan are shown (b and c). Errors in the low coverage Denisova genomes result in artificially long branches (c). Divergences of the molar genomes to *Denisova 3* are therefore calculated as the percent of all differences between *Denisova 3* and the human-chimpanzee ancestor that are not shared with the molar genomes, b/(a+b)x100. **B.** Autosomal divergences of *Denisova 4* and *Denisova 8* to the *Denisova 3* genome, the Neandertal genome, and ten present-day human genomes calculated as in A. All estimates are based on DNA fragments from the two molars that carry putative deamination-induced C to T substitutions. Bars indicate 95% confidence intervals.

Table 13. Divergences for *Denisova 4*. Divergences for the deaminated sequences, not deaminated sequences as well as all sequences combined are shown. Divergence is given the percent divergence of *Denisova 4* along the branch to the human-chimpanzee ancestor from the high-coverage genomes given in the first column. Percent divergences and 95% CI are given.

	Deaminated fragments				**Not deaminated fragments**				**All fragments**			
High-coverage genomes	Shared[1]	Genome[2]	*Den4*[3]	%	Shared	Genome	*Den4*	%	Shared	Genome	*Den4*	%
Denisova 3	3,699	109	3,767	**2.86** *2.28-3.44*	121,663	11,775	77,551	**8.82** *8.66-8.99*	126,716	11,990	81,920	**8.64** *8.48-8.81*
Altai Neandertal	3,471	340	4,029	**8.92** *8.01-9.83*	120,142	13,796	79,546	**10.30** *10.11-10.48*	124,952	14,290	84,303	**10.26** *10.08-10.44*
French	3,482	481	4,164	**12.14** *11.10-13.17*	126,237	11,133	76,306	**8.10** *7.94-8.27*	131,123	11,749	80,963	**8.22** *8.05-8.39*
Sardinian	3,448	489	4,095	**12.42** *11.37-13.47*	124,622	11,049	75,208	**8.14** *7.97-8.31*	129,262	11,634	80,055	**8.26** *8.09-8.42*
Han	3,455	477	4,111	**12.13** *11.06-13.2*	125,153	11,464	76,061	**8.39** *8.21-8.57*	129,955	11,919	80,724	**8.40** *8.23-8.57*
Dai	3,442	452	4,120	**11.61** *10.56-12.66*	124,793	11,519	75,590	**8.45** *8.28-8.62*	129,623	11,993	80,407	**8.47** *8.31-8.63*
Papuan	3,445	456	4,087	**11.69** *10.69-12.69*	124,182	11,617	75,444	**8.55** *8.38-8.73*	129,005	12,275	80,101	**8.69** *8.53-8.85*
Australian	3,418	449	4,098	**11.61** *10.56-12.67*	124,613	11,252	75,620	**8.28** *8.11-8.45*	129,368	11,845	80,360	**8.39** *8.23-8.55*

Dinka	3,418	448	4,159	**11.59** *10.58-12.59*	123,200	12,939	77,631	**9.50** *9.32-9.69*	127,989	13,397	82,318	**9.48** *9.3-9.66*
Mbuti	3,433	473	4,129	**12.11** *11.08-13.14*	122,769	13,726	78,122	**10.06** *9.87-10.24*	127,615	14,241	82,765	**10.04** *9.86-10.22*
Yoruba	3,473	515	4,146	**12.91** *11.88-13.95*	123,623	13,188	78,107	**9.64** *9.46-9.82*	128,425	13,890	82,882	**9.76** *9.57-9.95*
San	3,407	455	4,095	**11.78** *10.76-12.81*	121,951	13,989	77,901	**10.29** *10.10-10.48*	126,739	14,558	82,650	**10.30** *10.11-10.49*

1. The number of allelic states shared by the genome and *Densiova 4* but not shared with the human-chimpanzee ancestor.
2. Allelic states specific to the genome analyzed.
3. Allelic states specific to *Denisova 4*.

Table 14. **Divergences for *Denisova 8***. See Table 13 for explanations.

	Denisova8 deaminated				Denisova8 not deaminated				Denisova8 all			
Individual#1	Shared	Genome	*Den8*	%	Shared	Genome	*Den8*	%	Shared	Genome	*Den8*	%
Denisova 3	88,315	3102	33,574	**3.39** *3.25-3.53*	507,405	26,224	210,931	**4.91** *4.83-5*	637,505	31,657	261,670	**4.73** *4.64-4.82*
Altai Neandertal	84,101	7598	38,370	**8.29** *8.09-8.48*	486,591	47,274	234,493	**8.86** *8.73-8.97*	611,034	58,838	292,030	**8.78** *8.66-8.9*
French	82,999	10741	40,898	**11.46** *11.23-11.69*	486,909	60,026	243,442	**10.97** *10.86-11.09*	609,735	75,858	303,855	**11.02** *10.95-11.17*
Sardinian	82,188	10641	40,463	**11.46** *11.24-11.68*	481,113	59,575	240,320	**11.02** *10.9-11.13*	602,610	74,671	299,982	**11.05** *10.92-11.13*
Han	82,694	10661	40,505	**11.42** *11.2-11.64*	483,764	60,157	242,418	**11.06** *10.95-11.17*	606,187	75,989	302,355	**11.13** *11.03-11.24*
Dai	82,488	10633	40,676	**11.42** *11.2-11.64*	482,321	59,659	242,036	**11.01** *10.89-11.12*	604,506	75,249	302,505	**11.10** *10.97-11.17*
Papuan	82,423	10515	40,375	**11.31** *11.1-11.54*	481,045	59,090	240,568	**10.94** *10.83-11.05*	602,992	74,518	300,472	**11.00** *10.89-11.11*
Australian	82,513	10150	40,374	**10.95** *10.73-11.18*	482,594	57,825	240,792	**10.70** *10.59-10.81*	604,910	72,637	300,738	**10.76** *10.61-10.83*
Dinka	82,250	10846	40,385	**11.65** *11.43-11.87*	480,376	61,308	243,261	**11.32** *11.21-11.43*	601,643	76,990	303,706	**11.31** *11.24-11.45*
Mbuti	82,646	10858	40,571	**11.61**	480,838	62,446	244,989	**11.49**	603,063	78,469	305,286	**11.51**

				11.4-11.82				*11.37-11.61*				*11.4-11.63*
Yoruba	82,598	10875	40,745	**11.63** *11.42-11.85*	482,785	62,201	244,267	**11.41** *11.29-11.53*	604,950	77,960	304,739	**11.41** *11.31-11.52*
San	82,173	10985	40,645	**11.79** *11.57-12.01*	478,377	62,644	243,639	**11.58** *11.46-11.69*	599,764	79,290	304,396	**11.65** *11.57-11.78*

Table 15. Divergences for *Denisova 3*. See Table 13 for explanations.

	***Denisova 3* deaminated**				***Denisova 3* all**			
Individual#1	Shared	Genome	*Den3*	%	Shared	Genome	*Den3*	%
Denisova 3	-	-	-	-	-	-	-	-
Altai Neandertal	4531663	418624	1180396	**8.46** *8.37-8.54*	6040420	560355	1424811	**8.49** *8.4-8.57*
French	4439597	591694	1303961	**11.76** *11.68-11.84*	5908484	793950	1585350	**11.85** *11.76-11.93*
Sardinian	4391458	584609	1288694	**11.75** *11.67-11.82*	5842629	786441	1568908	**11.86** *11.78-11.94*
Han	4421887	587375	1295063	**11.73** *11.64-11.81*	5882753	788594	1576327	**11.82** *11.73-11.9*
Dai	4431008	587058	1299791	**11.70** *11.62-11.78*	5893395	788793	1581617	**11.80** *11.72-11.89*
Papuan	4410486	577448	1283146	**11.58** *11.49-11.66*	5867117	774918	1559910	**11.67** *11.58-11.75*
Australian	4433269	565793	1285982	**11.32** *11.23-11.4*	5899300	759006	1564065	**11.40** *11.31-11.49*
Dinka								

				11.69-11.87				*11.92-12.27*
Mbuti	4427808	593721	1301891	**11.82** *11.74-11.9*	5889250	795352	1585013	**11.90** *11.82-11.98*
Yoruba	4422950	592266	1297910	**11.81** *11.72-11.89*	5884572	794419	1581895	**11.89** *11.81-11.98*
San	4413422	595874	1297860	**11.90** *11.81-11.98*	5870882	798906	1580382	**11.98** *11.89-12.06*

Table 16. Divergences for Neandertals to the high coverage Altai Neandertal genome. See Table 13 for explanations of labels. All *Mezmaiskaya1* fragments were used for this analysis, because UDG treatment left C to T substitutions at only 4% of fragment ends.

	Neandertal deaminated				Neandertal all			
Neandertal	Shared	AltaiNea	Neandertal	%	Shared	AltaiNea	Neandertal	%
Feldhofer 1	447	6	576	1.32 *0.28-2.37*	2,581	67	3,446	2.53 *1.96-3.1*
Sidron 1253	893	29	1026	3.15 *2.00-4.29*	2,716	73	3,158	2.62 *1.97-3.26*
Vindija33.16	569,284	14,610	750,801	2.50 *2.44-2.57*	1,611,437	42,324	1,991,958	2.56 *2.5-2.61*
Vindija33.25	500,325	12,729	560,651	2.48 *2.41-2.55*	1,730,545	43,780	1,918,680	2.47 *2.41-2.52*
Vindija33.26	477,869	12,296	585,208	2.51 *2.44-2.58*	1,591,266	40,910	1,829,657	2.51 *2.45-2.56*
Mezmaiskaya1	-	-	-	-	2,331,784	59,473	772,431	2.49 *2.43-2.54*

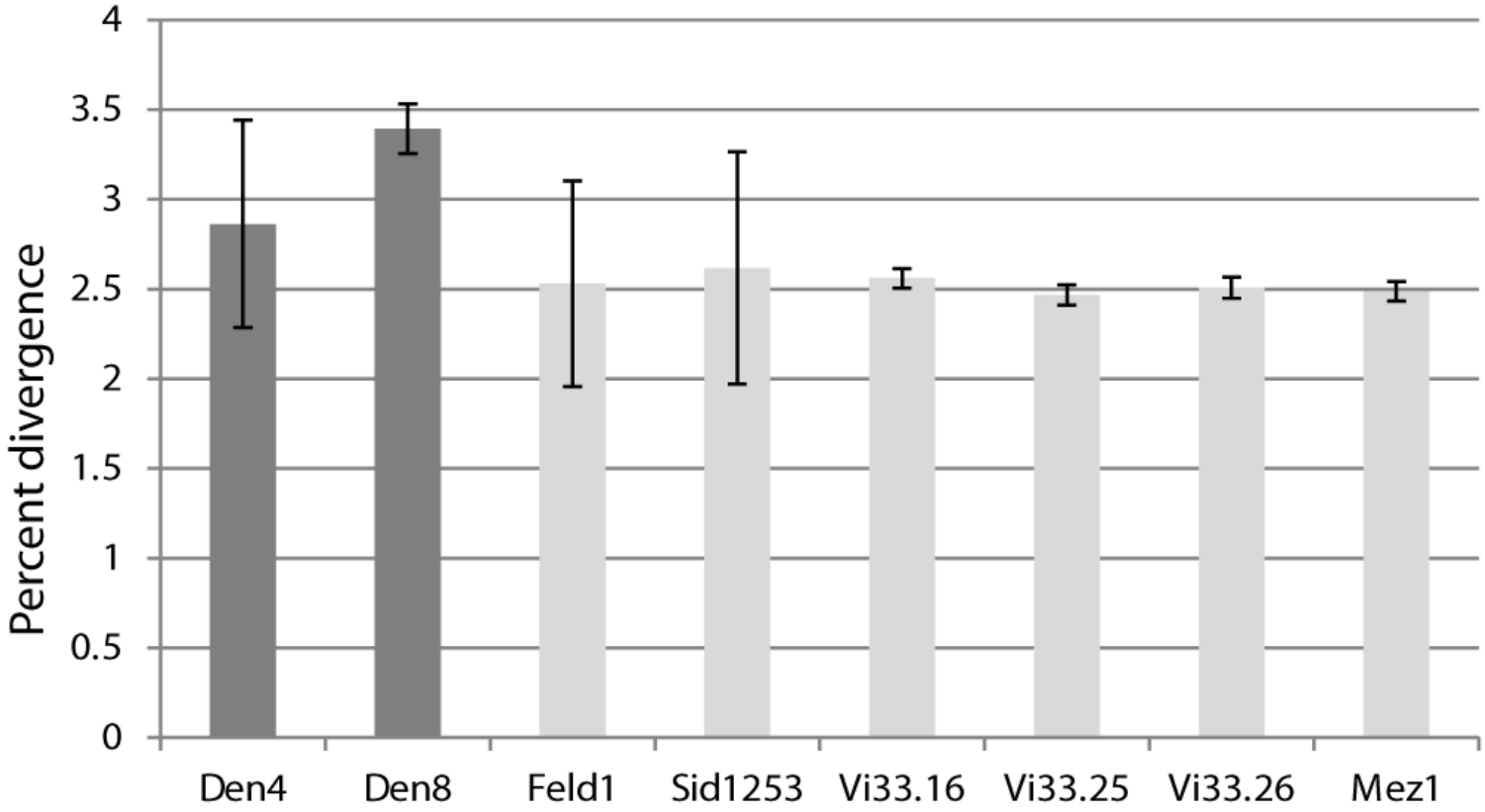

Figure 18. Divergences to *Denisova 3* and Altai Neandertal reference genomes. The percent divergence of the *Denisova 4* and *8* genomes to the *Denisova 3* genome (dark gray) and of six low-coverage Neandertal genomes to the Altai Neandertal genome (light gray) estimated as in main text Fig. 3A. Error bars indicate 95% CIs.

Table 17. Sharing of derived alleles between *Denisova 8* and Eurasian populations. Only *Denisova 8* fragments carrying a C to T substitutions at the first or last two positions are used.

	Type of sites	$AADA_a$	ADDA	DADA	DDDA	(ADDA-DADA)/ (ADDA+ADDA)	Z_b
Papuan, French, *Den8*, Chimp	all sites	43,502	1,311	1,473	205,735	-0.06	-3.03
	no cpg sites	36,640	906	1,022	179,687	-0.06	**-2.55**
	only cpg sites	6,862	405	451	26,048	-0.05	-1.57
	transitions	25,093	913	1,004	136,322	-0.05	-2.03
	transversions	18,409	398	469	69,413	-0.08	-2.33
Papuan, Sardinian, *Den8*, Chimp	all sites	43,387	1,358	1,454	205,685	-0.03	-1.90
	no cpg sites	36,519	930	1,023	179,680	-0.05	**-2.24**
	only cpg sites	6,868	428	431	26,005	0.00	-0.10
	transitions	25,031	944	1,010	136,224	-0.03	-1.56
	transversions	18,356	414	444	69,461	-0.03	-1.02
Papuan, Han, *Den8*, Chimp	all sites	43,255	1,232	1,352	204,023	-0.05	-2.32
	no cpg sites	36,435	832	951	178,188	-0.07	**-2.84**
	only cpg sites	6,820	400	401	25,835	0.00	-0.03
	transitions	24,989	855	913	135,233	-0.03	-1.36
	transversions	18,266	377	439	68,790	-0.08	-2.09
Papuan, Dai, *Den8*, Chimp	all sites	43,215	1,199	1,356	204,110	-0.06	-3.31
	no cpg sites	36,360	833	956	178,239	-0.07	**-3.01**
	only cpg sites	6,855	366	400	25,871	-0.04	-1.32
	transitions	24,981	816	927	135,238	-0.06	-2.81

	transversions	18,234	383	429	68,872	-0.06	-1.66
Australian, French, *Den8*, Chimp	all sites	43,027	1,224	1,451	204,126	-0.08	-4.32
	no cpg sites	36,314	861	966	178,308	-0.06	**-2.43**
	only cpg sites	6,713	363	485	25,818	-0.14	-4.19
	transitions	24,847	865	979	135,092	-0.06	-2.66
	transversions	18,180	359	472	69,034	-0.14	-3.82
Australian, Sardinian, *Den8*, Chimp	all sites	43,118	1,313	1,482	204,409	-0.06	-3.19
	no cpg sites	36,335	915	1,023	178,581	-0.06	**-2.40**
	only cpg sites	6,783	398	459	25,828	-0.07	-2.13
	transitions	24,892	896	1,009	135,205	-0.06	-2.60
	transversions	18,226	417	473	69,204	-0.06	-1.91
Australian, Han, *Den8*, Chimp	all sites	43,016	1,228	1,389	202,806	-0.06	-3.06
	no cpg sites	36,281	844	944	177,174	-0.06	**-2.34**
	only cpg sites	6,735	384	445	25,632	-0.07	-2.07
	transitions	24,852	875	927	134,265	-0.03	-1.20
	transversions	18,164	353	462	68,541	-0.13	-3.92
Australian, Dai, *Den8*, Chimp	all sites	42,767	1,243	1,391	202,727	-0.06	-2.98
	no cpg sites	36,047	894	969	177,058	-0.04	**-1.80**
	only cpg sites	6,720	349	422	25,669	-0.09	-2.75
	transitions	24,757	848	929	134,146	-0.05	-2.04
	transversions	18,010	395	462	68,581	-0.08	-2.20
Papuan, Han, *Den3*, Chimp	all sites	71,720	8,606	9,909	1,397,467	-0.07	-9.5

	no cpg sites	60,186	6,052	7,040	1,225,758	-0.08	**-8.60**
	only cpg sites	11,534	2,554	2,869	171,709	-0.06	-4.211
	transitions	48,439	5,944	6,801	927,866	-0.07	-7.52
	transversions	23,281	2,662	3,108	469,601	-0.08	-5.87
Papuan, French, *Den3*, Chimp	all sites	71,440	8,886	10,258	1,397,118	-0.07	-10.0
	no cpg sites	59,920	6,284	7,224	1,225,378	-0.07	**-8.30**
	only cpg sites	11,520	2,602	3,034	171,740	-0.08	-5.69
	transitions	48,215	6,168	7,094	927,573	-0.07	-8.08
	transversions	23,225	2,718	3,164	469,545	-0.08	-6.00

a. 'A' refers to an ancestral state and 'D' refers to a derived state. Thus, this column shows the number of sites where populations 1 and 2 share the ancestral allele with population 4 (Ancestral), and population 3 (Derived) has a derived state.
b. The *Z*-score is the difference between the *D*-statistics using all data and the mean of the same statistics for bootstrap replicates divided by the standard deviation for those replicates.

Table 18. Sharing of derived alleles between *Denisova 8* and Neandertals. *Denisova 8* fragments carrying a C to T substitutions at the first or last two positions (*Den8*_deaminated) as well as all fragments (*Den8*_all) are used. Only estimates based on transversions can be used due to errors in the low coverage *Mezmaiskaya1* genome.

	Type of sites	**AADA**	**ADDA**	**DADA**	**DDDA**	**(ADDA-DADA)/ (ADDA+ADDA)**	**Z**
Mez, AltaiNea, *Den8*_deam, Chimp	all sites	15,245	511	376	77,110	0.15	4.49
	no cpg sites	12,142	179	139	64,649	0.13	2.27
	only cpg sites	3,103	332	237	12,461	0.17	4.00
	transitions	8,898	431	313	52,358	0.16	4.34
	transversions	**6,347**	**80**	**63**	**24,752**	**0.12**	**1.44**
Mez, AltaiNea, *Den8*_all, Chimp	all sites	104,707	3,586	2,532	521,739	0.17	13.88
	no cpg sites	87,986	1,382	1,138	441,125	0.10	4.92
	only cpg sites	16,721	2,204	1,394	80,614	0.23	14.12
	transitions	56,272	3,063	2,041	354,226	0.20	14.75
	transversions	**48,435**	**523**	**491**	**167,513**	**0.03**	**1.02**
Mez, AltaiNea, *Den3*, Chimp	all sites	3,392	296	498	77,271	-0.25	-7.15
(sites covered by *Den8*_deaminated)	no cpg sites	2,655	121	177	64,648	-0.19	-3.21
	only cpg sites	737	175	321	12,623	-0.29	-6.75
	transitions	2,371	234	420	52,530	-0.28	-7.44
	transversions	**1,021**	**62**	**78**	**24,741**	**-0.11**	**-1.32**
Mez, AltaiNea, *Den3*, Chimp	all sites	23,573	3,463	2,024	523,579	0.26	20.23
(sites covered by *Den8*_all)	no cpg sites	18,757	1,348	957	441,914	0.17	8.38
	only cpg sites	4,816	2,115	1,067	81,665	0.33	20.29
	transitions	16,333	2,977	1,599	355,784	0.30	21.29
	transversions	**7,240**	**486**	**425**	**167,795**	**0.07**	**2.04**
Mez, AltaiNea, *Den3*, Chimp	all sites	295,159	42,000	24,746	6,550,020	0.26	70.25
(all *Den3* sites, not conditioned on	no cpg sites	232,239	16,149	11,699	5,547,171	0.16	27.62
***Den8*)**	only cpg sites	62,920	25,851	13,047	1,002,849	0.33	67.79
	transitions	205,685	36,111	19,517	4,490,038	0.30	76.13
	transversions	**89,474**	**5,889**	**5,229**	**2,059,982**	**0.06**	**6.35**

5.4 Discussion

The nuclear DNA sequences retrieved from *Denisova 4* and *8* are more closely related to the *Denisova 3* genome used to define the Denisovans as a hominin group than to present-day human or Neandertal genomes. Furthermore, the mtDNA of the two molars form a clade with *Denisova 3*. Thus, the present work extends the number of Denisovan individuals identified by mitochondrial and nuclear DNA from one to three. Although the number of Denisovan individuals is small, restricted to one locality, and differ in age, it is nevertheless interesting to note that the nuclear DNA sequence diversity among the three Denisovans is slightly higher than that found among seven Neandertals although these are widely geographically distributed, but lower than that seen among present-day humans world-wide or among Europeans.

Although the three Denisovans come from a single cave, they may differ significantly in age as indicated by the branch-length of the mtDNA of the *Denisova 8* molar which is shorter than those of *Denisova 4* and the *Denisova 3*, an observation that is congruent with the stratigraphy. If we assume that the mtDNA mutation rate of ~2.5 x 10^{-8}/site/year (95% confidence interval 1.8-3.2) estimated for modern humans (106) applies also to Denisovan mtDNA, *Denisova 8* is in the order of 60,000 years older than *Denisova 3* and *Denisova 4*. A similar or even larger age difference between *Denisova 8* and the other two teeth is suggested by a Bayesian analysis (Suppl. Material; Table 12). Although it is unclear if the mtDNA mutation rate in archaic humans is similar to that in modern humans and thus if the difference in age is as large as this, it is clear that *Denisova 8* is substantially older than *Denisova 4* and *Denisova 3*. This is of interest from several perspectives.

First, the two molars are very large and their morphology is unlike what is typical for either Neandertals or modern humans. Since they differ substantially in age this reinforces the view that Denisovan dental morphology was not only distinct from that of both Neandertals and modern humans, but also a feature typical of Denisovans over an extended period of time, at least in the Altai region. This may prove useful for the identification of potential Denisovan teeth at other sites.

Second, the difference in age between the two Denisovan molars as well as their similar morphology suggests that Denisovans used Denisova Cave at least twice and possibly over a long time, perhaps interrupted by Neandertal occupation(s) (10). Denisovans may therefore have been present in southern Siberia over an extended period. Alternatively, they may have been present in neighboring regions from where they may have periodically extended their range to the Altai.

Third, the *Denisova 8* molar is not only older than *Denisova 4* and *Denisova 3*, its mtDNA differs substantially from the other two. The mtDNA diversity among the three Denisovan individuals is larger than that among seven Neandertals from which complete mtDNA sequences are available (Figure 15C), despite the fact that the Denisovans all come from the same site while the Neandertals are broadly distributed across western and central Eurasia. Notably, the nuclear genome of *Denisova 8* also shows a tendency to be more deeply diverged from the genome of *Denisova 3* than is *Denisova 4* (Figure 17B). Given that the high-coverage genome from the *Denisovan 3* phalanx carries a component derived from an unknown hominin who diverged 1-4 million years ago from the lineage leading to Neandertals, Denisovans and present-day humans (10), it is possible that this component differs among the three Denisovan individuals. In particular, it may be that the older Denisovan population living in the cave carried a larger or different such component. It is also possible that the two diverged mtDNA lineages seen in *Denisova 8* on the one hand and *Denisova 3* and *4* on the other were both introduced into the Denisovans from this unknown hominin as has been suggested for the mtDNA of *Denisova 3* (8, 9). However, more nuclear DNA sequences from Denisovan specimens of ages similar to *Denisova 4* and *8* are needed to address this question fully.

6. A Neandertal from Denisova Cave with ancient spotted hyena contamination

This manuscript is in preparation for publication

Authors: Susanna Sawyer[a], Bence Viola [b,c,d], Michael V. Shunkov [d,e], Anatoly P. Derevianko [d,f], Matthias Meyer[a], Kay Prüfer [a], Svante Pääbo [a].

[a] Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany

[b] Department of Anthropology, University of Toronto, Toronto, ON M5S 2S2, Canada

[c] Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany

[d] Institute of Archaeology and Ethnography, Russian Academy of Sciences, Siberian Branch, Lavrentieva Avenue, 17 Novosibirsk, RU-630090, Russia.

[e] Novosibirsk National Research State University, Pirogova st., 2, Novosibirsk, RU-630090, Russia

[f] Altai State University, Pr. Lenina, 61, Barnaul, RU-656049, Russia

Author contributions:

Contributed materials: Michael Shunkov, Anatoly Derevianko

Performed experiment: Susanna Sawyer

Analyzed morphological data: Bence Viola

Analyzed mitochondrial data: Susanna Sawyer, Kay Prüfer, Matthias Meyer

Analyzed nuclear data: Susanna Sawyer

Wrote the paper: Susanna Sawyer, Bence Viola and Svante Pääbo

6.1 Abstract

To date, four hominin remains have been published from Denisova Cave. A piece of a finger phalanx (*Denisova 3*) and two upper molars (*Denisova 4* and 8) belong to a recently discovered sister group of Neandertals, the Densiovans. The fourth remain, a toe phalanx, belongs to a Neandertal (*Altai 1*). Here we present an almost complete mitochondrial genome as well as 18.4 Mb of the autosomal genome of a second Neandertal from Denisova Cave (*Altai 2*). Based on mtDNA, *Altai 2* is most closely related to *Altai 1* compared to all other published Neandertal mtDNA genomes. We also present partial mtDNA genomes of spotted hyenas from the Pleistocene that contaminated the *Altai 1* and *Altai 2* Neandertals as well as *Denisova 3* and *4*. The hyena contamination is heaviest in *Altai 2*, and belonged to a hyena population that until now has only been found in eastern Russia and China. The hyenas that contaminated the Denisovans, came from a population found in Pleistocene Europe as well as northern Africa today.

6.2 Introduction

Denisova Cave is located in the Siberian Altai Mountains on the Anui River. In 2008, a small piece of a child's finger phalanx (*Denisova 3*) was found in the east Gallery of the cave. Ancient DNA was extracted from the bone, and the nuclear and mtDNA genomes were sequenced to high coverage (7, 8). The nuclear DNA showed that this bone belonged to a new hominin group, the Denisovans, a sister group to Neandertals. Surprisingly, the mtDNA is far deeper diverged, with a divergence from the early modern human-Neandertal ancestor twice as deep as the Neandertal divergence from early modern humans (7). The low-coverage genomes from two molars (*Denisova 4* and 8) showed that an additional two Denisovans left remains in the cave. *Denisova 8* belonged to a Denisovan that lived many millennia earlier than *Denisova 3* or *4*, showing that Denisovan presence in the region occurred at least twice over a long period of time (Chapter 5). In 2010, a second phalanx was found in the East Gallery of Denisova Cave, this time from an adult toe. The DNA preservation proved to be remarkable enough to produce another high coverage genome, which revealed that the individual to whom the toe phalanx belonged was a Neandertal (10), called *Altai 1* here. *Denisova 3* shows more than 0.5% Neandertal admixture from a population more closely related to *Altai 1* than other Neandertals (10). Present-day non-African humans also show Neandertal admixture (1.5-2.1%), from an admixture event that

happened 37-86kya (140) from a Neandertal population that was more closely related to *Mezmaiskaya1*, a 65,000 year old Neandertal from the Russian Caucasus (10).
Although none of the hominin remains from Denisova Cave are dated, Denisovan and Neandertal occupation of the region most likely happened over a long time period in the Pleistocene, possibly between 150,000 and 50,000 years ago (Chapter 5). During this time, the Altai Mountains were also home to a large number of large carnivores, especially the spotted hyena (2, 3, 63). The Pleistocene spotted hyena had a vast range from Europe to eastern Asia (12). Due to its larger size than the current spotted hyena found in sub-Saharan Africa, it was often hypothesized to be a separate species (12, 64), however genetic data shows that the Pleistocene spotted hyenas fall within the variation of present-day spotted hyenas (13, 14). Spotted hyenas are impressive carnivores, with a diet consisting completely of meat (65), and jaws able to crush and digest entire skeletons, leaving behind only teeth, hair, horns and often smaller metapodials (16). They live in clans, and dig large dens, and both hunt and scavenge for food (17). Spotted hyenas today are not fearful of humans (3), suggesting that the Pleistocene spotted hyena was a terrifying opponent to hominins of the time.
In 2011, a second finger phalanx from an adult was found in layer 12 of the East Gallery of Denisova Cave. Here we present the almost complete mtDNA genome and 18.4 Mb of the nuclear genome from this phalanx, called *Altai 2*. We also present the partial mtDNA genome of the spotted hyena contaminant of *Altai 1*, *Altai 2*, *Denisova 3* and *Denisova 4*.

6.3 Results

6.3.1 *Altai 2* morphology

Altai 2 is a distal manual phalanx found in 2011 in the East Gallery of Denisova Cave. The phalanx derives from Layer 12 in square G-3. The official excavation name of this phalanx is *Denisova 9*, however I refer to this phalanx as *Altai 2* in this thesis, to avoid confusion.
The specimen is well preserved, with a complete distal end. Proximally, the palmar aspect of the articular surface is missing. The presence of a large and flattened apical tuft, as opposed to a claw indicates that the specimen is a human (Figure 19). Similarly broad and rounded apical tufts are usually present in Neanderthals, while modern humans usually have narrower and oval apical tufts. There is no indication of acid etching, thuis no indication of the bone having been exposed to stomach acids.

Figure 19. The manual distal phalanx from *Altai 2*. Each square represents 1 cm.

6.3.2 Neandertal mtDNA analyses

Ten libraries were produced from three DNA extractions (Figure 9, Table 2) of bone powder from three areas of the *Altai 2* bone (Figure 7). After enrichment for human mtDNA, sequences from each of the ten libraries were filtered as described in section 4.2.3, and aligned to the human reference mtDNA genome (rCRS, (76)). The sequences were then compared to 69 diagnostic positions where 10 Neandertals differ from 311 present-day humans from around the world. Each of the ten libraries has more sequences containing Neandertal diagnostic bases than human bases (54.3 to 93.3 percent Neandertal, Table 19). The same analysis was done for the combined libraries for Denisovan versus human diagnostic positions. 99.5% of the sequences carry human-like bases. After filtering for sequences with a putative C to T change at the ends of molecules, the percentage of sequences that carry the present-day human base decreases to 0.46 to 8.8% in the ten libraries, indicating that this bone belongs to a Neandertal. After merging of the putative C to T filtered libraries the present-day human contamination rate is 1.9% (95%CI: 1.6-2.3%) (Table 19).

Table 19. Contamination estimates of human mtDNA captured and aligned *Altai 2* libraries. Estimates are given per library as well as after merging of the ten libraries (TOTAL). Present-day human contamination (PD-human cont) is given for all sequences as well as sequences filtered for terminal C to Ts. Spotted hyena contamination is after terminal C to T filtering. 95 % CI in brackets.

			PD-human cont (All seq)			PD-human cont (terminal C>T seqs)			Spotted hyena contamination		
Parent Library	Library	Description	Human-like seqs	Neandertal-like seq	% contamination	Human-like seqs	Neandertal-like seq	% contamination	Hyena-like seqs	Human-like seqs	% contamination
L9366	L9632	UDG	2844	8404	25.3 (24.5-26.1)	12	866	1.4 (0.8-2.4)	413	1968	17.3 (15.9-18.9)
L9367	L9605	UDG	1201	3483	25.6 (24.4-26.9)	3	287	1.0 (0.4-3.0)	33	109	23.2 (17.1-30.8)
L9575	L9586	UDG, u-selec fraction	131	1821	6.7 (5.7-7.9)	3	644	0.46 (0.2-1.4)	50	239	17.3 (13.4-22.1)
L9576	L9587	UDG, u-selec fraction	138	1863	6.9 (5.9-8.1)	4	688	0.57 (0.2-1.5)	48	259	15.6 (12.0-20.1)
L9580	L9591	UDG, u-dep fraction	2470	8957	21.6 (20.9-22.4)	3	104	2.8 (1.0-7.9)	5	31	13.9 (6.1-28.7)
L9581	L9592	UDG, u-dep fraction	2224	7886	22.0 (21.2-22.8)	2	80	2.4 (0.7-8.5)	14	26	35.0 (22.1-50.5)
L9467	R3108	UDG	278	1353	17.0 (15.8-19.8)	4	121	3.2	347	392	47.0 (43.4-50.6)
A9231	L9627	UDG	245	1139	17.7 (15.8-19.8)	5	73	6.4 (2.8-14.1)	269	338	44.3 (40.4-48.3)
A9232	L9628	$NaHPO_4$ wash, no UDG	77	126	37.9 (31.5-44.8)	3	77	3.8 (1.3-10.5)	6	31	16.2 (7.7-31.1)
A9233	L9629	after $NaHPO_4$ wash, no UDG	768	913	45.7 (43.3-48.1)	53	549	8.8 (6.8-11.3)	476	559	46.0 (43.0-49.0)
	TOTAL		11,405	37,661	23.2 (22.9-23.6)	104	5370	1.9 (1.6-2.3)	584	1216	32.4 (30.3-34.6)

The sequences of these ten libraries were subsequently mapped to the Vindija 33.16 Neandertal mtDNA (AM948965, (59)). The average coverage across the mtDNA genome ranged in the ten libraries from 3.9 to 211.2-fold, with a total coverage of 946.3-fold. After C to T filtering the coverage was reduced to 102.8-fold (Table 20).

We calculated the spotted hyena contamination level in the conserved regions of the mtDNA genome as described in section 4.2.5. The spotted hyena contamination levels varied from 13.8 to 46.9% in the ten libraries with a combined contamination level of 32.4% (95%CI: 30.3-34.6%) (Table 19).

The Neandertal mtDNA genome was built as described in section 4.2.6, leaving an average coverage of 98-fold and six unresolved positions (positions 185, 189, 195, 2951, 2964 and 5767 in rCRS coordinates) (Figure 20). A Bayesian phylogenetic mtDNA tree, based on 16,446 positions, shows that the *Altai 2* Neandertal falls together with the *Altai 1* Neandertal (posterior probability of 1) (Figure 21). The *Altai 1* and *Altai 2* Neandertals have ten differences between their mtDNAs. All ten differences are at positions where the *Altai 2* alingment has a coverage above 44 and a consensus support above 94%. Six of these differences are unique to *Altai 2*. Both *Altai 1* and *Altai 2* carry more differences to the other Neandertals (Table 21). There are 30 differences between the *Altai 2* and the MRCA of Neandertals (calculated exactly as in section 4.1.10), which is the highest number of differences to the MRCA when compared to seven other Neandertals (Table 10), five more than to the next highest, *Mezmaiskaya1*.

Table 20. Coverage of the Neandertal and spotted hyena mtDNA genome of the ten libraries from *Altai 2*.

				Human mtDNA capture			Spotted hyena mtDNA capture		
Extract	Parent Library	Description	mg bone used for lib	Avg coverage (all;CT filtered)	Capture efficiency	% uniq	Avg coverage, no human regions (all;CT filtered)	Capture efficiency	% uniq
E1114	L9366	UDG	3	211.2 ; 16.0	14.2	30.1	78.4; 10.5	17.4	19.1
	L9367	UDG	3	168.3 ; 12.3	5.2	32.9	22.4; 3.1	6.6	78.5
	L9575	UDG, u-selec fraction	4.5	35.9 ; 28.7	5.1	15.4	18.7; 15.7	14.4	
	L9576	UDG, u-selec fraction	4.5	36.9 ; 30.0	5.4	15.2	20.5; 17.5	16.8	4.8
	L9580	UDG, u-dep fraction	4.5	209.2 ; 2.7	8.5	44.1	98.1; 3.5	15.6	19.2
	L9581	UDG, u-dep fraction	4.5	186.2 ; 2.2	4.8	59.5	103.7; 3.7	14.7	22.4
E1269	L9467	UDG	1.2	32.1 ; 2.5	13.6	4.6	59.4; 6.8	22.6	8.8
	A9231	UDG	1.2	26.9 ; 2.0	7.9	4.0	43.1; 4.9	17.5	17.5
E3000	A9232	$NaHPO_4$ wash, no UDG	4	3.9 ; 0.6	0.1	80.5	8.4; 1.7	0.8	26.5
E3001	A9233	after $NaHPO_4$ wash, no UDG		35.7 ; 5.9	4.1	9.5	90.0; 17.6	17.0	18.7
	TOTAL	-	-	946.3 ; 102.8	-	-	543; 85.0	-	-

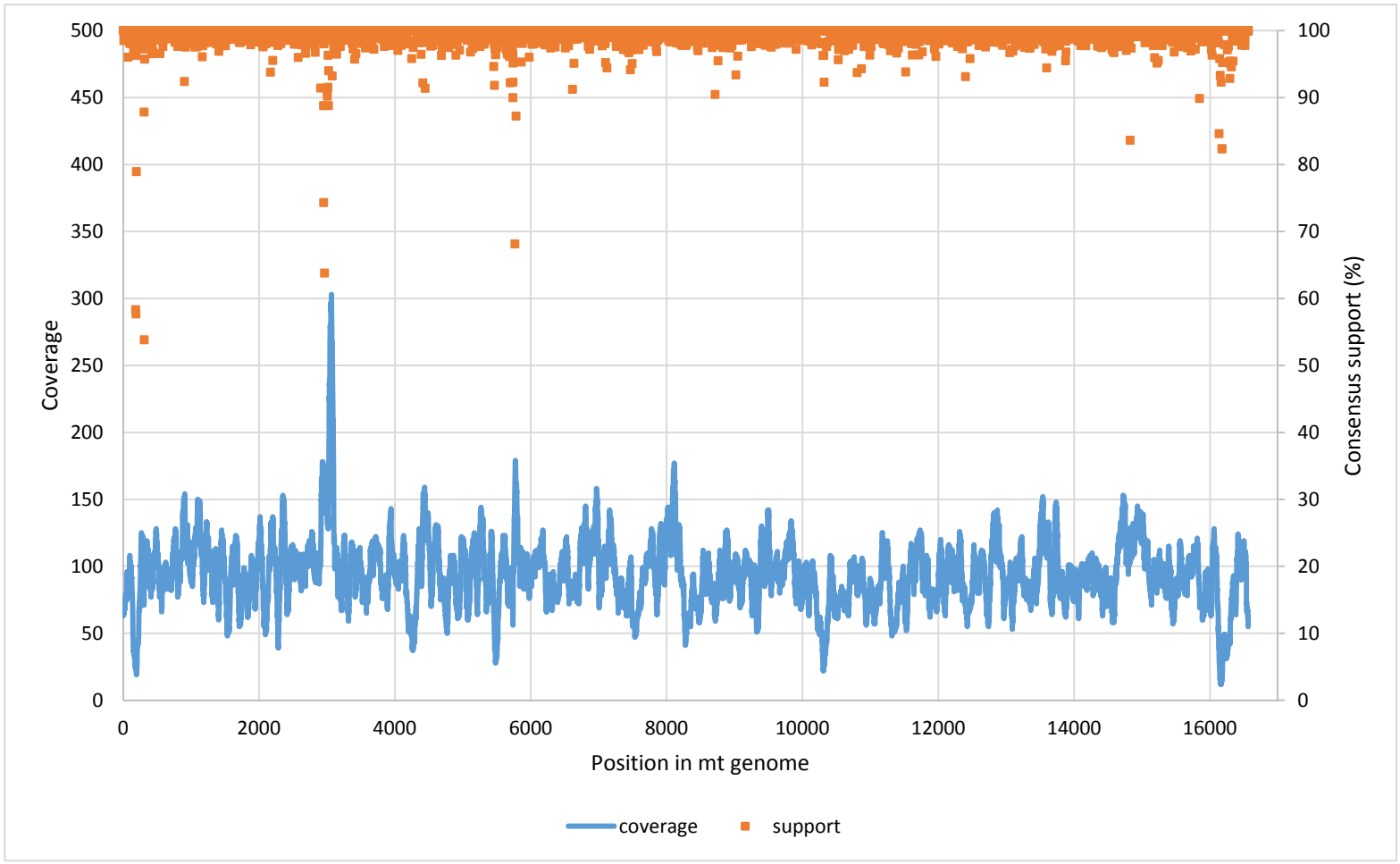

Figure 20. Coverage and consensus support of the merged *Altai 2* libraries after human mtDNA capture, alignment to Neandertal mtDNA and filtering (see section 4.2.6).

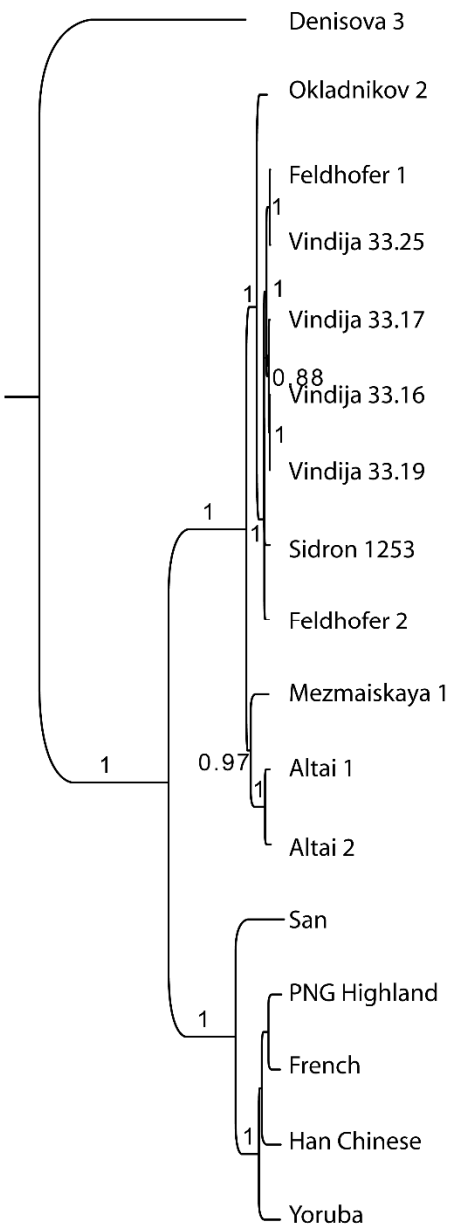

Figure 21. Bayesian tree of eleven Neandertal mtDNAs (including the *Altai 2*), five present-day humans, one Denisovan and a Chimpanzee as an outgroup (not shown). Posterior values are shown.

Table 21. Pairwise differences among ten Neandertals, including the *Altai 2* Neandertal.

	Mez1	**Feld1**	**Vindija 33.25**	**Feld2**	**Sidron**	**Vindija 33.16**	**Vindija 33.19**	**Vindija 33.17**	***Altai 1***
Feldhofer1	43								
Vindija 33.25	43	0							
Feldhofer2	38	9	9						
Sidron1253	41	10	10	5					
Vindija 33.16	44	9	9	8	9				
Vindija 33.19	44	9	9	8	9	0			
Vindija 33.17	43	8	8	7	8	1	1		
Altai 1	32	45	45	40	43	46	46	45	
Altai 2	38	53	53	48	51	54	54	53	10

6.3.3 Autosomal analyses

The ten *Altai 2* libraries were each sequenced without enrichment. Percent endogenous was calculated for each of the ten libraries, after alignment to the human genome (see section 4.2.10 for details) (Table 22). The percent endogenous for the regular libraries of E1114 is highest at 1.2%. The U-selected libraries, L9575 and L9576, have a slight decrease in percent endogenous, but are still double than that of the U-depleted libraries (L9580 and L9581). The two libraries from E1269 have a percent endogenous of 1.0%. The $NaHPO_4$ wash (A9232) has the lowest percent endogenous, half that of the extraction after the $NaHPO_4$ wash (A9233).

Table 22. Percent endogenous (% end) for each of the ten libraries.

Extract	E1114						E1269		E3000	E3001
Library	L9367	L9366	L9575	L9576	L9580	L9581	L9467	A9231	A9232	A9233
% end	1.2	1.2	0.84	0.82	0.46	0.46	1.0	1.0	0.26	0.56

The two U-selected libraries (L9575 and L9576) were pooled and sequenced on one lane of a HiSeq. After filtering for sequences with putative C to T changes (see section 4.1.12) there exist 18.4 Mb of data. Divergence was calculated to *Altai 1*, *Denisova 3* and ten present-day humans on the lineage to the human-chimpanzee ancestor as described in sections 4.1.13 and Figure 17a. The divergence of *Altai 2* is lowest to *Altai 1* (3.31, 95% CI 3.13-3.49), second lowest to *Denisova 3* (10.11, 95% CI: 9.86-10.36), and highest to the ten present-day humans (12.7-13.22) (Figure 22 and Table 23).

In addition the divergence of the *Altai 2* to the *Altai 1* was also compared to the divergences of seven low-coverage Neandertals to *Altai 1* (as described in sections 4.1.13 and 5.3.5). The divergence of *Altai 2* to *Altai 1* is highest (3.31, 95% CI 3.13-3.49) than that of the other Neandertals to *Altai 1* (2.47-2.62, Table 16).

The *Altai 2* sequences were also compared to derived positions on the Denisovan, Neandertal, present-day human (based on a single Mbuti individual), Denisovan-Neandertal, human-Neandertal, and human-Denisovan lineages. The fraction of sequences that share the derived states of each lineage were calculated as described in Meyer *et al* (113). The *Altai 2* shares the

highest fraction of derived positions with the Neandertal, Denisovan-Neandertal and human-Neandertal lineages (>0.7), while it shares less with the other three lineages (<0.1) (Table 24).

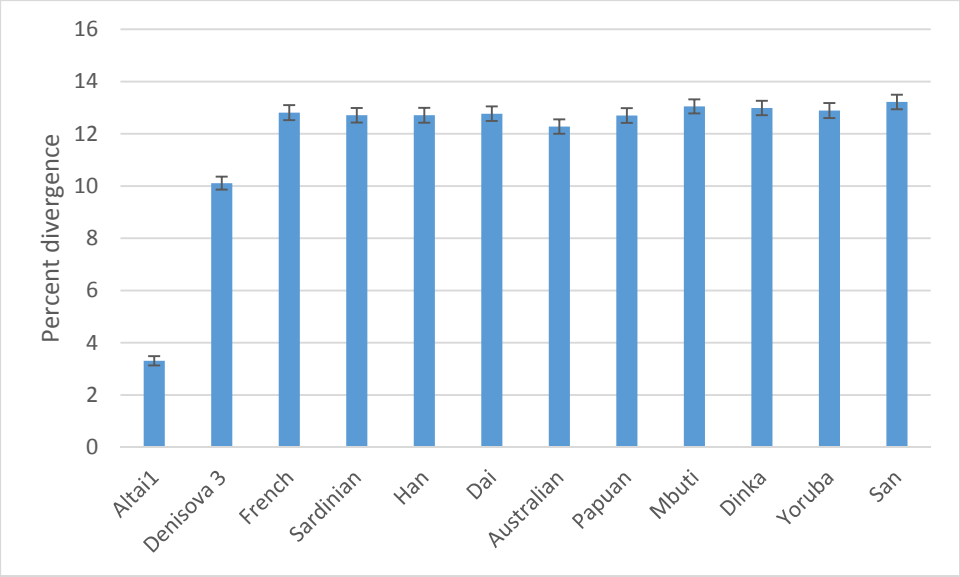

Figure 22. Autosomal divergences of *Altai 2* to the *Altai 1* Neandertal genome, the *Denisova 3* genome, and ten present-day human genomes calculated as in Figure 17A. All estimates are based on DNA fragments that carry putative deamination-induced C to T substitutions. Bars indicate 95% confidence intervals.

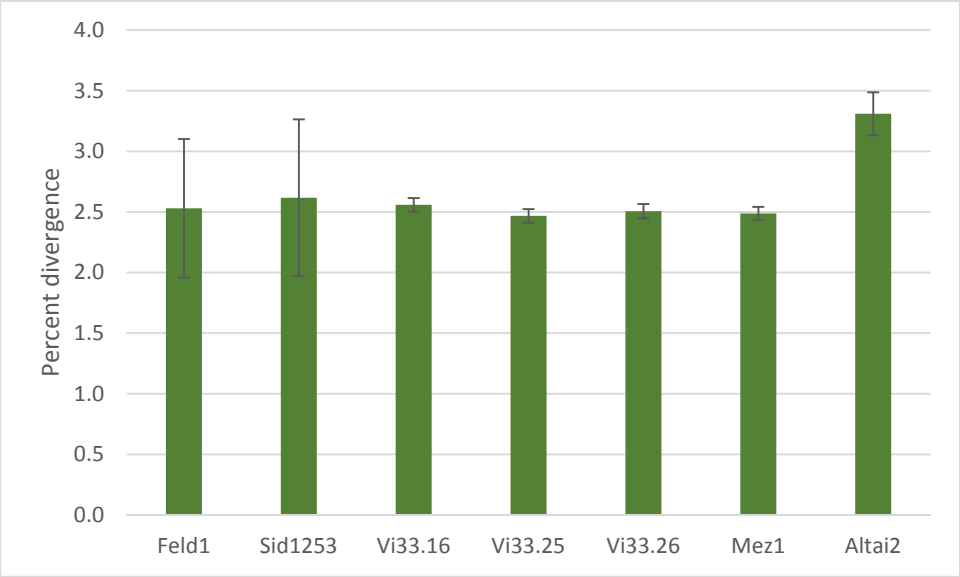

Figure 23. Autosomal divergences of seven low-coverage Neandertal genomes to the high-coverage *Altai 1* Neandertal genome calculated as in Figure 17A. All estimates are based on DNA fragments that carry putative deamination-induced C to T substitutions, except for *Mezmaiskaya1*. Bars indicate 95% confidence intervals.

Table 23. Divergences for the deaminated sequences. Divergence is given as the percent divergence of *Altai 2* along the branch to the human-chimpanzee ancestor from the high-coverage genomes given in the first column. Percent divergences and 95% CI (low and high) are given.

High-coverage genomes	Shared	Genome	*Altai 2*	% div	% div low	% div high
Altai 1	66250	2268	66250	3.31	3.13	3.49
Denisova3	61154	6876	61154	10.11	9.86	10.36
French	60777	8930	60777	12.81	12.52	13.10
Sardinian	60242	8772	60242	12.71	12.43	12.99
Han	60653	8829	60653	12.71	12.42	13.00
Dai	60738	8893	60738	12.77	12.49	13.05
Australian	60724	8499	60724	12.28	12.00	12.56
Papuan	60356	8782	60356	12.7	12.42	12.98
Mbuti	60636	9102	60636	13.05	12.78	13.32
Dinka	60171	8985	60171	12.99	12.71	13.27
Yoruba	60519	8956	60519	12.89	12.60	13.18
San	60487	9215	60487	13.22	12.94	13.50

Table 24. Number of nuclear sequences from the *Altai 2* that look derived or ancestral on various lineages. The total number of sequences covering the derived positions for each lineage are also given, as well as the fraction of sequences that are derived for each lineage.

Lineage	Ancestral	Derived	Total	Fraction derived
Denisova	2250	51	2301	0.02
PD Human - Denisova	403	35	438	0.08
Neandertal	453	1611	2064	0.78
Neandertal - Denisova	111	1104	1215	0.91
Neandertal - PD Human	43	539	582	0.93
All	541	23948	24489	0.98
PD Human	2941	41	2982	0.01

6.3.4 Animal contamination analyses based on mtDNA

A coverage distribution of the combined libraries aligned to the Vindija 33.16 Neandertal mtDNA reveals 3 major spikes in coverage: position 850-900, which falls into the 16S rRNA gene, 2200-3000, which falls into the 12S rRNA gene and 5600-5750, which falls into the tRNA-Asn gene (Figure 20). Both the 12S and 16S gene of the mtDNA are well known for being highly conserved, so this is a possible signal of animal contamination. We therefore recaptured the ten libraries using probes designed from 242 mammalian genomes (98). After capture, the filtered sequences were aligned using BLAST and ranked by taxanomic identification number. The ranking of each library as well as the combined sequences of all ten libraries shows that most sequences align to either modern humans or Neandertals (Table 25). In all ten libraries the taxon with the third most sequences aligned is the spotted hyena, which has 2.4 to 5.6 times more aligned sequences than the next mammal. When the ten libraries are combined, 6392 sequences align to either Neandertal or modern human. 477 sequences align to the spotted hyena, 4.8 times more than the 100 sequences that align to the woolly rhinoceros.

Table 25. BLAST alignment of each library from *Altai 2* after capture with 242 mammalian mtDNA genomes. The six taxons with the most hits are shown. Results after combining the libraries are also given.

L5483		**L5484**		**L5485**		**L5486**		**L5487**		**L5488**		**L5489**		**L5490**		**L5491**		**L5492**		**All combined**	
Sp.[1]	#[2]	Sp.	#	Sp.	#	Sp.	#	Sp.	#	Sp.	#	Sp.	#	Sp.	#	Sp.	#	Sp.	#	**Sp.**	**#**
Nea	439	Nea	690	Nea	118	Nea	750	Nea	410	Nea	1254	Nea	460	Nea	80	Nea	93	MH	72	**Nea**	**4344**
MH	286	MH	403	MH	44	MH	156	MH	108	MH	622	MH	236	MH	36	MH	85	Nea	50	**MH**	**2048**
Cro	42	Cro	45	Cro	42	Cro	78	Cro	34	Cro	112	Cro	35	Cro	33	Cro	6	Cro	50	**Cro**	**477**
Bos	8	Coe	12	Equ	8	Vul	32	Coe	14	Bos	20	Bos	10	Coe	11	Cer	2	Ovi	12	**Coe**	**100**
Urs	8	Bos	10	Equ	7	Cer	22	Cer	8	Coe	18	Urs	5	Equ	10	Vul	2	Equ	9	**Bos**	**81**
Cer	5	Cer	8	Urs	6	Cer	11	Cer	7	Cer	15	Vul	5	Bos	8	Bos	2	Coe	9	**Vul**	**71**

1. Sp. – Species Name:
   Nea: Neandertal, MH: Modern human, Cro: Spotted hyena (*Crocuta crocuta*), Bos: Cow (*Bos taurus*), Urs: Brown bear (*Ursus arctos*), Cer: Deer (*Cervus elaphus* or *albirostris*, or *Dama dama*), Coe: Woolly rhinoceros (*Coelodonta antiquitatis*), Equ: Horse (*Equus caballus* or *ovodovi*), Vul: Fox (*Vulpes vulpes*), Ovi: Sheep (*Ovis ammon hodgsoni*)
2. # – Number of sequences that aligned in the BLAST search to the specific species

6.3.5 Spotted hyena mtDNA analyses in *Altai 2*

Based on the all mammalian capture, the spotted hyena contributed the most DNA (aside from humans) to the sequences in the *Altai 2* bone. Therefore the ten libraries of the *Altai 2* were also captured with and aligned to the spotted hyena mtDNA genome (NC020670, (97)) as described in section 4.2.7. After filtering (see section 4.2.3), the coverage across the spotted hyena genome showed significant peaks in conserved regions, due to the large amount of Neandertal mtDNA (Figure 24). Therefore regions that map to the human mtDNA genome were removed (see section 4.2.7), reducing the spotted hyena mtDNA genome from 17,138 to 14,478 positions. After human-mappable regions were removed, the average coverage for each of the ten libraries varied from 8.4 to 103.7-fold, with a combined coverage of 543-fold (Table 20, Figure 24). After filtering for of C to T sequences, the coverage drops to 85-fold (Table 20).

Sequences from the ten libraries were combined and a consensus of the spotted hyena sequences of both before and after C to T filtering were called, requiring at least 5-fold coverage and 80% consensus support. 13,724 positions of the 14,478 possible were used to call a consensus for all sequences, while 12,485 positions were used to call a consensus of the C to T filtered sequences.

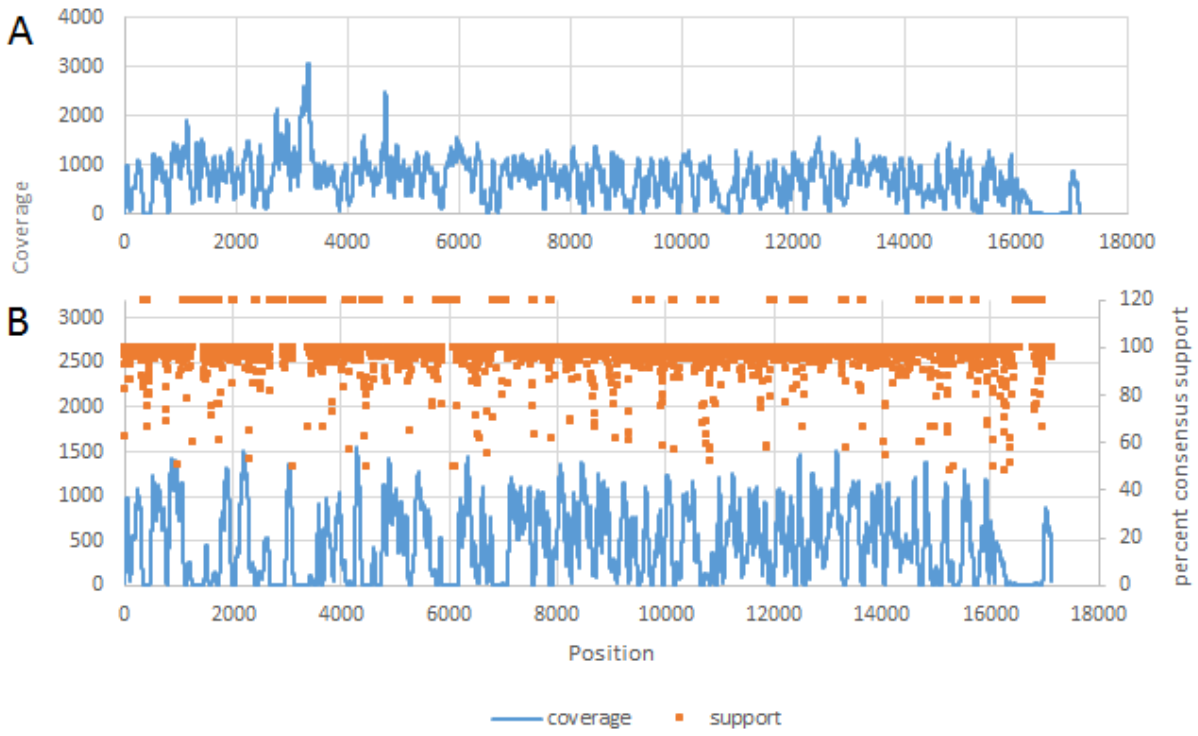

Figure 24. Coverage and consensus support across the spotted hyena captured and aligned sequences of *Altai 2*. Positions with (A) and without (B) human-mapped regions are shown. Positions with a coverage below 5-fold are shown with a consensus support of 120%.

The library made from SP3388, the spotted hyena bone from the same gallery and layer of Denisova Cave as *Altai 2*, was captured using the same spotted hyena probes used to capture the libraries from *Altai 2*. After sequencing, the sequences were filtered and aligned and used to create a consensus sequence as described in section 4.2.7. The average coverage is 1596-fold (Figure 25). The consensus is made up of 17,093 of the 17,138 positions possible when a minimum of 5-fold coverage and consensus support of 80% are required. The 3' C to T percent is 35.0% (95% CI: 34.6-35.4%) and the 5' C to T percent is 31.0% (95% CI: 30.7-31.3%).

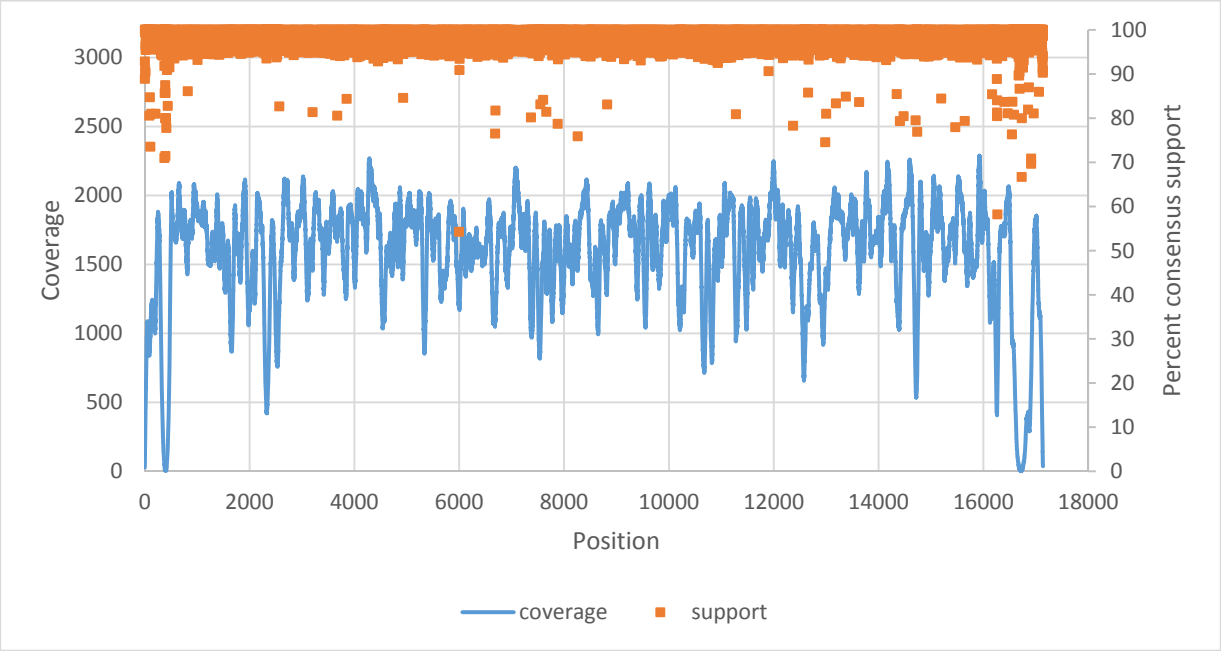

Figure 25. Coverage and consensus support across the spotted hyena captured and aligned sequences of SP3388.

The complete mtDNA genomes of two Pleistocene spotted hyenas from Coumere Cave, France (CC8 (NC020670) and CC9 (JF894379), (97)), an extant spotted hyena (Kira (zoo name), JF894377, (97)), an extant striped hyena (Cerza (zoo name), NC020669, (97)), a mongoose (NC006835), and the partial mtDNAs of the *Altai 2* spotted hyena sequences and the Denisova Cave spotted hyena from SP3388 (see section 4.2.7 for more details), were used to calculate the pairwise differences among hyenas (Table 26). The two Pleistocene spotted hyenas, CC8 and CC9, have identical mtDNA genomes. The Denisova Cave spotted hyena from bone SP3388 is the closest to the two French Pleistocene hyenas with 21 differences to both, while carrying 82 differences to the extant spotted hyena, JF894377. The spotted hyena sequence consensus gained

from the *Altai 2* Neandertal bone carries between 382 and 404 differences to the other spotted hyenas, about five times higher than the second largest number of differences within spotted hyenas and three times less than the number of differences between striped and spotted hyena. The number of differences between striped and spotted hyenas does not differ greatly between the spotted hyena individuals, including the *Altai 2* spotted hyena. Filtering for C to T sequences does not have a significant effect.

Table 26. Pairwise differences among five spotted hyenas, one striped hyena and one mongoose. Number of differences for all sequences are shown. The number of differences for sequences filtered for C to T differences are shown in parentheses. The accession number of the individual is given where available.

#	Individual	Group	1	2	3	4	5	6
1	*Altai 2*	Spotted hyena						
2	NC020670 (CC8 ref genome)	Spotted hyena	383 (349)					
3	JF894379 (CC9)	Spotted hyena	383 (349)	0				
4	SP3388	Spotted hyena	382 (350)	21 (19)	21 (19)			
5	JF894377 (Kira)	Spotted hyena	404 (370)	81 (59)	81 (59)	82 (62)		
6	NC020669 (Cerza)	Striped hyena	1297 (1191)	1292 (1167)	1292 (1167)	1299 (1174)	1294 (1176)	
7	NC006835	Mongoose	2245 (2062)	2239 (2045)	2239 (2045)	2233 (2041)	2233 (2043)	2204 (2011)

A Bayesian phylogenetic tree was constructed of these five spotted hyenas, the striped hyena and the mongoose as an outgroup (see section 4.2.9 for more details) (Figure 26). The SP3388 Denisova Cave spotted hyena falls together with the CC8 and CC9 spotted hyenas from France (posterior probability of 1). The spotted hyena sequences from the *Altai 2* falls together with the other spotted hyenas with a posterior probability of 1, but is the most diverged (Figure 26).

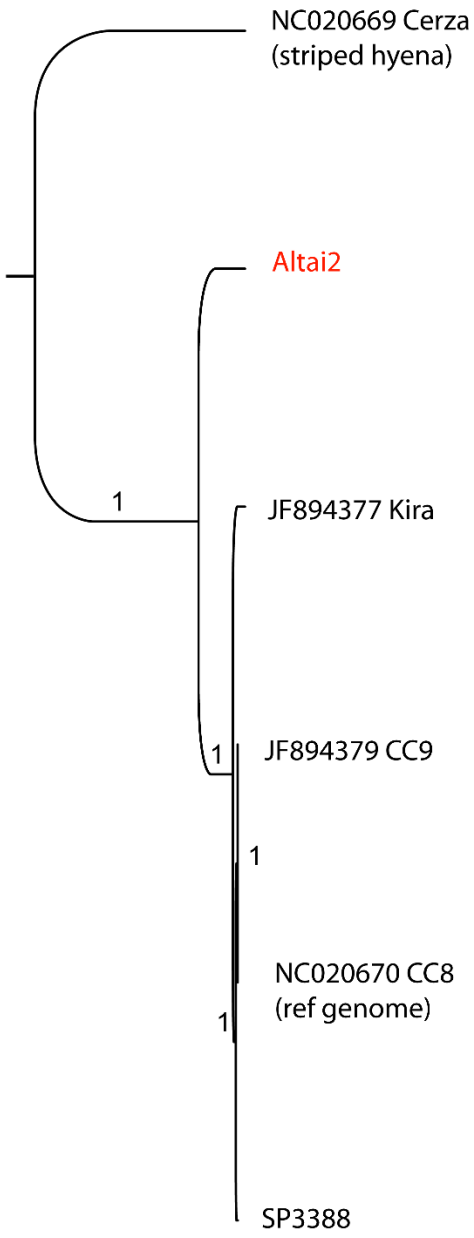

Figure 26. A Bayesian mtDNA tree of the *Altai 2* spotted hyena consensus (not filtered for C to T changes), the spotted hyena from Denisova Cave (SP3388), three spotted hyenas (CC8, CC9, Kira), one striped hyena (Cerza) and a mongoose as outgroup (not shown). Posterior values are given. The tree is based on 12,849 positions.

A pairwise comparison was done of 234 bps of the cytochrome b gene of the mtDNA genomes of 57 spotted hyenas, 2 aardwolves, 18 striped hyenas, 6 brown hyenas and one mongoose (see section 4.2.9 for method and accession numbers). The 57 spotted hyenas are represented by 13 sequences as many of the individuals carry the same 234 bp sequence. They have been labeled as sequences A-M (Table 27). The most differences (9 to 11 differences) are between sequence I (the sequence shared by three from teeth from Da'an Cave, China (14) and one individual from the Geographical Society cave near Vladivostok, Russia, (13)) and sequences B-G and J-M. The *Altai 2* Neandertal spotted hyena is the only individual with sequence H and only differs from sequence I at one position. The two aardwolves each have a distinct sequence (N and O), the 6 brown hyenas all share one sequence (sequence P), and the 18 striped hyenas reduce to 3 sequences (Q, R and S) (see Table 27).

A Bayesian tree was constructed of the 234 bps of the cytochrome b gene of the mtDNA genomes of the 13 spotted hyenas sequences (A-M), the 2 Aardwolf sequences (N and O), the brown hyenas sequence (P), the 3 striped hyena sequences (Q, R and S) and one mongoose as an outgroup (see section 4.2.9 for method and Appendix Table 1 for accession numbers) (Figure 27). Clades within the larger spotted hyena clade are grouped by color if the posterior probability is above 0.94. Thus there are four major clades (green, yellow, pink and blue). Spotted hyena samples from Africa are from living individuals or from museum samples that are at most 150 years old, and therefore represent present-day spotted hyenas. Samples from outside Africa are from spotted hyenas from the Pleistocene.

The pink clade is found exclusively in southern Africa in present-day spotted hyenas with one individual coming from farther north, in Sudan. The yellow clade is found only in Pleistocene spotted hyenas from Europe (Figure 27). The blue clade has the most individuals and is present in both present-day northern-Africa as well as Pleistocene Europe (Figure 27). Two of the spotted hyena individuals from Denisova Cave (the individual sequenced here, SP3388, and the ~42k year old AJ809327, (141)) fall into the larger blue clade. The green clade is made up of five individuals, three from the teeth from Da'an Cave, China (14), one from the Geographical Society cave near Vladivostok, Russia (13), and the consensus sequence of the spotted hyena sequences from the *Altai 2* bone from Denisova Cave.

Table 27. Pairwise differences in 234 bps of the cytochrome b gene among the mtDNA genomes of 57 spotted hyenas, 2 aardwolves, 18 striped hyenas, 6 brown hyenas and one mongoose. Individuals with no sequence differences between them are grouped and highlighted with the same color.

	Accession #	Group	Sequence Type	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85
1	AJ809332	Spotted Hyena	A																																																																																				
2	AJ809331	Spotted Hyena	A	0																																																																																				
3	AJ809330	Spotted Hyena	A	0	0																																																																																			
4	AJ809329	Spotted Hyena	A	0	0	0																																																																																		
5	AJ809328	Spotted Hyena	A	0	0	0	0																																																																																	
6	AJ809327	Spotted Hyena	B	5	5	5	5	5																																																																															
7	AJ809326	Spotted Hyena	B	5	5	5	5	5	0																																																																														
8	AJ809325	Spotted Hyena	B	5	5	5	5	5	0	0																																																																													
9	AJ809324	Spotted Hyena	B	5	5	5	5	5	0	0	0																																																																												
10	DQ157554	Spotted Hyena	B	5	5	5	5	5	0	0	0	0																																																																											
11	DQ157592	Spotted Hyena	B	5	5	5	5	5	0	0	0	0	0																																																																										
12	DQ157575	Spotted Hyena	B	5	5	5	5	5	0	0	0	0	0	0																																																																									
13	DQ157565	Spotted Hyena	B	5	5	5	5	5	0	0	0	0	0	0	0																																																																								
14	DQ157564	Spotted Hyena	B	5	5	5	5	5	0	0	0	0	0	0	0	0																																																																							
15	DQ157562	Spotted Hyena	B	5	5	5	5	5	0	0	0	0	0	0	0	0	0																																																																						
16	DQ157559	Spotted Hyena	B	5	5	5	5	5	0	0	0	0	0	0	0	0	0	0																																																																					
17	DQ157558	Spotted Hyena	B	5	5	5	5	5	0	0	0	0	0	0	0	0	0	0	0																																																																				
18	DQ157567	Spotted Hyena	B	5	5	5	5	5	0	0	0	0	0	0	0	0	0	0	0	0																																																																			
19	DQ157556	Spotted Hyena	B	5	5	5	5	5	0	0	0	0	0	0	0	0	0	0	0	0	0																																																																		
20	NC020670	Spotted Hyena	B	5	5	5	5	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0																																																																	
21	JF894379	Spotted Hyena	B	5	5	5	5	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																																																																
22	AF153053	Spotted Hyena	B	5	5	5	5	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																																																															
23	AY048811	Spotted Hyena	B	5	5	5	5	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																																																														
24	AJ809323	Spotted Hyena	C	4	4	4	4	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1																																																													
25	AJ809322	Spotted Hyena	C	4	4	4	4	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0																																																												
26	AJ809321	Spotted Hyena	C	4	4	4	4	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0																																																											
27	AJ809320	Spotted Hyena	C	4	4	4	4	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0																																																										
28	AJ809319	Spotted Hyena	C	4	4	4	4	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0																																																									
29	AJ809318	Spotted Hyena	C	4	4	4	4	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0																																																								
30	DQ157574	Spotted Hyena	C	4	4	4	4	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0																																																							
31	DQ157563	Spotted Hyena	D	4	4	4	4	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2																																																						
32	DQ157571	Spotted Hyena	E	7	7	7	7	7	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3																																																					
33	DQ157557	Spotted Hyena	F	6	6	6	6	6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	1																																																				
34	DQ157573	Spotted Hyena	G	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	4	7	6																																																			
35	DQ157572	Spotted Hyena	G	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	4	7	6	0																																																		
36	DQ157570	Spotted Hyena	G	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	4	7	6	0	0																																																	
37	DQ157569	Spotted Hyena	G	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	4	7	6	0	0	0																																																
38	DQ157568	Spotted Hyena	G	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	4	7	6	0	0	0	0																																															
39	DQ157566	Spotted Hyena	G	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	4	7	6	0	0	0	0	0																																														
40	DQ157561	Spotted Hyena	G	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	4	7	6	0	0	0	0	0	0																																													
41	DQ157560	Spotted Hyena	G	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	4	7	6	0	0	0	0	0	0	0																																												
42	AF153051	Spotted Hyena	G	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	4	7	6	0	0	0	0	0	0	0	0																																											
43	AY048812	Spotted Hyena	G	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	4	7	6	0	0	0	0	0	0	0	0	0																																										
44	AY048807	Spotted Hyena	G	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	4	7	6	0	0	0	0	0	0	0	0	0	0																																									
45	AY048808	Spotted Hyena	G	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	4	7	6	0	0	0	0	0	0	0	0	0	0	0																																								
46	AY048809	Spotted Hyena	G	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	4	7	6	0	0	0	0	0	0	0	0	0	0	0	0																																							
47	AY048804	Spotted Hyena	G	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	4	7	6	0	0	0	0	0	0	0	0	0	0	0	0	0																																						
48	AY048805	Spotted Hyena	G	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	4	7	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0																																					
49	Altai2	Spotted Hyena	H	6	6	6	6	6	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	8	8	8	8	8	8	8	8	9	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8																																				
50	DQ157555	Spotted Hyena	I	7	7	7	7	7	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	9	9	9	9	9	9	9	9	10	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	1																																			
51	KC117381	Spotted Hyena	I	7	7	7	7	7	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	9	9	9	9	9	9	9	9	10	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	1	0																																		
52	KC117380	Spotted Hyena	I	7	7	7	7	7	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	9	9	9	9	9	9	9	9	10	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	1	0	0																																	
53	KC117379	Spotted Hyena	I	7	7	7	7	7	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	9	9	9	9	9	9	9	9	10	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	1	0	0	0																																
54	SP3388	Spotted Hyena	J	6	6	6	6	6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	3	2	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	10	11	11	11	11																															
55	JF894377	Spotted Hyena	K	6	6	6	6	6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	3	2	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	10	11	11	11	11	2																														
56	AF153052	Spotted Hyena	L	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	5	8	7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	9	10	10	10	10	7	7																													
57	AY048806	Spotted Hyena	L	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	5	8	7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	9	10	10	10	10	7	7	0																												
58	AY048810	Spotted Hyena	M	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	4	5	4	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	6	7	7	7	7	6	6	3	3																											
59	AY928679	Aardwolf	N	22	22	22	22	22	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	22	22	22	22	22	22	22	22	25	24	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	21	21	21	21	24	24	21	21	22																										
60	AY048791	Aardwolf	O	25	25	25	25	25	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	25	25	25	25	25	25	25	25	26	25	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	21	20	20	20	20	27	25	22	22	23	13																									
61	DQ157591	Brown Hyena	P	31	31	31	31	31	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	29	29	29	29	29	29	29	31	30	29	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	30	30	30	30	29	31	31	31	30	31	30																								
62	DQ157590	Brown Hyena	P	31	31	31	31	31	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	29	29	29	29	29	29	29	31	30	29	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	30	30	30	30	29	31	31	31	30	31	30	0																							
63	DQ157589	Brown Hyena	P	31	31	31	31	31	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	29	29	29	29	29	29	29	31	30	29	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	30	30	30	30	29	31	31	31	30	31	30	0	0																						
64	DQ157588	Brown Hyena	P	31	31	31	31	31	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	29	29	29	29	29	29	29	31	30	29	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	30	30	30	30	29	31	31	31	30	31	30	0	0	0																					
65	AY048790	Brown Hyena	P	31	31	31	31	31	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	29	29	29	29	29	29	29	31	30	29	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	30	30	30	30	29	31	31	31	30	31	30	0	0	0	0																				
66	AY048789	Brown Hyena	P	31	31	31	31	31	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	29	29	29	29	29	29	29	31	30	29	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	30	30	30	30	29	31	31	31	30	31	30	0	0	0	0	0																			
67	DQ157587	Striped Hyena	Q	31	31	31	31	31	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	31	31	31	31	31	31	31	32	31	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	30	30	30	30	33	33	31	31	30	28	29	17	17	17	17	17	17																		
68	DQ157586	Striped Hyena	Q	31	31	31	31	31	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	31	31	31	31	31	31	31	32	31	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	30	30	30	30	33	33	31	31	30	28	29	17	17	17	17	17	17	0																	
69	DQ157585	Striped Hyena	Q	31	31	31	31	31	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	31	31	31	31	31	31	31	32	31	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	30	30	30	30	33	33	31	31	30	28	29	17	17	17	17	17	17	0	0																
70	DQ157576	Striped Hyena	Q	31	31	31	31	31	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	31	31	31	31	31	31	31	32	31	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	30	30	30	30	33	33	31	31	30	28	29	17	17	17	17	17	17	0	0	0															
71	DQ157579	Striped Hyena	Q	31	31	31	31	31	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	31	31	31	31	31	31	31	32	31	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	30	30	30	30	33	33	31	31	30	28	29	17	17	17	17	17	17	0	0	0	0														
72	DQ157580	Striped Hyena	Q	31	31	31	31	31	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	31	31	31	31	31	31	31	32	31	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	30	30	30	30	33	33	31	31	30	28	29	17	17	17	17	17	17	0	0	0	0	0													
73	DQ157581	Striped Hyena	Q	31	31	31	31	31	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	31	31	31	31	31	31	31	32	31	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	30	30	30	30	33	33	31	31	30	28	29	17	17	17	17	17	17	0	0	0	0	0	0												
74	DQ157582	Striped Hyena	Q	31	31	31	31	31	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	31	31	31	31	31	31	31	32	31	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	31	30	30	30	30	33	33	31	31	30	28	29	17	17	17	17	17	17	0	0	0	0	0	0	0											
75	DQ157584	Striped Hyena	R	32	32	32	32	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	32	32	32	32	32	32	32	33	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	31	31	31	31	34	34	32	32	31	29	30	16	16	16	16	16	16	1	1	1	1	1	1	1	1										
76	DQ157578	Striped Hyena	R	32	32	32	32	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	32	32	32	32	32	32	32	33	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	31	31	31	31	34	34	32	32	31	29	30	16	16	16	16	16	16	1	1	1	1	1	1	1	1	0									
77	DQ157583	Striped Hyena	R	32	32	32	32	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	32	32	32	32	32	32	32	33	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	31	31	31	31	34	34	32	32	31	29	30	16	16	16	16	16	16	1	1	1	1	1	1	1	1	0	0								
78	DQ157577	Striped Hyena	R	32	32	32	32	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	32	32	32	32	32	32	32	33	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	31	31	31	31	34	34	32	32	31	29	30	16	16	16	16	16	16	1	1	1	1	1	1	1	1	0	0	0							
79	AY048787	Striped Hyena	R	32	32	32	32	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	32	32	32	32	32	32	32	33	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	31	31	31	31	34	34	32	32	31	29	30	16	16	16	16	16	16	1	1	1	1	1	1	1	1	0	0	0	0						
80	AF153054	Striped Hyena	R	32	32	32	32	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	32	32	32	32	32	32	32	33	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	31	31	31	31	34	34	32	32	31	29	30	16	16	16	16	16	16	1	1	1	1	1	1	1	1	0	0	0	0	0					
81	AY928678	Striped Hyena	S	32	32	32	32	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	32	32	32	32	32	32	32	33	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	31	31	31	31	34	34	32	32	31	29	30	18	18	18	18	18	18	1	1	1	1	1	1	1	1	2	2	2	2	2	2				
82	NC020669	Striped Hyena	S	32	32	32	32	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	32	32	32	32	32	32	32	33	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	31	31	31	31	34	34	32	32	31	29	30	18	18	18	18	18	18	1	1	1	1	1	1	1	1	2	2	2	2	2	2	0			
83	AF153055	Striped Hyena	S	32	32	32	32	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	32	32	32	32	32	32	32	33	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	31	31	31	31	34	34	32	32	31	29	30	18	18	18	18	18	18	1	1	1	1	1	1	1	1	2	2	2	2	2	2	0	0		
84	AY048788	Striped Hyena	S	32	32	32	32	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	32	32	32	32	32	32	32	33	32	33	33	33	33	33	33	33	33	33	33	33	33	33	33	33	32	31	31	31	31	34	34	32	32	31	29	30	18	18	18	18	18	18	1	1	1	1	1	1	1	1	2	2	2	2	2	2	0	0	0	
85	NC006835	Mongoose		41	41	41	41	41	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	39	39	39	39	39	39	39	39	40	39	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	41	40	40	40	40	41	39	41	41	40	35	36	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	38	38	38	38	

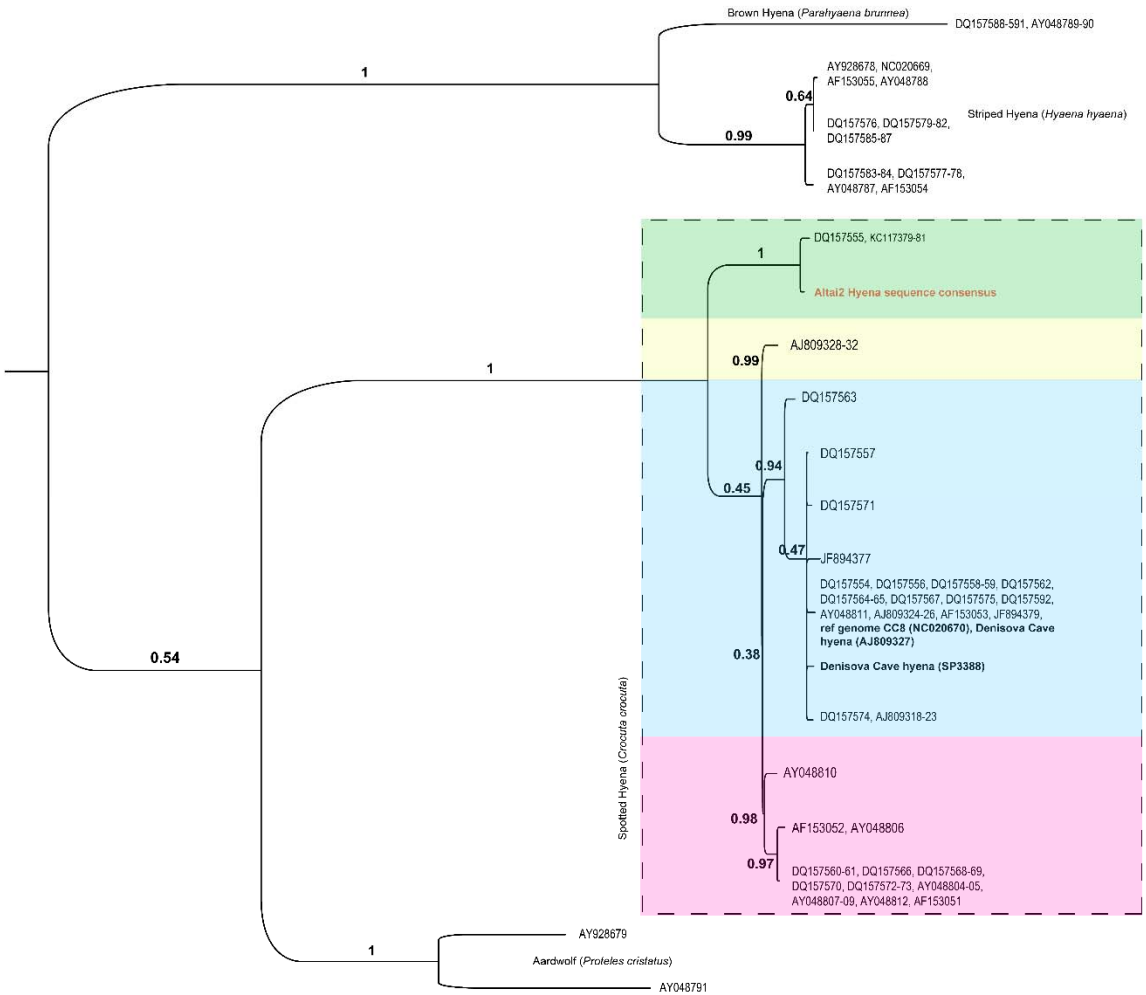

Figure 27. A Bayesian tree of the 234 bps of the cytochrome b gene in the mtDNA of 13 spotted hyena sequences (based on 57 individuals), two aardwolves, three striped hyena sequences (based on 18 individuals), one brown hyenas sequence (based on six individuals) and one mongoose (outgroup: not shown). Posterior values are shown if above 0.8 or at important nodes. Spotted hyenas were grouped together by color if their clade had a posterior probability above 0.94. The colors correspond to the colors in the map in Figure 28. Accession numbers of all individuals are given.

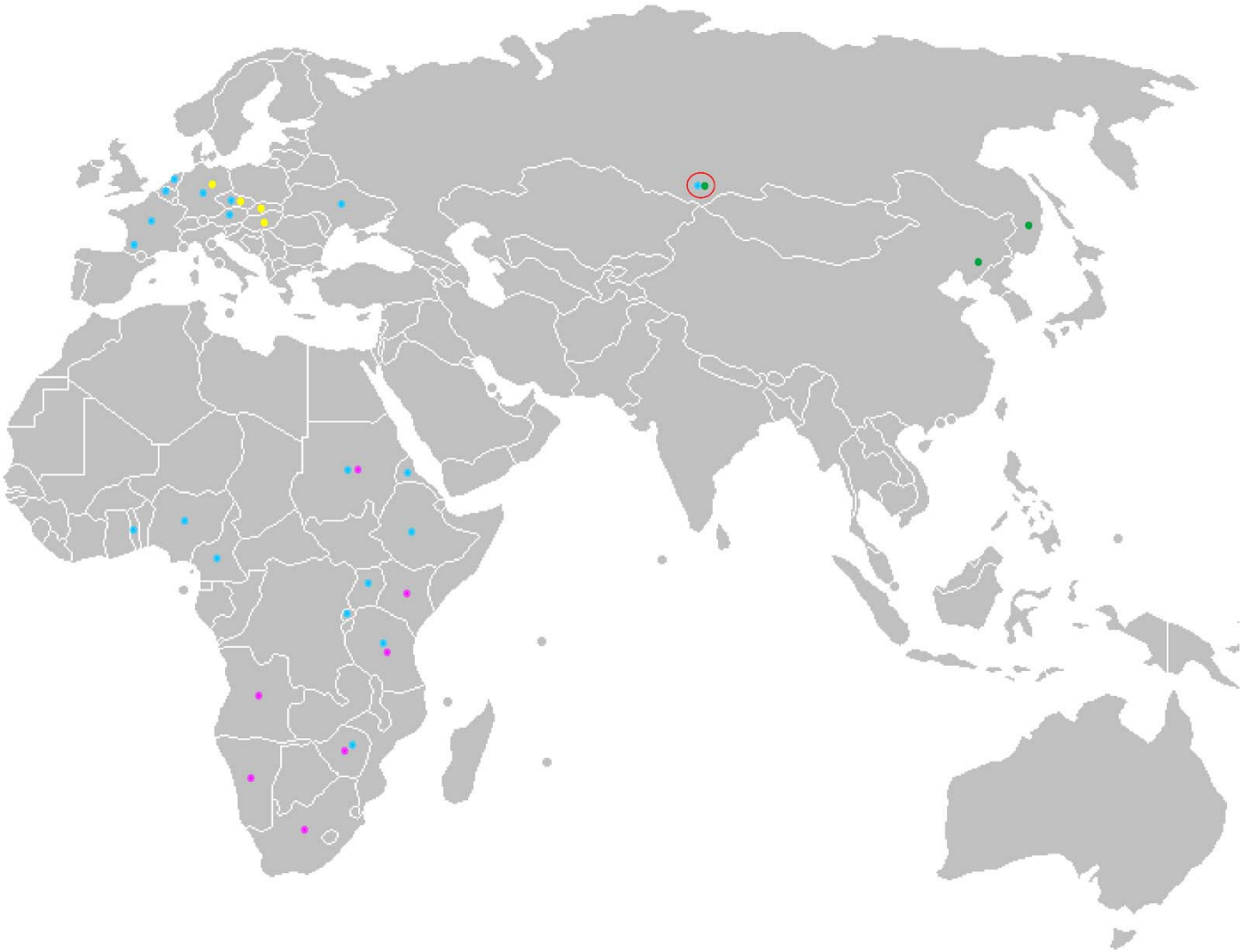

Figure 28. A map of the old world showing the spotted hyena for which the 234 bps of the cytochrome b gene in the mtDNA of 13 sequences from 57 individuals were determined. Colors are based on the groupings in Figure 27. Denisova Cave is marked with a large red circle.

We used a newly described method (15) to wash bone powder from the *Altai 2* bone with a phosphate wash and subsequently extracted after the phosphate wash. The sequences gained in the phosphate wash have almost the same average coverage for both the human capture/alignment and spotted hyena capture/alignment (Table 28). The sequences extracted after the phosphate wash on the other hand have twice as many spotted hyena sequences as human sequences (Table 28). The percent spotted hyena contamination is also significantly higher in the extract after phosphate washing than in the phosphate wash (Table 28). These observations are confirmed in the all-mammalian capture (section 6.3.4), where the number of sequences that align to the spotted hyena are less than one tenth the amount aligned to human or Neandertal for the sequences from the phosphate wash. After the phosphate wash, the same number of sequences align to spotted hyena and Neandertal, with ~25% more aligning to modern human (Table 28).

The spotted hyena mtDNA captures have a higher C to T rate than the human mtDNA captures, even when the human mtDNA captured C to T rate is filtered for C to Ts on the opposite fragment ends (3' filtered) (Table 29). We compared the sequence lengths of sequences captured and aligned to the spotted hyena mtDNA genome and sequences captured with human mtDNA probes and aligned to the Neandertal mtDNA genome. While the two sequences sets have approximately the same peak at 45 bps, the Neandertal sequences are longer and the spotted hyena sequences are shorter (Figure 29).

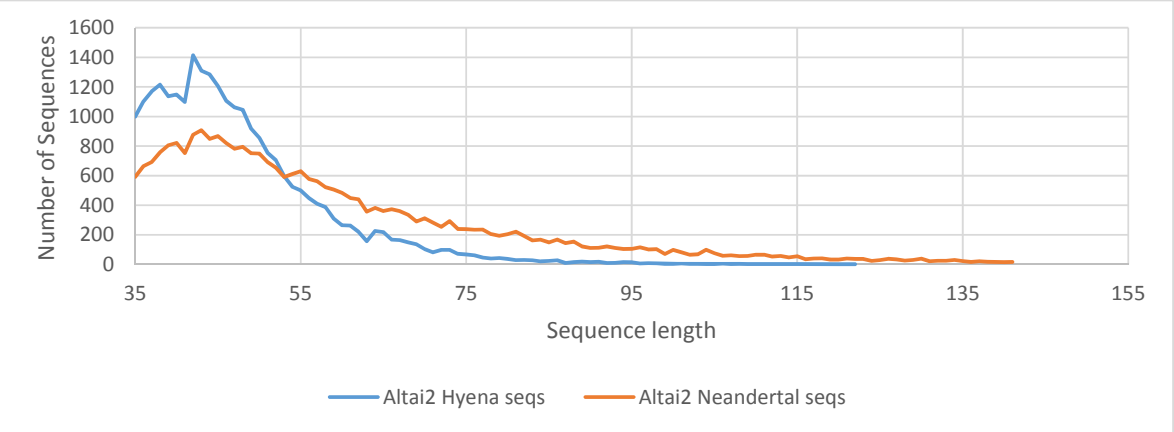

Figure 29. Length distribution differences between spotted hyena and Neandertal sequences in the *Altai 2* bone. Only sequences with C to T differences to the reference genome are shown here.

Table 28. Summary of relevant data from phosphate treatment of bone powder from the *Altai 2* bone. 'Avg cov exhausted' refers to the average coverage in this library for either human/Neandertal or spotted hyena capture/alignment if the library were sequenced to exhaustion. PD-human refers to present-day human. Percent spotted hyena contamination was calculated for the Neandertal sequences as described in section 4.2.5; 95% CI are shown in italics. The top hits in BLAST were calculated after the libraries were captured using the all-mammalian probe set as described in section 4.2.3.

Parent library	**Treatment**	**Avg cov exhausted Neandertal**	**Avg cov exhausted spotted hyena**	**% PD-human contamination in Nea seqs**	**% spotted hyena contamination in Nea seqs**	**Top hits in BLAST**	
						Species	**# seqs align**
A9232	$NaHPO_4$ wash	10.5	9.2	37.9 (*31.5-44.8*)	16.2 *(7.7-31.1)*	Neandertal	93
						Modern human	85
						Crocuta crotuta	6
						Dama dama	2
A9233	After $NaHPO_4$ wash	36.9	95.8	45.7 (*43.3-48.1*)	45.9 *(42.9-49.0)*	Modern human	72
						Neandertal	50
						Crocuta crotuta	50
						Ovis ammon	12

Table 29. Terminal C to T substitutions. "3' filtered" and "5' filtered" refer to fragments that carry C to T substitutions at their 3'- and 5'-ends, respectively. Note that libraries with UDG treatment have a lower 5' C to T rate than 3' rate as described in (8).

		Human mtDNA captures				Spotted spotted hyena mtDNA captures	
		All		**5' filtered**	**3' filtered**	**All (regions with no human seq)**	
Parent Library	**Treatment**	**5' C>T percent**	**3' C>T percent**	**3' C>T percent**	**5' C>T percent**	**5' C>T percent**	**3' C>T percent**
L9366	UDG	8.1 (*7.8-8.6*)	31.9 (*31.0-32.8*)	8.9 (*7.1-11.2*)	37.9 (*31.1-45.3*)	18.6 (*17.4-19.8*)	47.4 (*45.9-48.9*)
L9367	UDG	7.9 (*7.4-8.5*)	31.5 (*30.5-32.6*)	10.5 (*8.2-13.3*)	38.4 (*31.0-46.4*)	19.8 (*17.6-22.2*)	47.8 (*45.0-50.5*)
L9575	U-selected	34.9 (*33.3-36.6*)	94.7 (*94.1-95.2*)	NA	NA	51.6 (*49.1-54.0*)	97.2 (*96.6-97.7*)
L9576	U-selected	37.6 (*36.0-39.3*)	94.9 (*94.3-95.4*)	NA	NA	51.2 (*48.9-53.5*)	97.0 (*96.4-97.5*)
L9580	U-depleted	3.8 (*3.5-4.2*)	3.9 (*3.4-4.5*)	NA	NA	9.7 (*8.9-10.6*)	18.1 (*16.5-19.8*)
L9581	U-depleted	3.2 (*2.9-3.6*)	3.4 (*2.9-4.0*)	NA	NA	9.3 (*8.5-10.1*)	14.3 (*13.0-15.7*)
L9467	UDG	9.9 (*8.6-11.4*)	52.5 (*49.2-55.7*)	14.0 (*8.7-21.8*)	60.0 (*40.7-76.6*)	16.5 (*15.2-17.9*)	66.3 (*64.2-68.3*)
A9231	UDG	10.5 (*9.0-12.2*)	52.4 (*48.7-56.0*)	7.8 (*3.6-16.0*)	60.0 (*31.3-83.2*)	16.5 (*15.0-18.2*)	66.6 (*64.1-69.0*)
A9232	non-UDG NaHPO4 wash	22.5 (*17.8-28.1*)	32.5 (*25.7-40.1*)	21.4 (*7.6-47.6*)	37.9 (*10.8-60.3*)	41.6 (*37.6-45.8*)	50.7 (*45.9-55.5*)
A9233	non-UDG after NaHPO4 wash	27.6 (*25.7-29.6*)	38.0 (*35.5-40.6*)	31.9 (*24.2-40.8*)	38.4 (*29.4-48.2*)	41.5 (*40.2-42.8*)	52.8 (*51.2-54.4*)

6.3.6 Spotted hyena mtDNA analyses in other Denisova Cave individuals

Shotgun sequences from three Denisovans (*Denisova 3*, *4* and *8*) (8), three Neandertals (*Altai 1*, *Altai 2* and *Mezmaiskaya1*) (10), one early modern human (Ust Ishim) (106) and one present-day human were mapped to the mtDNA genomes of the spotted hyena, an arctic ringed seal (NC008428), a cave bear (NC011112) and human. The ratio of hyena to human mapped sequences for *Denisova 3*, *Denisova 4*, *Altai 1* and *Altai 2* are at least one order of magnitude higher than the ratios of seal or cave bear to human (Table 30). There is also a significant difference between the number of sequences that align to hyena and seal or cave bear for *Denisova 3*, *Denisova 4*, *Altai 1* and *Altai 2* when compared to the present-day human or Ust-Ishim. None of the other individuals show this significant difference (Table 31). When looking only at perfectly aligned sequences, *Denisova 3*, *4*, *Altai 1* and *Altai 2* again have at least an order of magnitude more sequences align to hyena than to seal or cave bear. The other individuals have either no sequences or very few, as is the case with Mezmaiskaya where four sequences align to the hyena (Table 32).

We aligned the consensus of the mtDNA spotted hyena sequences of *Denisova 3*, *4* and *Altai 1* each to three published spotted hyenas, a striped hyena, a mongoose as an outgroup as well as the partial mtDNAs of the *Altai 2* spotted hyena sequences and the Denisova Cave spotted hyena from SP3388 (see section 4.2.7 for more details) (Figure 30). The *Altai 2* spotted hyena consensus falls outside the variation of the other spotted hyena mtDNA genomes in every case. Based on 7,507 positions, *Altai 1* also falls outside the variation of the other spotted hyenas (with the exception of *Altai 2*). Both of the Denisovans fall within the variation of the western Pleistocene spotted hyenas, the spotted hyena from Denisova Cave and an extant spotted hyena (based on 8,592 positions for *Denisova 3* and 468 positions for *Denisova 4*) (Figure 30).

The *Denisova 3* hyena sequence covers 143 of the 234 positions of the cytochrome b gene discussed in section 6.3.5. These 143 positions are identical to the JF894377 and CC8 spotted hyenas. The *Altai 1* hyena sequence covers only 62 of these positions and is also identical to the JF894377 and CC8 spotted hyenas at these positions. The *Denisova 4* hyena sequence does not overlap with any of the 243 cytochrome b positions.

Table 30. Number of sequences that align to spotted hyena, arctic ribbon seal, cave bear or human mtDNA genomes for three Denisovans (*Denisova 3*,*4* and *8*), three Neandertals (*Altai1*, *2 Mezmaiskaya1*), one early modern human (Ust Ishim) and one present-day human.

Individual	# seq align to hyena	# seq align to seal	# seq align to cave bear	# seq align to human	ratio hyena/human	ratio seal/human	ratio cave bear/human
Present-day human	33	43	29	391,582	8.43E-05	1.10E-04	7.41E-05
Denisova 3	1262	120	82	481,784	2.62E-03	2.49E-04	1.70E-04
Denisova 4	142	8	8	11,001	1.29E-02	7.27E-04	7.27E-04
Denisova 8	2	5	3	17,503	1.14E-04	2.86E-04	1.71E-04
Altai 1	1583	112	104	465,621	3.40E-03	2.41E-04	2.23E-04
Altai 2	1009	25	31	940	1.07	2.66E-02	3.30E-02
Mezmaiskaya1	6	2	2	64,746	9.27E-05	3.09E-05	3.09E-05
Ust Ishim	42	68	47	130,820	3.21E-04	5.20E-04	3.59E-04

Table 31. P-values of a fisher exact test of number of sequences shown in Table 30. Comparison was between either a present-day human and archaic individuals or the Ust Ishim early modern human and archaic individuals.

Individual 1	Individual 2	Hyena vs. Seal	Hyena vs. Cave bear	Seal vs. Cave bear
Present-day human	*Denisova 3*	< 0.0009	< 0.0009	1
Present-day human	*Denisova 4*	< 0.0009	< 0.0009	0.58
Present-day human	*Denisova 8*	0.69	0.66	1
Present-day human	*Altai 1*	< 0.0009	< 0.0009	0.28
Present-day human	*Altai 2*	< 0.0009	< 0.0009	0.11
Present-day human	*Mezmaiskaya1*	0.14	0.14	1
Present-day human	Ust Ishim	0.54	0.51	1
Ust Ishim	*Denisova 3*	< 0.0009	< 0.0009	1
Ust Ishim	*Denisova 4*	< 0.0009	< 0.0009	0.59
Ust Ishim	*Denisova 8*	0.71	1	1
Ust Ishim	*Altai 1*	< 0.0009	< 0.0009	0.25
Ust Ishim	*Altai 2*	< 0.0009	< 0.0009	0.1
Ust Ishim	*Mezmaiskaya1*	0.06	0.16	1

Table 32. As Table 30, except the number of sequences reported here are only sequences that align perfectly.

Individual	# seq align to hyena	# seq align to seal	# seq align to cave bear	# seq align to human	ratio hyena/human	ratio seal/human	ratio ursus/human
Present-day human	0	0	0	294,538	0	0	0
Denisova 3	900	1	1	94,588	9.51E-03	1.06E-05	1.06E-05
Denisova 4	92	0	0	3,255	2.83E-02	0	0
Denisova 8	0	0	0	5,417	0	0	0
Altai 1	541	1	11	162,907	3.32E-03	6.14E-06	6.75E-05
Altai 2	29	0	1	95	0.31	0	1.05E-02
Mezmaiskaya1	4	0	0	27,748	1.44E-04	0	0
Ust Ishim	0	0	0	111,780	0	0	0

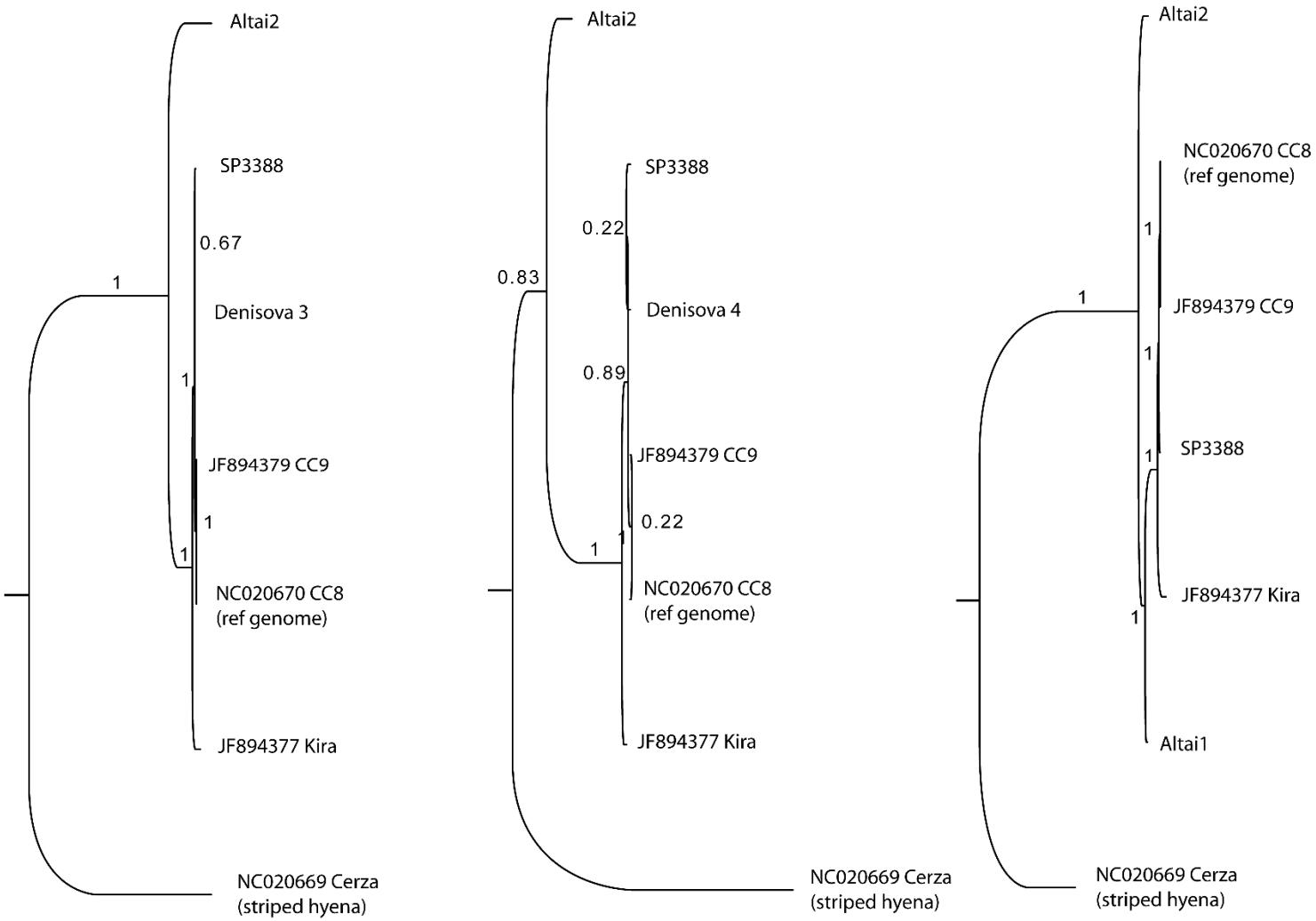

Figure 30. Separate bayesian phylogenetic mtDNA trees of the spotted hyena sequences of *Denisova 3*, *4* and *Altai 1*, each with the *Altai 2* spotted hyena consensus, the spotted hyena from Denisova Cave (SP3388), three additional spotted hyenas, one striped hyena and a mongoose as outgroup (not shown). The number of positions used to build each tree were 8,592 for *Denisova 3*, 468 for *Denisova 4* and 7,507 for *Altai 1*.

6.4 Discussion

The finger bone from the *Altai 2* individual shows clear morphological signs of being human. When aligned to the present-day human mtDNA and nuclear genomes, the *Altai 2* is clearly a Neandertal.

This Neandertal bone is substantially contaminated with ancient spotted hyena DNA. Nevertheless we were able to build a Neandertal mtDNA genome free of hyena and present-day human contamination with only six positions missing, which can therefore be used for comparative mtDNA analyses. The *Altai 2* falls together with the *Altai 1* Neandertal from the same cave. The *Altai 2* finger bone was found in a deeper layer of the same gallery as the *Altai 1* (layer 12 versus layer 11.4), however the two mtDNA genomes only have 10 differences between them, three to five times less than the differences they carry to other Neandertal individuals. So far, in the nine Neandertal mtDNAs published, all of the western European Neandertals cluster together, while the eastern Neandertals from Mezmaiskaya, Okladnikov and Denisova fall outside the variation of the European Neandertals (Figure 21).

The nuclear genome data generated from the *Altai 2* finger bone is not enough to effectively eliminate hyena or present-day human contamination. We examine only sequences with putative cytosine deamination at the fragment ends, which reduces contamination (see Chapter 5). Filtering for deaminated Cs does not however eliminate the ancient hyena sequences. The data generated shows that based on nuclear DNA, the divergence of the *Altai 2* is lower to the high-coverage *Altai 1* Neandertal than to the Denisovans or present-day humans. When comparing divergences of other low-coverage Neandertals and the *Altai 2* to the *Altai 1*, the divergences of the other Neandertals are all significantly lower. It is possible that either residual present-day human contamination, or more likely, substantial hyena contamination is inflating the divergence of the *Altai 2* to the *Altai 1* Neandertal.

We had hoped to reduce the hyena contamination from our extract by first washing the bone powder with sodium phosphate before extracting DNA from the now washed bone powder. It has been shown that sodium phosphate washing preferentially removes the microbial contamination in ancient bones, allowing endogenous DNA to be extracted after washing, possibly due to the microbial DNA being bound to the outer Ca^{2+} of a hydroxyapatite compound, while the endogenous DNA, which bound first, is bound farther inside and harder to wash off (15). Present-day human contamination is not preferentially washed off however, maybe due to

the fragments being protected in skin flakes (15). Contrary to our expectations the sodium phosphate wash did not reduce the hyena contamination. In fact the extraction after sodium phosphate washing has significantly more hyena contamination as well as higher present-day human contamination (Table 28). One explanation for this could be that the hyena DNA was already present in large amounts while the Neandertal cells were lysing, maybe as the hyena was eating a freshly killed Neandertal, and so both populations of DNA had equal chances to bond to the hydroxyapatite. This extract may have also come from a pocket of the bone where there happened to be more contaminant hyena DNA than endogenous Neandertal DNA, as is also the case for E1269 (Table 20), explaining the higher amounts of hyena DNA. However the hyena DNA does seem to be more degraded, since the hyena DNA is more fragmented and has a higher C to T deamination at the fragment ends than the Neandertal DNA (Table 29 and Figure 29). This degradation could be an indication that the DNA was more exposed than the endogenous Neandertal DNA and could also be due to the source of the hyena DNA, from saliva or feces/urine. The sodium phosphate wash does contain more microbial background than the extraction after washing (0.26% endogenous in the $NaHPO_4$ wash to 0.56% endogenous afterwards).

The substantial amount of hyena contamination in the *Altai 2* finger bone gives us the opportunity to study the hyena that contaminated the bone. We are able to reconstruct a mostly complete mtDNA of a hyena, although the low consensus support across some of the mtDNA genome indicates that there were possibly more hyenas or other animals that also contributed to the contamination (Figure 24). The contaminating hyena is clearly a spotted hyena, as expected since the Pleistocene hyenas that inhabited Europe and Asia have been shown to have mtDNAs that fall within the variation of present-day spotted hyenas (13, 14). However this spotted hyena sequence is quite diverged from the four other complete mtDNA genomes of extant and Pleistocene spotted hyenas, including the mtDNA genome of a spotted hyena bone from the same layer and gallery from Denisova Cave.

A comparison between the cytochrome b gene of 57 spotted hyenas both recent and ancient, including a second spotted hyena from Denisova Cave (AJ809327, (141)), shows that the *Altai 2* hyena contaminant only has one difference to the cytochrome b gene of Pleistocene hyenas from the far east, from caves near Vladivostok and north eastern China.

When we examined the other individuals from Denisova Cave, *Denisova 3*, *4*, *8* and *Altai 1*, we find that all except *Denisova 8* also show detectable levels of hyena contamination (Tables 30-32). A closer look at these sequences reveals that the two Denisovans were contaminated with spotted hyenas closer to the western European and African hyenas, while the *Altai 1* contaminating hyena falls outside the variation of these spotted hyenas, although it is not as diverged as the *Altai 2* hyena.

These results could stem from various scenarios. First it is possible that Pleistocene hyenas scavenged or ate both Denisovans and Neandertals living in the area and brought the remains into Denisova Cave. Since none of the bones have any detectable bite marks or acid etching from stomach acids, these small bones may have been regurgitated by the hyenas in hair balls (16). Hyenas are known for leaving behind both teeth and metapodials after eating large prey (12, 16). This could be a possible explanation for why only such small remains of Denisovans and Neandertals are found in the cave. The larger bones would have been easily crushed by the hyenas while eating, leaving behind only morphologically indistinct remains. Denisova Cave has thousands of such small bone fragments (3).

It is also possible however that the remains from Denisovans and Neandertals were deposited in other ways, and that the hyena contamination occurred afterwards. The hyenas could have found or dug up the remains, as spotted hyenas in Africa are known for grave robbing (3). Hyenas could have peed and defecated in the cave as well, and this excrement could have mixed with water to form a soup in which the archaic human remains sat and absorbed the hyena DNA. The differing amount of hyena contamination in the various bones could have been a result of either differing amounts of hyena contact or of the remains lying in parts of the cave with more or less hyena defecation. However, since it is possible that more than one hyena contaminated the *Altai 2* bone, it is not possible to predict which scenario occurred.

Regardless of how the archaic human remains were contaminated with hyena DNA, this study is the first example of such extreme animal DNA contamination in a Neandertal or Denisovan. While Denisova Cave does show evidence of human occupation (2) some of the homins may have been dragged into the cave by hyenas. Based on spotted hyenas in Africa today, their ranges can vary widely based on prey density, anywhere between 40 and 1000 km^2 (66), but they do not often drag prey very far from a kill site (17), and thus the hominds were most likely within a few kilometers of the cave. It is already known that Neandertals lived in the region (Okladnikov and

Chagyrskaya Caves (15, 44)), but this indicates that Denisovans also must have been close to the cave if they did not occupy it.

Denisova Cave was a meeting point of both eastern and western hominids (Denisovans and Neandertals) as well as eastern and western populations of Pleistocene spotted hyenas. Of note is that the hyena DNA found in the eastern hominid group (Denisovans) comes from a western clade of hyenas, while the hyena DNA in the western hominid group (Neandertals) comes from an eastern clade of hyenas. Direct dating of hyenas from the Denisova cave could reveal when these hyena populations and, by extension, when Neandertals and Denisovans inhabited the region. One of the hyenas from Denisova Cave (AJ809327 (141)), whose cytochrome b gene falls into the blue clade (Figures 27 and 28) has been directly C^{14} dated to 42,300 +940/-840 years old. In addition, one of the three hyena teeth from Da'an Cave in northern China was dated to 35,520 +/- 230 years old, while the hyena from the Geographical Society Cave from eastern Russia was dated to 48,650 +2380/-1840 years old. The cytochrome b genes of both of these eastern hyenas fall into the green clade (Figures 27 and 28) with *Altai 2*. These initial dates indicate that the two populations overlapped in time, however a direct date of a hyena from Denisova Cave that stems from the far eastern hyena population would be needed to shed light on this question. More genetic data, including nuclear DNA data, from Pleistocene hyena populations across Eurasia would also be of interest.

7. Discussion

7.1 Hominin occupation of the Denisova Cave region

Denisovans are a new hominin group that are the first to be identified based solely on DNA (7-9). The only known remains before the work presented here were a small morphologically indistinct piece of a finger phalanx from a young child and possibly a large upper third molar, although only the mtDNA was known (9). With this work we can confirm that the large molar did indeed belong to a Denisovan. We can also confirm that the large size of this molar is not a morphological aberration. Instead it was a constant Denisovan feature over a long period of time since an additional large upper third molar has been described here that belonged to an individual that lived much earlier. No other morphology can be associated with Denisovans. It is possible that they were otherwise not morphologically distinct from Neandertals or even more archaic hominins, and have been misclassified based on morphology, but this can only be confirmed with further DNA analysis.

Denisovans populated the Denisova Cave region on at least two occasions over a span of about 60,000 to 100,000 years (Table 12, as well as calculations in section 5.4). It is possible that they lived in the region continuously during this time, and that none of their remains have to date been found. Or they only occasionally populated the Altai Mountains and used them as a refugium during extreme climate changes in the lower steppe lands.

Denisovans show a higher diversity in their autosomal DNA than Neandertals (Figure 18). This higher diversity could be due to the great difference in age between the *Denisova 8* and the two other Denisovans. The Neandertals to which we compared the Denisovans range in age from 65,000 years for *Mezmaiskaya1* (62) to 38,000 for Vindija 33.16 (73), and show no significant difference in divergence to the *Altai 1* Neandertal (Table 16, Figure 18). However it should be noted the Neandertals cover a geographical range from Spain to Siberia, while the three Denisovans are all from the same cave.

Denisovans show an even higher degree of diversity in their mtDNA genomes compared to Neandertals (Figure 15, Table 10). The number of differences between *Denisova 8* and *Denisova 3* is 73% of the largest number of differences measured between 311 present-day humans (Figure 15). Again this could be due to the large age difference among the Denisovans, however both the Neandertals and humans come from large geographic ranges.

Denisova 8 is much older than *Denisova 3*, and shows a higher divergence both in autosomal and mtDNA data to *Denisova 3* compared to divergences among Neandertals. The Denisovan population that introgressed with the ancestors of present-day Oceanians diverged from the *Denisova 3* population 100 to 400 kya, depending on which mutation rate is used (10). Since *Denisova 8* could be up to 100ky older than *Denisova 3*, it is possible that *Denisova 8* belonged to a population that is more closely related to the population of Denisovans that introgressed with Oceanians. Unfortunately there are not enough data from *Denisova 8* to answer this question. There are also not enough data to ascertain whether *Denisova 8* had an admixture signal from the *Altai 1* Neandertal, as is shown in *Denisova 3*. Since this admixture signal comes from a recent event, post-dating almost all genetic drift in the population to which *Denisova 3* belonged (10), it is possible that such an admixture signal would not be seen in *Denisova 8*. *Denisova 8* may also have more admixture from more archaic hominins, but again the lack of data prevents an answer to these questions.

Altai 2 is a second Neandertal from Denisova Cave. This new Neandertal comes from the deepest layer of all of the remains in the East Gallery, however the disturbance of stratigraphy of the East Gallery could have occurred in layer 12 as well. Based on the mtDNA, there is no branch shortening to be seen in *Altai 2*. In fact *Altai 2* has the longest branch from the MRCA of Neandertals when compared to seven other Neandertals by five differences (section 6.3.2 and Table 10). Therefore it is unlikely that *Altai 2* is significantly older than *Altai 1*; if anything it is more likely younger.

The region around Denisova Cave also has other caves in which Neandertal remains have been found. One of these is Okladnikov cave (44, 103), from which an almost complete mtDNA exists from an individual which has been dated multiple times and produced dates from 29,990 $\pm$ 500 to 37,800 $\pm$ 450 years before present (uncalibrated C^{14} dates) (44). Another cave, Chagyrskaya Cave (15, 142), also has Neandertal bones which are beyond the C^{14} dating boundary (so greater than 50,000 years old) (143), however no published DNA exists from Neandertals from this site.

We now have mtDNA sequences from four Neandertals outside of Western Europe: *Mezmaiskaya1* (dated ~65ky old (62)), Okladnikov (dated ~30-38ky old (44)) and two undated Neandertals from Denisova Cave. All four of these Neandertals fall basal to the Neandertals from Western Europe (Figure 21). Based on this data, it seems that Eastern Neandertals fall

outside the variation of European Neandertals, as has been previously shown (103). Previous work with the hypervariable region of the mtDNA has shown that some older European Neandertals also fall outside the variation of the European Neandertals used in this study (144), however the use of the hypervariable region for Okladnikov has been shown to result in unreliable placement in the tree (103). While it is tempting to start speculating about Neandertal population movements based on these results, it would take a large scale autosomal DNA study of these Neandertals to more accurately look at this question.

7.2 The deposition of hominin remains in Denisova Cave

Many of the hominin bone pieces found in the Altai Mountains are small, morphologically indistinct pieces (3). No complete or even semi-complete skeletons have been excavated to date (3). One reason for only small pieces of remains being left is that the population sizes of hominins in the region could have been small. If only small groups of hominins existed, they would have left few dead behind. If we follow the hypothetical calculations done by Tuner *et al* in 2013, then 100 bands with 20 people each would leave only 300 dead to be discovered over a 30,000 year time frame (3). Both Denisovans and Neandertals had very small population sizes (8, 10). In addition, the *Altai 1* Neandertal was the product of extreme and recent inbreeding (10). Based on coding regions, the *Altai 1* has 79% and 89% of the heterozygosity of a Neandertal from El Sidron, Spain or Vindija Cave, Croatia, respectively (145). Therefore it is possible that both the Denisovan and Neandertal populations in the Altai Mountains were small. Turner *et al* 2013 (3) and Wrinn 2010 (63) both argue that the number of lithics found in the Altai mountains are small. Based on the calculations from Wrinn, layer 12 of the main chamber of Denisova Cave has the most lithics compared to surrounding sites such as Okladnikov and Anui 3, with about 600 artifacts per m^3 over 10,000 years, compared to the other sites that rarely reach 200 artifacts per m^3 over 10,000 years (63). The authors argue that these results suggest a sporadic occupation of the Altai Mountains. However no comprehensive comparison of the Altai region with other regions has been done to show that the lithic assemblages are indeed smaller than in other regions. A study from Mellers and French in 2011 compared archeological evidence from Neandertal and anatomically modern human sites in the Aquitaine region of southwestern France (146). They calculated stone tool densities between 6.6 to 17.6 tools per m^2 per 1000 years (146), however it is not clear if their methods for calculations are similar to the

calculations done by Wrinn 2010, and it would thus not be prudent to draw conclusions based on these data. In addition it has been cautioned that using stone tool densities to try to calculate population size is extremely complex (147).

The spotted hyena of the Pleistocene was a formidable opponent and obstacle to hominins. They were larger than spotted hyenas today (12), who are not fearful of humans (3). Reports have shown that the number of attacks on humans by hyenas are higher than by other terrestrial carnivores (3). They hunt in packs and are easily able to take down large prey (17). Since spotted hyenas dig dens for their young, they, along with other cave-dwelling animals such as the cave bear, would have been harsh competitors for hominins for cave-use. It is hard to imagine small hominin groups expelling an established hyena den from a large cave with ample digging room such as Denisova Cave.

The only hominin remains from the Pleistocene in Denisova Cave are phalanxes and teeth. Studies of bones left behind at dens of African spotted hyena today show that hyenas often leave behind metapodials and teeth, as well as occasional phalanxes (12, 16). Although hyena dens can show large assemblages of other bones, most of the unmodified bones are metopodials, possibly an indication of being swallowed whole and then regurgitated (16). We see hyena contamination in four of the five hominin remains in Denisova Cave that have been published so far. Only *Denisova 8* shows no detectable amount of hyena contamination. *Denisova 8* is much older than the other Denisovans and Neandertals, so it is possible that this individual was not eaten by hyenas, or that the location in the cave was somehow not tainted by hyena refuse. However, lack of detection of contamination does not mean that this tooth was not also disturbed by hyenas.

There are various possible reasons for the hyena contamination on the hominin remains in Denisova Cave. First it is possible that the remains were deposited during occupation of the cave by Denisovans or Neandertals. It is unclear whether Neandertals buried their dead (148, 149), but if remains were left in an occupied cave, it is likely that they were put under the ground in some manner to avoid the smell. The cave could have also been used as a site of deposition of dead bodies, while the hominins lived outside the cave. After the hominins left the cave (or were chased out), hyenas could have moved in. It has been speculated, based on hyena and animal bones left at Okladnikov cave, that hominins only used the cave in spring and fall, and lived in the forests during the winter, where they were closer to firewood, while the hyenas then moved into the caves in the winter as a protection against the cold (3).

It is unclear how much later the remains at Denisova Cave could have been disturbed, whether it was a season later or thousands of years later. Whether hyenas eat old bones is not mentioned in the literature. However it has been observed that hyenas very occasionally chew on dried out skeletons (personal communication with Dr. Gus Mills, Hyaena Specialist Group) and are not averse to chewing on crunchy material such as wood or plastic car tail lights (17). Hyenas can easily go many days without food and large numbers of bleached bones have been seen in areas where hyena density is low and food density is high, meaning that they are not eaten by the hyenas (17). In addition, spotted hyenas in Africa today will take off with a chunk of a kill, store it in water for a few days to preserve (17), and return for it later on. Thus they most likely prefer well preserved meat that has not begun to degrade in the hot African sun. But it is possible that a starving or bored hyena may chew on a bone that is thousands of years old.

Thus hyenas may have come across recent carcasses and dug them up, or even have stolen carcasses from where they were placed by the hominins. This phenomenon is observed countless times in the wild today, as hyenas, like other carnivores, often scavenge for food and will steal food away from other animals as well as each other (17). As the hyena DNA did not wash off more than the endogenous Neandertal DNA during the phosphate washing of the *Altai 2* bone (Table 28), it is possible that the hyena and Neandertal DNA were present at the same time, while the cells were lysing, and thus both populations of DNA had the same chances to bind to the hydroxyapatite in the bone. However there is an indication that more than one individual contaminated the *Altai 2* bone, making this scenario less likely.

It is entirely possible that the hyenas never directly interacted with the remains in Denisova Cave. Hyenas probably used the cave as a den multiple times over the span of thousands of years, causing disturbance of the layers in the East Gallery in the process, and may have defecated there. Spotted hyenas use latrines, specific areas often on the outer perimeter of their ranges, where they will defecate (17). The young, however, must defecate in the dens, especially when there is danger and they cannot leave the cave. Either Denisova Cave was used as a latrine or as a den or both, many times over the thousands of years after the remains were deposited. The feces then decomposed and released the DNA into the soil (18). The combination of the wet environment and hyena waste could have caused the hyena DNA to leech into the Neandertal and Denisovan remains. If such a situation occurred, it is unclear why the hyena DNA did not preferentially release off of the bone during the phosphate washing as did the microbial DNA.

Since remains from both the East and South Gallery show hyena contamination, the hyenas must have used both galleries for such activities at some point in time.

The last possibility is that the large Pleistocene hyenas hunted the hominins, as they have been known to do in Africa today (3, 17). There is evidence that hyenas ate hominins in Siberia from signs of acid erosion on bones (3), as would be expected from such a large predator. Hyenas are nocturnal (65) and pose a great threat to humans that sleep out in the open in Africa today. Hominins in the Pleistocene were not agriculturalists with established protected buildings as most Africans are today, and unless they were sleeping in a cave, they would have been exposed to hyena attacks at night. Therefore if hyenas were using Denisova Cave as a den or were otherwise occupying it, making it inaccessible to the hominins as a place of shelter, the hominins may have lived out in the open. While hyena territories today span between 40 and 1000 km^2 (66), they do not necessarily drag their food back to their den for their young, since spotted hyenas exclusively nurse their young for many months (17). The killing or scavenging of the hominin would therefore have happened within a few km of the cave, and during the feeding frenzy a hyena could have taken an arm foot or head with them to eat it in the shelter of the cave in peace, as is common behavior for spotted hyenas today (17). Therefore even if the Denisovan and Neandertal remains that we find today did not belong to individuals that occupied the cave, they must have been within a few km of the cave.

Due to the phosphate washing results, it is likely that the hyena and Neandertal DNA in the *Altai 2* bone bonded to the hydroxyapatite at the same time, so for the *Altai 2* at least the Neandertal and hyena may have had direct contact, either before or right after death. This could also explain why the amount of hyena DNA is so much higher in the *Altai 2* bone than in the others. However since the *Altai 2* hyena mtDNA sequences show a potential mixture of contributers, it is very possible that more than one of these scenarios happened in combination, since Denisova Cave was used heavily by hyenas as is evidenced by the large amounts of hyena bones and even coprolites in the cave (2, 3).

7.3 The Pleistocene spotted hyenas of Denisova Cave

Based on the mtDNA cytochrome b data, there are three clades of Pleistocene hyenas (Figure 27). Of the two that are no longer found in Africa today, one occurred exclusively in Europe

while the other occurred in eastern Asia and Denisova Cave. The third Pleistocene hyena clade occurs in African spotted hyenas today as well as in Pleistocene hyenas in Europe and Denisova Cave. Since there are no hyena sequences from hyenas that lived between western Ukraine and Denisova Cave, or between Denisova Cave and Vladivostok, it is unclear how much these two clades overlap. However it is clear that they overlap at Denisova Cave, making Denisova Cave a possible meeting point of two Pleistocene hyena populations. Whether these populations existed simultaneously is unclear. Of the 24 Pleistocene hyena sequences, 11 come from individuals that have been dated with C^{14} dating. The far-east hyenas have dates between 35kya (Chinese) (14) and 48kya (eastern Russian) (13). This age range overlaps with the one dated hyena from Denisova Cave, 42 kya (141), which is one of the Denisova Cave hyenas that fall into the clade found also in Europe and in extant African spotted hyenas. The age range also overlaps with the ages of the European hyenas, which range from 38kya to 51kya (13, 141). The exclusively European clade only has two dated individuals, and both are the oldest dated hyenas in Europe at 51kya and older than 48kya, so it is possible that this clade was made up of older hyenas. However the difference between eastern and western Eurasian hyenas does not seem to be due to age.

Previous work on the cytochrome b gene of the mtDNA of Pleistocene hyenas has been used to hypothesize about the origin of the spotted hyena. Rohland *et al* concluded that spotted hyenas populated Eurasia from an African origin in three waves. The first wave left southern Africa, the site of the some of the earliest spotted hyena fossils in Africa (3.46 million years ago (150, 151)), crossed northern Africa, then exited Africa and went across Asia to East Asia. They must have traveled through Pakistan, where the oldest Asian spotted hyena fossils exist (2.6-3.7 million years ago (152)). They could have also traveled over the Altai Mountains, where a group could have stayed. Since this migration happened up to 3.5 million years ago, it is likely though that the hyena sequence consensus found in *Altai 2*, which carries only one difference to the eastern hyena sequences, was either part of a large and often mixing eastern population, encompassing the far-east and Denisova Cave, or represents a back migration later in time. The second wave left Africa and came to Europe, where the earliest hyena fossils are dated to 0.8 million years ago (153). This group then died out in Africa, leaving only a European clade. Around this time, the hyena populations were separating roughly into northern and southern African populations. The

last wave left northern Africa and populated both Europe and Siberia (13). Our additional results do not refute this hypothesis.

With the addition of three more eastern Pleistocene hyena cytochrome b sequences (14), a new hypothesis arose, where all four clades were panmictic in Eurasia. First the eastern clade separated 400 to 230 kya, then the exclusively African clade as well as the African and European clades entered Africa between 145 and 50 kya, and subsequently they separated into northern and southern populations. The eastern hyena contamination of *Altai 2* does not negate this hypothesis, since the eastern clade of hyenas could have migrated back to Denisova Cave, however the 3.46 million year old spotted hyena fossils from Africa (150, 151) complicate the picture.

Both hypotheses state that the East Asian spotted hyena clade was the earliest to split off. We now show that this clade also existed much farther west than previously thought; thus it is possible that this eastern population had a much larger geographical distribution than previously thought, encompassing both eastern Russia and central Siberia.

Denisova Cave was not only a possible meeting point for Pleistocene hyena populations, but also for hominin groups. Denisova Cave represents the farthest eastern location with Neandertal remains found to date, and is possibly on the far East of the Neandertal range in Europe. No Denisovans have been found outside of Denisova Cave. An admixture signal from Denisovans is seen in Oceania and to a lesser degree in East Asia, so it is tempting to hypothesize that Denisovans lived in the East. However the split between the introgressing Denisovan population and *Denisova 3* is quite deep, as they diverged at least 140kya (10). Therefore it is possible that the introgressing Denisovans were from a very diverged or archaic population, and the population to which the three Denisovans from Denisova Cave belonged never came into contact with the ancestors of present-day Oceanians. Thus it is unclear how far East or West Denisovans lived, although it is less likely they were widespread in Europe.

Of interest is that the Pleistocene hyenas that contaminated the Denisovans came from the western hyena clade, while the hyena that contaminated the *Altai 2* and possibly *Altai 1* Neandertals came from an eastern population.

These hyena data give a glimpse into possible populations of Pleistocene hyenas. However it is important to note that the results here are mostly based on a small part of the mtDNA. Even the complete mtDNA data are only giving information on this one maternally-inherited locus. As

shown in Denisovans, the mtDNA alone can show a very different picture from autosomal data (9). Therefore it would be of great interest to sequence the autosomal genomes of these Pleistocene hyenas, as well as more populations across Russia and the Middle East. Before such sequencing is done, the question of spotted hyena origin and movement cannot be fully answered.

8. Appendix

8.1 Index of Figures

Figure: Page:

8.2 Index of Tables

Table: Page:

8.4 Appendix of Tables

Appendix Table 1. List of spotted hyena sequences used in section 6.3.5. Clade color refers to the color of the clade the individual falls into, see Figure 27. Age is for samples greater than 1000 years old, collection date is given for samples collected in the 19th and 20th century. Samples from live individuals were taken in the last 20 years.

Accession number	Museum number/Name	Location	Clade color	Age/collection date	Reference
AJ809318	Teufel 2	Teufelslucke (Austria)	blue	38,060	(141)
AJ809320	Winden	Winden cave (Austria)	blue	38,680	(141)
AJ809319	Irpfel 4	Irpfel cave (Germany)	blue	-	(141)
AJ809324	Igric (V10529)	Igric (Romania)	blue	41,800	(141)
AJ809331	Kiske M1 (V14484)	Kiskevelyi (Hungary)	yellow	>48,500	(141)
AJ809321	Vypustek p	Vypustek (Czech Rep.)	blue	46,000	(141)
AJ809330	Linde 1	Lindenthal cave (Germany)	yellow	-	(141)
AJ809327	Altai D19	Denisova Cave (Russia) Asia	blue	42,300 +940/-840	(141)
AJ809326	Bukovinka	Bukovinka cave (Ukraine)	blue	41,300	(141)
AJ809328	Certova 1	Certova pec (Slovakia)	yellow	51,200	(141)
AJ809329	Sveduv 2	Sveduv stul (Czech Rep.)	yellow	-	(141)
AJ809332	Tmava (TS 250)	Tmava skala (Slovakia)	yellow	-	(141)
AJ809322	PLU 681 E3 II	Les Plumettes (France)	blue	-	(141)
AJ809325	RDV 01H10 23	Les Roches de Villeneuve (France)	blue	40,700	(141)
AJ809323	Niederlande 1	North Sea (The Netherlands)	blue	-	(141)
DQ157554	2812 37	Goyet cave (Belgium)	blue	-	(13)
DQ157555	895 (34490 (1))	Geographical Society cave (Vladivostok, Russia) Asia	green	48,650 +2380/-1840	(13)
NC020670	CC8	Coumere Cave, France	blue	-	(97)

JF894379	CC9	Coumere Cave, France	blue	-	(97)
KC117379	DARD1	Da'an cave, Tonghua county, Jilin province, China	green	35,520 +/- 230	(14)
KC117380	DARD2	Da'an cave, Tonghua county, Jilin province, China	green	-	(14)
KC117381	DARD3	Da'an cave, Tonghua county, Jilin province, China	green	-	(14)
DQ157556	41225	Dagana (Senegal)	blue	1925	(13)
DQ157557	46123	Dire Dawa (Ethiopia)	blue	1928	(13)
DQ157558	82399	(N-Cameroon)	blue	1913	(13)
DQ157559	82400	Kete Kratshi (Togo)	blue	1899	(13)
DQ157563	82404	Ikoma (Tanzania)	blue	1906	(13)
DQ157568	82412	Windhuk (Namibia)	pink	1898	(13)
DQ157562	82423	Lake Kivu (Rwanda)	blue	1902	(13)
DQ157567	82432	Ikoma (Tanzania)	blue	1913	(13)
DQ157564	82455	Sefane (Eritrea)	blue	1913	(13)
DQ157573	82467	Singa (Sudan)	pink	1912	(13)
DQ157566	82472	Malindi (Zimbabwe)	pink	1911	(13)
DQ157592	82477	(NE-Rwanda)	blue	1907	(13)
DQ157560	82482	Masai steppe (Tanzania)	pink	1905	(13)
DQ157561	82533	Otawi (Namibia)	pink	1902	(13)
DQ157570	82537	Loanda (Angola)	pink	1901	(13)
DQ157569	1244	Welkom (South Africa)	pink	1880s	(13)
DQ157571	1755	Somaliland (Somalia)	blue	1898	(13)
DQ157572	3919	Stony Athi (Kenia)	pink	1908	(13)

DQ157565	6061	upper course of Dinder (Sudan)	blue	1925	(13)
DQ157575	7397	S shore of lake Tschad (Cameroon)	blue	1928	(13)
DQ157574	21495	Mutir (Uganda)	blue	1882	(13)
JF894377	Kira	French zoo	blue	From live individual	(97)
AF153053	Ibo	Berlin zoo	blue	From live individual	(112)
AY048811	Nigeria-TPIbo	Berlin zoo	blue	From live individual	(111)
AF153052	Kr.L.097	Ngorongor crater, wild	pink	From live individual	(112)
AF153051	M119	serengeti, wild	pink	From live individual	(112)
AY048812	Suedafrika-Muen	Munich zoo, poss South African origin	pink	From live individual	(111)
AY048807	KrX012	northwest tanzania, wild (ngorongoro?) (GPS: 03 10 80/35 34 76)	pink	From live individual	(111)
AY048808	Lemuta-Z056	Serengeti, no GPS	pink	From live individual	(111)
AY048809	Seronera-Z098	Serengeti, no GPS	pink	From live individual	(111)
AY048805	Central-M119	Serengeti, wild (GPS: 02 27 30 / 34 46 00)	pink	From live individual	(111)
AY048806	Ngorongoro-KrM17	Ngorongoro, wild (GPS: 03 09 64 / 35 36 63)	pink	From live individual	(111)
AY048804	Camsite-C004	Serengeti, wild (GPS: 02 24 82 / 34 50 48)	pink	From live individual	(111)
AY048810	Ruaha-Z095	Ruaha park, no GPS	pink	From live individual	(111)

9. References

1. Derevianko A (2011) The upper paleolithic in Africa and Eurasia and the origin of anatomically modern humans. *Novosibirsk, Institute of Archeology and Ethnography SB RAS Press, MV Shunkov (Ed)*.
2. Derevianko A*, et al.* (2003) Paleoenvironmnet and paleolithic human occupation of Gorny Altai. *Novosibirk, Institute of Archeology and Ethnography, SB RAS Press, AP derevianko and MV Shunkov (Eds)*.
3. Turner C, Ovodov N, & Pavlova O (2013) Animal teeth and Human tools: A taphonomic Odyssey in ice age Siberia. *Cambridge University Press*.
4. Turner C (1990) Upper Paleolithic Mal'ta child (Siberia) *News of siberian Branch, USSR Academy of Sciences* 2:70-71.
5. Viola B (2009) New hominid remains from central Asia and Siberia: the eastern-most Neanderthals? *PhD Dissertation: Vienna University*.
6. Mednikova MB (2011) A proximal pedal phalanx of a paleolithic hominin from Denisova cave, Altai. *Archaeology Ethnology & Anthropology of Eurasia* 39(1):129-138.
7. Krause J*, et al.* (2010) The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* 464(7290):894-897.
8. Meyer M*, et al.* (2012) A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338(6104):222-226.
9. Reich D*, et al.* (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468(7327):1053-1060.
10. Prufer K*, et al.* (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505(7481):43-49.
11. Green RE*, et al.* (2010) A draft sequence of the Neandertal genome. *Science* 328(5979):710-722.
12. Kurten B (1968) Pleistocene mammals of Europe. *Weidenfeld and Nicholson (eds), London*.
13. Rohland N*, et al.* (2005) The population history of extant and extinct hyenas. *Mol Biol Evol* 22(12):2435-2443.
14. Sheng GL*, et al.* (2014) Pleistocene Chinese cave hyenas and the recent Eurasian history of the spotted hyena, Crocuta crocuta. *Mol Ecol* 23(3):522-533.
15. Korlević P, Gerber, T., Gansauge, M.T., Hajdinjak, M., Nagel, S., Aximu-Petri, A., Meyer, M (2015) Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* (in press).
16. Hill A (1989) Bone modification by modern spotted hyenas. *In: Bonnichsen R, Sorg MH (Eds), Bone modification, Peopling of the Americas, Center for the study of the first Americans, Orono*.
17. Kruuk H (1972) The spotted hyena, A study of predation and social behavior. *In: Schaller GB (Ed), Wildlife behavior and ecology, University of Chicago press*.
18. Horwitz L & Goldberg P (1989) A study of pleistocene and holocene hyaena coprolites. *Journal of Archaeological Science* 16:71-94.
19. Saint-Hilaire IG (1859) Histoire Naturelle Generale des Regres Organiques, Tome II. *Librarie de Voctor Masson, Paris*.
20. Huxley TH (1863) Evidence as to Man's place in nature *Williams and Norgate, London*.
21. Darwin C (1871) The descent of man. *John Murray, London*.
22. King W (1864) The reputed fossil man of the Neandertal. *Quarterly Journal of Science* 1:88-97.
23. Chimpanzee S & Analysis C (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437(7055):69-87.

24. Prufer K*, et al.* (2012) The bonobo genome compared with the chimpanzee and human genomes. *Nature* 486(7404):527-531.
25. Goodman M (1999) The genomic record of Humankind's evolutionary roots. *Am J Hum Genet* 64(1):31-39.
26. Wood B (2010) Colloquium paper: reconstructing human evolution: achievements, challenges, and opportunities. *Proc Natl Acad Sci U S A* 107 Suppl 2:8902-8909.
27. Gabunia L*, et al.* (2000) Earliest Pleistocene hominid cranial remains from Dmanisi, Republic of Georgia: taxonomy, geological setting, and age. *Science* 288(5468):1019-1025.
28. Wolpoff MH*, et al.* (1988) Modern human origins. *Science* 241(4867):772-774.
29. Cavalli-Sforza LL (1966) Population structure and human evolution. *Proc R Soc Lond B Biol Sci* 164(995):362-379.
30. Cann RL, Stoneking M, & Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* 325(6099):31-36.
31. Tishkoff SA & Verrelli BC (2003) Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu Rev Genomics Hum Genet* 4:293-340.
32. Schlebusch CM*, et al.* (2012) Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *Science* 338(6105):374-379.
33. Li H & Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493-496.
34. Stringer C (2002) Modern human origins: progress and prospects. *Philos Trans R Soc Lond B Biol Sci* 357(1420):563-579.
35. Atkinson QD (2011) Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 332(6027):346-349.
36. McBrearty S & Brooks AS (2000) The revolution that wasn't: a new interpretation of the origin of modern human behavior. *J Hum Evol* 39(5):453-563.
37. Hublin JJ (2009) Out of Africa: modern human origins special feature: the origin of Neandertals. *Proc Natl Acad Sci U S A* 106(38):16022-16027.
38. Stringer CB & Hublin J (1999) New age estimates for the Swanscombe hominid, and their significance for human evolution. *J Hum Evol* 37(6):873-877.
39. Schwarcz HP, Simpson JJ, & Stringer CB (1998) Neanderthal skeleton from Tabun: U-series data by gamma-ray spectrometry. *J Hum Evol* 35(6):635-645.
40. F B & E T (2006) Middle Paleolithic Human remains from Bisitun cave, Iran. *Paleorient*:105-111.
41. Ovchinnikov IV*, et al.* (2000) Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* 404(6777):490-493.
42. Glantz M, Viola, B., Wrinn, P., Chikisheva, T., Derevianko, A., Krivoshapkin, A., Islamov, U., Suleimanov, R. & Ritzman, T. (2008) New hominin remains from Uzbekistan. *Journal of Human Evolution* 55(2):223-237.
43. Gunz P & Bulygina E (2012) The Mousterian child from Teshik-Tash is a Neanderthal: a geometric morphometric study of the frontal bone. *Am J Phys Anthropol* 149(3):365-379.
44. Krause J*, et al.* (2007) Neanderthals in central Asia and Siberia. *Nature* 449(7164):902-904.
45. Finlayson C*, et al.* (2006) Late survival of Neanderthals at the southernmost extreme of Europe. *Nature* 443(7113):850-853.
46. Grun R*, et al.* (2006) ESR and U-series analyses of enamel and dentine fragments of the Banyoles mandible. *J Hum Evol* 50(3):347-358.
47. Stringer CB, Grun R, Schwarcz HP, & Goldberg P (1989) ESR dates for the hominid burial site of Es Skhul in Israel. *Nature* 338(6218):756-758.
48. Higuchi R, Bowman B, Freiberger M, Ryder OA, & Wilson AC (1984) DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312(5991):282-284.

49. Paabo S (1985) Molecular cloning of Ancient Egyptian mummy DNA. *Nature* 314(6012):644-645.
50. Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* 362(6422):709-715.
51. Green RE*, et al.* (2006) Analysis of one million base pairs of Neanderthal DNA. *Nature* 444(7117):330-336.
52. Paabo S*, et al.* (2004) Genetic analyses from ancient DNA. *Annu Rev Genet* 38:645-679.
53. Briggs AW*, et al.* (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A* 104(37):14616-14621.
54. Brotherton P*, et al.* (2007) Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res* 35(17):5717-5728.
55. Sawyer S, Krause J, Guschanski K, Savolainen V, & Paabo S (2012) Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* 7(3):e34131.
56. Krause J*, et al.* (2010) A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr Biol* 20(3):231-236.
57. Skoglund P*, et al.* (2012) Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* 336(6080):466-469.
58. Krings M*, et al.* (1997) Neandertal DNA sequences and the origin of modern humans. *Cell* 90(1):19-30.
59. Green RE*, et al.* (2008) A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* 134(3):416-426.
60. Reich D*, et al.* (2011) Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am J Hum Genet* 89(4):516-528.
61. Gansauge MT & Meyer M (2013) Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature protocols* 8(4):737-748.
62. Skinner AR*, et al.* (2003) ESR dating at Mezmaiskaya Cave, Russia. *American Journal of Physical Anthropology*:193-193.
63. Wrinn P (2010) Middle Paleolithic settlement and land use in the Altai Mountains, Siberia. *Settlement dynamics of the Middle Paleolithic and Middle Stone Age vol. III, Tubingen kerns Publishing, N Conrad and A Delagnes (Eds)*.
64. Werdelin L & N S (1991) The hyaenidae: taxonomy, systematics and evolution. *Fossils and Strata* 30:1-104.
65. Gittleman J & Harvey P (1982) Carnivore home-range size, metabolic needs and ecology. *Behav Ecol Sociobiol* 10:57-63.
66. Mill M (1989) The comparative behavioral ecology of hyenas: the importance of diet and food dispersion *Carnivore behavior, ecology and evolution, Gittleman J (Ed), Cornell University*.
67. Rohland N & Hofreiter M (2007) Comparison and optimization of ancient DNA extraction. *Biotechniques* 42(3):343-352.
68. Briggs AW*, et al.* (2010) Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res* 38(6):e87.
69. Dabney J & Meyer M (2012) Length and GC-biases during sequencing library amplification: A comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* 52(2):87-94.
70. Maricic T, Whitten M, & Paabo S (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* 5(11):e14004.
71. Kircher M, Stenzel U, & Kelso J (2009) Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol* 10(8):R83.

72. Kircher M (2012) Analysis of high-throughput ancient DNA sequencing data. *Methods in molecular biology* 840:197-228.
73. Briggs AW*, et al.* (2009) Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* 325(5938):318-321.
74. Genomes Project Consortium*, et al.* (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491(7422):56-65.
75. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754-1760.
76. Andrews RM*, et al.* (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23(2):147.
77. R Core Team (2014) R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*.
78. Ingman M, Kaessmann H, Paabo S, & Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408(6813):708-713.
79. Arnason U, Xu X, & Gullberg A (1996) Comparison between the complete mitochondrial DNA sequences of Homo and the common chimpanzee based on nonchimeric sequences. *J Mol Evol* 42(2):145-152.
80. Katoh K, Misawa K, Kuma K, & Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30(14):3059-3066.
81. Katoh K & Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* 30(4):772-780.
82. Koichiro Tamura GS, Daniel Peterson, Alan Filipski, and Sudhir Kumar (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. . *Molecular Biology and Evolution* 30:2725-2729.
83. Posada D & Crandall K (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* 14(9):817-818.
84. Huelsenbeck JP, Ronquist F, Nielsen R, & Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294(5550):2310-2314.
85. Ronquist F & Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12):1572-1574.
86. Drummond AJ, Suchard MA, Xie D, & Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29(8):1969-1973.
87. Meyer M*, et al.* (2014) A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* 505(7483):403-406.
88. Fu Q*, et al.* (2013) A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr Biol* 23(7):553-559.
89. Kass R & Raftery A (1995) Bayes Factors. *Journal of the American Statistical Association* 90(430):773-795.
90. Librado P & Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451-1452.
91. Paten B, Herrero J, Beal K, Fitzgerald S, & Birney E (2008) Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Res* 18(11):1814-1828.
92. Paten B*, et al.* (2008) Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res* 18(11):1829-1843.
93. Patterson N*, et al.* (2012) Ancient admixture in human history. *Genetics* 192(3):1065-1093.
94. Dabney J*, et al.* (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):15758-15763.

95. Gansauge MT & Meyer M (2014) Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res* 24(9):1543-1549.
96. Kircher M, Sawyer S, & Meyer M (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* 40(1):e3.
97. Bon C*, et al.* (2012) Coprolites as a source of information on the genome and diet of the cave hyena. *Proc Biol Sci* 279(1739):2825-2830.
98. Slon V*, et al.* (2015) Mammalian mitochondrial capture, a tool for rapid screening of DNA preservation in faunal and undiagnostic remains, and its application to Middle Pleistocene specimens from Qesem Cave (Israel). *Quaternary International* In press.
99. Fu Q*, et al.* (2013) DNA analysis of an early modern human from Tianyuan Cave, China. *Proc Natl Acad Sci U S A* 110(6):2223-2227.
100. Renaud G, Kircher M, Stenzel U, & Kelso J (2013) freeIbis: an efficient basecaller with calibrated quality scores for Illumina sequencers. *Bioinformatics* 29(9):1208-1209.
101. Renaud G, Stenzel U, & Kelso J (2014) leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res* 42(18):e141.
102. Renaud G, Stenzel U, Maricic T, Wiebe V, & Kelso J (2015) deML: robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics* 31(5):770-772.
103. Skoglund P*, et al.* (2014) Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc Natl Acad Sci U S A* 111(6):2229-2234.
104. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. . *Nucl. Acids. Symp.* 41:95-98.
105. Camacho C*, et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
106. Fu Q*, et al.* (2014) Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 514(7523):445-449.
107. Arnason U*, et al.* (2006) Pinniped phylogeny and a new hypothesis for their origin and dispersal. *Mol. Phylogenet. Evol.* 41(2):345-354.
108. Krause J*, et al.* (2008) Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evol Biol* 8:220.
109. Drummond AJ & Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
110. Koepfli KP*, et al.* (2006) Molecular systematics of the Hyaenidae: relationships of a relictual lineage resolved by a molecular supermatrix. *Mol. Phylogenet. Evol.* 38(3):603-620.
111. Albert R (2002) Genstruktur und Genfluss in ausgewaehlten Populationen der Tuepfelhyaene (Crocuta crocuta). *PhD Dissertation, Freie Universitaet, Berlin*.
112. Albert R, Hofer H, East M, & Pitra C (2000) Genetic identification of the geographic origin of spotted hyaenas (Crocuta crocuta). *Zool. Gart.* 70:1-10.
113. Meyer M*, et al.* (2016) Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Submitted to Nature*.
114. Scally A & Durbin R (2012) Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet* 13(10):745-753.
115. Bailey S, Glantz, M., Weaver, T.D. & Viola, B (2008) The affinity of the dental remains from Obi-Rakhmat Grotto, Uzbekistan. *Journal of Human Evolution* 55(2):238-248.
116. Martinón-Torres M, Bermúdez de Castro, J.M., Gómez-Robles, A., Prado-Simón, L. & Arsuaga, J.L. (2012) Morphological description and comparison of the dental remains from Atapuerca-Sima de los Huesos site (Spain). *Journal of Human Evolution* 62(1):7-58.
117. Kaifu Y (2006) Advanced dental reduction in Javanese Homo erectus. *Anthropol Sci* 114:35-43.

118. Xing S, Martinon-Torres, M., Bermudez de Castro, J.M., Wu, X., Liu, W (2015) Hominin Teeth From the Early Late Pleistocene Site of Xujiayao, Northern China. *Journal of Physical Anthropology* 156(2):224-240.
119. Frayer D (1976) Evolutionary Dental Changes in Upper Palaeolithic and Mesolithic Human Populations *PhD Dissertation, University of Michigan, Ann Arbor*.
120. Wood BA (1991) Hominid cranial remains. Koobi Fora Research Project, Vol. 4. *Clarendon Press, Oxford*.
121. Tobias PV (1991) Olduvai Gorge (The skulls, endocasts and teeth of Homo habilis, Vol. 4). *Cambridge: Cambridge University Press*.
122. Martinón-Torres M, Bermúdez de Castro, J.M., Gómez-Robles, A., Margvelashvili, A., Prado, L., Lordkipanidze, D. & Vekua, A. (2008) Dental remains from Dmanisi (Republic of Georgia): Morphological analysis and comparative study. *Journal of Human Evolution* 55(2):249-273.
123. Brown B & Walker A (1993) The dentition. *The Nariokotome Homo erectus Skeleton*:161-192.
124. Suwa G, Asfaw, B., Haile-Selassie, Y., White, T., Katoh, S., WoldeGabriel, G., Hart, W.K., Nakaya, H. & Beyene, Y. (2007) Early Pleistocene Homo erectus fossils from Konso, southern Ethiopia. *Anthropological Science* 115(2):133-151.
125. Weidenreich F (1937) The dentition of Sinanthropus pekinensis: a comparative odontography of the hominids. Paleaontologia Sinica new Series D. *Chinese Academy of Sciences, Beijing*.
126. Wu X & Poirier FE (1995) Human Evolution in China. *Oxford University Press, Oxford*.
127. Grine FE & Franzen JL (1994) Fossil hominid teeth from the Sangiran dome (Java, Indonesia). *Courier Forsch Senckenberg* 171:75-103.
128. Condemi S (2001) Les Néanderthaliens de La Chaise (Abri Bourgeois-Delaunay). *Éditions du Comité des Travaux Historiques et Scientifiques, Paris*.
129. Rougier H (2003) Étude descriptive et comparative de Biache-Saint-Vaast 1 (Biache-Saint-Vaast, Pas-de-Calais, France). *PhD Thesis, University of Bordeaux I*.
130. de Lumley H, & de Lumley, M.A (1982) L'Homo erectus et sa place de l'homme de Tautavel parmi les hominidés fossiles. *Nice, Palais des expositions*.
131. Suzuki H & Takai F (1970) The Amud Man and his Cave Site. *Academic Press of Japan, Tokyo.*
132. Tillier A (1979) La dentition de l'enfant moustérien Châteauneuf 2, découvert á l'Abri de Hauteroche (Charente). *L'Anthropologie* 83(3):417-438.
133. Wolpoff MH (1979) The Krapina dental remains. *American Journal of Physical Anthropology* 50(1):67-113.
134. Trinkaus E (1983) The Shanidar Neanderthals. *Academic Press, New York*.
135. McCown T & Keith A (1939) The Stone Age of Mount Carmel II: The human remains from the Levalloiso-Mousterien. *Oxford: Clarendon Press*.
136. Vandermeersch B (1981) Les hommes fossiles de Qafzeh (Israel). *Paris: Editions du CNRS* 319.
137. Hublin J-J, Verna, C., Bailey, S., Smith, T., Olejniczak, A., Sbihi-Alaoui, F. Z., & Zouak, M. (2012) Dental evidence from the Aterian human populations of Morocco. In J.- J. Hublin & S. McPherron (Eds.). *Vertebrate Paleobiology and Paleaanthropology: Modern origins - A North African perspective. Dordrecht: Springer*:189-204.
138. Sladek V, Trinkaus, E., HIllson, S.W. & Holliday, T.W. (2000) The People of the Pavlovian: Skeletal Catalogue and Osteometrics of the Gravettian Fossil Hominids from Dolní Vestonice and Pavlov. *Inst. of Archaeology, Academy of Sciences of the Czech Republic, Brno.*
139. Trinkaus E (2010) Denisova Cave, Peştera cu Oase, and Human Divergence in the Late Pleistocene. *PaleoAnthropology* 2010:196-200.
140. Sankararaman S, Patterson N, Li H, Paabo S, & Reich D (2012) The date of interbreeding between Neandertals and modern humans. *PLoS Genet* 8(10):e1002947.

141. Hofreiter M*, et al.* (2004) Lack of phylogeography in European mammals before the last glaciation. *Proc Natl Acad Sci U S A* 101(35):12963-12968.
142. Mednikova MB (2013) AN ARCHAIC HUMAN ULNA FROM CHAGYRSKAYA CAVE, ALTAI: MORPHOLOGY AND TAXONOMY. *ARCHAEOLOGY, ETHNOLOGY & ANTHROPOLOGY OF EURASIA* 41(1):66-77.
143. Rudaya N, Vasilyev S, Viola B, Talamo S, & Markin S (in press) Paleoenvorinments during the period of the Neanderthals settlement in Chagyrskaya cave (Altai mountains, Russia). *Paleogeography, Paleoclimatology, Paleaoecology*.
144. Dalen L*, et al.* (2012) Partial genetic turnover in neandertals: continuity in the East and population replacement in the West. *Mol Biol Evol* 29(8):1893-1897.
145. Castellano S*, et al.* (2014) Patterns of coding variation in the complete exomes of three Neandertals. *Proc Natl Acad Sci U S A* 111(18):6666-6671.
146. Mellars P & French JC (2011) Tenfold population increase in Western Europe at the Neandertal-to-modern human transition. *Science* 333(6042):623-627.
147. Dogandzic T & McPherron SP (2013) Demography and the demise of Neandertals: a comment on 'Tenfold population increase in Western Europe at the Neandertal-to-modern human transition'. *J Hum Evol* 64(4):311-313.
148. Gargett R (1999) Middle Palaeolithic burial is not a dead issue: the view from Qafzeh, Saint-Césaire, Kebara, Amud, and Dederiyeh. *Journal of Human Evolution* 37:27-90.
149. Belfer-Cohen A & Hovers E (1992) In the Eye of the Beholder: Mousterian and Natufian Burials in the Levant. *Curr Anthropol* 33(4):463-471.
150. Barry J (1987) Larger Carnivores (Canidae, Hyaenidae, Feidae) from Laetoli. *In: MD Leakey and JM Harris (eds). Laetoli: A Pliocene site in Tanzania. Oxford Clarendon Press*:235-258.
151. Turner A (1990) The evolution of the guild fo larger terrestrial carnivores during the Pliopleistocene in Africa. *Geobios* 23:349-368.
152. de Vos J, Leinders J, & Hussain S (1987) A historical review of the Siwalik Hyaenidae (Mammalia, Carnivora) and description of two new finds from the Upper Siwalik of Pakistan. *Palaentology, Proc B* 90:333-369.
153. Garcia N & Arsuaga JL (1999) Carnivores from the Early Pleistocene hominid-bearing Trinchera Dolina 6 (Sierra de Atapuerca, Spain). *J Hum Evol* 37(3-4):415-430.

Acknowledgements

I would like to thank my husband, Spencer, who never really understood what I did, but nodded his head and pretended to listen very loyally. Without him I would not be at this point in my life. We have both had tremendously stressful lives the past few years, not helped by the addition of our daughter, but we stuck together through all the hard times. Nothing can make me forget the craziness of our lives more than spending a few hours with him and Emma.

I would like to thank my sister, Johanna, who has always been my crutch in life. Without her I would never have come this far. Regardless of what she is doing, she listened to me, helped me deal with hundreds of hurdles and life crises, and has always been at my side. Throughout my life she has always been My Person. I cannot thank her enough.

I would like to thank my mother and father, Sabine and Haywood, who gave me such an amazing childhood. They nurtured my interest for science despite the fact that neither of them are science-inclined. They encouraged my various science hobbies and fascinations, from gorillas to blowing up chemicals with my chemistry set. They stood by side and cheered me on, they prodded me when I needed it, but also knew when to stand aside and let me make own mistakes.

I would like to thank my mentor, Svante Pääbo. He gave the amazing opportunity to pursue my PhD in his group. He expected me to do well and was therefore tough on me at times. This has given me the opportunity to grow as a person and a scientist and has been an invaluable experience. I would like to especially thank him for being so incredibly understanding towards me in regards to having a child during my PhD.

I would like to thank Aida and Janet for being there in times of crises, and always having an open door and listening while I poured my heart out. I would also like to thank Janet, Kay and Matthias for helping me with this thesis. A big thanks to the Ancient DNA group, both past and present members, especially Qiaomei, without whom I would have been totally lost those first few years.

I would like to thank my two favorite office mates, Tomi and Felix. You two have had to listen to my inane babbling for five years now, and I am sure you are happy to finally be rid of me. Our impromptu discussions have always been the best!

And of course I would like to thank Shona and Ana. Our weekly Friday night drinks, our much needed coffee breaks, our various forays into the Leipzig nightlife, they were all so special.

Curriculum Vitae

Susanna Katherine Sawyer
susanna_sawyer@eva.mpg.de
Karl-Liebknecht-Str. 63, 04275 Leipzig; Tel: +4917680216758

Academic Career

June 2010 -present	**PhD**, Department of Genetics, Max Planck Institute for Evolutionary Anthropology, Germany Supervisor: Svante Pääbo Dissertation focus: Nuclear and Mitochondrial DNA relationships within Denisovans and Neandertals of the Altai Mountains Relevant skills: Ancient DNA lab work (extraction, library preparation using Illumina platform, targeted capture, sequencing), genome analysis (mtDNA, low-coverage nuclear genomes, contamination assesment), basic population genetics, basic programing skills (perl, awk, bash, R)
August 2008 – June 2010	**M.S. in Evolutionary Biology** at Uppsala Universitet, Uppsala, Sweden Masters Project: "*Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA*" done at the Max Planck Institute for Evolutionary Anthropology September 2009-June 2010
August 2004 – May 2008	**Bachelors degree** at North Carolina State University, Raleigh, NC, USA B.S. in Zoology; GPA: 3.953/4.33 (total), 4.00/4.33 (major)
January 2005 – June 2008	**Undergraduate research on behavior of zebrafish** with Dr. John Godwin, NCSU Relevant skills: use of Open Field tests, developed own behavioral tests, Microarrays, Cortisol extractions and measurements
March 2004 – June 2004	**Internship**: Max Delbrück Center for Molecular Medicine, Berlin, Germany Relevent skills: Immunohistology of cells, Maintenance of small lab animals, Measurement of tumor growth, Mouse xenotransplantation

Publications

Research Papers

Sawyer S, Viola B, Shunkov M, Derevianko A, Meyer M, Prüfer K, Pääbo S. A Neandertal from Denisova Cave with ancient spotted hyena contamination. *In preperation.*

Sawyer S*, Renaud G* , Viola B, Gansauge MT, Shunkov M, Derevianko A, Prüfer K, Kelso J, Pääbo S (2015) Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *PNAS.*
* These authors contributed equally

Benazzi S, Slon V, Talamo S, Negrino F, Peresani M, Bailey SE, **Sawyer S**, et al. (2015) The makers of the Protoaurignacian and implications for Neandertal extinction. *Science.*

Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, **Sawyer S** et al. (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature.*

Castellano S, Parra G, Sanchez-Quinto FA, Racimo F, Kuhlwilm M, Kircher M, **Sawyer S** et al. (2014) Patterns of coding variation in the complete exomes of three Neandertals. *Proc Natl Acad Sci U S A.*

Lazaridis I, [18 people] **Sawyer S** et al. (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature.*

S. T. D. Consortium (2014) Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature.*

Thalmann O, Shapiro B, Cui P, Schuenemann VJ, **Sawyer S**, et al. (2013)., Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science*

Guschanski K, Krause J, **Sawyer S**, Valente LM, Bailey S, et al. (2013) Next-generation museomics disentangles one of the largest primate radiations. *Systematic biology.*

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

Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S (2012) Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One.*

Conferences

Sawyer S, Renaud G , Viola B, Gansauge MT, Shunkov M, Derevianko A, Prüfer K, Kelso J, Pääbo S. An Analysis of Three Denisovan Individuals Using Nuclear and Mitochondrial DNA. Abstract for **Poster**. The Biology of Genomes at CSHL, Cold Spring Harbor, NY, USA (2013)

Sawyer S, Viola B, Shunkov M, Derevianko A, Pääbo S. Neandertal and Denisovan Genomes from the Altai. Abstract for **Plenary Talk**. European Society for the study of Human Evolution, Bordeaux, France (2012)

Sawyer S, Viola B, Gansauge MT, Shunkov M, Derevianko A, Pääbo S. The complete mitochondrial genome of third individual from Denisova Cave. Abstract for **Poster**. Society for Molecular Biology and Evolution, Dublin, Ireland (2012), received a

Sawyer S. Denisovans, Neandertals and Modern Humans in the Altai. Abstract for **Talk**. International Symposium in Denisova Cave, Altai Mountains, Russia (2011)

Awards and Scholarships

2012	Graduate Student Poster Prize, Society for Molecular Biology and Evolution
2004-2008	Park Scholarship: merrit based full scholarship for undergraduate studies at North Carolina State University, $40,000
2006-2008	Howard Hughes Medical Institute Scholar, two year research grant, $9000
2004-2008	North Carolina State University Honors Program

Public Understanding of Science

2013	Documentary: *Neandethals Decoded* (Nova)
2012	Documentary: *Sex in the Stone Age* (National Geographic)

Selbständigkeitserklärung

Ich versichere, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Ich reiche sie erstmals als Prüfungsleistung ein. Mir ist bekannt, dass ein Betrugsversuch mit der Note "nicht ausreichend" (5,0) geahndet wird und im Wiederholungsfall zum Ausschluss von der Erbringung weiterer Prüfungsleistungen führen kann.

Name: Sawyer

Vorname: Susanna

Leipzig

14.12.2015